Customer Churn Prediction by Rough Neuro-Fuzzy Classifier with CA Defuzzification

Robert K. Nowicki
Częstochowa University of Technology/Department of Intelligent Computer Systems
Częstochowa, Poland *robert.nowicki@pcz.pl*

Magdalena M. Scherer
Częstochowa University of Technology/Faculty of Management
Częstochowa, Poland *magdalena.scherer@pcz.pl*

Abstract

Given that churn management is a crucial endeavour for firms aiming to retain valuable customers, the capacity to forecast customer churn is indispensable. We use rough and fuzzy set based classifier to predict customer churn on the example of the Bank Customer Churn dataset. Rough set theory offers techniques for handling incomplete or missing data. By utilizing lower and upper approximation concepts, the system can still perform prediction even when certain feature values are missing what we show in the paper for every combination of missing features. Moreover, we determine feature importance coefficient evaluated through two different means: directly from data and from the working classifier. Rough set-based systems can be integrated with other machine learning and data mining techniques, and we use the LEM-2 rule induction algorithm to create a rule base for the rough-fuzzy classifier.

Keywords: rough sets, rough neuro-fuzzy classifier, customer churn prediction.

1. Introduction

Highlighting the greater expenses involved in acquiring new customers in contrast to keeping current ones, along with the tendency for long-standing customers to generate greater profits, it becomes evident that maintaining customer loyalty enhances profitability. Numerous competitive companies have come to recognize that a fundamental tactic for thriving in the industry involves maintaining current clientele. Customer churn, also known as attrition or turnover customers, refers to the rate at which customers terminate their relationship with a company or cease using its products or services within a specified period of time. Customer churn can have various reasons, such as dissatisfaction with a product or service, competitive offerings, poor customer experiences, pricing issues, or changes in customer needs or circumstances. High churn rates can be detrimental to a business as they can lead to revenue loss, reduced market share, and increased customer acquisition costs.

Predicting customer churn is crucial in all industries as it enables a better understanding of customers and forecasting future revenues. Additionally, it allows companies to identify and improve areas lacking in customer service. Much research has been conducted in this field, but many industries still have untapped customer data to explore. It is worth noting that results vary depending on data from different industries. Customer churn poses a challenge for various sectors, especially in highly competitive markets. Losing customers leads to financial losses due to decreased sales and increases the need to acquire new customers [21].

The issue of Customer Churn has escalated across various sectors, encompassing the banking industry, prompting banks to consistently monitor customer engagement in order to identify those likely to depart. Customer Churn modeling predominantly targets potential defectors, enabling proactive measures to mitigate churn [2]. The optimal foundational marketing strategy for the future lies in retaining existing customers and avoiding customer churn [9]. Enhancing the retention rate by as little as 5 % can lead to a potential profit boost of up to 85 % for a bank [13]. Moreover, there is a growing recognition of the heightened significance of customer retention compared to previous perspectives. Churn management is integrating into customer relationship management [17]. Companies should prioritize this integration as they aim to cultivate enduring relationships with customers and optimize the value of their customer base.

Entering a competitive market requires companies to focus primarily on profits derived from customers. Therefore, Customer Relationship Management (CRM) always concentrates on confirmed customers, who are the most fertile source of decision-making data. These data reflect the actual individual behaviors of customers. Such behavioral data can be used to assess potential customer value, evaluate the risk of them ceasing to pay their bills, and predict their future needs [12]. Moreover, as customer churn can lead to business loss, predicting churn has gained increasing attention in marketing and management literature recently. Additionally, it is shown that even a small change in the retention rate can have a significant impact on business operations [19].

The fuzzy set theory was proposed by Lotfi A. Zadeh in 1965 as a means to represent and work with uncertainty and vagueness in data and decision-making [22]. The rough set theory was proposed by Polish mathematician and computer scientist, Zdzisław Pawlak, in [16] to deal with uncertainty and vagueness in data analysis and decision-making processes. In this paper we use rough and fuzzy setbased classifier to predict customer churn on the example of the Bank Customer Churn dataset. Rough set-based systems are a valuable tool for analyzing and classifying data, especially in situations where data may be incomplete, imprecise, or uncertain. By identifying the most relevant features or attributes, rough set-based systems can streamline the classification process and reduce computational complexity. They can generate classification rules directly from the data. These rules are typically expressed in a human-readable format, making them easy to interpret and understand. This interpretability is valuable for decision-making and gaining insights into the underlying patterns in the data.

