

# Explainability of Highly Associated Fuzzy Churn Patterns

D.Y.C. Wang<sup>1</sup>, Lars Arne Jordanger<sup>1</sup>, and Jerry Chun-Wei Lin<sup>1</sup>

Western Norway University of Applied Sciences, Bergen, Norway  
ycw@hvl.no; Lars.Arne.Jordanger@hvl.no; jerrylin@ieee.org

**Abstract.** Customer churn, particularly in the telecommunications sector, has a profound impact on costs and profits. As the explainability of models becomes increasingly important, this study emphasizes not only the explainability of customer churn through machine learning models, but also the importance of identifying multivariate patterns and setting soft bounds for intuitive interpretation. The main objective is to use a machine learning model and fuzzy-set theory with top-k HUIM to identify highly associated patterns of customer churn with intuitive identification, referred to as Highly Associated Fuzzy Churn Patterns (HAFCP). Additionally, this method aids in uncovering association rules among multiple features across low, medium, and high distributions. Such discoveries are instrumental in enhancing the explainability of findings. Experiments show that when the top-5 HAFCPs are included in four datasets, a mixture of performance results is observed, with some showing notable improvements. It becomes clear that high importance features enhance explanatory power through their distribution and patterns associated with other features. As a result, the study introduces an innovative approach that improves the explainability and effectiveness of customer churn prediction models.

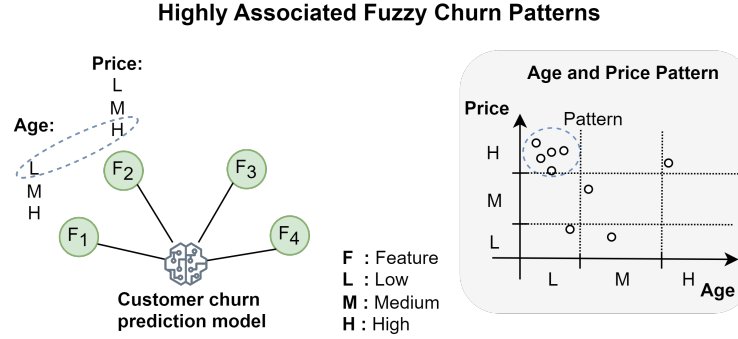
**Keywords:** Customer Churn Prediction · Top-K High Utility Itemset Mining · Fuzzy-set Theory

## 1 Introduction

Customer churn refers to consumers switching services, often due to dissatisfaction, better options or personal preference [1]. Fluctuations in churn rates can have a major impact on costs and profits. An example from the telecom sector: while churn rates are typically between 20% and 40%, it costs more than 5 - 6 times more to retain customers, and a mere 5% reduction can lead to a 25% increase in profits [1]. Customer churn prediction (CCP) not only determines whether a customer is likely to churn and what the reasons are, but also helps in implementing customer retention strategies to mitigate the economic loss. Nowadays, it is more important to understand how the models explain the results of churn prediction.

For the interpretation of CCP, it is essential to understand the complexity of the numerous influencing variables. These variables not only function independently of each other, but also exhibit interrelated patterns between different distributions of characteristics. However, relying only on the explainability of machine learning (ML) models may not fully uncover these nuanced associations, especially in the context of association rule mining. Furthermore, predictive models work primarily with numerical variables that lack definitive boundaries for intuitive interpretation, which presents a particular challenge. For example, the lack of a clear threshold for categorizing a customer’s call duration as high or low leads to ambiguity. Accordingly, there are two main needs in this study beyond pure prediction and explainability. First, to classify data by defining soft boundaries rather than using traditional unambiguous thresholds to provide a clearer and more actionable perspective for stakeholders, and second, to identify patterns that reveal underlying associations.

The main goal is to develop an integrative framework that connects ML models with top-k high utility itemset mining (HUIM). In a previous study, a churn prediction pattern (CPP) [22] was defined as a set of features that can influence several definable indicators for customer churn evaluation. This study, the extracted patterns are defined as Highly Associated Fuzzy Churn Patterns (HAFCP), which are conceptually illustrated in Fig. 1. This aims not only to provide revealing insights into the underlying data, but also to increase the predictive capability of the model. A study using a similar concept of HUIM and statistical learning has already been proposed in the field of inductive logic programming (ILP) [18], but while their focus was primarily on rule discovery and they used quite different datasets than this study, our intention is to improve the explainability and strengthen the predictive capabilities of the model.



