IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 27 January 2023, accepted 4 March 2023, date of publication 15 March 2023, date of current version 29 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3257352

RESEARCH ARTICLE

Development of a Customer Churn Model for Banking Industry Based on Hard and Soft Data Fusion

MASOUD ALIZADEH[1], DANIAL SADRIAN ZADEH[1], (Graduate Student Member, IEEE), BEHZAD MOSHIRI[1,2], (Senior Member, IEEE), AND ALLAHYAR MONTAZERI[3], (Member, IEEE)

[1]School of Electrical and Computer Engineering, University of Tehran, Tehran 1439957131, Iran
[2]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[3]School of Engineering, Lancaster University, LA1 4YW Lancaster, U.K.

Corresponding author: Allahyar Montazeri (a.montazeri@lancaster.ac.uk)

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/V027379/1.

ABSTRACT There has been an increase in customer churn over the past few years—customers decide not to continue purchasing products or services from an organization. Customers' data lie in two categories: soft and hard. The term "hard data" refers to the records generated by various devices and programs, including but not limited to smartphones, computers, sensors, smart meters, fleet management systems, call detail records (CDRs), and consumer bank transaction data. On the other hand, information that is subject to interpretation and viewpoint is known as "soft data." Fusing these two types of data leads to better customer's behavior analysis. This paper uses a supervised machine learning algorithm, namely a decision tree (DT), and the change mining method to model hard data. K-means clustering, an unsupervised machine learning algorithm, is also used along with the data preprocessing techniques. This paper also considers the Dempster-Shafer theory and other steps for soft data modeling. By fusing soft and hard data, the churn rate of customers compared with each other can be calculated. Besides, the customers' banking data are leveraged for data modeling. The results show that the banking industry will gain a more dynamic and efficient customer relationship management system by using this model.

INDEX TERMS Change mining, customer churn analysis, customer relationship management, hard and soft data fusion, machine learning.

I. INTRODUCTION

Organizations require innovation to meet new customer needs, and business executives must make the right decisions to succeed in today's competitive marketplace. Decision-making depends on data, information, and knowledge gathered from different sources. The hidden knowledge extracted from the data could be the source of competitive advantage and differentiation for organizations. Given the intense market competition and various choices of services and products facing customers, analyzing and predicting customer's behavior can be critical for organizations, especially banks. As one of the customer relationship management components, customer maintenance is an essential strategy for customer satisfaction and continuity of communication with customers.

Customers disconnect from an organization or may reduce their level of communication for obvious or hidden reasons. With churn management, those customers can be identified. "Churn" indicates the loss of clients over a specific time frame. Since acquiring a new customer is more costly than retaining a current one, managers should consider these customers for retention strategies. Churn may have different meanings in different businesses. In this study, customers

The associate editor coordinating the review of this manuscript and approving it for publication was Aasia Khanum.

are labeled as churn and non-churn, using the definition provided by experts based on the reduction in the number of annual transactions and the annual customer account's average balance.

So far, the majority of works have focused on modeling customer churn in the banking industry, but none of them has fused hard and soft data in their analysis. This paper is the first to study hard and soft data fusion using the decision tree (DT) and the change mining method to analyze customer churn behavior. Compared with previous studies, the main advantage of this study is the inclusion of hard and soft data fusion in customer churn analysis. Banking experts and managers can use these analytics to change or refine marketing strategies.

There are methods to handle some forms of uncertainty, such as missing data, outliers, and noise that may occur in real-world data. However, more than these methods is needed to manage all types of uncertainties because the data in the database do not include all the factors affecting customers' behavior and decisions, such as political, economic, and social factors. It is better to benefit from the diagnosis and opinions of expert people to overcome the existing limitations in diagnosing and measuring various uncertainties regarding customers' behavior with the help of a structured method. Therefore, fusing hard and soft data seems necessary to better analyze customer data, including banking customers on whom the organization's profit depends.

It is important to note that hard data cannot provide answers regarding real results due to their limitations, such as their inability to account for reasoning. A representative sample size is also essential for generating reliable hard data. In cases where it is impossible to collect a representative sample, it is necessary to build an appropriate sample size. Due to the necessity of representing a larger group well, this process can be time-consuming and tedious. Soft data are helpful for businesses because they offer researchers a deeper understanding of the subject matter. Hard data delivers facts, while soft data delivers context to participants' responses. It would therefore be possible to reduce the limitations associated with each type, increase the accuracy of the results, and analyze the results more realistically by combining these two types of data. As regards churn, various economic, political, and social factors influence customer churn, making it impractical to rely solely on hard data that only considers limited customer characteristics and ignores a wide range of customer factors affecting customer churn. Accordingly, soft data, such as expert opinions, can improve hard data results and enhance prediction accuracy and speed.

The proposed method is executed on the bank customers' data in a specific country, and the findings are elaborated. This method is transferable to other organizations or industries, including telecommunications, retail, and insurance, which may experience customer churn in a dynamic business environment.

Hence, the key contributions of this manuscript can be recapitulated as follows:

- This paper measures the churn rate of customers compared with each other with the help of fusing hard and soft data instead of focusing on predicting customers' churn separately.
- This paper considers the opinions of banking experts about customers as soft data to analyze customer churn as a source of decision-making and considers the financial, behavioral, and demographic data of customers as hard data.
- This paper leverages the change mining method's similarity measure and machine learning techniques for hard data modeling and the Dempster-Shafer theory for soft data modeling.

The remainder of this manuscript is organized as follows: In Section II, the related works are reviewed. The employed techniques are presented in Section III. The research methodology is described in Section IV, and the analysis of the results is provided in Section V. Finally, the manuscript is concluded in Section VI.

