FIRST: FINDING INTERESTING STORIES ABOUT STUDENTS AN INTERACTIVE NARRATIVE APPROACH TO EXPLAINABLE LEARNING ANALYTICS

by

Ahmad Al-Doulat

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Software and Information Systems

Charlotte

2021

Approved by:

____

Dr. Mary Lou Maher

____

Dr. Bojan Cukic

____

Dr. Albert Park

____

Dr. Harini Ramaprasad

ABSTRACT

AHMAD AL-DOULAT. FIRST: Finding Interesting stoRies about STudents An Interactive Narrative Approach to Explainable Learning Analytics . (Under the direction of DR. MARY LOU MAHER)

Learning Analytics (LA) has had a growing interest by academics, researchers, and administrators motivated by the use of data to identify and intervene with students at risk of underperformance or discontinuation. Typically, faculty leadership and advisors use data sources hosted on institutional databases to advise their students for better performance in their academic life. Although academic advising has been critical for the learning process and the success of students, it is one of the most overlooked aspects of academic support systems. Most LA systems provide technical support to academic advisors with descriptive statistics and aggregate analytics about students' groups. Therefore, one of the demanding tasks in academic support systems is facilitating the advisors' sensemaking of students at the individual level. This enables them to make rational, informed decisions and advise their students. To facilitate the advisors' sensemaking of individual students, large volumes of student data need to be presented effectively and efficiently.

Effective presentation of data and analytic results for sensemaking has been a major issue when dealing with large volumes of data in LA. Typically, the students' data is presented in dashboard interfaces using various kinds of visualizations like scientific charts and graphs. From a human-centered computing perspective, the user's interpretation of such visualizations is a critical challenge to design for, with empirical evidence already showing that 'usable' visualizations are not necessarily effective and efficient from a learning perspective. Since an advisor's interpretation of the visualized data is fundamentally the construction of a narrative about student progress, this dissertation draws on the growing body of work in LA sensemaking, data storytelling, creative storytelling, and explainable artificial intelligence as the inspiration for the development of FIRST, Finding Interesting stoRies about STudents, that supports advisors in understanding the context of each student when making recommendations in an advising session. FIRST is an intelligible interactive interface built to promote the advisors' sensemaking of students' data at the individual level. It combines interactive storytelling and aggregate analytics of student data. It presents the student's data through natural language stories that are automatically generated and updated in coordination with the results of the aggregate analytics. In contrast to many LA systems designed to support student awareness of their performance or support teachers in understanding the students' performance in their courses, FIRST is designed to support advisors and higher education leadership in making sense of students' success and risk in their degree programs. The approach to interactive sensemaking has five main stages: (i) Student temporal data Model, (ii) Domain experts' questions and queries, (iii) Student data reasoning, (iv) Student storytelling model, and (v) Domain experts' reflection. The student storytelling stage is the main component of the sensemaking model and it composes four tasks: (i) Data sources, (ii) Story synthesis, (iii) Story analysis, and (iv) User interaction.

The contributions of this dissertation are: (i) A novel student storytelling model to facilitate the sensemaking of complex student data, (ii) An anomaly detection model to enrich student stories with interesting, yet, insightful information and (iii) An explainable interactive LA model to inspire advisors' trust and confidence with the student stories. This study reports on four ethnographic studies to show the potential of the proposed LA sensemaking model and how it affects the advisor's sensemaking of students data. The user studies considered for this dissertation were focus group discussions, in-depth interviews, and diary study. These studies investigate if FIRST can improve and facilitate the advisor's sensemaking of students' success or risk by presenting individual student's data as a complete and comprehensive story.

DEDICATION

In memory of my Dad, Saleh, this achievement is yours. Your support, encouragement, and constant love have always sustained me throughout my life.

My world is my Mom, Shams, I cannot express your love and dedication. I salute you for the selfless love, care, pain, and sacrifice you did to shape my life. I love you and I appreciate everything that you have done for me.

I owe thanks to a very special person, my beloved wife, Sarah. Your continued and unconditional love, support, and understanding during my pursuit of my Ph.D. degree have made the completion of this dissertation possible. You were always around at times I thought that it is impossible to continue, you helped me to keep things in perspective. I greatly value your contribution and deeply appreciate your belief in me.

To my family, your dedication has led me to where I am now. Thanks for your love and support.

ACKNOWLEDGEMENTS

It is a genuine pleasure to express my deep sense of thanks and gratitude to my advisor and mentor Dr. Mary Lou Maher (Director of Center for Education Innovation, College of Computing and Informatics at UNCC). Her dedication and keen interest above all the overwhelming attitude to help me had been solely and mainly responsible for completing my work. Her timely advice meticulous scrutiny, scholary advice, and scientific approach have helped me to a very great extent to accomplish this dissertation.

I would like to express my sincere gratitude to my dissertation committee: Dr. Bojan Cukic, Dr. Harini Ramaprasad, and Dr. Albert Park, whose wisdom and experience were greatly appreciated. Thanks for your insightful and instructive guidance that made this research so much better.

I thank profusely all the professional and faculty advisors in the College of Computing and Informatics at the University of North Carolina at Charlotte for their kind help and co-operation throughout my user studies. Their insightful comments and suggestions have made this research so much better.

I am grateful for the assistance of the University of North Carolina at Charlotte for funding my Ph.D. study through the Graduate Assistant Support Plan. I am also grateful for the National Science Foundation (NSF) assistance for funding my research. This work was partially funded by NSF Award #1820862: EAGER: An Interactive Learning Analytics Framework based on a Student Sequence Model for understanding students, retention, and time to graduation.

Finally, I would like to extend my sincere thanks to my friends, lab mates, colleagues, and research team in the HCI Lab, who have provided friendship and support, and with whom I have shared laughter, frustration, and companionship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

As artificial intelligence in education becomes increasingly prominent, there is a growing need to consider augmented intelligence. This is the idea that artificial intelligence can and should be used to enhance human intelligence and abilities rather than attempt to replace them. The 2016 National Artificial Intelligence Research and Development Strategic Plan stated that "the walls between humans and AI systems are slowly beginning to erode, with AI systems augmenting and enhancing human capabilities. Fundamental research is needed to develop effective methods for human-AI interaction and collaboration" [6]. Popenici and Kerr further emphasize the importance of recognizing education as a "human-centered endeavor" and the idea that "solely rely[ing] on technology is a dangerous path, and... that humans should identify problems, critique, identify risks, and ask important questions..." [7]. Therefore, we should take on a human-centered approach in the era of AI. Human-centered AI [8] is a viewpoint discussing that "AI systems and algorithms must be designed with an awareness that they are part of a larger system involving humans" [9]. This includes the view that AI research should not just be technological, but humanistic and ethical as well [10]. One aspect of human-centered AI is to create systems that help humans understand the system itself [9]. Therefore, the goal is not simply to provide results through a black-box model. The focus is to help users understand those results and how those results are derived.

Sensemaking, as defined by Weick [11], is "making sense of what is happening". It is also defined by Dixon [12] as "the human ability to retrospectively find patterns in the continual flow of events that individuals encounter repeatedly in order to give those events meaning". The patterns that humans construct are mainly influenced by their knowledge base and experience. Additionally, in order for people to make sense of things, they need to engage in a continual revision of their understanding based on sequence events and based on the interpretation of others. Therefore, "making sense is not finding the "right" or "correct" answer, but finding a pattern that gives meaning to people that makes what has occurred sensible" [12]. Klein et al. [13] defined sensemaking as "a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively".

1.1 Motivation

Faculty leadership and advisors use data sources hosted on institutional databases to advise their students for better performance in their academic life [14, 15, 16]. Therefore, advisor awareness of the students' success or risk is a demanding task because of the number of advisees an advisor needs to deal with and the multitude of each advisor's data. Student advising is a critical challenge for advisors since they need to be able to make informed decisions and provide better advice to their students. Therefore, advisors need to constantly be aware of the information about their advisees and the progress they are making. In recent years, the research and development of LA systems that support professors and advisors have been rich and varied. These systems take the form of dashboards that capture, analyze and visualize student data and progress throughout their enrollment. These dashboards aim to deliver insights and actionable knowledge to advisors to enable an immediate translation of student information into intervention for students who might be at risk. As Clow [17] denotes, "the loop of the LA cycle is only closed effectively when the LA is used to instigate an intervention of some kind that influences students' learning". Therefore, the cycle of LA dashboards and interfaces would be closed when advisors have the ability to make sense of the student data and, subsequently, have the ability to have a positive impact on student outcomes and progress. The premise is that LA systems provide discovery of "actionable knowledge" [18] that enables immediate translation of student information into concrete intervention and support for students who might be at risk. Nevertheless, between the presentation of metrics and analytics about the student progress and the actual change in student outcomes lies a process of "sense-making" from the advisor. Advisors need to be able to make sense of these metrics and analytics and, as a result, be able to make informed decisions and advise their students [19, 1].

Effective presentation of analytics results for sensemaking and decision making has been a major issue when dealing with large volumes of data in LA. Typically, visualizations using tables, charts, and graphs are the most widely used. Although data visualization helps experienced users to explore large datasets in a relatively short time, and identify trends based on what they see, it is not always intuitive or information-rich for most users to digest what is going on [20]. Results from analytics models typically treat students at an aggregate level. Although useful, aggregate analytics does not necessarily help advisors with their most important task of interacting with and understanding students on an individual level.

Several research studies show the lack of work and research that contributes to advisors' sensemaking of student data in LA. Sergis et al. [21] stated that very few studies focus on how advisors and professors make sense of student data and how they translate student information into intervention. This raises the question of how advisors use LA. LA is an additional source of information about students' performances and their academic progression. Information about students is a magnitude of stimuli that requires careful attention from advisors. Hence student data needs to be presented in an effective way that enables advisors to effectively and efficiently make sense of their student situation.

Storytelling is an integral part of our communication skills, as we always tie facts together into stories in a memorable way [22]. This dissertation suggests that the use of concepts from storytelling is the next step in LA research. Presenting students' data in natural language stories can enable domain experts to understand complicated computational models. These stories can be produced automatically using Natural Language Generation (NLG) techniques. This study aims to provide advisors with stories that help them understand students beyond raw data and numerical risk scores.

Explainable Artificial Intelligence (XAI) is a research field that aims to make the outcome of an Artificial Intelligence (AI) system more understandable and interpretable by humans, either through introspection or through a generated explanation [23]. In other words, XAI is an AI system that can describe its purpose, behavior, and decision-making process in a way that can be easily understood by an average person. Although the term XAI is relatively new as it was first coined in 2004 by Van Lent et al. [24], to show their proposed approach capability in explaining AI-controlled entities behaviors in simulation game applications, the problem of explainability first appeared in the context of rule-based expert systems in the mid-1970s [25]. In many studies, users have expressed high desires for an explanation feature of decision-making systems that explains the system's decisions to the user [26, 27]. The explanation is directly related to the users' acceptance and satisfaction of these decisions. To facilitate the user's sensemaking, the system should be able to describe its purpose, behavior, and decision-making process in a way that can be easily understood by an average person. This means that the system does not only produce a decision for the user but also, explains why a decision is made. Besides, the system should allow the user to be part of the decision-making process through interaction.

1.2 Research Focus

This dissertation explores sensemaking in LA as an example of human-centered AI and presents how we address this challenge for advisors who are presented with large amounts of data and analytics about their students. LA is an interdisciplinary field that emerged to make sense of unprecedented amounts of data collected by the extensive use of technology in education. LA brings together researchers and practitioners from two main fields: data mining and education [28]. Effective presentation of analytical results for decision-making has been a major issue when dealing with large volumes of data in LA [29]. Furthermore, many systems for early alerts on student performance provide results without providing necessary explanations as to how the system derived those results. Results from an early alert system that are inconsistent with the advisor's expectation can be easily discounted or mistrusted if there is no explanation or justification on how the system arrived at those results [30]. Human sensemaking relies on developing representations of knowledge to help serve a task, such as decision-making, and on the design of AI approaches to better aid these tasks.

This dissertation introduces an interactive system called FIRST (Finding Interesting stoRies about STudents) designed to help advisors better understand student success and risk. This interactive system aims to promote the advisors' sensemaking of students' data at the individual level. It presents the student's data through stories that are automatically generated and updated in coordination with the results of the aggregate analytics. In contrast to many LA systems designed to support student awareness of their performance or to support teachers in understanding the students' performance in their courses, FIRST is designed to support advisors and higher education leadership in making sense of students' success and risk in their degree programs. The approach to interactive sensemaking has five main parts: (i) Student temporal data Model, (ii) Domain experts' questions and queries, (iii) Student data reasoning, (iv) Student storytelling model, and (v) Domain experts' reflection.

1.3 Thesis Statement and Research Questions

An interactive narrative storytelling model for learning analytics can improve faculty leadership and advisors' sensemaking of student success or risk and reveal hidden insights not apparent in visual analytics alone.

When designing decision support systems for faculty leadership and advisors, considering the factors that improve their sensemaking and decision-making of complex and heterogeneous student data should precede the design of accurate analytical results. This dissertation claims that interactive narrative storytelling about student data is more engaging, effective, and easier to understand the diverse and heterogeneous student data compared to other presentation styles. Storytelling in LA helps reveal hidden insights not apparent in visual analytics alone. Another storytelling aspect that makes it so effective in presenting complex student data is that it works for all types of decision-makers regardless of their level of expertise.

Based on this thesis statement, this dissertation addresses the following research questions:

- **RQ1:** What are the benefits and features of storytelling when compared to visual analytics?
  - **RQ1.1:** Do the student stories provide an effective way of presenting complex and heterogeneous student data to domain experts?
  - **RQ1.2:** Do the student stories provide an easier, more engaging and more understandable way for non-experts to make sense of complex and heterogeneous student data?
  - **RQ1.3:** Which presentation style do domain experts prefer to make sense of complex and heterogeneous student data?
  - **RQ1.4:** What are the effects of student storytelling on advising?
- **RQ2:** How do the student stories help domain experts in discovering actionable knowledge about their students?
  - **RQ2.1:** Do the student stories help domain experts discover actionable knowledge?
  - **RQ2.2:** What insights do domain experts learn from student stories?
- **RQ3:** Which story building blocks (contents and structures) are meaningful for

the domain experts?

- **RQ3.1:** Which features from the student data model do domain experts find helpful throughout their advising session?
- **RQ3.2:** What content of the student stories are more meaningful for the domain experts?
- **RQ3.3:** What story structures are more meaningful for domain experts?

- **RQ4:** How do explainability and interpretability support the domain experts' trustworthiness and satisfaction with the students' stories?
  - **RQ4.1:** What are the benefits of selecting student features to be included in the student stories?
  - **RQ4.2:** What role(s) do story explanations play in supporting domain experts' trustworthiness and satisfaction of students' stories?

1.4 Methods and Evaluation

This dissertation adopts a mixed-method approach to explore and answer the aforementioned research questions. This approach comprises three main activities: (i) Reviewing the literature to develop the research framework and highlighting the importance of creative and data storytelling in improving the decision-making and sensemaking of diverse, complex, and heterogeneous data. (ii) Designing and developing a storytelling model in the LA domain to improve domain experts' sensemaking of diverse, complex, and heterogeneous students' data. (iii) Analyzing domain experts' interaction with the student data and the storytelling model. This dissertation presents a series of evaluations to investigate domain experts' behavior and interaction with the students' stories to make sense of the student data. The methodologies considered for this dissertation were focus group discussions, in-depth one-on-one interviews, and diary study- in-situ and snippet technique.

1.5 Contributions

The contributions of this study are as follows:

- Development of a novel student storytelling model for academic advisors (Chapter 4).
- Identification of student story's contents and structures that are meaningful for academic advisors (Chapter 7).
- Development of an anomaly detection model to compare the performance of a student in the context of other students. This model aims to find interesting information and unexpected patterns in the student data to be presented in the student stories (Chapter 5).
- Development of an explainable interactive LA storytelling system for academic advisors. This system aims to increase the trustworthiness and effectiveness of the student stories (Chapter 6).

First, a novel student storytelling model in the LA domain is proposed. Rather than presenting student data and analytics results using visualizations, this study proposes to present the student data using natural language stories. These stories are automatically generated using NLG techniques and updated in coordination with the results of the aggregate analytics. Unlike other LA studies that tend to support student awareness of their performance or to support teachers in understanding the students' performance in their courses, this study aims to support advisors and higher education leadership in making sense of students' success and risk in their degree programs.

Second, this study identifies the key student story's contents and structures that are meaningful for domain experts. Ethnographic studies are used for this identification. Other data storytelling studies tend to generate summaries where the input data to these systems are the analytical results or numeric predictions. This study proposes a storytelling model that is capable of generating multi-paragraph stories from different sources of students' data that are complex, temporal, and heterogeneous. In this model, the identification of the story content is based on three different sources of data: (i) Students' temporal features, (ii) Aggregate analytics results, and (iii) User-selected features. This latter one gives the advisor the ability to be part of the story generation process. The identification of the story structure is done by employing two fundamental approaches from the creative storytelling field: the story representation structure proposed in [4] to identify the story elements and the Freytag pyramid [5] to decide the plot of the students' stories. The story structure is evaluated using an ethnographic study, where advisors are presented with three alternatives of story structures and asked questions regarding the circumstances they would prefer one structure over another.

Third, an anomaly detection model is proposed to find interesting and unexpected information in the student data. This information is presented in the student stories. The anomaly detection model aims to detect if there are extreme values (anomalies) in the student data compared to other students. This study proposes two models of anomaly detection- Personal Anomaly Detection (PAD) and Collective Anomaly Detection (CAD). The PAD model aims to detect if an individual student's data instance can be considered anomalous compared to the rest of the data (e.g. a student's GPA significantly decreased from one semester to another compared to other students in CCI). The CAD model aims to detect if a collection of student data instances is anomalous compared to other students in CCI, but not individual values. For instance, if a student follows a non-typical pattern for the number of credits passed each semester.

Last, an explainable interactive LA storytelling system for domain experts is proposed. To the best of our knowledge, this is the first study to make an explainable and interpretable storytelling system. Explainability and interpretability were used in the domain of artificial intelligence systems to expose the reasoning and data behind a machine learning model prediction. In this study, explainability and interpretability are used to show the users how the stories have been generated and how the contents of the story are selected from the student data model. This system aims to increase the advisors' trustworthiness and satisfaction with the generated stories.

1.6 Thesis Organization

The dissertation is organized as follows: Chapter 1 presents the purpose of the study, the research questions to be investigated, and the significance of the study. Chapter 2 reviews related work from the fields that are germane to this study including sensemaking in LA, data storytelling, creative storytelling, and explainable artificial intelligence. This chapter provides an overview of each field. Then, it presents the systems and models that are designed and developed in the literature. Afterward, it presents how those systems are evaluated for each field. Chapter 3 presents the proposed sensemaking model for the LA domain. Then, in Chapter 4, a novel student storytelling model in the LA domain is presented. Chapter 5 describes the analytical model used to find interesting and useful information from student data. Chapter 6 presents the approach used to make the student storytelling model interpretable and explainable for the domain experts. Chapter 7 presents four user studies that are conducted to evaluate the proposed storytelling model and to answer the research questions of this dissertation. Finally, Chapter 8 contains the discussion, limitations in the current study, and recommendations for future work.

CHAPTER 2: RELATED WORKS

2.1 Overview

Four fields are germane to this dissertation: sensemaking in LA, creative storytelling, data storytelling, and Explainable AI. This chapter presents a description of each field. Then, it presents the existing research and models for each field. Finally, it presents how these research and models are evaluated in the literature.

2.2 Sensemaking in Learning Analytics

Sensemaking as defined by Klein [13] "is a motivated, continuous effort to understand connections in order to anticipate their trajectories and act effectively". We believe that facilitating faculty leadership and advisors' sensemaking in LA helps them to make rational, informed decisions and advisement to their students. This section starts by presenting the body of work that has been done in the field of sensemaking in LA. Then, it presents the evaluation methods implemented and conducted in the literature to assess various sensemaking models.

2.2.1 Learning Analytics Sensemaking Models

Van et al. [31] stated that "sensemaking is a core component of LA dashboard interventions, as the purpose of these tools is to provide users with the ability to become aware of, reflect upon, and make data-based decisions". Echeverria et al. [29] proposed a learning design-driven data storytelling approach where they support user sensemaking by directing the user's attention to the critical features of the students' data using visualizations with data storytelling components. Their user study suggests that adding storytelling elements to the LA dashboards has the potential to help users make sense of the critical features of students' data with less effort. CALMSystem [32] is another example of a LA system that supports sensemaking, awareness, and reflection. It was developed on top of an intelligent tutoring system to give a learner insight into the learner model. Klein et al. [1] proposed a model of student sensemaking of LA dashboards to show how data and visualization inform user sensemaking and action. Verbert et al. [1] introduced a LA system for learners and teachers visualizing learning traces with four distinguished stages for the process model (as presented in Figure 2.1) - (i) awareness is only concerned with the students' data presented using various visualizations, (ii) reflection focuses on usefulness and relevance of the queries by the users, (iii) sensemaking is concerned with users responses in the reflection process and the creation of new insights, and (iv) impact is concerned with the induction of new meaning or changing behavior by the users. Additionally, researchers made contributions to better prediction and sensemaking of student progress trajectories. Learning Management Systems (LMSs) storing students' temporal data have been leveraged in various works to analyze students' progression throughout their whole program [15, 16, 33, 34, 14] and within a course level [35, 36, 37].

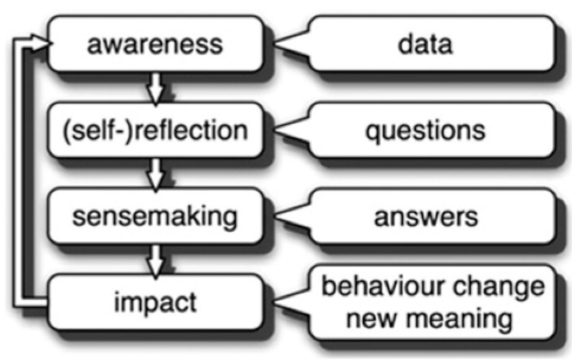


Figure 2.1: Learning Analytics process model by Verbert et al. [1]

Effective presentation of analytics results for sensemaking is a major issue when dealing with large volumes of data such as in LA. Typically, visualizations using tables, charts, and graphs are the most widely used (See figures 2.2 and 2.3). Data visualization in LA helps users to understand large datasets in a relatively short time by driving their attention to some important features of the students' data. However, these visualizations require some level of expertise to be interpreted correctly and on time based on their communicative power [29]. Results from analytics models typically treat students at an aggregate level. Although useful, aggregate analytics does not necessarily help advisors with their most important task of interacting with and understanding students on an individual level. This dissertation introduces a LA model-driven data storytelling approach called FIRST (Finding Interesting stoRies about STudents) for academic advisors using a temporal data model and storytelling techniques. This model aims to promote the advisors' sensemaking of students' data at the individual level. It presents the student's data through stories that are automatically generated and updated in coordination with the results of the aggregate analytics. Rather than helping advisors to make sense of the aggregate analytics of a group of students, this model helps advisors make sense of students at the individual level. Additionally, rather than driving the attention of users to some important features of the data, our model aims to help advisors with their most important task of interacting and understanding individual students as a complete and comprehensive story.

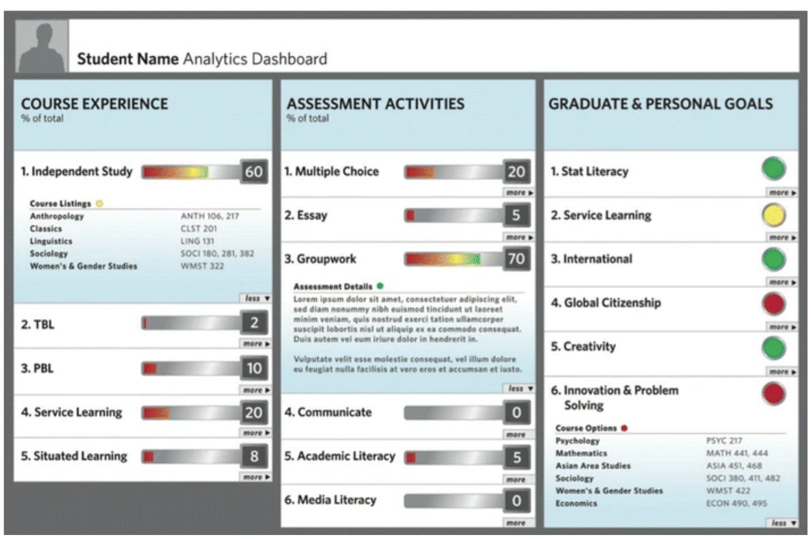


Figure 2.2: Visualizing students data in OLI dashboard [2]

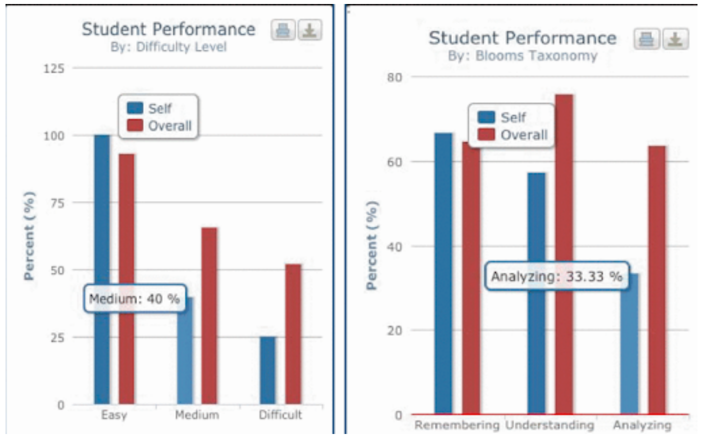


Figure 2.3: Visualizing students data in QMCU student's dashboard [3]

2.2.2 Evaluation of Learning Analytics Sensemaking Models

Evaluation of LA sensemaking is conducted for various goals including effectiveness, efficiency, usefulness, and usability of these applications and how these measures improve the users' sensemaking of the LA system. These applications have been evaluated in several ways in the literature. Verbert in [1] stated that measuring the LA systems' usability and usefulness is considered relatively easier than measuring the LAs systems' efficiency and effectiveness. For instance, in the form of learning impact, it is much harder to evaluate, as this requires longer-term and larger-scale evaluations. Several evaluation methods have been conducted in the literature. Some systems have been evaluated with teachers or learners, or both.

Course Signal [15] was evaluated to assess the effectiveness and potential impact of the system on the student performance over a three academic year period. The result of this evaluation indicates that using the Course Signal dashboard helps students to improve their grades and retention behavior. Other experiments are more limited and are often conducted in a controlled setting. To evaluate the effectiveness of CALM-system [32], a one-hour controlled experiment was conducted with thirteen students. The experiment was to evaluate the effectiveness of the system. The result of the experiment indicated that the CALMsystem helps students making sense of their performance, improving their self-assessment. A similar experiment was conducted to evaluate the effectiveness of Teacher ADVisor [38]. The results of this experiment indicated higher satisfaction with courses for students who used the dashboard. However, the results did not show any significant difference between students' grades who use the dashboard and students who do not use the dashboard. Another controlled group experiment was conducted to evaluate the effectiveness of CourseVis [39]. In their experiment, the authors measured the time required by teachers to answer questions about their students while using the system and without using the system. The results of this experiment indicated that teachers needed less time to understand their students' situations while using the system.

Similar evaluations were conducted with LOCO-Analyst [40], OLI Dashboard [2], TUT Circle Dashboard [41], Student Inspector [42], SAM [43], and StepUp! [44] to evaluate the usefulness of these systems. These evaluations include questions to teachers regarding their student success or risk.

2.3 Creative Storytelling

Text generation using computer programs has been an area of interest for many AI researchers. Although the first attempts in the 1960s and 70s were to generate stories and poems by computer [45, 46], the origin of computer-based storytelling models can be traced to TALE-SPIN and the story-grammar approach. TALE-SPIN [47] is a computer program that generates stories by first identifying goals for characters in the story and then recording their actions to reach these goals. It applies problem-solving techniques to generate stories. The problem-solving technique proposed in [47] became the model to follow for other storytelling models. The following section presents various story generation systems from the literature.

2.3.1 Creative Story Generation Systems

Automatic story generation is a part of a wider research area in AI named Computational Creativity (CC), which is the pursuit of creative behavior in machines [48]. A story generator algorithm refers to a computational procedure resulting in an artifact that can be considered a story [49].

Story generation models are called storytelling systems, which are computational systems used to tell stories. It is started with the development of story grammars, which is developed to create a theory of story understanding. Story grammars represent stories as linguistic objects which have a constituent structure and it can be represented by a grammar [50, 51]. This approach was first introduced in the context of story understanding and later was employed by some researchers to design automatic storytelling systems. For instance, GESTER [52] is a system that uses story grammar derived from medieval French epics to generate story outlines. GESTER was only able to generate stories that satisfy its grammar. This section reviews state-of-the-art story generation models and how these models represent the knowledge they need to create stories.

TALE-SPIN [53] is a planning solver that generates stories by narrating the steps performed for achieving the characters' goals. The stories were set in a forest. The input to the TALE-SPIN system is a collection of characters along with their associated goals. The system then generates a sequence of actions and events while resolving the characters' goals. Author [54] is also a planning solver that generates stories, but, unlike TALE-SPIN, it uses planning to achieve the author's goals instead of character goals.

Mexica [55] generated short stories about the early inhabitants of Mexico. It takes into account emotional links between the characters for driving and evaluating ongoing stories. Mexica's knowledge base included various types of structures for representing things like characters, actions, emotional links, and a literary base composed of previously generated stories.

Fabulist [56] is a framework for automatic story generation and presentation. It incorporates an author-centric technique with a representation of characters' intentionality and open-world planning for creating believable stories. Fabulist uses little prior knowledge built into the system. This feature allows Fabulist to generate novel stories that were not anticipated by the system's creator.

STellA (Story Telling Algorithm) [57] is a storytelling system that uses an exploration engine to generate a set of action simulations. These actions are represented as world states and used to form partial short stories. Then, the system selects the states that constitute a complete and coherent story.

PropperWryter [58, 59] is a storytelling system that creates a structure for a narrative for a single plotline. This structure is described in terms of an abstract description of events and actions that occur in a single plot. It uses Propp's generation rules [60] to generate Russian folktales. Propp identified a set of rules across a corpus of Russian folk tales in terms of the functions for each character in a story. Then, it selects the character functions in terms of story actions to produce the story's conceptual representation.

Another form of creative storytelling model is interactive storytelling; which is an interactive experience in which users participate in or affect the storytelling process through their choices and actions [61]. These storytelling systems aim to engage the users in a virtual environment so that they feel that they are essentially part of the unfolding story.

In addition to these single storytelling systems, Concepcion et al. [4] propose a common representation model that allows the free exchange of knowledge between different story generation systems. In their study, the authors identify the dimensions considered by the knowledge managed by storytelling systems. A set of dimensions was selected for identifying the common aspects of the representation of knowledge in storytelling systems as shown Figure 2.4. Their representation model is represented as a structure that consists of a specific set of concepts; in which each concept plays a particular role when generating a complete story. These concepts include space, plot, setting, location, happening, existents, and characters. Further details about these concepts are presented in Section 4.3.2.2.

This dissertation adopts the conceptual model driven by theoretical foundations in creative storytelling that aims to enhance the quality of the generated stories. Particularly, this dissertation adopts the hierarchical story representation structure proposed in [4] and maps the concepts into the students' data. The central intuition is to adopt models from the field of creative storytelling systems to generate compelling and engaging students' stories.

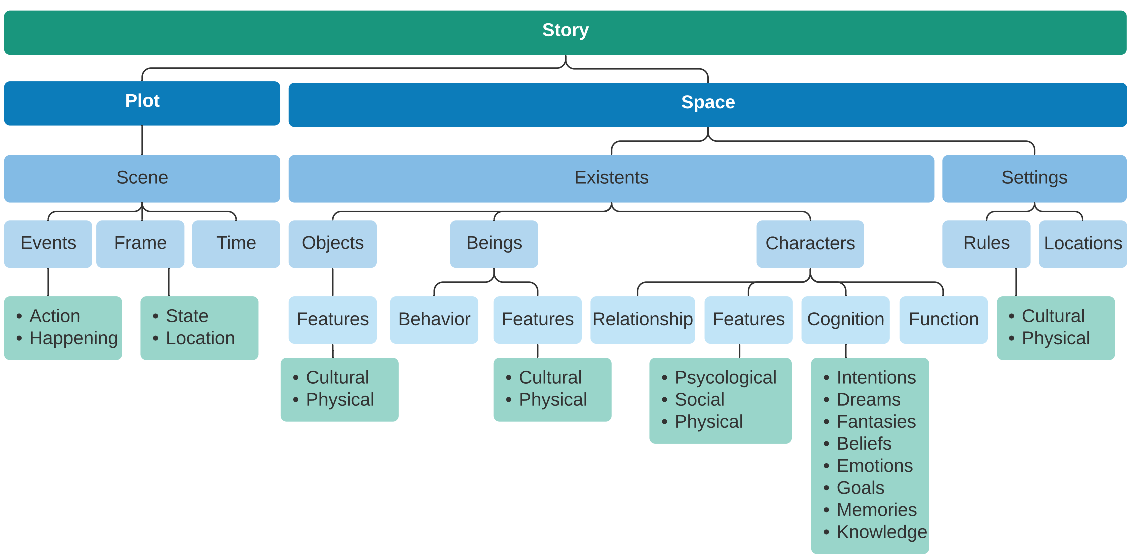


Figure 2.4: The structure of the common storytelling systems story representation [4]

2.3.2 Evaluation of Creative Storytelling Systems

Stories generated by creative storytellers are evaluated for different goals regarding the creativity of the storyteller in terms of (i) narrative flow and coherence, (ii) narrative structure, (iii) narrative content, and (iv) originality of the generated story [62].

Most creative storyteller systems in the literature evaluated their generated stories through user studies. These user studies ask participants to rate the generated stories on a Likert scale to assess the quality of the generated stories in terms of the evaluation goals mentioned above. For instance, a story generated by MINSTREL was evaluated through an Internet questionnaire. In this questionnaire, the participants were presented with stories generated using MINSTREL and asked to identify some characteristics about the hypothetical author of the stories. For example, the author's age and level of education. In addition, participants were asked to answer questions regarding the coherence and fluency of the narrative stories. MEXICA was also evaluated through a questionnaire. In this questionnaire, participants were asked to rate the coherence, fluency, and quality of four different stories generated using MEXICA on a 5-Likert scale. The results of their evaluation indicated that their system was able to generate short stories that are close to stories written by humans. Fabulist was also evaluated through a questionnaire, in which participants were asked to read stories generated by the system and make a goodness-of-answer assessment for question-answer pairs. The question-answer pair takes the form of a "why" question. It asks about the characters' actions in the stories. For example, "Why did Aladdin slay the dragon?", and the answer could be "Because King Jafar ordered Aladdin to get the magic lamp for him." Participants were asked to rate the goodness-of-answer on a 4-point scale.