**Fig. 1.** An illustrative example of highly associated fuzzy churn patterns. The figure simply illustrates the concept of HAFCP. In the diagram, 'L', 'M', and 'H' represent low, medium, and high respectively, which are calculated by the membership function in fuzzy-set theory.

## 2 Related Work

### 2.1 Explainability of SHAP Value

The latest developments [17] in eXplainable AI (XAI) are gradually making ML and artificial intelligence (AI) transparent. In particular, model-agnostic methods such as SHapley Additive exPlanations (SHAP) facilitate sophisticated modeling, understanding and representation of complex events and systems [11]. A unified approach that uses SHAP and feature differences to explain ML predictions based on game theory [14]. The overall importance of a feature in a model can be calculated by taking the mean of the absolute SHAP values for that feature in all instances. This gives a general indication of the importance of a particular feature in terms of its impact on the output of the model.

### 2.2 Fuzzy-set Theory and Fuzzy Pattern Mining

In contrast to traditional classification, where data is categorized into specific sets, fuzzy-set theory assigns data based on membership functions that determine the degree to which each piece of data belongs to a particular set [10], [24]. Common methods include the Gaussian fuzzy function and the triangular fuzzy function for classification. For data that follows a normal distribution, the Gaussian fuzzy function is usually preferred [25]. Leveraging the principles of fuzzy set theory [8] in the context of utility mining enables the development of a data-driven model that is not only intuitive for human understanding but improves efficiency in mining operations. Wang et al. first attempt to discover high utility patterns from quantitative database [21]. Then, Gan et al. proposed an explainable fuzzy utility mining (FUM) for fuzzy high-utility sequence patterns [5].

### 2.3 Top-K High Utility Itemset Mining

High Utility Itemset Mining (HUIM) is an essential topic in data mining that aims to discover itemsets that bring high profit or utility in transactional databases [4]. Traditional HUIM focuses on finding all itemsets that exceed a user-defined utility threshold. However, setting an appropriate utility threshold is challenging, which led to the emergence of the top-k model for mining high utility itemsets [20]. Traditional utility mining can be overwhelming due to the generation of numerous itemsets, while the top-k approach selectively highlights the most relevant k-itemsets [12]. By focusing on high utility itemsets using the top-k model, users can gain insights into highly profitable or important items and thus improve their decision-making processes accordingly.

## 3 The Developed HAFCP Model

In this section, this study describes the analytical framework for identifying strongly associated churn patterns with fuzzy logic using predictive feature importance, fuzzy-set theory, and top-k high-utility itemset mining.

### 3.1 Problem Statement

This paper focuses exclusively on the binary classification problem for predictive modeling in the methodology section. Starting from a labeled customer churn dataset, this study aims to uncover associated patterns from categorical and numerical features by constructing a predictive model using fuzzy-set theory [10] and leveraging top-k HUIM [12]. To provide a more intuitive understanding and explanation of the data, we convert numerical features into categorical levels such as high, medium and low using fuzzy-set theory [10]. This conversion facilitates the use of the HUIM approach by transforming the original numerical data into corresponding frequencies of low, medium and high linguistic terms, thereby improving the explainability and relevance of the patterns found. Furthermore, we intend to use the global feature importance — a measure that quantifies the overall contribution of each feature to the ML model performance. Then, the HAFCPs are identified using the proposed method as the final output. Here is the mathematical definition of HAFCP:

**Definition 1** (*Highly Associated Fuzzy Churn Pattern (HAFCP)*)

Consider a set of features  $V$ , where each numerical feature  $v_i \in V$  can be mapped to a linguistic term such as Low ( $L$ ), Medium ( $M$ ), or High ( $H$ ). Let  $B(v_i)$  represent the application of the maximum cardinality criterion [6] to the membership degree of the numerical feature  $v_i$ . Given a utility function  $U$  and a pattern  $P$ , with  $P \subseteq V$ , we can rank all patterns based on their computed values. The top- $k$  high utility patterns are then generated. HAFCP is defined as:

$$\begin{aligned} \text{HAFCP}(K) = \{P_1, P_2, \dots, P_K \mid P_i = U^{(i)} \wedge \\ \forall v_j \in P_i, B(v_j) \in \{L, M, H\}\}, \end{aligned} \quad (1)$$

where  $U^{(i)}$  represents the  $i^{\text{th}}$  highest value pattern,  $P_1$  is the top-1 high utility pattern,  $P_2$  the top-2 high utility pattern, and so forth. The condition  $B(v_j) \in \{L, M, H\}$  ensures that variables in each pattern are distinctly categorized, e.g., as “Low”, “Medium”, or “High” based on the defined membership functions.

