PREDICTING AN ELECTIVE COURSES IN STUDENT-CENTRIC EDUCATIONAL SYSTEM

Prof. Rahul Wantmure, Assistant Professor
NCRD's Sterling Institute of Management Studies
Navi Mumbai
rahul_wan2003@yahoo.co.in

Dr. Murlidhar Dhanawade, Professor and HoD, (MCA)
NCRD's Sterling Institute of Management Studies
Navi Mumbai
dr.murlidhar.dhanawade@gmail.com

Dr. Sandeep Ponde, Associate, Professor
NCRD's Sterling Institute of Management Studies
Navi Mumbai
emailponde@gmail.com

ABSTRACT
The role of electives in a student-centered educational system is significant. This enables the children to adhere to their interests and pick up important knowledge and expertise. Also, electives offer students the chance to learn new skills and expand their knowledge of different fields of study. As a result, it's critical to comprehend how to accurately estimate the students' electives. In a student-centered educational system, this research provides a high utility mining (HUM) strategy to predicting student electives. The suggested method makes use of HUM to track out commonly selected electives as well as connections between electives and other variables. The method's outcomes are utilised to build a customizable prediction model.
Keywords: Datamining, Frequency mining technique, high-utility itemset mining, predictions, electives, pruning, internal and external parameters

Introduction

Early-semester academic performance prediction of students is a highly valuable tool for early intervention to get better their performance and also to lower the failure rates of students at the conclusion of the semester. However, predicting students' college performance is a difficult task since a variety of characteristics, including academic background a student's prior academic accomplishments demographic characteristics, economic background, behavioural characteristics, and other elements can affect a student's performance. As a result, Educational Data Mining (EDM) is a key technique for resolving this issue. One of the most popular uses of EDM is the prediction of students' future performance using their academic data in the past years. It is a crucial tool that may be used to improve student performance, lower failure rates, and give a full picture of how pupils are learning.

By using data mining techniques, it is possible to analyze data from different sources and predict which electives will be the most beneficial for a student. This paper explores how data mining can be used to predict electives for students. Data mining is an important tool for predicting electives for students. By analyzing data from different sources, it is possible to identify patterns and correlations that can be used to make predictions about which electives will be the most beneficial for a student. For example, data mining can be used to analyze the data from past student performance and identify correlations between different electives and student performance. This data can then be used to make predictions about which electives a student should take.

High Utility Itemset Mining for Predicting Electives

In the context of Higher Education, High Utility Itemset Mining (HUIM) is a technique for mining data that can be used to precisely predict electives in the context of higher education. HUIM is based on the concept of discovering the most significant items or events that occur in a dataset. It is particularly useful for uncovering patterns in datasets that contain a large number of variables and data points. HUIM can be used to identify relevant features in the dataset that are important in predicting electives. For example, HUIM can be used to identify the most influential factors such as student preferences, academic performance, and program requirements that are associated with a student's decision to enrol in a particular elective. Additionally, HUIM can be used to generate new knowledge by uncovering interesting patterns and relationships in the data. HUIM can also be used to identify potential areas of improvement in the elective selection process, as well as to identify areas where additional resources might be needed.

Literature Review

Viger (2015) has spoken about a problem in data mining termed as High-Utility Itemset Mining. The author has given an outline of this hindrance and also makes clear how it is exciting, and has provided the algorithms for this problem and datasets. Kakaraddi & Bojewar (2017), the authors of this study relate HUM's current algorithms. In order to ensure predictive analysis on the education domain and predict student success based on their choice of elective topics, it designs and assesses an algorithm for HUM. This projection is compared to their historical data, or the performance of prior semesters.

