Bulletin of Pure and Applied Sciences
Zoology (Animal Science), Vol.43B, No.2s,
July-December 2024: P. 315-330

Print version ISSN 0970-0765
Online version ISSN 2320-3188

Original Research Article
Content Available online at: www.bpasjournals.com

ANALYZING BANK CUSTOMER CHURN: A COMPARATIVE EXPLORATION OF MACHINE LEARNING MODELS AND MULTIMODAL FUSION

Baliboina Charith[1] K.Pushpalatha[2] Kotha Dinesh Sai[3] U M Gopal Krishna[3]

[1,3]VIT – AP School of Computer Science, VIT – AP University, Andhra Pradesh, India.
[2]Master of Business Administration, Malla Reddy Engineering College (Autonomous), Telangana, India.
[4]VIT – AP School of Business, VIT – AP University, Andhra Pradesh, India.
Corresponding Author: gopalakrishna.gks@gmail.com

Abstract—In today's competitive business landscape, organizations strive to enhance their quality of service (QoS) to meet the increasing demands of customers. Customer Relationship Management (CRM) systems play a vital role in acquiring new customers, establishing lasting relationships, and improving customer retention for sustained profitability. By leveraging machine learning models, In order to achieve a competitive advantage, customer relationship management systems are able to analyze the personal and behavioral data of customers accurately predicting customer churn and identifying the underlying reasons. This paper aims to predict customer churn, analyze the associated factors, and provide insights for improvement. These predictions enable organizations to design targeted marketing plans and service offerings.

Four separate analytic approaches from different types of learning are chosen for this study so that the performance of different machine learning techniques for churn prediction can be compared and analyzed. Ensemble-based (Random Forest) and multi-model (XgBoost) and machine-learning (SVM) methods. These techniques are applied to a dataset containing 10,000 records of Bank Customer Data.

The results of the analysis demonstrate that XgBoost achieve a high accuracy rate of 85% for customer churn prediction. Additionally, Random Forest, ANN, Support Vector Machines and Multimodal Algorithm exhibit promising accuracy rates of approximately 86% and 87% respectively. The findings highlight the effectiveness of these machine-learning techniques in predicting customer churn.

Furthermore, the analysis of the churn predictions reveals noteworthy insights. Specifically, it is observed that customers in the age group of 50-60 exhibit a higher churn rate compared to the retained customers. Additionally, if precautionary measures are not taken, this trend is expected to continue for the age group of 40-50. Moreover, the analysis indicates that customers with credit cards are more prone to churn compared to those without credit cards, emphasizing the importance of credit card-related strategies in reducing churn.

In conclusion, this research paper underscores the significance of accurate churn prediction for organizations in their quest to enhance QoS and improve customer retention. The comparative analysis of various machine learning techniques provides valuable insights into their performance, with Random Forest and Support Vector Machines showcasing superior accuracy rates. The identified trends related to age groups and credit card usage further

inform targeted marketing efforts and service enhancements to mitigate customer churn.

Keywords—Multimodal Approach, Bank Customer Churn, Comparative Study.

Introduction

The market as it exists today is highly competitive and extremely dynamic. This is due to the fact that there are many different service providers to choose from. Identifying the shifting behaviors of customers and meeting their ever-increasing expectations are two of the challenges that service providers must overcome. The current generation of consumers has significantly higher aspirations than previous generations of consumers, as well as diverse demands for increased connectivity and novel, individualized approaches. These demands are generations of consumers. They have a strong educational foundation and are up to date on cutting- edge practices. Because of this new information, they now shop differently, and as a result, there is a growing trend of "analysis paralysis," which refers to the practice of excessively analyzing the selling and buying scenario. This, in the end, helps them to make better decisions regarding their purchases. Therefore, it is a significant obstacle for the new generation of service providers to think creatively in order to satisfy their customers and add value to their offerings.

The banking industry was one of the industries that processed and stored enormous amounts of data. The cost of storing and processing data is escalating at an exponential rate as a result of the banking industry's increased adoption of developing digital technologies. Every bank stores a wide variety of customer information. These may consist of transaction data, account information, and financial data. As a consequence of this, the tasks of storing, processing, arranging, and analyzing this kind of data are becoming increasingly important in the banking industry. The financial sector is the single most important component of the international economy. In the past few years, the banking industry has had to deal with a lot of problems. Some of these problems include predicting customer churn, analyzing risk, keeping customers, finding fraud, managing risk, and dividing customers into groups. We need to come up with solutions that will maximize profits for the banking industry if we are going to be successful. Customers are at the forefront of each and every one of the banking industry's operations. As a direct result of this, it is essential to conduct client data analysis in terms of behavior, status, and overall satisfaction. An improvement in the level of satisfaction experienced by customers in the banking industry will make it simpler to attract new customers. Therefore, keeping the customer will result in an increase in the value of the bank. Customer churn is usually defined as the loss of customers who switched from one service provider to another and left. This switch could be caused by better prices or the advantages that a competitor company offers at information exchange. In the end, this means that the company's profits and revenues will go down.

