Churn Prediction on Higher Education Data with Fuzzy Logic Algorithm

Supangat[1], Mohd Zainuri Bin Saringat[2], Geri Kusnanto[3], Anang Andrianto[4]

[1,3,4]Information Engineering Department, Indonesia

[2]Information Computer And Communications Technology, Malaysia

[1]supangat@untag-sby.ac.id, [2]zainuri@uthm.edu.my, [3]gerikusnanto@untag-sby.ac.id,
[4]anangandriano@gmail.com

Abstract— Colleges are an optional final stage of formal education. However, with time, the Management Section finds the fact that the student churn rate in the university scope becomes a problem. The purpose of this research is to predict whether the student will be churn or loyal in the future, the data will be taken from 2014. The analytical technique used in this research process is the Fuzzy method of C-Means. At a university, the variable that will be tested is the Length of students who are currently attending higher education at the University or campus, when the last student pays enrollment, the last payment period, and the student's total payment. The data used as many as 100 datasets were collected in this research, starting from 2014 to 2020. Of the 100 Datasets of informatics engineering, students gained 92% of loyal students, and 8% of students predicted to churn.

Keywords— Churn Redictin, Fuzzy C-Means, Higher Education.

I. INTRODUCTION

The dropout rate is the most significant indicator to assess the quality of a higher education institution. The dropout rate refers to the percentage change between the number of students enrolled, and the number of students enrolls the previous year [1]. Reasons to drop out can be varied, from mismatch prospective student orientation with the course of study selection [2], willingness to enter the labor market as soon as possible, lack of financial aid to study [3], educational background, academic performance, and characteristic of the student [4]. Students drop out can be considered customer churn in the service industry, which is a significant problem for service providers, including higher education institutions [5]. The student regretted their study choice and moved to another campus or out of the campus. Generally, the dropout rate indicates that an institution cannot meet the student's expectations who enroll there [6].

In Europe [1], more than 3 million students drop out from higher education with the highest dropout country in France (32%), Italy (15.8%), and UK (12%). In Indonesia, dropout rates reach only 2.8% in 2017 and 3% in 2018 [7]. Even drop out the rate is relatively low in Indonesia, the student leaving from higher education institute is still necessary to be predicted and prevented because of university ranking assessment and its reputation [6].
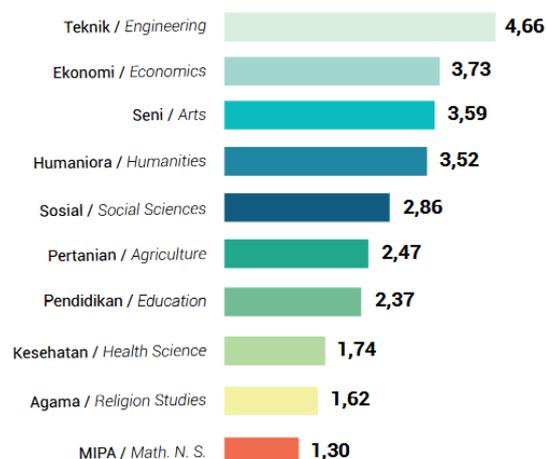
Figure1: Drop Out Rates based on Academic Fields

Figure 1 describes that engineering fields are academic fields with the highest drop out rate as 4,66%. In Italy, the engineering field is the second rank with 25% dropout rate after the professional field with 35% drop

out [1]. So, this research focuses on higher education institutes with the engineering field in Indonesia.

To maintain existing students, universities need to improve the service, improve the quality of teaching, improve campus security, improve the environmental quality inside-outside the campus and know earlier which students will have the possibility to leave the institution or retain until graduation [1]. The higher education management needs to identify earlier the student who has a higher tendency to enroll and retain in their institutions and therefore be able to allocate their money effectively [8]. Furthermore, the better prediction of a student who is willing to retain in higher education is crucial to formulate an effective CRM strategy that works on developing long-term and sustaining a profitable relationship with the customer and key stakeholders [9].