II. RELATED WORKS

In this section, customer relationship management (CRM) and customer churn concepts are discussed, and the related works are reviewed; next, the DT and change mining methods, along with the Dempster-Shafer theory, are described.

A. CUSTOMER RELATIONSHIP MANAGEMENT

CRM has been recognized as an essential concept in business recently. While no globally agreed definition for CRM exists [1], [2], [3], [4], [5], all definitions describe it as an integrated approach to customer acquisition and retention that uses business intelligence to increase customer value [3]. CRM can be characterized as a cycle with four aspects regarding the customer: identification, attraction, retention, and development. These aspects assist businesses in discriminating and directing resources toward the most profit-making customer group [3], [6]. As one of the CRM components, customer retention is an essential strategy for customer satisfaction and communication continuity [7].

B. CUSTOMER CHURN

Customer churn is a term that implies the loss of customers for various reasons. Churn management is a concept that identifies customers who would like to transfer their demands to a competing service provider [8]. In the following, some of the research on customer churn is reviewed.

Ahn et al. used the transaction and tax information of the customers of a mobile service company in Korea to investigate the factors affecting customer churn. In this study, changing the regular customer transaction trend to a downward trend or blocking by the company is considered a sign of partial churn. Their results show a relationship between changing the customer's situation and the possibility of churn. The technique used in this research is the logistic regression method [9].

Tsai and Lu research used one-year data from the communications industry to predict churn, where two hybrid models were used to combine two backpropagation neural networks and a self-organizing map. In the first hybrid model, the first multi-layer perceptron (MLP) was leveraged for data reduction and the second MLP for prediction. In the second model, the neural network used self-organized mapping, while MLP was employed for prediction. A five-part cross-validation method was used to assess the hybrid models. The results showed better performance in combining MLPs with other MLPs rather than in combining neural networks with self-organizing maps. It was also shown that the second model performed better than the case where the neural network was used alone [10].

To improve prediction accuracy, Basiri et al. used the ordered weighted averaging (OWA) operator. In this approach, two prediction algorithms, namely bagging and boosting classification trees as well as local linear model trees (LOLIMOT), were employed. Using the optimistic exponential OWA operator, the decision lists of the underlying learned classifiers were fused [11].

Miguéis et al. investigated the partial churn of store customers. The study had five periods, and the second period was considered as a reference. Compared with the reference period, customers who bought less than 40% in each period were considered customers with partial churn. Indeed, to detect a partial churn, each customer's behavior is compared with their past behavior [12].

In the study by Hassouna et al., researchers compared DTs and logistic regression at a mobile supplier in the UK. They used a balanced dataset (50% churner). The training data included 19919 customers, the experimental data included 15519 customers, and the number of variables was 17. CART, CHAID, and C5.0 DT algorithms were used for comparison with logistic regression. The researchers used the receiver operating characteristics (ROC) curve and accuracy to compare the models. According to the results, the C5.0 algorithm with an area under the curve (AUC) of 0.76 showed the best performance [13].

Azeem et al. used a fuzzy classification algorithm to predict the churn of customers in the communication industry. Features were selected based on domain knowledge, and 84 out of 722 variables were selected. The data were divided into two parts: training data and test data. The researchers used the sampling method to balance the dataset. Finally, they compared the fuzzy classifier with non-fuzzy algorithms, such as neural networks, logistic regression, DTs, and support vector machines (SVMs). In this study, the C4.5 algorithm with an AUC of 0.57 was the worst of the classifiers, and the fuzzy classifier with an AUC of 0.68 was recognized as the best algorithm [14].

Vijaya and Sivasankar research presented a methodology based on the rough set theory to identify effective variables for predicting churned customers in the telecommunications industry. Selected attributes were assigned to group classification techniques, such as bagging and boosting. The model's efficiency was evaluated by calculating its accuracy. It was found that the model presented with group classification techniques increased its accuracy to 95.13%, compared with the single-classification model [15].

Alboukaey et al. considered customers' daily behavior, instead of monthly behavior, to predict churn. They proposed four RFM-based (which is short for recency, frequency, and monetary), statistics-based, long short-term memory (LSTM)-based, and convolutional neural network (CNN)-based models to predict churn. According to the results, daily models performed significantly better than monthly models in predicting churners more quickly and accurately [16].

Vo et al. utilized unstructured call log data, as opposed to current approaches that are primarily based on structured data, such as demographics and account history, to predict customer churn [17].

Modelling and predicting customer churn have been considered in various industries, such as telecommunications, insurance, and banking. Considering the focus of the current research, a survey on customer churn research in the banking industry is provided. Since churn analysis is highly dependent on the definition of churn in the organization under consideration, a different definition is given to conduct churn research in each organization according to the organization's characteristics. The following are some of the churn definitions in the banking industry.

In Gür Ali and Arıtürk research, according to the bank's definition, customers are defined as churners if their portfolio size remains the same for six months below a threshold value. The term ''portfolio'' refers to the amount of performance of several customer accounts specified by the bank [18].

In Rosa research, weak relationships between the customer and the bank indicate involuntary churn. A weak relationship is a situation with the following conditions: 1) no transactions have been recorded in each customer's current account in the past three months, and 2) every customer's current account balance is less than $50 [19].