Customers are able to seamlessly switch from one institution (Bank) to another in order to take advantage of better service quality or lower prices. Many businesses are under the impression that bringing on board new customers is significantly more difficult and expensive than keeping the ones they already have. However, one of the most significant challenges they face is providing dependable service to customers on time and within the allotted budget, all the while preserving a positive working relationship with those customers. They have to take into account the requirements of the customers in order to overcome these challenges. Attrition rates among their customer base will be one of their primary focuses. The term "customer churn" refers to the process by which clients or subscribers sever their relationship with an organization or service. Utilizing a company's sales and marketing assets throughout the various stages of the sales pipeline is necessary for any business that wishes to increase its customer base.

Managers at the company have come to recognize the significance of accurately forecasting the churn rate among their clientele in today's business climate. It has been determined that keeping current customers is both easier and cheaper than seeking out new ones. In addition, given the current climate of intense competition, the loss of a single customer has developed into a significant barrier for the company. There are three perspectives from which one can examine the effects of customer churn. To begin, the loss of customers is a serious problem for a business because it represents the disposal of an important resource. It will be more difficult for us to acquire new customers as a result of the loss of the assets, which are the customers. In addition, it is of greater importance for our company to be able to draw in new customers. However, new customers almost never have the same level of loyalty as existing customers, and it takes a long time to establish loyalty in customers. It will be to everyone's advantage, as indicated by the preventative plan that was presented.

Numerous research papers in industries as diverse as telecommunications, insurance, and many others have used data mining and machine learning techniques to predict customer churn. Decision trees, SVM, xgboost, random forest, logistic regression, and neural networks are just a few examples of these methods. In contrast, deep learning approaches improved performance and accuracy, and even so, there is much left to discover about these techniques. Shallow learning and other older machine learning techniques are not as accurate as deep learning and other more modern approaches. The findings and outcomes of deep learning techniques applied to banking industry data to address the challenging problem of churn are therefore illuminated by this study.

On the other hand, customer retention tends to be more cost-effective than customer acquisition because the company has already earned the trust and loyalty of the customer who is already a client. As a consequence of this, it is essential for any company to have a system that is able to make reliable forecasts regarding the early stages of the loss of customers. In this paper, we seek to develop a framework that is capable of predicting the churn of banking clients by making use of Machine Learning techniques.

Research Methodology

Sample Frame and Sampling Technique

In the India, there are 12 scheduled banks in the public sector and 22 scheduled banks in the private sector. Including all banks in the study is not desired 34 banks were chosen for data collection, 6 from the public sector (covering 50% of banks), as shown in Table, and 11 from the private sector (100 percent banks are covered). Bank employees are the respondents to the data collection from banks.

The sample of bank employees was constructed in two stages: the first stage utilised probability sampling to pick banks, and the second stage used convenience sampling to choose individual bank employees. The researcher polled twenty to twenty- five staff from each bank.

Table 1 : Top Banks on the basis of their Market Capitalization

Public Sector Bank		**Private Sector Bank**	
Banks	**Market Cap (INR cr)**	**Banks**	**Market Cap (INR cr)**
SBI	417,448.70	HDFC	599,119.59
Bank of Baroda	43,879.01	Kotak Mahindra	283,825.08
Punjab National Bank	41,841.86	ICICI Bank	264,640.21
Indian Overseas Bank	39,127.99	Axis Bank	178,146.35
Canara Bank	36,672.64	IndusInd bank	98,382.88
Union Bank	30,141.24	Bandhan Bank	57,549.43
		IDFC First Bank	19,919.02

		Yes Bank	19,561.13
		Federal Bank	17,518.15
		RBL Bank	16,962.58
		City Union Bank	13,731.49

Source: www.moneycontrol.com

Sample Size

Stated by	**Criteria**	**Suggested Sample as per criteria**
Hair et al,1999	Number of factors * 25	7 X 25= 175 Sample Size.
General thumb-rule	Number of statements *5	37 X 5= 185 Sample Size.
Nunnally 1978; Friedel 2001	For factor analysis, 330 sample size is adequate	330

Many studies have looked into how to choose an appropriate sample size. A sample size of less than 30 participants is considered too small for study. If the population is large, consider a sample size of 100 or more (Sekaran, 2006). This study employed a sample size of 432 people. As stated by Zikmund (2003), sample size is crucial since the smaller the sample size, the higher the chance of inaccuracy. The researcher also argued that a larger sample size results in more accurate research results. The cost of data collection is an important component to consider while making decisions. The sample size is crucial for interpreting SEM results because it is used to estimate sampling error. For the sample size, we used the following criterion (Table 1).
Structured questionnaire was used for the collection of data.