To identify a user who tends to leave or remain in the institution, the churn prediction can be used so the campus possibilities of losing students can be avoided [1]. In churn analysis, one model is commonly applied is the analysis of Recency, Frequency, and monetary (RFM) [10],[11]. RFM model analysis generally processes the customer's transaction data include latest transaction date (Recency), transaction amount (Frequency), and monetary amount per transaction (monetary) [10]. The output of RFM analysis is consumption behavior [12], customer future value [13], and customer lifetime value [14],[15]. The most important output of RFM analysis for churn prediction is customer lifetime value [16]. Churner or customer with a high tendency to leave can be defined as someone with customer lifetime value decreasing over time [16]. In recent literature, RFM analysis is modified by adding customer relationships with the company [17]. Adding the L (Length of customer relationship) variable to the RFM model will lead to a more accurate customer analysis and increase the segmentation quality [18]. Therefore, this research use LRFM analysis to predict student churn or drop out in higher education based on customer lifetime value.

LRFM analysis implementation in churn prediction is very fit when combined with data mining techniques [1],[17],[19],[20]. The main objective of churn prediction with machine learning is to categorize the customers into two clusters, churners or non-churners, and suggest that prior customers be targeted with retention strategies [20]. Various machine learning techniques have successfully been used to predict customer churn in a recent study such as logistic regression (LR)[21], Artificial Neural Networks (ANN)[22], Fuzzy C-Mean algorithm [23],Decision Trees (DT)[24] and Support Vector Machine (SVM) [25]. Deep learning methods are applied in the machine learning section that can perform data processing. This research uses fuzzy C-means to test the eligibility of churn predictors. Fuzzy C-means process data input – user data with grouping analysis.

II. METHOD

In this research, classification is a quantitative approach, with experimental methods. The quantitative approach is objective, oriented to the verification conducted in this study to attempt the hypothesis that data mining techniques could be used to predict churn levels. This approach is made by the experimental method of investigation due to a causal relationship using the researcher's controlled trial. This experiment was done with the dataset test, and the data mining model validation is generated through the training process. In this stage, the data that has been inserted into the application will be processed. Several processes are done before the result of the predictions comes out, ranging from normalization, determination of random value, calculation of membership, looking for the total cost of membership, the maximum result of iteration, and minimum error margin.

A. Analysis

Analyze the problems that happened and see the needs of the users made to complete questions. Start by analyzing by coming to the University that research will do.

B. Study Literature

Literary studies are useful for obtaining supporting theories in research, such as identifying predictions using the Fuzzy C-Means. Literature studies are obtained from journals, scientific papers, books as well as research and articles that have been done before

C. Data Collection and Observation

At this stage will be conducted observation and collection of data done by going directly to the data source in the 17 Agustus University Surabaya scope, which will be used for objects in the manufacturing of predictive systems. We define the Length (L) variable as Length of student enrollment with higher education, Recency (R) is defined as the number of payment since last enrollment (months). Frequency is defined as the number of enrollment in higher education. Monetary (M) is defined as the total amount of money paid when students enroll in higher education.

D. Application Design

At this stage will be performed application design of the Fuzzy calculation of C-Means, determination of variables used for predictions, and application design. So the results are expected to know students who churn or do not churn.

E. Implementation

At the implementation stage, the design of the previously created applications will begin to be implemented in real-time in desktop applications.

F. System Trial

The trial process of this application will be based on the data that has been collected to see the number of requests that have been created.

G. Documentation

At this stage, drafting the report and documentation is obtained from all research stages, ranging from the initial stage to the analysis and evaluation stage.

III. RESULTS AND DISCUSSION

The research used the university students' data 17 Agustus 1945 Surabaya with a total of students in the analysis period from 1 January 2014 to 30 April 2020 as much as 100 data. Data will be a preprocessing process first. This stage is a useful process for improving the accuracy and efficiency of data modeling. This stage helps see if there are double, empty, so data will later damage the research results. The data is then selected according to the attributes Length, Recency, Frequency, and Monetary; Data that has been chosenaccording to Length, Recency, Frequency, and Monetary is further done normalization by using equations. Standardization of the research is used because the monetary value of M has a difference is much different compared to the cost of L, R, and F. This is because the value of M is the amount of money that is issued by the university student with its unit of Rupiah (RP). The very distant difference is that normalization is required to avoid disrupting research results. The Range used in this research is the value between 0-1.