Following the foundations used in the field of creative storytelling to evaluate the generated stories, this dissertation evaluates the generated stories in terms of the story narrative structure, coherence, and clarity of the text. The evaluation is based on different forms of user studies including a focus group study, interview study, and diary study. These evaluation methods are discussed in chapter 7.

2.4 Data Storytelling

There have been several research studies on summarizing or synthesizing structured data, ranging from summarizing statistical results [63, 64], stock market trends [65] and environmental data [66]. The underlying assumption behind these research efforts is that the presentation of tabular or numeric data using a natural language style story makes the data more understandable and memorable by human users [67]. The stories are capable of conveying essential information to users more naturally and familiarly.

2.4.1 Data Storytelling Systems

A wide range of applications has been deployed in the context of data to text (data storytelling) systems. Among the most successful applications to date is the weather forecasting systems which started in the mid-1990s. For instance, Forecast Generator (FoG) [68] is one of the earliest systems that attempted to automatically generate bilingual (English/French) textual weather forecasts. These textual forecasts are generated from data that is pre-processed and manipulated by human users through a graphical user interface called Forecast Production Assistant (FBA). The central intuition behind FoG is to reduce the routine tasks performed by human forecasters when writing these weather forecast reports. Another multilingual (English, French, German, Spanish, and Dutch) text generator is MULTIMETEO [69], which is an interactive tool that automatically generates textual weather forecast reports from structured data. It also provides a user interface that enables users to modify the style of the texts to be generated. The potential of weather forecasting systems has been demonstrated in SumTime system [70], which is a text generator that produces textual marine weather forecasts for oil platform applications. SumTime offers textual forecasts that can be tailored to specific user requirements, by allowing the forecasters to control the input data. An evaluation of the SumTime system shows that users preferred some of the automatically generated forecast texts over those produced by professional forecasters.

In the medical domain, NLG research has been used widely too. For instance, TOPAZ [71], a system that creates reports of blood cell and drug dosages data for lymphoma patients. TOPAZ uses a numerical model to detect differences between patient-specific values and population parameters. Then a temporal abstraction is done to group significant events into intervals and identifies a set of possible explanations. Then it uses a schema-based text generation system that converts the abstractions into a summary or report that is readable and understandable by clinicians. Another example is Suregen [72] which is a system that helps doctors in writing routine reports. Another example is the Narrative Engine [73] which is a system that helps doctors to create summaries and medical reports for the patients regarding their symptoms, lab tests, and prescriptions. Rather than creating textual reports for the medical staff, some NLG systems have been developed to produce textual reports for patients. For example, STOP [74], a system that generates personalized smoking-cessation letters for patients.

Several, other data storytelling systems have been developed to summarize small data sets, including summaries of statistical results [63, 64], air quality reports [66], and financial data [65]. One commonality among these systems is that they all tend to generate summaries in a well-defined domain where the input data to these systems are the analytical results or numeric predictions. For example, weather forecast systems only deal with numeric weather prediction data. Rather than relying only on the analytical results and the numeric predictions as shown in Figure 2.5(a), our proposed storytelling model is capable of generating multi-paragraph stories from different sources of students' data that are complex, temporal, and heterogeneous. In this model, the input to the storytelling algorithm consists of three different sources of information; temporal students' data, aggregate analytics results, and the user-selected features as shown in Figure 2.5(b).

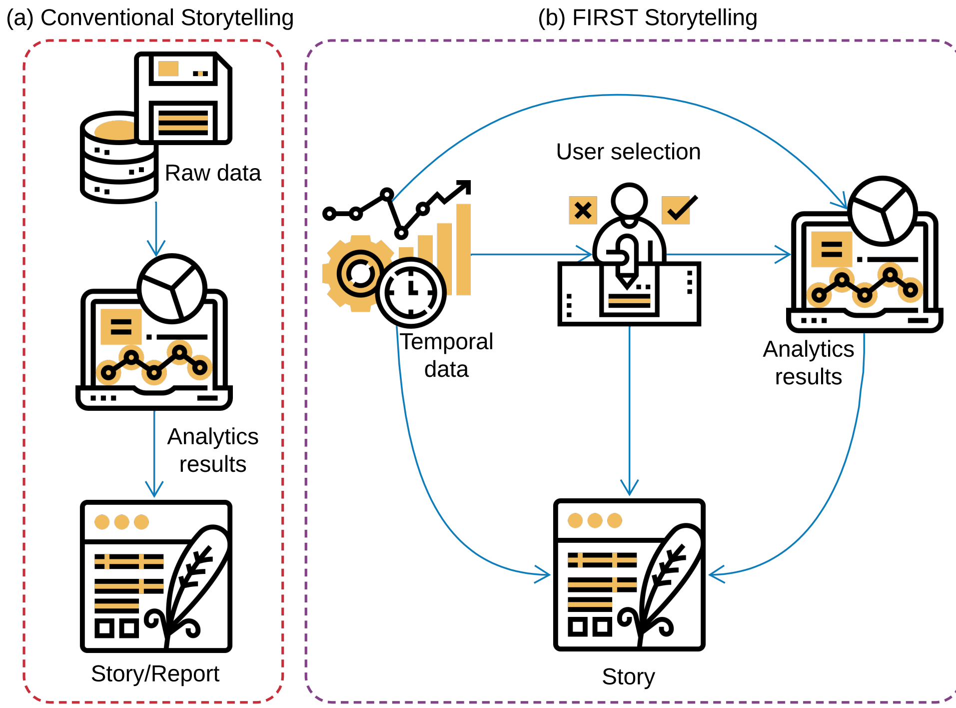


Figure 2.5: Data Storytelling Process: (a) Conventional Data Storytelling Process. (b): FIRST Storytelling Process

2.4.2 Evaluation of Data Storytelling Systems

Stories generated using data storytelling systems can be evaluated using intrinsic evaluation or extrinsic evaluation. The intrinsic evaluation assesses the generated story in terms of language quality, coherency, fluency, and fidelity of the generated story text. These intrinsic evaluations take two forms: Online evaluation and offline evaluation. Online evaluations are usually done in controlled group experiments. Participants in these evaluations are presented with stories generated using a storytelling system and asked to answer questions like "does the story text read naturally?", "Does the story text have clarity?" Offline evaluations usually rely on comparing the stories generated using a storytelling system with stories written by humans. This comparison is done using objective word-based metrics like BLEU (Bilingual evaluation understudy) [75], METEOR (Metric for Evaluation of Translation with Explicit ORdering) [76] and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [77].

The extrinsic evaluation assesses to what extent a storytelling system is capable of generating stories that support the purpose of the intended user. In other words, do the stories generated by the storytelling system help the user to make better decisions? These evaluations are usually done using online controlled group experiments. In these experiments, participants were asked to answer questions about the helpfulness, usefulness, and effectiveness of the generated stories, like "does the story help you make better decisions?". Some storytelling systems use computational methods to evaluate the stories generated by the generation model. For instance, SumTime [70] generates weather forecasts in three stages. After each stage of the generation process, a human forecaster edits some erroneous system predictions. Therefore, they evaluate the generation model by measuring the number of edits made by the forecaster after each stage. This means that the minimum number of problems that need to be fixed by humans the better the generation model performs.

This dissertation evaluates the generated stories using both intrinsic and extrinsic evaluations. The intrinsic evaluation assesses the linguistic correctness and fluency of the student story's text. The extrinsic evaluation assesses to what extent the generated story is well-equipped with insightful and interesting information that supports the advisors' sensemaking and decision-making when advising their students. These evaluation methods are discussed in 7.

2.5 Explainable Artificial Intelligence

Explainability refers to the algorithms that try to answer the question of "Why a decision is made by the system?". These algorithms do not only produce results or outputs for the users but also explain why such results are produced. Explanations of the system's decisions can serve multiple aims, such as exposing the reasoning and data behind a decision. Other aims of an explanation include increasing the users' trust and confidence in the system's decision, persuading the user to accept the decision, making it easier and faster for the user to find the most relevant information they want, and increasing user satisfaction. Explainable systems play an important role in enhancing the user experience [78]. However, it is still largely open to what a good explanation is and it is dependent on the general purpose of the system [78].

The HCI community often defines the interpretability of a system as its ability to make its decisions interpretable and understandable by study and investigation from the user [79, 80, 81, 82, 83]. In the literature, both interpretable and explainable systems refer to systems that provide information about the system's decision-making process using understandable terms to the user [79]. Doran et al. define an interpretable system as "a system where a user cannot only see but also study and understand how inputs are mathematically mapped to outputs" [84]. Montanavon et al. stated that "an interpretation is the mapping of an abstract concept into a domain that the human can make sense of" which in turn forms explanations [85].

2.5.1 Explainable Systems

The history of research and development of explainable systems has been rich and varied. Some research developed interfaces that are capable of explaining their underlying context-aware rules to users [86]. Some interfaces provide textual explanations [87, 88, 89, 90, 91]. Other interfaces provide visual explanations [92, 93]. Another stream of research has been done to investigate how to interpret and make sense of machine learning models. These research studies aim at developing and machine learning models that are interpretable and at the same time debuggable [94, 95]. Some research explored the interpretability, explainability, and understandability when interacting with autonomous vehicles [96, 97]. Explainability and interpretability have been also explored across many different application domains, such as in e-commerce systems [98, 99, 100, 101, 100, 102, 103, 104], social relation systems [105, 106, 107, 108], location systems [109, 110, 111], multimedia systems [112, 113, 114, 115, 116, 117], and healthcare [118].

Typically, these explainable systems tend to expose the reasoning and data behind their decision either through textual or visual explanations. Our proposed explainable storytelling model is capable of generating both textual and visual explanations to inspire user trust and satisfaction with the generated student stories. The explanations compose four parts: explanation title, explanation feature, explanation text, and explanation body. Further details about these explanations are presented in Chapter 6.

2.5.2 Evaluation of Explainable Systems

In the literature, there are many ways to evaluate a good explanation. For example, the ability of the explanation to inspire user trust and loyalty, make it simpler and faster for the user to find what they are looking for, improve target user satisfaction, and persuade the target user to receive and accept the system's decision. Several studies identified seven different metrics to evaluate explanations. These metrics are: (i) Transparency, (ii) Trustworthiness, (iii) Persuasiveness, (iv) Effectiveness, (v) Efficiency, (vi) Satisfaction, and (vii) Scrutability. This section introduces commonly used evaluation approaches for the system's explanations in terms of these seven metrics. Explanations can be evaluated using online evaluation, offline evaluation, or a combination of both.

Generally, in offline evaluation, there are two approaches to evaluate the explanations using offline evaluation. One is to evaluate the percentage of the system decisions that the system was able to generate explanations for, regardless of the quality of these generated explanations; and the second approach is to evaluate the quality of the generated explanations. For the first approach, there are some measures used in the literature to evaluate the explanations. For instance, Abdollahi and Nasraoui in [119], used two measures to evaluate their explanations. These measures are *Explainability Precision (EP)*; which is defined as the percentage of items explained in the top n system decisions relative to the number of system decisions for each user, and *Explainability Recall (ER)*; which is defined as the percentage of items explained in the top n system decisions relative to the number of all explainable decisions for a specific user. Another measure used by Peake and Wang in [120] is the *Model Fidelity (MF)*; which is defined as the proportion of system decisions that can be explained by the explanation model.

In the second approach, the evaluation of the quality of an explanation depends on the explanation type. For instance, one of the most common types of explanations is textual explanations, in which the explanation is presented to the target user as a complete and coherent sentence. Evaluation of such types of explanations can be conducted using text-based measures. Such as *BiLingual Evaluation Understudy (BLEU)* [75]; which is a measure used for automatic machine translation evaluation and it is highly correlated with human evaluation. Another text-based measure is the *Recall-Oriented Understudy for Gisting Evaluation (ROUGE)* [121]; which is a measure used to automatically determine the quality of a piece of text by counting the frequency of the overlapping units in a piece of text, word pairs, and word sequences and then comparing it to other ideal human-generated pieces of text. For instance, In [122], Lin et al. proposed Neural Outfit Recommendation (NOR); which is an explainable outfit recommender system, and conducted experiments to evaluate the explanations generated by their proposed system. They evaluated the generated explanations in terms of the BLEU and ROUGE measures, and their results showed that NOR achieves high ROUGE and BLEU scores compared to human-written comments.

Other studies evaluate the quality of explanations using readability measures, such as (i) *Gunning Fog Index* [123]; which estimates the years of education required by a person to comprehend and make sense of the text on the first reading. (ii) *Flesch Reading Ease* [124]; which finds what level of education someone will need to be able to read a piece of text easily, (iii) *Flesch Kincaid Grade Level* [125]; which is used to indicate how difficult a piece of text in English is to understand, (iv) *Automated*

Readability Index [126]; which estimates the approximate representation of the US grade level needed to comprehend a given text, and (v) *Smog Index* [127]; which estimates the years of education needed to understand a piece of text like Gunning Fog but it is more accurate and easier to be calculated. For instance, In [128], the authors proposed a framework to provide rating explanations. These explanations are generated based on users' reviews on a dataset of books from Amazon and evaluated using the readability measures. The results showed strong comprehensibility of the generated explanations.

In the online evaluation, the evaluation requires the user's interaction with the system. Usually, online evaluations are conducted using user studies by recruiting either volunteers or paid experiment users. Users for these user studies are hired either directly by the researchers or based on online crowdsourcing platforms such as Amazon Mechanical [129] and CrowdFlower [130]. In this evaluation, there are usually several measures used including, Conversion Rate (CR); which is the proportion of consumed items to the total number of checked (clicked) items, and Click Through Rate (CTR); which is the proportion of checked (clicked) items to the total number of items.

Evaluating the decisions' explanations depends mainly on the aims of these explanations, such as effectiveness, satisfaction, and persuasiveness. Typically, evaluating the explanations' persuasiveness is considered as the simplest measure; which refers to the user's acceptance to receive and accept a system decision. For instance, Zhang et al. in [99] conducted online A/B-tests to evaluate the persuasiveness of their system's explanations. In their experiments, they used three groups of users, an experimental group; that receives the proposed explanations, a comparison group; that receives baseline explanations (such as "People also like"), and a control group; that receives no explanations. Then, to evaluate the performance of their proposed explanation model, the authors calculated the CTR to compare between the three groups. As a result of their tests, they found that the CTR of the experimental group is significantly higher than that of the other two groups. Vig et al. developed four explanation interfaces (RelSort, PrefSort, RelOnly, and PrefOnly) based on users' feature preferences and feature relevance. They conducted an online evaluation to measure three metrics: justification, effectiveness, and mood compatibility. More specifically, they asked the recruited users to complete an online survey to evaluate these three metrics based on the four proposed interfaces.

This dissertation evaluates the explanations using online evaluation to measure the usefulness of the generated explanations in terms of the seven evaluation metrics presented at the beginning of this chapter. This evaluation is presented in Section 7.6.

2.6 Summary

This chapter presented the existing research on the core components of sensemaking in LA systems. Since the main purpose of these systems is to provide domain experts with the ability to become aware of, reflect upon, and make data-based decisions, research is needed to explore ways to facilitate and promote domain experts' awareness and sensemaking of diverse, complex, and heterogeneous student data. Effective presentation of analytics results to promote domain experts' sensemaking is a major issue when dealing with large volumes of data such as in LA. Typically, visualizations using tables, charts, and graphs are the most widely used. Data visualization in LA helps users to understand large datasets in a relatively short time by driving their attention to some important features of the students' data. However, these visualizations require some level of expertise to be interpreted correctly and on time based on their communicative power [29]. Results from analytics models typically treat students at an aggregate level. Although useful, aggregate analytics does not necessarily help advisors with their most important task of interacting with and understanding students on an individual level. This chapter also presented alternative approaches to present large volumes of data. One promising approach is to present the student data using storytelling techniques. In an attempt to facilitate the sensemaking of student data to domain experts with various levels of expertise, this dissertation proposes a novel approach to present student data. This approach presents the student data to domain experts using natural language stories that are automatically generated and updated according to the results of the students' aggregate analytics. For this purpose, two fields from the literature are discussed in this chapter: data storytelling and creative storytelling. Research in data storytelling aims to summarize or synthesize structured data. The underlying assumption behind these research efforts is that the presentation of tabular or numeric data using a natural language style story makes the data more understandable and memorable by human users. Moreover, research in creative storytelling aims to generate compelling and engaging stories. Chapter 3 presents a novel LA sensemaking model and highlights the importance of this model to promote the domain experts' sensemaking of student data. The notions and aspects of the four fields discussed in this chapter are considered when developing the LA sensemaking model.

CHAPTER 3: FIRST: A NOVEL SENSEMAKING MODEL FOR COMPLEX AND HETEROGENEOUS STUDENT DATA

3.1 Overview

This chapter introduces the proposed LA sensemaking design-driven conceptual model. Traditional sensemaking models involve carrying out the analysis and then show the final results of the analysis to the domain experts. In the proposed sensemaking model, the domain expert interaction is a major component in the analytical model. The model affords domain experts to first interact with the student data model to select the features that they find useful and interesting to be included in the analytical model. Then, after the analysis, domain experts can interact with and reflect on the student stories generated after applying the analytical model. As shown in Figure 3.1, the proposed model includes five main stages. These stages include: (i) Student data model, (ii) Domain experts' questions and queries, (iii) Student data reasoning, (iv) Student storytelling model, and (v) Domain experts' reflection. Based on the stage numbering in Figure 3.1, these stages are presented in the following sections.

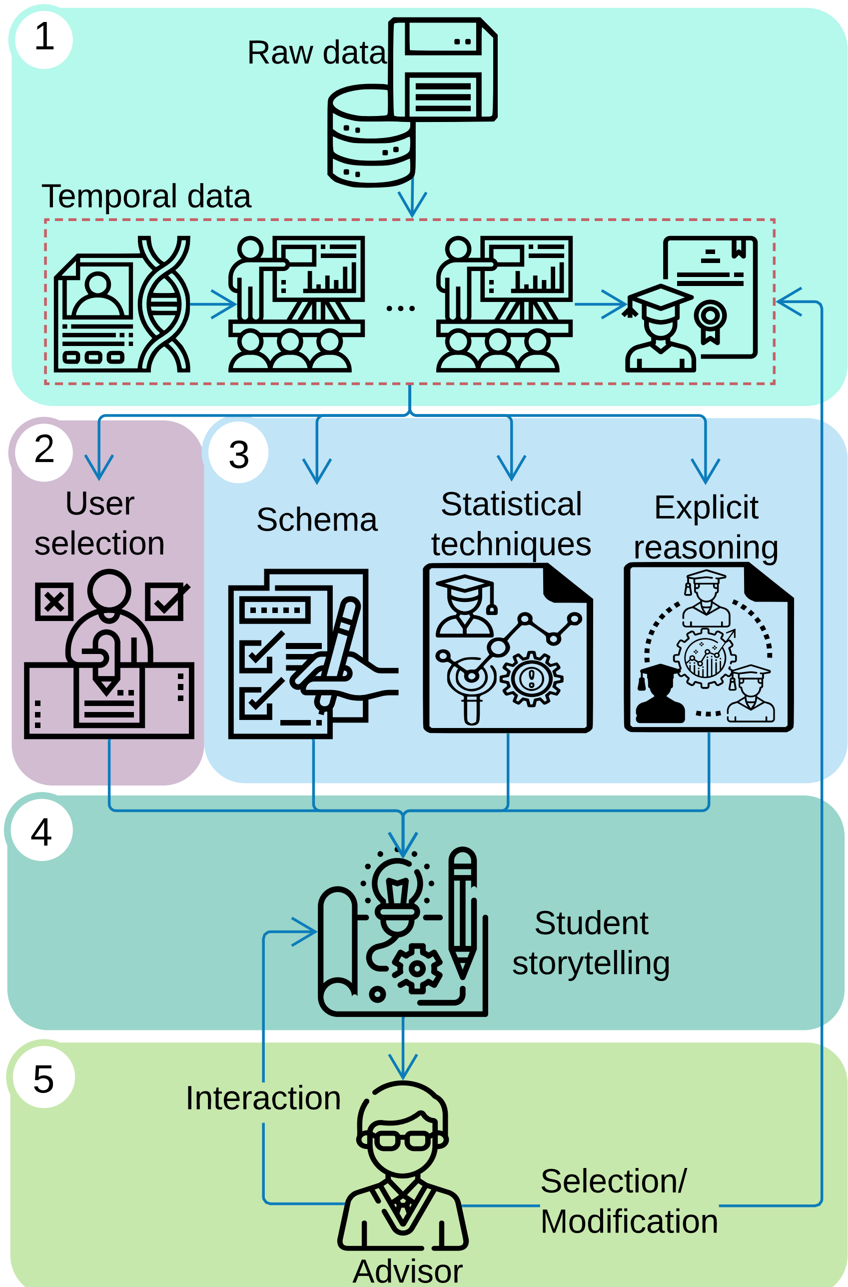


Figure 3.1: Sensemaking model architecture.

3.2 Student Temporal Data Model

This stage is concerned with the students' data available for the analysis. This dissertation adopts the temporal data model proposed in [14] to represent the students' data. This data model uses the time to sort heterogeneous and diverse sources of student data in a sequence of nodes for each student as shown in Figure 3.2. This temporal data model has several benefits when applying analytical models to student data. First, it allows the analytical models to consider the temporal dependency of student data throughout their enrollment. Second, it helps in identifying any unexpected or unusual patterns in student data. Finally, it gives flexibility when defining each temporal node, contextualizing information within a node, changing the granularity of a node, and interpreting sequences of nodes as stories.

The student sequence data model contains one sequence per student where a sequence is a representation of nodes arranged in the sequential order of the enrolled semesters. Each node in a sequence represents a period (e.g., a single semester) and contains a vector of features (variables, such as courses taken in that semester). In this study, there are three types of nodes for each student: the background node with demographic information, the semester node with semester-wise activities and information, and the outcome node with the value of the target variable. The student data model is shown in Figure 3.2.

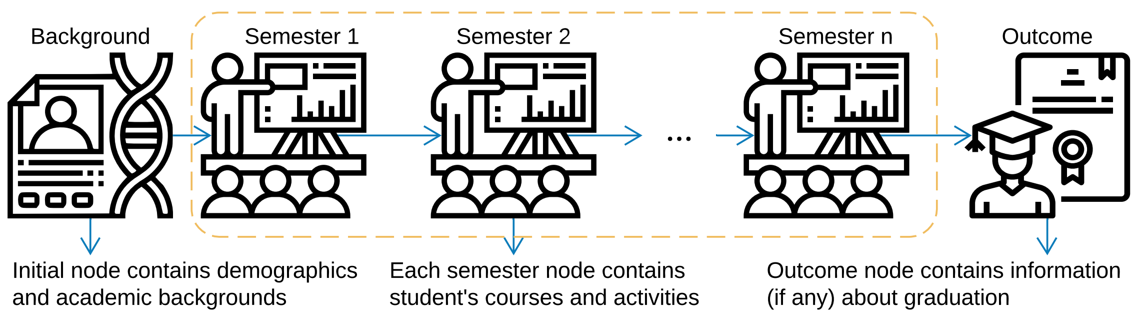


Figure 3.2: Student temporal data model.

3.3 Domain Expert's Questions and Queries

This stage is concerned with domain experts' questions and queries regarding the student data that they consider useful for their sensemaking. The proposed model involves the domain experts in the loop of sensemaking by affording them the ability to select the data that they find interesting and useful from the student data. The involvement of domain experts in the loop of LA sensemaking allows them to discover interesting patterns in student data that are not apparent to the data scientist. Therefore, the proposed sensemaking model allows the domain experts to select the features from the student temporal data model that they find interesting and useful for their sensemaking. These features then are used when deciding the contents of the student stories. The decision on how to report the selected features is based on the student data reasoning which is presented in the following subsection.

3.4 Student Data Reasoning

This stage is concerned with analyzing the student data to find interesting patterns, insightful information, and unexpected behaviors. Three types of reasoning are employed for these purposes. First, schemas are identified using a set of rules that either inspect any performance improvement or deterioration concerning some student features. These rules are constructed by analyzing the student dataset and identifying the points where students show performance improvement or deterioration. Second, statistical techniques aim to detect any big change (either increasing or decreasing) in student temporal behavior (performance) from one semester to another (anomaly if occurs at a certain time. e.g. large spike in the student's number of failed credits in a specific semester). Finally, explicit reasoning aims to examine the unusual or unexpected information from the student data. The input to the student data reasoning stage composes temporal student data and domain experts' questions and queries. The reasoning approaches used in the proposed sensemaking model are discussed in detail in Section 4.3.2.1

3.5 Student Storytelling

This stage is concerned with delivering the insights obtained through the student data reasoning stage to the domain experts. These insights are delivered through natural language stories that are generated from the temporal student data and analytics results. Student stories provide an effective, engaging, and easy way to understand complex and heterogeneous student data. Another important aspect that makes storytelling so effective in presenting complex student data is that it works for all types of users regardless of their level of expertise. In addition to the student stories, the sensemaking model supports the domain experts' trustworthiness and satisfaction by providing explanations and justifications on why, and how a story content is generated and included in the student story. These explanations aim to expose the reasoning and data behind the content of a student's story. These explanations are presented to the domain experts using visual components like scientific charts- pie charts, bar charts, and line charts, or they can also be presented using tabular components like a table of student courses attempted, passed, failed, or withdrawn throughout their enrollment.

3.6 Domain Expert's Reflection

This stage is concerned with the domain expert feedback regarding the results of the reasoning stage. Two types of reflections are included in the model. First, the domain experts can interact with the generated stories to expose the reasoning and data behind the information presented in the student stories. Second, the domain experts can change their questions or queries (student features), which in turn will update the student stories according to their new questions. These two types of reflection aim to increase the domain experts' trust in the model, find the information that can help them make sense of the student pattern of performance. This stage is discussed in detail in Chapter 6.

3.7 Summary

This chapter introduced the conceptual sensemaking model. Unlike traditional sensemaking models, in which the analytical process is carried out without the involvement of domain experts in the reasoning process, the domain expert interaction is a major component of the proposed sensemaking model. The conceptual sensemaking model composes five main stages. These stages are: (i) Temporal student data model, (ii) Domain experts' questions and queries, (iii) Student data reasoning, (iv) Student storytelling, and (v) Domain experts' reflection. The major component that governs the whole sensemaking process is the student storytelling model. This storytelling model is capable of automatically generating stories from student data. These stories aim to provide an effective and more natural way of data presentation to users with a wide range of expertise. The storytelling model aims to explain students' academic performance and progression throughout their enrollment at the individual level. This aims to provide advisors with an insightful overview of students before they go to advising sessions. The student storytelling model is discussed in Chapter 4.

CHAPTER 4: FIRST: A STORYTELLING MODEL FOR COMPLEX AND HETEROGENEOUS STUDENT DATA

4.1 Overview

This chapter introduces the student storytelling model (FIRST). This model is the major component that governs the whole sensemaking process presented in the previous chapter (Chapter 3). This storytelling model is capable of automatically generating stories from structured student data. These stories aim to provide an effective and more natural way of data presentation to advisors with a wide range of expertise. The storytelling model aims to explain students' academic performance and progression throughout their enrollment at the individual level. The storytelling model provides advisors with an insightful overview of students before they go to advising sessions.

The generated stories present information about the students' background, performance throughout their enrollment in a degree, and their outcome information. The determination of this information as well as the structure and organization of this information into a complete and coherent story has to be identified through user studies with the domain experts and extensive analysis of the student data.

The developed storytelling model is evaluated by conducting a series of user studies to understand the effectiveness of the model (See Chapter 7). The following subsections describe the methodologies for the proposed research.

4.2 Story Sentence Representation

This section presents how FIRST represents the sentences in the student stories. The story sentence representation is based on predefined templates that are associated with various kinds of features from the student data model. Sentences in student stories are structured as a group of consecutive clauses. A clause in a sentence can be one or more words that are labeled with a set of predefined labels. These labels are used to manage the presentation of the story sentences as well as clauses within a sentence. An example of a single semester credit count sentence template is shown in Figure 4.1. The sentence template composes static and dynamic clauses. As shown in this figure, the sentence structure is as follows:

- Sentence label: this label indicates the type of contents within the sentence. This label is used to decide the sentence section based on Freytag's pyramid discussed (see Section 4.3.2.2).
- Relation: this label indicates the relation of the sentence to the sentence label. For instance, the label COUNT in Figure 4.1 is used to indicate that this sentence describes the count of credits attempted by the student in a single semester. Other examples of relations in the student stories include OUTLIER, SIGNIFICANT_CHANGE, PERFORMANCE_IMPROVMENT, and PERFORMANCE_DETERIORATION.
- Raw features: this label includes the set of features from the student temporal data model used to generate the sentence clauses. These features include all the features from the student temporal data model.
- Engineered text: this label indicates how the text for the clause in a sentence is generated from the student's raw features. For example, in figure 4.1, the engineered text "*credits_count*" is decided by summing up the number of credit hours for all courses in Fall 2020.

The clause structure within a sentence is as follows:

- Text: the text to be displayed in the student story.
- Feature: the feature used to generate the text for the clause.
- Type: either dynamic (i.e., generated from student temporal data features) or

static (i.e., from a predefined template).

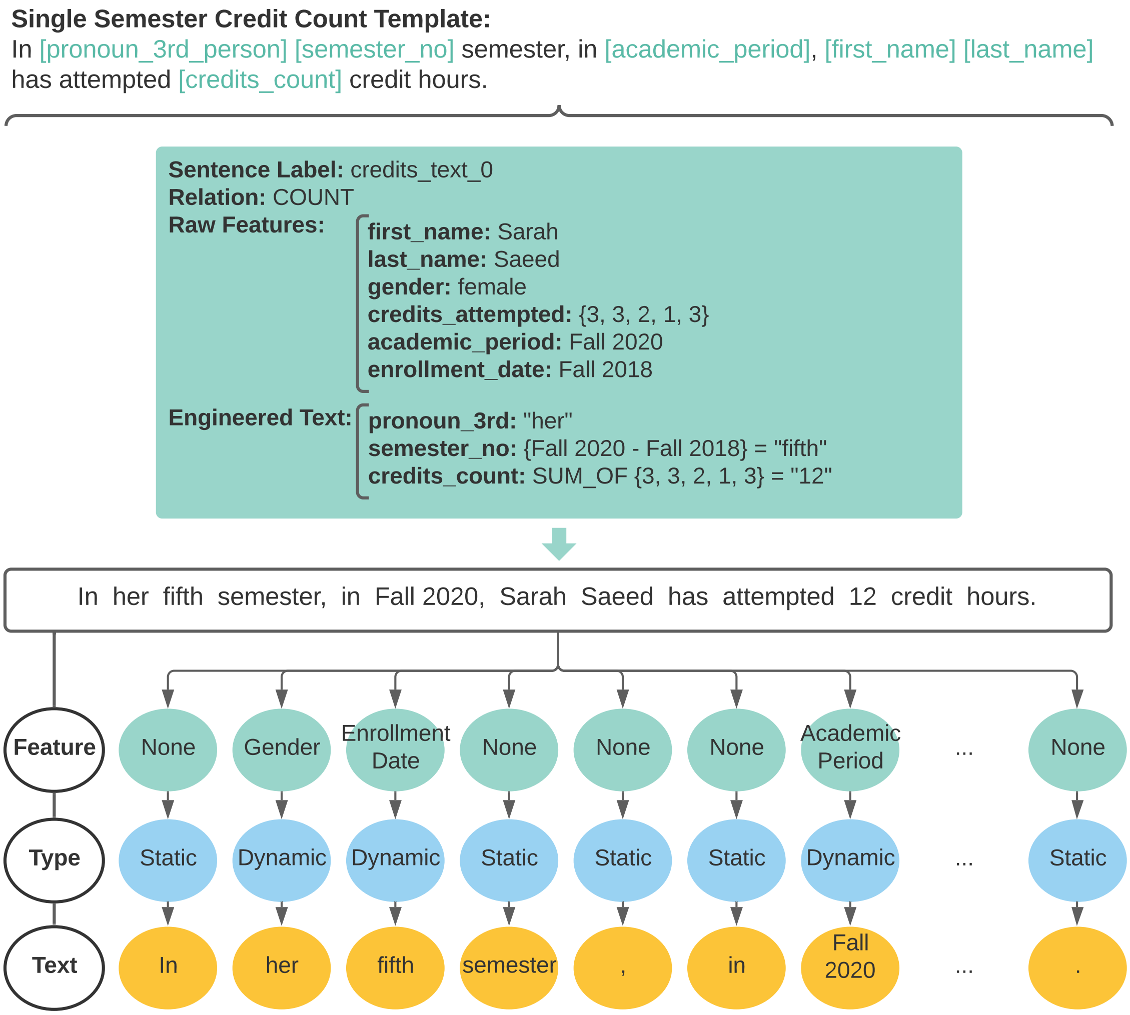


Figure 4.1: An example of a single semester credit count sentence template.

This representation serves multiple purposes when presenting the students' stories along with their respected explanations. These purposes are summarized as follows:

- Identifying the static and dynamic parts of the student stories. These parts can be highlighted for the advisor to enable them to differentiate between interactive and non-interactive parts of the student story.
- Deciding the different explanation components based on the sentence label, relation, raw features, type, and engineered text. The explanation generation algorithm (presented in Chapter 6) uses these parameters to decide the different

explanation components. For instance, if a part in a sentence is labeled with the feature "*credits_attempted*", this feature is used to give the advisor the option to view the student's credits attempted timeline.

4.3 Student Story Generation Model

This section presents the process of generating student stories from the student data model discussed in Section 3.2. Based on this data model, the student stories have three main components: (i) Background component, (ii) Semester component, and (iii) Outcome component.

- Background Component: presents the students' demographic data such as gender, age, primary ethnicity, citizenship type, etc. Besides, the background component presents the students' previous education such as institution type (college, high school), school rank, school GPA, etc.
- Semester Component: presents various important aspects about the student, such as the number of advisors, the number of attempted, passed, failed, and withdrawn credit hours, etc.
- Outcome Component: presents information regarding the graduation status of the student such as the number of semesters until graduation, graduation date, or expected graduation date. It also includes the students' total GPA at the time of graduation.

