Global Scientific JOURNALS

GSJ: Volume 9, Issue 2, February 2021, Online: ISSN 2320-9186
www.globalscientificjournal.com

CUSTOMER RETENTION STRATEGIES USING DATA MINING METHODS: A REVIEW

Aroosa Javaid, Malik Mubashir Hussain, Dr. Hamid Ghoos

Department of Computer Science, Institute of Southern Punjab Multan Pakistan
Aroosa Javaid, Ph # 0300-6869927, Email: arosajavaid1@gmail.com
Malik Mubashir Hussain

KeyWords

Retention, Data Mining, Supervised, Unsupervised, Machine Learning, loyality, features

ABSTRACT

Customer retention has become one of the main priorities for business these days. The more they retain customers the better they survive in competitive market. The goal of customer retention programs is to help companies retain many customers. To retain customer it's a tough job because customer data is so large and imbalanced that company can't handled it manually. There is a need for the system that can automatically predict leaving customers. Many researchers in past have tried to solve this problem by using data mining techniques. In this review paper we are reviewed the literature related to customer retention using data mining techniques. The structure of this paper is based on supervised, unsupervised, and hybrid models of data mining. At the end we have discussed the limitations of the previous work and future directions for the upcoming researchers.

INTRODUCTION:

According to the organization, customers are the foundation of its success and sales, which is why companies are becoming increasingly concerned about the importance of achieving the satisfaction of their customers. There are two aspects of marketing in the retail market: customer retention and acquisition. Customer acquisition is the process of attracting new customers to the product and retention is the process of keeping the customer continuously buys the product. The business holds different client details about their visit, purchasing experience, and so on. There is a need to process this huge customer's data to predict leaving customers. While using different strategies to determine the user's interest, they may miss some features and therefore the different techniques would face challenges because of their higher spatial property. The issue of customer's retention has been well discussed and studied. There are many methods has been discussed initially with regard to improving marketing. However, methods go beyond achieving high performance and require some strategic approach. So with the introduction of Machine Learning Techniques, the problem of high size and missing performance can be easily solved with big data. There are a number of machine learning algorithms available e.g. supervised, unsupervised, and Semi supervised machine learning.

SIGNIFICANCE:

The business holds different customers details about their visit, purchasing experience, and so on. In predicting user interest, huge data is used. While using different strategies to determine the user's interest, they might miss some features and therefore the different techniques would face challenges because of their higher spatial property. If the business already predicts which customer is about to leave then they try to offer great services or reward loyalty to their customers. There are many methods has been discussed initially with regard to improving marketing. However, methods are difficult to achieve high performance and require some strategic approach. So by using machine learning algorithms and feature selection methods key features are selected from huge data that helps the business increase customer retention.

BACKGROUND:

Customer retention means the willingness of an organization to combine business with a particular customer or to continuously adjust to their requirement. Retention can also be characterized as the willingness of love, identity, commitment, trust and client to recommended and repeat purchases. It is important for the success of any businesses to establish successful customer's relationship. For the growth of businesses, customer satisfaction, retention, good words of mouth and loyalty are important. In addition, existing research on online customer retention is limited. Many businesses currently fail to attract new customers in order to maintain a suitable marketing department and assign mangers to pay attention to their existing customers. Most previous studies have concentrated on customer loyalty in restaurants, telecommunication firms, hotels and other services, although less attention has been given to customer retention in a retail industry.

DATA MINING

Data mining is a method used to predict customer retention. Data mining involves extracting information from a huge data and converting it into an easily interpretable system that enables organization to evaluate complex problems that result in customer loyalty and turn over to companies. Data mining provided many algorithms that we can use in the prediction of customer retention.

Supervised Machine Learning	Unsupervised Machine Learning
a. Classification i. Naïve Bayes ii. Random Forest iii. Nearest Neighbor iv. Discriminant Analysis v. Support vector Machine b. Regression i. Linear Regression GLM ii. SVR, GPR iii. Ensemble Methodology iv. Decision Tree v. Neural Network vi. Boosting vii. Bagging viii. Stacking	c. Clustering Algorithm i. Hierarchal Clustering ii. Gaussian Mixture iii. Neural Networks iv. C-Means v. Fuzzy vi. Hidden Markov Model vii. K-medoids viii. K-Means ix. KNN

Supervised

Naveen et.al (2020) predicted customer churn and relation model by using Machine Learning methods on cell2cell and IBM datasets.

They had predicted 30 variables and implemented Naïve Bayes, Decision Tree (DT) and Support Vector Machine (SVM). The outcome of their model was measured by using the Area under the curve (AUC) and gained 0.87, 0.82, and 0.77 for IBM and 0.98, 0.99 and 0.98 for the cell2cell dataset. On the other hand, **Hyun et.al (2020)** utilized the US-based telecommunication company dataset for the analysis of customer switching behavior and they implemented data cleaning and data balancing Pre-processing techniques. They had utilized some machine learning approaches Logistic regression (LR), Vector machine, Random Forest (RF) and DT. The outcomes of their model demonstrate that the predicted performance of their work is greater than 86% and the LR has the highest accuracy rate. They suggested using other data mining methods to create a predictable model similar to the ANN and Bayesian networks that were able to improve their research.

Hemlata et.al (2020) classifies the customer churn by using Supervised Machine Learning. They utilized some preprocessing strategies for data acquisition and cleaning. They grouped the dataset into training and testing that is acquired from an online source. They had implemented KNN (K-nearest Neighbors) and XGBoost Algorithms. Thus as that conclusion, KNN gains 83.85% accuracy with high error and low sensitivity and specificity while on the other hand, XGBoost gains accuracy of 86.85% with low error and high sensitivity and clarity. While, **Shreyas Rajesh (2020)** implemented machine learning algorithms, Extra Trees Classifier, XGBoosting Algorithm, SVM, SGD Classifier, AdaBoost Classifier, Gaussian Naive Bayes and LR to recognize the customer churn. He used BigML churn Telecom dataset and utilized some preprocessing strategies handling missing values, encoding categorical data, dividing the dataset into test and train sets, and feature scaling. The detection of their work was the Extra Tree classifier, XGBoosting Algorithm, and SVM was the finest with AUC score (0.8439, 0.78, and 0.735) respectively.