Rough set theory offers techniques for handling incomplete or missing data. By utilizing lower and upper approximation concepts, rough set-based systems can still perform analysis and classification even when certain data points are missing what we show in the paper. Rough set-based systems can be integrated with other machine learning and data mining techniques to enhance classification performance. Namely, we use the LEM-2 rule induction algorithm [6] to create a rule base for the rough-fuzzy classifier. Intervals from LEM-2 rules are changed into Gaussian fuzzy sets, and we use ten-fold cross-validation creating ten classifiers and averaging their predictions for all the combinations of missing feature values. We applied the LEM-2 algorithm solely as a generator to obtain the rules. Moreover, the rules taken from LEM-2 have been fuzzified to obtain a better generalisation effect. Of course, various rule generation methods can used as well. The main goal of the work is to use a rough-set-based system to handle possible missing feature values and to research feature significance changes depending on the availability of values of other conditional attributes in the context of churn prediction.

The rest of the paper is organized as follows. In Section 2 we describe selected works on customer churn prediction. In Section 3 we describe the churn dataset we use, and in Section 4 we describe the rough-fuzzy classifier. The results of the experiments are presented in Section 5, and Section 6 concludes the paper.

2. Related Works

To effectively manage customer churn, it is important to develop a more efficient and accurate model for predicting customer churn. In the literature, statistical techniques and data mining have been used to create predictive models. Data mining tasks can be used for description (i.e., discovering interesting patterns or relationships in data) and prediction (i.e., forecasting or classifying model behavior based on available data) [4]. In other words, it is an interdisciplinary field with the overarching goal of predicting outcomes and applying advanced algorithms to primarily uncover hidden patterns, relationships, anomalies, and/or structures from extensive data stored in data warehouses or other repositories of information, as well as filtering necessary information from large datasets [7].

In the literature, hybrid data exploration models, combining clustering and classification techniques, can improve the performance of individual clustering or classification techniques. In particular, they consist of two learning stages, where the first serves for 'preprocessing' the data, and the second for the final output of predictions [11].

However, there is limited research on assessing the performance of hybrid data exploration techniques in predicting customer churn. Therefore, this article investigates two different methods of combination to create hybrid models in terms of predicting customer churn. The first is based on the combination of clustering, i.e., self-organizing maps (SOM), and classification techniques, i.e., artificial neural networks with backpropagation (ANN), namely SOM + ANN.

The modern business era, particularly in the telecommunications and banking sectors, focuses on building enduring customer relationships and understanding the factors influencing their loyalty. One of the key challenges facing enterprises is understanding and predicting customer churn - the phenomenon where customers terminate services or products offered by a company and switch to competitors. In this context, data analysis and the utilization of machine learning methods become indispensable tools for identifying patterns and factors determining customer churn.

In earlier studies [8], the technique of decision trees was applied to build a model identifying the characteristics of bank customers who churned their services. The study results depict the profile of customers who terminated services, with conclusions indicating the potential identification of future service-churning clients through decision trees. Authors suggest that bank managers should develop customer retention strategies, especially for those whose traits increasingly resemble those of customers who churn. Meanwhile, in the research by Wadikar [20], a comparative study of the most popular supervised machine learning methods - Logistic Regression, Random Forest, Support Vector Machine (SVM), and Neural Network - was conducted in the context of predicting churn among customers leaving the Credit Union financial institution. In the study by Vafeiadis et al. [18], a comparative analysis of popular machine learning methods such as Artificial Neural Network, Support Vector Machines, Decision Trees Learning, A Bayes classifier, and Logistic Regression Analysis was presented, applied in the challenge of predicting customer churn in the telecommunications industry.

3. Customer Churn Dataset

We use Bank Customer Dataset [1] with ten thousand cases as it contains a mix of categorical and numerical features, allowing the evaluation of how well the model handles different data types. It can also help in understanding customer behaviors and decision-making processes, which are essential for banks. Rough-based systems excel at handling redundant and irrelevant features, common in customer datasets. Typically, in each bank all pertinent customer information is gathered and validated, with each client distinguished by a unique customer ID and associated surname. The dataset encompasses various customer details, including credit scores, age, tenure, balance, number of products, and estimated salary. Boolean measurements, represented by values of 0 or 1, are included alongside other categorical sections featuring two or more classes, such as county, gender, possession of a credit card, active membership status, and churned status. The column labeled “exited” denotes the current customer state, with a value of 1 indicating customer attrition. The collected data exhibit variability based on factors such as customer location, economic status, and gender. The number of products utilized by a customer correlates with their loyalty and profitability to the bank. To facilitate modeling without losing information, categorical data were converted into one-hot encoding format. In our case, feature Geography is converted to one-hot encoded three variables (Germany, France and Spain). Additionally, irrelevant columns such as "RowNumbers", "CustomerID" and "Surname" were excluded from analysis.

