2021

Recommender systems in higher education: The effectiveness of meta-data analysis in predicting academic success

Michael J. Holmes
University of Northern Iowa

Recommended Citation
Holmes, Michael J., "Recommender systems in higher education: The effectiveness of meta-data analysis in predicting academic success" (2021). Dissertations and Theses @ UNI. 1195.
https://scholarworks.uni.edu/etd/1195

Offensive Materials Statement: Materials located in UNI ScholarWorks come from a broad range of sources and time periods. Some of these materials may contain offensive stereotypes, ideas, visuals, or language.
RECOMMENDER SYSTEMS IN HIGHER EDUCATION:
THE EFFECTIVENESS OF META-DATA ANALYSIS
IN PREDICTING ACADEMIC SUCCESS

An Abstract of a Dissertation

Submitted

in Partial Fulfillment

of the Requirements for the Degree

Doctor of Technology

Approved:

____________________________________
Dr. Jin Zhu
Committee Chair

____________________________________
Dr. V. Eugene Wallingford
Committee Co-Chair

____________________________________
Dr. Jennifer Waldron
Dean of the Graduate College

Michael J. Holmes
University of Northern Iowa

December 2021
ABSTRACT

A recommender system is: “a software technology that proactively suggests items of interest to users based on their objective behavior or their explicitly stated preferences” (Pu et al., 2011, p. 157).

Whether we realize it or not, much of our lives are influenced by recommender systems. These systems may recommend where we eat, what movie we watch, what music we listen to, what products we buy, or even what news and social media content we see.

Recommender systems have been effectively utilized in higher education to recommend courses as well as customizing content within an online course to meet specific student needs. The large volumes of data produced by e-learning systems, combined with other student and course data have been utilized with traditional statistical analysis as well other artificial intelligence techniques to predict student success.

Utilizing course participation data, and other data collected from the e-learning allows for universities to provide academic support for students, but often the at-risk students are not identified until after the course begins.

The purpose of this study was to develop and evaluate the effectiveness of content filtering and collaborative filtering recommender systems, in the grade prediction for university students. The information provided to the systems was limited to historical grade information.

These recommenders were trained with over 377,000 individual course grades, from four years of university courses. The recommenders were then used to predict
approximately 36,000 individual grades, with the predictions compared with the actual grades the students achieved.

The collaborative filtering recommender system successfully predicted within a half grade 61% of the time. The recommender also was able to correctly predict 11% of the D and F grades. The results of the content filtering recommender system demonstrated that it was no more effective than simply predicting the average grade of all students in the course. This recommender successfully predicted 46% of the grades within a half grade. It was ultimately only successful in predicting less than 1% of the D and F grades correctly.

The study demonstrated the potential for recommender systems to be utilized in grade prediction and early warning for students that may have trouble in a course, it opens several paths for future study.
RECOMMENDER SYSTEMS IN HIGHER EDUCATION:
THE EFFECTIVENESS OF META-DATA ANALYSIS
IN PREDICTING ACADEMIC SUCCESS

A Dissertation
Submitted
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Technology

Approved:

_______________________________
Dr. Jin Zhu, Committee Chair

_______________________________
Dr. V. Eugene Wallingford, Co-Chair

_______________________________
Dr. Shashi Kaparthi, Committee Member

_______________________________
Dr. Ali Kashef, Committee Member

_______________________________
Dr. J. Ben Schafer, Committee Member

Michael J. Holmes
University of Northern Iowa
December 2021
DEDICATION

Dedicated to my wife Misty, and children: Matthias, Mary, Marcus, and Matthew

for inspiration to always keep learning.
ACKNOWLEDGEMENTS

I would like to acknowledge all my past instructors at UNI throughout my bachelor and master coursework, as well as those who taught my doctoral courses. Without their knowledge and dedication to education, I could not have completed this achievement.

I owe a special debt of gratitude to the members of my graduate advisory committee, Dr. Kaparthi, Dr. Kashef, and Dr. Schafer, who were gracious in giving their time and assist me in accomplishing this effort. A special thanks to Dr. Zhu, from whom I was fortunate enough to learn from in class and receive her support in the final steps of this process. Finally, I would like to thank Dr. Wallingford for his academic and personal mentorship for so many years. He has provided consistent guidance from my first semester as an undergrad at UNI, which has allowed me to achieve great things.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................. vii

LIST OF FIGURES ............................................................................................................... viii

CHAPTER 1 INTRODUCTION ............................................................................................... 1
  Statement of Purpose ......................................................................................................... 2
  Statement of Need .............................................................................................................. 3
  Research Questions .......................................................................................................... 4
  Data Collection .................................................................................................................. 4
  Statement of Procedure ..................................................................................................... 5
  Limitations of Study .......................................................................................................... 6
    Delimitations of Study ...................................................................................................... 7

CHAPTER 2 REVIEW OF RELATED LITERATURE ............................................................... 8
  Recommender Systems ...................................................................................................... 9
  Types of Recommenders ................................................................................................. 11
    Content Filtering Recommenders ................................................................................ 11
    Knowledge Based Recommenders ................................................................................ 14
    Collaborative Filtering Recommenders .......................................................................... 16
    Hybrid Recommenders .................................................................................................. 19
  Evaluation of Recommender Systems ............................................................................. 20
  Novelty and Serendipity in Recommender Systems ....................................................... 21
  Diversity in Recommender Systems ................................................................................ 22
  Social Implications of Recommender Systems ............................................................... 23
# LIST OF TABLES

Table 1: Counts of grades used in the dataset................................................................. 38

Table 2: Grade Prediction Results by Content Filtering Recommenders ......................... 40

Table 3: D&F Grade Prediction Results by Content Filtering Recommenders .................... 40

Table 4: ANOVA analysis of Content Filter Recommenders ............................................. 41

Table 5: Turkey post hoc test on results of Content Filtering Recommender Systems ....... 41

Table 6: Grade Prediction Results by Collaborative Filtering Recommenders ................... 42

Table 7: D&F Grade Prediction Results by Collaborative Filtering Recommenders ............ 43

Table 8: ANOVA analysis of Collaborative Filter Recommenders .................................... 44

Table 9: Turkey post hoc test on results of Collaborative Filtering Recommender Systems .......................................................................................................................... 45

Table 10: Comparison of Grade Prediction Results by Recommenders ......................... 46

Table 11: Comparison of D&F Grade Prediction Results by Recommenders ................. 47

Table 12: ANOVA analysis of Recommenders ................................................................... 47

Table 13: Turkey post hoc test on results of Recommender Systems. ......................... 48
LIST OF FIGURES

Figure 1: General design of a content filtering recommender system. .......................... 12
Figure 2: Items clustered by similarities in a content-based recommender. ..................... 13
Figure 3: General design of a knowledge base recommender system. .......................... 14
Figure 4: General design of a collaborative filtering recommender system. .................. 17
Figure 5: Identifying k-nearest neighbors with movie rating data. .............................. 17
Figure 6: General design of a hybrid recommender. ..................................................... 20
Figure 7: Interface UML diagram of system .................................................................. 32
Figure 8: Class UML diagram of course recommender components ......................... 33
Figure 9: Object UML diagram for course recommender ............................................. 34
Figure 10: Class UML diagram of recommender system ............................................. 35
Figure 11: Groups statistically similar collaboration filter recommenders .................... 44
CHAPTER 1

INTRODUCTION

Whether we realize it or not, many of the decisions that we make each day are influenced by recommender systems. A recommender system is: “a software technology that proactively suggests items of interest to users based on their objective behavior or their explicitly stated preferences” (Pu et al., 2011, p. 157). These systems may recommend where we eat, what movie we watch, what music we listen to, what products we buy, or even what news and social media content we see. Sometimes we consciously choose to use the recommender and are aware of this influence but, often we are not.