Sahar F. Sabbe (2018) used a public dataset of customers in the Telecom industry, took 60% of training and 40% of the testing dataset. He implemented machine learning techniques, DT (CART), SVM, KNN, Adaboost, RF, Stochastic gradient boost, and MLP ANN and upholds some pre-processing techniques like data cleaning, transformation, and feature selection. Thus the outcome of their work that RF and AdaBoost output performance is the same with the accuracy of 96%. Multilayer Perceptron and SVM give an accuracy of 94%. DT gives 90%, Naive Bayesian 88%, and LR and LDA give 86.70%. Hence RF and Adaboost give the best output performance. This research is extended by including deep learning and a hybrid model. On the other hand, **Nagraj et.al (2018)** purposed customer retention by upholding some machine learning algorithms ANN, SVM, and Deep Neural Network DNN and they took two datasets one of German credit second of a bank customer. They had applied pre-processing technique Normalization in the customer bank dataset. They had achieved the accuracy of 98%, 92%, and 97% for ANN, SVM, and DNN respectively and for bank customer data and 72%, 76% and 72%, respectively for the credit dataset of German. ANN gives better accuracy for bank customer data while DNN gives better accuracy for the German credit dataset.

Machine learning algorithms KNN, CART, and SVM used by **Pushpa et.al (2019)** for customer retention. They had utilized Normalization on the customer dataset which is self-generated. The outcomes indicate that the KNN classifier is better than CART and SVM. The accuracy of KNN, CART, and SVM are 0.96, 0.95, and 0.94 respectively. They suggested that their work can be used to find applications in the 5th Generation of mobile communication where user retention is important. On the other hand, **Vijaya et.al (2019)** proposed feature selecting techniques by using two different datasets from French Telecommunication orange company KDD Cup (2009) and implemented some pre-processing data cleaning and random oversampling. They used DT, KNN, SVM, NB, and ANN. The outcome indicates that SVM gains higher accuracy of 91.66% with imbalanced data and KNN gains 93.9% with random oversampling. They had suggested in the future more advanced methodology is used for feature selection.

Zhen et.al (2012) Proposed Hierarchical Multiple Kernel Support Vector Machine (H-MK-SVM) for the customer churns using longitudinal behavioral data. They used three datasets extracted from AdventureWorks DW, real-world databases Foodmart 2000 and Telecom. They had implemented sampling and Normalization. The outcome of their Proposed was H-MK-SVM shows Superior performance on both imbalanced and balanced data as compare to SVM and MK-SVM. The accuracy of H-MK-SVM, SVM and MK-SVM, on balanced data is 76.31%, 70.0% and 70.65%, respectively. While XGBoost method is used by **Atallah et.al (2020)** for customer retention. They used the self-generated dataset of 5000 subscribers and perform some sampling methods oversampling, ADASYN, SMOTE, and Borderline –SMOTE. The outcomes of their research were oversampling method improved the performance of Gradient Boosted Trees 84% by SMOTE oversampling of ratio 20%. On the other hand, **Ahmet et.al (2020)** used a machine algorithm for user behavior prediction. They have used the self-generated dataset of 50 features and apply pre-processing feature encoding, feature extraction, pseudonymization, missing feature, and normalization. The outcomes of their work that the Gradient boosting algorithm performs better than other algorithms.

Machine Learning methods DT, Gradient boosted Machine Tree (GBM), RF, and Extreme Gradient boosting (XGBoost) used by **Abdelrahim et.al (2019)** for prediction of churn in Telecom. They had used SyriaTel telecom company dataset and apply some pre-processing techniques oversampling and Undersampling. The result of their work is that XGBoost gives a better performance of 93%. On the other hand, **Afaq et.al (2010)** implementing data mining in ISP for customer churn. They used a dataset from Spenta Co. and utilized some pre-processing techniques feature extraction and modeling. They uphold some methods DT, LR and NN algorithm. The outcome indicated that their model achieved an accuracy of 89.08% in churn prediction rate by using feed-forward Neural Networks. while, **Vafeiadis et.al (2015)** analyzed the comparison of machine learning by using ANN, SVMs, DTs, Naïve Bayes, and LR and they used a public domain dataset. The outcomes indicate that support machine classifier poly with AdaBoost gain accuracy of 97% and F-measure over 84%.

Kiansing et.al (2001) proposed customer retention via data mining they used dataset from a transactional database residing in Oracle and performed deviation analysis and feature selection techniques on the dataset. They had utilized DT induction. This model gives the proposal that data mining gives a good consequence for customer retention. This Model further demonstrates that the context of particle application of data mining is much an art. On the other hand, **Ridwan et.al (2015)** operated NN Analysis, Multiple Regression Analysis, and LR Analysis for the customer churn prediction and implemented feature extraction based on nine-month of the bills and normalization on the dataset that is extracted from the warehouse of a Telecommunication company. The outcome of their experiment is that NN gain the best accuracy of 91.28%.

Data mining techniques DT and CART (Classification and regression Tree) used by **Chitra et.al (2011)** for the customer retention of the bank. They had used the bank dataset and implemented data reduction on the given dataset. The outcome of their work that CART (Classification and regression Tree) gives the overall greater classification rate. On the other hand, RF and Regression Forest used by **Bart et.al (2005)** for customer retention and profitability. They make use of a dataset from warehouse of a large European financial Service Company and applied data validation and sampling on it. They conclude their model provided a better estimate and validation sample as a match to the regression model.

Preeti et.al (2016) used LR and DT in the Telecom industry for the customer churn prediction and they utilized the dataset from Telecom Company and applied data cleaning for making the system robust and feature extraction for the generation rules for the DT and estimation of parameters in LR The outcome of their work was by using both DT and LR is best to design customer retention and it will also easy to design or maintain the customer with a high probability to churn. On the other hand, UCI database of the University of California home telecommunication dataset was used and some preprocessing techniques data acquisition and oversampling also applied on it by **XIA et.al (2008)** for customer churn prediction. They had utilized SVM, ANN, DT, Naïve Bayesian Classifier, and LR, the outcome of their work were SVM give the better accuracy rat, strong generation ability, and good fitting Precision.

Yu Zhao at. el (2005) presented improved one-class SVM for customer churn and used the subscriber database provided by the Oracle dataset. They compare their model with traditional methods DT, ANN, and Naïve Bayesian Classifier and the accuracy rate of their comparison SVM, ANN, DT, and Naïve Bayesian Classifier gain were 78.1%, 62%, 83.24%, and 87.15% respectively. Their Model shows a better accuracy rate than the other. They suggested that more research be done on how to select the appropriate kernel parameters and input features to get accurate results. While, **Mohammed al. el (2015)** used a dataset of UK mobile Telecommunication operator data of warehouse and they applied some data preparation processes discretisation of numerical variables, imputation of missing values, transformation from one set off discrete values to another, new variable derivation, and feature selection of the most informative variables. They had utilized LR and DT and the outcome of their work DT was Preferable for the customer churn and accuracy rate of DT was 70% and LR was 68%.