In Safinejad et al. research, model indicators were weighted during interviews with banking experts; these indices cluster the customers. The value of each cluster was calculated, and the most valuable clusters were selected. A fuzzy inference model determined the degree of customer churn of each valuable cluster in each season, and a neural network predicted the future churn of every cluster. To create a fuzzy inference system, according to bank experts, three variables of account balance average (M), duration of the relationship with the bank (L), and the number of customer transactions (F) were considered as input variables for each customer, and the churn degree was considered as an output variable. For each of the three input variables, the states were assumed to be high, medium, and low. Regarding the three input variables and the three states for each one, this fuzzy system had a maximum of 27 rules. These rules were defined by the banking experts who represented the outputs for different input states. For example, if the average account balance is high, the duration of the relationship with the bank is high, and the number of transactions is high, then the degree of churn is high [20].

Some of the previous research used the specific definition provided by the bank to identify churning customers [18], [19]. In those research that did not have a specific definition by the bank, the opinion of banking experts was used to identify churning customers using techniques such as clustering and neural networks [20], [21].

In the current study, according to banking experts, a customer is a churner if their number of annual transactions and the average annual account balance are reduced by 30%.

III. EMPLOYED TECHNIQUES

A. MACHINE LEARNING

This subsection presents one supervised and one unsupervised machine learning technique, namely DT classification and k-means clustering, respectively, that are made use of in the current study.

Classification is a form of data analysis in which models are extracted to describe essential data classes. Such models, known as classifiers, predict class labels. Two steps are involved in data classification. In the learning step, a model is built/learned, and in the classification step, that model is used to predict class labels. Examples of classification algorithms are simple Bayes, Bayesian networks, nearest neighbors, neural networks, DTs, and logistic regression [22]. DT is the most widely employed algorithm in churn prediction, along with neural networks, due to its interpretability, deployment facility, and hierarchical structure; thus, it is presented as a set of rules for decision-makers. The structure of DT comprises a decision node and a leaf node; the former is the rule of decision for further data splitting, whereas the latter represents the prediction. A split criterion is used to split the data. C4.5, C5.0, CART, CHAID, and CTREE are the DT algorithms used in this study, which differ according to the type of target variable, the criterion of branching, and the number of branches [23].

Clustering is the process of splitting data points into several groups (i.e., clusters) so that data points in a given group are more comparable and have similar characteristics compared with data points in other groups. Clustering methods include partitioning, hierarchical, fuzzy, and density-based [22]. One of the partitioning methods is k-means, which is the most straightforward and popular unsupervised machine learning algorithm. The first group of randomly selected centroids is used as the starting point for every cluster in the k-means algorithm. The centroids are then optimized using iterative calculations. Clustering evaluation is essential to assessing the effectiveness of a clustering method in retrieving clusters and analogizing the performance of different clustering methods. There are two criteria for evaluating clustering results: external evaluation and internal evaluation. The former mainly tries to evaluate the performance of a cluster using

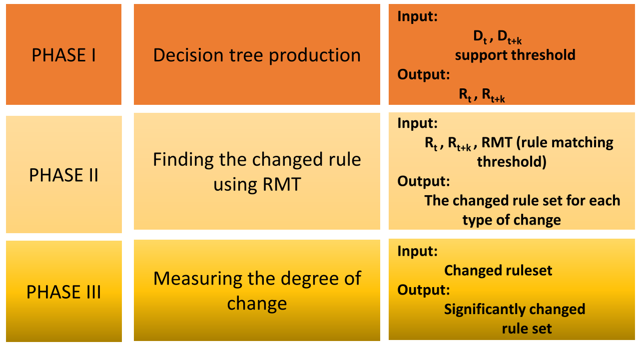

FIGURE 1. General procedure to recognize the change.

experts, while the latter evaluates the structure of clusters created by clustering algorithms. Clustering results are evaluated based on several different criteria. These indices measure the similarities and differences between members within the cluster. Therefore, the method that creates the most similarity or difference within the cluster is considered the appropriate method [24], [25].

B. CHANGE MINING

This approach usually divides the data into two categories according to time. The patterns obtained from the two sets are then compared. Different patterns are analyzed into different categories, such as emerging patterns, patterns that disappear over time, and patterns added in recent times [10], [26], [27]. This paper describes the method of change mining as described in [26].

According to Figure 1, recognizing changes in customer's behavior comprises three phases. Phase I extracts two sets of rules from each interval database separately using DT analysis. Phase II creates the changed rule set by utilizing the rule matching method based on comparing the two rules from each set. Emerging, unexpected, and added/perished patterns are types of rule changes extracted in Phase II. In order to comply with the rules, the criteria of difference and similarity are effectively defined. Phase III of the process involves a ranking of the changed rules identified in Phase II in accordance with the extent of the changes. Some symbols are defined in Table 1 for the description of our rule change mining method.

A similarity criterion is defined as follows:

$$s_{ij} = \begin{cases} \dfrac{l_{ij} \sum_{k \in A_{ij}} x_{ijk} c_{ij} y_{ij}}{|A_{ij}|}, & \text{if } |A_{ij}| \neq 0 \\ 0, & \text{if } |A_{ij}| = 0 \end{cases} \tag{1}$$

where $l_{ij} \sum_{k \in A_{ij}} x_{ijk} / |A_{ij}|$ represents the similarity of the conditional part and $c_{ij} y_{ij}$ indicates the similarity of the consequent part between r_i^t and r_j^{t+k}. When both parts between r_i^t and r_j^{t+k} are identical, the similarity degree equals 1. The similarity measure can range from 0 to 1.

C. DEMPSTER-SHAFER THEORY

The Dempster-Shafer theory provides a generalized expression for uncertainty by considering sets of propositions

TABLE 1. Some symbols in the change mining method.