Figure 4.2 shows an example of a student temporal data model used by the storytelling model. Figure 4.2(a) shows the nodes in the temporal data model, Figure 4.2(b) shows the features selected from each node, and Figure 4.2(c) shows the sentences that are constructed from each feature. In these sentences, the text in black is from a predefined template while the text in red is generated from the features. After generating the sentences for each of the selected features, these sentences are used to generate the story as discussed in the following subsections.

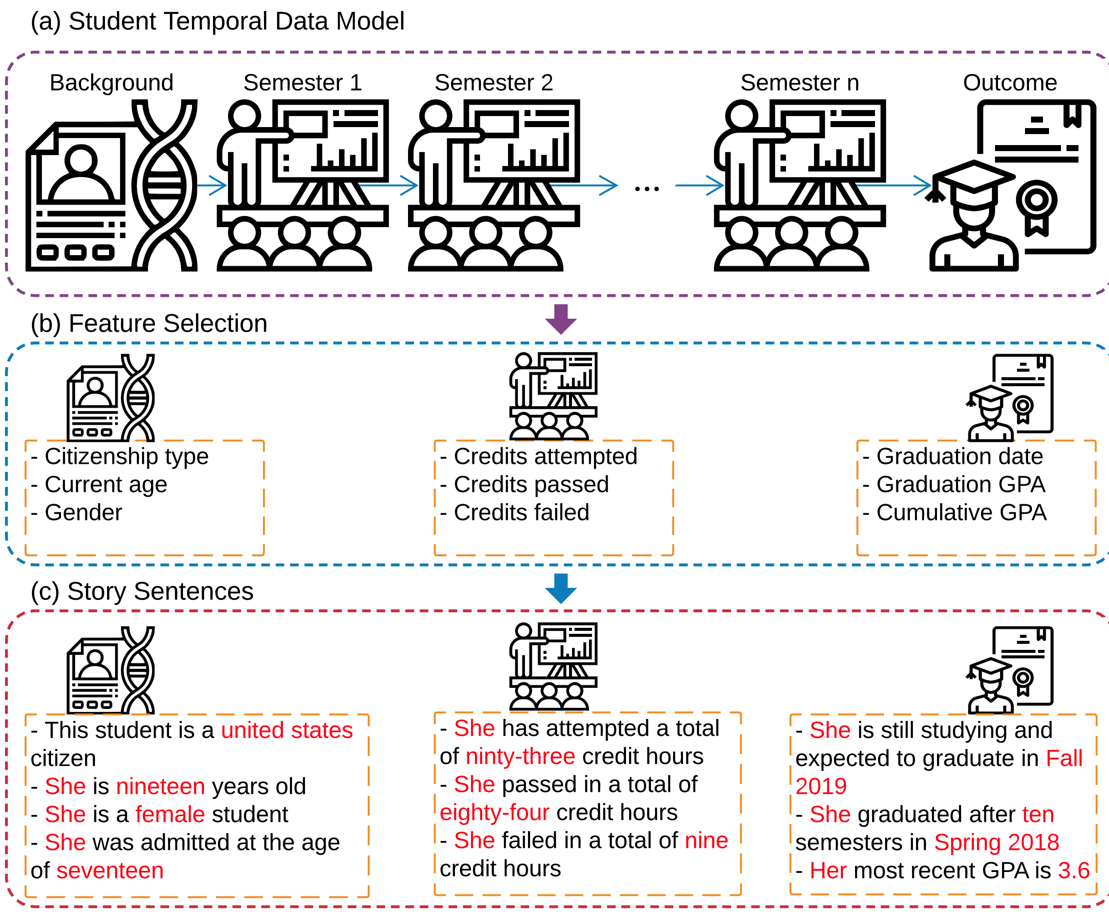


Figure 4.2: Components for generating students' stories: (a) temporal data model, (b) selected student features, and (c) examples of sentences in the story

The process of generating stories in natural language, as shown in Figure 4.3, has four main stages: data source, story synthesis, story analysis, and student story. The following subsections introduce these stages one by one.

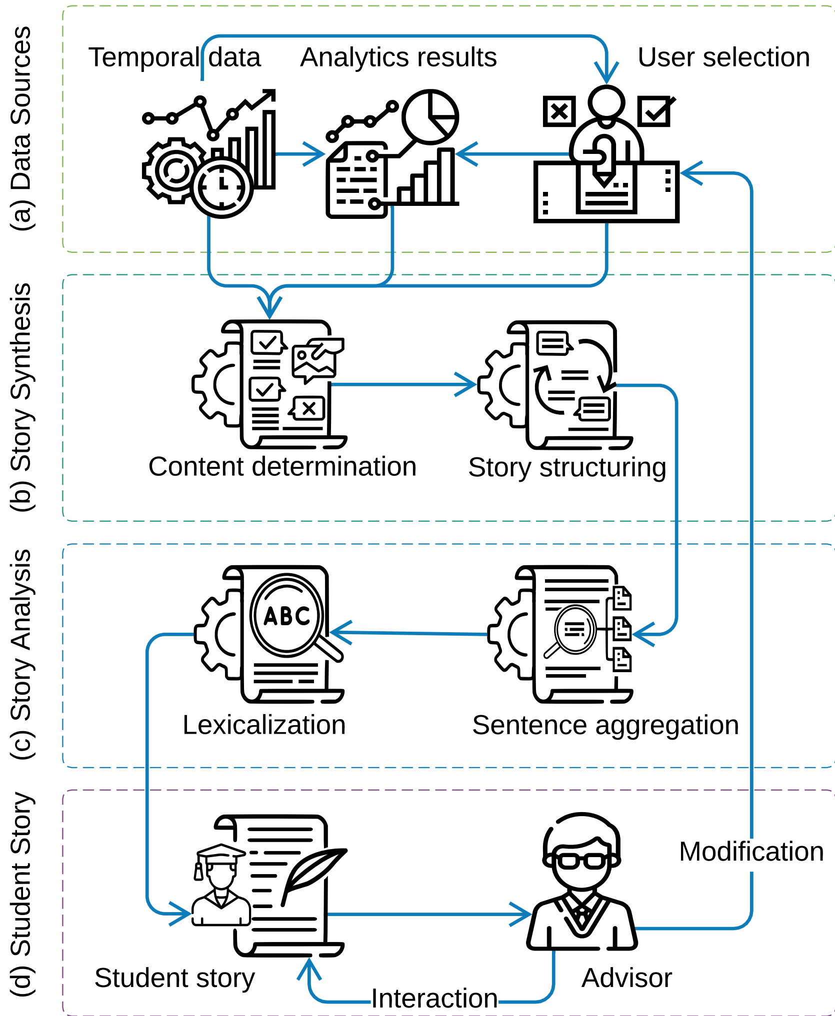


Figure 4.3: Student storytelling model

4.3.1 Student Story Data Sources

As shown in Figure 4.3(a), three primary sources of student data are used as input to the story generation. These sources are the raw variables in the student temporal data, user-selected features, and the aggregate analytics results. Each of these sources serves a role in the subsequent content determination task. The following sections present these data sources one by one.

4.3.1.1 Temporal Student Data

This data source includes all student variables and features in the student temporal student data model. According to the student data model presented in Section 3.2, these features are grouped into three types: background features, semester features, and outcome features. All of these features are used to run against a set of predefined performance rules in the subsequent content determination task. Consequently, if any of these features satisfy one of the rules, then it is included in the output student story. The performance rules are defined to spot any interesting information in the student data. For example, some of these rules aim to detect if the student academic performance has improved or deteriorated in terms of various features including student GPA, courses grades, number of passed credit hours, number of failed credit hours, number of withdrawn credit hours, number of credit hours with a D grade, academic standing.

4.3.1.2 User-Selected Features

To make the storytelling model scrutable and interpretable by the user and to make the student stories customized and tailored to a specific user, the storytelling model affords the user to select the features that they find interesting from the student temporal data model. Information about these selected features is included in the output student stories. This gives the domain experts the flexibility to be part of the student story generation process and allows them to focus on the features they find helpful and useful for their decision-making and sensemaking of students data.

4.3.1.3 Analytics Results

To make the student stories interesting and helpful for domain experts, the storytelling model uses an anomaly detection approach to compare the performance of a student in the context of other students. In other words, it detects if there are extreme values (anomalies) in the student data compared to other students and includes these anomalies in the student stories. Two models of anomaly detection are performed: (i) Personal Anomaly Detection (PAD) and (ii) Collective Anomaly Detection (CAD). The PAD model aims to detect if an individual student's data instance can be considered anomalous compared to the rest of the data. For instance, the PAD model detects if a student's GPA has extremely decreased from one semester to another compared to other students. The CAD model aims to detect if a collection of student data instances is anomalous compared to other students, but not individual values. For instance, the CAD model detects if a student follows a non-typical pattern for the number of credits passed each semester. A detailed discussion about these anomaly detection models is presented in Chapter 5.

4.3.2 Student Story Synthesis

The goal of this stage is to determine and sort the content presented in the student's story. Therefore, it includes two tasks (See Figure 4.3(b)); (i) content determination; and (ii) story structuring. The following subsections discuss these two tasks in more detail.

4.3.2.1 Content Determination

This is the task of choosing which pieces of information should be included in the student story, and which piece should be dropped. In this task of story generation, three general factors impact the content determination task. These issues can be illustrated using the following examples of sentences in the student story.

1. Regarding Sarah's number of failed credits, in her 5th semester, in Fall 2020, it has significantly increased from 2 to 13 credit hours.

2. Sarah's cumulative GPA is 3.9 and compared to other students in CCI, Sarah's number of withdrawn credit hours follows a non-typical pattern, in which, during her four enrolled semesters, she withdrew 7, 9, 6, and 8 credit hours respectively.

3. Compared to other students in CCI, Sarah's number of passed credit hours follows a typical pattern.

Based on the above sentence examples, the factors that impact the content determination are as follows:

- The first factor is the communicative goal of the text, i.e. its purpose and reader. In the above examples, for instance, an advisor who wants to decide the student risk status would probably be most interested in the student's significant increase of the number of failed credits each semester or the unusual number of credits withdrawn each semester (first and second examples). However, in the third example, knowing that the student number of credits passed is normal is less important than the other two examples.
- The second factor is the size and level of detail of the generated text. For instance, a short story about a successful student who has no major issues or troubles can be sufficient for the advisor to decide that this student is doing well. But, a longer story about an at-risk student who has major issues or troubles can be more helpful for the advisor to find out where the struggles are for a student and try to provide the necessary advice.
- The final factor is how unusual and unexpected the information in the student story is. For instance, in the second example above, the student has a high GPA and at the same time has a large number of withdrawn credit hours each semester. This unexpected pattern is helpful for the advisor to decide if the student is at-risk or not and therefore should be included in the student's story. For this purpose, the storytelling model uses cluster-based anomaly detection using k-means clustering to identify ways in which this student is different from other students. More discussion about this anomaly detection model is presented in Chapter 5.

Based on the aforementioned factors, there are three basic subtasks to content determination: schemas, statistical approach, and explicit reasoning. The following sections present these subtasks:

- **Content Determination Using Predefined Schemas:** Schemas are templates that explicitly specify the content of a generated text. These schemas are identified using a set of rules that either inspect any performance improvement or deterioration compared to some student features. These rules are constructed by analyzing the student dataset and identifying the points where students show performance improvement or deterioration. An example of this approach is shown in Figure 4.4.

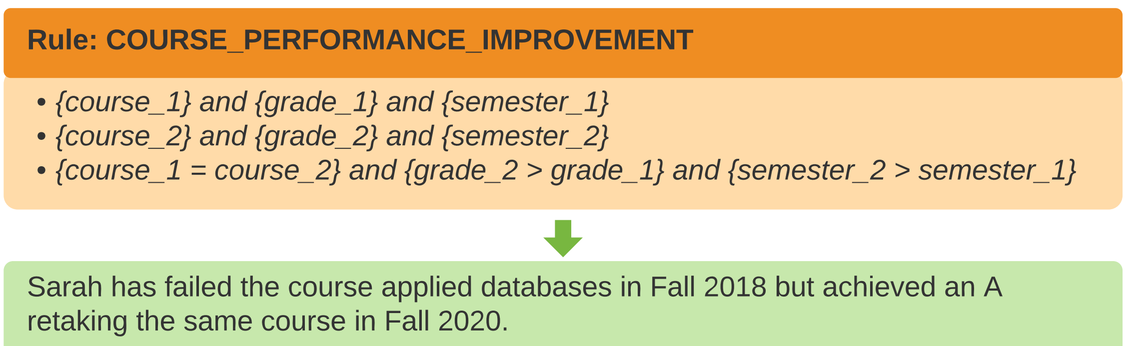


Figure 4.4: Story content determination using predefined schemas example

- **Story Content Determination Using Explicit Reasoning:** This subtask uses statistical analysis techniques to automatically determine the content of the generated texts. This subtask aims to detect any significant change (either increasing or decreasing) in student temporal behavior (performance) from one semester to another. This means that a student feature is considered anomalous if it occurs at a certain time, e.g. large spike in the student's number of attempted credits in a specific semester. For this purpose, this study uses the PAD model on data features from all students in CCI. An example of a sentence generated using this subtask is shown in Figure 4.5. The PAD model is discussed in more detail in Chapter 5.

> Regarding Sarah's number of failed credits, in her 10th semester, in Fall 2020, it has significantly increased from 2 to 13 credit hours.

Figure 4.5: Story content determination using statistical techniques example

- **Story Content Determination Using Explicit Reasoning:** The explicit reasoning sub-task uses clustering techniques to examine the unusual or unexpected information from the student data. Specifically, this study uses a clustering-based CAD model to examine any unusual patterns in the student data compared to other students. The underlying assumption is that when clustering students data, normal student data will belong to clusters while anomalous data will either belong to small clusters or not belong to any cluster. An example of a sentence generated using this sub-task is shown in Figure 4.6. The CAD model is discussed in more detail in Chapter 5.

> Compared to other students in CCI, Sarah's number of withdrawn credit hours follows a non typical pattern, in which, during her four enrolled semesters, she withdrew 7, 6, 9, and 8 credit hours respectively.

Figure 4.6: Story content determination using explicit reasoning example

This dissertation is an attempt to combine the three approaches for deciding the story content to be communicated to the user.

4.3.2.2 Story Structuring

This is the task of deciding the plot of the story presented to the reader. The main objective of this task is to generate compelling and engaging students' stories. The ultimate challenge in this task is to generate a good narrative. In other words, a story starts by setting the scene and giving an introduction or overview; then describes a set of events so readers can easily see how the individual events are related and linked together; and concludes with a summary or ending. To achieve this, this dissertation employs two fundamental approaches from the creative storytelling field: the story representation structure proposed in [4] to identify the story elements and the Freytag's pyramid [5] to decide the plot of the students' stories.

Concepcion's story representation structure [4] is used to identify the story elements from the student data. Their representation model is a structure that consists of a specific set of concepts; in which each concept plays a particular role when generating a complete story. These concepts include space, plot, setting, location, happening, existents, and characters. The mapping of students' data to Concepcion's structure is shown in Figure 4.7.

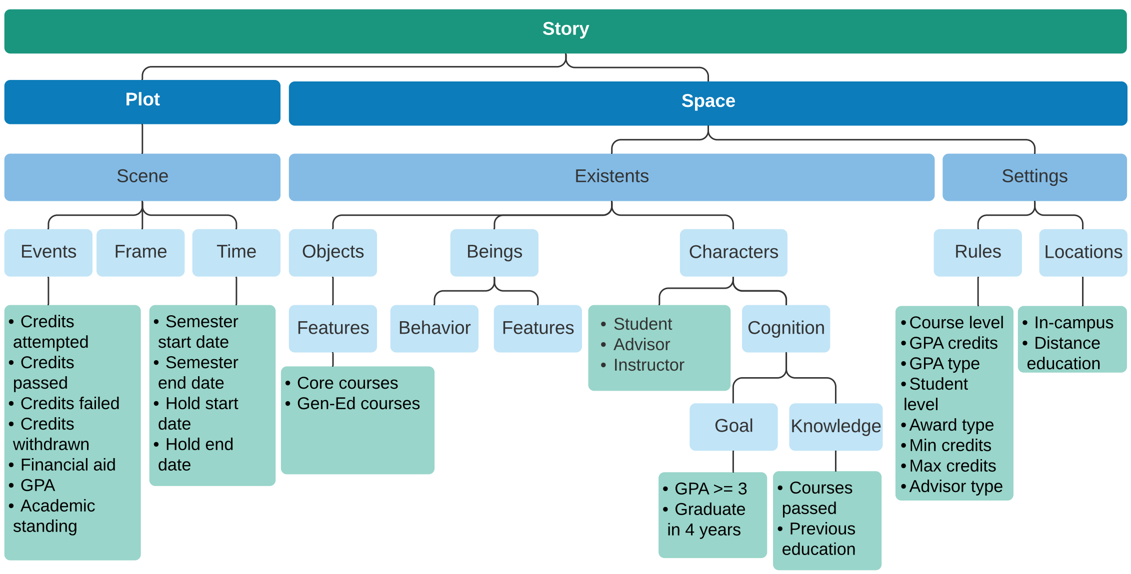


Figure 4.7: Student feature mapping to Concepcion's story representation structure [4]

The Freytag's pyramid [5] is used to decide the plot of the students' stories. As shown in Figure 4.8, it divides the story plot into five distinct sections:

- Introduction: sets out the background and historical information that is needed to encourage the reader to appreciate the storyline.
- Rising action: presents the basic conflict by introducing related secondary conflicts, including various challenges and hurdles that frustrate the main charac-

ters' from reaching their goal.

- Climax: the turning point that affects a change either for the better or for the worse in the protagonist's situation.
- Falling action: happens after the climax. It signals that the turning point of the story is over and the storyline is heading to the end.
- Conclusion: the end of the story is considered as the closing and clarification of a narrative plot.

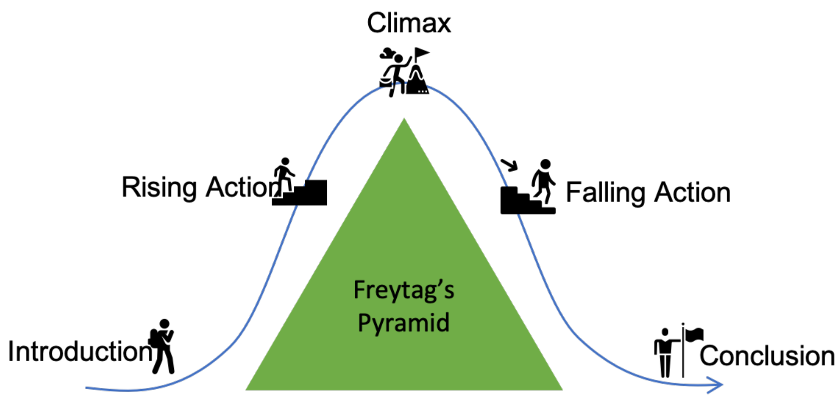


Figure 4.8: Freytag's pyramid [5].

In this dissertation, all possible sentences in the student stories are classified into their respected Freytag's section. The sentence classification is done when generating sentences to be included in the student story. The classification depends on the sentence label and the relation. Table 6 shows some examples of student story sentences and their classification based on Freytag's analysis.

Table 4.1: Examples of student story sentences and their classification based on Freytag's analysis.

Sentence Label	Relation	Freytag's Section
addmission_population	REPORT	Introduction
admission_date	REPORT	Introduction
primary_ethnicity	REPORT	Introduction

attempted_credit_by_semester	COUNT	Rising Action
total_attempted_credits	COUNT	Rising Action
advisor_info	COUNT	Rising Action
attempted_credit_by_semester	OUTLIER	Climax
gpa_by_semester	OUTLIER	Climax
retaken_course	PERF-IMPROVEMENT	Falling Action
academic_standing	PERF-DETERIORATION	Falling Action
gpa_by_semester	SIGNIFICANT-CHANGE	Falling Action
graduation_gpa	REPORT	Conclusion
graduation_status	REPORT	Conclusion

4.3.3 Student Story Analysis

This stage is to make the language of the stories more human-readable and coherent. This stage includes two tasks as shown in Figure 4.3(c): (i) Sentence aggregation, and (ii) Lexicalization and linguistic realization.

4.3.3.1 Sentence Aggregation

This task is to cluster multiple pieces of the same kind of information together into a single sentence instead of several ones. For instance, if there are a set of candidate sentences as "student achieved an A in the course X", "student achieved B in course Y", and "student achieved B in course Z", these sentences should be aggregated into one sentence "student maintained all his grades at B or above". Sentence aggregation is done on a set of features, such as $credit\_attempted$, $credit\_passed$, $credit\_failed$, $advisor\_count$, etc.

4.3.3.2 Lexicalization and Linguistic Realization

Lexicalization is choosing the proper words and phrases to transform the data into natural language text. Linguistic realization is inserting punctuation, functional words (such as prepositions and auxiliary words), and other elements required for the text to be fluid and coherent. In the proposed storytelling model, these are done using predefined templates for all possible sentences in the story.

4.4 Summary

This chapter described the process and tasks used to automatically generate stories from complex, diverse, and heterogeneous student data. Unlike the data storytelling models introduced in the literature, which rely only on the analytical results and the numeric predictions as input to the storytelling process. The proposed storytelling model uses three different sources of information; students' raw data, aggregate analytics results, and user-selected features. One of the essential tasks in the story generation process is the content determination. This task governs what information should be included in the student story. Three factors affect the story content determination: the communicative goal of the text, the size and level of detail of the generated text, and how unusual and unexpected the information in the student story is. Another important task in the story generation is story structuring. This task decides the plot of the story in which it is presented to the reader. Two fundamental approaches from the creative storytelling field are adopted for this purpose: the story representation structure proposed in [4] to identify the story elements and the Freytag's pyramid [5] to decide the plot of the students' stories. The next chapter presents the analytical models used to find interesting information and unexpected patterns in the student data to be presented in the student stories.

CHAPTER 5: ANOMALY DETECTION MODEL ON STUDENT DATA

5.1 Overview

This chapter presents the machine learning models that are used to find unexpected patterns in the student data. These models aim to find interesting information in the student data and present them to the domain experts. For this purpose, an anomaly detection model is developed to detect any data point in student data that does not fit with the rest of the student's data. The storytelling model uses an anomaly detection approach to compare the performance of a student in the context of other students. In other words, it detects if there are extreme values (anomalies) in the student data compared to other students. There are 3 categories of anomalies; point anomaly, contextual anomaly, and collective anomaly. The storytelling model uses two models of anomaly detection: (i) Personal Anomaly Detection (PAD) model which is based on the point anomaly and Collective Anomaly Detection (CAD) model which is based on the collective anomaly. The PAD model aims to detect if an individual student's data instance can be considered anomalous when compared to the rest of the data (e.g. a student's GPA decreased significantly from one semester to another when compared to other students). The CAD model aims to detect if a collection of student data instances (not individual values) can be considered anomalous when compared to other students. For instance, if a student follows a non-typical pattern for the number of credits passed each semester.

This dissertation uses the temporal features from the students' data to generate two types of feature vectors (one for each anomaly detection model) in terms of the features depicted in Table 1. The following sections present the two anomaly detection models.

Table 5.1: Student features for the anomaly detection models

No.	Feature/Semester
1	Number of credits attempted
2	Number of credits passed
3	Number of credits failed
4	Number of credits withdrawn
4	Number of credits with a D grade
4	Student's GPA

5.2 Collective Anomaly Detection Model

The CAD model aims to detect if a collection of student data instances can be considered anomalous when compared to other students. For instance, if a student follows a non-typical pattern for the number of credits attempted, passed, failed, or withdrawn each semester. This model includes three steps as shown in Figure 5.1: (i) data engineering, (ii) CAD model analysis, and (iii) story content. The following subsections describe these steps and present the challenges faced when building the CAD model.

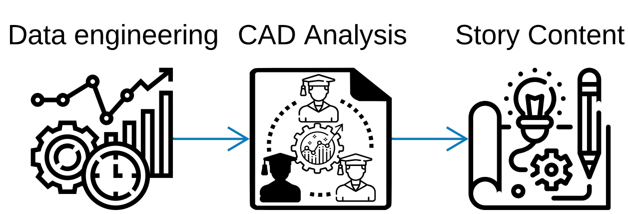


Figure 5.1: Collective anomaly detection model steps.

5.2.1 Collective Anomaly Detection Data Engineering

The data engineering in this model includes creating vectors from temporal student data throughout their enrollment. These vectors are used as input to the clustering model presented in the next subsection. Students have unequal vector sizes governed by the number of semesters they enrolled in. To overcome this issue, this model assembles students into different groups based on the number of semesters they enrolled in (the set of all students' groups referred to as G). For this purpose, the model considers all students who have at least 4 semesters, 5 semesters, 6 semesters, and so on. The minimum number of semesters considered for the analysis is 4 semesters and the maximum number of semesters is 19 semesters. The distribution of the number of students who have at least n number of enrolled semesters is shown in Figure 5.2. For each feature in Table 5.1, a feature vector is created for each student with the values from student data. Examples of feature vectors for the number of credits attempted feature for some students are shown in Figure 5.3.

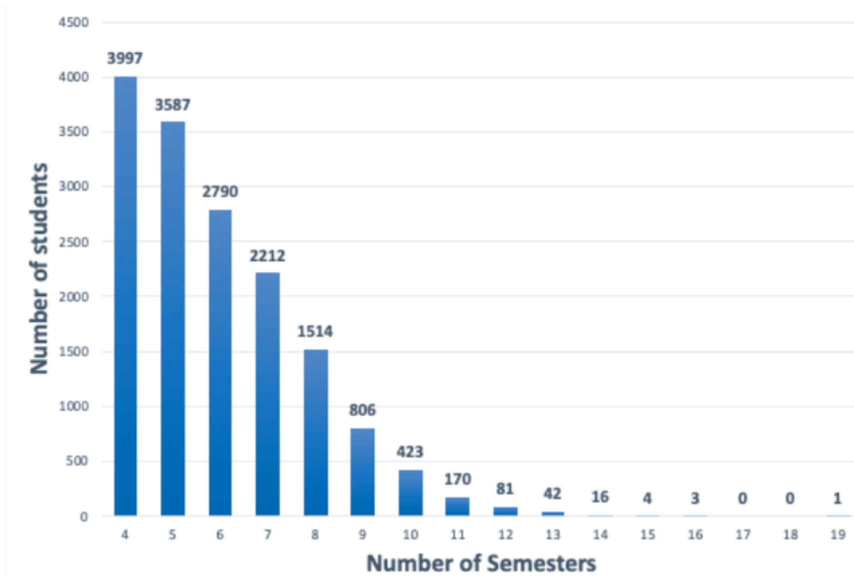


Figure 5.2: Distribution of the number of students who have at least n number of enrolled semesters; where $n \in [4, 19]$.

(a) Feature vector format (Sem_i: semester number, v_i: feature value for Sem_i)

Sem_1	Sem_2	Sem_3		Sem_n
v_1	v_2	v_3	...	v_n

(b) An example of a student's feature vector for the number of credits attempted each semester

Sem_1	Sem_2	Sem_3	Sem_4	Sem_5	Sem_6	Sem_7	Sem_8
15	12	9	6	12	9	9	14

(c) An example of the clustering algorithm input which is a set of equal-sized vectors for the number of credits attempted (8 semesters, *n* students)

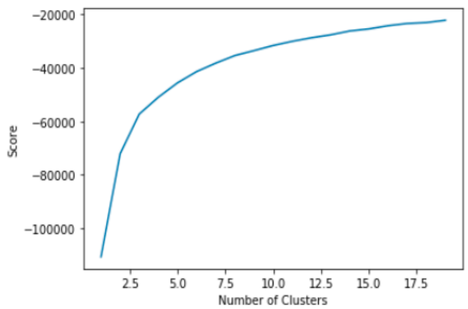
	Sem_1	Sem_2	Sem_3	Sem_4	Sem_5	Sem_6	Sem_7	Sem_8
$Student_1$	15	12	9	6	12	9	9	14
$Student_2$	10	14	15	6	6	10	6	12
$Student_3$	15	10	12	12	7	10	12	9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$Student_n$	15	12	8	10	10	6	6	6

Figure 5.3: Feature vector format for the CAD model.

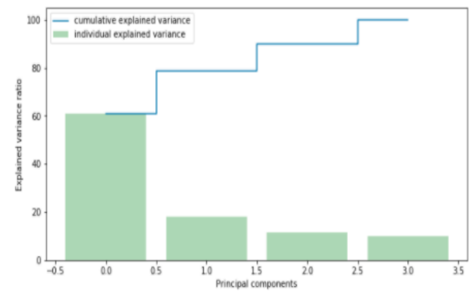
5.2.2 Collective Anomaly Detection Analysis

The CAD model aims to detect if a collection of student data instances is anomalous when compared to other students in CCI, but not individual values. A cluster-based anomaly detection using k-means clustering algorithm is used to identify ways in which a student is different from other students. K-means algorithm is applied on the student feature vectors presented in the previous section (See Figure 5.3(c)). It creates k similar clusters of feature vectors. Vectors that fall outside of these groups could potentially be marked as anomalies. For each group of students in G, the elbow method is used to determine the optimal number of clusters (See an example in Figure 5.4(a)). Then, a dimensionality reduction using the Principal Component Analysis (PCA) algorithm is performed to find the number of components (features) to keep from the students' vectors. An example of students with at least 4 enrolled semesters is shown in Figure 5.4(b). In this figure, the first component explains over 60% of the variance. The second component explains almost 20%. It is noticed that almost none of the components are negligible. The first 2 components contain over 80% of the information. So, the number of components is set to 2. After this, k-means clustering is applied on the student feature vectors using the determined number of clusters and number of components. Figure 5.4(c) shows a visualization of the obtained clusters for students with at least 4 enrolled semesters.

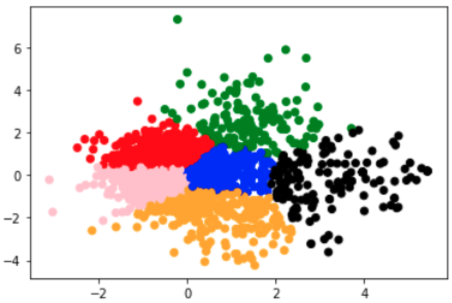
The underlying assumption in the clustering-based anomaly detection is that if we cluster the student data, normal student data will belong to clusters while anomalies will not belong to any clusters or belong to small clusters. For this purpose, the distance between each point and its nearest centroid is calculated. The biggest distances are considered anomalies. To determine the distance threshold for the anomalies in each cluster, a z-score is used to find the percentage of observations (number of anomalies) that should fall over the absolute value 3 in the z-score distance from the mean in a standardized normal distribution. Then, the distance threshold is set as the minimum distance of these anomalies. Any observation that falls over the threshold is considered an anomaly. Figure 5.4(d) shows a visualization of the anomalous student feature vectors. In this figure, normal students' patterns are colored in blue, while anomalous student patterns are colored in red.



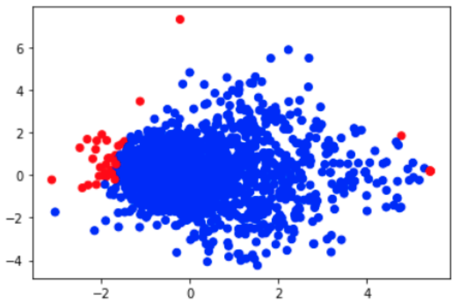
(a) Number of clusters using the Elbow method



(b) Number of components using PCA



(c) Plot of the obtained clusters



(d) Plot of anomalies with cluster view

Figure 5.4: CAD model for students with at least 4 enrolled semesters for the number of attempted credit hours/semester feature

5.2.3 Collective Anomaly Detection Story Content

The results of the CAD model are used to generate sentences in the student's story. For instance, a CAD model is used to generate a narrative sentence that describes the student pattern of performance in terms of the number of credits failed each semester as "Compared to other students, Luca follows a non-typical pattern for the number of credits failed each semester, in which, during his six enrolled semesters, he failed in 7, 3, 0, 3, 6, and 0 credit hours respectively".

5.2.4 Collective Anomaly Detection Challenges

The challenges faced while building the CAD model can be summarized as: (i) students have unequal vector sizes for the engineered features which makes it difficult to determine the clusters for students with small vector sizes. This dissertation tackles this problem by assembling students into different groups based on the number of semesters they enrolled in, and (ii) Students' data has high dimensionality and as the number of dimensions increases, k-means clustering converges to a constant value between any given observation. This means that the ratio of the standard deviation to the mean of distance between observations decreases as the number of dimensions increases. This convergence means k-means becomes less effective at distinguishing between observations. This dissertation tackles this problem by reducing the dimensionality by using PCA on the student feature vectors.

5.3 Personal Anomaly Detection Model

The PAD model aims to detect if an individual student's data instance can be considered anomalous compared to the rest of the data (e.g. a student's GPA decreased significantly from one semester to another when compared to other students). This model includes three steps as shown in Figure 5.5: (i) data engineering, (ii) PAD model analysis, and (iii) story content. The following subsections describe these steps in detail and present the challenges faced when building the PAD model

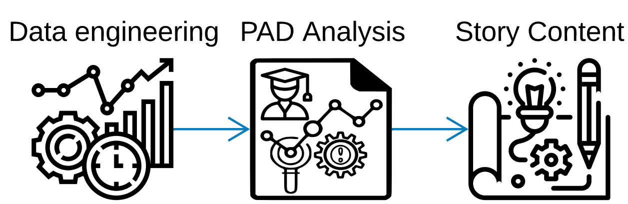


Figure 5.5: Collective anomaly detection model steps.

5.3.1 Personal Anomaly Detection Data Engineering

The data engineering in this model includes creating vectors from the temporal student data throughout their enrollment. These vectors are then used to find anomalies in the student temporal data. For this purpose, for each feature f from Table 5.1, the algorithm calculates the change of feature values Δv from one semester to another using Equation 5.1.

$$\Delta v_i = v_i - v_{i-1} \tag{5.1}$$

Where i is the semester number and v_i is the feature value for semester i. An example of a delta vector is shown in Figure 5.6(b). Then, for each student s, create a vector v_i of deltas as shown in Equation 5.2.

$$v_i =< \Delta v_2, \Delta v_3, \Delta v_4, ..., \Delta v_n > \tag{5.2}$$

Where n is the maximum semester number for students. After that, for each feature f, a feature feature vector V_{fi} is formed for each semester i that include Δv_i from all students in CCI using Equation 5.3.

$$V_{fi} =< \Delta v_{i.s_1}, \Delta v_{i.s_2}, \Delta v_{i.s_3}, ..., \Delta v_{i.s_n} > \tag{5.3}$$

Where n is the total number of students in CCI, $\Delta v_{i.s_j}$ is the delta value for student j in semester i. Example of semester feature vectors for the number of credits attempted each semester is shown in Figure 5.6(c).