Abbas et.al (2016) used Electronic bank customer data from the bank's database and applied it to preprocess data cleaning, feature selection, and sampling on it. They used DT and the accuracy of DT was 99.70%. on the other hand, Applications of AdaBoost (Real, General, and Modest) utilized by **Shao et.al (2007)** for the churn prediction and they used a dataset provided by anonymous bank in China. They implemented some preprocessing handling of missing values and sampling. As a result of their model, these algorithms are proven improved for the predicting accuracy than the other algorithm.

Yaya et.al (2009) used improved balanced RF for churn prediction and they used the real bank customer dataset. They applied a sampling technique to the given dataset. The outcome of their research IBRF produced better accuracy than the other RF algorithm (balanced and weighted random forest). It offered great potential due to scalability, running speed, and faster training. On the other hand, **Lonut et.al (2016)** presented churn prediction for pre-paid mobile industry by using Neural Network, SVM, and Bayesian Network. They used the dataset of pre-paid mobile telecommunication industry and implemented data modeling on it. The outcome of their model was overall accuracy of 99.10% for Bayesian Network, 99.55% for NN, and 99.70% for SVM.

Suban et.al (2016) presented customer retention of MCDR by using data mining approaches Naïve Bayes, Radial Basis Function Network, Random Tree, and J48 Algorithm. They used two different datasets Dataset PAKDD 2006 available in "dat" and dataset of 3G network by an Asian Telco operator. They had applied some pre-processing handling missing values and Chi-square feature selection on the given datasets. The outcome of their research was Random Tree with three staged classifiers Chi2 gives the high accuracy rate of 87.67%. while, **LI-SHANG et.al (2006)** proposed knowledge discovery for customer churn by using DT and Logistic Regression. They utilized had customer dataset from the data warehouse and implemented sample extraction on it. The outcome of their model was DT performed better than the Logistic Regression. Similarly, **Ali et.al (2004)** proposed customer churn and retention by using Simple DT, DT with cost-sensitive DT, Boosting, and Logistic Regression. They used the customer dataset of major Australian online fast-moving consumer goods and implemented sampling. The outcome of their proposed model was AUC measurement shows different performance by the area under the ROC curve are 0.83, 0.91, 0.85, and 0.92 for Cost-Sensitive DT, LR, Simple DT and Boosting respectively.

Hongmei et.al (2008) Proposed GA- based Naïve Bayesian Classifier for the prediction of customer retention. They used dataset from Japanese credit debit and credit companies. They used a genetic algorithm for feature selection and the outcomes was compared with NB, TAN, and APRI classifier and indicate GA-based identified customer churn is better for a large number of customers shows higher classifying precision. On the other hand, **Ali et.al (2015)** predicted customer churn by using Echo State Network (ESN) and SVM. They used two different datasets one KDD dataset and the other from a Jordanian cellular telecommunication company. The outcome of their work was accuracy rate for the KDD dataset of ESN with SVM-readout was 93.7% and 87.9% and for the dataset from Jordain cellular telecommunication company was 99.2% and 96.8% respectively.

Zaho et.al (2008) presented a support vector machine for churn prediction and compares their result with the DT (C4.5), Logistic regression, ANN, and Naïve Bayesian classifier. They used a dataset of VIP customer's domestic branch of CCB as the core data and implemented sampling. The output of their work SVM with the higher accuracy rate of 0.5974, 0.5148 for C4.5, 0.5890 for LR, 0.5549 for Naïve Bayesian Classifier, and 0.5479 for ANN. On the other hand, **Shin et.al (2006)** proposed telecom churn management by using DT and Neural networks. They utilized random sampling on a dataset gained from wireless telecom companies in Taiwan. The output of their work was Both DT and NN give accurate churn prediction. Similarly, **Michael et.al (2000)** predicted improving retention and subscriber dissatisfaction by using LR, DT, NN, and Boosting. They used the dataset provided by the wireless carriers and utilized the outcome of their prediction was NN gives better nonlinear structure in the sophisticated representation than the DT, LR, and Boosting.

Jae et.al (2005) detected the change in customer behavior by using DT and used Dataset from a Korean online shopping mall. They had utilized data cleaning and discretization. The output of their detection was DT based methodology can be used to detect the changes of customer behavior from sales data and customer profile at different times. They suggested in future, this methodology can be extended to discover changes in customer behavior for three or more dataset. On the other hand, **Kristof et.al (2017)** proposed customer churn prediction by using LR and used Dataset provided by a large European mobile telecommunication provider. They implemented data cleaning, data reduction, sampling, missing value handling, and outlier. The outcome of their proposed was LR is a more advanced and assembled algorithm. It correctly prepares data generally less cumbersome than other algorithms.

Aurelie (2006) proposed bagging and boosting classification and used a dataset provided by Teradata Center at Duke University trees to predict customer churn. They used oversampling and the outcome of their proposed as boosting and bagging provided better classifiers than a binary logistic model. Predicted churn performance gain 16% of gini coefficient and 26% for top -docile lift. Both are provided good diagnostic measures, partial dependence plots, and variable importance. On the other hand, **Dudyala et.al (2008)** predicted credit card customer churn by using Multilayer Perceptron (MLP), DT (J48), LR, RF, SVM, and Radial Basis Function (RBF) Network. They performed oversampling, Undersampling, and synthetic minority oversampling on dataset from Latin American bank. The output of their prediction model gives the best result for under and oversampling and when original data is synthetic minority oversampling. Synthetic minority oversampling produced excellent results with 91.90% overall accuracy.

Kristof et.al (2006) predicted churn in subscription services by using the application of SVM and compare parameters with LR and random forest. They used dataset from a Belgian newspaper publishing company and implemented randomly Undersampling. The outcome of their predicted are SVM display fine performance when applied to a new, noisy marketing dataset and the accuracy for a real test set of SVM(SVMacc and SVMauc) are 84.90% and 85.14 respectively, LR is 84.60% and are 87.21%. They suggested deriving a solution to select the correct kernel and parameter value according to the problem is an interesting topic for future research.

Anuj et.al (2011) predicted customer churn in cellular network services by using NN and used dataset from the UCI repository database at the University of California. The output of their prediction was NN can predict the customer churn with an accuracy of 92.35%. They suggested we should propose Pre-processing and we also implemented deep methods. On the other hand, **Weiyun et.al (2008)** proposed a model for the prevention of customer churn and used dataset provided by Chines Bank. They implemented sampling on the given dataset. They had utilized improved balance RF and compare their result with DT, ANN, and CWC-SVM. The outcome of their proposal was improved balance RF gives highest accuracy of 93.4% and DT, ANN and CWC-SVM gives 62%, 78.12%, and 87.15% respectively.