Symbol	Description	Formula
S_{ij}	Similarity measure: similarity degree between r_i^t and r_j^{t+k}	$0 \leq S_{ij} \leq 1$
l_{ij}	Degree of attribute match of the conditional parts	$I_{ij} = \frac{\lvert A_{ij}\rvert}{\max(\lvert x_i^t\rvert\lvert x_j^{t+k}\rvert)}$
c_{ij}	Degree of attribute match of the consequent parts	$c_{ij} = \begin{cases} 1, & \text{The same consequent attribute} \\ 0, & \text{O.W.} \end{cases}$
$\lvert A_{ij}\rvert$	Number of attributes common to the conditional parts of both r_i^t and r_j^{t+k}	
$\lvert X_i^t\rvert$	Number of attributes in the conditional parts of r_i^t	
$\lvert X_j^{t+k}\rvert$	Number of attributes in the conditional parts of r_j^{t+k}	
x_{ijk}	Degree of value match of the k-th matching attribute in A_{ij}	$x_{ijk} = \begin{cases} 1, & \text{The same value} \\ 0, & \text{O.W.} \end{cases}$
y_{ij}	Degree of value match of the consequent attribute	$y_{ij} = \begin{cases} 1, & \text{The same value} \\ 0, & \text{O.W.} \end{cases}$

(rather than just one) and assigning an interval to each set that includes the degree of belief for the set. It is particularly helpful when each piece of evidence supports multiple candidates' conclusions, and the support of each conclusion is determined by overlapping the contributions of various pieces of evidence. As opposed to classical probability theory, the Dempster-Shafer theory also allows some portion of belief to remain unassigned so as to reflect relative ignorance when information is incomplete, thus being suitable for knowledge representation in specific domains, notably legal reasoning [28].

Assume a set of n mutually exclusive and exhaustive propositions, $\Theta = \{X_0, X_1, \ldots, X_n\}$, where Θ is called a frame of discernment. 2^Θ refers to the set of all subsets of Θ [29]. The opinion about a proposition x normally has the form of

$$\omega_x = \big(b(x), d(x), u(x), a(x)\big). \quad (2)$$

According to the Dempster-Shafer theory, evidence is assigned based on mass probability, denoted by $m(x)$, where

$$m(x) \geq 0; \quad m(\emptyset) = 0; \quad \sum_{x \in 2^\Theta} m(x) = 1. \quad (3)$$

The support for a proposition is the total degree of belief for that proposition, defined as

$$b(x) = \sum_{y \subseteq x} m(y); \quad x, y \in 2^\Theta, \quad (4)$$

and the disbelief for a proposition x is the belief for $\bar{x}$, defined as

$$d(x) = \sum_{y \cap x = \emptyset} m(y); \quad x, y \in 2^\Theta. \quad (5)$$

For a given set x, uncertainty is measured by the total belief masses on supersets or partially overlapped sets of x,

$$u(x) = \sum_{y \cap x \neq \emptyset, y \not\subset x} m(y); \quad x, y \in 2^\Theta; \quad x \neq \emptyset \quad (6)$$

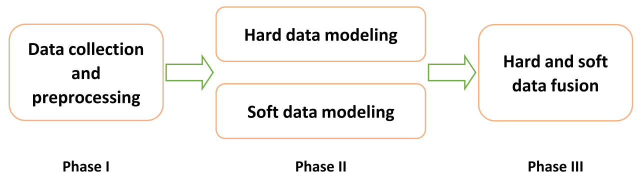

FIGURE 2. Proposed methodology phases.

where

$$b(x) + d(x) + u(x) = 1; \quad x, y \in 2^\Theta. \quad (7)$$

The fourth element of (2), $a(\cdot)$, is known as the relative atomicity, which is the number of atomic sets that any particular set x contains. The relative atomicity of a proposition x relative to another proposition y is defined as

$$a(x|y) = \frac{|x \cap y|}{|y|}. \quad (8)$$

Another quantity is the probability expectation, denoted by $E(\cdot)$, and is defined as

$$E(x) = \sum_{y} m(y)a(x|y); \quad x, y \in 2^\Theta. \quad (9)$$

Given any particular proposition x, $E(x)$ is a mapping of ω_x onto the probability space [0, 1] and is the expected probability of x being true such that

$$\sum_{i} E(x_i) = 1; \quad x_i \in 2^\Theta. \quad (10)$$

IV. RESEARCH METHODOLOGY

The proposed methodology comprises three main phases, detailed in the following and illustrated in Figure 2 and Figure 3.

In the following, a description of the presented methodology is provided.