In today’s world, companies generate and collect vast amounts of information each day. Retailers gather data on customer purchases and transactions. Online social networks collect personal information, our likes and dislikes and other items users post on their websites. Satellites, smart phones, global positioning devices (GPS) and other systems gather information on locations and movements of individuals. This explosion of digital information has developed into an industry of “big data.” Since the volume of data is overwhelming, organizations need to find a way to analyze this data effectively and identify the relevant data. One way this data can be utilized is by using a recommender system: “the aim of developing recommender systems is to reduce information overload by retrieving the most relevant information and services from a huge amount of data” (Lu et al., 2015).

Recommender systems developed from two major approaches. Content filtering applies similarities between items to help identify new items that may be of interest,
based on items the user has preferred in the past. In contrast, collaborative filtering utilizes preference data from all the users to identify groups of similar users. The system then recommends new items to a user, that other similar users preferred.

**Statement of Purpose**

Recommender systems have been utilized to personalize course content and assist identifying learning activities, they have not been widely applied to other areas in higher education for predicting in advance those individuals that may be at risk of failing courses.

Various tools have been utilized to help identify students at risk of failing courses. These approaches typically require information or feedback obtained during the course to identify students at risk of failing. Utilizing recommender system algorithms provides a way to use historic data to identify similar students and predict the student’s grades prior to the beginning of the course. This early warning would allow for early intervention to provide needed support.

The purpose of this study is to determine if recommender systems could be used as an effective means of providing decision support in two specific ways. The first is to provide a general prediction of a student’s academic success in each course. The second is to identify students that are potentially at risk of failing the course prior to the start of a term. An early warning would allow for better allocation of support services for those students.
Statement of Need

While there were many factors that led to the early development of recommender systems in specific domains, there has not been as much research of their application in higher education. The vast majority of work in educational data mining is related to e-learning course development and recommending learning resources that would be most applicable to the individual students (Aher & Lobo, 2013; Ghauth & Abdullah, 2011; Hiltz & Turoff, 2005; Romero & Ventura, 2007; Vialardi et al., 2009). E-learning systems provide a ready-made set of online data which facilitates the adoption of recommender systems.

These systems have been used for recommending online courses (Aher & Lobo, 2013), developing customized content for individual students in a course (Peiris & Gallupe, 2012), and recommending learning activities in an e-learning environment (Ghauth & Abdullah, 2011). This focus on e-learning courses allows for the utilization of readily available electronic data, and enables recommending customized electronic resources to construct a curriculum that would not be available with traditional classes.

While thus far, the focus has revolved around readily available e-learning data, there is a large amount of academic data on a student that is maintained by universities that has not yet been utilized. This information could be harnessed to help identify patterns among students and provide useful recommendations within the higher education arena.
Research Questions

1. Are there any significant differences between a content filter or collaboration filter recommender system in predicting a grade in a course for a student?

2. Are there any significant advantages between a content filter or collaboration filter recommender system in predicting students that may at risk of failing courses?

Data Collection

The data used in this study consisted of historical academic data collected over a period of 5 years from a medium sized Midwestern comprehensive university. This data set includes course enrollment data, and grades for all students enrolled at the university over that time period. This data will be utilized by the recommender systems to recommend courses and predict student grades in specific courses.

By using existing data a model can be built and tested against existing information to validate the results. This allows for an objective way to measure the success of the recommender by comparing the predicted action to actual data that has been collected. This is more practical that attempting to evaluate the recommendations in a live system, where requires collecting additional data to confirm the accuracy of the predictions.

Utilizing a common machine learning theory model, the data will be divided into two sets: the first will be utilized for training the system and building the recommender’s prediction model, while the second will be used for testing and validating the results (Mitchell, 1997; Goldman, 1999).
Multiple recommender systems will be created utilizing collaborative filtering, knowledge based, and hybrid approaches. The results from each of these systems will be gathered and compared and contrasted with the others.

Statement of Procedure

Recommender systems that provide recommendations for products and movies, such as Amazon or Netflix, can be evaluated by whether the user chooses to purchase a recommended product, or based on follow-up feedback provided by a customer rating. This feedback not only validates the accuracy of the recommendation, but it also provides additional information to improve future recommendations (Konstan & John, 2012).

For this study the results will be compared with the test data for validation. While predicting a student’s grade in the course, the predicted grade will be compared to the actual grade the student received in the test data. For example, if the system predicts that a student will earn a B in the course, but in reality, the grade was a C-, it will be evaluated as an error of 4 (the difference in the grade points). These individual prediction errors will then be used to calculate an overall mean error score for the recommender system.

To provide a baseline of comparison, the mean error score will be calculated for a system that simply predicts the average grade for each student in the class. For example, if historically the average grade in a particular course was a C, the system will predict a grade of a C for all the students taking the course. For a recommender system to demonstrate success, it should produce overall significantly better grade predictions than the baseline. To determine if there are any significant different in the overall mean error
rates, analysis of variance (ANOVA) tests will be run. These results of this analysis will demonstrate if a particular recommender system approach produces better predictions than the rest.

**Limitations of Study**

At an abstract level, all recommender systems share a common goal: identify which items in a collection are the most relevant. While these systems share some common obstacles like the cold start problem, each domain introduces its own unique constraints.

Within higher education there are several requirements that may affect the applicability of a recommendation beyond the general issues of expectedness and lack of diversity. For example, many courses have pre-requisite courses or requirements that a student must meet to enroll. There are also sequences of courses that may lead to a bias for certain courses. Program requirements will also create a pattern in the data. If correctly utilized this domain knowledge may be used to enhance the diversity and increase unexpectedness.

In addition, every university offers new courses as curriculums change and develop over time. For the recommender system to be effective it must have a way of introducing these new courses into the recommendations or it may potential, prevent these new courses from gaining popularity.

Another potential concern with student grade data is trends in grades over time. This could be the result of grade inflation over time, pedagogical changes within the
course, or other broad changes in course delivery implemented in response to a pandemic.

**Delimitations of Study**

One of the goals of this study was to ascertain the possibilities of utilizing recommender systems to identify at risk students prior to the beginning of the course. Previous research in identifying students in need of additional academic support utilized data collected during the course.

As research on recommender systems has progressed, additional approaches that introduce domain knowledge into the algorithms have been developed. Additional domain information about the courses and students were not included, and the focus of this study was to determine if the content filtering or collaborative filtering approaches would be effective with grade prediction.

This study was delimited in two ways:

1. This study was limited to a minimal data set including only student’s course history, and grades from courses taken.

2. The study was limited to utilizing the two basic methodologies of recommender systems: content filtering and collaborative filtering.
CHAPTER 2
REVIEW OF RELATED LITERATURE

Turing published one of the first papers that explored the possibility of machine learning. He discussed the possibility of a machine that could think and described a test for machine intelligence that still bears his name. In concluding his work, Turing (1950) stated, “The learning process may be regarded as a search for a form of behavior which will satisfy some criterion” (p. 460).

While Turing’s work began the discussion of artificial intelligence and machine learning, it took some time until a formal theory developed. Valiant (1984) introduced a mathematical construct to describe machine learning more formally. Building on his computational and mathematical proofs, Valiant remarks, “we have considered learning as the process of deducing a program for performing a task, from information that does not provide an explicit description of such a program” (Valiant, 1984).

This view of machine learning was later described as probably approximately correct (PAC) learning. For this type of learning the machine receives both positive and negative examples, and the system must construct a general formula for correctly classifying new cases within specified probability of accuracy (Goldman, 1999).

Over time other theorems of machine learning developed and the field of computational learning theory developed. The Association for Computational Learning (2018) defined computational learning theory as:

A research field devoted to studying the design and analysis of machine learning algorithms. In particular, such algorithms aim at making accurate predictions or
representations based on observations. The emphasis in COLT is on rigorous mathematical analysis using techniques from various connected fields such as probability, statistics, optimization, information theory and geometry. While theoretically rooted, learning theory puts a strong emphasis on efficient computation as well.