Literature Review

K. In 2021, Veningston, P. V. Venkateswara Rao, C. Selvan, and M. found that financial institutions prioritize customer attrition prediction. Ronalda. We considered the possibility that a major international bank's customers had decided to leave. The bank investigated why customer churn was so high. We search a 10k-record dataset for potential customers who are more likely to stop using the bank's value additions soon. This paper uses supervised classification models trained with cutting-edge machine learning algorithms to predict future customers and churn. The massive historical banking data trained these models. Dataset has 13 attributes and a class label. Nave Bayes had the highest accuracy at 86.29 percent. Telecommunication companies can use churn prediction to predict which customers will switch to competing networks, and human resources departments can use it to predict which employees will leave so they can start replacing them.
In the telecommunications industry, where customer churn has become a CRM and customer retention issue, Ying Huang and Tahar Kechad (2021) say churn prediction is crucial to keeping customers and minimizing losses. Classification models are also assessed by their predictive accuracy and interpretability. Classification methods based on a single model are becoming less practical. To better predict customer behavior, we present a model-based learning system that combines supervised and unsupervised techniques. FOIL and k-means clustering are used in the system. Three experiments used telecom datasets. Three groups of experiments test whether weighted k- means clustering improves data partitioning, classification results, and the proposed hybrid- model system. We used UCI repository benchmarks to compare our results to other researchers'. The hybrid model-based learning system outperforms state-of-the-art alternatives.
In today's competitive market, predicting customer turnover is becoming more important (Yu

Zhao, Bing Li, Xiu Li, Wenhuang Liu, Shouju Ren, 2005). The accuracy of churn prediction studies is too low to be useful to businesses. Data mining, machine learning, computer vision, and pattern recognition all benefit from the accuracy and generalizability of Support Vector Machines (SVMs), which are grounded in statistical learning theory. Predicting customer attrition with SVM has not been attempted. Due to the scarcity of negative data points in the churn dataset, an improved one-class SVM is required. We used our methodology to analyze churn rates in the wireless industry. We achieve better results than artificial neural networks, decision trees, and naive bayes.

According to Awe M. Oluwatoyin, Sanjay Misra, John Wejin, Abhavya Gautam, Ranjan Kumar Behera, and Ravin Ahuja (2022), the technology of the twenty-first century has resulted in an enormous influx of data. In 2020, daily Internet data production will reach 2.5 quintillion bytes. With an expected 5.3 billion Internet users by 2023, it will be necessary to develop sophisticated and effective methods for discovering and extracting insights from massive data sets. Logistic regression, decision trees, and clustering are just some of the machine learning techniques that are gaining traction in the field of churn prediction. Customer churn prediction predicts how many customers will stop using or cancel a company's product or service. Many prediction models have been proposed, but most studies have focused on their accuracy rather than how they can promote long- term economic growth. This paper uses Power BI to examine the causes of customer churn at United Bank of Africa (UBA), Nigeria. Power Bi's decision tree algorithm trained and tested. Churn was predicted by high customer account balances. Male customers also churn less than female customers. This study will help banks improve customer retention. A reliable customer churn prediction model is important to academics and industry professionals.

Ketaki Patil, Shivraj Patil, Riya Danve, and Ruchira Patil (2022) say customer satisfaction is crucial to a business's growth. Customer churn is a top priority for all businesses because it causes customer loss. Because customer churn affects revenue, industries are actively developing and adopting new methods to predict it. This identifies and addresses the many causes of high customer churn. This study builds and analyzes banking and telecommunications churning rates to predict customer churn. Using an open-source dataset of telecom and banking customers and machine learning techniques such as decision tree, random forest, KNN, kernel SVM (K-SVM), naive Bayes, and logistic regression, this study predicts churning rates. The best model is used for churn forecasting after being compared using different metrics. Random forest (87.05%) and artificial neural network (81.93%) performed best on the Bank and telecom datasets, respectively. This study examines customer churn and offers solutions. This will boost productivity and profits.

Himani Jain, Garima Yadav, and R. Manoov 2020 explains. Customer attrition occurs when customers gradually stop using a company's products or services. Our paper describes how exploratory data analysis and targeted promotions can predict customer churn and prevent it. We compare four algorithmic models—logistic regression, random forest, support vector machine, and XGBoost— across three industries to predict churn. Because few studies objectively compare algorithms across application areas, this comparison is being made. Exploratory data analysis helps us develop retention strategies.

Bright Senanu and Bedman Narteh (2023) examined customer bank switching after regulatory pressure and banking sector reforms forced retail banking consolidation. Survey-based quantification was used. After a literature review identified five antecedents of switching intentions, 392 affected bank customers were surveyed. Partial least square structural equation modeling tested the research model. Price, reputation, and ineffective communication influenced customers' switching intentions in regulatory-induced mergers and acquisitions. Price-switching intentions correlated with viable alternatives. The first study of customers' intentions to switch banks after a banking regulatory change. It extends previous empirical research on switching in regulatory- driven bank mergers and acquisitions. The study also shows that accessible alternatives moderate switching.

