24-osios Lietuvos jaunųjų mokslininkų konferencijos „Mokslas – Lietuvos ateitis" teminė konferencija
Proceedings of the 24th Conference for Junior Researchers "Science – Future of Lithuania"

EKONOMIKA IR VADYBA / ECONOMICS AND MANAGEMENT

2021 m. spalio 21 d., Vilnius, Lietuva
21 October 2021, Vilnius Lithuania

eISSN 2029-7149
eISBN 978-609-476-279-6
Article Number: vvf.2021.030

Šiuolaikinio verslo aktualijos / Actualities of Modern Business

http://jmk.vvf.vilniustech.lt

RISING COMPANY'S EFFICIENCY THROUGH USING CUSTOMER'S DATA ANALYTIC TOOLS: A CASE STUDY OF USING CHURN ANALYSIS TO PREDICT STUDENT'S DROPOUT

Aleksei IURASOV[1], Ogonna Joy ANGUS[2*]

[1]Department of Business Technology and Entrepreneurship, Business Management, Vilnius Gediminas Technical University, Saulėtekio al., LT-10223 Vilnius, Lithuania
[2]Department of Business Management, Business Management, Vilnius Gediminas Technical University, Saulėtekio al., LT-10223 Vilnius, Lithuania
[*]E-mail: ogonna-joy.angus@stud.vilniustech.lt

Abstract. To increase the efficiency of higher education institutions (HEI), the student centered education plan is introduced. The management of higher education institutions is inextricably linked to the rise and fall of the school's overall growth. How the leadership can rely on the functioning of the evaluation mechanism to enable the school to construct strong programmes and activities to adapt to the competitive circumstances of the education market, these must be thoroughly thought-out strategies for problem-solving. The increase of e-learning resources, instrumental educational software, the use of the Internet in education, and the construction of student information databases has resulted in massive reservoirs of educational data. A well-done churn prediction model can help the higher education institutions track student's academic progress, enrolment and drop-out in the most effective way in which the best result can be achieved. This purpose of this paper is to analyse the commonly used method of decision tree to predict and reduce the likelihood of students dropping out from higher education institutions (HEI). Businesses spend countless amounts in Informational Technologies (IT) deployment and update in the world of technology. The research methodology used will be based on different journals, articles, and reports to investigate the effectiveness of customer churn analysis using a qualitative approach.

Keywords: customer data analysis, customer relationship management, churn prediction, educational data mining, higher education institution, technologies, information technologies.

Introduction

The article by (Tinto & Cullen, 1973) explained the need to differentiate between the variety of meanings offered to the word dropout:

– Individuals who leave the University in which they are enrolled

This is classified as a drop-out by any person leaving his or her institution of enrolment, which is specifically tailored to the needs and policies of higher education institutions. From their point of view, the inability of people to complete the degree program for which they are enrolled reflects an inappropriate use of scarce institutional resources. In addition, each drop-out reflects a loss to the institution not only of a position that could have been taken over by another person capable of completing the program of instruction but also of a wide variety of intellectual capital expended in its development as a student. As established, Dropout was, therefore, a requirement both for admission officers, administrative planners, guidance, and counselling workers, and for social scientists and those concerned with student morality, institutional engagement, and the estimation, interpretation, and prevention of student turnover in higher education institutions.

Smith and Naylor (2001) argued the fact that the non-completion rate for students has been the subject of much more study in the United States than in the United Kingdom may partly reflect the fact that the non-completion rate in the United States is much higher (at around 37%) than in the United Kingdom (currently about 18 percent in the expanded higher education sector). It usually takes three years in England to earn a bachelor's degree and unlike the United States, it is not common for transfers between institutions (fewer than 3 percent switch institutions). This is partially because there is no accepted college credit program, so it is difficult to pass. A minimum academic level of achievement is subject to the advancement of students from one year to the next. This implies that students may drop out because they do not achieve the appropriate academic level (the degree to which they must pass progress exams varies by school, but all have some sort of evaluation) or because they expect failure, family history is likely to impact not only the financial capacity of students to finish their studies, but also their academic preparedness and determination are linked to this and their career ambitions after college. There are limits (that differ by the institution) to the number of times a student can take a class again, however. Students who repeatedly fail are not permitted to proceed.