This research area is focused on applied machine learning and provides a framework for evaluating machine learning systems.

As Goldman (1999) stated, computational learning theory, “provides a formal framework in which to precisely formulate and address questions regarding the performance of different learning algorithms so that careful comparisons of both the predictive power and the computational efficiency of alternative learning algorithms can be made” (p. 30-2).

The goal of data mining in general and recommender systems in particular, is to learn to identify relevant data from a large data set, and develop a model using that data to predict future outcomes.

**Recommender Systems**

One area where machine learning has been most successfully applied (Jannach & Jugovac, 2019) are the recommender systems designed to utilize the vast amount of data available, to filter and recommend content that will be most applicable to the users.

The idea of gathering data from others, to develop recommendations existed long before computers were created. One can imagine a group of primitive humans coming across a new plant. It would lend to their survival if one of them ate the new plant in
question to determine if it is poisonous. In this example the recommender is the one who first tries out the potential food (Konstan, 2012).

This idea of recommending a product to others developed into the concept of critics. A critic’s job is to test out a restaurant, movie, service or item and then write a review for others who are making a decision about that item. This allows individuals to make a decision based on the knowledge of other’s recommendations (Konstan, 2012).

In addition to critics, indexing systems were developed in many areas to help identify similar items. Libraries, for example, grouped and stored books utilizing a system of classification and categorization. When finding a book on a shelf, other books on similar topics would be near it, as a way of recommending other potential books of interest. Indexes were created of periodicals based on subject areas so researchers could easily find recommendations of articles to read when conducting research.

With the dawn of the computer age, these indexes were improved upon by utilizing the power of computers to organize and index information to make retrieval and filter by topics easier (Konstan, 2012).

While indexing allowed for the grouping of similar items together, it was not able to utilize additional subjective information to further aid in the recommendations. With the birth of the internet, collaborative ranking became possible.

When news articles were first released on Usenet, readers were able to rate the articles they read. This allowed readers to not only find articles on the topic they were looking for, but also to get those with the higher overall ranking, to better recommend articles they may be interested in (Konstan, 2012).
Modern recommender systems utilize a variety of approaches to further enhance and personalize the recommendations it provides individuals. These systems make recommendations based on an individual’s past experience and feedback, in addition to others like him. These systems attempt to utilize the large volume of information available and provide specific recommendations for each individual (Tran, 2009).

**Types of Recommenders**

The goal of a recommender system is to provide a user with a list of items that they may be interested in, or to predict whether or not you will like a specified item (Tran, 2009). Users want to see information that is the most relevant or interesting to them. This role has developed with the introduction of the internet and social media websites. As Bostandjiev et al. (2012) stated:

> Recommendation systems play an increasingly important role in this domain as they serve to filter and refine a user’s information space according to the personal tastes and current requirements (p. 35).

To accomplish these goals a variety of techniques and algorithms have been developed to offer the best results. The design is based on the amount of information available, user feedback and the specifics characteristics of the domain.

**Content Filtering Recommenders**

One approach to recommender systems is content filtering. This approach uses a user’s own preference information about previous items to predict how well they will like new items (Figure 1). The basic concept is that individuals will like items that are similar to ones they have liked before (Veras et al., 2015; Deschenes, 2020).
A common example is recommending products to a user while they are shopping online. These websites track the items you are viewing and may provide a list of suggestions based on items that are similar to what you have been looking at. For example, if a customer is researching refrigerators at an online home improvement store, the site may suggest other related items, such as ice cube makers or water supply lines.

If the user has purchased items in the past, the recommender can utilize this purchasing history to further enhance the recommendations. If for example, a customer has previously purchased a number of Star Wars related toys, as well as a number of video games, the system may recommend a Star Wars video game.

If the customer adds to their profile data by providing the site with ratings on the items they have purchased, the recommender system can further enhance the recommendations. To do this in the system are clustered based on similarity on a number of dimensions. If a customer has a strong negative or positive opinion on the item
purchased, the system uses when making a prediction about how the individual may like a new item. By identifying the items that a user rates highly, a set of items that share some common features can be recommended for that individual (Figure 2) (Adamopoulos & Tuzhilin, 2015).

One advantage of the content filtering approach is that it allows for an explainable recommendation. Filtering similar items and utilizing tags with content filtering allows for a user understood explanation of the recommendations (Lops et al., 2019).

Figure 2: Items clustered by similarities in a content-based recommender.

There are some potential drawbacks to using content filtering. First, when making initial recommendations there can be a “cold start” problem. Due to the fact that the system has not yet gathered profile information about the user’s preference on items, it cannot evaluate how they may like other items (Tran, 2009; Bobadilla et al., 2013).
Another issue that can occur is in some cases products may have similar features, but they are not particularly useful recommendations. For example, a customer may purchase a large volume of whole milk. In this situation, recommending skim milk may not be helpful even though whole and skim milk share much in common (Konstan, 2012; Lee & Lee, 2015).

Preference information gathered from a user may not be applicable across domains. For example, information on the types of music a customer purchases may not be beneficial in recommending movies that they may wish to see (Knijnenburg et al., 2011).

Knowledge Based Recommenders

One attempt to overcome the challenges with content filtering was to introduce additional knowledge about the domain. These knowledge based recommenders utilize the profile and rating information that the content filtering systems use, and in addition, the knowledge model to aid in the recommendations (Figure 3; Bobadilla et al., 2013).

Figure 3: General design of a knowledge base recommender system.
The knowledge model may include contextual information that is gathered from the user regarding the specifications or features they are looking for in the particular item or service. This information is then used as another dimension to further differentiate the potential items (Champiri et al., 2015).

For example, if a customer is looking for replacement brake pads for their vehicle on at an auto parts store, they will receive better recommendations by providing the year, make and model of their vehicle. The recommender system can then use the knowledge model to correctly recommend only the parts that will fit that vehicle.

Knowledge based recommenders can improve effectiveness when compared content filtering recommenders, but they do not completely overcome the issues. The characteristics of the domain have an affect the impact of the domain knowledge model. A domain that is generally more organized, benefits from a more defined way to apply the knowledge. When recommending auto parts, the items are clearly applied to a very specific set of applications. When given the context of the search by the user, the vast inventory of items can be systematically reduced down to a few appropriate ones. In recommending movies, however, the data model may not be as effective based on the subjectiveness of the user’s requirements. Simply knowing the genre, director and actors of a particular movie, may not provide a strong level of certainty that an individual will like the film.

In this study, the system generated a list of recommended courses. Using a content filtering approach, it would be expected to recommend courses similar to those that they have previously taken. Additionally, the recommender utilized the grades
earned as a way to further refine the recommendation to suggest courses similar to those that the student was previously successful in.

**Collaborative Filtering Recommenders**

Collaborative filtering recommenders take a very different approach to recommending items. This approach attempts to utilize other individual’s profile data to create recommendations. The goal is to recommend items preferred by other similar users (Figure 4), utilizing explicit user feedback, or other implicit data gathered on user activity within the system or with the items (Sharma et al., 2019; Deschenes, 2020).

Collaborative filtering utilizes profile data about the individual to identify a set of others who are similar to and have similar preferences as the individual. The system then estimates a recommendation based on how these others evaluate items. As (Tran, 2009) stated it, “you may like it because your friends liked it.”

One challenge with collaborative filtering is identifying the group who may best predict your opinion. This algorithm is referred as finding the “k nearest-neighbors” (Konstan, 2012; Srifi et al., 2020). The goal is to find the most significantly similar individuals that match your preferences.
In this example (Figure 5), Joe's data may not be useful at all, because he rates all movies as 5s. When selecting neighbors for Tim, Matt and Pat would be the best options if we wanted to predict what Tim's opinion of the movie Frozen maybe. In addition, Susan would serve as a good negative "neighbor" because they have common ratings of movies that are opposite of each other.
Recommender systems that utilize collaboration filtering suffer from the same cold start issues that occur with content based approaches. Until there is enough profile data established for a user, it is difficult to identify others who have common opinions (Tran, 2009; Srifi et al., 2020).

