AdVizor: Using Visual Explanations to Guide Data-Driven Student Advising

Riley Weagant
Ontario Tech University

Zixin Zhao [*]
Ontario Tech University

Adam Bradley
Ontario Tech University

Christopher Collins [†]
Ontario Tech University

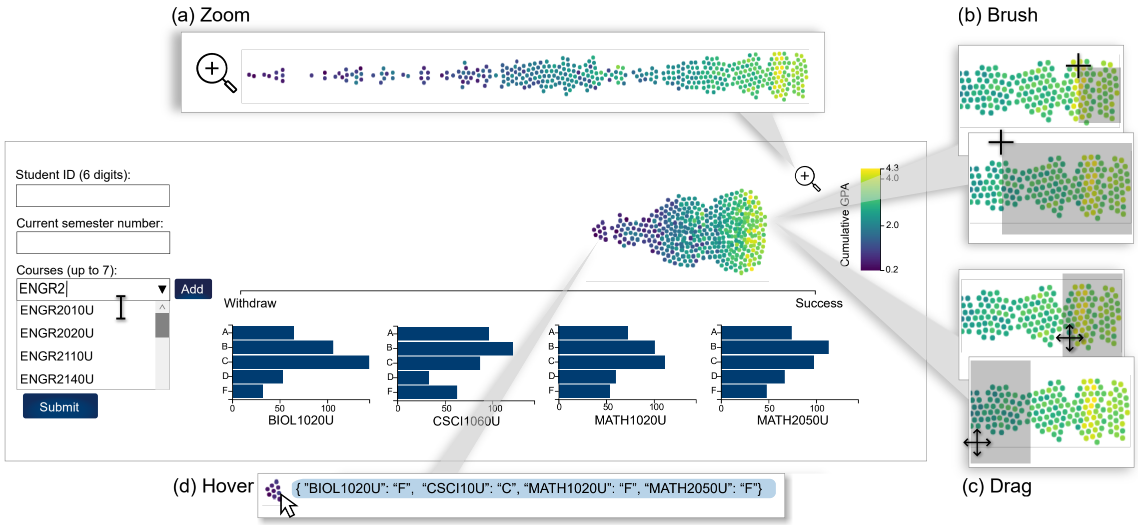

Figure 1: The main AdVizor interface consisting of a drop-down input menu and an interactive visualization. Users can (a) zoom into the beeswarm graph, (b) + brush a selection of plot points to inspect selected area's the grade distributions, (c) drag the selected area to somewhere else along the graph, and (d) hover over a point on the plot to see the grades predicted for that particular point

Abstract

Academic advising can positively impact struggling students' success. We developed **AdVizor**, a data-driven learning analytics tool for academic risk prediction for advisors. Our system is equipped with a random forest model for grade prediction probabilities uses a visualization dashboard to allows advisors to interpret model predictions. We evaluated our system in mock advising sessions with academic advisors and undergraduate students at our university. Results show that the system can easily integrate into the existing advising workflow, and visualizations of model outputs can be learned through short training sessions. AdVizor supports and complements the existing expertise of the advisor while helping to facilitate advisor-student discussion and analysis. Advisors found the system assisted them in guiding student course selection for the upcoming semester. It allowed them to guide students to prioritize the most critical and impactful courses. Both advisors and students perceived the system positively and were interested in using the system in the future. Our results encourage the development of intelligent advising systems in higher education, catered for advisors.

Keywords: Information Visualization, Academic Risk Prediction, Academic Advising, Learning Analytics

[*]e-mail:zixin.zhao@ontariotechu.net
[†]e-mail:Christopher.Collins@ontariotechu.ca

1 Introduction

Advising synthesizes and contextualizes students' educational experience at higher education institutions by providing them with the help needed to be successful pursing their degree. Student attrition is a complex subject, and although the reason behind withdrawing varies [29], it is often attributed to lower grades or feeling isolated [18]. Providing frequent advising to at-risk students can improve their likelihood of passing a course [32]. However, despite academic advising proving to be an important aspect of student retention, few technological tools were made to aid advisors [7, 15].

Past models have used machine learning are used to predict grades in open learner models (OLM) [6] to allow students view the grade predictions to support their learning process. However, showing predicted to students have been shown to negatively affect their performance [8]. Therefore, we propose advisors to serve as intermediaries to interpret the predictions and provide suggestions based on data-driven methods. So, to aid in identifying and providing guidance to at-risk students, we present a learning analytics dashboard (LAD) system to assist advisors during advising workflows and help identify possible student performance problems early. Our system, **AdVizor**, presents advisors with many potential future scenarios generated and classified by a machine learning algorithm. AdVizor presents advisors with an interactive visualization dashboard populated with potential future scenarios generated and classified by a machine learning algorithm. During advising sessions, advisors can use the dashboard to explore the scenarios to determine potential factors affecting the student's performance and guide their decisions towards futures with a higher likelihood of success. Our work shows that academic advisors and students benefit by having a system to facilitate course planning and monitoring of student learning using interactive visualization of machine learning model outputs. Our work's main contributions are: a learning analytic dashboard for advisors with interactive visualization for easy interpretation of ML model outputs to support decision-making, and insights from a mock advising study involving our expert-focused system with advisors and students.

2 Related Work

2.1 Predicting Attrition

Identifying risk factors for student attrition has been well investigated for many years and continues to be an area of active research [23]. The complex nature of student retention has been explored in different disciplines. The decision to withdraw can be described as a combination of social integration, academic performance, and commitment to personal goals and the institution [29].

Data mining techniques have often been employed to determine the causes of attrition. Superby et al. found that the factors correlated with student withdrawal include personal history, student involvement in their studies, and student perception [27]. In addition to causes for attrition, prediction algorithms like decision tree-based classifiers, logistic regression, linear regression, naive Bayes, random forest, explainable boosting machines, and neural networks were used to predict student attrition [9–11,16,21,24]. Delen showed that using decision trees, random forest, neural networks, and linear discriminant analysis leads to an overall 81% accuracy for attrition prediction [10], informed by these results, we chose to these these four models with our student data for retention prediction.