Thomas et.al (2014) presented profit optimizing customer churn by using Bayesian network classifiers (Naïve Byes Classifier, Bayesian Network, Augmented Naive Bayes classifiers, Search-and-score algorithms, General Bayesian network classifiers, Constraint-based algorithms, and Hybrid methods). They used four real-time datasets three datasets from the center for customer relationship management at Duke University, one from a European telecommunication operator, and one synthetic dataset. They used some preprocessing techniques feature selection (to limit the available attributes and Markov Blanket based algorithm for feature selection is used) and Undersampling (removing non-churn from the dataset). The outcome of their work was classification performance was measured area under the receiver operating characteristic curve (AUC) and the maximum profit (MP) criterion and both give a different ranking of the classification algorithm. The Naive Bayes methods do not lead to compatible networks, while the General Bayesian Network algorithms lead to simpler and more flexible networks and the Bayesian Network dividers cannot bypass traditional planning. On the other hand, **Koh et.al (2019)** predicted customer churn by using particle swarm optimization and extreme Learning Machine. They used the telecommunication dataset from Kaggle and implemented random oversampling and feature scaling. The outcome of their model retention plan can be built based on these features to reduce the churn rate from a company. The non-scaled feature gives training accuracy of 50.03% and testing accuracy of 49.95% similarly scaled feature gives training accuracy of 84.71% and testing accuracy of 81.16%. PSO and ELM determined high testing accuracy.

Author/ Year	Dataset	Pre-processing	Methods	Result	Future Work.
Year 2020 Mr. V Naveen Kumar Mr. G Ramesh Babu Mr. A Ravi Kishore	➢ cell2cell dataset ➢ IBM dataset	➢ 30 Predicted variables	➢ Naive Bayes ➢ SVM ➢ decision tree (DT)	Outcome measured by (AUC) and achieved 0.82, 0.87 and 0.77 for IBM and 0.98, 0.99 and 0.98 for the cell2cell dataset.	
Year (2020) Mohammed Al-Mashraiea Hyun Woo Jeonb Sung Hoon Chunga	➢ US-based telecommunication company	➢ data cleaning ➢ data balancing	➢ logistic regression (LR) ➢ vector machines ➢ random forest (RF) ➢ decision tree (DT)	Predicted performance of their work is greater than 86% and the LR has the highest accuracy rate.	
Year 2020 Hemlata Dalmia CH V S S Nikil Sandeep Kumar	➢ Dataset acquired from an online source	➢ data Acquisition ➢ data Cleaning	➢ XGBoost Boosting ➢ XGBoost Algorithm ➢ K Nearest Neighbors (KNN)	KNN gains 83.85% and XGBoost gains 86.85% accuracy	

year 2020 Shreyas Rajesh Labhsetwar	➢ BigML churn Telecom dataset	➢ Handling Missing Values ➢ Encoding ➢ Categorical Data ➢ Feature Scaling	➢ Extra Trees Classifier ➢ XGBoosting Algorithm ➢ SVM ➢ SGD Classifier ➢ Adaboost Classifier ➢ Gaussian Naive Bayes ➢ LR	Extra Tree classifier, XGBoosting Algorithm, and SVM with AUC Score (0.8439, 0.78, and 0.735) respectively.	
Year 2018 Sahar F. Sabbe	➢ Public dataset of customer in Telecom industry ➢ 60% of training & 40% testing dataset	➢ Data transformation ➢ Data cleaning ➢ Feature selection	➢ Decision Tree (CART) ➢ SVM ➢ KNN ➢ Adaboost ➢ Random forest ➢ Stochastic gradient boost ➢ MLP ANN	RF and Adaboost with same accuracy of 96%. Multilayer perceptron and SVM give an accuracy of 94%. DT gives 90%, Naive Bayesian 88%, and LR and LDA give 86.70%.	In future this research is extended by including deep learning and hybrid model.
Year 2018 Nagraj V. Dharwadka and Priyanka S. Patil	➢ German Credit dataset	➢ Normalization	➢ Support Vector Machine (SVM) ➢ Deep Neural Network (DNN) ➢ Artificial Neural Network (ANN)	accuracy of 98%, 92%, and 97% for ANN, SVM, and DNN respectively and for bank customer data and 72%, 76% and 72%, respectively for the credit dataset of German.	
Year 2019 Pushpa Singh Vishwas Agrawal	➢ Self generated customer dataset	➢ Normalization	➢ KNN ➢ CART ➢ SVM	KNN classifier is better than CART and SVM. The accuracy of KNN, CART, and SVM are 0.96, 0.95, and	In future this work can be used to find applications in fifth Generation (5G) of

				0.94 respectively.	Mobile communication where user retention is importance.
Year 2019 E.Sivasankar J. Vijaya	➢ Dataset from French Telecommunication orange company KDD Cup (2009)	➢ Data cleaning ➢ Random Oversampling	➢ DT ➢ KNN ➢ SVM ➢ NB ➢ ANN	SVM gains higher accuracy of 91.66% with imbalanced data and KNN gains 93.9% with random oversampling.	In future more advanced methods are used for feature selection techniques.
Year 2012 Zhen-Yu Chen Minghe Sun Zhi-Ping Fan	Three real-word databases datasets ➢ Foodmart 2000 ➢ Adventure Works ➢ DW Telecom	➢ Sampling ➢ Normalization	➢ Hierarchical Multiple Kernel Support Vector Machine (H-MK-SVM) ➢ SVM ➢ Multiple Kernel Support Vector Machine (MK-SVM)	Accuracy of H-MK-SVM, SVM, and MK-SVM on balanced data are 76.31%, 70.0% , and 70.65% respectively.	
Year 2020 Atallah M. AL-Shatnwai Mohammad Faris	➢ Self generated dataset consist the information of 5000 subscribers	➢ Over Sampling ➢ Random -oversampling ➢ SMOTE ➢ ADASYN ➢ Border line -SMOTE	➢ XGBoost (Gradient Boosted Trees Algorithm)	Performance of Gradient Boosted Trees 84% by SMOTE oversampling of ratio 20%.	
Year 2020 Ahmet Turkmen Cenk Anil Bahcevan Youssef Alkhanafseh Esra Karabiyik	➢ Self generated dataset consist of 35 features	➢ Feature encoding ➢ Feature extraction ➢ Pseudonymization ➢ Missing-feature ➢ Normalization	➢ Machine leaning algorithm	The outcomes of their work that the Gradient boosting algorithm performs better than other algorithms.	
Year 2019	➢ SyriaTel tele-	➢ Over sampling	➢ RF	XGBoost gives a	