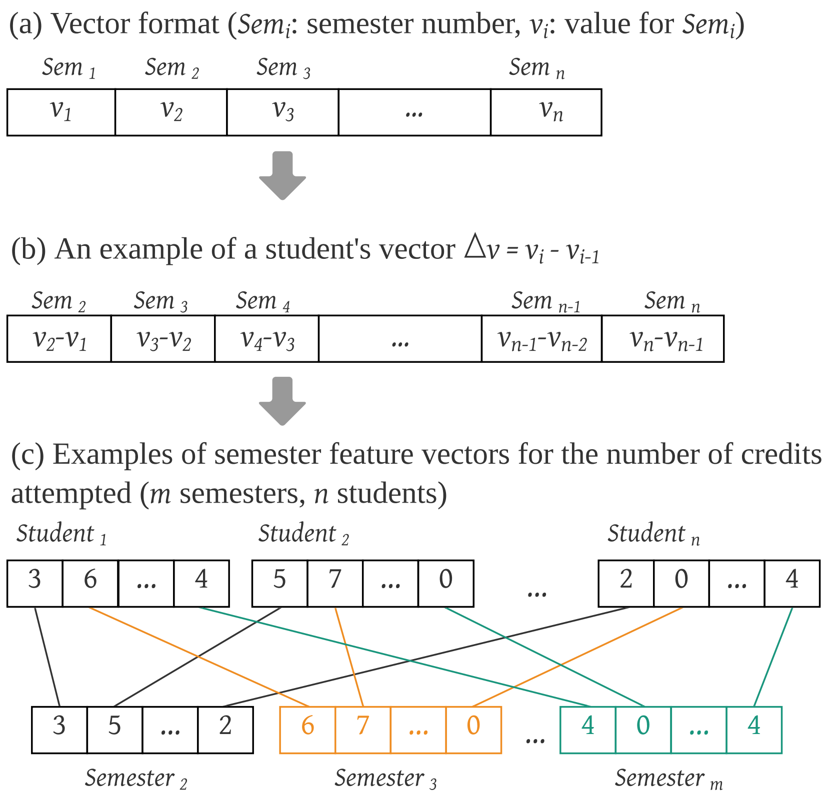
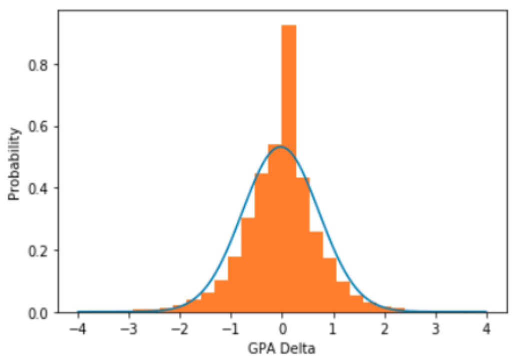


Figure 5.6: Feature vector format for the PAD model.

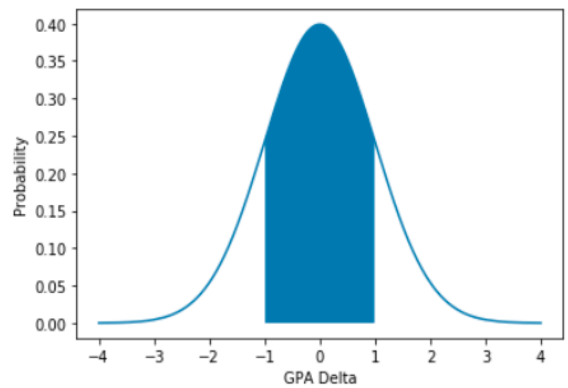
5.3.2 Personal Anomaly Detection Analysis

This model aims to detect any big change (either increasing or decreasing) in student behavior (performance) from one semester to another (i.e., anomaly if occurs at a certain time. e.g., a large spike in the student's number of attempted credits in a specific semester). Although contextual anomaly detection may look more appropriate for detecting such an anomaly, there are two issues in using contextual anomaly detection for students' data. First, individual student's data does not have sufficient historical data to help in detecting anomalies. However, the length of history that is used for detection is critical in detecting anomalies in the contextual anomaly methods. Second, students' data is not generated from a statistical process. Therefore, contextual anomaly detection will perform poorly in capturing anomalies in the student data [131]. In light of these two issues, a PAD model is performed on vectors generated from all students in CCI. The rationale behind this is two folds: (i) Rather than dealing with vectors with small lengths for each student with the feature values for each semester (maximum usable length is 10), vectors for each semester are created with a length equivalent to the number of students in the whole dataset (maximum usable length is 6203), and (ii) using vectors generated from all students data helps to identify anomalies of the student behavior compared to all students in CCI.

A z-score statistical analysis is performed to detect any anomaly observation in the feature vectors of deltas for each engineered feature. This analysis determines the observations that should fall over the absolute value 3 in the z-score distance from the mean in a standardized normal distribution. An example of z-score analysis for the GPA feature vector for the fourth semester is shown in Figure 5.7. Figure 5.7(a) shows the normal distribution of the GPA feature vector and Figure 5.7(b) shows the z-score analysis performed on this feature vector. In this figure, any observation that falls over the shaded area is considered an anomalous change in the student GPA in their fourth semester (i.e., if the student GPA increased by 1 or decreased by 1 from the third semester to the fourth semester, then this is considered as anomalous). Based on this figure, if the student's GPA change falls to the right of the shaded area, then this indicates that the student is improving. On the contrary, if it falls to the left of the shaded area, then this indicates that the student is underperforming for that particular semester.



(a) Normal distribution of GPA deltas over semesters



(b) z-score statistical analysis for GPA deltas over semesters

Figure 5.7: PAD Model using z-score analysis on the GPA feature vector for the fourth semester.

5.3.3 Personal Anomaly Detection Story Content

The results of the PAD model are then used to generate sentences in the student's story. For instance, a PAD model is used to generate a narrative sentence that describes the student's significant change of performance in terms of their GPA as "Regarding Luca's GPA, in his 7th semester, in Fall 2019, it has significantly decreased 2.5 points, from 3.0 to 0.5".

5.3.4 Personal Anomaly Detection Challenges

The challenges faced while building the PAD model can be summarized as (i) individual student's data has limited length governed by the number of semesters they spent in the university, and (ii) students' data is arbitrary (i.e., it is based on student's choice or personal whim, rather than any statistical process). Because of these two characteristics, it is hard to apply contextual anomaly detection to detect any big change in student behavior. This dissertation tackles these issues by generating feature vectors by combining student vectors from all students' data on a semester basis and applying a PAD model using statistical methods (e.g., z-score).

5.4 Summary

This chapter presents the anomaly detection models that are used to find unexpected patterns in the student data. These models aim to find interesting and insightful information in the student data and present them in the student stories to the domain experts. These models work by detecting if there are extreme values (anomalies) in student data compared to other students in CCI. The rationale behind this approach is the assumption that normal and anomaly student behaviors form different clusters in the features space. In this dissertation, two models of anomaly detection are performed: (i) Personal Anomaly Detection (PAD) and (ii) Collective Anomaly Detection (CAD). PAD aims to detect if an individual student's data instance can be considered anomalous compared to the rest of the data. CAD aims to detect if a collection of student data instances is anomalous when compared to other students in CCI, but not individual values.

The main contributions of the CAD model are threefold: (i) the anomalies detected by k-means clustering were either some of the very high rates or some of the very low rates. This gives the ability to decide if a student is performing well or poorly and enables early intervention for students who are performing poorly, (ii) k-means clustering easily adapt to new observations which makes this model adaptable to new students' data, and (iii) clustering students into a set of disjoint groups makes it easier to interpret complex and heterogeneous students' data. Such a process enables finding similarities between students, drawing inferences, and finding hidden patterns among successful and at-risk students.

The main contributions and implications of the PAD model are twofold: (i) detecting big changes in students' data among several features can help in detecting relationships and correlations between those features. For example, we can infer that a student frequently attempting a lot of credit hours per semester, are likely to prevent them from achieving a high GPA or to force them to withdraw more credits hours. This, in turn, might prevent them from graduating on time, and (ii) identifying major changes in students' performance over time can help in defining a set of rules that can be used to assess if a student is potentially at risk of not graduating on time.

CHAPTER 6: EXPLAINABLE AND INTERPRETABLE INTERACTIVE STUDENT STORYTELLING

6.1 Overview

This chapter introduces the concepts and procedures performed to increase the advisors' trustworthiness of FIRST. To achieve this goal, this study adopts the fundamental approaches of eXplainable Artificial Intelligence (XAI); a research field that aims to make the outcome of an Artificial Intelligence (AI) system more understandable and interpretable by humans, either through introspection or through a generated explanation [23]. In other words, XAI is an AI system that can describe its purpose, behavior, and decision-making process in a way that can be easily understood by an average person.

To increase the advisors' trustworthiness of FIRST, the storytelling model needs to be more transparent to the advisor on how the stories are generated and how the contents of the story are selected from the student data model. In other words, making the storytelling model understandable and interpretable by advisors. In FIRST, this is achieved as follows:

- Giving the advisors the ability to select the features that they are interested in. These features are then used to decide the content of the student stories. This helps in generating students' stories that are customized and tailored to a particular advisor.
- Adding explanations to the dynamic parts (i.e., parts that are generated from student features) of the student stories to make them more understandable to advisors. An advisor can interact with the dynamic parts of the students' stories. This interaction shows a pop-up message that tells why this feature

is included in the story and other information about that part of the story. Adding an explanation to the generated stories provides additional information to expose the reasoning and data behind the generation process. This aims to increase the transparency, efficiency, user satisfaction, and trustworthiness of the storytelling model.

This chapter starts by presenting the different components of the story explanations in Section 6.2). Then, Section 6.3 presents the goals and objectives of these explanations in FIRST. These explanations are generated based on the content determination task in the storytelling model discussed in Section 4.3.2.1.

6.2 Explanation Components

Explanations in FIRST aim to expose the reasoning and data behind a system decision to include contents in the student story. To expose the reasoning behind the content of a story, FIRST describes the procedure to generate the story content. For instance, if the story content is about the students' clustering data analytics, FIRST tells the advisor that this piece of information is generated using a clustering algorithm that is performed on all CCI students. To expose the data behind a story, FIRST shows the advisor the features from the student's data model that are used to generate a sentence in the story. These features can be presented to the advisor using visual components like scientific charts- pie charts, bar charts, and line charts, or they can also be presented using tabular components like a table of student courses attempted, passed, failed, or withdrawn throughout their enrollment. Therefore, the story explanation composes four parts as shown in Figure 6.1. The process of selecting the different explanation components is based on the story sentence representation presented in Section 4.2. The explanation generation algorithm uses the parameters in the sentence representation to decide the different explanation components. These parameters include sentence label, relation, raw features, type, and engineered text. The following subsections introduce the explanation components one by one.

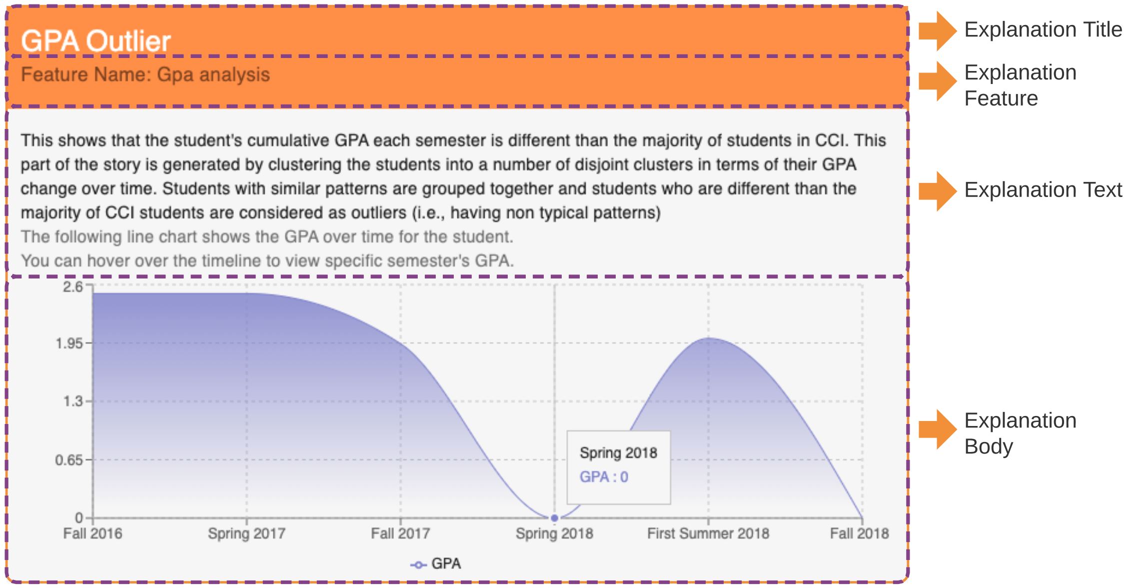


Figure 6.1: Story explanation components.

6.2.1 Explanation Title

The title of the explanation indicates the type of information in the story sentence. It shows the advisor what part of the student data model this piece of information belongs to. For instance, Figure 6.2 shows an explanation of the student's citizenship and the title tells the advisor that this information is part of the student's demographic information. This means that this information about the student is generated from the student background node in the student temporal data model. Other examples of explanation titles are "Student Number of Transferred Credits", "Student Credits Information", "Student Passed Credits Significant Change", "Student Academic Standing Information", "Student Graduation Status Information".

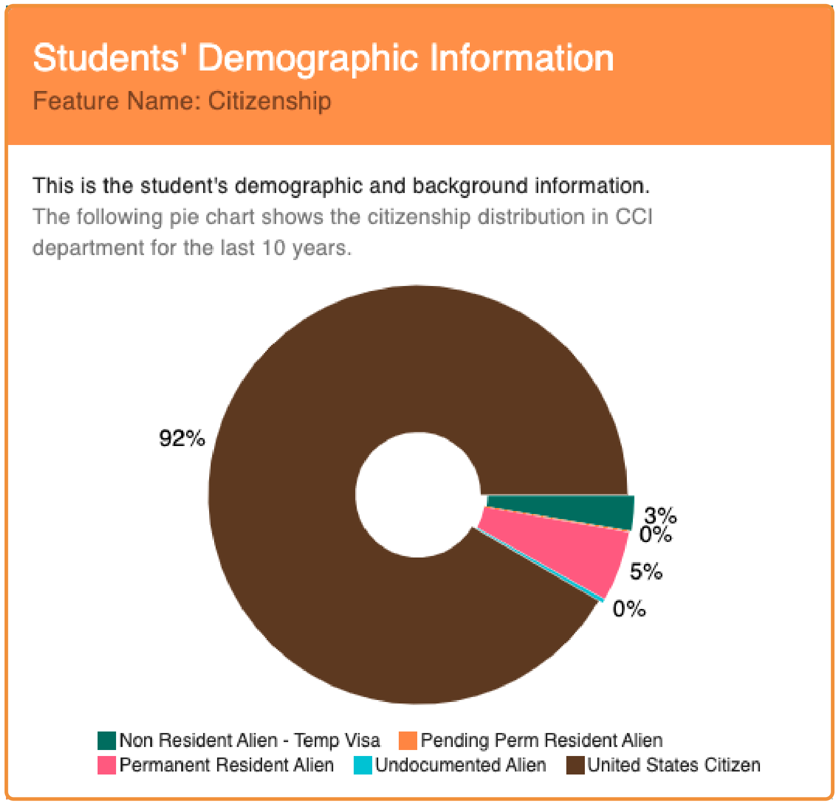


Figure 6.2: An example of a story explanation for the student citizenship information.

6.2.2 Explanation Feature

Rather than information about the complete sentence in the story, this component shows the advisor the feature that is used to generate the selected part of the student story. As shown in Figure 6.1, the feature that is used in this example is the student's GPA analysis from the student data model. Therefore, this part of the explanation could be any of the features in the student data model. This helps the advisor to decide what features from the student temporal data model to select for any particular student.

6.2.3 Explanation Text

This component shows the advisor a narrative explanation about the selected part of the student's story. This narrative could be a description of the selected part of the story or it could be a description of the procedure used to generate this part of the story. It also includes instructions on how to interact with the explanation body. For instance, the explanation text shown in Figure 6.3 is presented to the advisor for the sentence "Regarding Sarah's number of credits with a D grade, in her 2nd semester, in Spring 2020, it has significantly decreased 7 credit hours, from 7 to 0 credit hours". As shown in this figure, the explanation informs the advisor that this sentence is generated by comparing the change of student's number of credits with a D grade in comparison with all students in CCI.

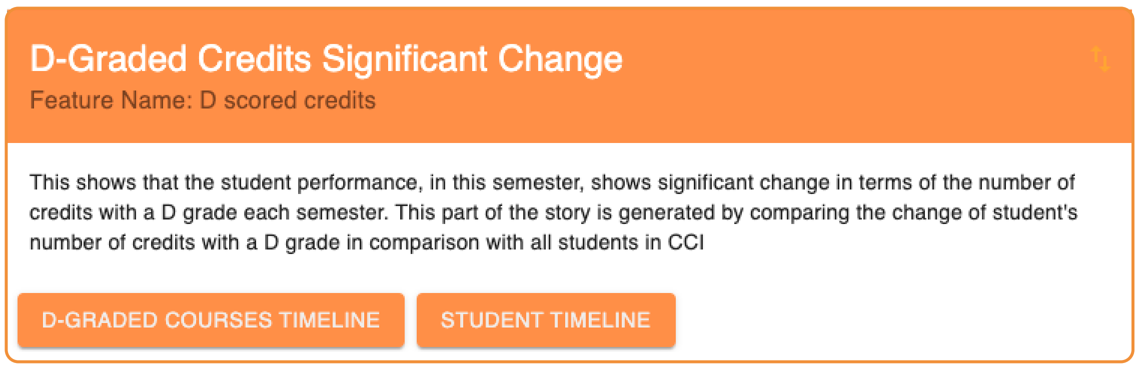


Figure 6.3: An example of a story explanation about the significant change of a student's number of credits with a D grade.

Another example of an explanation text is shown in Figure 6.4. This explanation is presented to the advisor for the sentence "Compared to other students in CCI, Sarah follows a non-typical pattern for the number of credits failed each semester, in which, during her seven enrolled semesters, she failed 0, 6, 0, 3, 6, 9, and 12 credit hours respectively". As shown in this figure, the explanation tells the advisor that this sentence is generated by clustering the students into a number of disjoint clusters in terms of the change of the number of credit hours failed over time. Students with similar patterns are grouped together and students who are different from the majority of CCI students are considered as outliers (i.e., having non-typical patterns).

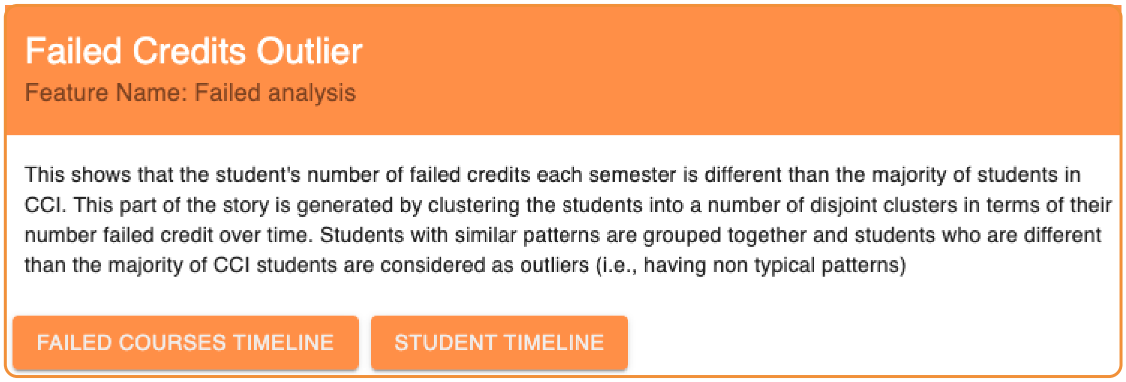


Figure 6.4: An example of a story explanation about a non-typical pattern in terms of a student's number of failed credit hours each semester.

It is worth noting that the explanation text describes the reasoning and procedure of generating a story content in nontechnical terms to make it easier and more understandable to advisors regardless of their levels of expertise.

6.2.4 Explanation Body

The explanation body aims to expose the data that is used to generate the selected story content. Two types of presentation are used to present the student data: visual and tabular presentation. The visual presentation includes scientific charts like pie charts, line charts, and bar charts. It also includes other kinds of visual components like cards and timelines. The tabular presentation includes tables of student data throughout their enrollment. The following two subsections introduce these two types of data presentation.

6.2.4.1 Visual Presentation

In this presentation style, data is displayed using either scientific charts or timelines. The scientific charts include pie charts which are used to visualize group distributions like CCI student gender, ethnicity, and citizenship distributions (see Figure 6.2). Scientific charts also include histograms (bar charts) to display CCI student age distribution (see Figure 6.5). Another kind of chart is a line chart which is used to display student data over time like a student's GPA (see Figure 6.1). Visual presentation could also be a timeline such as a student's academic standing change over time (see Figure 6.6).

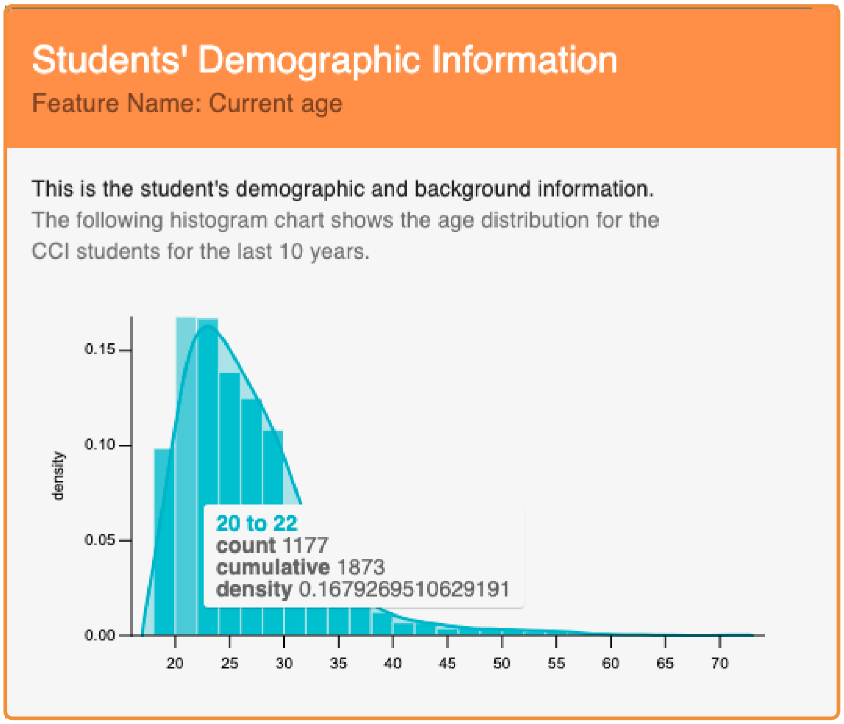


Figure 6.5: An example of a story explanation for the student age information.

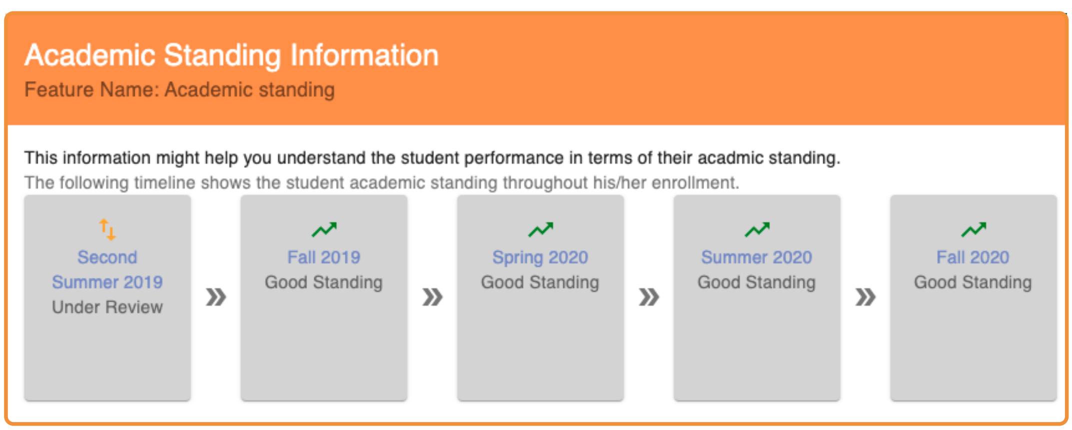


Figure 6.6: An example of a story explanation for the student academic standing timeline.

6.2.4.2 Tabular Presentation

This presentation style is used to display two types of information about the selected student as follows:

- Student timeline; which includes all features about the selected student throughout their enrollment. These features include GPA, Cumulative GPA, Academic Standing, Credits Attempted Count, Credits Passed Count, Maximum Required Credits, Minimum Required Credits, Major, Advisor Count, Primary Advisor Type, Current Time Status, Department, Enrollment Status, Expected Graduation Date, Fin Aid Applicant Indicator, Housing Indicator, Student Level, Student Population, Student Status (see Figure 6.7).
- Course timelines; which include student courses, course Identification number, the course title, course number of credit hours, and course grade (see Figure 6.8). Depending on the type of story content, this timeline of courses can be filtered to display all student courses, transferred student courses, passed student courses, failed student courses, withdrawn student courses, or student courses with a D grade.

Feature Name	Spring 2015	Fall 2015	Spring 2016	Second Summer 2016
GPA	4.0	4.0	3.7	1.0
Cumulative GPA	4.0	4.0	3.8	3.6
Academic Standing	Probation	Good Standing	Good Standing	Good Standing
Credits Attempted Count	6	15	19	3
Credits Passed Count	6	15	19	3
Maximum Required Credits	18	18	19	7
Minimum Required Credits	0	0	0	0
Major	Computer Science	Computer Science	Computer Science	Computer Science
Advisor Count	2	2	3	3
Primary Advisor Type	Major Advisor	Major Advisor	Major Advisor	Major Advisor

Figure 6.7: An example of a student timeline.

11 Spring 2015

Course ID	Course Title	Number of Credits	Grade
ANTH1101	Intro to Anthropology	3	TB
INFO2130	Intro to Business Computing	3	TA
COMM1101	Public Speaking	3	TC
ECON2102	Principles of Economics-Micro	3	TB
UWRT1101	Writing & Inquiry in Acad I	3	TB
ENGL2104	Major American Writers	3	TA
ESCI1101	Physical Geography	3	TA
ESCI1101L	Physical Geography Lab	1	TA
HIST1121	European History since 1660	3	TC
ITCS3160	Data Base Design & Implem	3	A
ITCS2214	Data Structures	3	A

Figure 6.8: An example of a student's courses timeline.

6.3 Story Explanation Goals

Explainability refers to the algorithms that try to answer the question of "Why a decision is made by the system?". These algorithms do not only produce results or outputs for the users but also explain why such results are produced. Explanations of the system's decisions can serve multiple aims, such as exposing the reasoning and data behind a decision. Other aims of an explanation include increasing the users' trust and confidence in the system's decision, persuading the user to accept the decision, making it easier and faster for the user to find the most relevant information they want, and increasing user satisfaction. Explainable systems play an important role in enhancing the user experience [78].

This dissertation draws on the body of work on explainability in the literature to identify the main goals of explanations in FIRST. There are seven main goals of explanations in FIRST. The following subsections introduce these goals and how they interact with each other to improve the advisors' and faculty leaderships' experience when interacting with the student stories.

6.3.1 Transparency

Transparency in FIRST aims to explain how the system works by exposing the reasoning and data behind a decision to include a piece of specific information in the student story. For example, the story explanation shown in Figure 6.9 shows several useful information to the advisor about a sentence in the story: First, it shows the feature from the student data model that is used to generate the sentence (i.e., GPA). Second, it shows the student data about this feature throughout the student enrollment (i.e, the line chart). Third, it shows an illustration of how this content of the story is generated (i.e., This part of the story is generated by comparing the change of student's GPA in comparison with all students in CCI).

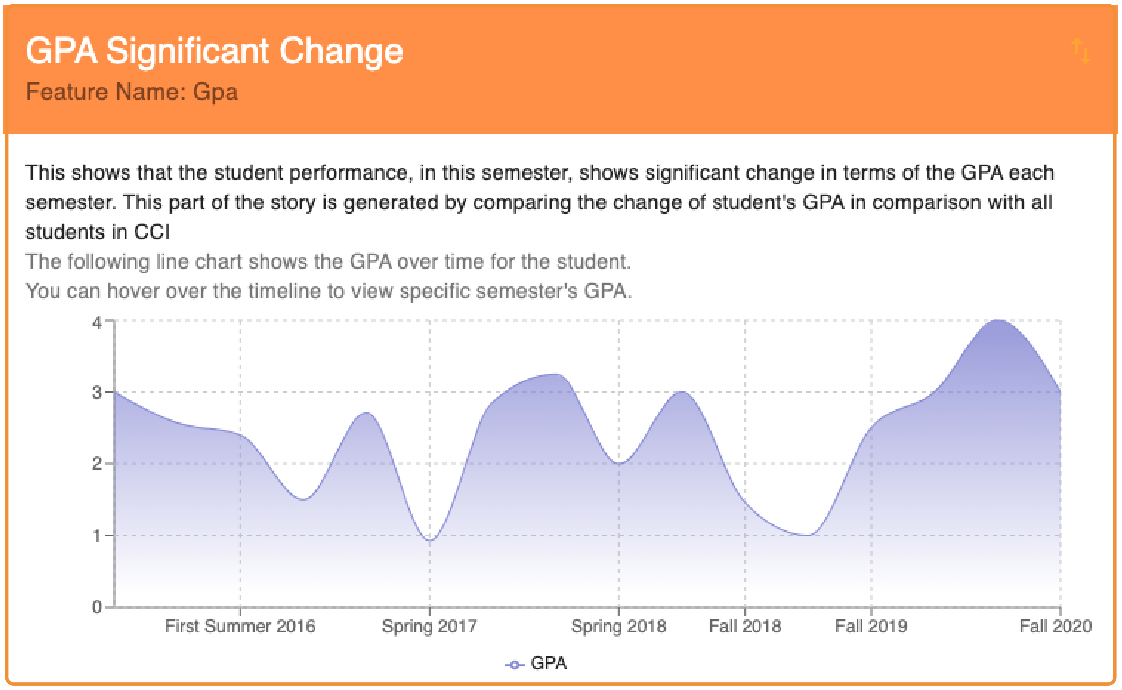


Figure 6.9: An example of an explanation in FIRST for the GPA significant change.

6.3.2 Trustworthiness

Trustworthiness means increasing the advisors' confidence in the story content. Story trustworthiness is linked with the transparency of FIRST, in which the transparency of FIRST and the advisors' ability to interact with the student features to influence the story content determination increases the advisors' trust in FIRST.

6.3.3 Scrutability

Scrutable systems refer to the system that enables the user to understand the system by careful study and investigation of the system. The Scrutability of FIRST is built on transparency and can be achieved by allowing advisors to interact with the dynamic parts of the student stories to expose the reasoning and data that are used to generate those parts of the story.

6.3.4 Effectiveness

Effective explanation increases the advisors' ability to ignore irrelevant or uninteresting story content while assisting them to choose the relevant features for the student story. For example, if an advisor finds that knowing about a student's citizenship type or ethnicity is not helpful in advising, they can ignore selecting such features from the student data model for future advising sessions. Effective explanations can also introduce new insights to new advisors by helping them to understand the full range of features about the student.

6.3.5 Persuasiveness

Persuasiveness means convincing the advisor to receive and accept the story content. Story explanations might increase the advisor's acceptance of the story content. For instance, if there is a sentence in the student story that says "Sarah follows a non-typical pattern for the number of credits failed each semester". In this sentence, an explanation, that aims to convince the advisor that the student pattern of credits failed each semester is not typical, is shown in Figure 6.10. In this figure, the explanation aims to expose the reasoning of why this student follows a non-typical pattern for the number of failed credits each semester. The explanation tells the advisor that the student's number of credits failed each semester is different when compared to other typical students. The explanation also gives the advisor the ability to expose the student data that supports the claim in the sentence by giving the advisor the option to view the student's failed courses timeline.

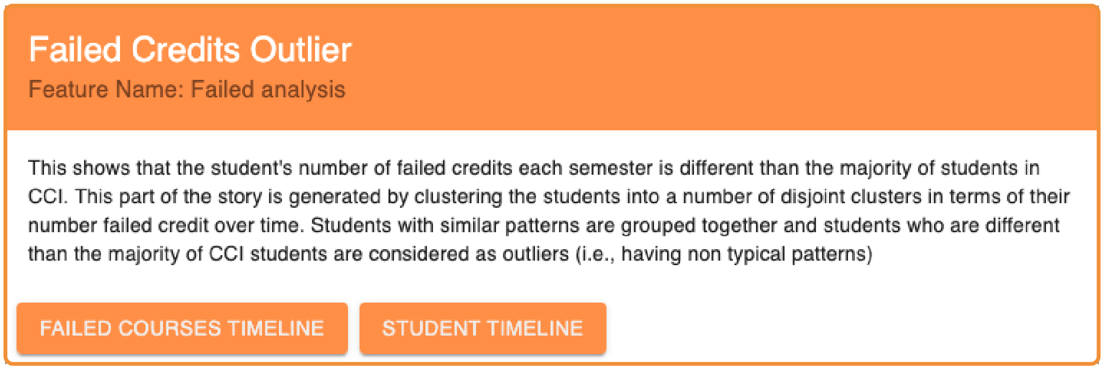


Figure 6.10: An example of an explanation in FIRST for the failed credit hours.

6.3.6 Efficiency

Story explanations may help advisors make decisions faster, by providing them with the information needed to help them make informed decisions when advising students. Efficiency can be increased by letting the advisor understand the relation between student features and the story that is generated from those features.