Past works identified demographic information, course difficulty (e.g. annotating "gatekeeper course") [1], and social behaviours [3] to be important features for predicting attrition. Superby et al. identified and grouped factors correlated with student withdrawal like personal history, student involvement in their studies, and student perception [27]. Other methods, like the one proposed by Sweeny et al., focused on academic performance used student grades for a course given their past academic performance and contextual information about the courses to predict attrition [28]. Their proposed model reports promising results but requires large amounts of up-to-date information about students' demographics, instructors, and course details. As these sources are expensive to gather and maintain, our work explores ways to use minimal data to predict attrition. Additionally, although all models proposed attempt to improve prediction fairness and mitigate bias, systems are oftentimes unable to take personal circumstances into account when making decisions [19], and there is higher fairness by excluding demographic data in the predictive model [2]. Therefore, we chose to mitigate model bias due to personal information by excluding it in model training.

2.2 Visualizations for Machine Learning

In AdVizor, we aimed to make predictions made by machine learning models more easily interpretable to academic advisors who do not come from machine learning backgrounds. Most work involving visualization and machine learning attempts to improve model performance by analyzing the input data and trying to interpret a machine learning algorithm's "black box". Frank et al. explored visualizing predictor class probabilities. Their approach involves plotting the class probability estimates and colouring the rectangular background accordingly [14]. The user chooses the class colours and plotting attributes. While this method is not new, they provided details on generalizing it to other classification models that can produce class probability estimates. Our approach borrows from this idea and uses the class probabilities to represent the uncertainty of our predictive model as a likelihood of success for a student.

Other work at the intersection of machine learning and visual analytics focuses on explaining and optimizing machine learning algorithms (e.g., [25,26]). Visual analytics to support decision-making has a long history (Ruppert provides a comprehensive review [22]). Similar to the tool proposed by Wexler et al. [30] which lets non-expert users probe, visualize, and analyze machine learning systems with little coding, our work uses interactive visualizations as an auxiliary tool to explore predictive analytics.

2.3 Visual Analytic Tools in Education

Visual analytics is often used for decision-making as it allows for multidimensional data to be encoded to make patterns apparent, using color, shape, position, and other attributes. In education research, visualizing student models and student learning is mainly done using LADs and OLMs. Works with OLMs tend to have students as the target user [4,6], however a study by Chaturapruek et al. [8] showed that students with direct access to course planning applications showing potential outcome information led to a lower overall GPA. In comparison, LADs often have the predictive tools that OLMs lack but are mainly used by institution decision-makers, so to allow students access to predicted performance indirectly, we looked to advisors to serve as a mediator between the system and the student.

Earlier research in LADs for academic advisors includes work by Du et al. [12], which uses multivariate correlation visualizations to generate and evaluate potential outcomes in academic advising settings. They present a detailed analytic dashboard to use during advising sessions and allow advisors to formulate a temporal plan to optimize the likelihood of a desired student degree outcome [12]. However, due to the high computational power needed to create the visualizations, their system is unsuitable for short advising sessions common at many universities.

Other works focused on tools for real-time short-duration advising sessions showed that visual analytic dashboards with student grades were able to support dialogue during advising sessions and helped advisors give more specific feedback to students [7]. Additionally, systems that used predictive algorithms for academic risk as a measurement are useful for less experienced advisors [15]; however, it should be noted that advisors with more experience preferred to rely on their expertise and knowledge instead of a model prediction. More recently, Mendéz et al. created an interactive visualization tool to aid students in course selection by maximizing their predicted GPA [17] and later adapted their system for advisors [19]. Their prediction model used gradient boosting trees, with the student's previous grades, the number of times that course was taken, aggregated course difficulty, and term workload as inputs. Student users rely heavily on GPA prediction when choosing courses, which may hinder explorations of courses that could be more valuable to their future development. Therefore, Mendéz et al.'s results suggest that the presence of an advisor is crucial when grade prediction tools are used. Additionally, they found that advisors believed that dashboards are important are useful as they provide them with more credibility.

3 AdVizor System

3.1 Academic Advising at Our University

Our university, a public research university focused on science and technology located in Canada, separates an academic year into 4 semesters: Fall (Sep–Dec), Winter (Jan–Apr), Spring (May–Jun), and Summer (Jul–Aug). A 4.3 GPA scale is used, and there are three categories related to student GPA: clear standing (GPA $>$ 2.00), probation (first instance GPA $<$ 2.00), and suspension (failed conditions of probation). A student with probation status is required to achieve a GPA $>$ 2.00 and contact an academic advisor.

An average advising session lasts around 20 minutes and is conducted one-on-one. Advising at our university is initiated by student request and is optional, except in the case of probation. During advising sessions, advisors assist students with course planning and adjusting to university life, provide resources, and answer general

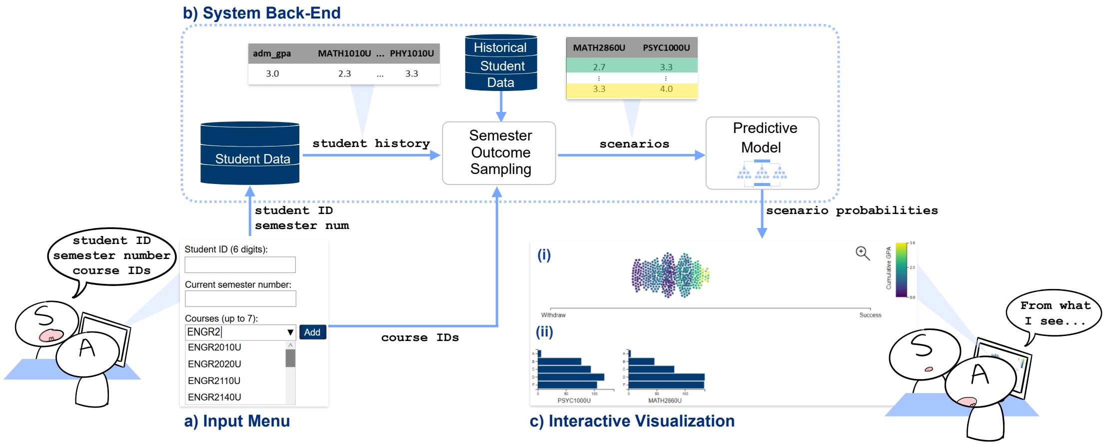

Figure 2: Application schematic and data flow through the system. The student information is first entered into the **a) input menu**. The student ID, a numerical number for the current semester, and courses planned for the next semester are used in the **b) system back-end** to fetch the student grade history and used to generate possible scenarios. Scenarios possibilities are generated by a predictive model and passed back to the front-end. Scenario possibilities are shown in an **c) interactive visualization** composed of a *(i) beeswarm visualization* for a global overview of the predicted grade scenarios *(ii) grade possibility bar graphs* give a more granular grade distribution for each course.

inquiries. Advisors may also redirect students to other resources like the career and peer tutoring centres.