Anita, Deshpande & Dhabu (2018), A unique selective database projection-based HUI mining algorithm (SPHUI-Miner) and an effective data format, called HUI-RTPL, which is an ideal and packed in depiction of data requiring little space, are both proposed by the researchers in this study. In order to reduce the search space for HUI mining, they also suggest two brand-new data structures, namely the Tail-Count list and the selected database projection utility list. Our suggested method is more effective thanks to selective projections of the database that shorten database scanning time. It produces distinct data instances and fresh predictions for data with fewer dimensions, which speeds up HUI mining.

Nabil, Seyam & Ahmed (2021), In order to identify students who are at risk for failing, this research investigates the effectiveness of deep learning in the field of education data mining. To address the issue of an asymmetrical dataset, they considered various resampling techniques. Mengash (2020), this research concentrates on strategies to help colleges make admissions decisions by predicting applicants' educational outcomes at universities using data mining tools. The findings show that based on specific pre-admission factors, applicants' early university achievement may be anticipated before admission.

Krishnamoorthy (2014), the researcher offers a high utility mining approach that makes use of unique pruning algorithms in an effort to advance the state-of-the-art. Experimentation is used to show how useful the proposed approach is. Also, a comparison of the method to a cutting-edge method is offered. The experimental findings show how well the suggested strategy works to eliminate unqualified candidates. Tewari, Panwar (2018), the scholars have presented an assessment among various association rule mining algorithms that deals with high utility patterns mining.

Tseng, Shie, Wu & Yu (2013),provides two algorithms for mining high utility itemsets, UP-Growth and UP-Growth+, along with a number of efficient ways for candidate itemset pruning. Utility Pattern Tree (UP-Tree), a tree-based data structure, is used to maintain the information of high utility itemsets so that candidate itemsets can be efficiently generated with just two database scans. Using various real and synthetic datasets, the performance of UP-Growth and UP-Growth+ is matched with that of cutting-edge algorithms.

Liu, Wang and Fung (2016), suggests a novel approach that, without producing candidates, identifies high utility patterns in a single phase. The innovations are a look-ahead mechanism, a linear data structure, and a high utility pattern growth approach. The reverse set enumeration tree is searched during the pattern growth phase, and utility upper bounding is used to condense the search space. Macarini, Cechinel, Machado, Ramos & Munoz (2019), this study tries to identify at-risk pupils early on in introductory programming classes. In order to determine which datasets (collection of variables) and classification algorithms work best together, they present a comparison analysis. Findings exhibit that there are no statistically major differences between models created from the various datasets, and that interaction counts and derived attributes are sufficient for the work.

Objectives Of The Study

1. Precisely predict any elective to students of Higher Education Institute (HEI). This is to enable students to make informed decisions about their academic choices.
2. To help students to identify the best suited electives for them and make accurate decisions about their educational goals. This can help students to maximise their chances of success in their studies.
3. By using High utility mining, students are more likely to select the most suitable electives for their degree programs, which can lead to better academic performance and increase their job prospects.
4. Based on these results, suggestions to the students for the elective can be done in their forthcoming semesters. This can help to improve substantially in their performance.

High Utility Itemset Mining

High utility itemset mining is a technique for accurately predicting electives for students in higher education based on the student's academic history, interests, goals, intelligence, emerging industry trends, and other peripheral parameters. Utility is the overall satisfaction or value derived from using a certain good or service. Understanding utility values is crucial to understanding why different products have varying prices and levels of demand. Higher utility products are frequently in greater demand, which enables them to command higher prices.

A transaction database is a database that holds a collection of consumer transactions. A transaction is a combination of the products that customers have purchased. In the database below, for instance, the first client purchased "apples, bread, cheese, gherkin, and egg," whereas the second purchased "apples, bread, and egg."

Transaction Items	
T1	{apples, bread, cheese, pickle, egg}
T2	{apples, bread, egg}
T3	{cheese, pickle, egg}
T4	{apples, bread, pickle, egg}

Table 1: A transaction database

Itemset	Support
{egg}	4
{pickle, egg}	3
{bread, pickle, egg}	2
{apples}	3
…..	