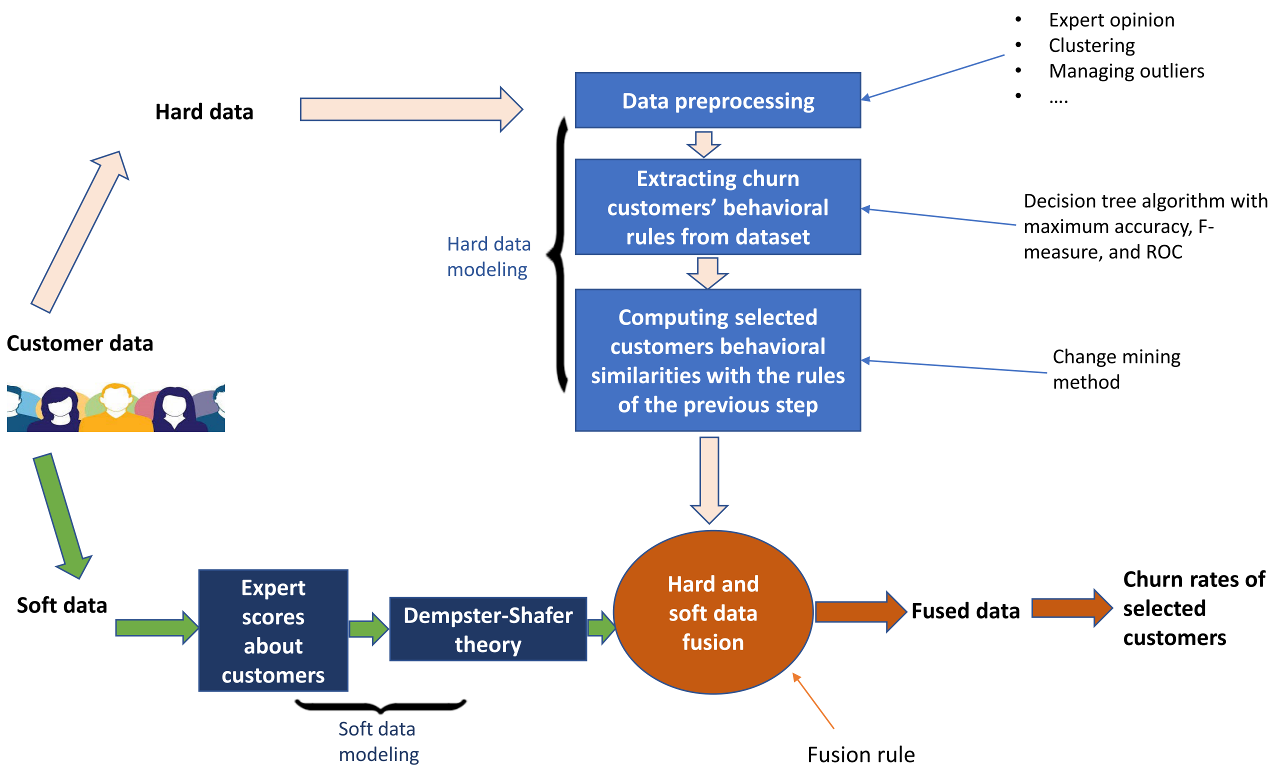

FIGURE 3. Proposed methodology steps.

A. PHASE I - DATA COLLECTION AND PREPROCESSING

In the first phase, data must be collected and preprocessed. For this study, data have been extracted from a bank's database in a specific country. In order to identify suitable variables for customer churn modeling in banking, past research in the churn prediction of the banking industry has been studied, and the variables used in them have been extracted. In interviews with banking experts, these variables and their impact on customer churn modeling were scrutinized. In these interviews, various variables were mentioned for modeling bank customer's churn behavior. The essential variables in describing and examining customer's churn behavior, according to the literature, are as follows:

- L: Length of the period of customer's relationship with the bank (year);
- F: Total number of annual customer's account transactions;
- M: Annual average of customer's account balance.

These variables, along with some behavioral and demographic variables, were provided to the researcher. Data preprocessing is essential to ensuring the quality of the findings. Some of the main preprocessing tasks are handling missing values, outliers, and duplicate data as well as feature selection, normalization, and discretization. The general purpose of discrete methods is to convert continuous features into several categories to extract knowledge from that information. There are different discretization methods, such as frequency-based discretization, entropy, and clustering. This study makes use of the k-means clustering algorithm to discretize and classify the LFM variables. Some criteria provided by experts discretize other variables. This study also uses the average silhouettes to select the best number of clusters. Clustering with the highest average silhouette is considered the best. With the help of experts, the results of clustering the LFM variables (external clustering criterion) have been evaluated.

B. PHASE II - DATA MODELING

1) HARD DATA MODELING

By extracting the behavioral patterns of churned customers from the database with the help of a classification method, such as the DT, a comparison has been made with the behavioral patterns of selected customers, and the similarity between them has been calculated with the help of the change mining method's similarity criteria. It is then possible to determine whether the selected customers are churners. The similarity value has been used as the output of the hard data source. This paper names some customers as selected customers to rate their churn, which is not included in the database.

2) SOFT DATA MODELING

In soft data modeling, according to Acharya and Kam's research, the following five steps must be followed [29]:

1) Ask some banking experts to rate the selected customers based on their confidence in being churn, non-churn, or uncertain. This study uses the opinions of 18 banking experts.
2) Randomly sample the data generated by banking experts twice (each as a soft source W_1 and W_2).
3) Calculate (4)-(8) for the selected customers based on the soft source of information using the Dempster-Shafer theory.
4) Calculate the combined values of (4)-(8) obtained from two soft sources for the selected customers based on the soft source of information with the following relations:

$$b_x^{1,\ldots,n} = \frac{\sum_{i=1}^{n} b_x^i (\prod_{j\neq i} u_x^j)}{\sum_{i=1}^{n}(\prod_{j\neq i} u_x^j) - (n-1)(\prod_{i=1}^{n} u_x^i)}, \quad (11)$$

$$d_x^{1,\ldots,n} = \frac{\sum_{i=1}^{n} d_x^i (\prod_{j\neq i} u_x^j)}{\sum_{i=1}^{n}(\prod_{j\neq i} u_x^j) - (n-1)(\prod_{i=1}^{n} u_x^i)}, \quad (12)$$

$$u_x^{1,\ldots,n} = \frac{\prod_{i=1}^{n} u_x^i}{\sum_{i=1}^{n}(\prod_{j\neq i} u_x^j) - (n-1)(\prod_{i=1}^{n} u_x^i)}, \quad (13)$$

$$a_x^{1,\ldots,n} = \frac{\sum_{i=1}^{n} [a_x^i (\prod_{j\neq i} u_x^j)(1-u_x^i)]}{\sum_{i=1}^{n}(\prod_{j\neq i} u_x^j) - n(\prod_{i=1}^{n} u_x^i)}. \quad (14)$$

5) Calculate the values of $E_P^{W_1W_2}$, $E_S^{W_1W_2}$, and …for the selected customers, including P, S, and etc., based on

$$E(x)^{1,\ldots,n} = b_x^{1,\ldots,n} + (d_x^{1,\ldots,n} \times 0) + (u_x^{1,\ldots,n} \times a_x^{1,\ldots,n}). \quad (15)$$

6) Perform steps 2 to 4 ten times and calculate the average of the values obtained to ensure the reliability of the results.