In Nigeria according to (Kainuwa & Yusuf, 2013) the socio-economic support of parents, cultural tradition, and practice, as well as the parent's religious beliefs, are some of the parental factors that impact the system. Most rural residents are farmers who have very little socio-economic support to the extent that they are always struggling to survive and talk less about their daughters' education. They have traditionally attached less importance to female children's education, so any attempt to contribute to their development is rendered useless. The female net enrolment ratio (NER) is as high as 70% in some states in the South, while some are as low as 10% in the North. The dropout percentage in rural schools was as high as 35.39%. In rural schools, female dropouts were higher than males, 42.10% versus 28.67%.

Research Questions: Generally, predicting university dropouts and transfers can be used to assist with marketing and other measures to hold or reduce dropout statistics. However, this thesis seeks to answer the important questions related to the following:

- What is churn prediction?
- How churn prediction can be used to construct a student drop-out predictive model?

Research Objective: There are several different approaches for modelling churn predictions which are the best way to predict the accuracy. The objective of this study is to compare the use of various methods and to assess the accuracy of the two most popular methods.

The task to achieve the objectives:

- Proposing and studying in practice various mathematical approaches to prediction.
- Comparing the approaches with their benefits and drawbacks.
- Churn prediction model development using two separate models to assess methods' accuracy and to recommend possible options to help reduce the churning statistics by examining the use of internationalization programmes to satisfy students and to have a significant impact on student satisfaction.

1. Literature review

Before we begin, we believe that it is essential to discuss some of the research foundations that have been developed. Previous research will help us understand what has been scrutinized in the past and how these findings (or paucity thereof) can be valuable to advance the progression of literature. The future investigation of these research issues would at best offer accelerated contributions.

1.1. Customer Relationship Management (CRM)

Customer relationship management (CRM) is regarded as a broad business approach that employs technical knowledge to manage long-term relationships between the industry and its customers (customer behaviour and retention) (Gopalsamy & Gokulapadmanaban, 2021). According to (Farooqi & Raza, 2011) CRM is a software-based approach to managing customer relationships. Most CRM software vendors emphasize the importance of a comprehensive approach to CRM strategy. CRM initiatives often fail when implementation is restricted to software installation without offering enough incentives for workers to understand, contribute, and fully use the information system. The customer relationship is a corporate strategy that seeks to identify, predict, and manage the needs of an organization's existing and future customers. It is neither a term nor a project.

CRM encompasses a wide range of elements that are inextricably linked:

- Front-office processes: face-to-face contact with customers, email, online services, phone calls, and so on.

– Back-office processes: Operations that have an indirect impact on front-office activities, such as billing, maintenance, marketing, planning, financing, manufacturing, and advertising.
– Interaction with other businesses and partners, such as suppliers/vendors and retail stores, distributors, and business networks. This external network is used to support both front and back-office operations.
– Analysis: Key CRM data is analysed to plan marketing campaigns, develop business strategies, and assess the performance of CRM operations, such as market share, customer number and form, sales, profitability, and so on.

Customer Relationship Management (CRM) techniques have been developed and used to enhance customer satisfaction and retention, increase profitability, and support critical analytical activities such as predictive analytics and classification. CRM applications usually store a massive amount of information about each customer. This information is obtained through customer activity at the company, data entered by the customer during the registration process, call to help hotlines, and so on. Proper data analysis will yield impressive results not only for marketing purposes but also for finding customers who are likely to cancel their contract (Lazarov & Capota, 2007).

CRM is a business policy that has existed from the beginning of the conception of company. It is a strategy and approach used to develop a relationship with consumers, which assists businesses in increasing revenue, customer value, and service quality by knowing and addressing each student's needs. The requirement to customize any product or service given by a firm has become a necessity due to the continual increase in market demand. CRM connects front- and back-office services to customers. CRM is a collection of many customer-focused management practices that aid in the development of a long-term relationship with the client (Guerola-Navarro et al., 2021).