For a given decision table, as well as for any defined decision system, the significance of attributes or their groups [3] is determined by

$$\sigma_{(C,D)}\left(C'\right) = \frac{\gamma_C\left(D^*\right) - \gamma_{C"}\left(D^*\right)}{\gamma_C\left(D^*\right)} \tag{1}$$

where $C', C" \subseteq C$ are subsets of conditional attributes (can be even empty)

$$C" = C \setminus C' \tag{2}$$

and

$$\gamma_C\left(D^*\right) = \frac{\overline{\overline{\mathrm{Pos}_C\left(D^*\right)}}}{\overline{\overline{U}}} \tag{3}$$

where D^* is a family of sets determined by decision attributes D and P is the subset of conditional attributes C. Significance $\sigma_{(C,D)}(C') \subset [0;1]$ and the value 0 occurs when the lack of values of attributes C' does not affect classification. Value 1 means that without values of attributes C' the classification will be impossible. The significance values for the used dataset are given in Table 1. In Section 5 we compute significance from ten created classifiers by removing all possible features apart one each time.

Table 1. Normalized significance coefficient computed directly from the data for bank churn across ten attributes. In the experiments, Geography is treated as one-hot encoded three variables but these coefficients stay the same after encoding.

Feature	Significance
CreditScore	0.0134
EstimatedSalary	0.0105
Tenure	0.0095
Age	0.0078
Balance	0.0054
IsActiveMember	0.0022
NumOfProducts	0.0020
HasCrCard	0.0008
Geography	0.0000
Gender	0.0000

4. Rough Neuro-Fuzzy Classifiers with CA Defuzzification

The structure of the rough neuro–fuzzy classifier with the Centre Average (CA) defuzzification is based on its corresponding neuro-fuzzy architecture

$$\overline{z}_j = \frac{\sum\limits_{\substack{r=1 \\ r:\, \overline{z}_j^r=1}}^{N} \mu_{A^r}(\mathbf{v})}{\sum\limits_{r=1}^{N} \mu_{A^r}(\mathbf{v})}. \tag{4}$$

where $\overline{z}_j$ is a representative of set B_j in classifiers. Such neuro-fuzzy system was first proposed by Wang in [10]. The system infers about the membership degree of the object x to class ω_j, i.e. $\overline{z}_j = \mu_{\omega_j}(x)$ and using a knowledge base consisting of fuzzy rules

$$\begin{gathered} R^r\colon \text{IF } c_1 \text{ is } A_1^r \text{ AND } c_2 \text{ is } A_2^r \text{ AND} \ldots \\ \ldots \text{ AND } c_n \text{ is } A_n^r \text{ THEN } z_1 \text{ is } B_1^r, z_2 \text{ is } B_2^r, \ldots \\ \ldots, z_m \text{ is } B_m^r, \end{gathered} \tag{5}$$

The rough neuro–fuzzy classifier with the Centre Average (CA) defuzzification consists of two parallel connected structures with common knowledge. The part leading to the lower value of the output membership degree (the left bound of the output interval) is defined as follows

$$\underline{\overline{z}_j} = \frac{\sum\limits_{r=1}^{N} \overline{z}_j^r \cdot \mu_{\underline{\widetilde{P}}A^r}(\mathbf{v})}{\sum\limits_{r=1}^{N} \overline{z}_j^r \cdot \mu_{\underline{\widetilde{P}}A^r}(\mathbf{v}) + \sum\limits_{r=1}^{N} \neg\overline{z}_j^r \cdot \mu_{\overline{\widetilde{P}}A^r}(\mathbf{v})}, \tag{6}$$

The upper output value (the right bound of the output interval) is calculated using the following formula

$$\overline{\overline{z}_j} = \frac{\sum\limits_{r=1}^{N} \overline{z}_j^r \cdot \mu_{\overline{\widetilde{P}}A^r}(\mathbf{v})}{\sum\limits_{r=1}^{N} \overline{z}_j^r \cdot \mu_{\overline{\widetilde{P}}A^r}(\mathbf{v}) + \sum\limits_{r=1}^{N} \neg\overline{z}_j^r \cdot \mu_{\underline{\widetilde{P}}A^r}(\mathbf{v})}, \tag{7}$$

where $\neg\overline{z}_j^r = 1 - \overline{z}_j^r$ and Pos_C is a set of available conditional attributes applied during the reasoning process. Thus, $\widetilde{P}$ is an equivalence relation based on set P, $\underline{\widetilde{P}}A^r$ is a lower and $\overline{\widetilde{P}}A^r$ is the upper approximation of fuzzy set A^r. The correctness of such construction has been proved in [14]. Figure 1 presents the architecture of such dual system. When the values of parameters $\overline{z}_j^r$ have been finally set, the structures of the classifiers can be simplified by eliminating these weights and some connections as follows

$$\underline{\overline{z}_j} = \frac{\sum\limits_{\substack{r=1 \\ r\colon \overline{z}_j^r=1}}^{N} \mu_{A_{\mathrm{L}}^r}(\mathbf{v})}{\sum\limits_{r=1}^{N} \mu_{A_{\mathrm{L}}^r}(\mathbf{v})}, \tag{8}$$

$$\overline{\overline{z}_j} = \frac{\sum\limits_{\substack{r=1 \\ r\colon \overline{z}_j^r=1}}^{N} \mu_{A_{\mathrm{U}}^r}(\mathbf{v})}{\sum\limits_{r=1}^{N} \mu_{A_{\mathrm{U}}^r}(\mathbf{v})}. \tag{9}$$