6.3.7 Satisfaction

Story explanations make the advisor more satisfied with FIRST by making it easier and more enjoyable to use the system. As discussed earlier, the perceived transparency by the user is positively correlated with the user satisfaction with the explanation interfaces. Therefore, providing detailed explanations for the story content is positively correlated with the advisor-perceived usefulness and the simplicity and ease of use of FIRST.

6.4 Summary

This chapter introduces the concepts and procedures implemented to make FIRST more understandable and interpretable. Explainability and interpretability of FIRST can serve multiple aims, including exposing the reasoning and data behind a decision, increasing the advisors' trust and confidence in the system's decision, persuade the advisor to accept the decision, make it easier and faster for the advisor to find the most relevant information they need to advise their students, and to increase advisors' satisfaction with the generated stories. Generally, making FIRST an explainable system plays an important role in enhancing the advisor experience.

Two main approaches are used to make FIRST explainable and interpretable: (i) giving the advisors the ability to select the features that they are interested in to be included in the students' stories. (ii) adding explanations to the dynamic parts of the student stories to make them more understandable to advisors. These explanations include four main components: (i) explanation title, (ii) explanation feature, (iii) explanation text, and (iv) explanation body. Each component serves a task in enhancing the advisor's experience when using FIRST.

CHAPTER 7: EVALUATION STUDIES

7.1 Overview

This chapter presents the evaluation methodologies used to evaluate FIRST. Stories generated using FIRST are evaluated using intrinsic and extrinsic evaluations. The intrinsic evaluation assesses the generated story in terms of language quality, coherency, fluency, and fidelity of the generated story text. This dissertation evaluates the stories generated by FIRST using online evaluation by presenting stories to participants and asking them to answer questions like "does the story text read naturally?", "Does the story text have clarity?". The main focus in this dissertation is the extrinsic evaluation which assesses to what extent FIRST is capable of generating stories that support the purpose of the advisors. In other words, to what extent do the student stories generated by FIRST help the advisor to make sense of student data.

This chapter starts by presenting the demographic characteristics of the participants who are recruited for all user studies- who are they, what roles do they have in student advising, what current experiences do they have, and what tools are they currently using. Then, this chapter presents the evaluation methodologies used to evaluate FIRST. The methodologies that are considered in this dissertation include: First, a focus group study using version 1.0 of FIRST to evaluate the impact of the student storytelling model in the LA. Second, a one-on-one interview study using version 2.0 of FIRST to identify the critical features of the student storytelling model in LA. Third, a diary study using version 3.0 of FIRST to contextually understand advisors' experiences with the student storytelling model. Finally, a focus group study using version 4.0 of FIRST to evaluate the explainability of the student storytelling model. The following subsections present these four evaluation methods.

7.2 Demographic

All studies in this thesis include academic advisors from the CCI at the University of North Carolina at Charlotte. These advisors are committed to quality academic advising for all students in CCI and they comprise professional and faculty advisors. Professional advisors are believed to play essential roles in student retention, progression, and timely graduation of CCI students. Faculty advisors teach courses as well as help students create and achieve their academic goals. Beyond helping students develop an academic plan, faculty advisors are in a position to engage students in the academic environment and educate them on research and career opportunities. Both professional and faculty advisors are responsible for: (i) providing accurate and timely information about degree and career-related requirements, (ii) empowering each student to make independent and informed decisions, (iii) being knowledgeable about policies and procedures of the college, (iv) serving as a guide, teacher, facilitator, coach, and counselor, (v) encouraging active engagement in the curriculum-based advising process by using degree audit tools, (vi) advising from an integrated perspective of general education, major(s), minor(s), experiential learning, study abroad, (vii) ensuring a smooth transition for students declaring and changing majors, (viii) providing realistic options for students' decision making and encouraging reasonable time to degree.

Advisors in CCI are already familiar with multiple tools that provide data, analytics, and risk scores for the students that they advise. These tools include (i) DegreeWorks, which is a degree audit tool that allows students and their advisors to view progress toward a degree based on the catalog year of a degree, major, concentration, or minor. (ii) Connect; which is an academic early alert and advising software system. It allows advisors to send systematic notifications to students regarding their academic progress in their courses referred to as at-risk alerts. (iii) Banner; which is a system used by faculty and advisors to view students' grades and financial aid status, and advise students. It includes information about the students like: registered for the current term indicator, first time attended date, last term attended date, residence status, student citizenship, student type, student class, student's advisors' names, student's advisors' types, and student's expected graduation date.

7.3 Focus Group Study to Evaluate the Impact of FIRST in LA

7.3.1 Overview

The purpose of this focus group study is to evaluate the claim that storytelling extends and complements existing approaches to interactive learning analytics. The goal of this user study is to demonstrate FIRST and explore how storytelling complements existing data and dashboard-style tools for advisors. The study includes professional and faculty advisors since they are already familiar with multiple tools that provide data, analytics, and risk scores for the students that they advise. A focus group was selected for this study because it is effective in collecting user opinions and attitudes through group discussion and conversation dynamics. Compared to one-on-one methods such as interviews or surveys, the focus group study results in richer and varied insights because listening to others' experiences stimulates memories, ideas, and experiences in participants.

Six advisors were recruited for the focus group study. Three were professional advisors. The other three participants were faculty advisors. The participants comprised two females and four male advisors. The evaluation of this study was designed to test the storytelling model and gather insights about features that domain experts look for.

7.3.2 User Experience Design

This section presents the user experience design of FIRST version 1.0 for this user study. The user experience aims to provide the domain experts the following functionalities:

- The ability to select features from the student temporal data model that they find useful for their advising sessions.
- The ability to change the feature selection for each student.
- The ability to view aggregate analytics (Course Number Means Through Semesters) about students at the group level.
- The ability to view statistical information about a cohort of students (e.g., student group average GPA, standard deviation, minimum and maximum GPA)
- The ability to select students to view their experience at the individual level.
- The ability to view an automatically narrative story about the selected students.
- The ability to view an interactive timeline of the student per semester and view information about each of the semesters.

Based on these functionalities, the user experience includes three pages: students data model, student aggregate analytics, and student stories pages. These pages are presented in the following subsections.

7.3.2.1 Student Data Page

The student data page is shown in Figure 7.1. This page includes instructions on how to select student features as shown in Figure 7.1(a). Also, on this page, the student data is arranged according to the student temporal data model (i.e., background, semester, and outcome) as shown in Figure 7.1(b). The domain expert can select the data features that they are interested in. Their feature selection is used when generating the student stories. At the bottom of this page, there is a non-sequential navigation menu as shown in Figure 7.1(c). This navigation menu is accessible from all pages of the application, which enables the user to go back and modify the settings at any time. Upon finishing the feature selection, domain experts can navigate to the aggregate analytics page.

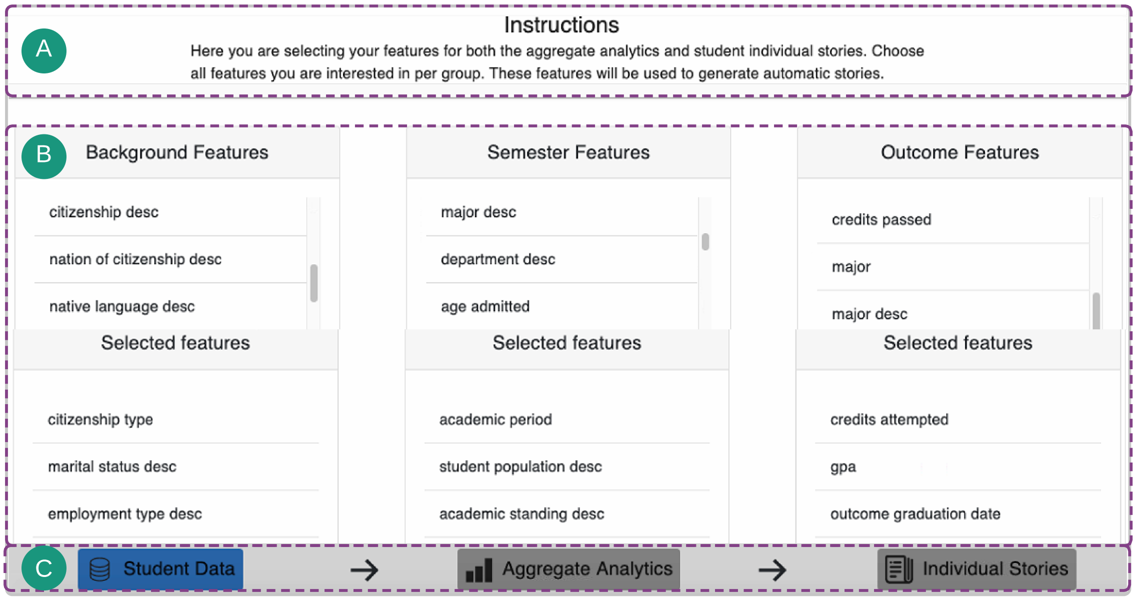


Figure 7.1: FIRST version 1.0 student data page.

7.3.2.2 Aggregate Analytics Page

The aggregate analytics page is shown in Figure 7.2. On this page, the domain expert can view groups of students that have similar patterns for several temporal engineered features. The instructions about how to select different temporal engineered features and view their visualizations are at the top of the aggregate analytics page as shown in Figure 7.2. Two types of visualizations are presented on this page, the first visualization as shown in Figure 7.2(a) shows clusters of students based on the selected temporal engineered feature. The second visualization as shown in Figure 7.2(b) shows the average GPA per student cluster. These visualizations are included to enable a comparison of the domain expert's sensemaking through visualization versus their sensemaking through storytelling. Finally, at the bottom of this page, a table of all CCI students is presented as shown in Figure 7.2(c). The domain expert can select the student they want to know more about and then go to the student stories page.

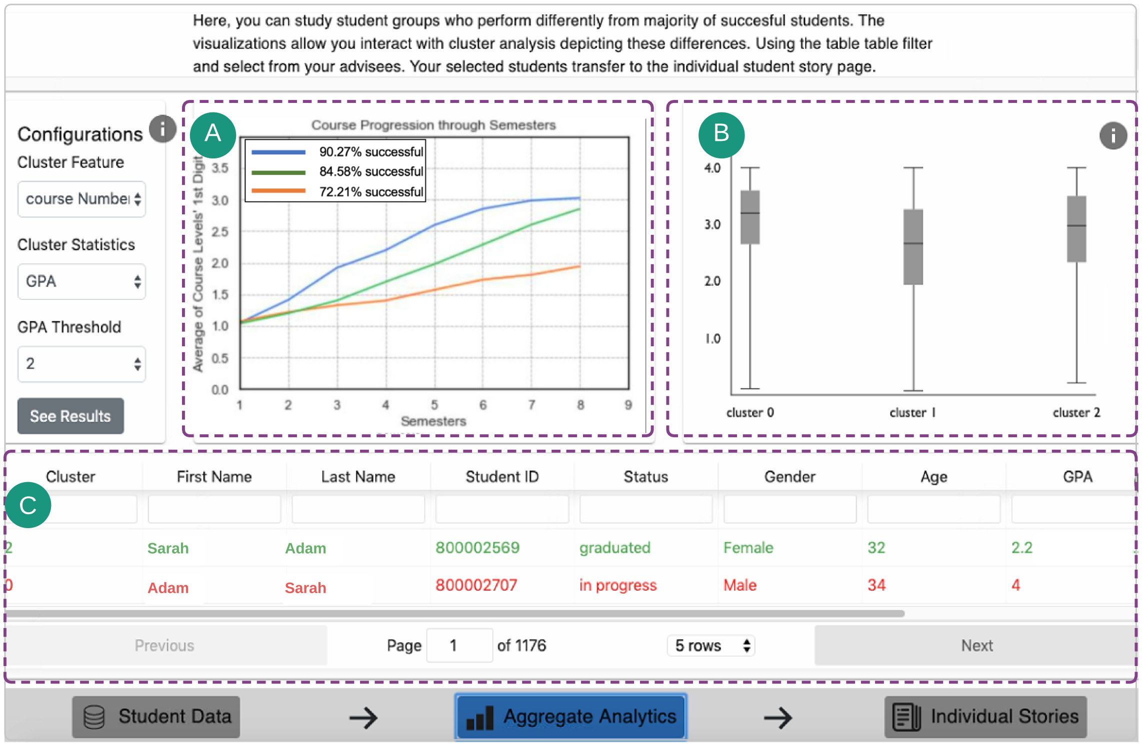


Figure 7.2: FIRST version 1.0 aggregate analytics page.

7.3.2.3 Student Stories Page

The student stories page is shown in Figure 7.3. On this page, the domain expert can view an automatically generated story for the students who are selected from the aggregate analytics page. The instructions about how to interact and select students to view their stories as well as their timeline are at the top of the student stories page. The selected students' names from the aggregate analytics page are presented in the selected students' panel at the left as shown in Figure 7.3(a). By clicking on the student name from this panel, the domain expert can view the student stories as shown in Figure 7.3(b). Also, the domain expert can view an interactive timeline of the student per semester and view information about each of the semesters as shown in Figure 7.3(c). At any time, the domain expert can use the navigation menu at the bottom of the page to go back and change their features and student selection.

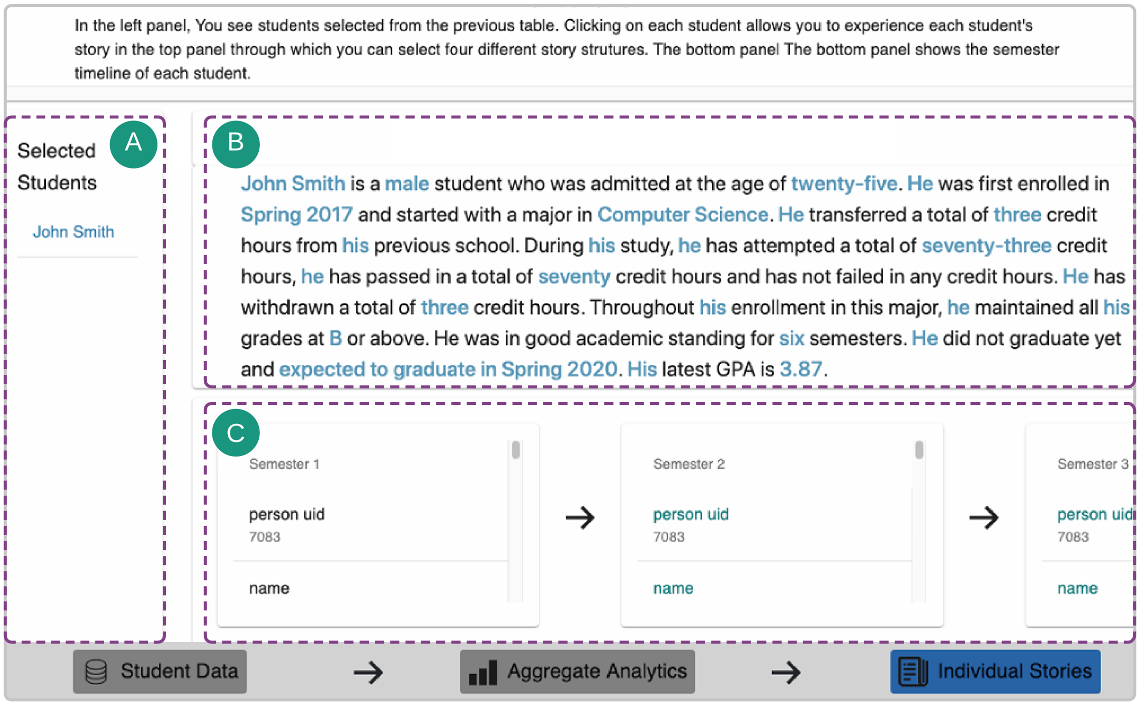


Figure 7.3: FIRST version 1.0 student stories page.

7.3.3 Study Design

The focus group study is designed to take almost an hour in total. It is divided into four parts: preliminary discussion, initial questions, demonstration, and follow-up questions. Throughout these parts, the participants were involved in verbal discussions. The study facilitator is the proxy for interacting with the system. There are no direct interactions with the FIRST system required from the participants. The following subsections present these parts of the study one by one.

7.3.3.1 Preliminary Discussion

The study starts with an overview of the project and the study. Then, the study facilitator presents an explanation of the student temporal data model and the student storytelling model. The facilitator presents how the student stories are generated from the student temporal data model. The explanation includes a discussion about the students' aggregate analytics and how it is performed on the student data.

7.3.3.2 Initial Questions

Following the preliminary discussion, a series of questions are administered to know about the previous experience of the participants in terms of advising. The first question was "what are some current tools you use during advising?", in which three participants mentioned Connect as their current tool, while some other participants mentioned DegreeWork and Banner. The second question asked was "what is useful about current tools?", in which one participant mentioned that a useful feature of DegreeWork is the ability to know students' course progression in terms of credit hours they already took and what they still have to take. Also, participants find that having multiple sources of information (like students' courses, GPAs, etc.) in one place is a useful feature of the current advising tools. Another participant finds at-risk reports based on success markers as a useful feature of Connect. The third question asked was "what could be improved in current tools?", in which one participant stated that although the student reports generated by Connect are very useful, it would be better if it provides more flexibility to the advisors and gives them the ability to customize the information it provides about students. For example, the number of students who took a particular course and did not do well or failed multiple times in this course, the number of transfer students, and the information about students who spent more than five years and still have not graduated. The last question was "What are some questions about students you usually try to answer before/during advising?", in which one participant mentioned reaching out to students who applied to graduate and look at their records and ask if they are missing any requirements.

7.3.3.3 Demonstration

Following the initial questions, FIRST was demonstrated by the study facilitator with some scenarios for several students from CCI. The scenarios include successful students who are doing well towards their degree and at-risk students who are struggling to complete their degree or underperforming. All the data used in this focus group is real students' data from the CCI. In this part of the study, the user experience presented in Section 7.3.2 is demonstrated to the participants.

7.3.3.4 Follow-up Questions

After the demonstration, the participants asked questions about the system and the study facilitator demonstrated additional interactive features of FIRST. Upon finishing the demonstration, participants are asked the following questions.

- What insights were you able to gain through viewing this tool?
- What are the differences between what you learned about the students from the analytics versus the stories?
- Explaining student success or risk is typically done with charts and graphs. In our tool, we also explain the student data using stories. Which one did you prefer and why?
- What is the value of the analytics results?
- What is the value of the student stories?
- How can the student stories help you with advising?
- Can you think of other features that would be good predictors of student success?

7.3.4 Qualitative Thematic Analysis

To analyze the focus group, the audio recording of the follow-up questions is transcribed and then a thematic analysis is performed. Figure 7.4 shows a word cloud of the transcribed audio recording. Then, the transcription is divided into segments. These segments' boundaries start when a participant starts talking about a particular concept and end when another participant starts talking about another concept. A revision is performed on the obtained segments to make sure that each segment includes only one concept. For instance, in a segment, if a participant starts talking about the student temporal data model, and then continues talking about the student stories, then the transcription for this discussion would belong to two segments. One segment includes sentences about the student temporal data model, and the other includes sentences about the student stories.

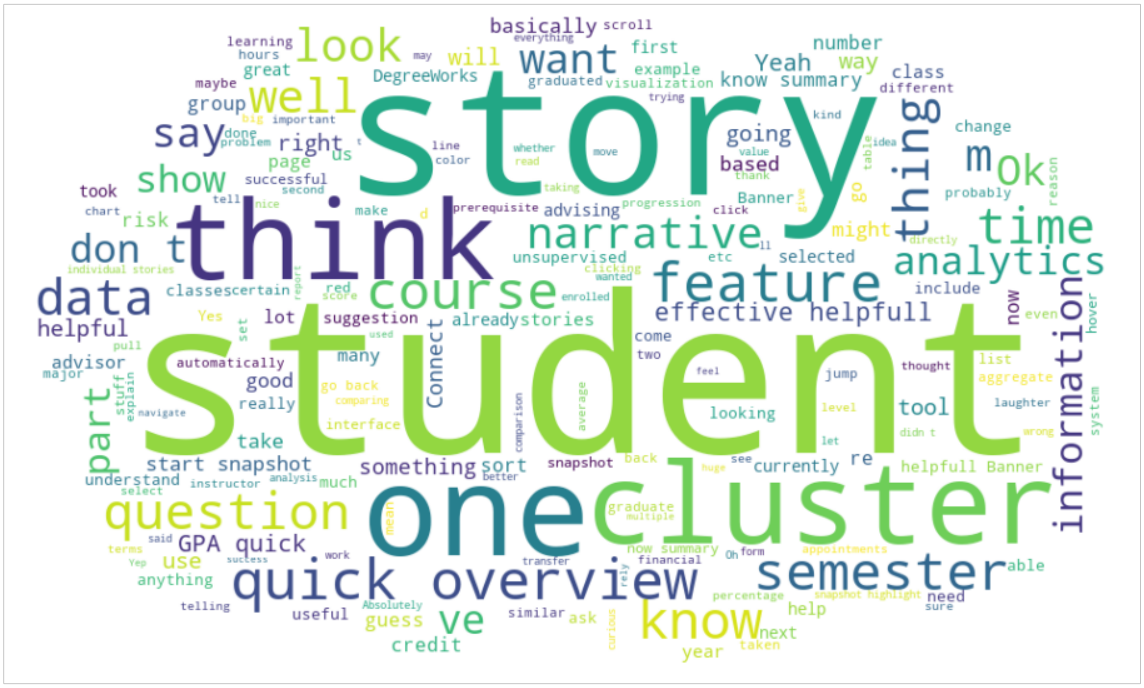


Figure 7.4: A word cloud visualization of the transcribed audio recording of the focus group follow up questions.

Upon finishing the segmentation of the transcription, each segment is coded. Initially, the codes are chosen based on the research questions about the storytelling model. Further, three categories are developed: (1) understanding student story contents, (2) research questions 1.1, 1.2, 2.1, and 2.2, and (3) future research. Table 7.1 shows the codes that are used for labeling segments. For each category, a number of codes are defined as shown in Table 7.1. A segment can be labeled by one or more codes. The codes are based on the concepts discussed in each segment. Therefore, these codes can be used to quantify the concepts discussed in each segment. For example, counting how many times participants talked about "student stories" or "aggregate analytics". Segments that contain questions asked by the study facilitator are not coded and excluded from the analysis.

Table 7.1: Interviews follow up transcription codes, categories, use cases, and examples.

Code	Category	Use cases	Segment example
Clarification /Question	Understanding the features of FIRST	When a participant asks a question or the focus group leader clarifies a concept, or answers a particular question.	"Can I look for things like how many of my students have taken 100 credits or more, and actually pass them?"
Story effectiveness	Research question	When a participant refers to the usefulness and effectiveness of student stories	"We understand stories so much better than many other things. The stories speak to me"
Actionable knowledge discovery	Research question	When a participant could discover actionable knowledge from the student story	"Just from the narrative. 25 credits passed and 145 were done. She should be gone... She should have graduated already."
Visual Analytics Issues	Research question	When participants identify or refer to an issue related to visual analytics.	"Without an illustration, the cluster is not useful... It's not useful for me to see the clusters at all"
Feature Selection	Research question	When a participant refer to or suggest new features that can be added to the stories	"Withdrawal reason would actually be helpful as a part of the story because it could be financial reasons, it could be grade reasons."
Suggestion	Future Research	When participants suggest changes or improvements to the current work or propose new approaches in analyzing student data.	"I like having that story just to get a big picture up front and then if there's also graphs and more details that we can look at later, that's fine, but I like having that upfront"

In some situations, a single code does not capture the concepts discussed in a segment, those segments are revised and multiple codes are assigned. For instance, a segment about "story effectiveness" is also related to "future work". One may talk about the effectiveness of the student stories for advising, while another may talk about future work that adds new student data to the student stories. Therefore, this segment is split into two segments with a single code for each segment. Based on this the total number of segments is 74 segments.

Coding the segments allows quantifying the concepts discussed in the focus group follow-up questions by counting the occurrences of each code. Figure 7.5 shows the distribution of codes in the focus group follow-up questions. Based on this figure, the discussions are more directed towards the effectiveness of student stories (27%) and the discovery of actionable knowledge about students viewed throughout the demonstration scenarios (23%). A notable amount of the discussion is directed towards questions and issues in the visualization components (15%). This finding supports the research questions about presenting complex and heterogeneous student data, which is the effectiveness of student stories in understanding student data. This finding also supports the research questions about the discovery of actionable knowledge from student stories that can guide further decisions.

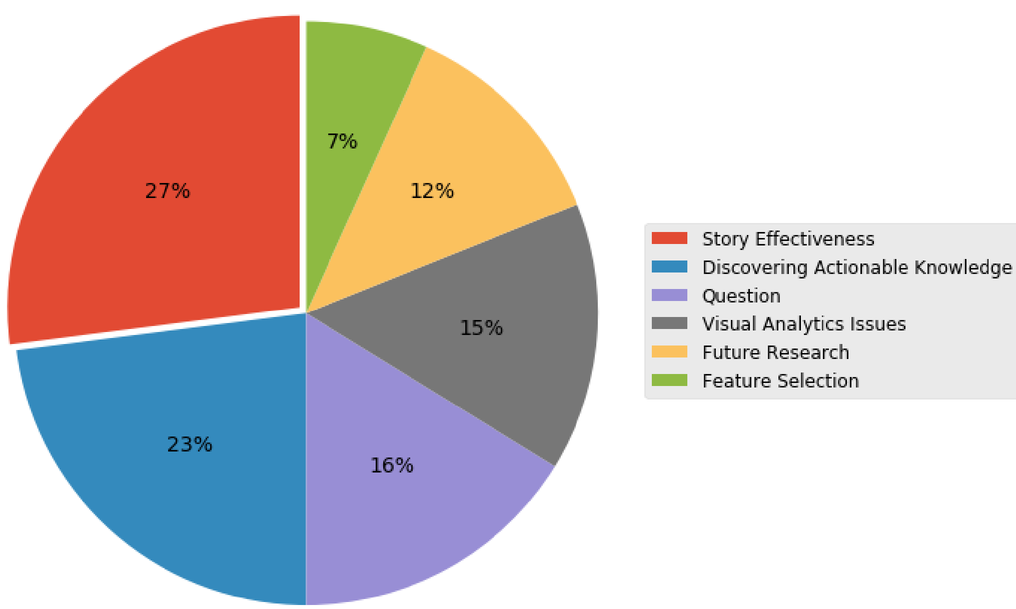


Figure 7.5: Distribution of the codes in the entire focus group.

7.3.5 Summary and Contributions

The result of the focus group follow-up questions analysis is four high-level themes. The following subsections present these themes and answer some of the research questions asked based on the thesis statement.

7.3.5.1 Effectiveness of Student Stories

This theme answers the research question RQ1.1 (Do the student stories provide an effective way in presenting complex and heterogeneous student data to domain experts?). In which most of the participants expressed that student stories are useful and effective. They expressed that the student stories provide a high-level overview or snapshot of the student. One participant stated: *"I like having that story just to get the big picture upfront".* One participant mentioned that the stories are helpful when advising many students in a sequence of meetings since it is faster to read than other ways of data presentation. *"Since all of us advise a lot of students, it is hard to remember each student sometimes, so having that student story there will refresh your memory about the student".*

7.3.5.2 Discovery of Actionable Knowledge

Student stories help domain experts to discover actionable knowledge from the student stories. This theme answers the research questions RQ2.1 (Do the student stories provide an effective way in presenting complex and heterogeneous student data to domain experts?) and RQ2.2 (What insights do domain experts learn from student stories?). For RQ2.1, participants agreed that stories have provided a good understanding of students in terms of their demographic information as well as their academic performance. One participant said about the entire FIRST experience: *"I like the stories the best - knowing that the story was created using analytics is reassuring".* For RQ2.2, the stories helped discover insights and understand the student quickly, with an explanation for that understanding. For example, one participant concluded quickly after reading a student's story. The participant said: *"Just from the narrative. 25 credits passed and 145 done. She should be gone. She shouldn't be here. She should have graduated already".* Another participant said, *"You can see that something went wrong for this student. This student changes majors 3 times, it may be that they need more credits because they took a bunch of credits in a major that doesn't help them, and that might delay their graduation".* Another example of actionable knowledge by the domain experts when a participant said: *"This student is a freshman. Just based on his age as mentioned in the story. I don't need to go and look for this by myself. It's good to have such kind of information in the student story."* Another example is when a participant said: *"This student has withdrawn from a lot of credit hours as shown in the narrative. This is a good indicator that this student might be at risk. This really helps me to understand that this student is not doing well."*

7.3.5.3 Visualizations Usability Issues

Participants expressed confusion about some of the visualization components. For example, one participant states *"I had trouble understanding the figures at the top. And also the table at the bottom. ... computer scientist advisors want to see the raw data and can slice and dice on their own".* Another participant said: *"I actually struggled to understand, I guess, the course level progression diagram you were trying to explain earlier. Or when you explained it, what is the difference between, I guess, the green, blue, and red".* The participants suggested that adding some kind of hover over some parts of the diagrams to show more information makes it more understandable. Another participant had similar thoughts by stating: *"my problem is actually not with the visualization, it's with just knowing what each cluster means, you know? Without that meaning, the cluster is not useful. It's not useful for me to see the clusters at all".* A suggestion from the focus group was to include a storytelling feature to describe the dashboard components such as line graphs, bar charts, and cluster illustrations.

This theme in conjunction with the first theme (Section 7.3.5.1) answers the research question RQ1.3 (Which presentation style do domain experts prefer to make sense of complex and heterogeneous student data?). For instance, one participant said: *"It takes way less time to read the student story than it does to like, scroll through DegreeWorks and figure out like, ok where is this student. Go back and look at the notes in Connect."* Another participant said: *"I like the stories the best. Of all of them. Because the computer scientists in the room probably understand the analytics a lot better than I do."* One participant who refers to the value of the student stories said: *"I've done research in this area and we understand stories so much better than many other things. The stories speak to me."*

7.3.5.4 System Improvement Suggestions

Participants also suggested some improvements that could be added to the system to make it more helpful and useful for advising. One participant suggested that the student stories can be enriched by adding some visual components for more details about the student. The participant said: *"I like having that story just to get a big picture upfront, and then if there's also graphs and more details that we can look at later, that's fine, but I like having that upfront."* Other participants suggested some features about the student that could be added to the student stories. For instance, student course withdrawal reasons. The participant said: *"Withdrawal reason would actually be helpful as a part of the story because it could be financial reasons, it could be grade reasons."* Other features include student housing information and mailing addresses.

7.4 Interview Study to Identify the Critical Features of FIRST in LA

7.4.1 Overview

This user study is to address the third research question. This user study has two main aims. First, finding the building blocks of student stories that are meaningful for the domain experts in terms of story content and story structure. For the story content, this user study attempts to find what content and how much content and special cases for determining content, check whether the information provided in the students' stories is sufficient to give the domain expert the insights they want to know about the students. For the story structure, this user study attempts to find which story structuring is more preferable by the domain experts. For this purpose, three-story alternatives are to be presented for each student to the advisors. One alternative is the default story that has basic information about the student. Another alternative is based on the student temporal data model, in which the story starts with the background information about the student, then with the semester information, and ends with the outcome information. And the last alternative starts the story with the outcome information first and followed by the student's background and semester information. Second, evaluating domain experts' sensemaking through interaction. In other words, evaluating the affordances for interactivity of the stories that assist the sensemaking process, in which advisors can select features from students' data that they are interested in. Information about the selected features is reported in the students' stories.

One-on-one Interviews are a method of data collection that involves two people exchanging information through a series of questions and answers. Interviews are designed to collect a richer source of information from a small number of people about their attributes, behavior, preferences, feelings, attitudes, opinions, and knowledge. A one-on-one interview study was selected for this study because it is an efficient way to gather detailed information from participants. Interviews also have an advantage over other types of user studies like surveys; with a survey, if a participant's response sparks some follow-up question in the researcher's mind, the researcher generally does not have an opportunity to ask for more information. In an interview, however, because the researcher is talking with his/her study participants in real-time, they can ask that follow-up question. Thus interviews are a useful method to use when you want to know the story behind responses you might receive in a written survey. Interviews help explain, better understand, and explore research subjects' opinions, behavior, experiences, phenomenon, etc. Additionally, interview questions are usually open-ended questions so that in-depth information will be collected.

Sixteen advisors from the CCI were recruited for this one-on-one interview study. Ten were professional advisors and six were faculty advisors. The participants comprised seven females and nine male advisors.

7.4.2 User Experience Design

This section presents the user experience design of FIRST version 2.0 for this user study. Some changes have been made to the user experience of FIRST version 1.0 presented in Section 7.3.2. These changes include:

- Adding more student features to the student temporal data model. These features are suggested by the participants in the focus group study discussed in Section 7.3. For example, the student's financial aid information, the student's age admitted, and the student's current age.
- Adding three alternatives of story structures for the advisors to select the one they prefer as shown in Figure 7.6(b). These structures are: (i) default story which has basic information about the student, (ii) temporal story that starts with the background information about the student, then with the semester information, and ends with the outcome information, and (iii) outcome story which starts the story with the outcome information first and followed by the student's background and semester information.