Table 1: Raw Data

	A	C	D	E
1	nim	nominaltagihan	tgllunas	periode
2	11114085:	1500000	2014-06-02	20141
3	11114085:	400000	2014-12-08	20141
4	11114085:	400000	2014-06-02	20141
5	11114085:	400000	2014-10-08	20141
6	11114085:	400000	2014-12-08	20141
7	11114085:	400000	2014-10-08	20141
8	11114085:	400000	2015-06-26	20142
9	11114085:	400000	2015-02-03	20142
10	11114085:	400000	2015-04-29	20142

After the raw data is transformed into transformation data, Length is transformed from first entry enrollment to last enrollment date, Recency is transformed from last enrollment to late enrolment date, Frequency

is transformed from every student enrollment date, and Monetary is transformed from the total amount of student enrollment fee.

Table 2: Transformation Data

	A	B	C	D	E
1	nim	periodemasukLength	tgllunasRecency	totaltransaksiFrequency	totalPembayaranM
2	1461404	13	4	13	3
3	1461404	13	28	13	3
4	1461404	13	3	13	3
5	1461404	13	4	13	4
6	1461404	13	4	13	4
7	1461404	13	3	13	3
8	1461404	13	3	13	4
9	1461404	13	3	13	3
10	1461404	13	3	13	3

Tabel 2 describes transformation data from students' big data based on churn predictor (Length, Recency, Frequency, and monetary). Once the normalization stage is performed, the centroid value is calculated by the Fuzzy C-Means (FCM) method. In FCM analysis, the data were divided into two clusters: 1) for loyal student and 2) for churners or potential students will drop out. FCM is needed to apply to calculate the cluster centers or centroid values to set up two clusters. These cluster centers are used for partitioning the input data into two different clusters.
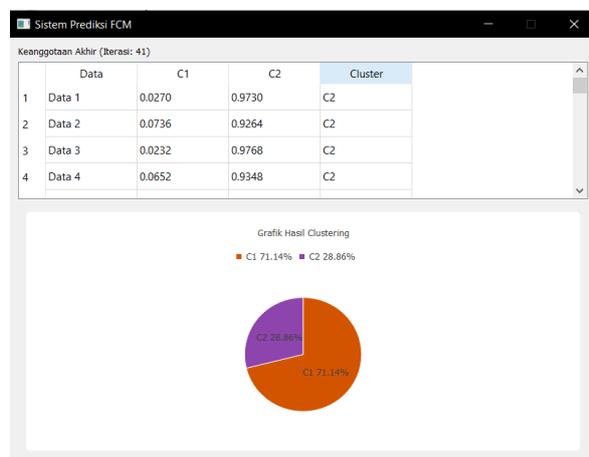
Figure 2: Result Cluster

The Result of a fuzzy C-means calculation based on LRFM show that 71.14 % is grouped in cluster 1 as loyal students of 17 Agustus 1945 University, and 28.86% is grouped in cluster 2 as churner students of 17 Agustus 1945 Universitywho will drop out in the future. This result shows that 28.86% of students are predicted to drop out

of the 17 Agustus 1945 University for varied reasons, including a willingness to work in the labor market, incorrect student orientation, insufficient motivation to finish the study, and difficulties in the course of study.

This churn rate prediction is relatively high, compare to actual drop out data on higher education in Indonesia, which is only 4% in 2018. This study's churn rate is prediction, which means students have the intention or tendency to leave higher education in the future. So, 28.86% tend to leave 17 Agustus University in the future. The prediction is almost near the previous study, a 29% churn customer predicted [11].