C. PHASE III - HARD AND SOFT DATA FUSION

Through the use of data obtained from the hard and soft sources, the following fusion rule can be applied in order to combine the sources [29]. This fusion rule is based on the Bayesian law and total probability, and is demonstrated as

$$\eta(q_1,\ldots,q_m) = \left(1 + \left[\frac{p_1}{1-p_1}\frac{f(q_1,\ldots,q_m|1)}{f(q_1,\ldots,q_m|0)}\right]^{-1}\right)^{-1}, \quad (16)$$

where p_1 is the prior probability, $f(\cdot)$ is the joint probability density function, and $q_1,\ldots,q_m$ corresponds to a specific proposition (i.e., customer) and consists of values achieved from all the sources regardless of their type. The data sources are assumed to follow the beta distribution. Since the sources are independent, the value of $f(\cdot)$ is the product of the values of the probability density function of each source.

V. IMPLEMENTATION AND RESULTS

This section presents the methodology implementation and results analysis.

A. DATA COLLECTION AND PREPROCESSING

One of the methods for discretizing data is clustering. Also, the rules extracted from DT algorithms are compared with selected customers' behavior rules using the change mining method that uses categorized features. Therefore, it is necessary to have two rules that generate similar or approximately similar ranges. Hence, the three most essential variables (i.e., L, F, and M) are classified using the k-means clustering method, and the numerical ranges for the clusters are also used to classify selected customer features.

All other features are discretized according to the criteria stated by the banking experts. The clustering of LFM variables in the dataset is described below. The k-means clustering algorithm clustered the set of 4995 customers into two clusters with very high and very small (outlier) data, so the large cluster was re-clustered, and this process was repeated over several steps. The results of clustering variable M are presented in Table 2.

Since all customers' average account balance is \$33399.4, the outliers of the first, second, and third stages are considered as one cluster (high-mean cluster); the outliers of the fourth stage are considered as another cluster (medium-mean cluster), and the big cluster of the fourth stage is considered as the third cluster (low-mean cluster). According to what was mentioned about variable M, variable F was also clustered. The results of clustering variable F are summarized in Table 3.

The total customer average is 1204 transactions. Therefore, the outliers of the first and second stages are considered as the first cluster (high transactional cluster), the outlying cluster of the third stage is considered the second cluster (medium-mean cluster), and the large cluster of the third stage is considered as the third cluster (low-mean cluster). Variable L is also classified into three clusters with a high (21 years), average (10 years), and low (4 years) mean by the k-means clustering algorithm. The mean and number of members per cluster in the clustering of variable L are given in Table 4.

B. HARD DATA MODELING

This paper uses the bank dataset to employ the change mining method and its similarity measure to calculate the similarity between dataset churn rules and selected customers' behavior rules. The following steps extract customers' behavioral rules for this dataset. The selected customers' behavioral similarities with those rules are calculated by similarity measure.

1) EXTRACTING CUSTOMERS' BEHAVIORAL RULES FROM THE DATASET

In this step, different DT algorithms are compared with different criteria, and the best algorithm is selected.

Table 5 compares different DT algorithms based on three criteria: accuracy, F-measure, and ROC. As can be seen in Table 5, the CTREE algorithm has higher values of accuracy, F-measure, and ROC compared with the other algorithms. These criteria have been also calculated for different

TABLE 2. Clustering steps for variable M.

Step	Number of Clusters Based on Silhouettes	Number of Each Cluster	Average of Large Cluster	Average of Outlying Cluster
1	2	5 and 4990	$26801.7	$6617915.8
2	2	52 and 4938	$18314.4	$833072.3
3	2	176 and 4762	$11123.09	$212871.5
4	2	496 and 4266	$5798.67	$56903.03

TABLE 3. Clustering steps for variable F.

Step	Number of Clusters Based on Silhouettes	Number of Each Cluster	Average of Large Cluster	Average of Outlying Cluster
1	2	57 and 4938	829	33674
2	2	205 and 4733	529	7740
3	2	521 and 4212	521	2104

TABLE 4. Characteristics of each cluster in the clustering of variable L.

	Low Average Cluster	Medium Average Cluster	High Average Cluster
Average	4	10	21
Members	1574	2948	473

TABLE 5. Comparison of DT algorithms (dataset is split into a 70 : 30 ratio).

Algorithm	Accuracy	F-measure	ROC
CTREE	0.8572	0.9228	0.67
C4.5	0.8474	0.9161	0.55
C5.0	0.8557	0.9207	0.53
CHAID	0.8414	0.9138	0.65
CART	0.8514	0.9190	0.53

algorithms based on train-test ratios of 70 : 30, 75 : 25, and 80 : 20, and the CTREE algorithm gained maximum values.

DT is plotted for data using the best algorithm gained from the previous step, and the customers' behavioral rules are extracted. The rules are in the if-then form, where the conditional part represents the customer's behavioral, financial, and sociological features, and the consequent part represents the customer's churn status. For example, the rule below shows a non-churn, individual, and foreign customer with a low annual average in the account balance and average annual transaction who does not own a point of sale (POS) and is not a member of the bank customer club:

- IF type = individual & foreigner AND annual average = low AND transaction = average AND POS = no AND club membership = no, THEN churn = no.