Using a customer dataset with features like ‘Age’ and ‘Price’, we aim to identify patterns linked to customer churn. Fuzzy-set theory categorizes ‘Age’ and ‘Price’ into ‘Low’, ‘Medium’, and ‘High’, where ‘Age’ represents different age groups and ‘Price’ different spending levels. With HUIM, we uncover patterns like  $\{\text{‘Age.L’}, \text{‘Price.H’}\}$ , suggesting that younger customers making high-priced purchases are more likely to churn.

### 3.2 Analytical Framework for HAFCP Mining

In this section, we explain the methodology for identifying HAFCP using fuzzy-set theory and feature score evaluation. This approach utilizes the insights and knowledge presented in the previous section to provide a comprehensive analytical framework. Fig. 2 shows the comprehensive structure of the proposed framework and the links between each phase.

**A. Feature Importance Calculation** The process begins with a specific dataset on customer churn with the label churn or do not churn. Once this dataset has been captured, encoding of the labels for potentially categorical data is required to convert it into a machine-readable format. After data conversion, the dataset is divided into two parts: 80% is assigned to the training set, while the remaining 20% is assigned to the testing set. In Fig. 2 part 1 outlines a series of steps for model construction and evaluation. The process begins with model training using the XGBoost [2], a gradient boosting framework, to train the model using the specified training data. After training, the model is applied to the testing dataset and its performance is evaluated using various metrics such as accuracy, recall rate, precision, etc. At the same time, the importance of the features of the model is calculated using various techniques. Particular attention is paid to SHAP values, which are a standardized measure of feature importance.

**B. Data Transformation** The second phase is primarily concerned with converting numerical data columns into a more interpretable fuzzy dataset, which is then translated into linguistic terms based on fuzzy set theory. The process begins with the identification of the numerical data columns, which are subjected to a normality test called "Shapiro-Wilk test" [19]. Depending on the result of the normality test, the data distributions are classified as either Gaussian or non-Gaussian. For data with a Gaussian distribution, a Gaussian fuzzy membership function is used [10]. Conversely, a triangular fuzzy membership function [10] is used for data with a non-Gaussian distribution due to simplicity and clearly defined range. Specifically, these numerical features are translated into three linguistic variables, including High, Medium and Low with their own memberships. The culmination of this phase is a fuzzy dataset that categorically represents the originally identified numerical data in the form of these linguistic variables. Subsequently, one-hot encoding [7] is inevitably used to implement the data for mining. One-hot encoding is a process in which categorical data is converted into a binary matrix, with each category represented by a unique binary column. After one-hot encoding, the data columns can be left with defuzzified binary values. This study discusses triangular membership functions in fuzzy set theory, similar in concept to Gaussian membership functions, as referenced in [10]. These functions are essential in determining the membership degree of elements in a fuzzy set. We define a fuzzy set  $A$  on a universe of discourse  $X$ . The triangular membership function of  $A$ , denoted  $\mu_A(x)$ , uses three parameters  $a$ ,  $b$ , and  $c$  to represent the vertices of the triangle. It is defined as:

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x \leq b \\ \frac{c-x}{c-b} & \text{if } b < x < c \\ 0 & \text{if } x \geq c \end{cases} \quad (2)$$

where  $\frac{x-a}{b-a}$  represents the ascending side of the triangle (low to medium),  $\frac{c-x}{c-b}$  the descending side (medium to high), and 0 for values outside  $[a, c]$ . Using  $\mu_A(x)$ , we categorize elements into low, medium, or high classes. The linguistic term for a given  $x$  is determined by:

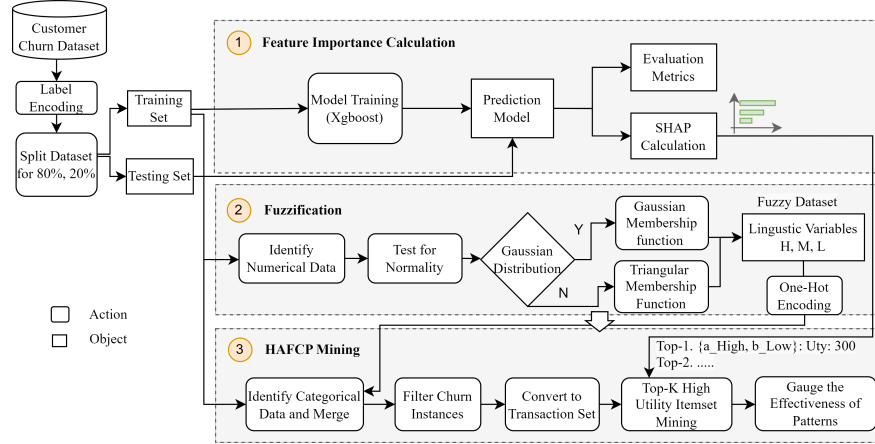
$$\text{Linguistic Term}(x) = \begin{cases} L & \text{if } \mu_A(x) \text{ is maximum in the low set} \\ M & \text{if } \mu_A(x) \text{ is maximum in the medium set} \\ H & \text{if } \mu_A(x) \text{ is maximum in the high set,} \end{cases} \quad (3)$$

where  $\mu_A(x)$  represents a membership function used in this paper.

**C. Top-K High Utility Itemset Mining** In the third phase, the process focuses mainly on extracting meaningful patterns from the quantitative data. The process begins with the identification of original categorical features within the customer churn dataset. Following this identification, the categorical data is then merged with the previously created fuzzy dataset to create a consolidated dataset. Since our goal is to find churn patterns, we filter out the data that contains churn instances. These filtered data instances are then converted into a frequency itemset format, a standard format that is compatible with mining algorithms. At the end of this phase is the implementation of the top-k HUIM [20] process. Specifically, the top-k HUIM algorithm begins by calculating the utility of each itemset in the dataset. A single feature of the data belongs to one of the categories low, medium, or high, we will treat the quantitative value of the feature as 1; if not, it will be treated as 0. The unit of profit is determined using the corresponding feature importance. The utility of an itemset is usually defined as the sum of the products of the individual utilities of the items and their respective quantities in the transactions. This calculation reflects both the importance and the frequency of the items within the dataset. After calculating the utility values, the algorithm creates a ranking list of these item groups based on their utility values. From this ranking, the top-k itemsets with the highest utility values are selected. From now on, we refer to the patterns identified using top-k HUIM as HAFCPs.

**D. Pattern Evaluation and Model Enhancement** In the final stage, before we move on to explaining the patterns obtained from the mining process, it is crucial to understand how we measure their meaning and effectiveness. This step is about understanding the key metrics to evaluate the quality and relevance of the identified patterns. With the utility values calculated, HAFCP can examine the identified patterns in more detail, particularly those associated with churn events. The aim is not only to identify which patterns occur frequently, but also to understand their relationship and significance in relation to churn events. Unlike traditional ML explainability tools, we can intuitively use labels such as low, medium and high, which are based on membership functions and allow us to clearly evaluate these combination patterns. Once we have understood

and explored the patterns associated with churn, it is important to evaluate how these patterns and insights can improve the predictive model. To do this, we need to repeat the feature engineering phase of the model to integrate the insights from the patterns. Essentially, the aim here is to determine whether the inclusion of these patterns as features improves the performance of the model or provides a more refined understanding of churn events. In this study, we utilize the identified top-k patterns to facilitate feature development. These patterns are converted into a new feature for use in ML models. This approach is also the reason why we exclusively use the training set for utility mining. In this study, we create a new feature that characterizes the presence of a particular HAFCP. In our experiments, we will observe whether these newly introduced features can effectively improve the model performance.



**Fig. 2.** The proposed framework for mining HAFCPs: This comprehensive framework outlines the process of extracting HAFCPs, detailing the methodology from initiation to conclusion. To facilitate understanding, the extension work will present a straightforward algorithmic example. This example will elaborate each step of the process, providing a clear explanation of how HAFCPs are identified and derived within our framework.