Another issue is the possibility of limited data. For example, an online commerce site may have thousands of items in its catalog. If many of these items do not have ratings from many users, it may be difficult identifying users with enough common ratings to utilize collaboration successfully (Konstan, 2012).

In addition, popular items may receive a bias in the recommendations. If there are items that are highly popular, these items will begin to be recommended for everyone (Lee & Lee, 2015). For example, if everyone rates a summer movie blockbuster highly, it will begin to show up as a recommendation in everyone’s results. In the same way, a high volume product like bread may be recommended to all customers based on the fact that all users commonly purchase bread.

In this study, the recommender will utilize the course selections of other similar students as a way to recommend courses. In higher education there is a different dynamic at play than in a consumer based application. Students enjoy some freedom in their choice of classes, but they are also constrained by their chosen curriculum. There may be required courses to complete a major. In this way students take some courses due to requirements and not due to choice. This distinction could potentially skew the recommendations made.
In addition, the structure of curriculum with a shared set of general education courses that all students are required to take, could challenge traditional approaches of identifying a set of neighbors. Just as other domains have popular items, these required courses may be recommended to all students.

There are also social implications that must be considered. A collaboration filtering system attempts to recommend items based on others users that are similar. This approach could potentially perpetuate situations for underrepresented groups. For example, if there is a field of study that has been historically gender specific, those courses would not be highly recommended, which would lead to a continuation of the pattern.

Hybrid Recommenders

Hybrid recommender systems combine the content-based, knowledge base and collaborative filtering together (Figure 6). The hybrid systems can be implemented in various ways. They may be combined in a single algorithm or used separately with the results combined (Konstan, 2012; Burke, 2002).
Figure 6: General design of a hybrid recommender.

These hybrid recommenders rely on multiple techniques and data approaches to help provide a more balanced result set. In this way all information, user preferences, other user’s data, item features, and domain knowledge are combined to try provide a more complete picture of the data. The goal is to try and capitalize on the strengths of each approach and overcome inherit weaknesses (Champiri et al., 2015).

Evaluation of Recommender Systems

While the goal of recommender systems is evident, it has been more difficult to develop an effective way to evaluate them. In real world applications, the evaluation is based on how useful the recommendations are. To get a complete picture requires feedback from the users after they used the recommendations. In order for a recommender to be successful the users must have a level of trust in the recommendations.
In research projects, the typical approach to evaluation systems is to split the data into a set of training data and another set of testing data. The recommender is then built utilizing the training data, and an analysis can be done on how effect it is at predicting the testing data.

**Novelty and Serendipity in Recommender Systems**

One of the challenges evaluating recommender systems goes beyond just the accuracy of the recommendations. In most applications, simply provided a list of recommended items that the user expects, while accurate, may not be useful (Lee & Lee, 2015). If for example, a movie recommender suggested a number of movies the user would have already watched otherwise, the user does not receive any benefit.

Novelty is the concept of recommending items that are new to the user, but the novelty of the item recommended to the user, does not mean that it is entirely unexpected (Adamopoulos & Tuzhilin, 2015). When recommending movies, it would not be unexpected to recommend a new sequel to a user who was a fan of the original film. In this scenario, the user would expect this recommendation.

So an accurate recommendation may be novel, but not considered a useful suggestion by the user. The user is looking for serendipity. A serendipitous recommendation is both novel and unexpected (Adamopoulos & Tuzhilin, 2015). When a user receives a recommendation for something that they would not find another way, and it is something they like, they are more likely to trust future recommendations. As de Gemmis et al. (2015) describes:
Let us consider a recommender system that simply suggests movies directed by the user’s favorite director. If the system recommends a movie the user was not aware of, the movie would be novel, but not serendipitous. On the other hand, a movie by a young, not very popular director is more likely to be serendipitous (and also novel) (p. 697)

For this reason, novelty is not a goal on its own.

**Diversity in Recommender Systems**

Another aspect that is important when evaluating a recommender system is the diversity of the results. There are different approaches to introducing diversity to the recommendations that are generated. A simple technique is to utilize some level of randomness into the recommendations (de Gemmis et al., 2015). One approach is to remove some of the expected items from the list and include other items that may be more unexpected (Adamopoulos & Tuzhilin, 2015).

A more robust method is to utilize a knowledge infusion (KI) process (de Gemmis et al., 2015). The approach draws on concepts from machine learning to build a general knowledge repository to utilize to enhance recommendations. This data model is not directly related to the domain knowledge that is part of a knowledge based recommender, but instead tries to utilize relationships external to the domain.

Suppose a customer has previously purchased a biography of Winston Churchill, the board game Risk. Using the knowledge infusion process, the recommender may suggest the World War II board game Axis and Allies. By utilizing the general
knowledge linking Winston Churchill to World War II and Risk to a genre of a war related board game, it is able to recommend a game that intersects with those concepts.

When evaluating recommender systems, accuracy alone is not always the best measure. As the accuracy of recommenders has improved, and the data they rely on are increasing, one concern is that user’s options are being limited. Social media system’s algorithms are so precise, that users are only presented with information that is consistent with their current preferences (de Gemmis et al., 2015). With this perspective the concepts of unexpectedness, serendipity and diversity provide another important measurement of the recommender system.

Social Implications of Recommender Systems

As the usage of recommender systems has increased, so has the concerns regarding the social implications of their use. Recommender systems require vast amounts of data, and as with any data collection there are questions related to privacy and data security. In addition to these concerns, are how the resulting predictions and recommendations can impact society as a whole.

Privacy Concerns

Several years ago, Target made national news with its attempts to send personalized coupons to customers. In addition to sending coupons related to previous purchases, Target developed a system to identify pregnant shoppers, and specifically market to them.

Target developed a system that was able to predict if a customer was pregnant based on their purchase history. The analysis revealed a set of products, that when
purchased together could predict that a customer was pregnant. Target then used this information to send offers for products that a pregnant woman might need (Duhigg, 2012).

This program led to privacy concerns when a teenage girl was identified by the system to receive the pregnancy mailer. The girl’s father was initially upset and contacted Target to complain, but eventually had to accept the fact that Target discovered his daughter’s pregnancy before he did (Duhigg, 2012).

With the amount of data that is generated on a person as part of modern day life, the question of privacy is more complex. Not only is there a question on how and what data is collected, but also how that data is used and analyzed to predict future behavior.

**Concerns with Confirmation Bias**

Eli Pariser described another potential area of concern that he called, “The Filter Bubble”. This occurs as websites increase their ability to affectively personalize the content we are presented, and filter out any content that is inconsistent with our preferences. Pariser (2011, p. 88) stated

The filter bubble tends to dramatically amplify confirmation bias—in a way, it’s designed to. Consuming information that conforms to our ideas of the world is easy and pleasurable; consuming information that challenges us to think in new ways or question our assumptions is frustrating and difficult. This is why partisans of one political stripe tend not to consume the media of another. As a result, an information environment built on click signals will favor content that supports our existing notions about the world over content that challenges them.
As recommender systems increase their effectiveness in predicting content they users prefer, they are left with content that confirms their previous bias, to the exclusion of any contrary recommendations. The social implications of this are obvious when considering more and more individuals are getting news and information via their social networking sites.

A study to test the filter bubble theory was completed using a movie recommender system. The researcher’s data did demonstrate a filter bubble effect. Over time, “the items recommended by the system and the items rated by users both became slightly narrower (less diverse)” (Nguyen et al., 2014).

The use of recommenders controlling online content and news articles has also raised concerns. For example, YouTube’s creates a play list designed to provide content a user is likely to watch, based on the internal recommendation system, this can lead to a user only being presented with videos that support a specific perspective (Alfano et al., 2020). In order to prevent a filter bubble, or confirmation bias in certain domains additional safeguards should be included in the recommender system, to introduce some variability.