Anusmita Bose, K. T. Thomas (2022) foresee that "churn" will be defined as "the loss of a customer due to that person switching service providers. With the rising level of internal market competition, key banks are placing a greater emphasis on customer relationship management. Banks would benefit greatly from having access to a robust, real- time churn analysis of their credit card customers. Keeping an existing customer is more than five times easier than finding a new one, according to numerous studies. Consequently, this paper suggests a technique for doing so using data from a financial institution. In order to deal with the lopsided data set, "Synthetic Minority Oversampling Technique" (SMOTE) has been implemented in this study. With the help of random forest, k-nearest neighbor, and the boosting algorithms XGBoost and CatBoost, we can forecast how likely a customer is to cancel their credit card. The accuracy has been improved through the use of hyperparameter tuning via grid search. Catboost outperforms competing models in experiments, achieving an accuracy of 97.85%.

Online banking has increased competition among financial institutions (Xiaofeng Li, Zhongwei Chen, 2022). Large financial institutions now prioritize customer retention and churn prevention. The paper begins with a descriptive statistical analysis of each feature based on an open dataset, then applies Logistic, Random Forest, and Scv models to predict customer churn using metrics like area under the receiver op.

Muhamed Hassen Seid and Michael Melese Woldeyohannis (2022) say identifying churned customers is essential to business growth. Identifying churned customers helps understand customer churn and develop market strategies for expansion. This research will create a machine learning model for the Commercial Bank of Ethiopia (CBE) to accurately identify at-risk customers and recommend retention strategies. This study used 204,161 datasets with eleven attributes. This experiment selected the best classifier by performance. Thus, Commercial Bank of Ethiopia (CBE) used Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbor, and Deep Neural Network to predict customer attrition. Research supports these classifier algorithms for predicting customer churn. This study selected features using a correlation matrix and importance measure. The SMOTE method balances data, and the algorithm's results were compared. A Deep Neural Network (DNN) won with 79.32% accuracy, 85.08% precision, and 78.19% recall across multiple experiments.

Varsha Agarwal, Shwetkranti Taware, Suman Avdhesh Yadav, Durgaprasad Gangodkar, ALN Rao, VK Srivastav (2022), Customer churn, a term used in business and finance, refers to the slow but steady loss of customers. A business can retain customers by predicting who will leave. The bank benefits from knowing which customers may switch banks soon, both theoretically and practically. Read this article to learn what machine learning algorithms to employ in order to detect banking customers who might be switching service providers. In this article, we see how Logistic Regression (LR) and Naive Bayes (NB) are two machine learning models that can use demographic data such as age, location, gender, credit card information, balance, etc., to forecast which customers will decide to leave the bank. Personal information such as reader's age, location, gender, credit card number, balance, etc. This article demonstrates the process of making probabilistic predictions using Naive Bayes (NB) and Logistic Regression (LR). Evidence from this study favours NB over LR.

Tomás Ferreira, Pedro Pita, and Isabel Sofia Brito (2022) found that some banks and corporate managers are seeing credit card usage decline. Customer satisfaction required new methods. This paper investigates a Kaggle-hosted fictitious data source to find the root of the problem and identify the clients most likely to defect so financial institutions can better serve their customers.

Banks prioritize customer retention because losing customers is expensive (Seyed Jamal Haddadi, Mohammad Ostad Mohammadi, Mojtaba Bahrami, Elham Khoeini, Mehdi Beygi, Mehrdad Haddad Khoshkar, 2022). We use time series Deep Neural Networks (DNNs) to retain retail banking customers in Iran in this paper. The dataset includes 50,000 Pasargad Bank customers' November and December 2021 daily transaction records.. The purpose of this research is to develop a highly churned customer predictor by collecting data from customers over the course of 30 days and making predictions about their behavior over the subsequent 30 days. In addition, unlike previous studies in this area, which use predetermined customer labels, we provide a novel definition of the churned banking customer to classify the information. After that, a Bi-LSTM neural network is fed the cleaned, preprocessed data. The proposed model clearly outperforms the state-of-the-art in machine learning. This paper can help direct AI and banking researchers in the right direction, while also educating banking industry managers on how to better serve their customers.

Predicting whether a bank customer will leave using machine learning and statistical models, Animesh Shukla (2021). The bank trains a machine learning model using data from 10,000 customers. We use statistical methods to analyze the data and draw conclusions about its connections. The web application uses the trained model stored in the database to calculate the likelihood of a customer leaving the bank.