Table 2: A frequent itemsets

Finding frequent itemsets is the aim of a prior mining method called frequent itemset mining. These methods require a transaction database and the minimum support threshold parameter "minsup" as input. Once all sets of things (itemsets) that appear in minsup transactions have been returned, these algorithms stop.

Consider the itemset “bread, pickle, egg." For instance, if we set minsup = 2, we would find multiple such itemsets (referred to as frequent itemsets), such as the following. Because it appears in three transactions, it is said to have a support of 3, and because the support of "bread, pickle, egg" is greater than minsup, it is considered to be common.

Significant Constraints of Frequent Itemset Mining are:

1. The absence of consideration of buying numbers is a key constraint. Hence, an item may only show up once or not at all in a transaction. Hence, if a consumer buys five, ten, or twenty loaves of bread, they are all treated equally.
2. The fact that all items are given the same weight, relevance, and utility is a second major constraint. For instance, it doesn't matter if a buyer buys a costly bottle of saffron or a loaf of bread—both are seen as equally significant. Thus, frequent pattern mining may find many frequent patterns that are not interesting.

High-Utility Itemset Mining

The issue of frequent itemset mining has been renamed as the issue of high-utility itemset mining to solve these drawbacks. In this issue, a transaction database holds transactions where both the unit profit of each item and the purchase amounts are considered. Take the following transaction database as an example.

Trans	Items
T0	{apples (1), bread(5), cheese(1), pickles(3), egg(1) }
T1	{ bread(4), cheese(3), pickles(3), egg(1) }
T2	{ apples (1), cheese(1), pickles(1) }
T3	{ apples (2), cheese(6), egg(2) }
T4	{ bread(2), cheese(2), egg(1) }

Item	Unit Profit
Apples	Rs. 5
Bread	Rs. 2
Cheese	Rs. 1
Pickle	Rs. 2
Egg	Rs. 3

Table 3: A transaction table with quantities and unit profit information for items
(Source: https://data-mining.philippe-fournier-viger.com/introduction-high-utility-itemset-mining/)

Think about transaction T3. The relevant consumer has purchased two units (kg) of apples, six units (packs), and two units (dozens) of eggs, according to this information. The unit profits of each of these products are now shown in the table on the right. For instance, the unit profits for the goods "apples," "bread," "cheese," "pickles," and "egg" are, respectively, Rs. 5, Rs. 2, Re. 1, Rs. 2, and Rs. 3, and they are all shown as Rs. This indicates, for instance, that every sold "apple" unit makes a profit of Rs.5.

Finding the itemsets (groups of items) in a database that sell well together is the challenge of high-utility itemset mining. The user must enter a value for the "minutil" threshold (the minimum utility threshold). All high-utility itemsets, or those that produce at least "minutil" profit, are output by a high-utility itemset mining algorithm. Consider, for instance, that the user set "minutil" to Rs.25. The following would be the output of a high utility itemset mining algorithm.

High Utility Itemsets	Profits (Rs.)
{ apple, cheese }	28
{ apple, bread, cheese, pickle, egg }	25
{ bread, cheese, pickle }	34
{ bread, cheese, egg }	37
{ bread, pickle, egg }	36
{ cheese, egg }	27
{ apple, cheese, egg }	31
{ bread, cheese }	28
{ bread, cheese, pickle, egg}	40
{ bread, pickle }	30
{ bread, egg }	31

Table 4: high-utility itemsets

Consider the combination of bread and pickles as an example. It is regarded as a high-utility itemset since its utility of 40 (which results in a profit of Rs. 40/-) is greater than the minutil barrier, which the user has set at Rs. 25/-.
Let's now examine the methodology used to determine an itemset's utility (profit) in more depth. The number of each item from the itemset multiplied by its unit profit, in general, determines how useful an itemset is in a transaction. Consider the figure below as an illustration. In transaction T0, the profit for "apple, egg" is 1 x 5 + 1 x 3 = 8 Rupees. Similar to this, the profit from the transaction T3 item "apple, egg" is 2 x 5 + 2 x 3 = 16 Rupees. A set's utility throughout the entire database is now calculated as the sum of all transactions in which it appears. As "apple, egg" only appears in transactions T0 and T3, its usefulness is the total of Rs. 24 (Rs. 8 + Rs. 16).