Abdelrahim Kasem Ahmad Assef Jafar Kadan Aljoumaa	com company dataset	➢ Under sampling	➢ Gradient boosted Machine Tree (GBM) ➢ DT ➢ Extreme Gradient boosting (XGBoost)	better performance of 93%.	
Year 2010 Afaq Alam Khan M.M Sepehri Sanjay Jamwal	➢ Dataset from spenta co.	➢ Feature extraction ➢ Modeling	➢ DT ➢ LR ➢ Neural Network Algorithm (NN)	Accuracy of 89.08% in churn prediction rate by using feed-forward Neural Networks.	
Year 2015 K.I. Diamantaras G. Sarigiannidis K.Ch. Chatzisavvas T. Vafeiadis	➢ Public domain dataset	-	➢ ANN ➢ SVM ➢ DT ➢ Naïve Bayes ➢ LR	Support Machine classifier Poly with AdaBoost gain accuracy of 97% and F-measure over 84%.	
Year 2001 Kiansing NG Huan Liu	➢ Dataset from transactional database residing in Oracle.	➢ Deviation analysis ➢ Feature selection	➢ Decision Tree induction	Data mining gives a good consequence for customer retention.	
Year 2015 M.ridwan ismail M.Nordin a rehman M.Khalid Awang Mokhairi Makhtar	➢ Warehouse of a telecommunication company	➢ Feature Extraction ➢ Normalization	➢ Neural Network Analysis ➢ Multiple Regression Analysis ➢ Logistic Regression Analysis	NN gain the best accuracy of 91.28%.	
Year 2011 K. Chitra B. Subashini	➢ Bank dataset	➢ Data reduction	➢ DT ➢ CART(Classification and regression Tree)	The outcome of their work that CART (Classification and regression Tree) gives the overall greater classification rate.	
Year 2005 Dirk Van den Poel	➢ Warehouse of a large European	➢ Data validation ➢ Data sampling	➢ Random Forest (RF)	They conclude their Model pro-	

Bart Lariviere	financial Service Company		➢ Regression Forest	vided a better estimate and validation sample as a match to the regression model.	
Year 2016 Siddhi K. Khandge Preeti k.Dalvi Aditya Bankar Ashish Deomore Prof V.A Kanade	➢ Telecom company dataset	➢ Data cleaning ➢ Feature extraction	➢ Logistic Regression (LR) ➢ DT	DT and LR is best to design customer retention and it will also easy to design or maintain the customer with a high probability to churn.	
Year 2008 XIA Guo-en JIN Wei-dong	➢ UCI database of University of California Home Telecommunication dataset	➢ Data acquisition ➢ Over-sampling	➢ SVM ➢ Artificial Neural Network (ANN) ➢ DT ➢ LR ➢ Naïve Bayesian Classifier	the outcome of their work were SVM give the better accuracy rat, strong generation ability, and good fitting Precision.	
Year 2005 Yu Zhao, Bing Li Xiu Li Wenhuang Liu Shouju Ren	➢ Subscriber database provided by Oracle dataset	-	➢ ANN ➢ DT ➢ Naïve Bayesian Classifier ➢ SVM	SVM, ANN, DT, and Naïve Bayesian Classifier gain were 78.1%, 62%, 83.24%, and 87.15% respectively.	They suggested that more research be done on how to select the appropriate kernel parameters and input features to get accurate results.
Year 2015 Mohammed Hassouna Tariq Elyas Mohammad Saeed	➢ UK mobile Telecommunication operator data warehouse	➢ Imputation of missing values ➢ Discretisation of Numerical variables	➢ Logistic Regression ➢ Decision Tree	the outcome of their work DT was Preferable for the customer churn and accuracy rate	

AbouTrab Ali Tarhini		➢ Feature Selection of the most informative variables ➢ New Variable derivation ➢ Transformation from one set off discrete values to another		of DT was 70% and LR was 68%.	
Year 2016 Abbas Keramati Hajar Ghaneei Seyed Mohammad Mirmohammadi	➢ Electronic bank customer data from bank's database	➢ Data Cleaning ➢ Feature Selection ➢ Sampling	➢ Decision Tree	They used DT and the accuracy of DT was 99.70%.	
Year 2007 Shao Jinbol Li Xiu Liu Wenhuang	➢ Dataset Provided by Anonymous bank in China	➢ Handling of missing values ➢ Sampling	AdaBoost ➢ real-AdaBoost, ➢ Gental AdaBoost ➢ Modest AdaBoost	These algorithms are proven improved for the predicting accuracy than the other algorithm.	
Year 2009 Xiu Li Yaya Xie Weiyun Ying E.W.T. Ngai	➢ Real bank customer dataset	➢ Sampling	➢ Improved balanced Random Forest	IBRF produced better accuracy than the other RF alogorithm (balanced and weighted random forest).	
Year 2016 Ionuț Brândușoiu Gavril Toderean Horia Beleiu	➢ Dataset of prepaid mobile telecommunication industry	➢ Data Modeling	➢ Neural Network ➢ SVM ➢ Bayesian Network	Accuracy of 99.10% for Bayesian Network, 99.55% for NN, 99.70% for SVM.	
Year 2016 Suban Ravichandran Chandrasekaran Ramasamy	➢ Dataset PAKDD 2006 available in "dat" ➢ Dataset of 3G	➢ Handling Missing Values ➢ Chi-square	➢ Naïve Bayes ➢ Radial Basis Function Network	Random Tree with three staged classifiers Chi2 gives the high accuracy	

	Network by an Asian Telco Operator	➢ Feature Selection	➢ Random Tree ➢ J48 Algorithm	rate of 87.67%.	
Year 2006 LI-SHANG YANG CHAOCHANG CHIU	➢ Customer data-set from data warehouse	➢ Sample extrac-tion	➢ Decision Tree (DT) ➢ Logistic Regres-sion	The outcome of their model was DT Performed better than the Logistic Regression.	
Year 2014 Stanislav Stakhovych Michael Ewing Ali Tamaddoni Ja-hromi	➢ Customer data-set of a major Australian on-line Fast mov-ing consumer Goods	➢ Sampling	➢ Decision Tree with Cost Sensi-tive ➢ DT ➢ Simple Decision Tree ➢ Boosting ➢ Logistic Regres-sion	performance by the area under the ROC curve are 0.83, 0.91, 0.85, and 0.92 for Cost-Sensitive DT, LR, Simple DT and Boosting respec-tively.	
Year 2008 Hongmei Shao Gaofeng Zheng Fengxian An	➢ Dataset from Japanese credit debit and credit company	➢ GA algorithm for feature selection	➢ GA- based Novel Bayesian Classifi-er	outcome was compared with NB, TAN, and APRI classifier and indi-cate GA-based identified custom-er churn is better for a large number of customers shows higher clas-sifying Precision.	
Year 2015 Ali Rodan Hossam Faris	➢ Dataset from Jordanian cellu-lar telecommu-nication com-pany. ➢ KDD dataset	-	➢ Echo State Net-work (ESN) ➢ SVM	accuracy rate for the KDD dataset of ESN with SVM-readout was 93.7% and 87.9% and for the dataset from Jordan cellular telecommunication company was	