2) COMPUTING CUSTOMER's BEHAVIORAL SIMILARITIES

The similarity of a customer's behavioral pattern to the churn patterns in the database can be considered hard data. The higher the degree of similarity, the more likely a customer will be a churner.

C. SOFT DATA MODELING

The soft data is also calculated according to the method described in Phase III for two customers, P and S. The results are shown in Table 6. After following steps one to six to generate reliable soft data that well covers the opinions of all participating experts and create a reliable average as the final soft data, the final value obtained is combined with the hard data. As can be seen in the table, the soft data value for customers P and S is 0.585 and 0.384, respectively. These values are actually the average of expectations, $E(x)$, for each of the customers, which contain the values of belief, disbelief, and uncertainty of experts regarding the churn situation of different customers.

TABLE 6. Hard and soft data modeling outputs.

	P	S
Soft	0.585	0.384
Hard	0.8	0.3

TABLE 7. Values of α and β for each source and hypothesis.

Hypothesis	Soft Data	Hard Data
$h = 1$	$\alpha = 4, \beta = 3$	$\alpha = 3, \beta = 2$
$h = 0$	$\alpha = 3, \beta = 4$	$\alpha = 2, \beta = 3$

D. HARD AND SOFT DATA FUSION

α and β values for each source are also listed in Table 7. These values are considered symmetric for simplicity. The prior probability for two customers, P and S, is assumed to equal $1/2$ for both. Finally, the values 0.956 and 0.671 are obtained for P and S, respectively, indicating that customer P is more churn than customer S with respect to hard and soft data sources. This model can be used for any number of customers, and their churn rate can be calculated. The results can be used to improve the organization's customer relationship management system.

VI. CONCLUSION

The current study presented a customer churn model for the banking industry based on fusing hard and soft data for churned customers in order to achieve high accuracy. This was a gap not addressed in past research. Therefore, in this study, considering a customer dataset, churned customer's behavior patterns were extracted using one of the most commonly used classification algorithms, the decision tree (DT). For this purpose, a relatively appropriate algorithm was selected by comparing different DT algorithms. The change mining method and its similarity measure were then employed to compare the behavioral patterns of selected customers with churned customers' behavioral rules extracted from the dataset to obtain hard data. The expert opinions about the selected customers were also leveraged in order to model soft data using the Dempster-Shafer theory. The results of soft and hard sources were fused with the fusion rule. The outcomes illustrated that the usage of the proposed model by bank decision-makers can lead to a precise analysis of customer churn rate and can modify banking policies to have more efficient customer relationship management.

REFERENCES

[1] R. Ling and D. C. Yen, "Customer relationship management: An analysis framework and implementation strategies," *J. Comput. Inf. Syst.*, vol. 41, no. 3, pp. 82–97, Feb. 2001.

[2] E. W. T. Ngai, "Customer relationship management research (1992–2002): An academic literature review and classification," *Marketing Intell. Planning*, vol. 23, no. 6, pp. 582–605, Oct. 2005.

[3] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2592–2602, Mar. 2009.

[4] J. U. Becker, G. Greve, and S. Albers, "The impact of technological and organizational implementation of CRM on customer acquisition, maintenance, and retention," *Int. J. Res. Marketing*, vol. 26, no. 3, pp. 207–215, Sep. 2009.

[5] E. Akhondzadeh-Noughabi and A. Albadvi, "Mining the dominant patterns of customer shifts between segments by using top-*k* and distinguishing sequential rules," *Manage. Decis.*, vol. 53, no. 9, pp. 1976–2003, Oct. 2015.

[6] R. Swift, *Accelerating Customer Relationships: Using CRM and Relationship Technologies*, 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[7] L. G. Renart, "Relationship marketing: Gaining competitive advantage through customer satisfaction and customer retention," *J. Consum. Marketing*, vol. 18, no. 1, pp. 75–83, 2001.

[8] A. Ghorbani and F. Taghiyareh, "CMF: A framework to improve the management of customer churn," in *Proc. IEEE Asia–Pacific Services Comput. Conf. (APSCC)*, Dec. 2009, pp. 457–462.

[9] J.-H. Ahn, S.-P. Han, and Y.-S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," *Telecommun. Policy*, vol. 30, nos. 10–11, pp. 552–568, Nov. 2006.

[10] C.-F. Tsai and Y.-H. Lu, "Customer churn prediction by hybrid neural networks," *Expert Syst. Appl.*, vol. 36, no. 10, pp. 12547–12553, Dec. 2009.

[11] J. Basiri, F. Taghiyareh, and B. Moshiri, "A hybrid approach to predict churn," in *Proc. IEEE Asia–Pacific Services Comput. Conf.*, Dec. 2010, pp. 485–491.

[12] V. L. Miguéis, D. Van den Poel, A. S. Camanho, and J. F. E. Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11250–11256, Sep. 2012.

[13] M. Hassouna, A. Tarhini, T. Elyas, and M. S. A. Trab, "Customer churn in mobile markets: A comparison of techniques," *Int. Bus. Res.*, vol. 8, no. 6, pp. 224–237, May 2015.

[14] M. Azeem, M. Usman, and A. C. M. Fong, "A churn prediction model for prepaid customers in telecom using fuzzy classifiers," *Telecommun. Syst.*, vol. 66, no. 4, pp. 603–614, Dec. 2017.

[15] J. Vijaya and E. Sivasankar, "Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector," *Computing*, vol. 100, pp. 839–860, Aug. 2018.

[16] N. Alboukaey, A. Joukhadar, and N. Ghneim, "Dynamic behavior based churn prediction in mobile telecom," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113779.