Trans	Items
T0	{ apples (1), bread(5), cheese(1), pickles(3), egg(1) }
T1	{ bread(4), cheese(3), pickles(3), egg(1) }
T2	{ apples (1), cheese(1), pickles(1) }
T3	{ apples (2), cheese(6), egg(2) }
T4	{ bread(2), cheese(2), egg(1) }

Item	Unit Profit
Apples	Rs. 5
Bread	Rs. 2
Cheese	Rs. 1
Pickle	Rs. 2
Egg	Rs. 3

Table 5: illustration of how to calculate the utility of itemset {apple,egg}

The issue of high-utility itemset mining is highly intriguing for the following reasons.

1. Rather than looking for groups of goods that are regularly purchased, it would be more useful to find those that produce significant profits from client transactions.
2. From a research standpoint, the high-utility itemset mining issue is more difficult. A well-known property of frequency (support) of itemsets in frequent itemset mining states that for every given itemset, all of its supersets must have a frequency (support) that is lower or equal. This is known as the "Apriori property" or "anti-monotonicity" trait, and it is particularly good at pruning the search area since we know that all other itemsets are rare.

Problem Statement

While candidates / students are very naïve in selecting any elective, he/she selects it at random with no knowledge of how it will help the candidate in future. If the electives are wrongly chosen it may hamper their educational performance in future. On the contrary, if the candidate is guided to select the right elective, it will be beneficial in many aspects. The factors that can impact his / her selection of elective could be:

1) The intelligence of the individual
2) Domain knowledge
3) Theory and Practical marks obtained by the candidate in the curriculum
4) Emerging trends in the industry

These factors can be used to evaluate a candidate and a certain pattern can be created from his/her performance on these factors. This pattern can be matched with the previous year's students who had taken a particular elective and scored well or not. Thus, the candidate can be guided to opt in or opt out a particular elective.

Proposed Work

In single phase association rule mining, a unique approach has been put forth that makes use of factors such as statistical threshold-based pruning to find high utility patterns. Here, pruning is utilised to cut down on the amount of memory and time needed to mine high utility itemsets. Setting a threshold value—which is frequently calculated after numerous runs or experiments with the algorithm—enables the discovery of high utility patterns from a dataset. A pattern will be boring if the utility of an itemset is less than a minimum threshold utility. These are the measures we can take to find high utility itemsets:

The algorithm's first inputs are the Transaction dataset, External Utility value, and Minimum Threshold value. Following database projection for the itemset, identical transactions are then combined. Each database projection simply requires linear time. The utility value is subsequently subjected to two upper bound techniques. After computing the upper bound, high utility itemsets are finally mined.

Algorithm On Data Mining For Predicting Electives For Students

1. **Gather and clean data:** The first step is to collect the information of the student whose elective is to be predicted. The parameters needed are as follows:
    a. The marks of theory courses taken by the candidate in his all earlier semesters
    b. The marks of practical courses taken by the candidate in his all earlier semesters
    c. The aptitude of the candidate where there will be scrutiny on various factors around him w.r.t. the curriculum he / she is learning
    d. The market trends w.r.t. the job opportunity and new emerging technologies
    e. The historical data of previous year students' marks (theory, practical, aptitude) who had studied the same curriculum.

There are some direct parameters and some indirect parameters that will affect the study.