				99.2% and 96.8% respectively.	
Year 2008 Zhao Jing Dang Xing-hua	➢ Dataset of VIP customers domestic branch of CCB as the core data	➢ Sampling	➢ SVM ➢ Logistic regression (LR) ➢ Decision Tree (C4.5) ➢ ANN ➢ Naïve Bayesian classifier	SVM with higher accuracy rate 0.5974, 0.5148 for C4.5, 0.5890 for LR, 0.5549 for Naïve Bayesian Classifier, and 0.5479 for ANN.	
Year 2006 Shin-Yuan Hung David C. Yen Hsiu-Yu Wang	➢ Dataset from wireless telecom company in Taiwan	➢ Random Sampling	➢ Decision Tree ➢ Neural Network (NN)	The output of their work was Both DT and NN give accurate churn prediction.	
Year 2000 Michael C. Mozer Richard Wolniewicz David B. Grimes Eric Johnson Howard Kaushansky	➢ Dataset Provided by wireless carrier	-	➢ Logistic regression ➢ DT ➢ Neural Network ➢ Boosting	NN gives better nonlinear structure in the sophisticated representation than the LR, DT, and Boosting.	
Year 2005 Hee Seok Song Hyea Kyeong Kim Jae Kyeong Kim Tae Seong Kim	➢ Dataset from Korean online Shopping mall.	➢ Data cleaning ➢ Discretization	➢ Decision Tree	DT based methodology can be used to detect the changes of customer behavior from sales and customer profile data at different times.	In Future this methodology can be extended to discover changes of customer behavior for three or more dataset.
Year 2017 Kristof Coussement Stefan Lessmann Geert Verstraeten	➢ Dataset provided by a large European mobile telecommunication provider	➢ Data cleaning ➢ Data reduction ➢ Sampling ➢ Missing value handling ➢ outlier	➢ Logistic Regression	LR is a more advanced and assembled algorithm. It correctly prepares data generally less cumbersome than other	

				algorithms.	
Year 2006 Aurelie lemmens Christophe croux	➢ Dataset provided by Teradata Center at Duke University	➢ Over sampling	➢ Bagging ➢ boosting	Boosting and Bagging Provided better classifiers than a binary logistic model. Predicted churn performance gain 26% for top-decile lift and 16% of Gini coefficient .	
Year 2008 Dudyala Anil Kumar V. Ravi	➢ Dataset from Latin American Bank.	➢ Over Sampling ➢ Under Sampling ➢ Synthetic Minority Oversampling	➢ Multilayer Perceptron (MLP) ➢ Decision Tree (J48) ➢ Radial Basis Function (RBF) Network ➢ Random Forest (RF) ➢ Support Vector Machine ➢ Logistic Regression (LR)	Synthetic Minority Oversampling as well for under and oversampling. Synthetic Minority oversampling produced excellent results with 91.90% overall accuracy.	
Year 2006 Kristof Coussement Dirk Van den Poel	➢ Dataset from Belgian Newspaper publishing company	➢ Randomly Undersampling	➢ Support Vector Machine	accuracy for a real test set of SVM(SVMacc and SVMauc) are 84.90% and 85.14 respectively, LR is 84.60% and are 87.21%.	Deriving a solution to select correct kernel and parameter value according the problem is an interesting topic for the future research.
Year 2011 Anuj Sharma Dr. Prabin Kumar Panigrahi	➢ Dataset from UCI Repository database at the university of	-	➢ Neural Network	The output of their Prediction was NN can Predict the customer churn	They suggested we should proposed Pre-processing and

	California.			with an accuracy of 92.35%.	we also implemented deep methods.
Year 2008 Ellis Johnson Weiyun Ying Yaya Xie Xiu Li	➢ Dataset Provided by Chines Bank	➢ Sampling	➢ Improved balance random forest	Improved balance(RF) gives highest accuracy of 93.4% and DT, ANN and CWC-SVM gives 62%, 78.12%, and 87.15% respectively.	
Year 2014 Thomas Verbrakena Wouter Verbekea Bart Baesensa,	➢ Dataset from European telecommunication operator ➢ Three datasets from Center for Customer Relationship Management at Duke University ➢ Synthetic dataset	➢ Undersampling (removing non churn from the dataset) ➢ Feature selection (to limit the available attributes and Markov Blanket based algorithm for feature selection is used)	➢ Bayesian Network ➢ Naïve Byes Classifier ➢ Augmented Naive Bayes classifiers ➢ General Bayesian network classifiers ➢ Search-and-score algorithms ➢ Constraint based algorithms ➢ Hybrid methods	classification performance was measured area under the receiver operating characteristic curve (AUC) and the maximum profit (MP) criterion and both give a different ranking of the classification algorithm.	
Year 2019 Koh Guan Li P. M. Booma	➢ Telecom churn dataset from kaggle	➢ Random oversampling ➢ Feature Scaling (Particle swarm Optimization)	➢ Extreme Learning Machine ➢ Particle swarm Optimization (PSO) algorithm	Testing accuracy of 49% and Training accuracy of 50% and also Scale feature Training accuracy 84% and Test accuracy of 81.16%.	

Unsupervised

Musadig et.al (2020) in the field of banking for the retention of their customers used unsupervised Machine Learning Methodology

(K-mean clustering algorithm) and they also utilized the real-time dataset of a bank. They had utilized some Pre-processing techniques data reduction, data cleaning, data transformation and normalization. Conclude of their work is decided to avoid this model due to high computational complexity and by using the different clustering algorithm which generates similar outcomes. They suggested methodology used in this paper dilated, so that it applied to the other behaviors of bank customers, such as a behavior based on deposits, loans, investments, etc. in addition to the ones extracted from the transactions history. On the other hand, **Priyanka et.al (2017)** used Artificial Neural Network for the analysis of banking data. They used two different datasets German credit dataset from UCI and the dataset of a bank customers and implemented threshold. The outcome of their work was ANN gives accuracy for dataset 1 and dataset 2 are 72% and 98% respectively. The proposed Model works efficiently for both datasets.