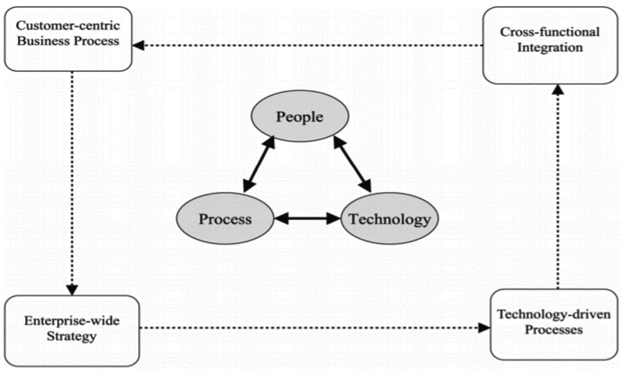


Figure 1. CRM implementation model (source: Chen & Popovich, 2003)

In Figure 1, research by (Chen & Popovich, 2003) mentioned that a CRM business plan combines marketing, operations, distribution, customer support, human resources, Research & Development, and finance, as well as information technology and the information systems, to increase the profitability of customer interactions. CRM provides consumers with customization, simplicity, and ease when performing transactions, regardless of the medium used for interaction. Technology advancements, competitive markets, and the Internet are only a few of the factors that have contributed to the success of one-to-one programmes. Companies may use these relationships to tailor the consumer experience, better anticipate online purchasing habits, entice consumers with exclusive deals or services, assess each customer's economic benefit, and create long-term mutually beneficial relationships.

1.2. The use of data mining in education

The school information management system stores a large amount of potentially relevant performance data. For a large amount of achievement data, data mining technology's association rules can be used to obtain the association relationship between subjects and find the neglected content in learning, which can provide targeted help and academic early warning for each student, as well as teaching guidance for teachers and administrators. Data mining is defined as a group of technologies that enable the automatic or semi-automatic extraction of a large amount of usable information, models, and trends from large datasets, such as "clustering", "classification", "association" and "regression"; intelligent artificial algorithms (Wang & Chung, 2021).

Educational Data Mining (EDM) is the use of data mining (DM) techniques to a specific type of dataset derived from educational contexts to resolve critical educational concerns. EDM evaluate data created by any type of information system that aides in the process of learning or education (in schools, colleges, universities, and other academic or professional learning institutions providing traditional and modern forms and methods of teaching, as well as informal learning). These data are not limited to interactions between individual students and educational systems; they may also include data from collaborating students, administrative data (e.g., school, school district, teacher), demographic data (e.g., gender, age, school grades), and student affectivity (e.g., motivation, emotional states). Figure 2 shows common characteristics such as several hierarchical levels (subject, assignment, and question levels) and fine-grained information (recording of data at different resolutions to facilitate different analyses) (Romero & Ventura, 2013).

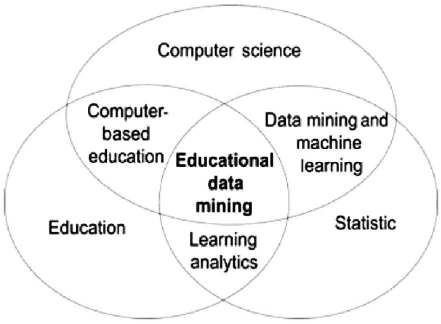


Figure 2. EDM related areas (source: Romero & Ventura, 2013)

Britto and Gobinath (2020) mentioned that data mining is a critical component of any CRM framework since it enables the analysis of business challenges, the preparation of data requirements, and the construction, validation, and evaluation of business model models. The data mining technique and algorithms enable businesses to search for, uncover, and extract useful knowledge buried in commercial data warehouses to obtain a wider understanding of business. Data mining employs sophisticated statistical data search algorithms to uncover hidden patterns and relationships to extract knowledge buried in corporate data warehouses or information left by visitors about their experience, the majority of which can result in improvements to the data's understanding and use to detect significant patterns and rules underlying consumer behaviour.

Data mining is the process of collecting relevant knowledge from data to address business challenges. It necessitates the use of a significant amount of science and technology, but the appropriate application also necessitates the use of art. However, like with many developed crafts, there is a well-understood procedure that structures the problem, allowing for sufficient consistency, repeatability, and objectivity. Figure 3 shows the Cross Industry Standard Process for Data Mining, abbreviated CRISP-DM, provides a valuable codification of the data mining process (Provost & Fawcett, 2013).