## 4 Case Study and Discussion

In this section, we first introduce the datasets and then explain the analytical framework tailored to the [1], [3], [13], [16] used in this study and discuss the evaluation measures. At the end of this section, several experiments are then conducted to evaluate the performance of the proposed approach and the compared models. In this study, AUC, precision, accuracy, F1 score, and recall rate will be employed as metrics for measuring model prediction performance [9]. We

use four publicly available datasets [1], [3], [13], [16] from the telecommunications sector. The target class denoting the churned customer is labeled "True" in these datasets. These datasets provide information about customers' telephony usage, including call durations, charges, frequency of different types of calls, customers' geographical location, account tenure, subscription plans, and interactions with customer service. Dataset 1 contains 3,333 entries with 15 numerical attributes and 5 categorical attributes. Dataset 2 contains 5,000 entries with the same attributes as the first dataset. Dataset 3 contains 3,150 instances with 13 attributes. Dataset 4 contains 51,047 labeled data entries with 57 attributes.

We first used the XGBoost [2] ML model with the parameter settings as:  $max\_depth = 6$ ,  $learning\_rate = 0.3$ , and  $n\_estimators = 100$ . The performance metrics of the baseline model for datasets 1-4 are shown in Table 1. We then extract 12 numerical columns from the dataset. For datasets 1 and 2, four features relating to the "Total Charge" are filtered out directly to save computational effort, as the importance of the feature is zero. In addition, features with low importance and several categories, namely "STATE" and "AREA", are eliminated. After performing normality tests on these columns, Gaussian membership functions [10] (for columns with normal distribution) and triangular membership functions [10] (for columns without normal distribution) are applied to categorize each column based on the predefined membership functions. This process of fuzzification is followed by one-hot encoding of the fuzzy classification columns. While the same logic and methodology is applied to datasets 3 and 4, the specifics of the data differ, so the membership functions must be adjusted to accurately reflect the unique distributions in these additional datasets.

From the above results, we have already obtained a dataset that combines categorical data with a fuzzy classification dataset. Now we have a table of frequent itemsets and unit profits. Algorithm 1 provides a detailed explanation of how HAFCP is processed, computed, and generated. The results of the top-k HUIM on the datasets show the mining results of datasets that consider the maximum cardinality as a binary value for fuzzy columns. In this context, the suffix "\_L" after a column name stands for the linguistic expression of "low" after fuzzy classification, "\_M" stands for "medium" and "\_H" for "high". For instance, in dataset 1, the top-1 HAFCP found was {'TTL DAY MIN\_L', 'TTL INTL MIN\_M'}.

Finally, we extract the association rules of the top-5 patterns to see if they can improve the performance of the model in these datasets. In this experiment, we tested by generating the binary column resulting from multiplying different features into the dataset and then training the dataset with the same parameter settings. By including the top-5 HAFCP separately, the models for both datasets underwent subtle but notable changes, as shown in Table 1. It shows the evaluation results from two previous studies [15,23] (one of which is the baseline) and compares them with the results obtained by including the top-5 HAFCP methods.

The extended evaluation over four datasets shows the robustness and adaptability of our model when integrating HAFCP. Remarkably, a detailed review

---

**Algorithm 1** HAFCP Mining
 

---

```

1: Input: Dataset  $D$  with customer features and churn labels, Machine learning model
    $M$ , SHAP Explainer  $E$ 
2: Output: Top- $k$  highly associated churn patterns  $P$ 
3: procedure MINE HFACPs( $D, M, E$ )
4:    $D \leftarrow$  Read dataset
5:    $C \leftarrow$  Factorize categorical features in  $D$ 
6:    $X, y \leftarrow$  Split  $D$  into features and target
7:   Initialize  $M$ 
8:   Train  $M$  on  $(X, y)$ 
9:    $\hat{y} \leftarrow$  Predict churn using  $M$ 
10:  SHAP values  $\leftarrow E(M, X)$ 
11:   $\bar{S} \leftarrow$  Calculate mean absolute SHAP values for  $X$ 
12:   $T \leftarrow$  Form profit table from  $\bar{S}$ 
13:  for each  $x_i \in X$  do
14:    normality  $\leftarrow$  Test normality of  $x_i$ 
15:     $\mu \leftarrow \begin{cases} \mu_T & \text{if normality is false} \\ \mu_G & \text{otherwise} \end{cases}$ 
16:     $x'_i \leftarrow$  Fuzzy transform  $x_i$  using  $\mu$ 
17:     $x''_i \leftarrow$  Assign  $L, M$ , or  $H$  to  $x'_i$  based on max membership
18:    One-hot encode  $x''_i$ 
19:  end for
20:   $F \leftarrow$  Collect transformed features
21:   $F \leftarrow F \cup C$ 
22:  Define HUIM algorithm  $H$ 
23:   $P \leftarrow H(F, T, k)$ 
24:  return  $P$ 
25: end procedure
    
```