Fuzzy sets A_{L}^r and A_{U}^r represent the lower or upper approximations of sets A^r depending on the considered part of the structure and the contents of the consequent part of the rule, according to the following assignments

$$A_{\mathrm{L}}^r = \begin{cases} \underline{\widetilde{P}}A^r & \text{gdy } \overline{z}_j^r = 1 \\ \overline{\widetilde{P}}A^r & \text{gdy } \overline{z}_j^r = 0. \end{cases} \tag{10}$$

and

$$A_{\mathrm{U}}^r = \begin{cases} \overline{\widetilde{P}}A^r & \text{gdy } \overline{z}_j^r = 1 \\ \underline{\widetilde{P}}A^r & \text{gdy } \overline{z}_j^r = 0. \end{cases} \tag{11}$$

Assignments (10) and (11) have been proved in [14, 15]. In Fig. 1 v_i is a value of conditional attribute c_i, $\underline{\overline{v_i}}$ is the set of all possible values of conditional attribute c_i, usually the interval, A_i^r

is a fuzzy set defined in antecedent part of r-th rule according to conditional attribute c_i. The outputs of the blocks depicted by A_i^r, $\sup A_i^r$, and $\inf A_i^r$ are the membership values $\mu_{A_i^r}(v_i)$ or $\sup \mu_{A_i^r}(v_i)$, or $\inf \mu_{A_i^r}(v_i)$, respectively, and x is a cartesian product realised by any T-norm. z_j^r are centres of fuzzy sets B_j^r defined in consequent part of the r-th rule, according to the j-th decision attribute. Σ depicts the simple summing block, and $\div$ is a block that divides the value from the upper input by the value from the lower input.