Author/ Year	**Dataset**	**Pre-processing**	**Methods**	**Result**	**Future Work.**
Year 2020 Musadig Aliyev Elvin Ahmadov- Sayel Abualigah Habil Gadirli Arzu Mammadova Emin Alasgarov	➢ Real time dataset of a bank	➢ data Reduction ➢ data cleaning ➢ data Transformation ➢ Normalization	➢ K-means clustering algorithms	They decided to avoid this model due to high computational complexity and by using the different clustering algorithm which generates similar outcomes.	it applied to the other behaviors of bank customers, such as a behavior based on deposits, loans, investments, etc. in addition to the ones extracted from the transactions history.
Year 2017 Priyanka S. Patil Nagaraj V. Dharwadkar	➢ German credit dataset from UCI. ➢ Dataset of a bank customers	➢ Threshold	➢ Artificial Neural Network	ANN gives accuracy for dataset 1 and dataset 2 are 72% and 98% respectively. The proposed Model works efficiently for both datasets.	

Hybrid

Jina et.al (2021) offers a deep hybrid learning model for customer repurchase behavior for smart phone company. They used customer dataset set which gain by the two survey of a company in December 2017 and in December 2019. The outcome of their model for the customer purchase behavior gains the accuracy of 90.71%. On the other hand, **Saad et.al (2013)** offers a model for the churn prediction by using Regression Analysis, DT, ANN, and K-mean clustering and they used a dataset from the customer DNA website.

They applied re-sampling and data acquisition on the given dataset. The follow-up of their work is that DT is the most accurate algorithm to identified customer churn and the accuracy achieved by DT is 70%.
Jorge et.al (2004) used Neural Network, DT, Neuron Fuzzy, and (GA) rule evolver for the evaluation of wireless churn. They used dataset from major wireless carrier operating in Brazil and implemented oversampling technique. The output of their result was NN with 15 hidden units that accomplish the best classification. On the other hand, **Duyen et.al (2017)** used AdaBoost, KNN, Extra Trees, NN, and XGBoost for the customer churn in ISP. They used a dataset of ISP companies in Vietnam and implemented oversampling. The outcome of their work was XGBoost gives the better result with the precision and recall of 45.71% and 42.06% respectively. Similarly, **S.Ummugulthum et.al (2014)** proposed customer relationship management by using Naïve Bayes, SVM, KNN, and J48. They used a dataset of KDD Cup 2009 and implemented feature selection by Information gain values. Information gain measures the amount of information for split over a feature. The output of their work was accuracy rate of Naïve Bayes, SVM, KNN and J48 are 93.78%, 98.84%, 98.25, and 98.78 respectively. SVM gives the highest accuracy.

Author/ Year	Dataset	Pre-processing	Methods	Result	Future Work.
Year 2021 Angel P. del Pobil HongGeun Ji Soyoung Oh Syjung Hwang Jina Kim Eunil Park	➢ Customer dataset of a smart phone company	-	➢ Deep hybrid learning model	Proposed model gain the accuracy of 90%.	
Year 2013 Aatif Kamal Ammar Saleem Rehman Ali mustafa Qamar Saad Ahmed Qureshi	➢ Dataset from Customer DNA Website	➢ Re-sampling ➢ Data acquisition	➢ K-mean Clustering ➢ Decision Tree ➢ Artificial Neural-Networks ➢ Regression Analysis	DT is the most accurate algorithm to identified customer churn and the accuracy achieved by DT is 70%.	
Year 2004 Jorge B. Ferreira Marley Vellasco Marco Aurélio Pacheco Carlos Hall Barbosa	➢ Dataset from Major wireless carrier operating in Brazil	➢ Oversampling	➢ Decision Tree ➢ Neuroz Fuzzy ➢ (GA) Rule Evolver ➢ Neural Network	The output of their result was NN with 15 hidden units that accomplish the best classification.	
Year 2017 Duyen DO Phuc HUYNH Phuong VO Tu VU	➢ Dataset of ISP companies in Vietnam.	➢ Over Sampling	➢ XGBoost ➢ Extra Trees ➢ KNN ➢ AdaBoost ➢ Neural Net-	XGBoost gives the better result with the recall and Precision of 42.06% and 45.71% re-	

			work	spectively.	
Year 2014 S.Ummugulthum Natchiar Dr.S.Baulkani	➢ KDD Cup 2009	➢ Feature Selection	➢ Naïve Bayes ➢ Support Vector Machine ➢ KNN ➢ J48	Accuracy rate of Naïve Bayes, SVM, KNN and J48 are 93.78%, 98.84%, 98.25, and 98.78 respectively.	

LIMITATIONS:

In previous research we have found some limitations such as lack of unsupervised learning methods being has been for customer retention. We have also struggled to find the review on deep learning methods and hybrid models based of feature selection methods.

In future we suggest that there is a need to develop hybrid model based on feature selection methods to handle the large customer's datasets. Feature selection methods can help in finding the important features without compromising on the quality of result. The combination of feature selection methods and deep learning can produce better results.

Conclusion

Customer retention has become more important topic for the business now days. To gain more customers is very important for the success of business. Customer data is imbalanced and huge that is difficult to handle manually. For the prediction of customer retention and handling of imbalanced data many paper has been reviewed in this paper. The limitations shows lack of supervised and hybrid models used for customer retention. There is a need to develop hybrid model based on feature selection methods. The combination of feature selection methods and deep learning can produce better result.

References

[1] Kumar, M. V. N., Babu, M. G. R., & Kishore, M. A. R. An Enhanced Methodology to predict Customer Churn and Retention model using Machine Learning.

[2] Al-Mashraie, M., Chung, S. H., & Jeon, H. W. (2020). Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers & Industrial Engineering*, *144*, 106476.

[3] Dalmia, H., Nikil, C. V., & Kumar, S. (2020). Churning of Bank Customers Using Supervised Learning. In *Innovations in Electronics and Communication Engineering* (pp. 681-691). Springer, Singapore.

[4] Labhsetwar, S. R. PREDICTIVE ANALYSIS OF CUSTOMER CHURN IN TELECOM INDUSTRY USING SUPERVISED LEARNING.

[5] Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. *International Journal of advanced computer Science and applications*, *9*(2).

[6] Dharwadkar, N. V., & Patil, P. S. (2018). Customer retention and credit risk analysis using ANN, SVM and DNN. *International Journal of Society Systems Science*, *10*(4), 316-332.

[7] Singh, P., & Agrawal, V. (2019). A collaborative model for customer retention on user service experience. In *Advances in Computer Communication and Computational Sciences* (pp. 55-64). Springer, Singapore.