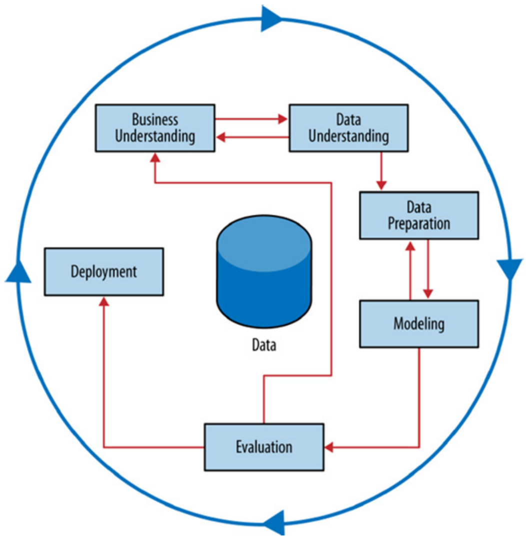


Figure 3. Data mining process (source: Provost & Fawcett, 2013)

According to (Alsagheer et al., 2017) the broad applicability of data mining is divided into two key groups:
– Descriptive data mining.
– Predictive or perspective data mining.

The two key groups can be used to create the following types of models to solve business problems:

Anomaly detection (Outlier/change/deviation detection): The process of identifying anomalous data records that may be intriguing or that include data flaws that require additional analysis. Anomaly detection is based on models of users' and applications' expected behaviour and interprets departures from this "normal" behaviour as evidence of malicious activity (Kruegel & Vigna, 2003).

Association facilitates the discovery of secret ties between disparate variables in databases. It identifies fuzzy patterns in data and uses various indicators of significance to pick the best rules from other rules. The best indicator is the lowest support and trust thresholds.

Clustering is a technology that identifies data sets that are comparable among themselves to explain the differences and similarities within the data. It is dependent on distance measurement.

Classification refers to the organisation of data into categories to make it easier to use and more effective. It aims to speed up the collection and retrieval of data, as well as to predict a specific effect based on the information provided.

Regression is a technology that allows data analysis to classify the relationships between variables. It relies on presenting values to get new values. It uses linear regression for simple cases but relative decline for complex cases that are difficult to predict because they depend on complex interactions of multiple variables.

Summarization: It is the process of reducing the size of a data set through visualisation and report generating.

2. Data mining techniques

Some researchers have merged a few algorithms to create an advanced algorithm that produces a higher accuracy rate. However, the accuracy output of each algorithm varies depending on the dataset used and the input variables chosen for the experiment. Most of the literature focuses on data mining algorithms, but only a few of them distinguish the essential input variables for churn prediction that will be used for data mining optimization technique (Iurasov & Stanelyte, 2020).

There are a plethora data mining algorithms to explore from, however, in article the data mining algorithms that will be analysed is Decision Tree Algorithm.

2.1. Decision tree

The most frequently used type of prediction model is the decision tree. It has developed into a significant knowledge structure that is used to classify future events. Typically, a decision tree is constructed in two stages: tree construction and tree pruning. The tree-building stage entails recursively splitting the training sets by their attribute values. The partitioning operation is repeated until all or most entries in each partition have similar values. Certain branches may be eliminated if they include noisy data. Pruning is the process of identifying and deleting branches with the highest estimated error rate. It is well established that tree pruning improves the predictive accuracy of the decision tree while reducing its complexity (Shaaban et al., 2012).

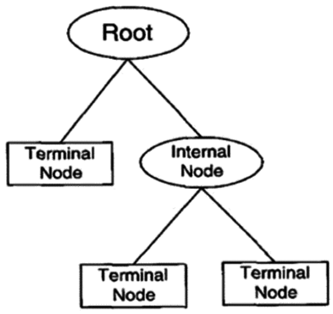


Figure 4. Decision tree form (source: Alsagheer et al., 2017)

A decision tree is a recursive partition of the instance space that expresses a classifier. The decision tree is made up of nodes that form a rooted tree, which means it is a guided tree with no incoming edges. Each of the other nodes has exactly one incoming edge. An internal or test node is a node with outgoing edges. The other nodes are referred to as leaves (also known as terminal or decision nodes) described in Figure 4. Each internal node in a decision tree divides the instance space into two or more sub-spaces based on a discrete feature of the input attribute values (Rokach & Maimon, 2006).

3. Methodology: The use of churn prediction to analyse students drop-out

This analysis would apply a quantitative approach into the research. The dataset used was gotten from VilniusTech moodle system, which contains 844 student data. Table 1 shows 31 attributes to predict the likelihood of drop-out using the Konstanz Information Miner (KNIME) analytic platform was used for the analysis.