3.2 System Design

We defined student success to mean the completion of an academic term without voluntary or involuntary withdrawal from the institution. Our system, **AdVizor**, predicts the likelihood a student remains in university in the following semester given the courses they want to take, and the system diagram can be seen in Figure 2. AdVizor allows advisors to explore outcome scenarios based on the student's own past academic performance, historical student performance, and the courses they wish to take. Advisors can try different combinations of the courses to view varying prediction results. The visualization conveys the prediction as probabilities to match the underlying representation, we did this to allow advisors to come to their own conclusion. To design AdVizor, we used an iterative design process with stakeholders at our university to formulate design criteria. Using these design criteria, we then selected several machine learning models suggested by prior works which fit our criteria. We chose the best performing model to use in our final system. We then ran some pilot testing on the system to ensure its usability.

3.2.1 Developing System Design Criteria

We used an iterative design process to develop an advisor-focused learning analytics dashboard. Our initial iterative steps involved developing medium-fidelity prototypes for meetings with stakeholders at our university, during the meetings, suggestions and problems were pointed out to be fixed. The following design criteria (DC) were set for our system.

DC1. The system must integrate with the existing advising workflow.

DC2. The users of this system should be trained advisors rather than students.

DC3. There must be a way to input courses quickly and provide results within 10 minutes.

DC4. Results should be presented as probabilities, and definite results should be avoided since it could discourage students.

DC5. Student retention is the main goal and should be used to measure success.

Parameter	Value
`n_estimators`	50
`criterion`	gini
`max_leaf_nodes`	860
`max_depth`	33
`min_samples_split`	9
`min_samples_leaf`	9
`max_features`	sqrt
`oob_score`	False
`class_weight`	None

Table 1: Parameters for the `RandomForestClassfier` model and the values used for our prediction model.

3.2.2 Building an Academic Prediction Model

To build our prediction model, we first processed student data and then selected the best performing machine learning model which fit our criteria. The institutional data we used was gathered from our university between 2003 and 2015. This totalled to over 758k rows of course records stored in a MySQL database, where each row contains records for a course for a specific semester. We had a separate database for student data containing over 148k entries. The student data includes student demographic information, admission average, courses taken that semester, grades obtained in courses taken, and academic standing history. However, we did not use demographic variables in our predictor to allow experienced users, advisors in our case, to make informed decisions, and echo Baker's [2] sentiment that there is greater fairness if demographic information is not used as predictors.

We merged the the course dataset with student identifiers to create a course history dataset made of *student vectors*. Each student vector contains the student ID, semester number, and GPA numerical equivalents of the letter grades received for each course offered at our university as a feature. If the student did not take a course, the feature value was set to 0. In the end, we had 92,633 samples and split the data into a 60:40 split for training and testing.

Based on models used in prior student retention literature [10, 21, 27, 28], we tested three lightweight predictive models (**DC3**): logistic regression, random forest, and neural network. We used KNIME, a visual analytics platform, and the scikit-learn library

Algorithm	Accuracy	Precision	Recall
Logistic Regression	81.6%	89.6%	88.1%
Random Forests	89.8%	91.7%	96.6%
Neural Network	84.0%	84.6%	98.8%

Table 2: Performance metrics of the tested algorithms using 10-fold cross-validation.

(RandomForestClassfier model) with Python to compare the model performance. We found that the random forest model performed the best overall with our training and testing dataset at 89.8% accuracy. The averaged performance of the algorithms is shown in Table 2. It should be noted that the computational complexity of random forest models increases proportionally to the data size so it is not suitable for large datasets, in our case, we are using minimal input data during prediction so it did not slow down prediction time during testing. After selecting our model, we performed parameter tuning using a smaller subset of data to select the best values, our results are shown in Table 1.

3.2.3 Generating Student Outcome Scenarios

From Figure 2b showing the back-end of the system. AdVizor uses student IDs and current semester numbers to query the student database to fetch the current student's past academic records. The student's past academic records contains a row for each of the past semesters they have attended our university along with their admission average and grades obtained from courses they took that semester (a snippet of the record can be seen in Figure 2b). The student's history, past student records, and the courses they wish to take in the following semester are fed into a semester outcome sampler, producing several discrete predictions of the student's grades in their chosen courses. The sampler creates a sub-sample of the historical students' academic performances by creating a normal distribution centred at the mean of the current student's predicted GPA with the historical student dataset's standard deviation. Although the historical grade distribution is not normal, we assume each individual student is independent of the distribution of all students and that their predicted performance is normal. Each sample from the sub-sampled normal graph is treated as a *scenario* prediction. The scenarios are fed into the predictive model as *student vectors*, and the confidence level of each prediction is displayed on an interactive visualization.

3.2.4 Interface Workflow

At the start of a session, the advisor needs to ask for the student ID, semester number, and courses the student wishes to to take so they can enter the information in the input menu, as shown in Figure 2. When the advisor inputs course's name, a drop-down list populated with courses with matching text segments is displayed for ease of input (**DC3**). The information is then submitted into the back-end, Figure 2b, and the possible scenarios and probabilities of each one is returned to the interactive visualization component, Figure 2c. On the top of the visualization is an overview of all the possible scenarios shown using a beeswarm plot, Figure 2(i).