Data science and machine learning methods aid banks in pinpointing the causes of customer attrition so they can implement strategies to minimize it (Noviyanti Tri Maretta Sagala & Syarifah Diana Permai, 2022). There is a lot of literature on banking customer retention, but we couldn't find any studies that used a consistent data set, methodology, or machine learning algorithms to predict which customers would leave. This study investigates, improves, and evaluates boosted-based techniques to find the best bank customer churn model. This study categorises customer churn data using boosting-based models like XGBoost, LightGBM, and CatBoost in two cases. These processes used default and hyperparameter settings. Grid search 10-fold cross-validation hyperparameters are good. We evaluate six machine learning models using four metrics and propose a comprehensive model based on bank customer turnover data. LightGBM produces the best model for the current task with 91.4% AUC, 94.8% precision, and 87.7% recall.

Today's banking customers have many options, according to Ilham Huseyinov and Omobola Okocha (2022). Most banks worry about customer retention and churn. This research uses machine learning to reduce customer churn. Statistics, SVM, Random Forest, Gradient Boosting, Extreme Gradient Classifiers include Boosting and Light Gradient Boosting. To achieve this goal, we first employ a feature selection technique to filter out superfluous information and zero in on the features that truly matter, and then we use the SMOTE technique to normalize the resulting data set. We compare the accuracy, recall, precision, and performance of classifiers trained on balanced data to those trained on imbalanced data. No classifier was found to be superior to others in cases of imbalanced data (before the SMOTE was applied). When applied to evenly distributed data (via SMOTE), the Random Forest classifier clearly outperformed its competitors.

According to research by Amany Zaky, Shimaa Ouf, and Mohamed Roushdy (2022), customer churn is a major problem for financial institutions. Since attracting new customers is challenging, CRM strategies typically center on retaining current ones. Churn occurs when a company loses customers because of higher prices and inferior services offered by its competitors. Using machine learning methods, numerous academic papers have identified and proposed remedies for the issue of customer attrition. This paper proposes a banking customer attrition framework. We predicted customer defections using deep learning and an artificial neural network on bank customer data. The ANN algorithm yielded 87% accuracy for bank customer data in churn modelling. The suggested framework a low-cost strategy for keeping banking clients, which boosts the financial institution's bottom line.

Business strategies have shifted to place more emphasis on retaining customers who are at risk of defecting from the company, as Victor Chang, Xianghua Gao, Karl Hall, and Emmanuel Uchenna (2022) point out. Preemptive retention strategies targeted at these customers are needed to solve this problem. If you want to save money without sacrificing efficiency, you need to make sure your retention efforts are only going to the customers who are planning to switch service providers. This report's research aims to devise a system for early prediction of churn with as little misclassification as possible. In order to maximize future attrition capture, the proposed methodology incorporates a temporal dimension into customer churn prediction. The proposed methodology is tested with a bank credit card dataset using six different machine learning algorithms. Finally, results from the proposed methodology are compared to those from previously published churn prediction techniques. The data suggests that different types of service agreements can be used to group customers together. Forecasting when a customer is likely to cancel their service with your company is possible.

According to Soumi De, P. Prabu, and Joy Paulose (2021), a company's growth strategy should focus on keeping existing customers happy. Several sectors are obviously seeing an increase in customer churn as a result of the global pandemic. Therefore, the foundation of any industry's long- term growth strategy should be customer retention, the focus of CRM. A company's bottom line can benefit greatly from a decrease in customer turnover. Researchers have found promising progress in predicting customer churn across industries including telecommunications, banking, e-commerce, and the energy sector. This paper's primary goal is to provide a comprehensive analysis of the various machine learning approaches currently in use to deal with churn. Fifty-five papers on the topic of churn classification from 2004 to 2020 are compiled and analyzed in this study. There are five overarching topics that connect the reviewed papers. Methods for selecting features, dealing with class imbalance, trying out different ML algorithms, hybrid models, and model ensembles are some of the topics covered. Finally, a few recommendations for future study are offered.

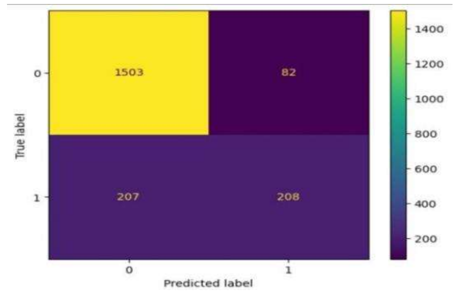
Financial institutions must now devote significant resources to analyzing the vast quantities of data produced by the banking sector (Amira Marouani and Andrea Tick, 2023). Companies that don't take advantage of the massive amounts of data they're generating risk falling behind the competition and failing to reap the full benefits that a comprehensive and thorough analysis can provide. Data mining has demonstrated its utility and enormous benefits, and is now widely acknowledged as an essential component of banks. This is due to its ability to help institutions better understand customer behavior and needs and to make decisions that are both accurate and profitable. The purpose of this research is to demonstrate the value of using data science algorithms and techniques to analyze banking clients' habits. Based on a case study of a Tunisian bank, this study examines how financial institutions in that country evaluate the risk and creditworthiness of loan applicants to decide whether or not to retain them. Increasing levels of competition and customer migration make customer churn a pressing issue for businesses of all sizes today (Bansal, Singh, Jain, & Verma, 2022).