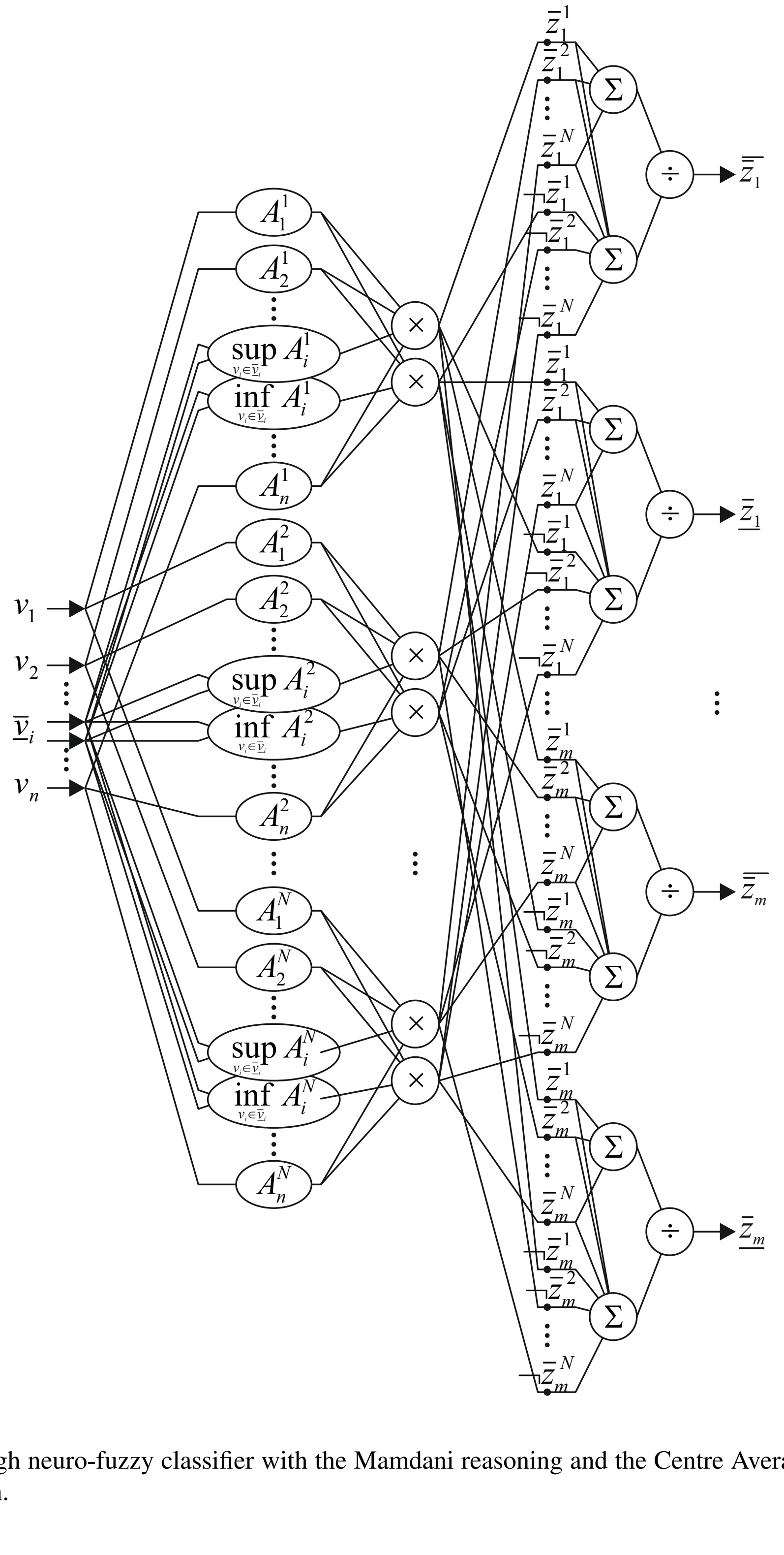

Fig. 1. Rough neuro-fuzzy classifier with the Mamdani reasoning and the Centre Average defuzzification.

5. Results

As aforementioned, we used the LEM-2 rule induction algorithm [5, 6] to create a rule base for the rough-fuzzy classifier. The generation of rules is performed in two phases. The first phase involves data quantization. The introduced set is treated as a decision table under the assumption of being deterministic (consistent), where the variability space of individual numerical attributes is divided into as few intervals as possible, yet such that after replacing the values in the decision table with labels of these intervals, the table remains deterministic. Such processed data can be directly used for rule generation. In essence, the process of generating the rule base can be described as reducing the decision table resulting from quantization. It is worth emphasizing that the method employed guarantees that all samples occurring in the dataset used during rule generation will be classified correctly. Such guarantees are not provided by gradient and evolutionary methods. The algorithm created about four thousand rules in the form of intervals, and 0.6 of the interval width was taken as a width of the Gaussian fuzzy sets in the classifier fuzzy rules. The algorithm was set to use 90 % of the most meaningful objects for the dataset. Our experience showed this ratio is a good trade-off keeping the rule base smaller and getting rid of outliers. The information about the generated rules is presented in Table 2.

Table 2. Numbers of generated rules for each fold of 10-fold data cross-validation. All the experiments in the paper were performed for each generated fold (i.e. ten times) and averaged.

Fold #	No of rules	No of conditions	Avg. no of conditions per rule
1	2848	22228	7.80
2	2832	22054	7.79
3	2814	21939	7.80
4	2773	21649	7.81
5	2900	22512	7.76
6	2796	21842	7.81
7	2974	23738	7.98
8	3519	26960	7.66
9	2865	22349	7.80
10	2790	21677	7.77
average	2911	22694	7.80

After preparing the rough-fuzzy classifier by LEM-2 we introduced and computed for every combination of data, averaged coefficient of significance of the attributes C'

$$\overline{\sigma}_{(C,D)}\left(C',k\right) = \frac{\overline{\gamma_{C,C',k}}\left(D^*\right) - \overline{\gamma'_{C,C',k-k'}}\left(D^*\right)}{\overline{\gamma_{C,C',k}}\left(D^*\right)} \tag{12}$$

where

$$\overline{\gamma_{(C,C',k)}}\left(D^*\right) = \frac{\sum_{C"\subseteq C:\, C'\subseteq C"\wedge\overline{\overline{C"}}=k}\gamma_{C"}\left(D^*\right)}{\overline{\overline{\left\{C"\subseteq C\colon C'\subseteq C"\wedge\overline{\overline{C"}}=k\right\}}}} \tag{13}$$

$$\overline{\gamma'_{(C,C',k)}}\left(D^*\right) = \frac{\sum_{C"\subseteq C:\, C'\cap C"=\emptyset\wedge\overline{\overline{C"}}=k}\gamma_{C}\left(D^*\right)}{\overline{\overline{\left\{C"\subseteq C\colon C'\cap C"=\emptyset\wedge\overline{\overline{C"}}=k\right\}}}} \tag{14}$$

$$\overline{\overline{\left\{C"\subseteq C\colon C'\subseteq C"\wedge\overline{\overline{C"}}=k\right\}}} = C_{n-k'}^{k-k'} = \left(\begin{array}{c} n-k' \\ k-k' \end{array}\right) = \frac{(n-k')!}{(k-k')!(n-k)!} \tag{15}$$

$$\overline{\overline{\left\{C"\subseteq C\colon C'\cap C"=\emptyset\wedge\overline{\overline{C"}}=k\right\}}} = C_{n-k'}^{k} = \left(\begin{array}{c} n-k' \\ k \end{array}\right) = \frac{(n-k')!}{(k)!(n-k'-k)!} \tag{16}$$

where $n = \overline{\overline{C}}$ and $k' = \overline{\overline{C'}}$. In our work $k' = 1$. In (15) we choose set C" from C, however a priory $C' \subseteq C$". Thus we choose $\overline{\overline{C"}} - \overline{\overline{C'}} = k - k'$ elementary set from $\overline{\overline{C}} - \overline{\overline{C'}} = n - k'$ elementary set. The number of possible choices is a number of combinations without repetitions $C_{n-k'}^{k-k'}$. In (16) we choose set C" from $C \setminus C'$, due to condition $C' \cap C" = \emptyset$. Thus we choose $\overline{\overline{C"}} = k$ elementary set from $\overline{\overline{C}} - \overline{\overline{C'}} = n - k'$ elementary set. The number of possible choices is a number of combinations without repetitions $C_{n-k'}^{k}$.