---

reveals that Dataset 1 experienced a substantial increase in AUC, especially noting a 1.5% lift from the top-5 patterns. Additionally, the recall metric saw significant improvement with the top-3 patterns, indicating more accurate positive predictions. Dataset 2 displayed modest but consistent enhancements in accuracy and precision, despite minimal fluctuations in AUC and recall, suggesting a nuanced but positive effect of HAFCP inclusion. The high stability of performance metrics in Dataset 3, along with marked improvements in recall and precision in Dataset 4—even from lower starting points—demonstrates the model’s consistency. With this approach, prediction performance is not only maintained but even improved in several datasets, confirming the integration of fuzzy churn patterns as a valuable strategy for customer churn prediction.

With regard to the explainability of the prediction model, the meaning of individual features was extended to the association rules of multiple features under different distributions by HAFCP, which are determined by the meaning of the features. This can help us to subsequently confirm the correlation or causality between features with high association. Fig. 3 shows a top-10 single feature visualization of the explainable models for four datasets. The associative patterns

**Table 1.** Comparison of performance metrics with adding top 1-5 highly associated fuzzy churn patterns for datasets 1-4. Results indicating improved performance compared to the XGboost [15] and RF [23] as the baseline models.

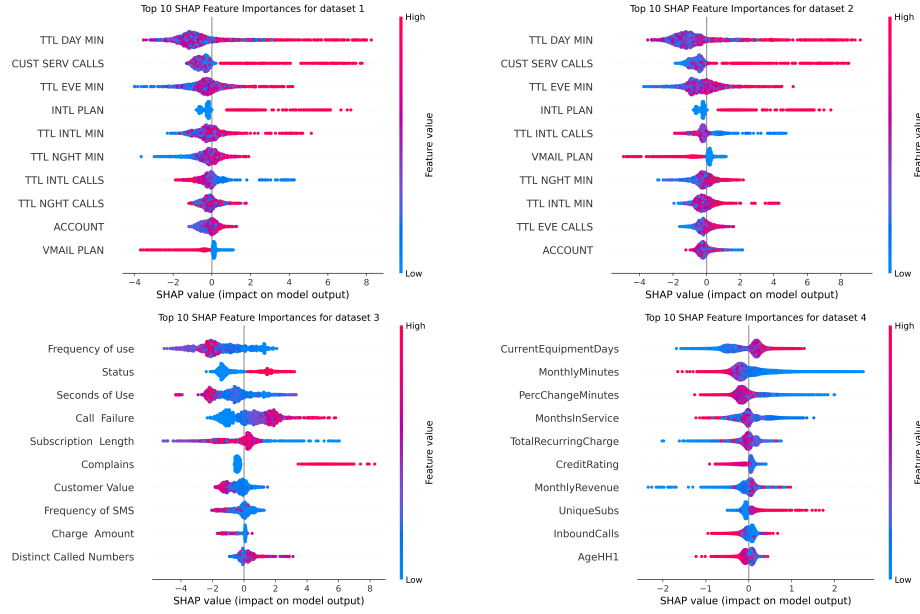
Performance Metrics									
Dataset	Metrics	RF	XGBoost	Top-1	Top-2	Top-3	Top-4	Top-5	AVG
1	AUC	91.3%	91.2%	<b>92.5%</b>	<b>93.1%</b>	<b>92.9%</b>	<b>92.9%</b>	<b>92.4%</b>	<b>92.8%</b>
	Accuracy	95.3%	95.7%	95.7%	95.7%	<b>96.0%</b>	95.7%	95.7%	<b>95.8%</b>
	Recall	71.9%	76.2%	76.2%	76.2%	<b>79.2%</b>	76.2%	76.2%	<b>76.8%</b>
	Precision	91.7%	93.9%	93.9%	93.9%	93.0%	93.9%	93.9%	93.7%
	F1 Score	80.4%	84.2%	84.2%	84.2%	<b>85.6%</b>	84.2%	84.2%	<b>84.5%</b>
2	AUC	89.9%	90.6%	90.6%	90.5%	<b>90.9%</b>	90.6%	<b>90.9%</b>	<b>90.7%</b>
	Accuracy	94.5%	95.6%	95.6%	95.4%	95.6%	<b>96.0%</b>	<b>95.7%</b>	<b>95.7%</b>
	Recall	78.9%	77.9%	77.9%	76.9%	<b>79.0%</b>	<b>80.0%</b>	<b>79.0%</b>	<b>78.6%</b>
	Precision	84.1%	92.1%	92.1%	92.0%	91.1%	<b>92.7%</b>	91.7%	<b>92.0%</b>
	F1 Score	81.5%	84.4%	84.4%	83.8%	<b>84.6%</b>	<b>86.0%</b>	<b>84.9%</b>	<b>84.7%</b>
3	AUC	97.6%	98.5%	<b>98.6%</b>	<b>98.6%</b>	<b>98.6%</b>	<b>98.6%</b>	<b>98.6%</b>	<b>98.6%</b>
	Accuracy	94.3%	96.2%	<b>96.7%</b>	<b>96.7%</b>	<b>96.7%</b>	<b>96.7%</b>	<b>96.7%</b>	<b>96.7%</b>