[8] Sivasankar, E., & Vijaya, J. (2019). A study of feature selection techniques for predicting customer retention in telecommunication sector. *International Journal of Business Information Systems*, *31*(1), 1-26.

[9] Chen, Z. Y., Fan, Z. P., & Sun, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of operational research*, *223*(2), 461-472.

[10] AL-Shatnwai, A. M., & Faris, M. Predicting Customer Retention using XGBoost and Balancing Methods.

[11] Turkmen, A., Bahcevan, C. A., Alkhanafseh, Y., & Karabiyik, E. (2020). User behaviour analysis and churn prediction in ISP. *New Trends and Issues Proceedings on Advances in Pure and Applied Sciences*, (12), 57-67.

[12] Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1), 1-24.

[13] Khan, A. A., Jamwal, S., & Sepehri, M. M. (2010). Applying data mining to customer churn prediction in an internet service provider. *International Journal of Computer Applications*, *9*(7), 8-14.

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

[15] Ng, K., & Liu, H. (2000). Customer retention via data mining. *Artificial Intelligence Review*, *14*(6), 569-590.

[16] Ismail, M. R., Awang, M. K., Rahman, M. N. A., & Makhtar, M. (2015). A multi-layer perceptron approach for customer churn prediction. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(7), 213-222.

[17] Chitra, K., & Subashini, B. (2011). Customer retention in banking sector using predictive data mining technique. In *ICIT 2011 The 5th International Conference on Information Technology*.

[18] Larivière, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert systems with applications*, *29*(2), 472-484.

[19] Dalvi, P. K., Khandge, S. K., Deomore, A., Bankar, A., & Kanade, V. A. (2016, March). Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)* (pp. 1-4). IEEE.

[20] Xia, G. E., & Jin, W. D. (2008). Model of customer churn prediction on support vector machine. *Systems Engineering-Theory & Practice*, *28*(1), 71-77.

[21] Zhao, Y., Li, B., Li, X., Liu, W., & Ren, S. (2005, July). Customer churn prediction using improved one-class support vector machine. In *International Conference on Advanced Data Mining and Applications* (pp. 300-306). Springer, Berlin, Heidelberg.

[22] Hassouna, M., Tarhini, A., Elyas, T., & AbouTrab, M. S. (2016). Customer churn in mobile markets a comparison of techniques. *arXiv preprint arXiv:1607.07792*.

[23] Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, *2*(1), 1-13.

[24] Jinbo, S., Xiu, L., & Wenhuang, L. (2007, June). The Application ofAdaBoost in Customer Churn Prediction. In *2007 International Conference on Service Systems and Service Management* (pp. 1-6). IEEE.

[25] Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, *36*(3), 5445-5449.

[26] Brânduşoiu, I., Toderean, G., & Beleiu, H. (2016, June). Methods for churn prediction in the pre-paid mobile telecommunications industry. In *2016 International conference on communications (COMM)* (pp. 97-100). IEEE.

[27] Ravichandran, S., & Ramasamy, C. (2016). Customer Retention of MCDR using Three-Stage Classifier based DM Approaches. *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, *5*, 11190-11196.

[28] Yang, L. S., & Chiu, C. (2006, November). Knowledge discovery on customer churn prediction. In *Procedings of the 10th WSEAS International Conference on Applied Mathematics, Dallas, Texas, USA, November 1-3, 2006, pp523* (Vol. 528).

[29] Jahromi, A. T., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, *43*(7), 1258-1268.

[30] Shao, H., Zheng, G., & An, F. (2008, October). Construction of Bayesian classifiers with GA for predicting customer retention. In *2008 Fourth International Conference on Natural Computation* (Vol. 1, pp. 181-185). IEEE.

[31] Rodan, A., & Faris, H. (2015, November). Echo state network with SVM-readout for customer churn prediction. In *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)* (pp. 1-5). IEEE.

[32] Zhao, J., & Dang, X. H. (2008, October). Bank customer churn prediction based on support vector machine: Taking a commercial bank's VIP customer churn as the example. In *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing* (pp. 1-4). IEEE.

[33] Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, *31*(3), 515-524.

[34] Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on neural networks*, *11*(3), 690-696.

[35] Kim, J. K., Song, H. S., Kim, T. S., & Kim, H. K. (2005). Detecting the change of customer behavior based on decision tree analysis. *Expert Systems*, *22*(4), 193-205.

[36] Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27-36.

[37] Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, *43*(2), 276-286.

[38] Anil Kumar, D., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. *International Journal of Data Analysis Techniques and Strategies*, *1*(1), 4-28.

[39] Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert systems with applications*, *34*(1), 313-327.

[40] Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert systems with applications*, *34*(1), 313-327.

[41] Ying, W., Li, X., Xie, Y., & Johnson, E. (2008, July). Preventing customer churn by using random forests modeling. In *2008 IEEE International Conference on Information Reuse and Integration* (pp. 429-434). IEEE.

[42] Verbraken, T., Verbeke, W., & Baesens, B. (2014). Profit optimizing customer churn prediction with Bayesian network classifiers. *Intelligent Data Analysis*, *18*(1), 3-24.

[43] Li, K. G., & Marikannan, B. P. (2019). Hybrid Particle Swarm Optimization-Extreme Learning Machine Algorithm for Customer Churn Prediction. *Journal of Computational and Theoretical Nanoscience*, *16*(8), 3432-3436.

[44] Aliyev, M., Ahmadov, E., Gadirli, H., Mammadova, A., & Alasgarov, E. (2020). Segmenting Bank Customers via RFM Model and Unsupervised Machine Learning. *arXiv preprint arXiv:2008.08662*.

[45] Patil, P. S., & Dharwadkar, N. V. (2017, February). Analysis of banking data using machine learning. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 876-881). IEEE.

[46] Kim, J., Ji, H., Oh, S., Hwang, S., Park, E., & del Pobil, A. P. (2021). A deep hybrid learning model for customer repurchase behavior. *Journal of Retailing and Consumer Services*, *59*, 102381.

[47] Ferreira, J., Vellasco, M. M., Pacheco, M. A. C., Carlos, R., & Barbosa, H. (2004, April). Data Mining Techniques on the Evaluation of Wireless Churn. In *ESANN* (Vol. 28, pp. 483-488).

[48] Do, D., Huynh, P., Vo, P., & Vu, T. (2017, December). Customer churn prediction in an internet service provider. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 3928-3933). IEEE.

[49] Natchiar, S. U., & Baulkani, S. (2014, November). Customer relationship management classification using data mining techniques. In *2014 International Conference on Science Engineering and Management Research (ICSEMR)* (pp. 1-5). IEEE.