The high churn rate in higher education should be considered by 17 Agustus 1945 University to make preemptive action to prevent actual churn in the future. Perchinunno et al. [1]suggest several preemptive actions might be used to prevent student drop out of higher education, which is 1) academic tutoring activities – to reduce difficulties in passing the exam, 2) guidance and counseling for matriculation – to improve student's convergence between study goal and their objective, so that student keep attend and continue the studies, 3) financial support, to support student focus on study and implement their project.

IV. CONCLUSION

The conclusion of the Data of higher education Churn prediction analysis with the Fuzzy Logic algorithm is the Fuzzy method C-means can use to clustering Churn the customers imported from the data set preprocessing so that the normalization results are obtained from each data. After that is calculating the membership degrees of each information using the Fuzzy method C-means. Once the cluster value is obtained for each input, it will look for the highest membership value, the highest amount will determine where this data will be in which cluster. From the sample, churn prediction result show that 71.14% is loyal students, and 28.86% wouldchurn.

Churn prediction rate in the 17 Agustus University is quite higher than average higher education student churn rate in Indonesia. So that the 17 Agustus University should develop retention strategy to retain student such as academic tutoring activities, guidance and counseling for matriculation, and financial support.

For future study, we suggest using another technique to predict churner on higher education such as Decision Trees (DT), Artificial Neural Networks (ANN), logistic regression (LR), and Support Vector Machine (SVM).

REFERENCES

[1] Perchinunno, P., M. Bilancia, dan D. Vitale, "A Statistical Analysis of Factors Affecting Higher Education Dropouts". Social Indicators Research, 1-22, 2019

[2] Smith, J. P., dan Naylor, R. A., "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, 2001.

[3] Murray, M.. "Factors affecting graduation and student dropout rates at the University of KwaZulu-Natal",South African Journal of Science, 110(11/12), 1–6, 2014

[4] Montmarquette, C., Mahseredjian, S., & Houle, R., "The determinants of university dropouts: A bivariate probability model with sample selection",Economics of Education Review, 20(5), 475–484, 2001

[5] Becker, Jan U., Martin Spann, and Christian Barrot, "Impact of proactive post-sales service and cross-selling activities on customer churn and service calls".Journal of Service Research 23.1, 53-69, 2020

[6] Thebestschools.org, "HowWe Rank Schools at TBS", https://thebestschools.org/rankings/ran

[7] Pddikti.kemdikbud.go.id, Statistik Pendidikan Tinggi. https://pddikti.kemdikbud.go.id/asset/d ata/publikasi/Statistik%20Pendidikan% 20Tinggi%20Indonesia%202018.pdf, [Accessed on 25 sept 2020, 2018]

[8] A. Slim, D. Hush, T. Ojah, and T. Babbitt, Predicting student enrollment based on student and college characteristics,Proc. 11th Int. Conf. Educ. Data Mining, EDM 2018, 2018.

[9] Xu,L.D, "Enterprise systems: state of the art and future trends".IEEE Transactionson Industrial Informatics, 7(4),630–640, 2011.

[10] Chen, Y. L., M. H. Kuo, S. Y. Wu, and K. Tang, "Discovering recency, frequency, and monetary(RFM) sequential patterns from customers'purchasing data",Electronic Commerce Researchand Applications, vol. 8, no. 5, pp. 241–251, 2009.

[11] Hidayat, R. Rismayati, M. Tajuddin, and N. L. P. Merawati, Segmentation of university customersloyalty based on RFM analysis using fuzzy c-means clustering", JurnalTeknologi dan SistemKomputer, vol. 8, no.2, pp. 133-139, 2020

[12] Chang, H. and H. Tsai, "Group RFM analysis as a novel framework to discover better customer consumption behavior". Expert Systems with Applications 38.12, 14499-14513, 2011.

[13] Khujand, M. and M. J. Tarokh, "Estimating customer future value of different customer segments based on adapted RFM model in retail banking context".Procedia Computer Science 3, 1327-1332, 2011

[14] Safari, F., N. Safari, and G. Ali Montazer. "Customer lifetime value determination based on RFM model".Marketing Intelligence & Planning. 2016.