	Recall	85.5%	85.5%	<b>86.7%</b>	<b>86.7%</b>	<b>86.7%</b>	<b>86.7%</b>	<b>86.7%</b>	<b>86.7%</b>
	Precision	82.5%	85.5%	<b>87.8%</b>	<b>87.8%</b>	<b>87.8%</b>	<b>87.8%</b>	<b>87.8%</b>	<b>87.8%</b>
	F1 Score	83.9%	85.5%	<b>87.2%</b>	<b>87.2%</b>	<b>87.2%</b>	<b>87.2%</b>	<b>87.2%</b>	<b>87.2%</b>
4	AUC	62.9%	65.0%	<b>65.5%</b>	<b>65.3%</b>	65.0%	65.0%	<b>65.6%</b>	<b>65.3%</b>
	Accuracy	69.2%	64.0%	63.9%	<b>64.1%</b>	63.7%	63.7%	<b>64.2%</b>	63.9%
	Recall	18.3%	51.6%	<b>53.1%</b>	<b>52.5%</b>	<b>51.7%</b>	<b>51.7%</b>	<b>53.3%</b>	<b>52.5%</b>
	Precision	40.8%	39.4%	<b>39.9%</b>	<b>40.0%</b>	<b>39.5%</b>	<b>39.5%</b>	<b>40.4%</b>	<b>39.9%</b>
	F1 Score	25.3%	44.7%	<b>45.6%</b>	<b>45.4%</b>	<b>44.8%</b>	<b>44.8%</b>	<b>45.9%</b>	<b>45.3%</b>

among the key variables influencing model performance. When combined with the visual insights from Fig. 3, these patterns facilitate a more comprehensive and intuitive interpretation of the model’s behavior.

While these patterns seem to improve the performance of the model to a great extent, they can also have certain negative effects. So when incorporating new patterns to improve the performance of the model, it is important to carefully evaluate the pros and cons of the different metrics. Obtaining expert opinions on this topic would be of great benefit. In addition, it is essential to identify potential areas for refinement of this methodology. Categorical variables often play an important role in ML models. However, since high-utility mining focuses on frequency, certain categorical variables, even those with significant feature importance, could be sidelined if they are not sufficiently frequent. An example from this dataset is the variable “INTL PLAN”. Future efforts should address more efficient strategies for handling categorical variables in this framework.

## 5 Conclusion

To achieve a better understanding of customer churn models and uncover the associative patterns between different feature distributions, this paper presents



**Fig. 3.** Top-10 features of explainability in four datasets

a method for identifying strongly associated fuzzy churn patterns in tabular datasets with numerical values and provides a comprehensive guide and examples for its application. The method effectively utilizes fuzzy set theory and enables the transformation of numerical data into frequent item sets in a defensible manner. In addition, the associated patterns can provide insights for the interpretation of prediction models. As part of a case study, we used four publicly available customer churn datasets to validate the feasibility and pilot runs of our approach and its potential to improve model performance.