**Adoption of Recommender Systems**

To be successful, recommender systems require a significant amount of data. For this reason, much of the early research in recommender systems were in domains where data could be easily collected, such as movies, publications, or other retail products (Konstan, 2012; Park et al., 2012).
When these systems began to show promise, many large companies began to explore their use as a marketing tool. Recommender systems could be utilized with the customer purchase history that the companies already had. These recommendations could be used to market new products to existing customers or expand their customer base. The financial incentives to improve recommendations for selling products led to more research with recommender systems.

Recommender systems have become ubiquitous in today’s culture. Nearly all online retailers utilize systems to increase sales and revenue, and the news and social media use systems to filter and provide personalized content to the user.

While much of the initial scholarly research in recommender systems focused on movie recommendations, there is much research in the domains of music, television, books, documents, eCommerce, social media, and health care (Bobadilla et al., 2013; Lu et al., 2015; DeCroon et al., 2021). This research has demonstrated the diversity and flexibility of recommender systems. In addition to finding articles to read, recommender systems have been developed to help researchers identify relevant journals and organizations to submit their research papers to for publication (Dehdarirad et al., 2020).

With the demonstrated promise of recommender systems their usage expanded beyond academia and were quickly adopted by online retailers. Netflix capitalized early with their use of recommender systems (Konstan, 2012), and soon other retailers like Amazon and Barnes and Noble developed recommenders for their products on their web sites.
One additional source of data for online retail sites are to utilize the textual comments that customers leave in their reviews. This information can be parsed and employed to provide additional context for use by the collaborative filtering recommenders to recommend products (Srifi et al., 2020).

From the very beginning, streaming services have relied on recommender systems, to identify content that will interest a particular user, and present to the user to keep them engaged with the platform. YouTube accounts that 60% of the clicks are from provided recommendations, and Netflix attributes 75% of the shows users watch is from a recommendation (Jannach & Jugovac, 2019).

The health care industry began researching the effectiveness of recommender systems in earnest in the 2000s. Recommender systems have been designed to aid doctors and nurses in designing treatment (Duan et al., 2011), as well for patients to recommend lifestyle and nutritional recommendations (DeCroon et al., 2021), or providing personalized motivation plans (Hors-Fraile et al., 2019).

Research in Education

As with other fields, educational systems have attempted to capitalize on the use of new technology. With the introduction of computer technology into education, more data was collected and stored electronically which allowed researchers attempt to harness the power of this information.

Romero and Ventura (2007) examined how this new field of educational data mining was developing. The first forays into this arena attempted to apply common data mining techniques that had been utilized in knowledge discovery in large databases. This
was a precursor to recommender systems, and was implemented using simple knowledge
based reasoners and statistical analysis to discover patterns within the data.

**Recommenders in Course Selection**

Recommender systems can be utilized to identify courses that best meet a
student’s needs and aptitudes. As higher education has adopted online tools to provide
students self-service enrollment, recommender systems have been utilized to help
students select courses (Lynn & Emanuel, 2021; Warnes & Smirnov, 2020).
Recommender systems have also been demonstrated to be effective in recommending
seminars for faculty looking for additional training (Paytaren, 2020).

Developing a course plan for college students can be a complex undertaking. One
area of research in recommender systems is to provide decision support in developing a
long term course plan. A system of this type can, “generate a set of optimal or near
optimal alternate plans that are of similar quality and yet structurally different”
(Mohamed, 2015).

**Recommenders to Support Teachers/Learners**

One area explored within education is to utilize recommenders to help self-
learners identify a variety of educational resources. Deschenes (2020) found that these
recommenders suggest, “books, learning content, publications (forum or blog posts,
articles, etc.), learning material, learning objects, papers and videos” (p. 10). Systems
have also been developed to help students identify peers that may be helpful in tutoring
or assisting them (Khalid et al., 2020). The goal of these recommenders was to filter
through a large set of potential information and identify the specific content that would best aid the specific learner.

**Recommenders in e-Learning and MOOCs**

With the development of electronic learning systems and Massive Open Online Classes (MOOCs), a large amount of electronic data is now captured from students. Because of this large volume of course available online, recommenders were developed to help students identify courses that may be of interest to them (Bousbahi & Chorfi, 2015; Aher & Lobo, 2013; Khalid et al., 2020).

With the large number of students enrolled in a MOOC, this leads to a large amount of data, navigating the course content is not as straightforward as a traditional online course, and recommender systems have been utilized to recommend discussion threads most pertinent to the student, and the content delivered to the student in the course can be customized by the recommender system, based on meta data collected in the e-Learning system (Khalid et al., 2020).

**Data Analysis to Identify At-Risk Students**

One of the challenges within higher education today, is identifying students at risk of failing the course. A goal of research in this area, it to identify at risk students early to provide additional support earlier in the course.

One such study developed a learning analytics algorithm for a single course that utilized student data collected from the learning management system to identify at risk students and provide interventions with additional tutoring modules in the course. This analysis required constructing the online course content in a way to allow for the
collection student interactions with course content and instructor and tutoring modules when the system identified students that were at risk (Simanca et al., 2019).

In another study, several statistical modeling methods were used to attempt to predict what students are at risk for failing at week 5 of the course. The variables selected included midterm exam grade as well as other learning objective scores. The various methods yielded overall accuracy rates between 86% and 94%, with accuracy rates of predicting students that would fail the course between 38% and 86% (Marbouti et al., 2016).
CHAPTER 3

DEVELOPMENT OF COURSE RECOMMENDER SYSTEM

Multiple recommender systems were developed for comparison of different techniques. In addition to utilizing both content and collaborative filtering, multiple thresholds were tested for each approach for comparison. A statistical analysis was conducted to compare the results to determine if one recommender was significantly better at grade predictions than others.

System Architecture

An object-oriented design was developed, with the implementation coded in Java. The grade data was provided in flat text files, and the system loaded these files and constructed recommenders utilizing both content filtering and collaborative filtering approaches. The design provided flexibility to adjust the tolerances and within the implemented algorithms.

System Interface Design

Regardless of domain, all recommender systems function in a similar manner: they recommend items from a catalog by predicting an individual’s assessment of the items they have not yet encountered. A system interface was designed at an abstract level to facilitate flexibility in the implementation (Figure 7). A common interface for the recommender system was defined, as well as the components of a catalog to maintain the collection of items, and individuals that made up the population of users.

This abstract interface supports recommender systems of all types, regardless of whether the recommendations are for movies, music, shopping, or university courses.
The catalog supports adding and removing items, and each item maintains a list of other items in the catalog that are similar to it, as well as a list of individuals that have interacted with that item.

Likewise, the population supports adding and removing individuals, and each individual maintains a set of ratings for various items, and maintains its set of nearest neighbors, which can be implemented with unique algorithms.

Figure 7: Interface UML diagram of system
System Implementation

Classes specific to a course recommender were then created, implementing the interfaces in the design (Figure 8).

Figure 8: Class UML diagram of course recommender components
In addition to the classes that directly implemented the interfaces in the design, DataSet and DataSetDAO classes were created to specifically handle the data processing of the grade files (Figure 9). The data access object encapsulates all the code required to read the files and parse the data into the objects needed for the recommender. The data set object creates a composite that contains both the population of students as well as the catalog of courses. This data set is then consumed by the recommender system objects as the training data required for making its predictions.

![Object UML diagram for course recommender](image-url)
Three different classes were created to implement the Recommender System interface (Figure 10). Each of these applied different techniques for making the recommendation and included flexibility to adjust the parameters as necessary.
**Average Grade Predictor**

To serve as a basis of comparison, grades for the test population were predicted using the average grade off all the students who took the course in the training data. This methodology would predict the same grade for any student for the course.