This visualization shows each possible outcome or scenario represented by a point in the plot with its horizontal position corresponding to the predicted level of success. Advisors can view the plotted points over the entire probability of success range [0,1] or view a zoomed-in portion containing the predicted scenarios. Advisors can zoom into the beeswarm plot using the magnifying glass icon at the top left of the beeswarm visualization, see Figure 1a. The beeswarm plot was chosen due to its ability to display multiple points on its axis while reducing over-plotting (multiple items drawn at the same position) [13,31] and it is beneficial over alternatives like

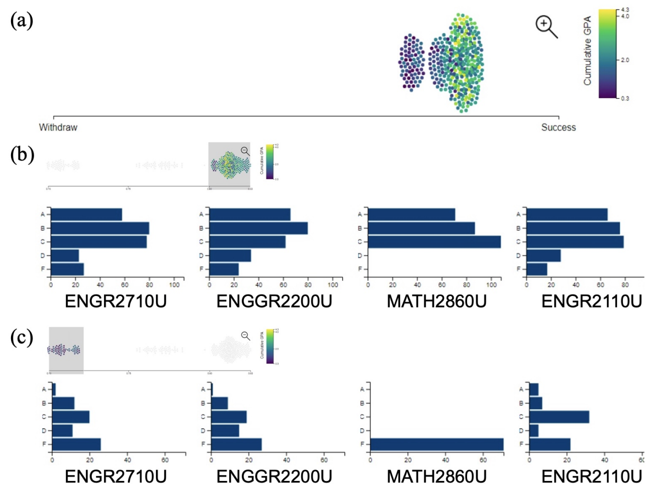

Figure 3: Use case study with a persona named Holly. (a) Beeswarm plot showing Holly's likelihood of success given the four required courses. All of the scenarios lie above the mid-point, indicating a high likelihood of success. (b) Higher likelihood scenarios selected on the zoomed-in beeswarm plot show similar grade distributions for the three engineering courses but only scenarios with no possibility of D or F in math. (c) Lower likelihood scenarios were selected on the beeswarm, and the bar graph showed that these points correlate with a low math course grade.

discrete histograms because it reduces the frequency of misleading clusters. To avoid setting a definite bar for acceptable GPAs, the axis endpoints are labelled *Withdraw* and *Success* and left for the advisor to interpret using their expertise (**DC4**, **DC5**). Each scenario plotted is coloured according to the overall GPA from the combination of courses for the given semester. To see the details about grades received for each course in a particular scenario, advisors can hover over the point and then a tooltip with the course codes and predicted grade is displayed, see Figure 1d. Advisors can click and horizontally drag (+ brush) on the x-axis to select points on the chart that fall within the brushed range, updating the bar charts below to represent grade distributions in the selected scenario(s), see Figure 1b. The brushed segment is highlighted and a lower opacity is applied to the rest of the scenarios outside of this segment. The segment can also be dragged to highlight different portions of the beeswarm plot, see Figure 1c.

Figure 2(ii) shows that the bar charts summarize the courses and grade information embedded within the beeswarm plot. A horizontal bar chart is created for each course the student wishes to take the following semester. The advisor may brush over the sections in the bar charts to show the exact number of occurrences for each course, as shown in Figure 1b. The main goal of these charts paired with the beeswarm plot is to allow advisors to see the success given a set of courses as a likelihood rather than with certainty and to determine the course with the most influence on a particular student's success. The bar charts can help correct misconceptions and show students that a scenario where they have a lower relative GPA can still lead to a good probability of success. Through interaction, advisors can explore and compare course grade distributions contributing to different scenarios in the beeswarm to discover key courses and grades most associated with overall success. This, in turn, can help students focus their study efforts.

3.3 System Use Case Study: Holly

We will use a persona named Holly to illustrate the usage of AdVizor in an example advising session. Holly is a software engineering student and just finished her first year. She took five courses in her first semester and received a cumulative GPA of 2.88. Although she performed well in most courses, she struggled with Calculus I and Physics I. In the second semester, she took six courses and received a 1.67 GPA leaving her with an overall cumulative GPA of 2.27. Of the six courses she took in the second semester, she did poorly in Engineering Dynamics, Calculus II, and Physics II. She was concerned about her falling GPA and possibly ending up on academic probation. Before registering for second year courses, Holly scheduled a meeting with an academic advisor to discuss her future.

Holly's meeting with her advisor starts with discussing her reason for coming in and her concerns about her grades. During this discussion, the advisor can enter the required courses AdVizor. Holly is still unsure which elective to take, so the advisor simply enters the 4 required courses. The resulting plot is shown in Figure 3a with all scenario predictions appearing right of the mid-point, meaning that Holly is predicted to succeed in all generated cases. Therefore, the advisor feels confident with the current course selection. To further investigate how to improve her likelihood of success, the advisor zooms in and brushes a portion of the bee swarm plot with a high likelihood of success with more yellow-coloured (higher GPA) points, Figure 3b. They see using the bar graphs that in all the highlighted scenarios, MATH2860U requires a grade above C. Then the advisor drags the selection window to the far left of the chart with lower likelihoods of success as shown in Figure 3c. The summary bar charts update to show which grades are present for each course in the selected scenarios with a high frequency of F's in MATH2860U. The advisor concludes that Holly needs to focus on doing well in MATH2860U, and a grade below a C in this course could lower her likelihood of success. This visual analysis could allow further discussion about reducing course load to spend more time on critical courses, and extracurricular activities. In comparison, during a traditional advising session, the advisor may not be able to come to conclusions as quickly. Additionally, given the stress Holly may be feeling during advising sessions, having both verbal and visual confirmation of the possibility of higher likelihood of success could be beneficial. Moreover, providing advisors with a system to affirm their opinions can help convince doubtful students.

4 Participant Study

4.1 Study Design

To evaluate the system's integration with the existing advising workflow we used mock advising sessions with real advisors working with undergraduate students given student personas from their faculty. Mock advising sessions were used due to privacy concerns about releasing student information. The rationale is that students playing personas from their faculties would be familiar enough with the course maps, graduation timetables, and course selections to represent students adequately in real advising sessions. We ran ten mock advising sessions, transcribed the recordings, and kept notes from interviews. We were able to conduct interviews with all the advisors. However, due to scheduling challenges, we could only conduct interviews with 5 out of the 10 student participants. All sessions were conducted on campus under a protocol approved by our institutional Research Ethics Board. Participants were free to opt out of any part of the study session.