Customers are leaving banks in search of alternatives that may offer them more favorable terms, lower fees, higher returns on investments, etc. To address this issue, we employ multiple prediction models to foresee which users will churn from the service and why. Ten ensemble learners, including bagging, boosting, voting, and stacking, are investigated here. Ensembles typically use algorithmic diversity with a fixed set of training instances to improve performance. Five open- source datasets are subject to ETL and exploratory data analysis. After resampling to achieve statistical balance, we put the datasets through empirical validation. Using five different performance metrics, we find that random forest and histogram- based gradient boosting outperform other ensemble methods, particularly in terms of specificity and geometric mean.

Approaches

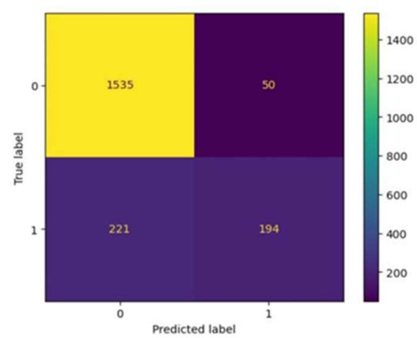
A. XGBoost

XGBoost is an ensemble learning algorithm that employs gradient boosting and regularization techniques for customer churn prediction in banks. By combining multiple decision trees and sequentially minimizing errors, it creates a strong predictive model. XGBoost optimizes an objective function, typically binary logistic regression, and provides insights into feature importance for identifying key factors affecting churn. With its ability to capture complex patterns and prevent overfitting, XGBoost enables accurate churn prediction and empowers banks to implement targeted retention strategies.



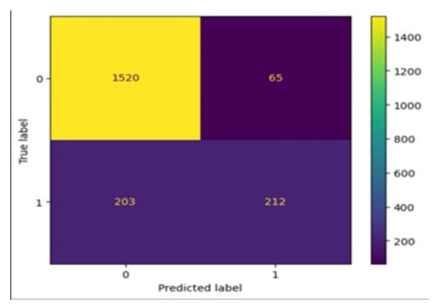
B. Random Forest

In the banking industry, predicting customer churn using the flexible machine learning algorithm Random Forest is a common practice. It operates by constructing an ensemble of decision trees, each trained on different subsets of the data. Through a majority voting mechanism, Random Forest aggregates the predictions of these trees to make accurate churn predictions. By leveraging the strength of multiple trees and reducing the impact of individual decision trees' biases, Random Forest offers robustness, handles high-dimensional data, and captures intricate relationships between features. Its effectiveness in customer churn prediction enables banks to proactively address churn risks and enhance customer retention and improved business performance.



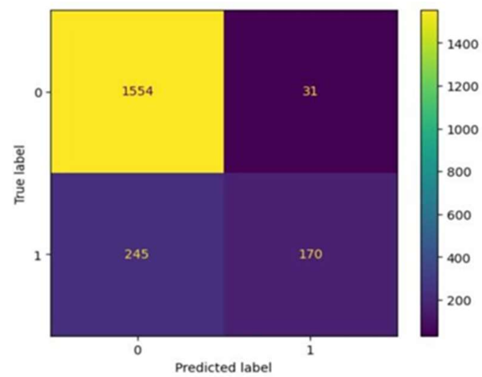
C. ANN

Strong machine learning algorithms like Artificial Neural Networks (ANN) are widely used by financial institutions for churn prediction. ANN is a model of the brain's structure and operation, consisting of stacked layers of interconnected nodes (neurons). By feeding historical customer data into the network, ANN learns complex patterns and relationships to make churn predictions. It utilizes forward and backward propagation algorithms to adjust the weights and biases, optimizing the model's performance. ANN's ability to capture nonlinear relationships and adapt to evolving patterns makes it an effective tool for identifying customers at risk of churn and facilitating targeted retention strategies in the banking industry.



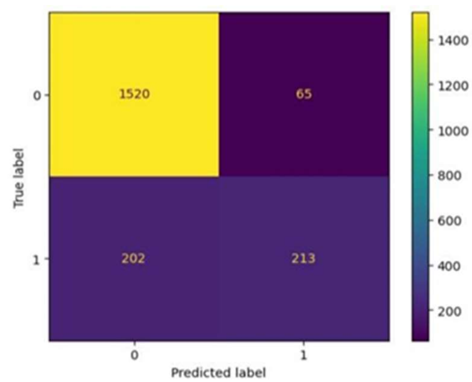
D. Support Vector Machines

The financial services sector makes heavy use of the robust machine learning algorithm known as Support Vector Machines (SVM) to predict customer churn. SVM works by finding a hyperplane in a high-dimensional feature space that most effectively distinguishes between churned and non-churned customers. It leverages a kernel function to transform the data into a higher-dimensional space, enabling the detection of complex patterns and decision boundaries. By selecting support vectors, SVM focuses on crucial data points that contribute significantly to churn prediction. Its ability to handle both linear and non-linear relationships, coupled with its robustness against overfitting, makes SVM a valuable tool for accurate customer churn prediction and assisting banks in implementing effective retention strategies.