- Adding eight more engineered features along with their visualizations so advisors can select the engineered feature they are interested in. These engineered features are:
    1. Total number of skipped semesters
    2. Course withdrawal throughout semesters
    3. Percentage of A grades throughout semesters
    4. Percentage of B grades throughout semesters
    5. Percentage of C grades throughout semesters
    6. Percentage of D grades throughout semesters
    7. Percentage of F grades throughout semesters
    8. Transferred courses throughout semesters

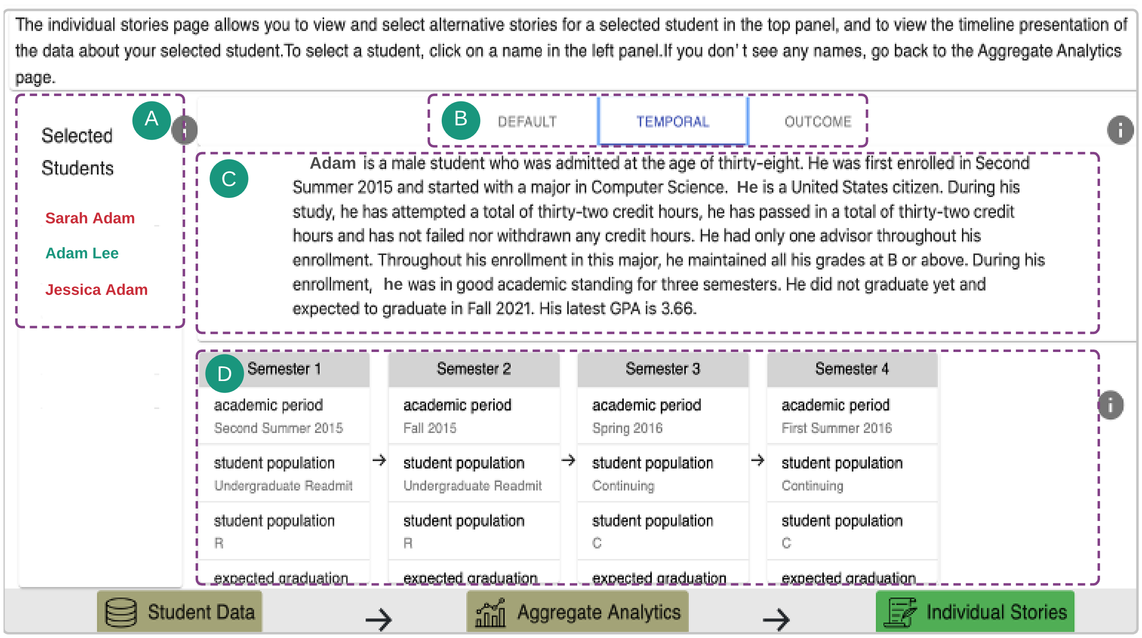


Figure 7.6: FIRST version 2.0 student stories page.

7.4.3 Study Design

The interview study is divided into five parts: preliminary discussion, initial questions, demonstration, user interaction, and follow-up questions. Throughout these parts, the participants were involved in verbal discussions with the study facilitator.

Participants are asked to directly interact with the FIRST system during the user interaction part of the study. The following subsections present the parts of the study one by one

7.4.3.1 Preliminary Discussion

The study starts with an overview of the project and the study. Then, the facilitator presents a brief explanation of the student temporal data model and the student storytelling model. The facilitator also presents how the student stories are generated from the student temporal data model. The explanation also includes a discussion about the students' aggregate analytics and how it is performed on the student data.

7.4.3.2 Initial Questions

Following the preliminary discussion, a series of questions are administered to know about the previous experience of the participants in terms of advising. The first question was "How many students do you advise each semester?", in which 2 participants advise more than 150 students each semester, 4 participants advise 100 students on average, 2 participants advise between 80 to 90 students each semester, 4 participants advise between 50 to 80 students each semester, and 4 participants advise less than 50 students each semester. The second question was "How long do you spend with each student?", in which two participants stated that they spend about 30 minutes with each student, the rest of the participants stated that they spend between 5 to 20 minutes with each student. The third question was "How much time do you have available to learn about the student before an advising session?", in which all participants mentioned that they have between 5 to 10 minutes for each student. The fourth question was "What information do you want to know about a student before advising them?", in which all participants mentioned that they want to know about a student's GPA, course information like courses taken, retaken courses, prerequisites, and courses to take next semester. 6 participants mentioned that they want to know previous notes about the students. Also, 6 participants mentioned that they want to know if the students have troubles, issues, red flags, or personal circumstances. Finally, 4 participants mentioned that they want to know about the student's previous education, like if the student came from a community college, university, or high school.

7.4.3.3 Demonstration

Following the initial questions, FIRST was demonstrated by the study facilitator with scenarios for two students from CCI. The scenarios include one successful student who is doing well towards his/her degree and one at-risk student who is struggling to complete his/her degree or underperforming. All the data used in this study are real students' data from the CCI.

7.4.3.4 User Interaction

Upon finishing the demonstration and answering the participant questions, participants are requested to interact with the FIRST and use it as if they were preparing for an advising session with some of their students. In this part of the study, the participants are required to try all the functionalities that are implemented in FIRST. During this user interaction, participants can ask questions, comment, or suggest any changes and the study facilitator answers their questions. The average interaction time for each participant is 15 minutes. Some participants use the system to look for only one student, while some participants use the system for more than one student. The most number of students looked at in a single session is 4 students. The total number of students looked at using FIRST for the entire user study is 31 students.

7.4.3.5 Follow-up Questions

After the user interaction, the participants asked questions about the system and the study facilitator demonstrated additional interactive features of FIRST. Upon finishing the user interaction, participants are asked the following questions:

- What did you like or dislike about the tool? Can you compare this with the previous tools you have used?
- What insights were you able to gain through viewing the information presented in this tool?
- In the data model, there are various features available for describing students. How do these features help you better understand the student?
- How do visual analytics help you in advising?
- What were you able to learn about the students in the storytelling part of the tool?
- Under what circumstances would you prefer a story structure over another?
- Which features (parts of the stories) do you find more useful to know about? And Would you like the story to have more or different information?
- Does the text of the individual student's stories read naturally?
- Does the text of the individual student's stories have clarity?
- Does the text of the individual student's stories convey what it should convey regarding the selected features?

7.4.4 Qualitative Thematic Analysis

To analyze the interview study, the audio recordings of all the interviews' follow-up questions are transcribed and then a thematic analysis is performed. Figure 7.7 shows a word cloud visualization of the transcribed audio of the follow-up questions. Then, the transcriptions are labeled with a number of codes. Table 7.2 contains all the codes that are used for labeling the transcriptions. Initially, the codes are chosen based on the third research question (Which story building blocks (contents and structures) are meaningful for the domain experts?). Further, three categories are developed: (1) understanding student story contents, (2) research questions 1.1, 1.2, 2.1, and 2.2, and (3) future research. For each category, a number of codes are defined as shown in Table 7.2. The codes enable quantifying the concepts discussed in a transcription.

For example, counting how many times participants talked about "temporal story" or "outcome story".

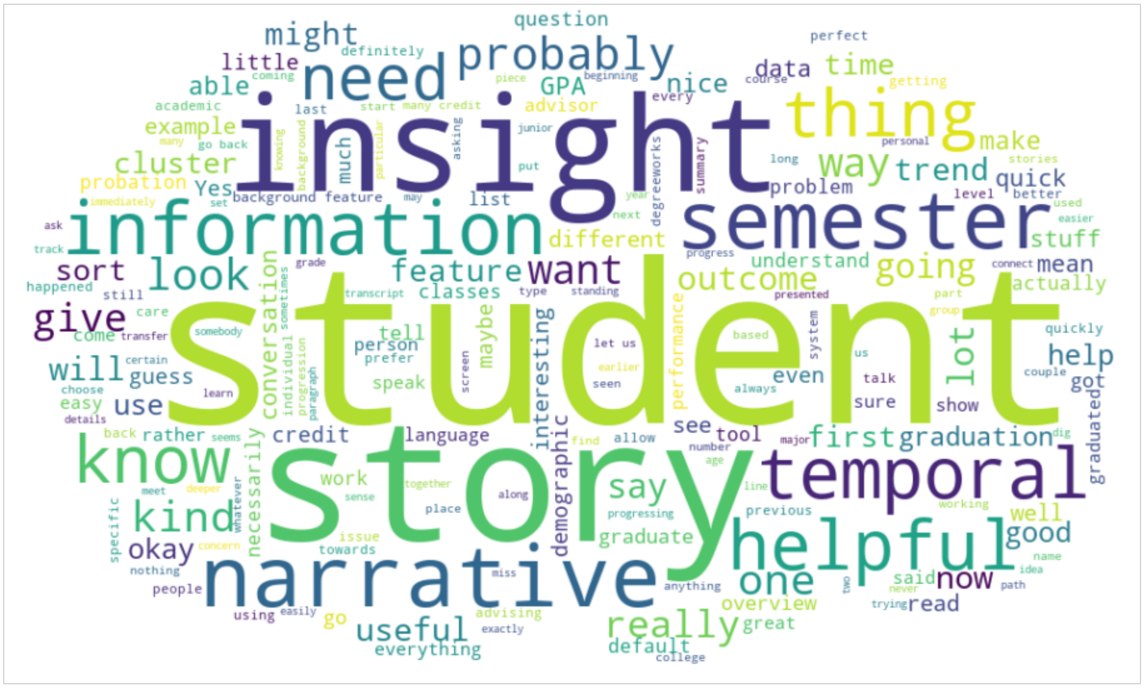


Figure 7.7: A word cloud visualization of the transcribed audio recording of the interviews follow up questions.

Table 7.2: Interviews follow up transcription codes, categories, use cases, and examples.

Code	Category	Use cases	Segment example
Clarification /Question	Understanding the features of FIRST	When a participant asks a question or the focus group leader clarifies a concept, or answers a particular question.	"Should I go back to the student data model page and see what are the features I wanted to include?"
Story content	Research question	When a participant refers to useful story content	"The background features allow me to sort of know if veterans are having particular problems in CCI, or international students facing specific challenges."

Story structure	Research question	When a participant refers to their preferred story structure	"Personally, I like the temporal one better. But I do like the fact that you have given that option if someone prefers to see the outcome first."
Story language quality	Research question	When participants asses the language of the student stories.	"the text is very clear. It does not have any missing things... the story sentence construction! That looks okay."
Usability suggestions	Future research	When participants suggest improvements or changes to the current work.	"It would be more convenient to be able to save the feature selection for future advising sessions."

Coding the transcriptions allows quantifying the concepts discussed in the interviews follow up questions by counting the occurrences of each code. Figure 7.8 shows the distribution of codes in the interview follow up questions. Based on this figure, the discussions are more directed towards the content of the student stories (37%) and the student story structuring (26%). A notable amount of the discussion is directed towards the quality of the student stories' language quality (15%).

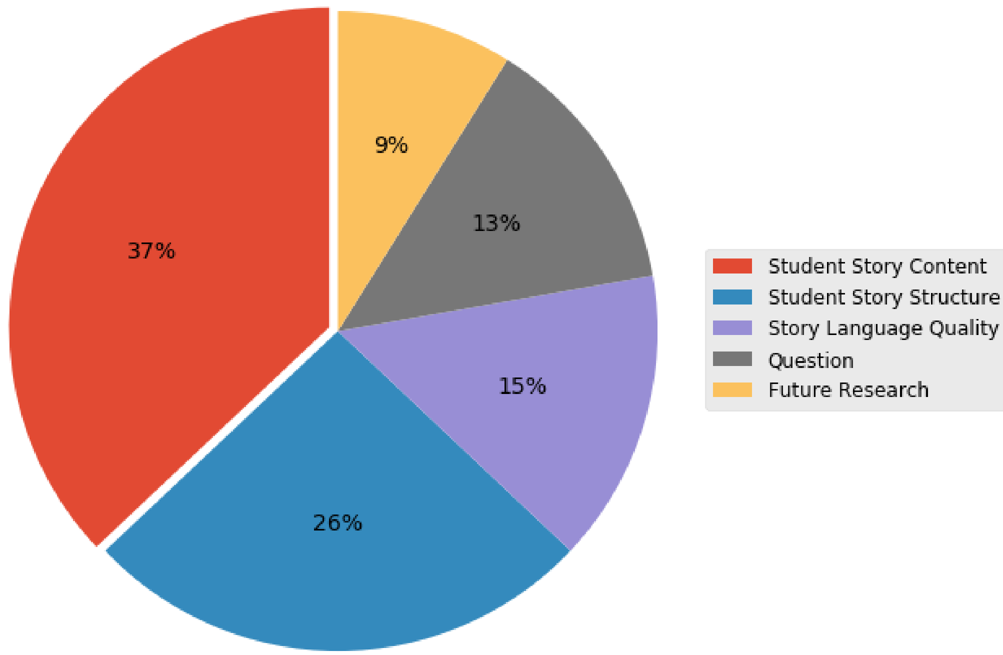


Figure 7.8: Distribution of the codes in the interview user study follow up questions.

7.4.5 Summary and Contributions

The result of the user study analysis is four high-level themes. The following subsections present these themes and answer some of the research questions asked based on the thesis statement.

7.4.5.1 Student Stories Content Identification

This theme answers the research questions RQ3.1 (Which features from the student data model do domain experts find helpful throughout their advising session?) and RQ3.2 (What content of the student stories are more meaningful for the domain experts?). Figure 7.9 shows a word cloud visualization of the contents that are meaningful for the participants. As shown in Figure 7.9, In which most of the participants expressed their preference in several kinds of contents they find meaningful and helpful in their advising sessions as follows:

- Student background and demographic information. Most of the participants show their interest in several features from the student background node. These features include citizenship, student native language, and ethnicity. Some participants mentioned that they find the student's previous education helps in assessing their performance.

- Student course information: all participants that the student courses information is essential for their advising sessions. For example, the student taken, passed, failed, withdrawn courses. In addition, they want to know about the students' courses with a D grade, retaken courses, prerequisites, number of courses taken each semester, and the order of the student courses in which they are taken.
- Student personal information: some participants show their interest in knowing about the students' personal information like students' personal interests, abilities, housing information, marital status, financial information, and any personal troubles or issues. On the other hand, some participants find the student work information helpful like if the students have done an internship or been invited for a company interview.
- Student grades and GPAs: evidently, all participants would like to know about the students' grades, GPA, and their transcript.
- Student outcome information: all of the participants show that they want to know about the students' graduation information like expected graduation data, delayed graduation, or student's plan for graduation.
- Student major information: most of the participants stated that knowing about the student's majors, minors, concentration, and early entry is helpful for their advising sessions.
- Student academic standing: most of the participants find that the student academic standing is important for their advising sessions.
- Comparison with other students: some participants mentioned that it is helpful

to know the pattern of performance of a student in the context of their peers (if a student performs well or is underperforming in comparison to other students in CCI).

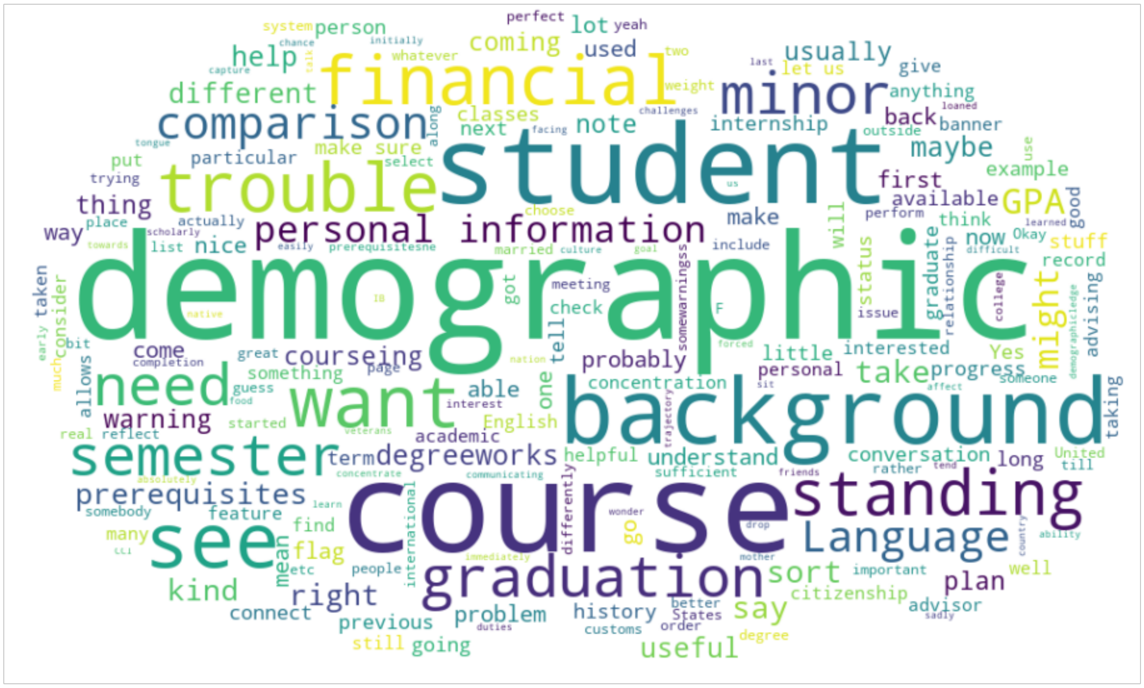


Figure 7.9: A word cloud visualization of the transcribed audio recording of the student story contents that are meaningful for participants in the interview study.

7.4.5.2 Student Stories Structure Identification

This theme answers the research questions RQ3.3 (What story structures are more meaningful for domain experts?). In the interviews, the participants are asked about the circumstances that would make them prefer a story structure over another, or if they have an alternative way of structuring the students' stories. Most of the participants stated that they like the temporal story structure; which starts by presenting the student's demographic information, then their semester and academic information, and ends up with their outcome information. For instance, one participant stated: *"The temporal order. The way it is set to have a nice conversation in and out enough about the students so I can help her. It just happens to make sense the way it* *is set up right now."* Another participant mentioned that the temporal structure is good in presenting the student information as a sequence of events. The participant said: *"I like the temporal one better, but that is just the way my head works. I like sequential things, you know, I like to see. Alright, well, how did they start? And then how did they finish up?"* Some participants liked the outcome story structure since it starts with the information that is most important to them. For instance, one participant stated: *"I would probably go with the outcome most of the time because you are trying to make sure that students graduate on time. And it is the first thing that you want to know more about. I definitely want to know whether they are on track to graduate or not."* Another participant stated: *"I do not necessarily care too much about the age as the first piece of information. Probably know the outcome most of the time."* Some participants mentioned that their preference depends on some factors like if the students are juniors or seniors. For instance, one participant said: *"I think it depends if a student is in their junior year, I would, or they have just come to me like transferring from the advising center to me, I would look into the temporal, but if a student is in their final semester or their second last semester, then I would look into the outcome."* Another participant stated: *"I think if I was most concerned about a student who was meeting goals or failing to meet goals or failing to meet expectations of when they are going to graduate, then I would be most concerned with the outcome one. Otherwise, I would prefer the temporal one."* Some participants suggested that it is good to have another story structure that starts with the student semester information. For example, one participant stated: *"I prefer to see the semester information first. If it is the first time I meet with the student, I probably want to see the demographic information first. But for an existing student who has been coming to me every semester for several semesters, I probably care more about how he did last semester, and how do we look towards his graduation?"*

Most of the participants appreciated that they have the option to choose the preferred structure. For example, one participant said: *"Personally, I like the temporal one better. But I do like the fact that you have given that option if someone prefers to see the outcome first. It is nice to have that option."*

7.4.5.3 Student Stories Language Quality

The participants are asked to rate the story language quality in terms of 3 metrics: (i) story fluency (does the text read naturally?), (ii) story comprehensibility (does the text in the story have clarity?), (iii) story correctness/fidelity (does the text convey what it should convey regarding the selected features. The average ratings for these metrics are depicted in Table 7.3.

Table 7.3: The average ratings for the student stories language quality

Metric	Rating (Out of 5)
Story fluency	4.6
Story comprehensibility	4.4
story correctness/fidelity	4.8

7.4.5.4 Usability Suggestions

Three participants suggested that the student stories page should be the first page when they open the system instead of the student data model. Most of the participants suggested some changes to the student data model page. For example, having options like select all and deselect all features, or the ability to save the feature selection for future advising sessions. Some participants suggested having an option to load their own advisees on the student stories page, instead of selecting the students from the aggregate analytics page. These suggestions are addressed for the next user study presented in the following section.

7.5 Diary Study for Contextual Understanding of Advisors Experiences with FIRST

7.5.1 Overview

This diary study aims to contextually understand advisors' experiences with the creative student storytelling model over time. This longitude diary study gathers advisors' insights about the proposed storytelling model and assesses how this model facilitates their sensemaking of students' success or risk over time. This study collects qualitative data about advisors' behaviors, activities, and experiences, how these behaviors evolve over time, and what influences these behaviors. Throughout this time, participants were asked to keep a diary and log of the activities being studied.

Sixteen advisors were recruited for the diary study. Nine were professional advisors. The other seven participants were faculty advisors. The participants comprised eight females and eight male advisors. This study was designed to evaluate the contextual understanding of advisors' experiences with the student storytelling model over time.

7.5.2 User Experience Design

This section presents the user experience design of FIRST version 3.0 for this user study. Some changes have been made to the user experience of FIRST version 2.0 presented in Section 7.4.2. In addition to the functionalities presented in the Sections 7.3.2 and 7.4.2, this user experience adds the following functionalities:

- The ability for advisors to view their advisees as a separate list on the student stories page. However, they are still able to view and select from all CCI students from the aggregate analytics page.
- The ability to select all student features, deselect all features, save current feature selection, and load saved feature selection in future advising sessions.
- The ability to log the daily experience and interaction using the daily log page.

Based on these functionalities, the user experience includes four pages: student stories, student aggregate analytics, student data, and daily log pages. These pages are presented in the following subsections.

7.5.2.1 Student Stories Page

The student stories page is the landing page and it is shown in Figure 7.10. On this page, the domain expert can view an automatically generated story for their students or for CCI students who are selected from the aggregate analytics page. The students' names are presented in "My Student" panel at the left (Figure 7.10(a)). By clicking on the student name from this panel, the domain expert can view and their automatically generated stories as shown in Figure 7.10(b). Also, the domain expert can view an interactive timeline of the student per semester and view information about each of the semesters (Figure 7.10(c)). At any time, the domain expert can use the navigation menu at the top of the page (Figure 7.10) to go to the student data page and select the features they want to include in the student story.

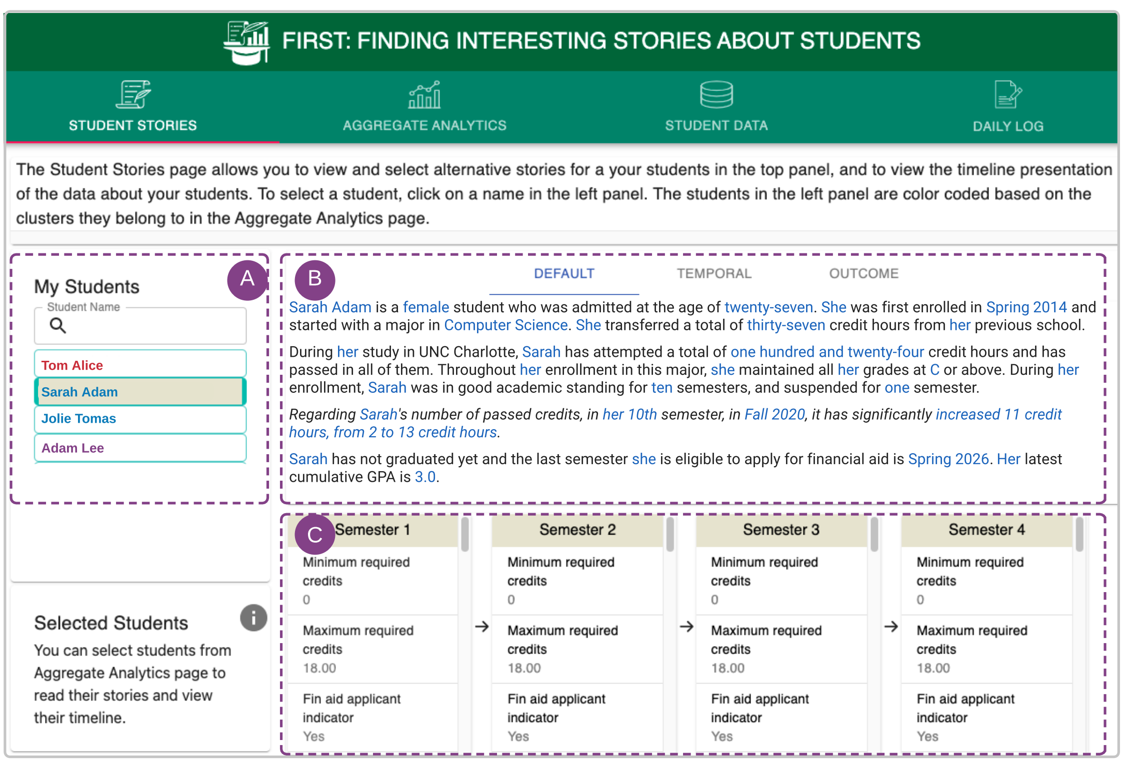


Figure 7.10: FIRST version 3.0 student stories page.

7.5.2.2 Aggregate Analytics Page

The aggregate analytics page is shown in Figure 7.11. On this page, the domain expert can view groups of students that have similar patterns for several temporal engineered features. The instructions about how to select different temporal engineered features and view their visualizations are at the top of the aggregate analytics page as shown in Figure 7.11. The domain expert can select the engineered feature they are interested in from the drop-down menu shown in Figure 7.11(a). Based on their selection, two types of visualizations are presented on this page for each engineered feature, the first visualization (Figure 7.11(b)) shows clusters of students based on the selected temporal engineered feature. The second visualization as shown in Figure 7.11(c) shows the average GPA per student cluster. These visualizations are included to compare the domain expert's sensemaking using visualization versus their sensemaking using storytelling. Finally, at the bottom of this page, a table of all CCI students is presented as shown in Figure 7.11(d). The domain expert can select the student they want to know more about and then go to the student stories page to read their stories.

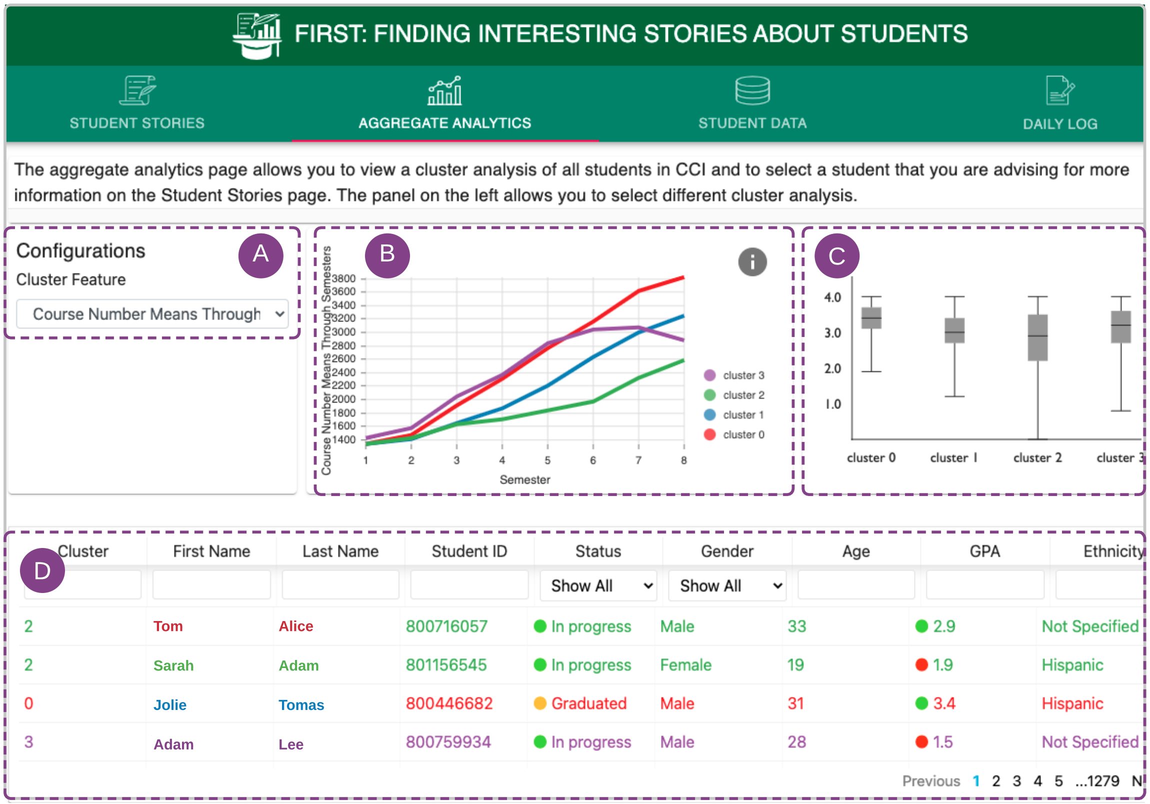


Figure 7.11: FIRST version 3.0 aggregate analytics page.

7.5.2.3 Student Data Page

The student data page is shown in Figure 7.12. At the top of this page, there is a non-sequential navigation menu as shown at the top of Figure 7.12. This menu is accessible on all pages of the application, which enables the user to navigate and modify the settings at any time. This page includes instructions on how to select student features at the top of the page. Also, on this page, the student data is arranged according to the student temporal data model (i.e., background, semester, and outcome) as shown in Figure 7.12(a). The domain expert can select the data features that they are interested in. Their feature selection is used when generating the student stories. After the selection of the features that the domain expert is interested in, they can navigate to the aggregate analytics page. At the bottom of this page, there are options to select all student features, deselect all features, save current feature selection, and load saved feature selection in future advising sessions. At the bottom of the page, as shown in Figure 7.12(c), there are some options for the feature selection, e.g., select all, clear selection, save current selection, and load saved selection.

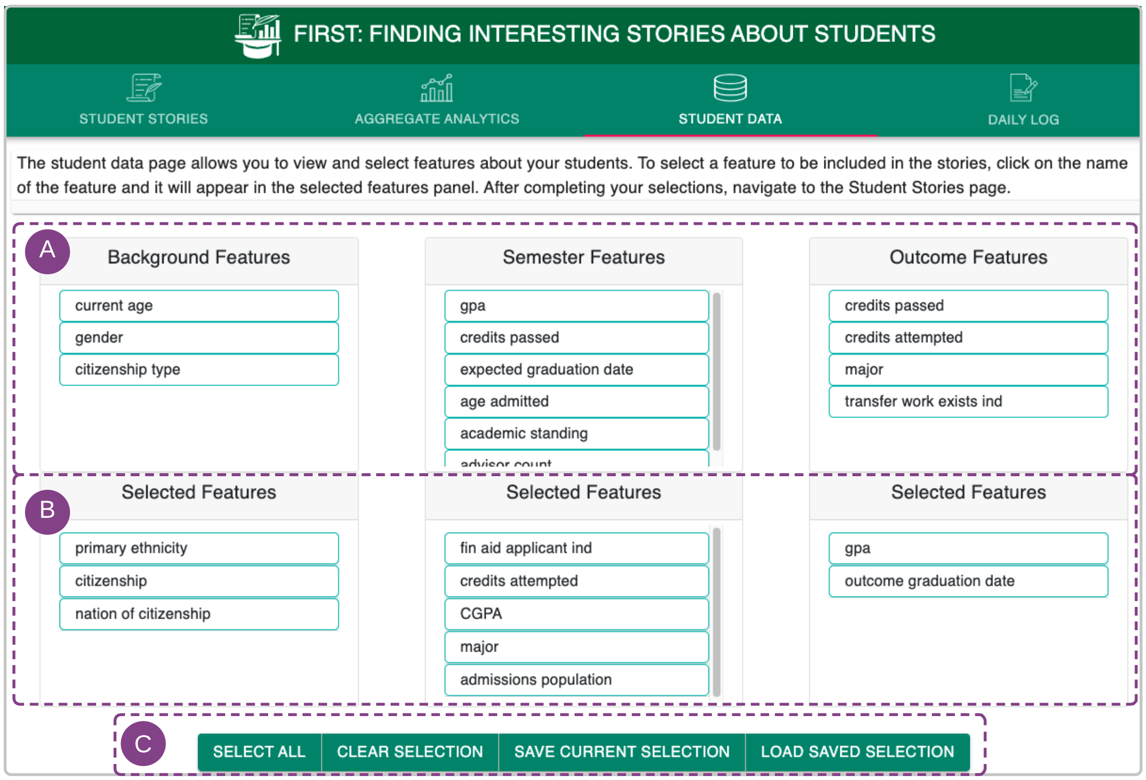


Figure 7.12: FIRST version 3.0 student data page.

7.5.2.4 Daily Log Page

The daily log page is shown in Figure 7.13. This page is for the user study participants to be able to log their experience and interaction with FIRST. A daily log is a form of survey composed of three pages: (i) instruction page as shown in Figure 7.13. This page gives the participant instructions and directions on how to fill and submit the survey, (ii) questions page as shown in Figure 7.14. This page asks the participant how many students they advise and how many students advised using FIRST on the logging day, and (iii) rating page as shown in Figure 7.15. This page asks the participant to rate several features of FIRST in terms of the following aspects:

- The student stories were helpful in advising
- The feature selection for the student stories helps me choose features I am interested in
- The student timeline was helpful in advising
- The aggregate analytics were helpful in advising
- The website is easy to navigate

The rating page also asks the participant if they would like to comment or suggest any feature.

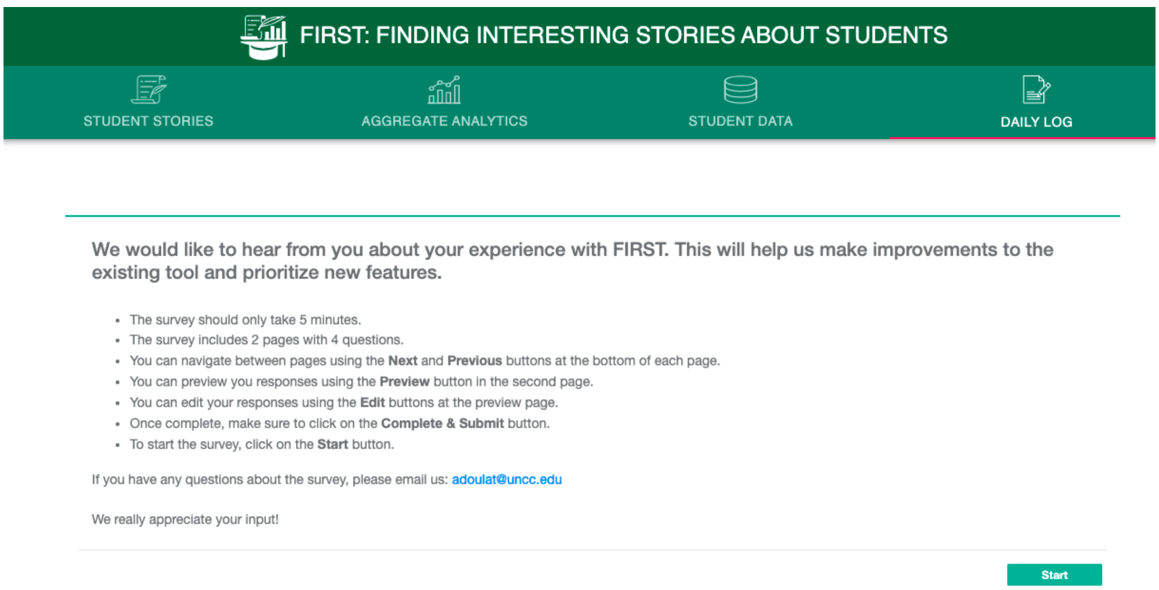
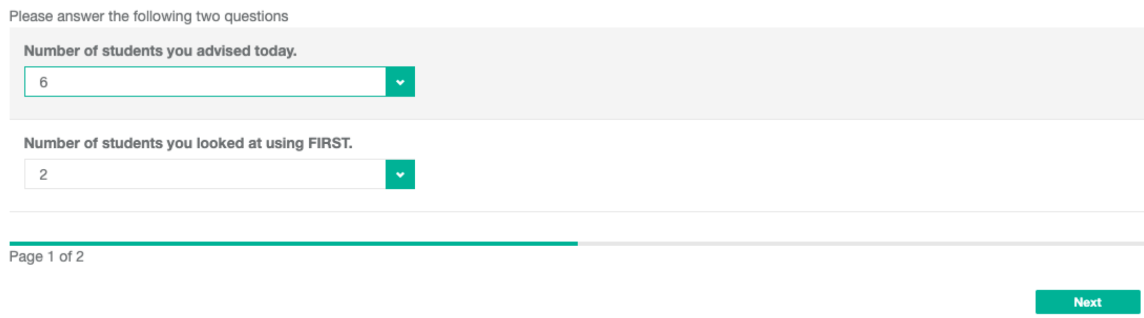


Figure 7.13: FIRST version 3.0 daily log instruction page.