We recruited four advising participants with at least five years of experience, through our professional circle, from the following three faculties: the Faculty of Engineering and Applied Science, the Faculty of Energy Systems and Nuclear Science (FESNS), and the Faculty of Science. Due to availability, two of the advisor participants were from the Faculty of Science. They were scheduled for 2-hour sessions of three phases: tutorial, mock advising sessions, and a follow-up interview. The tutorial used student personas to explain system usage, which has been shown to help data representation comprehension [5]. Then, multiple mock advising sessions were scheduled to observe the effect of increased familiarity with the tool over time. The first three columns of Table 3 summarize the session assignments.

Students in the second year or higher with prior academic advising experience were recruited through an open call using posters put around the university. Students were then screened to ensure we had sufficient and an even distribution of students for each faculty the advisor participants specialize in. The student participants were randomly assigned personas belonging to their faculty. They were provided a student ID, current semester number, a course and grade history, and a list of courses to take in the upcoming semester. Personas also included a short narrative containing information on academic standing and personal details (e.g. how long their commute to campus is). For student participants, the study consisted of three phases: learning about the assigned persona, a mock advising session, and a follow-up interview.

The observation room was set up to resemble an academic advising office, where the advisor sat on the far side of the table in front of the computer monitor, and the student sat across from them as they would during an advising session. We used a camera to record video and audio of the sessions, while we observed the sessions from the next room. Participants were asked to answer post-session interview questions after the mock session. The student participants were interviewed by a research team member immediately following their session, while advisor participants were interviewed after completing all of their scheduled sessions. We used the quantitative results obtained through analysis of the time usage during each session and qualitative analysis of the interview responses to evaluate the effect of AdVizor.

4.2 Study Results

To see how AdVizor integrates within the advising workflow we collected video recordings and transcripts of each session. Using the videos, we analyzed the advising workflow during session by categorizing what took place, when, and how it affected advising. Additionally, we conducted qualitative analysis using thematic coding on session transcripts to explore the effects of AdVizor. The thematic coding was conducted by researchers using NVivo, using an iterative process. Each researcher first did a pass to tag all the transcripts into specific categories. Then the two researchers gathered to discuss the themes they observed from the tags. After a round of discussion, previous tags were categorized into the larger themes. This process was repeated three times. We found through coding that AdVizor modified the course selection process, increased advisor's perceived credibility, and helped motivate students. We also present a summary of the features requested in post-task interviews and feedback on the study.

4.2.1 Advising Session Pattern with AdVizor

AdVizor was used in every session to aid the advising process. By colour encoding the common events during sessions and mapping them onto a timeline, Figure 4, we could compare and analyze how time was spent during the study sessions. The timelines in Figure 4 are scaled to reflect the ratio of time spent on each event across the entire session.

For most sessions, advisors started by introducing themselves to make the students feel more comfortable, followed by students giving their persona's student ID and courses for the next semester or explanations of their circumstances. In all the sessions, the advisor used the generated visualizations to give data-informed advice to students. In three separate sessions, see grey portions in Figure 4, there were miscellaneous chats between the student and advisor about the future implementation of AdVizor.

On average, 87% of the time was spent conversing, while the remaining 13% was spent in silence or the researcher stepping in to

Advisor	Department	Sessions	Showed screen	Changed advising
A	Engineering	1, 2	No	No
B	Science	3, 4, 5	Yes	Yes
C	Science	6, 7, 8	No	Yes
D	FESNS	9, 10	No	No

Table 3: Overview of advisor session assignments and results from interviews after conducting their mock advising sessions using AdVizor.

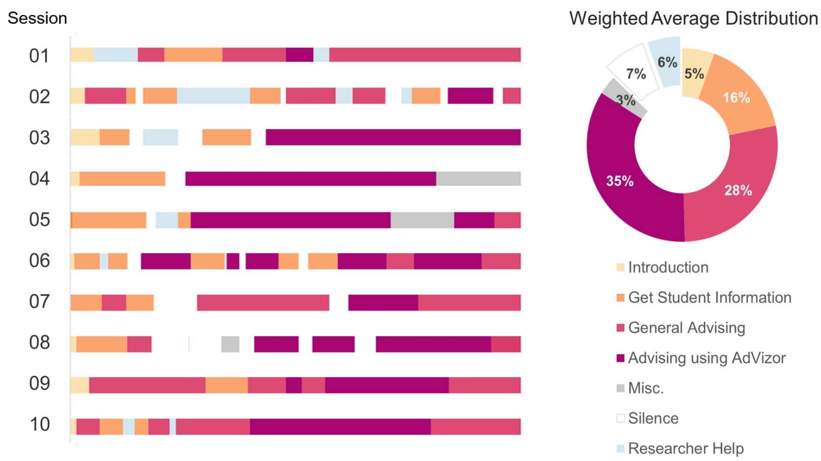

Figure 4: *(right)* Color-coded timeline of events that occurred during each mock session. *(left)* Time breakdown of how time was spent during sessions, represented using weighted average percentage.

troubleshoot. The time during the mock advising sessions was spent mainly on advising **(DC3)**, showing that AdVizor integrated into the session without the need for silent interpretation of the visualization at the expense of silence during advising. There were two events directly related to using AdVizor: getting student information to input into the system, and advising using the system. As shown in Figure 4, the portion of events influenced by AdVizor made up around 51% of the advising sessions. The system's usage was spread throughout the session, showing that it was easy to integrate into the advising session and allowed advisors to switch between different advising methods when needed **(DC1)**. The duration of the advising session did not differ from the usual expected 10-15 minutes, with sessions ranging from 7 to 21 minutes. A more controlled comparison experimentation would be needed to investigate the effect of AdVizor on the duration of advising.

Advisor B mentioned that automatic prediction methods lack context on the students' circumstances, so preferred that students did not have direct access to the interface but rather had an advisor act as a trained interpreter **(DC2)**. The tool seemed to fit seamlessly into the advising flow from the student's perspective **(DC1)**. From our interviews with students after the sessions, all students felt that advising with AdVizor did not feel different from usual, one student noted that they did not know that the advisor was using an additional tool. The experience was different for advisors in comparison, around half of the advisors found that the system changed how they conducted advising sessions, as seen in Table 3. Some advisors used the tool more than others, as can be seen when comparing session 1 (Advisor A) and session 5 (Advisor B). However, despite the differences in experience and usage, all the advisor participants said they would use such a system in the future.