[15] Heldt, R., C. S. Silveira, and F. B. Luce, "Predicting customer value per

product: From RFM to RFM/P".Journal of Business Research, 2019.

[16] Glady, N., B. Baesens, and C. Croux, "Modeling churn using customer lifetime value",European Journal of Operational Research, 197.1: 402-411, 2009.

[17] Khodabandehlou, S., "Designing an e-commerce recommender system based on collaborative filtering using a data mining approach",International Journal of Business Information Systems 31.4: 455-478, 2019.

[18] Li, D., W. Dai, and W. Tseng, "A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business,Expert Systems with Applications 38.6, 7186-7191, 2011.

[19] Schaeffer, S. E., and S.V. R. Sanchez, Forecasting client retention—A machine-learning approach,Journal of Retailing and Consumer Services 52, 101918. 2020.

[20] Khodabandehlou, S, and M. Z. Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior",Journal of Systems and Information Technology, 2017.

[21] De Caigny, A., K. Coussement, and K. W. De Bock. "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees".European Journal of Operational Research 269.2, 760-772, 2018.

[22] Sivasankar, E., and J. Vijaya. "Hybrid PPFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network".Neural Computing and Applications 31.11, 7181-7200, 2019

[23] Karahoca, A., and D. Karahoca, "GSM churn management by using fuzzy c-means clustering and adaptive neuro-fuzzy inference system", Expert Systems with Applications 38.3, 1814-1822, 2011.

[24] Lee, E., J. Kim, and S. Lee, "Predicting customer churn in mobile industry using data mining technology".Industrial Management & Data Systems,2017.

[25] Gordini, N., and V.Veglio, "Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry",Industrial Marketing Management 62, 100-107, 2017.

[26] Rajamohamed, R., and J. Manokaran, "Improved credit card churn prediction based on rough clustering and supervised learning techniques",Cluster Computing, 21.1, 65-77. 2018

[27] Milošević, M., N. Živić, and I. Andjelković, "Early churn prediction with personalized targeting in mobile social games".Expert Systems with Applications 83, 26-332, 2017.

[28] Ling,R., and Yen,D.C., Customerrelationshipmanagement:an analysisframeworkandimplementations trategies".J.Computer Information System., 41(3),82–97, 2001.

[29] Derby,N., "Reducingcustomerattritionwithmachin elearningforfinancialinstitutions".Proce edingsoftheSASGlobalForum,pp.1796, 2018.

[30] Al-Mashraiea, M., S. H. Chunga,H. W. Jeon, Customerswitchingbehavioranalysisint hetelecommunicationindustryviapush-pull-mooringframework:Amachinelearninga pproach", Computers and IndustrialEngineering144, 2020.

[31] Van den Poel, D. andLarivière, B., "Customer attrition analysis for financialservices using proportional hazard models",European Journal of OperationalResearch 157 (1), 196–217,2004.

[32] Buckinx, W. and Van den Poel, D., "Customer base analysis: Partial defection ofbehaviorally-loyal clients in a non-contractual FMCG retail setting",EuropeanJournal of Operational Research 164 (1), 252–268, 2005.

[33] Rust, R.T., Lemon, K.N. and Zeithaml, V.A., "Return on marketing: Using customer equity to focus marketing strategy",Journal of Marketing 68 (1), 109–127, 2004.

[34] Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Van Kenhove, and P.,Vanthienen, J., "Bayesian network classifiers for identifying the slope of thecustomer lifecycle of long-life customers",European Journal of OperationalResearch 156 (2), 508–523, 2003.

[35] Gupta, S., Lehmann, D.R., and Stuart, J.A., "Valuing customers". Journal of MarketingResearch 41 (1), 7–18, 2004.

[36] W. Anggraeni, "Penentuan Nilai Pangkat Pada Algoritma Fuzzy C-Means",Fakt. Exacta, vol. 8, no. 3, pp. 266–278, 2015.