The Average normalized feature significance coefficients (ANIC) for the training and testing datasets are presented in Figure 2 and 3. Each plot shows average features importance when certain number of features is missing, apart from removing the feature denoted on a plot. For example, we can find out what the average loss of accuracy is when a variable, for example Creditscore, was available, and what it is when we removed that variable (e.g. Creditscore). The average accuracy in % (OK), do not know decision (NO), and wrong answers (ERR) for the testing dataset (ten folds) is shown in Fig. 4. Each plot shows average predicted decisions for all combinations of removed features when certain number of features is missing, apart from removing the feature denoted on a plot.

6. Conclusions

The application of Rough Neuro-Fuzzy Classifiers, coupled with the LEM-2 rule induction algorithm, has demonstrated promising results in predicting bank customer churn. This approach offers a robust framework for effectively handling the complexities and uncertainties inherent in churn prediction tasks. The most important advantage of the proposed system is that it can handle missing feature values. There is no need to introduce any data imputation method as the system can still classify objects even when some input values are missing.

All the experiments were performed with ten-fold cross-validation, that is we built ten classifiers and the results are averaged across all ten folds. Through the utilization of the LEM-2 rule induction algorithm, we have successfully generated insightful rules for the classifier, providing valuable interpretability to the predictive model. These rules serve as actionable insights for decision-makers within the banking industry, facilitating proactive churn management strategies. The performance could be further improved using e.g. gradient learning.

By computing the importance of input features in two ways: directly from the data and from the classifier, we have gained deeper insights into the relevance and contribution of each feature towards predicting customer churn, enhancing the overall efficacy of the predictive model. The direct computation of feature importance from both data and classifier outputs provides a comprehensive understanding of the underlying factors driving customer churn. This dual assessment approach not only improves the interpretability of the predictive model but also enables informed decision-making processes tailored to mitigate churn risks effectively.

References

[1] *Bank Customer Dataset*. `https://www.kaggle.com/datasets/adammaus/predicting-churn-for-bank-customers`. Accessed April 2, 2024. 2024.

[2] Bin-Nashwan, S. A. and Hassan, H.: Impact of customer relationship management (CRM) on customer satisfaction and loyalty: A systematic review. In: *Journal of Advanced Research in Business and Management Studies* 6.1 (2017), pp. 86–107.

[3] Cios, K. J., Pedrycz, W., and Swiniarski, R. W.: Data mining methods for knowledge discovery. Vol. 458. Springer Science & Business Media, 2012.

[4] Fayyad, U. M.: *Data mining and Knowledge discovery in databases: Applications in Astronomy and Planetary Science*. Tech. rep. American Association for Artificial Intelligence, Menlo Park, CA (United States), 1996.

[5] Gryzmala-Busse, J.: An overview of the LERS1 learning system. In: *Proceedings of the 2nd International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*. 1989, pp. 838–844.

[6] Grzymala-Busse, J. W.: LERS-a system for learning from examples based on rough sets. In: *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory* (1992), pp. 3–18.

[7] Han, J., Pei, J., and Tong, H.: Data mining: concepts and techniques. Morgan kaufmann, 2022.

[8] Keramati, A., Ghaneei, H., and Mirmohammadi, S. M.: Developing a prediction model for customer churn from electronic banking services using data mining. In: *Financial Innovation* 2 (2016), pp. 1–13.

[9] Kim, M.-K., Park, M.-C., and Jeong, D.-H.: The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. In: *Telecommunications policy* 28.2 (2004), pp. 145–159.

[10] L. X. Wang: Adaptive Fuzzy Systems and Control. Englewood Cliffs: PTR Prentice Hall, 1994.

[11] Lenard, M. J., Madey, G. R., and Alam, P.: The design and validation of a hybrid information system for the auditor's going concern decision. In: *Journal of Management Information Systems* 14.4 (1998), pp. 219–237.

[12] Linoff, G. S. and Berry, M. J.: Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2011.

[13] Nie, G., Rowe, W., Zhang, L., Tian, Y., and Shi, Y.: Credit card churn forecasting by logistic regression and decision tree. In: *Expert Systems with Applications* 38.12 (2011), pp. 15273–15285.

[14] Nowicki, R.: Rough–neuro–fuzzy structures for classification with missing data. In: *IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics* 39.6 (2009), pp. 1334–1347.