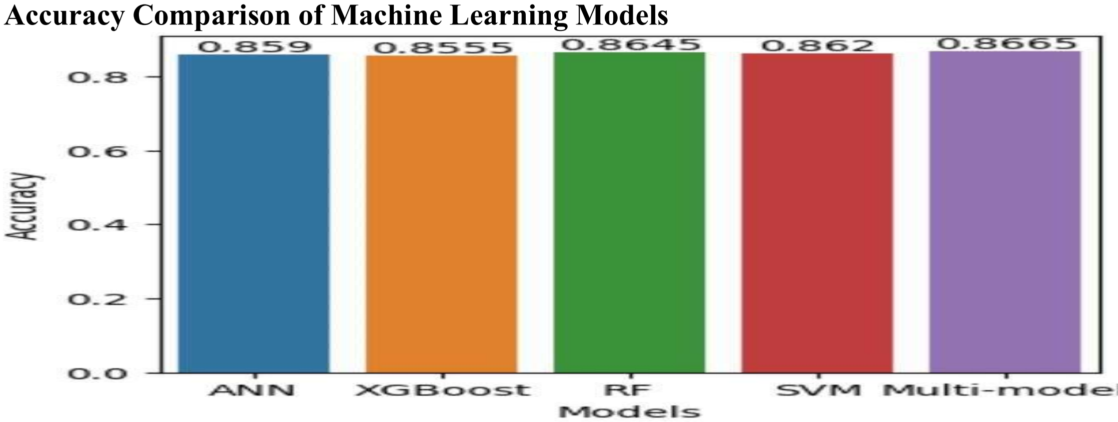


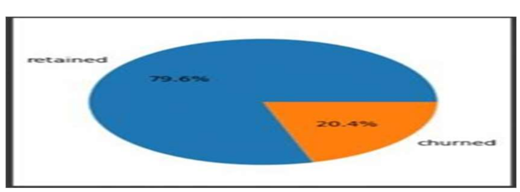
F. Multimodal Fusion

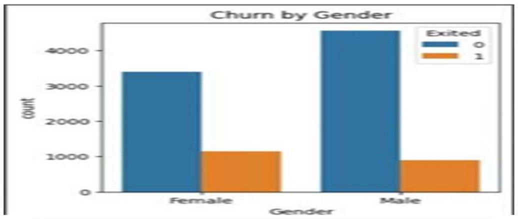
In the banking industry, multimodal fusion is used to predict customer churn by combining the results of several different machine learning models. These models can include ANNs, XGBoosts, Random Forests, and Support Vector Machine (SVM) classifiers. By integrating the outputs of these diverse models, it captures a comprehensive range of patterns and relationships from different perspectives. This fusion approach leverages the strengths of each model, such as ANN's ability to capture complex nonlinear patterns, XGBoost's gradient boosting optimization, Random Forest's ensemble learning, and SVM's robust classification. The multimodal fusion technique enhances churn prediction accuracy, provides a holistic view of customer churn risks, and enables banks to develop targeted strategies for customer retention and improved business performance.



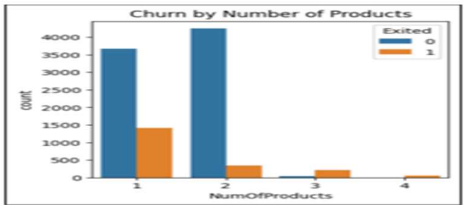
Accuracy Comparison of Machine Learning Models



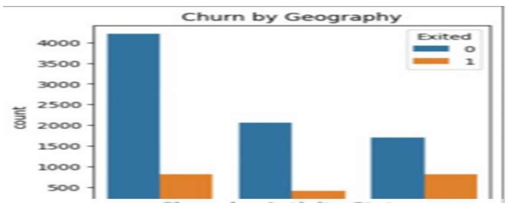




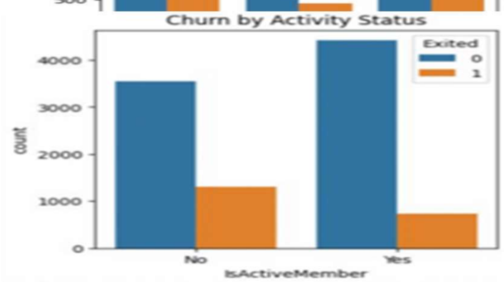
The displayed graph showcases the distribution of genders among individuals who have made the decision to either remain within the company or depart. Notably, the percentage of individuals who have chosen to stay exceeds the percentage of those customers.