## References

1. Amin, A., Adnan, A., Anwar, S.: An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and naïve bayes. *Applied Soft Computing* **137** (2023)
2. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp. 785–794 (2016)
3. Diamantaras, K.: Customer churn prediction 2020 (2020), <https://kaggle.com/competitions/customer-churn-prediction-2020>
4. Fournier-Viger, P., Chun-Wei Lin, J., Truong-Chi, T., Nkambou, R.: A survey of high utility itemset mining. *High-Utility Pattern Mining: Theory, Algorithms and Applications* pp. 1–45 (2019)
5. Gan, W., Du, Z., Ding, W., Zhang, C., Chao, H.C.: Explainable fuzzy utility mining on sequences. *IEEE Transactions on Fuzzy Systems* **29**(12), 3620–3634 (2021)

6. Giacometti, A., Markhoff, B., Soulet, A.: Mining significant maximum cardinalities in knowledge bases. In: International Semantic Web Conference. pp. 182–199. Springer (2019)
7. Hancock, J.T., Khoshgoftaar, T.M.: Survey on categorical data for neural networks. *Journal of Big Data* **7**(1), 1–41 (2020)
8. Hong, T.P., Lin, K.Y., Wang, S.L.: Fuzzy data mining for interesting generalized association rules. *Fuzzy sets and systems* **138**(2), 255–269 (2003)
9. Hossin, M., M.N., S.: A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining and Knowledge Management Process* **5**, 1–11 (2015)
10. Jang, J.S.R., Sun, C.T., Mizutani, E.: Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [book review]. *IEEE Transactions on Automatic Control* **42**(10), 1482–1484 (1997)
11. Li, Z.: Extracting spatial effects from machine learning model using local interpretation method: An example of shap and xgboost. *Computers, Environment and Urban Systems* **96**, 101845 (2022)
12. Liu, M., Qu, J.: Mining high utility itemsets without candidate generation. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. pp. 55–64 (2012)
13. Liu, R., Ali, S., Bilal, S.F., Sakhawat, Z., Imran, A., Almuhaimeed, A., Alzahrani, A., Sun, G.: An intelligent hybrid scheme for customer churn prediction integrating clustering and classification algorithms. *Applied Sciences* **12**(18) (2022)
14. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* **30** (2017)
15. Pamina, J., Raja, B., SathyaBama, S., Sruthi, M., VJ, A., et al.: An effective classifier for predicting churn in telecommunication. *Journal of Advanced Research in Dynamical & Control Systems* **11** (2019)
16. Repository, U.M.L.: Iranian Churn Dataset. UCI Machine Learning Repository (2020)
17. Saeed, W., Omlin, C.: Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems* **263**, 110273 (2023)
18. Shakerin, F., Gupta, G.: Induction of non-monotonic rules from statistical learning models using high-utility itemset mining. *ArXiv Preprint ArXiv:1905.11226* (2019)
19. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). *Biometrika* **52**(3-4), 591–611 (1965)
20. Tseng, V.S., Wu, C.W., Fournier-Viger, P., Philip, S.Y.: Efficient algorithms for mining top-k high utility itemsets. *IEEE Transactions on Knowledge and Data Engineering* **28**(1), 54–67 (2015)
21. Wang, C.M., Chen, S.H., Huang, Y.F.: A fuzzy approach for mining high utility quantitative itemsets. In: IEEE International Conference on Fuzzy Systems. pp. 1909–1913 (2009). <https://doi.org/10.1109/FUZZY.2009.5277408>
22. Wang, D.Y., Jordanger, L.A., Lin, J.C.W.: Explainability of leverage points exploration for customer churn prediction. In: IEEE International Conference on Big Data. pp. 5997–6004 (2023)
23. Wu, S., Yau, W.C., Ong, T.S., Chong, S.C.: Integrated churn prediction and customer segmentation framework for telco business. *IEEE Access* **9**, 62118–62136 (2021)
24. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems* **1**(1), 3–28 (1978)
25. Zhang, Y., Qin, C.: A gaussian-shaped fuzzy inference system for multi-source fuzzy data. *Systems* **10**(6) (2022)