4.2.2 Facilitating Course Selection

One way AdVizor was used during sessions was to help advisors gain a high-level and course-level view of a student's predicted performance. For the high-level overview, the beeswarm visualization was used to inform advisors about potential grade combinations the students could obtain and their predicted overall success rate. Three of the advisors (B, C, D) spent time testing different combinations of courses to maximize student success.

To show how more global effects are perceived, we can look at Figure 5 containing different graphs generated during an advising session with Advisor B. Due to the student persona's low GPA, the advisor first suggested trying out a lower course load of four courses, which yielded the beeswarm plot seen in Figure 5(a).

> *"So what if we were to look at a slightly different combination of courses... and maybe looking at taking three or four, as opposed to five?"*
> *- Advisor B*

After seeing many scenarios with high success rates, Advisor B inputted in the fifth course the student wanted to take. The resulting beeswarm, Figure 5(b), had a few scenarios spread across the left side of the visualization, so Advisor B expressed how if the student were to have a bad semester and let their grades fall, they would be at a much higher risk of probation. A few different course combinations were introduced, some involving courses they failed the previous semester. Finally, Advisor B felt confident with Figure 5(d) as all possible scenarios were on the right side of the beeswarm plot – correlated with a higher success rate – and due to the student's poor performance in the previous semester, the advisor also needed to ensure that the GPA required to have a high success rate was realistic, which in this case means fewer yellow coloured scenarios. Other advisors followed similar analysis methods, using the beeswarm plot to explore the effect of course combinations and gain a higher-level overview.

The high-level overview helped advisors with one dimension of course planning. The other common task facilitated by the system was assisting in the course suggestion process by helping advisors determine the students' important courses. During the mock advising sessions, advisors first varied the course inputted to see the

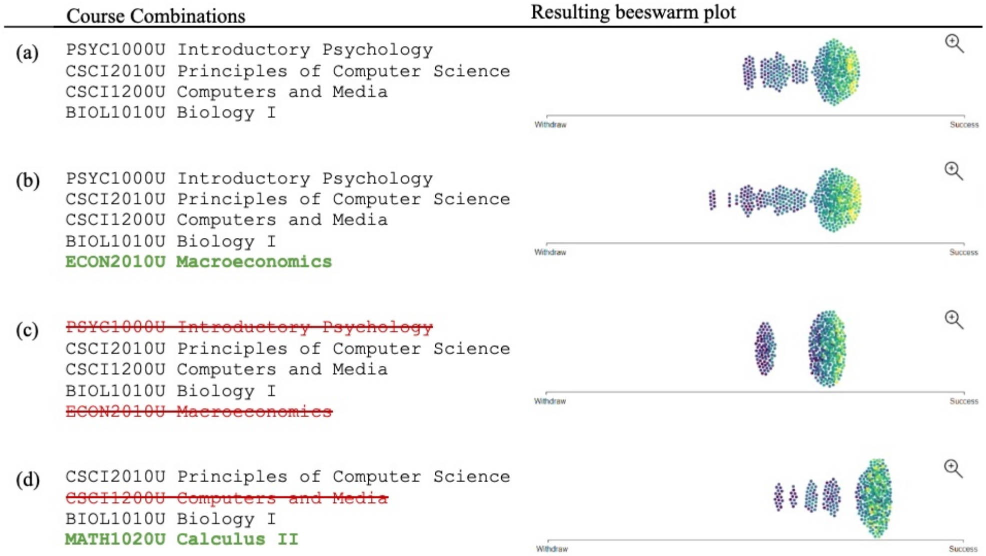

Figure 5: Advisor B's course combination iteration trials during Session 4 where the student came in with 5 desired courses. The advisor first (a) tried with a slightly reduced proposed course plan, then (b) added all the desired courses, then (c) tried a different combination but received a lower likelihood of success, and finally (d) settled with this course combination that maximized likelihood while not requiring a high GPA.

distribution of predicted scenarios using the beeswarm plot and then used the bar graphs to see exactly which courses the system thinks the student will struggle with the most. The suggestions are given using a combination of extracted information from the graphs and deductions from prior semester GPAs. Advisors often point out a particular course that has a large effect on the probability of success, for example,

> *"...based on your previous experience, you're going to probably have to put the most of your effort into physics."*
> *- Advisor B*

4.2.3 Increasing Advisor's Self-Perceived Credibility

Three advisors mentioned that although the system provides them with more credibility and confirms their deductions. They found that the visualizations confirmed their knowledge, giving them higher perceived confidence when speaking about their suggestions. For example, advisors would try to visualize the success rate of students if they were put on a reduced course load, since it

> *"...helped inform whether or not if a student had the option of reducing their course load."*
> *- Advisor C*

Advisors also found that they could make decisions based on their expertise, but the system was a helpful supplement. Three advisors mentioned that although the system provides them more validity and confirmation about their deductions during advising sessions, they still relied on prior experiences. Most found it useful as a complementary tool and that it gave them sufficient agency. One advisor described the system as getting a "glimpse into the future."

4.2.4 Helping Boost Student Motivation

We found that AdVizor also assisted advisors in motivating students by providing them with evidence of the possibility of success using data driven methods. Advisors generally use positive word choices and frame methods during advising to help with student motivation. For example, advisors will state something positive followed by something more negative. Students found that they felt the advising session boosted their confidence and enjoyed seeing the system show possibilities of obtaining high grades for courses that they thought their personas might struggle with. During the sessions where the advisor showed their monitor to the students, see Table 3, the students often asked more questions about clusters containing high GPA and asked which courses they were associated with. Students mentioned they would be more motivated in the course if they believed they could achieve higher grades. However, during the sessions, all advisors mentioned that although a high GPA is possible and students may seem to have a high probability of success, predictions are not absolute, and the student needs to put in the effort to realize it (**DC4**). The advisors communicated the severity and consequences of some potential outcomes, like academic probation and expulsion.

4.2.5 Requested Features

Interviews revealed several features that advisors felt would be useful additions to our system, including comparison screens and end-user controllable input features in the predictive model. Specifically, the advisor wanted to compare students' success if they sought supplemental external help versus no external help. We also saw advisors wait for the visualizations to reload after changing a single course and relying on memory to compare prior results. Therefore, allowing multiple visualizations to compare the success rate according to different course combinations would be useful. Some advisors expressed that cumulative GPA calculators would be useful to help students who want to maximize their GPA. Advisors found that showing the overall semester GPA prediction encoded using the colour bar was insufficient as students on probation are often worried about their cummulative GPA.