**Content Filtering Recommender System**

A content filtering recommender was constructed using the course history data, and for each course, a set of similar courses was identified based on the frequency that students took both courses. The similarity of two courses was calculated by identifying the set of students that took both courses and dividing that by the set of students that took either course. If this percentage was above the defined threshold, then the two courses would be included as similar.

Recommender systems were created using various thresholds to optimize the similar course lists. A lower threshold would create larger course similarity lists that were not as effective for predicting grades. A higher threshold would create limited course similarity lists, but too high of a threshold would effectively result with each course alone in its similarity and lead to the same predictions as the average grade prediction.

**Collaborative Filtering Recommender System**

Collaborative filtering recommender systems were constructed using the same course history data. For each individual student, a set of nearest neighbors were calculated. The nearness of two individuals was measured by calculating the average difference in grades for any courses that the two users have in common. For example,
two individuals with grades of (12, 10, 7) and (11, 12, 5) would have a difference of 1.667.

Recommender systems were created using various thresholds for identifying the neighborhood. A larger threshold identified more students within the neighborhood, and sometimes had a negative effect on the accuracy of the prediction. With a threshold that was too low, the neighbor size would be too small, and not provide enough similar students to predict a grade.
CHAPTER 4
RECOMMENDER SYSTEMS RESULT ANALYSIS

Each recommender was provided with a set of training data consisting over 377,000 individual grades for courses over multiple terms (Table 1). This data was then utilized to construct a recommender system to predict a grade for a student for each course in the test data, which consisted of over 36,000 grades from a subsequent term (Table 1). The predicted grades were then compared with the actual grades in the test data set and an analysis was made to determine how effective each recommender was at accurately predicting grades, as well as identifying students at risk of not passing the course.

The multiple recommender systems were then compared with others of the same type to determine if varying the thresholds of the filtering algorithms produced significantly different results. Finally, an analysis of the different types of recommenders was completed to determine if one approach provided significantly better results.

<table>
<thead>
<tr>
<th>Grades Used in Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total grades for training</td>
</tr>
<tr>
<td>Total grades for testing</td>
</tr>
<tr>
<td>Number of passing grades in test set</td>
</tr>
<tr>
<td>Number of D&amp;F grades in test set</td>
</tr>
</tbody>
</table>

Table 1: Counts of grades used in the dataset

When calculating the accuracy of the recommender systems, the predicted grade was compared with the actual grade the student received in the course. Additionally, the
accuracy was tested by expanding the tolerance by a half grade, a full grade, and a grade and a half. For example, with a tolerance of a full grade, an actual grade of a B would be counted as a correct if the predicted grade was between a C and an A.

Accuracy rates for passing and D&F grades were also calculated for each recommender system, utilizing the following formulas:

$$\text{Accuracy D&F} = \frac{\text{Number of correctly predicted D&F grades}}{\text{Number of D&F grades in test set}}$$

$$\text{Accuracy Passing} = \frac{\text{Number of correctly predicted passing grades}}{\text{Number of passing grades in test set}}$$

### Analysis of Content Filtering Recommender Systems

An initial analysis of the percentage of grades correctly predicted by each content filtering recommender yielded similar results (Table 2). Each recommender correctly predicted the exact grade approximately 15.9% of the time. The recommenders consistently demonstrated an accuracy of 69.1% when predicting within a full grade. When the accuracy tolerance was expanded to a range of a grade and a half, the systems were consistently correct 82.4% of the time.
Predicting At-Risk Grades

The content filtering recommenders proved ineffective at correctly predicting students that may be at risk of failing the course. Overall, the recommenders correctly predicted a passing grade for over 99% of the 34,831 passing grades in courses. However, the content filtering recommender systems were only able to correctly predict three D and F grades out of 1,547 actual D and F grades (Table 3) for well under a 1% accuracy.

Inherent to these types of recommender systems is predicting a grade for a student, based on the grades they achieved in similar courses. For the system to predict a D or F for a course, the student would have had to fail all off the similar courses identified.

<table>
<thead>
<tr>
<th>Course Similarity Threshold</th>
<th>80%</th>
<th>85%</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>15.92%</td>
<td>15.91%</td>
<td>15.93%</td>
<td>15.93%</td>
</tr>
<tr>
<td>Accuracy within half a grade</td>
<td>46.51%</td>
<td>46.49%</td>
<td>46.51%</td>
<td>46.51%</td>
</tr>
<tr>
<td>Accuracy within full grade</td>
<td>69.10%</td>
<td>69.09%</td>
<td>69.10%</td>
<td>69.10%</td>
</tr>
<tr>
<td>Accuracy within grade and a half</td>
<td>82.74%</td>
<td>82.46%</td>
<td>82.47%</td>
<td>82.48%</td>
</tr>
</tbody>
</table>

Table 2: Grade Prediction Results by Content Filtering Recommenders

<table>
<thead>
<tr>
<th>Course Similarity Threshold</th>
<th>80%</th>
<th>85%</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct D&amp;F grades</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>False D&amp;F grades</td>
<td>24</td>
<td>23</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>Correct passing Grades</td>
<td>34,806</td>
<td>34,807</td>
<td>34,807</td>
<td>34,812</td>
</tr>
<tr>
<td>False passing Grades</td>
<td>1,543</td>
<td>1,543</td>
<td>1,543</td>
<td>1,543</td>
</tr>
<tr>
<td>Accuracy – D&amp;F</td>
<td>0.19%</td>
<td>0.19%</td>
<td>0.19%</td>
<td>0.19%</td>
</tr>
<tr>
<td>Accuracy - Passing</td>
<td>99.93%</td>
<td>99.93%</td>
<td>99.93%</td>
<td>99.95%</td>
</tr>
</tbody>
</table>

Table 3: D&F Grade Prediction Results by Content Filtering Recommenders
Statistical Analysis

There was not a statistically significant difference between the content filtering recommender systems as demonstrated by one-way ANOVA ($p = 1.000$) (Table 4). A Tukey post hoc test (Table 5) showed that altering the threshold for determining the similarity of courses, produced no significant different in grade predictions between the four content filtering recommender systems.

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>.016</td>
<td>3</td>
<td>.005</td>
<td>.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>807512.198</td>
<td>152100</td>
<td>5.309</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>807512.214</td>
<td>152103</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: ANOVA analysis of Content Filter Recommenders

<table>
<thead>
<tr>
<th>Content Filter Recommender Similarity Threshold</th>
<th>Content Filter Recommender Similarity Threshold</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>85%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>.001</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td>85%</td>
<td>80%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>.001</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td>90%</td>
<td>80%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>85%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td>95%</td>
<td>80%</td>
<td>-.001</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>85%</td>
<td>-.001</td>
<td>.017</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>.000</td>
<td>.017</td>
<td>1.000</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.

Table 5: Turkey post hoc test on results of Content Filtering Recommender Systems.
Analysis of Collaborative Filtering Recommender Systems

Several collaborative filtering recommender systems were constructed with various thresholds to determine the nearest neighbors for each student. The largest neighborhood size with a similarity threshold of 2.0 preformed best with a wider tolerance range of a grade and a half, correctly predicting 87.71% of the time, while the recommender with a threshold of 1.0 preformed best at predicting the exact grade, correctly predicting 30.73% of the time (Table 6).