Table 1. The attributes of the datasets

S/N	ATTRIBUTES	TYPE
1.	St_Number	Number (integer)
2.	From_Country	String
3.	Admission to University	Number (integer)
4.	Graduation year	Number (integer)
5.	Sex	String
6.	Age enrolled	Number (integer)
7.	Dropout	String
8.	Nr of failures	Number (integer)
9.	Nr of retakes	Number (integer)
10.	Weighted average	Number (double)
11.	Study program	String
12.	School graduation year	Number (integer)
13.	Lithuanian	Number (double)
14.	Mathematics	Number (double)
15.	Foreign language	Number (double)
16.	Physics	Number (double)
17.	Chemistry	Number (double)
18.	History	Number (double)
19.	Information Technology	Number (double)
20.	Geography	Number (double)
21.	Biology	Number (double)
22.	Drawing	String
23.	Moral education	String
24.	Mother tongue	Number (double)
25.	Second foreign	Number (double)
26.	Political science	String
27.	Astronomy	String
28.	Music	Number (double)
29.	Physical education	Number (double)
30.	Art	Number (double)
31.	Arts	String

The purpose of this study is therefore to answer questions such as:

- What is churn prediction model?
- How churn prediction can be used to construct a student drop-out predictive model?

3.1. Customer churn

The principle of churn is the most widely used field of analysis of customer relations. This field of application is a subsection of customer analytics modelling. From the marketing perspective, the churn principle is linked to consumer retention and value proposition. When looking at the fundamental customers' value sets, they are an example of a customer's capital or expense in their lifetime in respect of the subsequent behavioural transactions (Seymen et al., 2021). For consumer turnover, two simple methods exist: targeted and unspecified. Customer turnover identifies specific methods. They offer customers direct advantages/incentives or tailor a service plan which allows them to stay. In contrast, unspecified approaches depend on superior products and publicity to enhance brand loyalty and maintain their clients (Mehwish et al., 2017).

Churn can be defined as the study of a customer's exit. In a more refined form, it means customers discard the product or service due to competitive pressure, and consumers choose the companies or organizations that offer it. A primary goal is to recognize the condition before the consumer gets out of the product or service. The measures are just a precaution against this occurring in the future. Attracting new customers costs ten times as much as keeping existing ones, according to recent research. It is used in areas such as the identification of current customers, application of customer rebounds, and the estimation of customer reversion in various market indices. Customers' involvement is essential to the significance of these undertakings. If you understand the number of clients that a firm has, you will realize the importance of the other parameters, which are proportional, like profitability, investment capacity, cash flow, and cost (Celik & Osmanoglu, 2019).

When customers switch service providers, they "churn", customer turnover is defined as the loss of customers. The main concern for businesses with multiple clients who are quick to switch rivals is that they may easily give up the customer, making customers question their competency. The phone company, the insurance providers, banks, and the schools are great examples of churning which ranges between $500 and $50,000 for a customer. Moreover, it is much more difficult to acquire new customers than to retain current ones. In general, it is proven that it is better to focus on keeping current clients rather than new ones (Richeldi & Perrucci, 2002).

Gorgoglione and Panniello (2011) mentioned that Knowledge Discovery in Databases (KDD) is the process of discovering new, true, understandable, and useful patterns in data. This procedure can be viewed as a potential method of targeting acts to customers. If the rules derived from KDD are "implementable", that is, if they present information on a customer that can be used to determine "what to do" to change that customer's conduct, then a manager can identify marketing activities tailored to that customer. The major drawback of these methods is that they are either inefficient due to the large number of discovered rules that require human oversight, or activities are not personal because when rules are integrated to increase performance, the goal is aggregated as well.

A successful churn forecast has many beneficial consequences for corporate profitability. Firstly, the recognition of potential customers in churning enables marketing decision-makers to prioritize marketing measures economically. Retention actions may be limited to a selection of customers but are intended to include a significant number of all customers. Secondly, the high retention of customers eases the pressure to attract many customers each time. It has proven to be typically more costly in acquiring new customers than maintaining an established customer base (De Bock & Van den Poel, 2011).

3.2. Types of churners

According to (AwoyeluI., 2015) there are two types of churners; voluntary churners and involuntary churners.

– Involuntary Churners

The easiest to spot are involuntary churners, these are the consumers who are churned because of fraud or/and non-payment.