[15] Nowicki, R.: On Combining Neuro–Fuzzy Architectures with the Rough Set Theory to Solve Classification Problems with Incomplete Data. In: *IEEE Trans. on Knowledge and Data Engineering* 20.9 (Sept. 2008), pp. 1239–1253.

[16] Pawlak, Z.: Rough Sets. In: *International Journal of Computer and Information Sciences* 11.5 (1982), pp. 341–356.

[17] Risselada, H., Verhoef, P. C., and Bijmolt, T. H.: Staying power of churn prediction models. In: *Journal of Interactive Marketing* 24.3 (2010), pp. 198–208.

[18] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., and Chatzisavvas, K. C.: A comparison of machine learning techniques for customer churn prediction. In: *Simulation Modelling Practice and Theory* 55 (2015), pp. 1–9.

[19] Van den Poel, D. and Lariviere, B.: Customer attrition analysis for financial services using proportional hazard models. In: *European journal of operational research* 157.1 (2004), pp. 196–217.

[20] Wadikar, D.: Customer Churn Prediction. In: (2020).

[21] Xia, G.-e. and Jin, W.-d.: Model of customer churn prediction on support vector machine. In: *Systems Engineering-Theory & Practice* 28.1 (2008), pp. 71–77.

[22] Zadeh, L. A.: Fuzzy sets. In: *Information and control* 8.3 (1965), pp. 338–353.

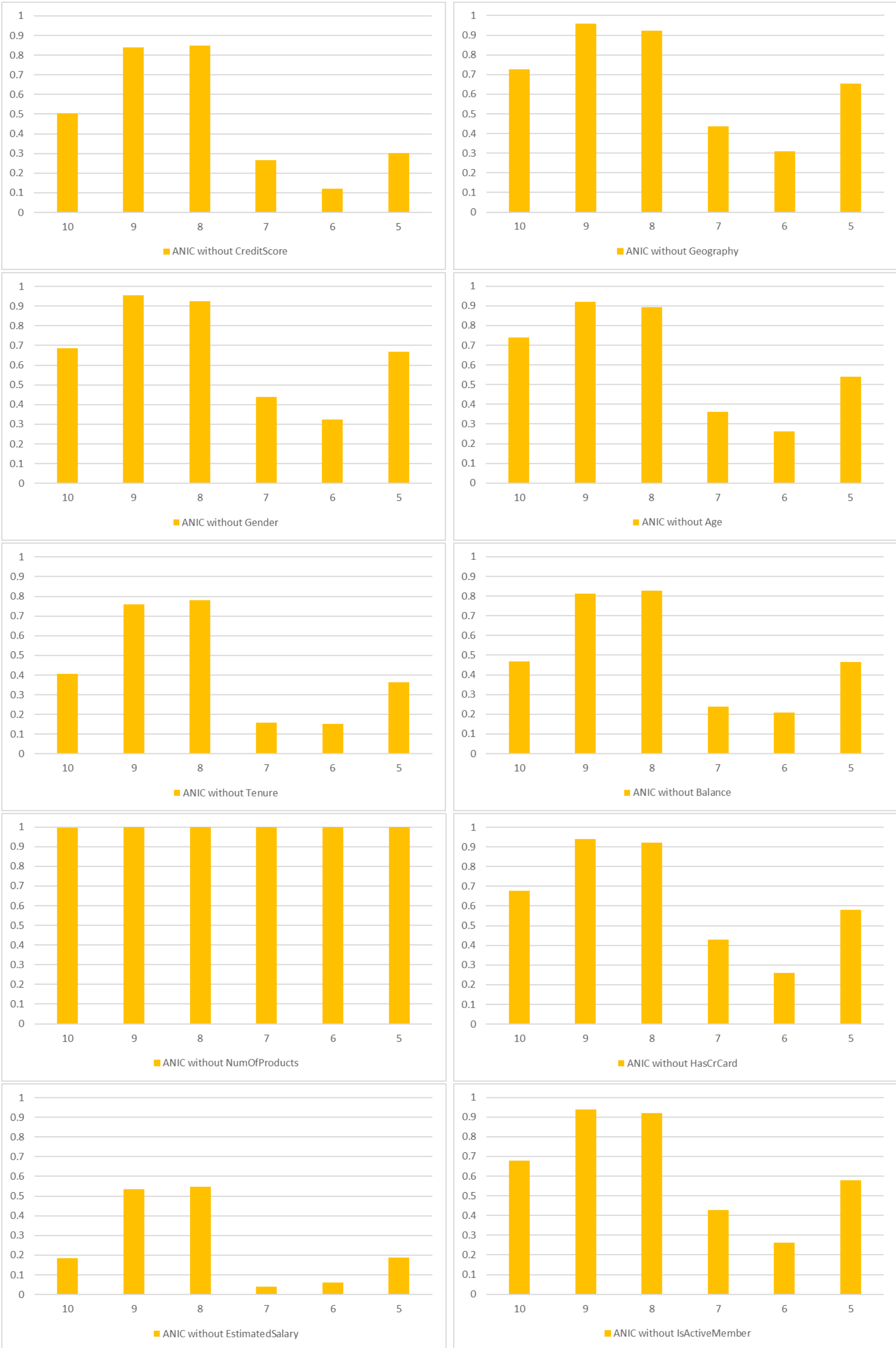

Fig. 2. Average normalized feature importance coefficient (ANIC) for the training dataset. Each plot shows average features importance when certain number of features is missing, apart from removing the feature denoted on a plot.