<table>
<thead>
<tr>
<th>Nearest Neighbors Threshold</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>1.0</th>
<th>1.1</th>
<th>1.2</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>29.24%</td>
<td>29.78%</td>
<td>29.91%</td>
<td>30.73%</td>
<td>29.84%</td>
<td>29.86%</td>
<td>28.30%</td>
</tr>
<tr>
<td>Accuracy within half a grade</td>
<td>59.77%</td>
<td>60.60%</td>
<td>60.73%</td>
<td>62.35%</td>
<td>62.10%</td>
<td>62.27%</td>
<td>61.69%</td>
</tr>
<tr>
<td>Accuracy within full grade</td>
<td>77.00%</td>
<td>77.54%</td>
<td>77.63%</td>
<td>79.07%</td>
<td>79.28%</td>
<td>79.43%</td>
<td>79.78%</td>
</tr>
<tr>
<td>Accuracy within grade and a half</td>
<td>86.13%</td>
<td>86.31%</td>
<td>86.35%</td>
<td>86.99%</td>
<td>87.20%</td>
<td>87.29%</td>
<td>87.71%</td>
</tr>
</tbody>
</table>

Table 6: Grade Prediction Results by Collaborative Filtering Recommenders

Predicting At-Risk Grades

The collaborative filtering recommenders fared better at correctly predicting at risk students, than the content filtering recommender. The collaborative filtering recommender systems were able to correctly predict between 7.66% and 11.58% of the actual 1,547 D and F grades (Table 7). These types of filters are designed to identify students with similar grades in the same courses, to predict the grade. For the system to
predict a D or F for a course, the student would have had to a neighborhood of students that did not do well in the course.

When looking at how often the recommender system incorrectly predicted a failing grade when the student passed the course, the collaborative filtering systems had a higher count of false positives than the content filtering recommender systems.

<table>
<thead>
<tr>
<th>Nearest Neighbors Threshold</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>1.0</th>
<th>1.1</th>
<th>1.2</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct D&amp;F grades</td>
<td>116</td>
<td>120</td>
<td>151</td>
<td>173</td>
<td>142</td>
<td>144</td>
<td>179</td>
</tr>
<tr>
<td>False D&amp;F grades</td>
<td>268</td>
<td>298</td>
<td>298</td>
<td>464</td>
<td>407</td>
<td>409</td>
<td>426</td>
</tr>
<tr>
<td>Correct passing grades</td>
<td>34,478</td>
<td>34,452</td>
<td>34,452</td>
<td>34,366</td>
<td>34,327</td>
<td>34,332</td>
<td>34,404</td>
</tr>
<tr>
<td>False passing grades</td>
<td>1,399</td>
<td>1,395</td>
<td>1,395</td>
<td>1,373</td>
<td>1,374</td>
<td>1,372</td>
<td>1,367</td>
</tr>
<tr>
<td>Accuracy – D&amp;F</td>
<td>7.66%</td>
<td>7.92%</td>
<td>9.77%</td>
<td>11.19%</td>
<td>9.37%</td>
<td>9.50%</td>
<td>11.58%</td>
</tr>
<tr>
<td>Accuracy – Passing</td>
<td>99.23%</td>
<td>99.14%</td>
<td>98.67%</td>
<td>98.83%</td>
<td>98.82%</td>
<td>98.78%</td>
<td>99.97%</td>
</tr>
</tbody>
</table>

Table 7: D&F Grade Prediction Results by Collaborative Filtering Recommenders

**Statistical Analysis**

There was a statistically significant difference between the collaborative filtering recommender systems as demonstrated by one-way ANOVA (p = 0.000) (Table 8). A Tukey post hoc test (Table 9) showed that adjusting the threshold for determining the nearest neighbors, generated overlapping clusters of recommender systems, that provided statistically similar results.
Table 8: ANOVA analysis of Collaborative Filter Recommenders

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>198.860</td>
<td>6</td>
<td>33.143</td>
<td>5.623</td>
<td>0.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>1569018.260</td>
<td>266175</td>
<td>5.895</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1569217.120</td>
<td>266181</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The recommender with a nearest neighbor threshold of .5 showed no statistically significant difference with any of the other collaborative filtering recommenders. The recommender with a 3 threshold was only statistically similar to the recommenders with .4 and .5 thresholds (Figure 11).

Figure 11: Groups statistically similar collaboration filter recommenders
<table>
<thead>
<tr>
<th>Nearest Neighbor Threshold</th>
<th>Nearest Neighbor Threshold</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>.3</td>
<td>.4</td>
<td>.020</td>
<td>.018</td>
<td>.908</td>
</tr>
<tr>
<td>.5</td>
<td>.024</td>
<td>.018</td>
<td>.809</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.069*</td>
<td>.018</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>.063*</td>
<td>.018</td>
<td>.007</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>.068*</td>
<td>.018</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.073*</td>
<td>.018</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>.4</td>
<td>.3</td>
<td>-.020</td>
<td>.018</td>
<td>.908</td>
</tr>
<tr>
<td>.5</td>
<td>.004</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.049</td>
<td>.018</td>
<td>.085</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>.042</td>
<td>.018</td>
<td>.199</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>.048</td>
<td>.018</td>
<td>.098</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.052*</td>
<td>.018</td>
<td>.049</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.3</td>
<td>-.024</td>
<td>.018</td>
<td>.809</td>
</tr>
<tr>
<td>.4</td>
<td>-.004</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.045</td>
<td>.018</td>
<td>.148</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>.038</td>
<td>.018</td>
<td>.309</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>.044</td>
<td>.018</td>
<td>.167</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.048</td>
<td>.018</td>
<td>.090</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.3</td>
<td>-.069*</td>
<td>.018</td>
<td>.002</td>
</tr>
<tr>
<td>.4</td>
<td>-.049</td>
<td>.018</td>
<td>.085</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>-.045</td>
<td>.018</td>
<td>.148</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>-.006</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>-.001</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.004</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>.3</td>
<td>-.063*</td>
<td>.018</td>
<td>.007</td>
</tr>
<tr>
<td>.4</td>
<td>-.042</td>
<td>.018</td>
<td>.199</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>-.038</td>
<td>.018</td>
<td>.309</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.006</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>.005</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.010</td>
<td>.018</td>
<td>.998</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>.3</td>
<td>-.068*</td>
<td>.018</td>
<td>.002</td>
</tr>
<tr>
<td>.4</td>
<td>-.048</td>
<td>.018</td>
<td>.098</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>-.044</td>
<td>.018</td>
<td>.167</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.001</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>-.005</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.004</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.3</td>
<td>-.073*</td>
<td>.018</td>
<td>.001</td>
</tr>
<tr>
<td>.4</td>
<td>-.052*</td>
<td>.018</td>
<td>.049</td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>-.048</td>
<td>.018</td>
<td>.090</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>-.004</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>-.010</td>
<td>.018</td>
<td>.998</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>-.004</td>
<td>.018</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.

Table 9: Turkey post hoc test on results of Collaborative Filtering Recommender Systems.
Comparison of Recommender Systems

A final analysis was done, comparing the results of the content filter and collaborative filter recommender systems, as well as the average grade recommender for a baseline comparison. The collaborative filtering recommender system, consistently performed better at correctly predicting grades at every tolerance level (Table 10). The collaborative filtering recommender successfully predicted within a half grade over 61% of the time, and within a full grade, nearly 80% of the time. The content filtering and average grade recommenders did not perform nearly as well, predicting within a half grade at 46% and within a full grade 69% of the time.

<table>
<thead>
<tr>
<th>Recommender Type</th>
<th>Collaborative Filter</th>
<th>Content Filter</th>
<th>Average Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>28.30%</td>
<td>15.92%</td>
<td>15.94%</td>
</tr>
<tr>
<td>Accuracy within half a grade</td>
<td>61.69%</td>
<td>46.51%</td>
<td>46.53%</td>
</tr>
<tr>
<td>Accuracy within full grade</td>
<td>79.78%</td>
<td>69.10%</td>
<td>69.11%</td>
</tr>
<tr>
<td>Accuracy within grade and half</td>
<td>87.71%</td>
<td>82.74%</td>
<td>82.50%</td>
</tr>
</tbody>
</table>

Table 10: Comparison of Grade Prediction Results by Recommenders

Predicting At-Risk Grades

The collaborative filtering recommender system correctly predicted grades of D and F at just over 11% (Table 11). In contrast, the content filtering recommender system was only able to correctly predict 3 D and F grades. The average grade recommender was unable to correctly predict any.

While the collaborative filter was more successful at identifying the students at risk of failing, it also incorrectly predicted 426 D and F grades, where the student
achieved a passing grade (Table 11). This was less than .01% of the overall 36,378 grades predicted.