2. **Prune the data**: This involves removing the unwanted data like, the data of students who had cancelled the admission or the data of those students who have year drop, from the historical data, Here, the pruning of data becomes very much essential as this will help to narrow-down on the results. In order to discover the best results, sampling with replacement can be used in conjunction with pruning to limit the number of candidate item sets. The data needs to be organized and formatted in a way that is easy to use for the data mining algorithm.

3. **Apply the data mining algorithm:** Once the features have been identified, the high utility mining algorithm can be applied. This involves using the features to develop a model that can be used to predict which electives a student is likely to take. The likely interesting and uninteresting patterns will be generated.

4. **Test the model:** The model should then be tested to see how accurate it is in predicting which electives a student is likely to take. This can be done by comparing the predictions to actual outcomes.
   To compare the observed and expected findings, a statistical testing method called the chi-square test can be very effectively applied. The goal of this test is to establish if a discrepancy between observed and anticipated data is the result of random variation or a correlation between the variables being examined.
   Once the model has been tested, it can be refined if necessary to improve its accuracy. This can involve adjusting the model parameters or adding more features. It is to be noted that refining can also include pruning the data below the threshold. A pattern will be boring if the utility of an itemset is less than a minimum threshold utility.

5. **Implement the model:** Once the model is refined, it can be implemented in the school to make predictions about which electives a student is likely to take. This can be done by inputting the student data into the model and using it to generate predictions.

Future Scope

The paper has proposed a very significant technique where a student can be guided to select an elective that can be helpful for his/her future in education as well as in the job perspectives also. The prediction will precisely help to choose a rightful elective rather than just select an elective depending upon some irrelevant parameters.
In the future a proper model can be designed and presented with software. This model can be tested on various programs in the field of technology like Masters of Computer Applications, Bachelor of Engineering, etc.

Conclusion

High utility mining is a data mining technique that can be used to precisely predict electives for students. By using the data, HEIs can suggest electives that are tailored to each student's interests, helping them to make the most of their studies. Additionally, this technique can be used to identify anomalies in student behaviour, which can help HEI to identify potential problems and intervene as needed.

References

Anita, Deshpande P. &Dhabu M.(2018), "Selective Database Projections Based Approach for Mining High-Utility Itemsets", Visvesvaraya National Institute of Technology, Nagpur, DOI 10.1109/ACCESS.2017.2788083

Kakaraddi S. &Bojewar S. (2017), "Application of High Utility Mining for Pattern Prediction", International Conference on Trends in Electronics and Informatics, ICEI 2017.

Krishnamoorthy S. (2014), "Pruning strategies for mining high utility itemsets", https://doi.org/10.1016/j.eswa.2014.11.001.

Liu J., Wang K. & Fung B. (2016), "Mining High Utility Patterns in One Phase without Generating Candidates", IEEE Transactions on knowledge and data engineering, volume 28, May 2016, DOI: 10.1109/TKDE.2015.2510012.

Macarini L., Cechinel C., Machado M., Ramos V. & Munoz R.(2019), "Predicting Students Success in Blended Learning-Evaluating Different Interactions Inside Learning Management Systems," Applied Sciences, vol. 9, no. 24, p. 5523, 2019.

Mengash H. (2020) , "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems", DOI: 10.1109/ACCESS.2020.2981905

Nabil A., Mohammed S. & Ahmed A. (2021). "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks", DOI 10.1109/ACCESS.2021.3119596, IEEE Access

Tewari V. &Panwar A. (2018), "An Efficient Algorithm for High Utility Pattern Mining from Transactional Databases", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, 2018 pp. 12326-12330.

Tseng V., Shie B., Wu C. & Yu P. (2013), "Efficient algorithms for mining high utility itemsets from transactional databases," IEEE Transactions on Knowledge and Data Engineering, Volume 25, NO.8 pp.1772-1786, Aug.2013, DOI: 10.1109/TKDE.2012.59

Viger P.(2015), "An Introduction to High-Utility Itemset Mining", The PAKDD 2015 Conference.