Please answer the following two questions

Number of students you advised today.

6

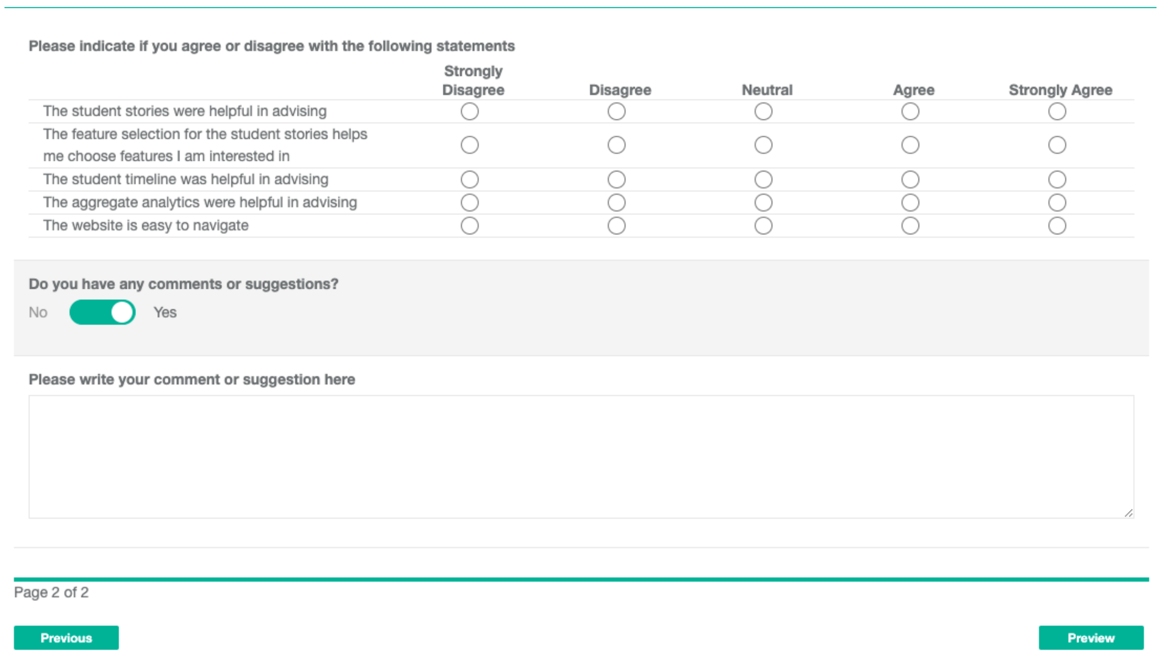
Number of students you looked at using FIRST.

2

Page 1 of 2

Next

Figure 7.14: FIRST version 3.0 daily log questions page.


Please indicate if you agree or disagree with the following statements

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The student stories were helpful in advising	○	○	○	○	○
The feature selection for the student stories helps me choose features I am interested in	○	○	○	○	○
The student timeline was helpful in advising	○	○	○	○	○
The aggregate analytics were helpful in advising	○	○	○	○	○
The website is easy to navigate	○	○	○	○	○

Do you have any comments or suggestions?

No Yes

Please write your comment or suggestion here

Page 2 of 2

Previous

Preview

Figure 7.15: FIRST version 3.0 daily log ratings page.

7.5.3 Study Design

The diary study composes five phases: (i) planning and preparation, (ii) pre-study brief, (iii) logging period, (iv) post-study interview, and (v) data analysis. The following sections present these phases one by one.

7.5.3.1 Planning and Preparation

This phase of the study is to define the focus of the study. In addition, it includes defining a timeline for the study, identifying how to collect data about users' behaviors, recruiting participants, and preparing instructions or support materials.

7.5.3.2 Pre-Study Brief

This phase aims to get participants ready for the study by scheduling meetings with each participant to discuss the details of the study. Additionally, present FIRST version 3.0 that they will be using and make sure each participant has familiarized themselves with the FIRST system. This phase of the study is divided into three parts: preliminary discussion, demonstration, and user interaction. In the preliminary discussion, an overview of the project and this study is presented. Then, the facilitator presents a brief explanation of the student temporal data model and the student storytelling model. The facilitator also presents how the student stories are generated from the student temporal data model. The explanation also includes a discussion about the students' aggregate analytics and how it is performed on the student data.

Following the preliminary discussion, FIRST version 3.0 was demonstrated by the study facilitator with some scenarios for several students from CCI. During this demonstration, the participants are allowed to ask questions about the system and the study facilitator answers their questions. Upon finishing the demonstration and answering the participant questions, participants are requested to interact with the FIRST and use it as if they were preparing for an advising session with some of their students. In this part of the study, the participants are required to try all the functionalities that are implemented in FIRST. During this user interaction, participants can ask questions, comment, or suggest any changes and the study facilitator answers their questions. The average interaction time for each participant is 15 minutes. Some participants use the system to look for only one student, while some participants use the system for more than one student. The most number of students looked at in a single session is 3 students. The total number of students looked at using FIRST for the entire user interaction is 24 students.

7.5.3.3 Logging Period

This phase of the study is designed to take almost a month to complete, i.e., over the advising period of Fall 2020. In this phase, the participants are asked to log their daily experiences with FIRST. Their daily log was to answer two questions regarding the number of students they advise in general and the number of students advised using FIRST on the logging day. The participants were also asked to rate the helpfulness of the following:

- Students' stories
- Story feature selection
- Students' timeline
- Students' aggregate analytics, and
- Students' data model

Throughout this logging period, the participants used FIRST to advise 298 students from CCI.

7.5.3.4 Post Study Interview

A follow-up one-on-one interview to discuss advisors' experience with FIRST in detail. This interview includes asking for feedback from the participant about their experience participating in the study in general. In this phase of the study, the participants were asked the following questions.

- What is your overall response to having stories about students available during advising?
- When in the advising process did you use FIRST?
- What value did reading the stories have for advising?
- What value did viewing the student timeline have for advising?
- What value did viewing the aggregate analytics have for advising?
- What did you find interesting in the students' stories?

- How do students' stories affect your advising?
- Do you have any questions, comments, or suggestions?

7.5.3.5 Data Analysis

This phase of the study aims to evaluate the participants' behaviors throughout the study. How do they evolve and change over time, and what influences these behaviors. The analysis includes the logging information and the post-study interviews. For the logging information, participants are asked to rate the helpfulness of students' stories, story feature selection, students' timeline, students' aggregate analytics, and students' data model. The average rating for each of these features is depicted in Table 7.4. As shown in the table, the student stories are rated as the highest (4.2) feature of the system in terms of its helpfulness in understanding students in advising sessions. Participants also find that the feature selection for the student stories is helpful in their advising sessions. Participants find that viewing student aggregate analytics using visualization is less effective in understanding students' data. For the post-study interviews, a qualitative thematic analysis is performed. The following section presents this qualitative thematic analysis.

Table 7.4: Participants daily log average rating in terms of feature helpfulness

Feature	Average Rating (out of 5)
Student Story	4.2
Story Feature Selection	4.1
Student Timeline	3.8
Student Data Model	3.7
Student Aggregate Analytics	2.8

7.5.4 Qualitative Thematic Analysis

To analyze the follow-up interviews, the audio recordings of all the post-study interviews are transcribed and then a thematic analysis is performed. Figure 7.16 shows a word cloud visualization of the transcribed audio. Then, the transcriptions are labeled with a number of codes. Table 7.5 contains all the codes that are used for labeling the transcriptions. Initially, the codes are chosen based on the first research question (Do the student stories provide an easier, more engaging, and more understandable way for non-experts to make sense of complex and heterogeneous student data through narrative stories?). Further, the transcripts are reviewed and the emerged themes are identified. The result is five high-level themes that were then used to revisit and code the transcript according to the themes. The following sections provide a summary of the interviews according to each theme.

Figure 7.16: A word cloud visualization of the transcribed audio recording of the post study interviews.

Table 7.5: Post study transcription codes, categories, use cases, and examples.

Code	Category	Use cases	Segment example
Clarification /Question	Understanding the features of FIRST	When a participant asks a question and the study facilitator answers the question.	"How does the system decide the expected graduation date?"
Story effectiveness	Research question	When a participant refers to the usefulness and effectiveness of student stories	"I start my advising session by looking at FIRST because it takes way less time in the narrative. That was a time saver."
Stories are more engaging and memorable	Research question	When a participant state that the stories are engaging and memorable	"... it is hard to remember each student, so having that student story there, will refresh your memory about the student"
Stories Content	Research question	When a participant refers to story content	"I like the student demographic information since we could not find this kind of information in other tools"
Story audience	Research question	When a participant refers to users who can make sense of student stories	"... the stories definitely were more helpful for us as a professional advisor. The faculty advisors probably love those analytics because they are computer science-minded folks"
Usability Issues	Future research	When participants suggest improvements or changes to the current work.	"... if there is a way to include student courses to the student story, it is going to be more useful"

7.5.5 Summary and Contributions

The result of the diary study analysis is six high-level themes. The following subsections present these themes and answer some of the research questions asked based on the thesis statement.

7.5.5.1 Student Stories Effects on Advising

Most of the participants expressed that the student stories should be the first thing to look at before meeting with students. When they are asked about the reason, they mentioned that the student stories help start a natural conversation with students. One participant said: *"The stories help us kind of start a conversation or continue a conversation more naturally than just looking at different databases and kind of piecing things together."* Some participants stated that having the student data presented in a story changes the way they prepare for an advising session, in which the stories provide them with information about the student that it is time-consuming to get in other tools. One participant said: *"When we are extremely busy as during the pre-registration period, a lot of us have 300 plus advisees and their appointments back up to each other so closely, and being able to look at those stories was a luxury that we didn't have because the timing was so tight, stories save me from looking into other tools that we used to look at."* Another participant said: *"I think it is a great idea to have the student's story available. It kind of summarizes everything that we would look at from different data points in the different parts of the student record. And it just gives us a quick overview of what they've done and where they currently stand and what they may be working toward in a story setting."* Some participants stated that student stories are a natural way to communicate with others and remind them that they are dealing with humans, not with numbers and graphs. One participant said: *"I think sometimes people can approach things very cut and dry or treat students like numbers. And I think the story goes a long way to remind you that they're a person and they have these different factors and they are coming from a different perspective."* Another participant said: *"Ultimately, I'm having a conversation with a real person, not a list of qualities or numbers in a table. The stories give me that impression."*

7.5.5.2 Student Stories are Effective

Most of the participants expressed that the student stories are effective in providing a quick overview or summary of the student. One participant stated: *"Well, I really liked the narrative portion, I thought that was very helpful. You know, being able to just get a quick glance at the students' performance in their courses, or what sort of their history since they attended the university. That was very helpful."* Another participant stated: *"I think I felt like that gave me a very quick way to get a sense for nearly a type of students, but really the sort of the challenges or the particular issues or the student might be facing, so I thought that was helpful."* One participant stated: *"I like having that story just to get the big picture up front."* Some participants expressed that student stories are faster to get insights about the student compared to other types of presentations (tabular and visual). One participant stated: *"It was way more efficient to read a story than to start scanning DegreeWorks or Connect to see what's going on."* Another participant stated: *"I start my advising session by looking at FIRST, because, you know, if I have to see for students struggling, I can see that from the DegreeWorks. But it takes way more time than in the narrative. That was a time saver. So I really like the narrative part."*

7.5.5.3 Student Stories are Engaging and Memorable

Participants agreed that stories are more engaging and easier to remember than numbers and graphs presented in tables and visual components. One participant mentioned that the stories are helpful when advising many students in a sequence of meetings since it is easier to remember than other ways of data presentation. *"Since all of us advise a lot of students, it is hard to remember each student sometimes, so having that student story there, will refresh your memory about the student."* Another participant stated: *"the value of the student's story for me is that we have a lot of advisees, and sometimes it's a little hard to remember things about the advisee, even* *though you're looking at their DegreeWords, that's not the whole picture. That's just telling you what courses they've taken. The story fills that gap."* Participants also expressed that the student stories make them more involved in and engaged with the student experience. For example, one participant stated: *"Other tools can provide me with information regarding students' courses and their risk scores. But that's not telling me more about the student. So the story tells you who the student is, where they're coming from, what their experience at UNC Charlotte so far has been etc. So that's kind of what I liked about it."*

7.5.5.4 Student Stories are for Everyone

Participants expressed that it is easier to make sense of data through storytelling especially for non-expert users compared to tabular or visual components. Most of the professional advisors who have little to no experience in data science expressed confusion about some of the visual components. For example, one participant states *"I had trouble understanding the figures at the top. And also the table at the bottom. ... computer scientist advisors want to see the raw data and can slice and dice on their own."* Another participant said when asked about the analytics visualization: *"So the stories definitely were more helpful for us as professional advisors. The faculty advisors probably love those analytics because they are computer science-minded folks. Yeah, we're education-minded. So analytics to us looks different, so I never could really wrap my head around that. So that seemed to be a little bit of something that wasn't as helpful for me."* In addition, another participant stated: *"I am not an analyst in any way. I'm not a computer science major. And with those analytics, I really couldn't even wrap my head around it enough to say whether I liked it or didn't like it because it was just more in-depth than I think I could go."*

Some participants suggested including a storytelling feature to describe the visual components such as line graphs, bar charts, and cluster illustrations.

7.5.5.5 Relevance of Student Stories' Content

Participants appreciate that the stories could provide a wide range of information about the student. Other tools they use like DegreeWorks and Connect provide information about the student courses, grades in these courses, or risk scores. However, FIRST provides information about the student's demographic information, their academic performance, comparison with other students, and their outcome information. For instance, one participant said: *"FIRST does a much better job of giving access to information that's relevant to my job as an advisor. Whereas Connect, it just, it's so hard to find information in Connect regarding the student demographic information. It is also hard to find a comparison with other students in DegreeWorks. But, the stories include such kind of information; which was really helpful".* Another participant stated: *"I like the student demographic information since we could not find this kind of information in other tools".* Participants also appreciated the analytics results in the students' stories. For example, one participant said: *"I like that the stories extract the trends or something abnormal, it highlights that. I found that to be the most useful if there's inconsistency or something that is not typical when it highlighted that or that was mentioned in the story, I found those to be most helpful."* Another participant said: *"I really liked the narrative part of FIRST - knowing that the story is built using analytics makes me trust it more."* Participants appreciated that they could select the features they thought should be part of the student story. For example, one participant stated: *"It's important that we have that flexibility to be able to put in different criteria to create the students' story."*

7.5.5.6 Student Stories Issues

Participants expressed that student stories lack some useful information about the student, like the student's courses taken each semester, their courses' grades, and their financial information. One participant stated: *"I use FIRST to read the student* *story and then if I need specific information about the student's courses, I go to Degreeworks. So if there is a way to include them, in a way or another, to the student story, it is going to be more helpful."* One faculty advisor who had experience with data visualization and aggregate analytics expressed that some kinds of data are better presented using charts and graphs by stating: *"Not all kinds of data can be presented in narrative text. For those kinds of data, I probably prefer charts and graphs. I tend to think like that, and I do a lot of that when I'm doing other parts of my job. So charts and graphs are things that I'm very comfortable with and things that I tend to go to get information. For those kinds of data, I think the narrative text isn't typically something I would use."*

7.6 Focus Group Study to Evaluate FIRST Explainability

7.6.1 Overview

The purpose of this focus group study is to evaluate FIRST explainability and interpretability. It aims to demonstrate the explainability and interpretability of FIRST to domain experts and evaluate the ability of the story explanations to inspire advisors' trust and confidence with the story content, make it simpler and faster for the advisor to find what they are looking for regarding student data, improve advisors satisfaction with the generated stories, and persuade the advisor to receive and accept the generated stories. The study includes professional and faculty advisors since they are already familiar with multiple tools that provide data, analytics, and risk scores for the students that they advise. A focus group was selected for this study because it is effective in collecting user opinions and attitudes through group discussion and conversation dynamics.

Three advisors from the CCI were recruited for this focus group study. Two were professional advisors and one was a faculty advisor. The participants comprised one female and two male advisors.

7.6.2 User Experience Design

This section presents the user experience design of FIRST version 4.0 for this user study. Essential changes have been made to the user experience of FIRST version 3.0 presented in Section 7.5.2. The updated user experience includes a single all-in-one page called student panel and it aims to provide the domain experts the various functionalities. Following are the major functionalities that are the focus of this user study:

- The ability to select the features they are interested in from the student data panel.
- The ability to read automatically generated stories for the selected features.
- The ability to select between three story structure alternatives: default, temporal, and outcome structures
- The ability to interact with the dynamic parts of the student stories to view the explanations for those parts. In those explanations, advisors can view students transferred, attempted, passed, failed, withdrawn, d-scored courses along with their number of credits hours and grades
- The ability to view several visual illustrations and explanations for student features like student academic standing, GPA timeline, gender distribution, age distribution, citizenship distribution, and ethnicity distribution.
- The ability to view a student data using a tabular timeline
- The ability to view aggregate analytics for all CCI students

7.6.3 Study Design

The focus group study is designed to take almost an hour in total. It is divided into three parts: demonstration, user interaction, and follow-up questions. Throughout these parts, the participants were involved in verbal discussions. Participants are asked to directly interact with the FIRST system during the user interaction part of the study. The following subsections present these parts of the study one by one.

7.6.3.1 Demonstration

The study starts with an overview of the project and this study. Following the overview, FIRST was demonstrated by the study facilitator with some scenarios for several students from CCI. The scenarios include successful students who are doing well towards their degree and at-risk students who are struggling to complete their degree or underperforming. All the data used in this focus group is real students' data from the CCI. During this part of the study, the participants ask questions about the FIRST system, and the study facilitator answers these questions.

7.6.3.2 User Interaction

Upon finishing the demonstration and answering the participant questions, participants are requested to interact with the FIRST and use it as if they were preparing for an advising session with some of their students. In this part of the study, the participants are required to try all the functionalities that are implemented in FIRST version 4.0. During this user interaction, participants can ask questions, comment, or suggest any changes and the study facilitator answers their questions. The average interaction time for each participant is 15 minutes. Some participants use the system to look for only one student, while some participants use the system for more than one student.

7.6.3.3 Follow-up Questions

After the user interaction, the participants were asked the following questions about their interaction with the FIRST system:

- What features did you select from the student data model? And why do you select those features?
- What did you find interesting in the student stories?
- What information in the student stories did you find not interesting?

  - Can you give me an example of some information in the student story that you find hard to understand?
  - Have you interacted with the colored text? Does the interaction make you change your mind about that information?
  - Have you found the explanations (when you interact with the story) helpful in understanding that information?
  - Have you changed your feature selection after seeing that information?
- What purpose do the explanations serve for you?
- In those explanations, there are several parts: the title, the feature, the text, and the body that have interactions. Which part do you find more useful and why?
- Would you like to see more justification? Give me an example?

7.6.4 Qualitative Thematic Analysis

To analyze the focus group, the audio recording of the follow-up questions is transcribed and then a thematic analysis is performed. Figure 7.17 shows a word cloud of the transcribed audio recording. Then, the transcription is divided into segments. These segments' boundaries start when a participant starts talking about a particular concept and end when another participant starts talking about another concept. A revision is performed on the obtained segments to make sure that each segment includes only one concept. For instance, in a segment, if a participant starts talking about the feature selection, and then continues talking about the story explanations, then the transcription for this discussion would be put into two segments. One segment includes the sentences about the feature selection, and the other segment includes the sentences about the story explanation.

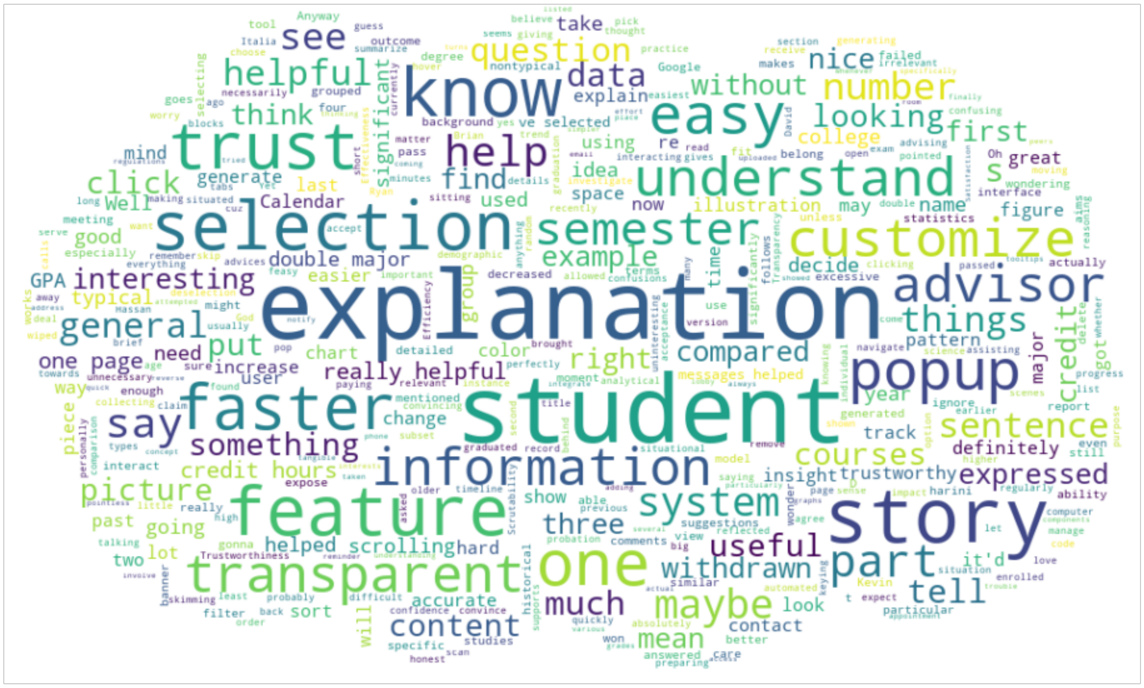


Figure 7.17: A word cloud visualization of the transcribed audio recording of the focus group follow up questions.

Upon finishing the segmentation of the transcription, each segment is coded. Initially, the codes are chosen based on the research questions about the FIRST explainability model. Further, two categories are developed: (1) understanding the features of FIRST, (2) research questions 4.1 and 4.2. Table 7.6 shows the codes that are used for labeling segments. For each category, a number of codes are defined as shown in Table 7.6. A segment can be labeled by one or more codes. The codes are based on the concepts discussed in each segment. Therefore, these codes can be used to quantify the concepts discussed in each segment. For example, counting how many times participants talked about "feature selection" or "story explanations". Segments that contain questions asked by the study facilitator are not coded and excluded from the analysis. The total number of segments is 43 segments.

Table 7.6: Focus group follow up transcription codes, categories, use cases, and examples.

Code	Category	Use cases	Segment example
Clarification /Question	Understanding the features of FIRST	When a participant asks a question and the study facilitator answers the question.	"If I click on the number of advisors for a student, do the advisors' names come up?"
FIRST Interpretability	Research question	When participants refer to story feature selection	"I find that selecting the student data that goes to the story useful because it helps me customize the story to tell me the information that I am looking for."
FIRST Explainability	Research question	When participants refer to story explanations	"I like the part that explains how this student is not typical compared to other students. Without this illustration, I won't be able to tell if this is accurate or not."

Coding the segments allows quantifying the concepts discussed in the focus group follow-up questions by counting the occurrences of each code. Figure 7.18 shows the distribution of codes in the focus group follow-up interview. Based on this figure, the discussions are more directed towards the FIRST explainability (54%) and then the FIRST interpretability (33%).

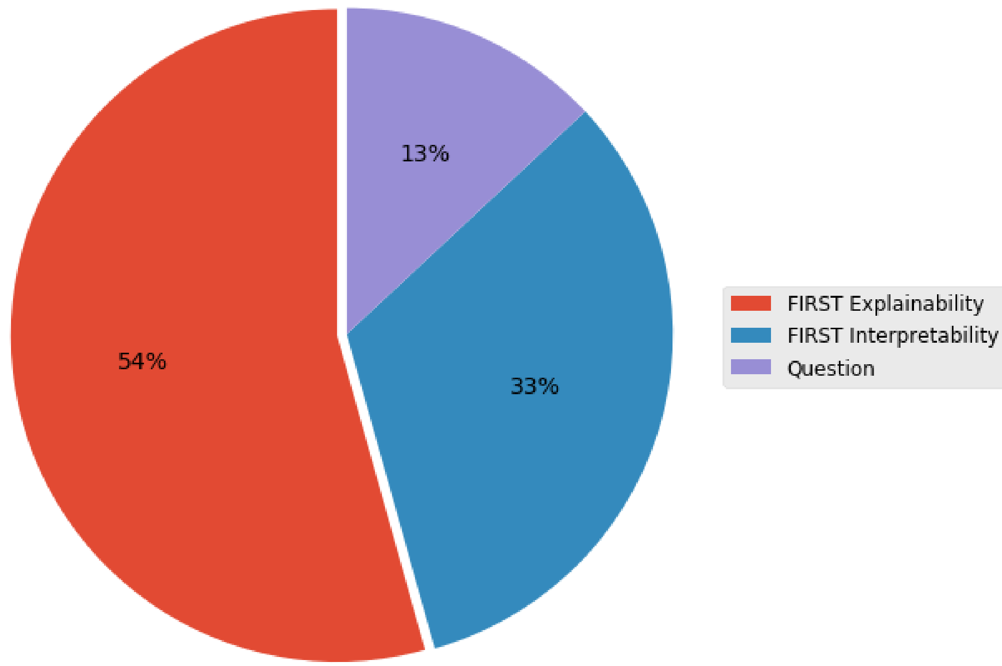


Figure 7.18: Distribution of the codes in the focus group follow up questions.

7.6.5 Summary and Contributions

The result of the focus group follow-up interview analysis is two high-level themes. The following subsections present these themes and answer some of the research questions asked based on the thesis statement.

7.6.5.1 Benefits of Selecting Student Features

This theme answers the research question RQ4.1 (What are the benefits of selecting student features to be included in the student stories?). In which participants appreciated having the ability to select the features that go into the student stories. It helped them look into the information they are most interested in. One participant said: "*I find that selecting the student data that goes to the story useful because, you know, it helps me customize the story to tell me the information that I am looking for.*". Participants expressed that the story explanations helped them decide which features from the student data are more meaningful to them. One participant stated:

"Those messages helped me understand why some parts of the story are there. Like, you know, it tells me that this piece of information is there because you've selected this feature or that feature. That's really helpful." Another participant stated: *"I like having those explanations in the story just to get an idea of what kind of information is used to generate different sentences."*

7.6.5.2 Benefits of Student Story Explanations

This theme answers the research question RQ4.2 (What role(s) do story explanations play in improving domain experts' sensemaking of students' data?). In which participants find the story explanations helpful in terms of the following aspects:

- **Transparency**: Story explanations helped advisors understand how the system works and how different content of the stories are generated. One participant stated: *"Those popup messages helped me understand why some parts of the story are there."* Another participant stated: *"I would say it does make the system more transparent and more trustworthy. So it's definitely helpful."*
- **Trustworthiness**: Story explanations help increase participants' confidence in the generated stories. Participants expressed that some stories explanations help them trust the content of the student stories. For instance, one participant stated that without the explanation, he could not make sure that the information in the student story is accurate or not. He stated: *".. it says that you grouped students into similar groups and this student is different because he doesn't belong to any group. Without this illustration, I won't be able to tell if this is accurate or not."*
- **Scrutability**: Story explanations help participants understand the system better when they interact with and investigate different parts of the story. One participant said: *"I understand from the pop-up message that this student is not typical compared to other students."* Another participant said: *".. Those messages helped me understand why some parts of the story are there. Like, you*

know, it tells me that this piece of information is there because you've selected this feature or that feature."

- **Effectiveness**: Story explanations increase the advisors' ability to ignore irrelevant or uninteresting story content while assisting them to choose the relevant features for the student story. Participants expressed that knowing how the system works helps them decide which features to select from the student data model. One participant said: *"I like the feature part of the explanation, the one just below the title. It helps me decide what features I should select and which ones to ignore."* Another participant stated: *"it tells me that this piece of information is there because you've selected this feature or that feature. That's really helpful."*
- **Persuasiveness**: Story explanations help convince participants about the reasoning behind some parts of the student story. For example, one participant stated: *"I like the part that explains how this student is not typical compared to other students ... Without this illustration, I won't be able to tell if this is accurate or not."* Another participant said: *"In one story, it has a sentence that says the student GPA has significantly decreased from one semester to another, and I wonder how you decide it is significant. When I clicked on the text, it said that it is significant compared to other students. So, it kind of answered the question in my mind."*
- **Efficiency**: Story explanations help participants make sense of student data faster, by providing them with detailed information about the features in the student stories. One participant stated: *"The charts that show a student's GPA, or passed courses change over semesters are really helpful. It shows the trend on how the student is making progress towards their degree which is useful."*. Another participant stated: *"I like skimming through the story just to get a general idea of the student situation, and then if I need more details for example the*

student failed courses, I can click on the text and see courses more specifically." Another example of an explanation that helped a participant get quick insight when he said: *"... it says that he has attempted 176 credit hours. So, I was wondering why he has this many credit hours. But when I clicked on the text, it showed me his courses with their grades, and it turns out that he withdrew several courses."*

- **Satisfaction**: Participants expressed that providing detailed explanations for the story content makes it easier and simpler to use the system. One participant stated: *"I used to look at charts and graphs and then have some kind of tooltips that explain what various parts of the chart mean. This tool does it in reverse. Anyway, it is always good to have some kind of illustrations when you have trouble understanding something."* Another participant said: *"Actually, I like that, you know, I can click on things and look at sort of general statistics for the college and so on. That makes it easier to get insights about this student compared to other peers in the college."* Participants also expressed that the explanations helped them customize the stories to their interests. One participant said: *"I find that selecting the student data that goes to the story useful because, you know, it helps me customize the story to tell me the information that I am looking for."*

7.7 Discussion

This chapter presents four ethnographic studies and the themes that emerged based on the researcher's experience of designing and developing FIRST, which is an explainable interactive learning analytics storytelling model, interviewing and interacting with the LA's domain experts, and analyzing the impact and values of having student stories on the domain experts. These themes provide answers to the research questions of this dissertation.

First, a focus group study is conducted to demonstrate FIRST and explore how the storytelling complements existing data and dashboard-style tools for advisors. This focus group study is conducted with 6 professional and faculty advisors from CCI at UNCC. The focus group study provided feedback on the implementation of FIRST version 1.0 and provided insights on the value of storytelling in the LAs system. The participants discussed the effectiveness of stories in providing a high-level understanding of the student and in the insight they can gain from a student that may be at risk or is taking too long to graduate. The student stories also help the participants to discover actionable knowledge about students. Moreover, the aggregate analysis, while useful for understanding groups of students, was improved with the storytelling feature since they could move from story to aggregate visualizations.

Second, a one-on-one interview study is conducted to find the building blocks that are meaningful for the domain experts in terms of student stories' content and structures. This interview study is conducted with 16 professional and faculty advisors from CCI at UNCC. This user study provided feedback on the implementation of FIRST version 2.0 and provided insights on the building blocks of students' stories that are meaningful for domain experts. The participants were appreciative of the ability to select from a larger range of student data than they have in other tools. Student demographic and background information helps participants to start conversations with students more engagingly and naturally. Moreover, the student semester, academic, and outcome information help participants to get insights and discover actionable knowledge about the students in terms of their academic performance. Participants were also appreciative of the ability to select different structures of the student stories. Their selections of different story structures were based on several factors like if the student was junior, senior, freshman, or transfer student.

Third, a longitude dairy study is conducted with the aim of understanding advisors' experiences with the proposed creative student storytelling model over time, gathering advisors' insights about the proposed storytelling model, and assessing how this model facilitates their sensemaking of students' success or risk over time. This study collects qualitative data about advisors' behaviors, activities, and experiences, how do these behaviors evolve over time? and what influences these behaviors? Throughout this time, the participants are asked to keep a diary and log of the activities being studied. This diary study is conducted with 16 professional and faculty advisors from CCI at UNCC. This user study provided feedback on the implementation of FIRST version 3.0 and provided insights on the contextual understanding of advisors' experiences with the student storytelling model over time. The participants expressed that the student stories are effective in providing a good understanding of student performance. Participants also expressed that student stories are faster to get insights about the student compared to other types of presentations (tabular and visual). Participants agreed that stories are more engaging and easier to remember than numbers and graphs presented in tables and visual components. Moreover, participants expressed that it is easier to make sense of data through storytelling especially for non-expert users compared to tabular or visual components.

Finally, a focus group is conducted to demonstrate FIRST explainability and interpretability to domain experts and evaluate the ability of the story explanations to inspire advisors' trust and confidence with the story content, make it simpler and faster for the advisor to find what they are looking for regarding student data, improve advisors satisfaction with the generated stories, and persuade the advisor to receive and accept the generated stories. This focus group study is conducted with three professional and faculty advisors from CCI at UNCC. The focus group study provided feedback on the implementation of FIRST version 4.0 and provided insights on the value of FIRST explainability and interpretability.