– Voluntary Churner

This is more difficult to identify; this form of churn happens when a customer decides to cancel the service contract. This is further subdivided into two categories: Deliberate and Incidental.

4. Research results

Data Pre-Processing: Missing data imputation or handling is performed during the data cleaning process. Because some of the suggested algorithms cannot handle missing data, missing values can be translated by median, mean, or zero. However, replacing missing data with a statistically calculated value is a superior option. The data set used includes missing values in several numerical variables as well as two categorical variables (Muhammad & Abubakr, 2020). Before training a model, one of the most important elements that might influence the model's success is feature selection as seen in Figure 5.

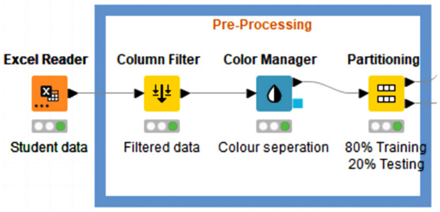


Figure 5. The data was cleaned before training and testing

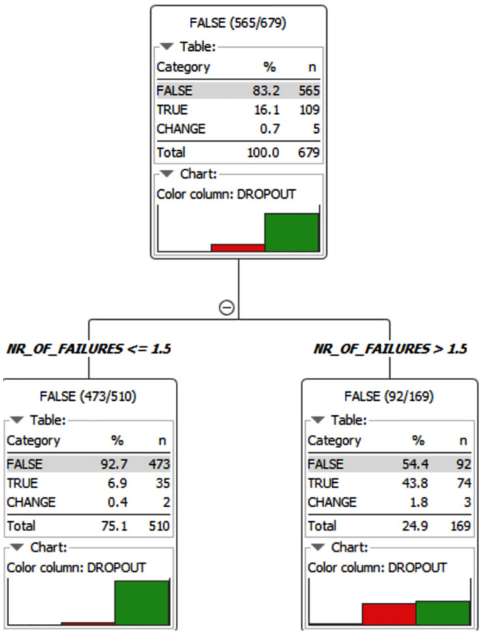


Figure 6. Decision tree model indicating the level of churn in each branch

Poor Performance: In Figure 6, the decision tree model is indicating the level of dropout in each branch. If the number of failures is >1.5 students are likely to drop-out because of poor performance.

Number of Retakes: In Figure 7, student with higher number of retakes <=4.5 are likely to drop out of the university.

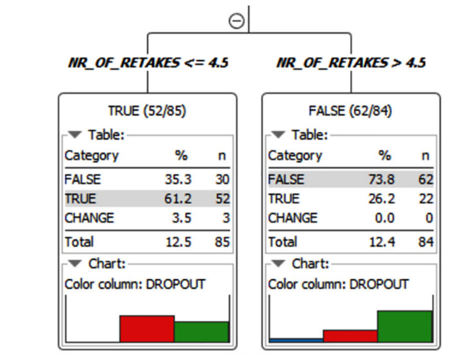
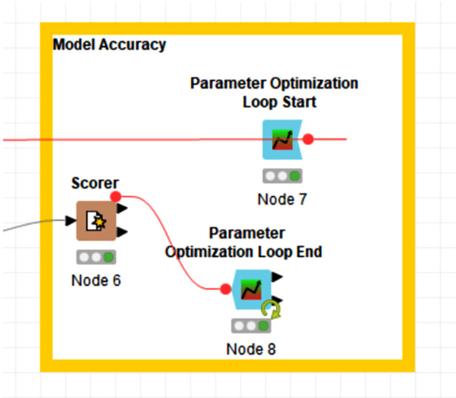


Figure 7. Decision tree indicating the level of churn in each branch

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specificity	F-meas...	Accuracy	Cohen'...
FALSE	129	13	16	12	0.915	0.908	0.915	0.552	0.912	?	?
TRUE	16	12	131	11	0.593	0.571	0.593	0.916	0.582	?	?
CHANGE	0	0	168	2	0	?	0	1	?	?	?
Overall	?	?	?	?	?	?	?	?	?	0.853	0.477



Row ID	MinRec...	Objecti...
Best parameters	20	0.876

Figure 8. Model accuracy and model was improved using the parameter optimization loop

From Figure 8 above, after training the model the model showed an accuracy of 85% which means that with the right parameters set, we can analyse the likelihood of students drop-out based on the number of retakes and number of failures. The model was further improved to the best parameters and an accuracy of 87% was achieved.