4.2.6 Study Realism Feedback

Due to privacy concerns, we conducted mock advising sessions. In interviews with the students, although three mentioned that the session felt natural overall, two felt that they struggled to immerse themselves in their persona at the start of the study. Many students noted that in real advising sessions, they would have a plan in place and have specific questions that they would ask the advisors; however, during the study, they felt unprepared and had to come up with plausible fake scenarios in a short period. Advisors also felt confused during the session due to unfamiliarity with the visualizations shown to them. Although we provided them with a 10-20 minute tutorial for the system, they preferred a refresher before each session to feel confident explaining the visualization to the students.

5 Discussion

Academic advising differs for each institution, with some universities having trained advisors to conduct advising, while others delegate the task to course professors. Due to the differences in how advising is conducted, it is difficult to produce comparative statistics between universities; accordingly, our results are presented as-is for context. We found that advisors were able to quickly learn how to interpret the visualizations, and by having advisors act as intermediaries to convey grade prediction allowed students to have a more holistic view of their academic performance and reduce negative emotions resulting from low success predictions.

There is also a learning curve to data literacy. Advisors struggled slightly initially during our study, but with continued use, they became more comfortable. For example, out of the ten sessions, six sessions needed the researcher to step in during the beginning, as shown in light blue in Figure 4, however after the initial help there were not many instances of advisors requiring help further into the session. Therefore, if the goal of a system is to be designed for advisors to use, it will be hard to design a short-term study to test out the system's full functionality. For future implementations, we recommend providing the advisor with a more comprehensive training period and allowing them to have extended experience with the system. Another option would be to add tool tips and help text next to visual components to refresh their memory.

Our findings show that AdVizor supported facilitating the dialogue between students and advisors by providing a wide range of personalized predictions for course performance, which the advisor could analyze to give students insights into which courses they should take in the upcoming semester. We found that showing outcomes as likelihood of success with a wide range of possible scenarios can aid the perception of dimensions that affect student success **(DC4, DC5)**. Advisors would examine all the possible "worst-case" and "best-case" scenarios to deduce which planned course's grades remained consistently low. As many of the actions advisors took during visualization exploration relied on intuition and experience, it is necessary to have trained professionals in control when interpreting the model outputs. In previous studies conducted with student systems [17], many of the students picked courses with the highest GPA outcomes per course to maximize their overall GPA. However, advisors during our study showed that it may be possible to maintain a higher overall semester GPA even if a student performs poorly in a particular course. Therefore, it may be beneficial for experienced and trained professionals to interpret predictions as it reduces the reliance on predictive modelling alone.Thus, by allowing advisors to be at the final step of model interpretation, experts are on-the-loop of system output interpretation. This enables them to give students advice based on both their expertise and supplemental help from visualizations of model predictions.

However, it should be noted that each advisor has their own method of advising, which is reflected by the varied usage of AdVizor during mock advising sessions. For instance, Advisor B used the system to test different combinations of courses and a lessened course load to see how that would affect student success. Students mentioned during interviews that they prefer personalized help, as it makes them feel like the advisors care about them. Having a bond with advisors and feeling like they belong at the university is a large part of student success [20, 29].

One of the reasons behind student withdrawal is the lack of motivation [29]; therefore, by picking out future courses with supplemental help from advisors, advising sessions can help students formulate a goal for themselves and create a purpose that propels them to become more successful. Since the system looks at students' performance in their next semester's courses, it forces them to think about the future. Given that the next semester's courses were predetermined in the mock sessions, we could not observe the process of advisors working with a student from scratch. However, we observed the advisors working with students to reorganize their course map according to each student's goals. Additionally, students with prior knowledge of course difficulty can negatively impact their performance [2, 8]. Therefore, having advisors interpret the predictions and maintain a positive attitude could inspire students to believe in themselves.

Through our work, we hope future researchers can explore more options involving trained educational experts as users of intelligent systems. With the vast amount of works using intelligent models for course recommends for students and grade prediction, there is a lack of focus on facilitating education provider's needs. Predictive AI models can use reckoning decisions based on calculative prediction but cannot use judgment based on practical wisdom. Unlike human advisors, models cannot consider individual preferences, like whether a student worked well with a professor's instruction style.

5.1 Limitations

One limitation of our study was the duration. Since the study duration was limited, the training time provided to teach advisors how to interpret the visualization was short. We deployed our model during the advising session, so we needed a model that could take in input and provide a prediction with minimal wait time, which may affected the predictive power of our system. We chose to minimize bias in the predictive model by removing demographic data from the training dataset, however, this does not account for human bias that advisors may have. Additionally, advisors who participated in our study were more open to the use of predictive models and incorporating new technology within their advising workflow, we acknowledge that this may not be the case for all advisors.

Our focus was on developing a tool that easily integrates into our university's current advising flow, and as such, we did not conduct a thorough study on the changes to the advising flow by comparing sessions with and without AdVizor. Previous research [19] revealed differences in performance between experienced advisors and junior (graduate student) advisors. Advisors who participated in our study all had at least 5 years of experience, by varying the experience level may reveal variations in the use and interpretation of AdVizor. We did not specifically gather participants who were not native English speakers. Cultural differences and social skills affect how students interact with advisors and their peers. This is a dimension of increasing inclusivity at institutions through advising, which we hope future researchers will contribute more to.

6 Conclusion

We built AdVizor, a combined machine learning and visualization system, to inform and empower advisors with knowledge from historical data and assist them during advising. Through a study with advisors and students, we explored usage and interaction patterns during advising sessions. The advisors appreciated the tool for providing targeted support on the more difficult parts of their job, like identifying the most crucial course to focus supplemental guidance and resources on, while allowing them to maintain agency. Advising sessions allow students to receive personalized support and guidance as they navigate their academic and professional goals. By working with advisors, students can feel more confident and prepared to make important decisions about their academic path. Future work on intelligent advising systems should consider the existing expertise of academic advisors and ensure that systems are designed to keep humans in the loop.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2021-04353].