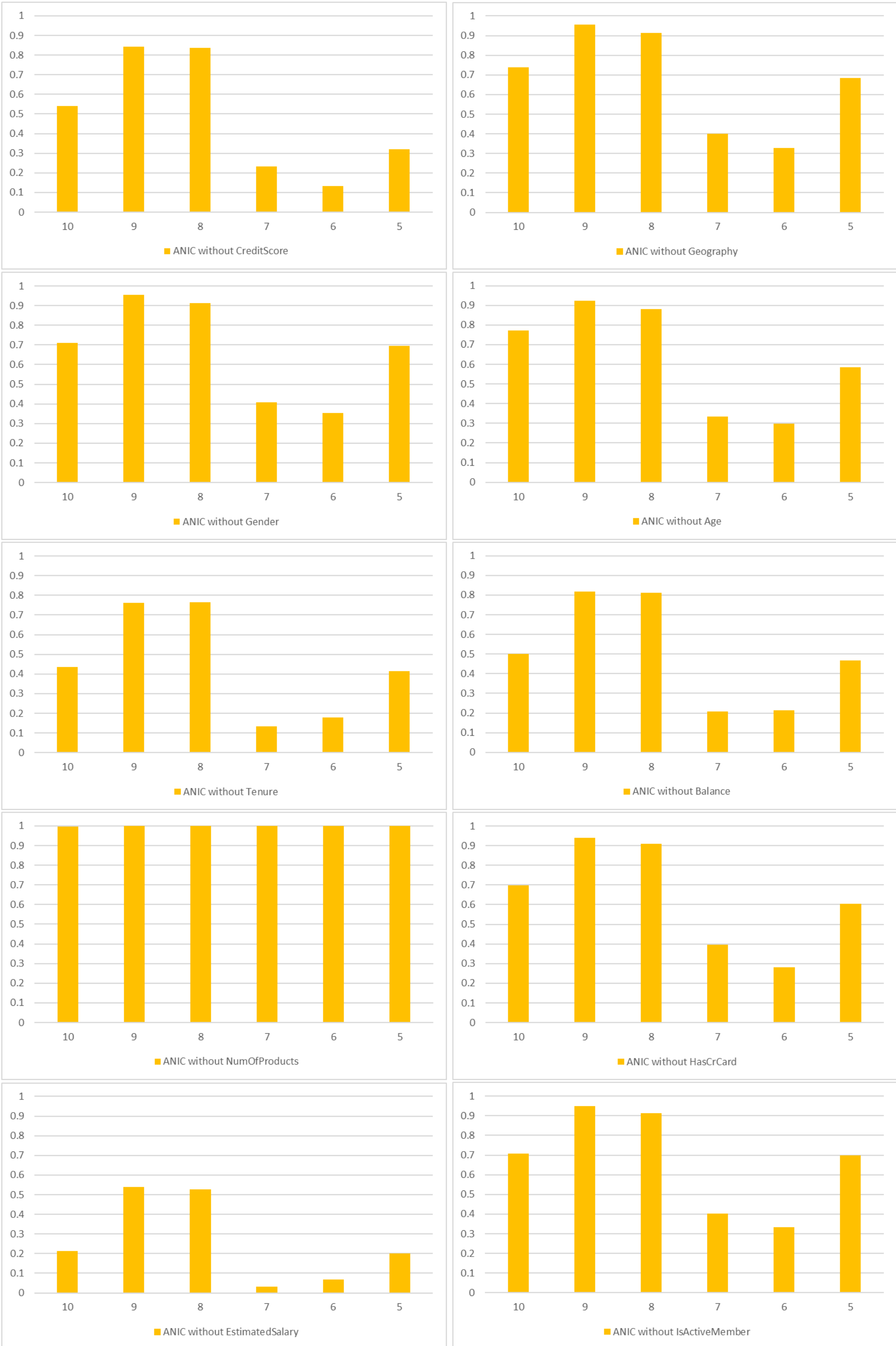

Fig. 3. Average normalized feature importance coefficient (ANIC) for the testing dataset. Each plot shows average features importance when certain number of features is missing, apart from removing the feature denoted on a plot.

Fig. 4. Average accuracy in % (OK), do not know decision (NO), and wrong answers (ERR) for the testing dataset. Each plot shows average predicted decisions for all combinations of removed features when certain number of features is missing, apart from removing the feature denoted on a plot.