CHAPTER 8: FUTURE WORK AND CONCLUSION

8.1 Overview

This chapter summarizes the findings and contributions of this dissertation to research and practice, as well as the limitations, directions for future research, and conclusions. These research trajectories are organized around the four main thrusts of this work, namely developing a new storytelling model for domain experts in the LA domain, identifying the key student story's contents and structures that are meaningful for domain experts, developing an analytic model to make the student stories more interesting and useful for domain experts, and developing an explainable and interpretable LA storytelling model. This dissertation presents directions for future research such as extending the storytelling model to the course level for faculty, and the group level for leadership, involving new sources of data about students in the process of story generation, and conducting a long term evaluation study to evaluate the potential impact of student storytelling on students' grades and retention behaviors.

8.2 Contributions

The major contributions of this study are fourfold. First, a novel sensemaking of complex, diverse, and heterogeneous data through storytelling techniques is proposed in Chapter 3. Second, identification of key student story's content and structures that are meaningful for domain experts using ethnographic studies in Chapter 7. Third, an anomaly detection model is proposed to enrich student stories with interesting, yet, helpful information for the domain experts in Chapter 5. Finally, an explainable and interpretable interactive LA model is proposed for the domain experts to facilitate their sensemaking of student data in Chapter 6. Each of these contributions is summarized in the following sections.

8.2.1 Development of Student Storytelling Model

This study proposed a novel aggregate analytics and dashboard presentation style of complex, diverse, and heterogeneous data in the LA domain. Rather than presenting student data and analytics results using visualizations like scientific charts and graphs, this study proposes to present the student data using natural language stories that are automatically generated using Natural Language Generation (NLG) techniques and updated in coordination with the results of the aggregate analytics. Unlike other LA studies that tend to support student awareness of their performance or to support teachers in understanding the students' performance in their courses, this study aims to support advisors and higher education leadership in making sense of students' success and risk in their degree programs.

8.2.2 Identification of Student Story Content and Structures

This study identifies the key student story's contents and structures that are meaningful for academic advisors, faculty, and leadership. Ethnographic studies are used for this identification. Unlike other data storytelling studies that tend to generate summaries where the input data to these systems are the analytical results or numeric predictions. This study proposes a storytelling model that is capable of generating multi-paragraph stories from different sources of students' data that are complex, temporal, and heterogeneous. In this model, the identification of the story content is based on three different sources of information; students' temporal features, aggregate analytics results, and user-selected features. This latter one gives the advisor the ability to be part of the story generation process. The identification of the story structure is also done in an ethnographic study, where advisors are presented with three alternatives of story structures and asked questions regarding the circumstances they would prefer one structure over another.

8.2.3 Development of Anomaly Detection Model

This study develops an anomaly detection model to detect if there are extreme values (anomalies) in the student data compared to other students. The central intuition behind this model is to make the student stories more interesting and useful for academic advisors, faculty, and leadership. This study proposes two models of anomaly detection- Personal Anomaly Detection (PAD) and Collective Anomaly Detection (CAD). The PAD model aims to detect if an individual student's data instance can be considered anomalous compared to the rest of the data (e.g. a student's GPA extremely decreased from one semester to another compared to other students in CCI). The CAD model aims to detect if a collection of student data instances is anomalous compared to other students in CCI, but not individual values. For instance, if a student follows a non-typical pattern for the number of credits passed each semester.

8.2.4 Development of Explainable Interactive LA System

This study develops an explainable interactive LA system for academic advisors, faculty, and leadership. To the best of our knowledge, this is the first study to make an explainable and interpretable storytelling system. Explainability and interpretability were used in the domain of artificial intelligence systems to expose the reasoning and data behind a machine learning model prediction. In this study, explainability and interpretability are used to show the users how the stories have been generated and how the contents of the story are selected from the student data model. This system aims to increase the advisors' trustworthiness and satisfaction with the generated stories.

8.3 Limitations

To facilitate the sensemaking of complex, diverse, and heterogeneous student data, research is needed to increase the effectiveness and efficiency of LA systems and dashboards. However, there are some limitations to the present research. First, dealing with sparsity in student data in terms of some student's demographic and personal data, such as race, ethnicity, generational social class, student body demographics, geographic location of the institution, and socio-economic status of students which can be essential factors for students to be successful. Sparsity makes it hard to build learning models that consider the full range of student data. To overcome this limitation, other data sources can be involved in the story generation process like students' self-reported interactions recorded in the LMS logs, social media, or advisors' insights, notes, and comments. Nonetheless, some student data are sparse; however, the storytelling model proposed in this dissertation is built in a way that makes it generalizable to include various sources of data. Second, another limitation that is directly related to the student data is that students' data is imbalanced since the majority of students are successful which makes it hard to build supervised learning models unless they overfit or impose an accuracy paradox due to a higher number of majority class examples caused by the imbalance. Therefore, we avoid using sampling techniques since undersampling might discard some potentially useful data about the students when building a predictive model. Oversampling works by making exact copies of existing examples, which makes overfitting likely. In fact, with oversampling it is quite common for a learner to generate a classification rule to cover a single, replicated, example. A second issue of oversampling is that it increases the number of training examples, thus increasing the learning time. To overcome this issue, this dissertation developed unsupervised learning models like anomaly detection models using clustering techniques. The anomaly detection aims to detect if an individual student's data instance can be considered anomalous compared to the rest of the data or if a collection of student data instances is anomalous compared to other students. Finally, conducting a large-scale evaluation over a long time is necessary to assess the effectiveness and potential impact of the student storytelling model on student learning, grades, and retention behavior. Most of the studies in the literature are limited and are more often in a controlled setting. However, this dissertation presents a longitudinal diary study and an in-depth contextual understanding of users' experiences with the creative student storytelling model over time. Unlike other common user research methods, such as surveys, or usability tests, this longitude diary study provides observations that are as rich or detailed as a true field study. However, this study does not assess the effectiveness of the storytelling model on student learning. Although there are limitations, this dissertation has developed a new model for sense-making from diverse, complex, and heterogeneous data that can be the basis of future research.

8.4 Future Research Directions

This dissertation has constructed factors that facilitate LA domain experts' sense-making of diverse, complex, and heterogeneous student data. There are three distinct avenues for future research that will be discussed in this section. First, Section 8.4.1 discusses research to extend the student storytelling model. Second, Section 8.4.2 discusses research to evaluate the long-term impact on student performance. Finally, Section 8.4.3 discusses research to generalize the storytelling model into different levels and areas.

8.4.1 Extending the Storytelling Model to Other LA Aspects

The storytelling model introduced in this dissertation deals with the student data at the individual degree level. It is directly related to understanding and making sense of student data at the individual level in terms of their degree towards graduation. However, a promising future research direction is the extension of the storytelling model to include other aspects at different levels. For example, a storytelling model at the course level for faculty and instructors teaching a class. Rather than making sense of student data at the degree level, faculty and instructors can utilize the stories to understand and make sense of the student performance at the course level. In a classroom setting, it is hard to recognize and subsequently attend to a student's weakness. With an effective storytelling model that facilitates the faculty and instructors' sensemaking, these deficiencies can be identified on time. Once an area of weakness is identified, faculty and instructors will have the chance to early intervene with underperforming students. Another example is a storytelling model at the group level for leadership. Facilitating leadership's awareness of students at the group level or cohort level through an effective storytelling model can potentially enable leadership to measure key indicators of student performance, support student development, understand and improve the effectiveness of teaching practices, inform curriculum decisions and inform institutional decisions and strategy.

8.4.2 Involving Other Data Sources in the Storytelling Model

As discussed in Section 8.3 of this chapter, one of the limitations of this study is that students' data is sparse in terms of some student's demographic data, such as race, ethnicity, generational social class, student body demographics, geographic location of the institution, and socio-economic status of students. This data can be essential for students to be successful and makes it hard to build learning models that consider the full range of student data. Consequently, future research should investigate other sources of data for the storytelling model such as students' self-reported interactions recorded in the LMS logs, social media, or advisors' insights, notes, and comments. The storytelling model proposed in this dissertation is built in a way that makes it generalizable to include various sources of data.

8.4.3 Developing a Long Term Evaluation Study

The factors discussed in this dissertation emerged from the behavior of domain experts who are presented with large volumes of data and analytics about their students. The loop of LA does not include only domain experts but also students who may be looking for course advice, career advice, and advice on how to be more successful if they are struggling to complete their degrees. Therefore, a critical future research direction to pursue is a long-term evaluation that measures and assesses the potential impact of student storytelling on students' grades and retention behaviors.

8.5 Conclusion

This dissertation is to explore the role and advantages of adopting fundamental approaches from four fields of study: sensemaking in LA, creative storytelling, data storytelling, and explainable AI as a component of interactive learning analytics. More specifically, the overarching goal of this adoption is to facilitate and improve the LA's domain experts' sensemaking and decision-making of diverse, complex, and heterogeneous student data. Sensemaking is a core component of LA dashboards and tools, as the purpose of these tools is to provide users with the ability to become aware of, reflect upon, and make data-based decisions. The theoretical foundations of sensemaking in LA helps faculty leadership and advisors to make rational, informed decisions when advising their students. Creative storytelling and data storytelling help with sensemaking and learning because stories are easy to remember. Learning from a well-told story is remembered more accurately, and for far longer, than learning derived from facts and figures. Through stories, advisors are more likely to engage with messages that make them feel personally involved in the student experience by triggering an emotional response. By including the students' demographic information and incorporating Concepcion's story structure and Freytag's pyramid to decide the structure and plot of the student stories, this study attempts to trigger the advisors' emotional response and as a result, make them personally involved and immersed in the student experience. Moreover, Explainable AI aims to increase the user's trustworthiness of the system decisions, either through introspection or through a generated explanation. Hence, the introduced storytelling model is transparent to the advisors on how the student stories have been generated and how the contents of the story are selected from the student data model. This makes it more understandable and interpretable by advisors.

The notions and aspects related to the four fields of the study presented in this dissertation are used as an aspiration to design and develop FIRST; Finding Interesting stoRies about sTudents, which is an interactive system designed to support advisors in their meetings with students who may be looking for course advice, career advice, and advise on how to be more successful if they are struggling to complete their degree. FIRST includes access to a large range of information about students and presents that information as features in three temporal categories: background data, semester data, and outcome data. The interactive components of FIRST enable the advisor to select specific features of interest and read the student stories. The student stories are automatically generated using the features that are selected by the user, the features that indicate significant changes, and additional data about the student using rules that present a more complete story. The process for generating stories has 3 stages: sourcing the data, selecting and structuring the story components, and aggregation and lexicalization of the sentences.

The ethnographic studies presented in this dissertation address the core research questions put forth in the beginning. Depending on the focus and purpose of the study, a mixed-method approach was adopted to address and explore these research questions, comprising three main types of studies: focus groups, one-on-one interviews, and longitudinal diary study. First, the focus groups were selected when we were collecting users' opinions and attitudes through group discussion and conversation dynamics. Focus groups result in rich and varied insights because listening to others' experiences stimulates memories, ideas, and experiences in participants. Second, a one-on-one interview study was selected when we were aiming to gather detailed information and ask open-ended follow-up questions to participants. Interviews help to explain, better understand, and explore research subjects' opinions, behavior, experiences, and phenomenon. Finally, a diary study was selected when we were looking for a contextual understanding of users' experiences over time. In the diary study, we collected qualitative data about advisors' behaviors, activities, and experiences, how these behaviors evolve over time, and what influences these behaviors? Throughout this time, we asked the participants to keep a diary and log of the activities being studied. Unlike other common user research methods, such as surveys (which are designed to collect self-reported information about a user's habits and experiences outside of the context of the scenarios being studied), or usability tests (which yield observational information about a specific moment or planned set of confined interactions in a lab setting), this longitude diary study provides observations that are as rich or detailed as a true field study.

The sensemaking model introduced in this dissertation makes a useful theoretical contribution, bridging the gap between heterogeneous student data and insight discovery while establishing a platform to support sensemaking and decision-making of domain experts about students who might be at risk. The utility of the proposed model lies in its ability to process multiple sources of data, extract insightful, interesting, and unexpected information, and tell the story that lies behind the data in a natural way that all users with a wide range of expertise can make sense of. We believe our model will enable further evaluation and integration of how the fields of sensemaking in LA, creative storytelling, data storytelling, and explainable AI can be fruitfully combined. Currently, it incorporates a common instrument for researchers and domain experts to guide the creation of relevant analytics that can serve educators in their learning design processes. Although, it is still early to base learning and education fully on LA systems alone. However, the opportunities this discipline has to offer are to provide new support for learning activities and stimuli for reflection and intervention and these are opportunities that LA should pursue.

REFERENCES

[1] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos, "Learning analytics dashboard applications," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1500–1509, 2013.

[2] A. Dollár and P. S. Steif, "Web-based statics course with learning dashboard for instructors," *Proceedings of computers and advanced technology in education (CATE 2012)*, 2012.

[3] N. R. Aljohani and H. C. Davis, "Learning analytics and formative assessment to provide immediate detailed feedback using a student centered mobile dashboard," in *2013 Seventh International Conference on Next Generation Mobile Apps, Services and Technologies*, pp. 262–267, IEEE, 2013.

[4] E. Concepción, P. Gervás, and G. Méndez, "A common model for representing stories in automatic storytelling," in *6th International Workshop on Computational Creativity, Concept Invention, and General Intelligence. C3GI*, vol. 2017, 2017.

[5] G. Freytag, *Freytag's technique of the drama: an exposition of dramatic composition and art.* Scott, Foresman and Company, 1908.

[6] S. PLAN, "The national artificial intelligence research and development strategic plan," 2016.

[7] S. A. Popenici and S. Kerr, "Exploring the impact of artificial intelligence on teaching and learning in higher education," *Research and Practice in Technology Enhanced Learning*, vol. 12, no. 1, p. 22, 2017.

[8] B. Shneiderman, "Human-centered ai," *Issues in Science and Technology*, vol. 37, no. 2, pp. 56–61, 2021.

[9] M. O. Riedl, "Human-centered artificial intelligence and machine learning," *Human Behavior and Emerging Technologies*, vol. 1, no. 1, pp. 33–36, 2019.

[10] W. Xu, "Toward human-centered ai: a perspective from human-computer interaction," *interactions*, vol. 26, no. 4, pp. 42–46, 2019.

[11] K. E. Weick, *Sensemaking in organizations*, vol. 3. Sage, 1995.

[12] N. Dixon, "Sensemaking guidelines–a quality improvement tool. med qic medicare quality improvement," 2003.

[13] G. Klein, B. Moon, and R. R. Hoffman, "Making sense of sensemaking 1: Alternative perspectives," *IEEE intelligent systems*, vol. 21, no. 4, pp. 70–73, 2006.

[14] M. J. Mahzoon, M. L. Maher, O. Eltayeby, W. Dou, and K. Grace, “A sequence data model for analyzing temporal patterns of student data.,” *Journal of Learning Analytics*, vol. 5, no. 1, pp. 55–74, 2018.

[15] K. E. Arnold and M. D. Pistilli, “Course signals at purdue: Using learning analytics to increase student success,” in *Proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 267–270, ACM, 2012.

[16] A. Essa and H. Ayad, “Student success system: risk analytics and data visualization using ensembles of predictive models,” in *Proceedings of the 2nd international conference on learning analytics and knowledge*, pp. 158–161, ACM, 2012.

[17] D. Clow, “The learning analytics cycle: closing the loop effectively,” 2012.

[18] L. Lockyer, E. Heathcote, and S. Dawson, “Informing pedagogical action: Aligning learning analytics with learning design,” *American Behavioral Scientist*, vol. 57, no. 10, pp. 1439–1459, 2013.

[19] I. Hoogland, K. Schildkamp, F. Van der Kleij, M. Heitink, W. Kippers, B. Veldkamp, and A. M. Dijkstra, “Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations,” *Teaching and Teacher Education*, vol. 60, pp. 377–386, 2016.

[20] F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes, “Automatic generation of textual summaries from neonatal intensive care data,” *Artificial Intelligence*, vol. 173, no. 7-8, pp. 789–816, 2009.

[21] S. Sergis and D. G. Sampson, “Teaching and learning analytics to support teacher inquiry: A systematic literature review,” in *Learning analytics: Fundaments, applications, and trends*, pp. 25–63, Springer, 2017.

[22] R. Kosara and J. Mackinlay, “Storytelling: The next step for visualization,” *Computer*, vol. 46, no. 5, pp. 44–50, 2013.

[23] Y. Zhang and X. Chen, “Explainable recommendation: A survey and new perspectives,” *arXiv preprint arXiv:1804.11192*, 2018.

[24] M. Van Lent, W. Fisher, and M. Mancuso, “An explainable artificial intelligence system for small-unit tactical behavior,” in *Proceedings of the National Conference on Artificial Intelligence*, pp. 900–907, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.

[25] J. D. Moore and W. R. Swartout, “Explanation in expert systemss: A survey,” tech. rep., UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY INFORMATION SCIENCES INST, 1988.

[26] R. L. Teach and E. H. Shortliffe, "An analysis of physician attitudes regarding computer-based clinical consultation systems," in *Use and impact of computers in clinical medicine*, pp. 68–85, Springer, 1981.

[27] L. R. Ye and P. E. Johnson, "The impact of explanation facilities on user acceptance of expert systems advice," *Mis Quarterly*, pp. 157–172, 1995.

[28] D. Gašević, V. Kovanović, and S. Joksimović, "Piecing the learning analytics puzzle: a consolidated model of a field of research and practice," *Learning: Research and Practice*, vol. 3, no. 1, pp. 63–78, 2017.

[29] V. Echeverria, R. Martinez-Maldonado, R. Granda, K. Chiluiza, C. Conati, and S. B. Shum, "Driving data storytelling from learning design," in *Proceedings of the 8th international conference on learning analytics and knowledge*, pp. 131–140, 2018.

[30] R. F. Murphy, "Artificial intelligence applications to support k–1 2 teachers and teaching," *RAND Corporation, DOI: https://doi. org/10.7249/PE315*, 2019.

[31] M. Van Harmelen and D. Workman, "Analytics for learning and teaching," *CETIS Analytics Series*, vol. 1, no. 3, pp. 1–40, 2012.

[32] A. Kerly, R. Ellis, and S. Bull, "Calmsystem: a conversational agent for learner modelling," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 89–102, Springer, 2007.

[33] K. T. Chui, D. C. L. Fung, M. D. Lytras, and T. M. Lam, "Predicting at-risk university students in a virtual learning environment via a machine learning algorithm," *Computers in Human Behavior*, 2018.

[34] N. Nur, N. Park, M. Dorodchi, W. Dou, M. J. Mahzoon, X. Niu, and M. L. Maher, "Student network analysis: A novel way to predict delayed graduation in higher education," in *International Conference on Artificial Intelligence in Education*, pp. 370–382, Springer, 2019.

[35] A. Wolff, Z. Zdrahal, D. Herrmannova, J. Kuzilek, and M. Hlosta, "Developing predictive models for early detection of at-risk students on distance learning modules," 2014.

[36] S. P. Choi, S. S. Lam, K. C. Li, and B. T. Wong, "Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions," *Journal of Educational Technology & Society*, vol. 21, no. 2, pp. 273–290, 2018.

[37] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," *Computers & Education*, vol. 51, no. 1, pp. 368–384, 2008.

[38] E. Kosba, V. Dimitrova, and R. Boyle, "Using student and group models to support teachers in web-based distance education," in *International Conference on User Modeling*, pp. 124–133, Springer, 2005.

[39] R. Mazza and V. Dimitrova, "Coursevis: A graphical student monitoring tool for supporting instructors in web-based distance courses," *International Journal of Human-Computer Studies*, vol. 65, no. 2, pp. 125–139, 2007.

[40] L. Ali, M. Hatala, D. Gašević, and J. Jovanović, "A qualitative evaluation of evolution of a learning analytics tool," *Computers & Education*, vol. 58, no. 1, pp. 470–489, 2012.

[41] K. Silius, T. Miilumäki, J. Huhtamäki, T. Tebest, J. Meriläinen, and S. Pohjolainen, "Studentsâ motivations for social media enhanced studying and learning," *Knowledge Management & E-Learning: An International Journal*, vol. 2, no. 1, pp. 51–67, 2010.

[42] O. Scheu and C. Zinn, "How did the e-learning session go? the student inspector," in *13th International Conference on Artificial Intelligence and Education (AIED 2007)*, IOS Press, 2007.

[43] S. Govaerts, K. Verbert, E. Duval, and A. Pardo, "The student activity meter for awareness and self-reflection," in *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp. 869–884, 2012.

[44] J. L. Santos Odriozola, K. Verbert, and E. Duval, "Empowering students to reflect on their activity with stepup!: two case studies with engineering students," in *Proceedings of ARETL'12 2nd workshop on Awareness and Reflection*, vol. 931, pp. 73–86, CEUR Workshop Proceedings, 2012.

[45] S. Klein, "Control of style with a generative grammar," *Language*, vol. 41, no. 4, pp. 619–631, 1965.

[46] J. R. Meehan, "The metanovel: writing stories by computer.," tech. rep., YALE UNIV NEW HAVEN CONN DEPT OF COMPUTER SCIENCE, 1976.

[47] J. Meehan, "Tale-spin, shank, rc and riesbeck, ck,(eds.), inside computer understanding: Five programs plus miniatures," *LEA*, 1981.

[48] T. Veale, "Creativity as a web service: A vision of human and computer creativity in the web era," in *2013 AAAI Spring Symposium Series*, 2013.

[49] P. Gervás, "Story generator algorithms," *The Living Handbook of Narratology*, vol. 19, 2012.

[50] G. Lakoff, "Structural complexity in fairy tales (the study of man)," in *Summer Meeting of the Linguistic Society of America*, pp. 128–150, 1972.

[51] J. M. Mandler and N. S. Johnson, "Remembrance of things parsed: Story structure and recall," *Cognitive psychology*, vol. 9, no. 1, pp. 111–151, 1977.

[52] L. Pemberton, "A modular approach to story generation," in *Proceedings of the fourth conference on European chapter of the Association for Computational Linguistics*, pp. 217–224, Association for Computational Linguistics, 1989.

[53] J. R. Meehan, "Tale-spin, an interactive program that writes stories.," in *Ijcai*, vol. 77, pp. 91–98, 1977.

[54] N. Dehn, "Story generation after tale-spin.," in *IJCAI*, vol. 81, pp. 16–18, 1981.

[55] R. Pérez y Pérez, *MEXICA: a computer model of creativity in writing.* PhD thesis, University of Sussex, 1999.

[56] M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *Journal of Artificial Intelligence Research*, vol. 39, pp. 217–268, 2010.

[57] C. León and P. Gervás, "Creativity in story generation from the ground up: Non-deterministic simulation driven by narrative.," in *ICCC*, pp. 201–210, 2014.

[58] P. Gervás, "Propp's morphology of the folk tale as a grammar for generation," in *2013 Workshop on Computational Models of Narrative*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013.

[59] P. Gervás Gómez-Navarro, "Reviewing proppâs story generation procedure in the light of computational creativity," *AISB 2014*, 2014.

[60] V. Propp, "Morphology ofthe folk tale. 1928," *Austin: U of Texas P*, 1968.

[61] M. O. Riedl and V. Bulitko, "Interactive narrative: An intelligent systems approach," *Ai Magazine*, vol. 34, no. 1, pp. 67–67, 2013.

[62] R. P. y Pérez and M. Sharples, "Three computer-based models of storytelling: Brutus, minstrel and mexica," *Knowledge-based systems*, vol. 17, no. 1, pp. 15–29, 2004.

[63] L. Ferres, A. Parush, S. Roberts, and G. Lindgaard, "Helping people with visual impairments gain access to graphical information through natural language: The igraph system," in *International Conference on Computers for Handicapped Persons*, pp. 1122–1130, Springer, 2006.

[64] L. Iordanskaja, M. Kim, R. Kittredge, B. Lavoie, and A. Polguere, "Generation of extended bilingual statistical reports," in *COLING 1992 Volume 3: The 15th International Conference on Computational Linguistics*, 1992.

[65] K. Kukich, "Design of a knowledge-based report generator," in *Proceedings of the 21st annual meeting on Association for Computational Linguistics*, pp. 145–150, Association for Computational Linguistics, 1983.

[66] B. Bohnet, F. Lareau, L. Wanner, *et al.*, "Automatic production of multilingual environmental information.," in *EnviroInfo (2)*, pp. 59–66, 2007.

[67] A. Nakasone and M. Ishizuka, "Storytelling ontology model using rst," in *Proceedings of the IEEE/WIC/ACM international conference on Intelligent Agent Technology*, pp. 163–169, IEEE Computer Society, 2006.

[68] E. Goldberg, N. Driedger, and R. I. Kittredge, "Using natural-language processing to produce weather forecasts," *IEEE Expert*, vol. 9, no. 2, pp. 45–53, 1994.

[69] J. Coch, "Interactive generation and knowledge administration in multimeteo," in *Proc. 9th International Workshop on Natural Language Generation (INLG-98), Aug.*, 1998.

[70] S. Sripada, E. Reiter, and I. Davy, "Sumtime-mousam: Configurable marine weather forecast generator," *Expert Update*, vol. 6, no. 3, pp. 4–10, 2003.

[71] M. Kahn, L. Fagan, and L. Sheiner, "Combining physiologic models and symbolic methods to interpret time-varying patient data," *Methods of information in medicine*, vol. 30, no. 03, pp. 167–178, 1991.

[72] D. Hüske-Kraus, "Suregen-2: A shell system for the generation of clinical documents," in *Demonstrations*, 2003.

[73] M. D. Harris, "Building a large-scale commercial nlg system for an emr," in *Proceedings of the Fifth International Natural Language Generation Conference*, pp. 157–160, 2008.

[74] E. Reiter, R. Robertson, and L. M. Osman, "Lessons from a failure: Generating tailored smoking cessation letters," *Artificial Intelligence*, vol. 144, no. 1-2, pp. 41–58, 2003.

[75] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311–318, Association for Computational Linguistics, 2002.

[76] S. Banerjee and A. Lavie, "Meteor: An automatic metric for mt evaluation with improved correlation with human judgments," in *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pp. 65–72, 2005.

[77] C.-Y. Lin, "Rouge: Recall-oriented understudy for gisting evaluation," in *Proceedings of the Workshop on Text Summarization*, 2003.

[78] N. Tintarev and J. Masthoff, "A survey of explanations in recommender systems," in *2007 IEEE 23rd international conference on data engineering workshop*, pp. 801–810, IEEE, 2007.

[79] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[80] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model," *The Annals of Applied Statistics*, vol. 9, no. 3, pp. 1350–1371, 2015.

[81] C. Rudin, "Algorithms for interpretable machine learning," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1519–1519, 2014.

[82] A. Vellido, J. D. Martín-Guerrero, and P. J. Lisboa, "Making machine learning models interpretable.," in *ESANN*, vol. 12, pp. 163–172, Citeseer, 2012.

[83] T. Wang, C. Rudin, F. Doshi-Velez, Y. Liu, E. Klampfl, and P. MacNeille, "A bayesian framework for learning rule sets for interpretable classification," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 2357–2393, 2017.

[84] D. Doran, S. Schulz, and T. R. Besold, "What does explainable ai really mean? a new conceptualization of perspectives," *arXiv preprint arXiv:1710.00794*, 2017.

[85] G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, pp. 1–15, 2018.

[86] A. K. Dey and A. Newberger, "Support for context-aware intelligibility and control," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 859–868, 2009.

[87] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Building interpretable classifiers with rules using bayesian analysis," *Department of Statistics Technical Report tr609, University of Washington*, vol. 9, no. 3, pp. 1350–1371, 2012.

[88] B. Y. Lim, O. Ayalon, and E. Toch, "Reducing communication uncertainty with social intelligibility: Challenges and opportunities," in *CHI 2017 Workshop on Designing for Uncertainty in HCI*, 2017.

[89] B. Y. Lim and A. K. Dey, "Toolkit to support intelligibility in context-aware applications," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pp. 13–22, 2010.

[90] J. Vermeulen, G. Vanderhulst, K. Coninx, and K. Luyten, "Answering why and why not questions in ubiquitous computing," 2009.

[91] J. Vermeulen, G. Vanderhulst, K. Luyten, and K. Coninx, "Pervasivecrystal: Asking and answering why and why not questions about pervasive computing applications," in *2010 Sixth International Conference on Intelligent Environments*, pp. 271–276, IEEE, 2010.

[92] A. Das, H. Agrawal, L. Zitnick, D. Parikh, and D. Batra, "Human attention in visual question answering: Do humans and deep networks look at the same regions?," *Computer Vision and Image Understanding*, vol. 163, pp. 90–100, 2017.

[93] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[94] T. Kulesza, S. Stumpf, M. Burnett, and I. Kwan, "Tell me more? the effects of mental model soundness on personalizing an intelligent agent," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1–10, 2012.

[95] T. Kulesza, W.-K. Wong, S. Stumpf, S. Perona, R. White, M. M. Burnett, I. Oberst, and A. J. Ko, "Fixing the program my computer learned: Barriers for end users, challenges for the machine," in *Proceedings of the 14th international conference on Intelligent user interfaces*, pp. 187–196, 2009.

[96] D. Norman, *The design of future things*. Basic books, 2009.

[97] A. Schmidt and T. Herrmann, "Intervention user interfaces: a new interaction paradigm for automated systems," *Interactions*, vol. 24, no. 5, pp. 40–45, 2017.

[98] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: understanding rating dimensions with review text," in *Proceedings of the 7th ACM conference on Recommender systems*, pp. 165–172, ACM, 2013.

[99] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 83–92, ACM, 2014.

[100] X. He, T. Chen, M.-Y. Kan, and X. Chen, "Trirank: Review-aware explainable recommendation by modeling aspects," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1661–1670, ACM, 2015.

[101] J. McAuley, C. Targett, Q. Shi, and A. Van Den Hengel, "Image-based recommendations on styles and substitutes," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 43–52, ACM, 2015.

[102] Y. Hou, N. Yang, Y. Wu, and S. Y. Philip, "Explainable recommendation with fusion of aspect information," *World Wide Web*, vol. 22, no. 1, pp. 221–240, 2019.

[103] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural attentional rating regression with review-level explanations," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pp. 1583–1592, International World Wide Web Conferences Steering Committee, 2018.

[104] X. Chen, H. Xu, Y. Zhang, J. Tang, Y. Cao, Z. Qin, and H. Zha, "Sequential recommendation with user memory networks," in *Proceedings of the eleventh ACM international conference on web search and data mining*, pp. 108–116, ACM, 2018.

[105] Z. Ren, S. Liang, P. Li, S. Wang, and M. de Rijke, "Social collaborative viewpoint regression with explainable recommendations," in *Proceedings of the tenth ACM international conference on web search and data mining*, pp. 485–494, ACM, 2017.

[106] L. Quijano-Sanchez, C. Sauer, J. A. Recio-Garcia, and B. Diaz-Agudo, "Make it personal: a social explanation system applied to group recommendations," *Expert Systems with Applications*, vol. 76, pp. 36–48, 2017.

[107] C.-H. Tsai and P. Brusilovsky, "Providing control and transparency in a social recommender system for academic conferences," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 313–317, ACM, 2017.

[108] C.-H. Tsai and P. Brusilovsky, "Explaining social recommendations to casual users: Design principles and opportunities," in *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, p. 59, ACM, 2018.

[109] Y. Wu and M. Ester, "Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pp. 199–208, ACM, 2015.

[110] K. Bauman, B. Liu, and A. Tuzhilin, "Aspect based recommendations: Recommending items with the most valuable aspects based on user reviews," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 717–725, ACM, 2017.

[111] S. Seo, J. Huang, H. Yang, and Y. Liu, "Interpretable convolutional neural networks with dual local and global attention for review rating prediction," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 297–305, ACM, 2017.

[112] B. Abdollahi and O. Nasraoui, "Explainable matrix factorization for collaborative filtering," in *Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 5–6, International World Wide Web Conferences Steering Committee, 2016.

[113] S. Chang, F. M. Harper, and L. G. Terveen, "Crowd-based personalized natural language explanations for recommendations," in *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 175–182, ACM, 2016.

[114] O.-J. Lee and J. J. Jung, "Explainable movie recommendation systems by using story-based similarity.," in *IUI Workshops*, 2018.

[115] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo, "Ripple network: Propagating user preferences on the knowledge graph for recommender systems," *arXiv preprint arXiv:1803.03467*, 2018.

[116] A. Passant, "dbrecâmusic recommendations using dbpedia," in *International Semantic Web Conference*, pp. 209–224, Springer, 2010.

[117] C. L. Kraus, *A news recommendation engine for a multi-perspective understanding of political topics*. PhD thesis, Master Thesis, Technical University of Berlin, 2016.

[118] J. Gao, N. Liu, M. Lawley, and X. Hu, "An interpretable classification framework for information extraction from online healthcare forums," *Journal of healthcare engineering*, vol. 2017, 2017.

[119] B. Abdollahi and O. Nasraoui, "Using explainability for constrained matrix factorization," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 79–83, ACM, 2017.

[120] G. Peake and J. Wang, "Explanation mining: Post hoc interpretability of latent factor models for recommendation systems," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2060–2069, ACM, 2018.

[121] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," *Text Summarization Branches Out*, 2004.

[122] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. de Rijke, "Explainable fashion recommendation with joint outfit matching and comment generation," *arXiv preprint arXiv:1806.08977*, 2018.

[123] R. Gunning, "The technique of clear writing," 1952.

[124] R. Flesch, "A new readability yardstick.," *Journal of applied psychology*, vol. 32, no. 3, p. 221, 1948.

[125] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," 1975.

[126] R. Senter and E. A. Smith, "Automated readability index," tech. rep., CINCINNATI UNIV OH, 1967.

[127] G. H. Mc Laughlin, “Smog grading-a new readability formula,” *Journal of reading*, vol. 12, no. 8, pp. 639–646, 1969.

[128] S. Ouyang, A. Lawlor, F. Costa, and P. Dolog, “Improving explainable recommendations with synthetic reviews,” *arXiv preprint arXiv:1807.06978*, 2018.

[129] T. J. Leeper, *MTurkR: Access to Amazon Mechanical Turk Requester API via R*, 2017. R package version 0.8.0.

[130] T. J. L. Pablo Barbera, Ken Benoit, *crowdflower: R Client for the CrowdFlower API*, 2017. vers.

[131] K. Golmohammadi and O. R. Zaiane, “Time series contextual anomaly detection for detecting market manipulation in stock market,” in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–10, IEEE, 2015.