Conclusions

The conclusion derived after analyzing the data are as follows:

– Dropout rates among students are now greater than prior studies indicated.
– In many cases, factors linked to dropout are multi-causal, and they are linked to psychological as well as scholastic aspects of the student.
– It is possible to construct a decision tree model that allows us to evaluate the risk of students dropping out depending on the previously described parameters.
– The presence of poor levels of academic performance and success implies a greater probability of dropping out of the degree programme.

Many HEI have come up with designs, programmes and strategies to keep the student rate high and dropout rate significantly low. Transnational education (TNE) is a tremendous promise to reduce dropout rate and improve access to higher education for people around the globe, with far less regional and geopolitical obstacles than conventional face-to-face college and university programs. Transnational curriculum systems have been professionally planned and carefully implemented. TNE has the prospects to significantly assist developing nations in the expansion of their intellectual infrastructure. These activities also can improve student preparation in many countries to operate collaboratively within and beyond highly competitive communities. The research showed that TNE is no longer considered to be exclusive to international mobility; now it has become a critical component of higher education policy at both institutional and national levels. It is also recommended that, instead of just ignoring these economic consequences, the developing world must develop and improve the management of cooperative governance. However, the future advantages of transnational education projects must be balanced against the risks involved. TNE is a multinational corporation that needs to focus on as a means for economic development, countries that can easily be penetrated with high demand for education can be the starting point of forming an alliance.

It is recommended that faculties keep a close eye on those students performing below average academically and provide new frontiers of development and remedial programmes to help through the studies before the student settles for dropping out due to inability to keep up with peers.

References

Alsagheer, R. H. A., Alharan, A. F. H., & Al-Haboobi, A. S. A. (2017). Popular decision tree algorithms of data mining techniques: A review. *International Journal of Computer Science and Information Technology*, *6*(6), 133–142.

AwoyeluI., O. (2015). *Modeling & simulation of a predictive customer churn model for telecommunication industry*. https://www.semanticscholar.org/paper/Modeling-%26-Simulation-of-a-Predictive-Customer-for-AwoyeluI./6f41c3d1dbd05567fa8b3e455bad5343647e94ed

Britto, M. M. J., & Gobinath, R. (2020). *A detailed review for marketing decision making support system in a customer churn prediction*. www.ijstr.org

Celik, O., & Osmanoglu, U. O. (2019). Comparing to techniques used in customer churn analysis. *Journal of Multidisciplinary Developments*, *4*(1), 30–38. https://www.researchgate.net/publication/337103029

Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business Process Management Journal*, *9*(5), 672–688. https://doi.org/10.1108/14637150310496758

De Bock, K. W., & Van den Poel, D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, *38*(10), 12293–12301. https://doi.org/10.1016/j.eswa.2011.04.007

Farooqi, M. R., & Raza, K. (2011). A comprehensive study of CRM through data mining techniques. In *Apeejay School of Management*. www.sgi.com/

Gopalsamy, S., & Gokulapadmanaban, S. (2021). Does implementation of customer relationship management (CRM) enhance the customer loyalty? An empirical researching banking sector. *Iranian Journal of Management Studies*, *14*(2), 401–417. https://doi.org/10.22059/IJMS.2020.302683.674075

Guerola-Navarro, V., Gil-Gomez, H., Oltra-Badenes, R., & Sendra-García, J. (2021). Customer relationship management and its impact on innovation: A literature review. *Journal of Business Research*, *129*, 83–87. https://doi.org/10.1016/J.JBUSRES.2021.02.050

Gorgoglione, M., & Panniello, U. (2011). Beyond customer churn: Generating personalized actions to retain customers in a retail bank by a recommender system approach. *Journal of Intelligent Learning Systems and Applications*, *3*(2), 90–102. https://doi.org/10.4236/jilsa.2011.32011

Gül, H., Sallan Gül, S., Kaya, E., & Alican, A. (2010). Main trends in the world of higher education, internationalization and institutional autonomy. *Procedia - Social and Behavioral Sciences*, *9*, 1878–1884. https://doi.org/10.1016/j.sbspro.2010.12.417

Iurasov, A., & Stanelyte, G. (2020). Study of different data science methods for demand prediction and replenishment forecasting at retail network. In *11th International Scientific Conference "Business and Management 2020"* (pp. 176–185). VGTU Press. https://doi.org/10.3846/bm.2020.604

Kainuwa, A., & Yusuf, N. A. B. (2013). Students' perspective on female dropouts in Nigeria. *International Journal of Academic Research in Education and Review*, *1*(2), 29–37.