REFERENCES

[1] L. Aulck, N. Velagapudi, J. Blumenstock, and J. West. Predicting student dropout in higher education. *Int. Conf. Machine Learning*, pp. 16–20, June 2016. 2

[2] R. S. Baker, L. Esbenshade, J. Vitale, and S. Karumbaiah. Using demographic data as predictor variables: a questionable choice. *J. Educational Data Mining*, 15(2):22–52, Jun. 2023. 2, 3, 8

[3] J. Bayer, H. Bydzovska, J. Geryk, T. Obsivac, and L. Popelinsky. Predicting drop-out from social behaviour of students. *Int. Conf. Ed. Data Mining*, June 2012. 2

[4] R. Bodily, J. Kay, V. Aleven, I. Jivet, D. Davis, F. Xhakaj, and K. Verbert. Open learner models and learning analytics dashboards: a systematic review. In *Proc. Int. Conf. Learning Analytics and Knowledge*, 10 pages, p. 41–50. ACM, New York, 2018. 2

[5] C. Brooks and J. Greer. Explaining predictive models to learning specialists using personas. In *Int. Conf. Learning Analytics and Knowledge*, pp. 26–30. ACM, 2014. 5

[6] S. Bull and J. Kay. Smilie: A framework for interfaces to learning data in open learner models, learning analytics and related fields. *Int. J. Artificial Intelligence in Education*, 26:293–331, 2016. 1, 2

[7] S. Charleer, A. V. Moere, J. Klerkx, K. Verbert, and T. D. Laet. Learning analytics dashboards to support adviser-student dialogue. *IEEE Trans. on Learning Technologies*, 11:389–399, 7 2018. 1, 2

[8] S. Chaturapruek, T. S. Dee, R. Johari, R. F. Kizilcec, and M. L. Stevens. How a data-driven course planning tool affects college students' gpa: evidence from two field experiments. In *Proc. ACM Conf. on Learning at Scale*, article no. 63, 10 pages. ACM, New York, 2018. 1, 2, 8

[9] C. F. de Oliveira, S. R. Sobral, M. J. Ferreira, and F. Moreira. How does learning analytics contribute to prevent students' dropout in higher education: a systematic literature review. *Big Data and Cognitive Computing*, 5(4):64, 2021. 2

[10] D. Delen. Predicting student attrition with data mining methods. *J. College Student Retention: Research, Theory & Practice*, 13(1):17–35, May 2011. 2, 3

[11] V. Dsilva, J. Schleiss, and S. Stober. Trustworthy academic risk prediction with explainable boosting machines. In *Int. Conf. Artificial Intelligence in Education*, pp. 463–475. Springer, 2023. 2

[12] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. Eventaction: Visual analytics for temporal event sequence recommendation. In *IEEE Conf. on Visual Analytics Science and Technology (VAST)*, pp. 61–70, Oct. 2016. 2

[13] A. Eklund. Beeswarm. R Software Package, 2015. 4

[14] E. Frank and M. Hall. Visualizing class probability estimators. In *Knowledge Discovery in Databases: PKDD 2003*, vol. 2838 of *LNCS*. Springer, 2003. 2

[15] F. Gutiérrez, K. Seipp, X. Ochoa, K. Chiluiza, T. D. Laet, and K. Verbert. Lada: A learning analytics dashboard for academic advising. *Computers in Human Behavior*, 107:105826, 6 2020. 1, 2

[16] M. Liz-Domínguez, M. Caeiro-Rodríguez, M. Llamas-Nistal, and F. A. Mikic-Fonte. Systematic literature review of predictive analysis tools in higher education. *Applied Sciences 2019, Vol. 9, Page 5569*, 9:5569, 12 2019. 2

[17] G. G. Mendez, L. G. rraga, and K. Chiluiza. Showing academic performance predictions during term planning: Efects on students' decisions, behaviors, and preferences. *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 5 2021. 2, 8

[18] L. Mu and K. Fosnacht. Effective advising: How academic advising influences student learning outcomes in different institutional contexts. *The Review of Higher Education*, 42(4):1283–1307, 2019. 1

[19] G. G. Méndez, L. Galárraga, K. Chiluiza, and P. Mendoza. Impressions and strategies of academic advisors when using a grade prediction tool during term planning. *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pp. 1–18, 4 2023. 2, 8

[20] H. Parkin, E. Heron, and M. Jacobi. What do students think? evaluating academic advising across an institution using the listening rooms method, 2022. 8

[21] K. Pittman. *Comparison of data mining techniques used to predict student retention*. PhD thesis, Nova Southeastern University, 2008. 2, 3

[22] T. Ruppert. Visual analytics to support evidence-based decision making, 2018. 2

[23] V. S. Sexton. Factors contributing to attrition in college populations: Twenty-five years of research. *J. General Psychology*, 72(2):301–326, Apr. 1965. 2

[24] N. Sghir, A. Adadi, and M. Lahmer. Recent advances in predictive learning analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, 28(7):8299–8333, 2023. 2

[25] T. Spinner, U. Schlegel, H. Schäfer, and M. El-Assady. explainer: A visual analytics framework for interactive and explainable machine learning. *IEEE Trans. on Visualization and Computer Graphics*, preprint, 2019. 2

[26] H. Strobelt, S. Gehrmann, H. Pfister, and A. M. Rush. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Trans. on Visualization and Computer Graphics*, 24(1):667–676, Jan. 2018. 2

[27] J. Superby, J. P. Vandamme, and N. Meskens. Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Int. Conf. Intelligent Tutoring Systems*, pp. 1–8, 2006. 2, 3

[28] M. Sweeney, H. Rangwala, J. Lester, and A. Johri. Next-term student performance prediction: A recommender systems approach. *J. Educational Data Mining*, 8(1):22–51, 2016. 2, 3

[29] V. Tinto. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1):89–125, 1975. 1, 2, 8

[30] J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viégas, and J. Wilson. The what-if tool: Interactive probing of machine learning models. *IEEE Trans. on Visualization and Computer Graphics*, 2019. 2

[31] L. Wilkinson. Dot plots. *The American Statistician*, 53(3):276–281, 1999. 4

[32] Y. Zhang, Q. Fei, M. Quddus, and C. Davis. An examination of the impact of early intervention on learning outcomes of at-risk students. *Research in Higher Education*, 26, 2014. 1