[17] N. N. Y. Vo, S. Liu, X. Li, and G. Xu, "Leveraging unstructured call log data for customer churn prediction," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106586.

[18] Ö. G. Ali and U. Aritürk, "Dynamic churn prediction framework with more effective use of rare event data: The case of private banking," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7889–7903, Dec. 2014.

[19] N. B. D. C. Rosa, "Gauging and foreseeing customer churn in the banking industry: A neural network approach," M.S. thesis, NOVA Inf. Manage. School, Campus de Campolide, Lisbon, Portugal, 2019. [Online]. Available: http://hdl.handle.net/10362/71584

[20] F. Safinejad, E. A. Z. Noughabi, and B. H. Far, *A Fuzzy Dynamic Model for Customer Churn Prediction in Retail Banking Industry*. Cham, Switzerland: Springer, Oct. 2018, pp. 85–101.

[21] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," *Financial Innov.*, vol. 2, no. 1, pp. 1–13, Aug. 2016.

[22] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2011.

[23] I. H. A. Kandel, "A comparative study of tree-based models for churn prediction: A case study in the telecommunication sector," M.S. thesis, NOVA Inf. Manag. School, Campus de Campolide, Lisbon, Portugal, 2019. [Online]. Available: http://hdl.handle.net/10362/60302

[24] M. Halkidi and M. Vazirgiannis, "A density-based cluster validity approach using multi-representatives," *Pattern Recognit. Lett.*, vol. 29, no. 6, pp. 773–786, Apr. 2008.

[25] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, "Internal versus external cluster validation indexes," *Int. J. Comput. Commun.*, vol. 5, no. 1, pp. 27–34, 2011.

[26] J. K. Kim, H. S. Song, T. S. Kim, and H. K. Kim, "Detecting the change of customer behavior based on decision tree analysis," *Expert Syst.*, vol. 22, no. 4, pp. 193–205, Sep. 2005.

[27] T. C.-K. Huang, "Mining the change of customer behavior in fuzzy time-interval sequential patterns," *Appl. Soft Comput.*, vol. 12, no. 3, pp. 1068–1086, Mar. 2012.

[28] T. L. Wickramarathne, "An analytical framework for soft and hard data fusion: A belief theoretic approach," Ph.D. dissertation, Dept. Elect. Comput. Eng., McArthur Eng. Building, Coral Gables, FL, USA, 2012.

[29] S. Acharya and M. Kam, "Evidence combination for hard and soft sensor data fusion," in *Proc. 14th Int. Conf. Inf. Fusion*, 2011, pp. 1–8.

MASOUD ALIZADEH received the M.Sc. degree in information technology engineering from Tarbiat Modares University, in 2019. He is currently pursuing the Ph.D. degree in information technology with the University of Tehran. He has over five years of experience working in data mining, machine learning, and hard/soft data fusion areas. His research interests include Industry 4.0, machine learning, data fusion, and business analytics.

DANIAL SADRIAN ZADEH (Graduate Student Member, IEEE) received the B.Sc. degree in control engineering from the Petroleum University of Technology, Iran, in 2018, and the M.Sc. degree in control engineering from the University of Tehran, Iran, in 2021. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include autonomous vehicles, control theory, machine learning, sensor data fusion, and state estimation.

BEHZAD MOSHIRI (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in control systems engineering from the University of Manchester Institute of Science and Technology (UMIST), in 1987 and 1991, respectively. He is currently a Full Professor of control systems engineering with the School of Electrical and Computer Engineering, University of Tehran, Iran. He has been an Adjunct Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, since 2014. He is the author/coauthor of more than 360 articles, including more than 130 journal articles and more than 20 book chapters. His research interests include advanced industrial control, advanced instrumentation systems, data fusion theory, and its applications in areas such as robotics, process control, mechatronics, information technology (IT), intelligent transportation systems (ITS), bioinformatics, and financial engineering. He has been a member of the International Society of Information Fusion (ISIF), since 2002, and a member of the Waterloo AI Institute, since 2019. He has also been serving as the Chairperson for the IEEE Control Systems Chapter in the IEEE Iran Section, since December 2018.

ALLAHYAR MONTAZERI (Member, IEEE) received the B.S. degree in electrical engineering from Tehran University, Tehran, Iran, in 2000, and the M.Sc. and Ph.D. degrees in electrical engineering from the Iran University of Science and Technology, Tehran, in 2002 and 2009, respectively. From 2010 to 2011, he joined the Fraunhofer Institute, Germany, as a Research Fellow, and then carried on his research with Fraunhofer IDMT and Control Engineering Group, Ilmenau University, Germany, from 2011 to 2013. He was a Visiting Research Scholar with the Control Engineering Group, ETH Zürich, Switzerland, and the Chemical Engineering Group, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. Since 2013, he has been an Assistant Professor with the Engineering Department, Lancaster University, U.K. His research interests include control theory and digital signal processing. Particularly, he is interested in adaptive signal processing and control, robust control, linear and nonlinear system identification, estimation theory, and evolutionary computing and optimization with applications in active noise and vibration control systems and robotics. His research is funded by different councils and industries in the U.K., such as the Engineering and Physical Research Council, Sellafield Ltd., the National Nuclear Laboratory, and the Nuclear Decommissioning Authority. He is currently serving on the IFAC Technical Committees ''Adaptive and Learning Systems'' and ''Modeling, Identification, and Signal Processing.'' He is a fellow of the Higher Education Academy. He was a recipient of the European Research Consortium for Informatics and Mathematics (ERCIM), in 2010, and the Humboldt Research Awards, in 2011.