Kruegel, C., & Vigna, G. (2003). Anomaly detection of web-based attacks. In *Proceedings of the 10th ACM Conference on Computer and Communications Security* (pp. 251–261). Association for Computing Machinery. https://doi.org/10.1145/948109.948144

Lazarov, V., & Capota, M. (2007). *Churn prediction*. https://www.semanticscholar.org/paper/Churn-Prediction-Lazarov-Capota/dbf15b7c5f766ef9f84ba83127c626d79b2087b2

Mehwish, F. A., Zaffar, A. S., & Sumaira, J. M. (2017). Research article autonomous toolkit to forecast customer churn. *International Journal of Current Research*, *9*(12), 62999–63006.

Muhammad, J., & Abubakr, B. (2020). *Customer churn prediction in telecom using machine learning in Big Data platform* [Thesis for by BEng (Hons) Telecommunications Engineering with Networking University of Mauritius]. https://doi.org/10.1186/s40537-019-0191-6

Provost, & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.

Rokach, L., & Maimon, O. (2006). Decision trees. In *Data mining and knowledge discovery handbook* (pp. 165–192). Springer-Verlag. https://doi.org/10.1007/0-387-25465-X_9

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27. https://doi.org/10.1002/widm.1075

Smith, J. P., & Naylor, R. A. (2001). Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *164*(2), 389–405. https://doi.org/10.1111/1467-985X.00209

Tinto, V., & Cullen, J. (1973). *Dropout higher education: A review and theoretical synthesis of recent* (Report No. OEC-0-73-1409). https://files.eric.ed.gov/fulltext/ED078802.pdf

Wang, L., & Chung, S. J. (2021). Sustainable development of college and university education by use of data mining methods. *International Journal of Emerging Technologies in Learning*, *16*(5), 102–115. https://doi.org/10.3991/ijet.v16i05.20303

BENDROVĖS EFEKTYVUMO DIDĖJIMAS NAUDOJANT KLIENTO DUOMENŲ ANALITIKOS ĮRANKIUS: MOKSLO STUDIJA NAUDOJANT CHURN ANALIZĘ STUDENTŲ IŠKRITIMUI NUSPĖTI

Aleksei IURASOV, Ogonna Joy ANGUS

Santrauka. Siekiant padidinti aukštųjų mokyklų (AMI) efektyvumą, įvedamas į studentą orientuotas mokymo planas. Aukštųjų mokyklų valdymas yra neatsiejamai susijęs su bendro mokyklos augimo kilimu ir kritimu. Kaip vadovybė gali pasikliauti vertinimo mechanizmo veikimu, kad mokykla galėtų kurti stiprias programas ir veiklas, kurios prisitaikytų prie konkurencinių švietimo rinkos aplinkybių, tai turi būti kruopščiai apgalvotos problemų sprendimo strategijos. Didėjant e. mokymosi ištekliams, instrumentinei mokomajai programinei įrangai, interneto naudojimui švietimo srityje, kuriant studentų informacines duomenų bazes, susidarė didžiuliai švietimo duomenų rezervuarai. Gerai atliktas studentų skaičiaus numatymo modelis gali padėti aukštojo mokslo institucijoms efektyviausiu būdu sekti studentų akademinę pažangą, priėmimą į studijas ir studijų metimą, kad būtų galima pasiekti geriausius rezultatus. Šio darbo tikslas – išanalizuoti dažniausiai taikomą sprendimų medžio metodą, leidžiantį prognozuoti ir sumažinti studentų iškritimo iš aukštųjų mokyklų (AMI) tikimybę. Įmonės išleidžia daugybę sumų informacinėms technologijoms (IT) diegti ir atnaujinti technologijų pasaulyje. Taikoma tyrimo metodika bus pagrįsta skirtingais žurnalais, straipsniais ir ataskaitomis, siekiant ištirti klientų lankomumo analizės efektyvumą taikant kokybinį metodą.

Reikšminiai žodžiai: klientų duomenų analizė, ryšių su klientais valdymas, Churn prognozavimas, švietimo duomenų gavyba, aukštoji mokykla, technologijos, informacinės technologijos.