[37] L. Warokaet al., "implementasialgoritma fuzzy c-means (fcm) dalampengklasterisasiannilaihiduppela nggandengan model lrfm pada barbershop omarjalandelimapekanbaruriau",Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi, vol. 6, no. 1, pp. 1–5, 2020.

[38] S. Monalisa, "Klusterisasi Customer Lifetime Value dengan Model LRFM menggunakanAlgoritma C-Means",J. Teknol. Inf. dan IlmuKomput., vol. 5, no. 2, p. 247, 2018.

[39] Hu, Y.-H. and Yeh, T.-W. "Discovering valuable frequent patterns based on RFM analysis without customer identificationinformation",Knowledge-Based Systems, Vol. 61, pp. 76-88, 2014.

[40] Nenonen, S. and Storbacka, K., "Driving shareholder value with customer assetmanagement: moving beyond customer lifetime value",Industrial Marketing Management,Vol. 52, pp. 140-150, 2015.

[41] Yeh, I.C., Yang, K.J. and Ting, T.M. "Knowledge discovery on RFM model using Bernoullisequence",Expert Systems with Applications, Vol. 36 No. 3, pp. 5866-5871, 2009.

[42] King, S.F., "Citizens as customers: exploring the future of CRM in UK local government",Government Information Quarterly, Vol. 24 No. 1, pp. 47-63. 2007.

[43] Coussement, K. and Poel, D.V. "Improving customer attrition prediction by integrating emotionsfrom client/company interaction emails and evaluating multiple classifiers",Expert Systems withApplications, Vol. 36 No. 1, pp. 6127-6134, 2009.

[44] Murakami, K. and Natori, S., "New customer management technique: CRM by'RFMþI'analysis",Nomura Research Institute, Vol. 186, 2013.

[45] Hosseini, S.M.S., Maleki, A. and Gholarmian, M.R., "Cluster analysis using data mining approachto develop CRM methodology to assess the customer loyalty",Expert Systems with Applications,Vol. 37 No. 7, pp. 5259-5264, 2010.

[46] Ortega, John Heland Jasper C., et al., "An Analysis of Classification of Breast Cancer Dataset Using J48 Algorithm", International Journalof Advanced Trends in Computer Science and Engineering 9.1.3, 2020.

[47] Durán, C., et al. "Fuzzy Knowledge to Detect Imprecisions in Strategic Decision Making in a Smart Port", International Journalof Advanced Trends in Computer Science and Engineering, 9.1.3, 2020.

[48] Mohamed, R. R., Razali Yaacob, M. A. M., Dir, T. A. T., & Rahim, F. A., "Data Mining Techniques in Food

Safety".International Journal of Advanced Trends in Computer Science and Engineering, 9.1.1, 2020.

[49] Awang, N., Xanthan, A., Samy, L. N., & Hassan, N. H. "A Review on Risk Assessment Using Risk Prediction Technique in Campus Network".International Journalof Advanced Trends in Computer Science and Engineering, 9.1.3, 2020.

[50] Dunn, J. C., "A fuzzy relative of the ISODATA process and its use in detectingcompact well-separated clusters", Journal of Cybernetics, 3, 32–57, 1973.

[51] Huang, Kuang Yu. "Applications of an enhanced cluster validity index method based on the Fuzzy C-means and rough set theories to partition and classification".Expert Systems with Applications 37.12, pp. 8757-8769. 2010.

[52] M. Muhammad Faizal, "Metode Clustering DenganAlgoritma Fuzzy C-Means UntukRekomendasiPemilihanBidangK eahlian Pada Program Studi Teknik Informatika",KaryaIlm., vol. 3, no. Data Mining, pp. 1–2, 2013.

[53] Supangat and Andrianto, A., "PerancanganSistemPrediksi Churn PT JATAYU MenggunakanAlgoritma Fuzzy Logic",dissertation,Dept. InformaticEngineering, Universitas 17 Agustus 1945, Surabaya, Indonesia, 2020.