<table>
<thead>
<tr>
<th>Recommender Type</th>
<th>Collaborative Filter</th>
<th>Content Filter</th>
<th>Average Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct D&amp;F Grades</td>
<td>179</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>False D&amp;F Grades</td>
<td>426</td>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td>Correct Passing Grades</td>
<td>34,404</td>
<td>34,806</td>
<td>34,821</td>
</tr>
<tr>
<td>False Passing Grades</td>
<td>1,367</td>
<td>1,543</td>
<td>1,546</td>
</tr>
<tr>
<td>Accuracy – D&amp;F</td>
<td>11.58%</td>
<td>0.19%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Accuracy -- Passing</td>
<td>98.78%</td>
<td>99.93%</td>
<td>99.97%</td>
</tr>
</tbody>
</table>

Table 11: Comparison of D&F Grade Prediction Results by Recommenders

**Statistical Analysis**

There was a statistically significant difference between the resulting predictions of the collaborative filtering, content filtering, and average grade recommender systems as demonstrated by one-way ANOVA (p = 0.000) (Table 12). A Tukey post hoc test (Table 13) showed that the collaborative filtering recommender provided statistically different results, than the content filtering and average grade recommenders, where there was not a statistically significant different between the content filtering and average grade recommenders.

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>4972.396</td>
<td>2</td>
<td>2486.198</td>
<td>449.862</td>
</tr>
<tr>
<td>Within Groups</td>
<td>630444.712</td>
<td>114075</td>
<td>5.527</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>635417.108</td>
<td>114077</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12: ANOVA analysis of Recommenders
<table>
<thead>
<tr>
<th>Recommender</th>
<th>Recommender</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Filter</td>
<td>Content Filter</td>
<td>-.443*</td>
<td>.017</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Average Grade</td>
<td>-.442*</td>
<td>.017</td>
<td>.000</td>
</tr>
<tr>
<td>Content Filter</td>
<td>Collaborative Filter</td>
<td>.443*</td>
<td>.017</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Average Grade</td>
<td>.001</td>
<td>.017</td>
<td>.998</td>
</tr>
<tr>
<td>Average Grade</td>
<td>Collaborative Filter</td>
<td>.442*</td>
<td>.017</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Content Filter</td>
<td>-.001</td>
<td>.017</td>
<td>.998</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.

Table 13: Turkey post hoc test on results of Recommender Systems.

Summary of Analysis

This research demonstrated that a collaborative filtering recommender system can yield a significantly better result predicting grades than simply predicting the average grade in the class. The collaborative filter recommender system was able to correctly predict the exact grade 28% of the time, and 80% of the time within a tolerance of a full grade. In addition, it was able to accurately predict the student would pass nearly 99% of the time. It was less successful in predicting D and F grades, only correctly predicting these 11.5% of the time.

One of the constraints of this study, was limiting the data provided to the system to only data available prior to the beginning of the course. This restricted using any additional data about the student’s interactions within the course, or assignment or midterm grades.

There may be some additional information, available prior to the course that could be included to enhance the recommender’s predictions. For example, additional meta data such as the student’s major, year in school, and standardized exam scores could be used to further enhance the nearest neighbor algorithm.
Similarly, there may be additional information about the courses themselves that could be utilized. For example, the instructor’s grading proclivities may be different, and may need to be factored into comparing student’s grades for the same course. Likewise, course content may evolve over time, which could alter grading patterns from semester to semester. Other aspects of the course, such as the delivery mode (online/in person) or teaching pedagogy (flipped classroom, lecture, lab, etc.) could be important in identifying similar courses.

Abstractly, recommender systems are preforming the same process in many domains, but as they are applied to each new area, there may be additional discipline specific knowledge that can be applied to enhance the systems.
CHAPTER 5

SUMMARY AND CONCLUSIONS

This study was designed to determine if various recommender system techniques could be utilized to accurately predict the final grades of students, based solely on their previous grade history. To complete this analysis, multiple recommender systems, utilizing content filtering and collaborative filtering techniques were constructed, as well as a recommender that would simply predict the average course grade for all students.

These recommenders were trained with over 377,000 individual course grades, from four years of university courses. The recommenders were then used to predict approximately 38,000 individual grades, with the predictions compared with the actual grades the students achieved. The accuracy was measured with several tolerances, from predicting the grade exactly, to predicting it within a full grade and a half. The accuracy of D and F grades were reviewed with special attention, with the goal of determining how effective the systems were identifying students that may be at academic risk.

A statistical analysis was conducted to identify if any of the recommender systems produced better results than a baseline of predicting the average course grade. The results of the analysis demonstrated that a content filtering approach, where similar courses are identified, and a grade is predicted based on the student’s grades in those similar courses, was not significantly better than the average grade prediction. The content filter recommender was unable to successfully identify those students who resulted in receiving a grade of D or F.
The analysis showed more success for the collaborative filtering recommender system, which did produce statistically better results than the average grade prediction. In addition, it was able to correctly predict 11% of the students that would receive a D or F in the course.

**Conclusions**

The study demonstrated the potential for recommender systems to be utilized in the use of predicting student outcomes in a course and to serve as an early warning for students at risk of failing, prior to the term. This would allow for the identification of students that may need additional academic support, prior to the course beginning.

Unlike more traditional methods of identifying students based on assignments, exams and course participation in an eLearning or classroom setting, the recommender system was constructed with only the student’s previous grade history in courses. With this limited information the collaborative filtering recommender system successfully predicted within a half grade 61% of the time. The recommender also was able to correctly predict 11% of the D and F grades.

The collaborative filtering approach identifies students with similar grades in courses and groups them into a neighborhood of similar students. The grade prediction for a new course for the student, can then be based on how other students in the neighborhood performed in the course.

The results of the content filtering recommender system demonstrated that it was no more effective than simply predicting the average grade of all students in the course.
This recommender successfully predicted 46% of the grades within a half grade. It was ultimately unsuccessful in predicting D and F grades correctly.

The content filtering approach identifies courses that are similar to other courses, and then bases the grade prediction for a new course, based on how the student previously performed in similar courses. This approach may have been less successful, due to the situation where a sequence of courses that progress on each other. These courses would be identified as similar, but the student may not achieve the same grades in the more advanced courses.

This also explains why the content filtering recommender performed poorly at predicting D and F grades, as only a student that consistently performed poorly in a group of similar courses would be predicted to fail a course.

**Future Research**

While this study demonstrated the potential for recommender systems to be utilized in grade prediction and early warning for students that may have trouble in a course, it opens several paths for future study.

Would the addition of other course meta-data enable the recommender system to be more effective? There are other data points that are known prior to the beginning of the course that could be used, such as the instructor teaching the course and whether the course in person or online. Additional domain knowledge could be introduced by grouping courses by subject to add additional context for the identification of similar courses.
Would additional information about the students benefit the recommender system? Utilizing information such as the student’s major or degree program, standardized test scores, or other demographic data may benefit the identification of similar students.

Would preprocessing the grade data benefit the recommender systems? Not all courses use the same grading scheme. For example, F+ grades, may not be valid, and could potentially skew the grade variations. Alternate approaches to convert and compare grades, may provide better success at predicting at risk students.

In addition, the grade history data is skewed towards passing grades. In the data set used, only 5% of the grades were D or Fs. Future research could be completed utilizing machine learning techniques to help address the skewed data.

Would altering the algorithms for the collaborative filtering produce better results? The similarity calculations could be adjusted to weight those students with more common courses higher, than those that may only have a single course in common. This would prevent students that have only a single course in common from being identified as closer than others that have a wider breadth of similar grades in multiple courses.

Would other techniques of constructing the recommender system be beneficial? A hybrid approach could be used to introduce domain knowledge to the system. This knowledge of the course catalog, and course sequences could prove beneficial in identifying similar courses.
REFERENCES