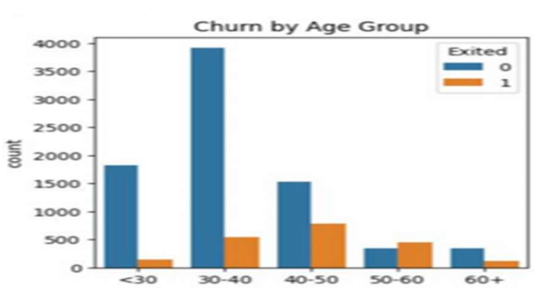
Customers with lower purchasing capacity are being retained at a higher rate, while those with higher purchasing capacity are experiencing a higher churn rate.



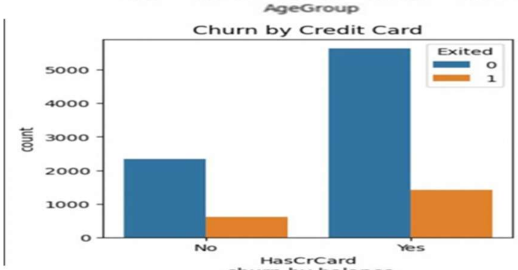
The provided graph showcases the spatial distribution of individuals who have made the decision to either continue their association with the company or terminate it, revealing a higher percentage of people who have left the bank specifically in Germany.



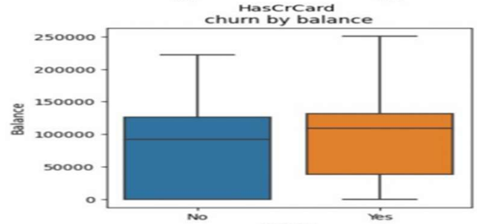
The displayed graph presents the churn rate based on activity status, differentiating between individuals who have decided to continue their association with the company and those who have opted to terminate it. The graph highlights that inactive individuals have a higher rate of attrition compared to those who left while being highly active.



The graph displays the churn status of individuals categorized by age groups. Notably, there is a notable retention rate in the age groups of less than 30 and 30-40, while the exit rate increases in the age group above 30. This suggests that the bank is strategically targeting and prioritizing the younger age groups.



Customers who possess credit cards exhibit a higher rate of retention compared to those who do not, while conversely, individuals without credit cards demonstrate a higher rate of churn in comparison to those with credit cards.



The average income for both retained and churned individuals is approximately equal, but the range of income levels among retained customers is broader compared to churned customers. Additionally, the average account balance among churned individuals is higher than that of retained individuals.

III. FINDINGS & SUGGESTIONS

A. Findings

1. Proportion of customers who have either remained with the company or chosen to discontinue their patronage.
2. The distribution of genders among individuals who have made the decision to either remain within the company or depart. Notably, the percentage of individuals who have chosen to stay exceeds the percentage of those who have chosen to leave.
3. Customers with lower purchasing capacity are being retained at a higher rate, while those with higher purchasing capacity are experiencing a higher churn rate.
4. Spatial distribution of individuals who have made the decision to either continue their association with the company or terminate it, revealing a higher percentage of people who have left the bank specifically in Germany.
5. Churn rate based on activity status, differentiating between individuals who have decided to continue their association with the company and those who have opted to terminate it. Highlights that inactive individuals have a higher rate of attrition compared to those who left while being highly active.
6. Churn status of individuals categorized by age groups. Notably, there is a notable retention rate in the age groups of less than 30 and 30-40, while the exit rate increases in the age group above 30. This suggests that the bank is strategically targeting and prioritizing the younger age groups.
7. Customers who possess credit cards exhibit a higher rate of retention compared to those who do not, while conversely, individuals without credit cards demonstrate a higher rate of churn in comparison to those with credit cards.
8. Customers who have credit cards display a higher retention rate in contrast to those without, whereas individuals without credit cards exhibit a higher churn rate when compared to those with credit cards.
9. The presence of credit cards among customers is associated with a higher likelihood of retention, whereas the absence of credit cards is linked to a higher probability of churn.
10. The mean income for both retained and churned individuals is roughly similar; however, retained customers exhibit a wider range of income levels compared to churned customers.
11. Churned individuals, on average, have a higher account balance than retained individuals.

B. Suggestions

1. Implement targeted marketing campaigns: Utilize machine learning techniques to identify customers at risk of churning. Design personalized marketing campaigns that offer incentives and tailored promotions to retain these customers.
2. Enhance customer experience: Focus on improving overall customer satisfaction by analyzing pain points and specific needs of different customer segments. Develop products and services that cater to their preferences, providing a seamless and personalized banking experience.
3. Develop credit card strategies: Pay special attention to customers with credit cards, as they are more prone to churn. Offer exclusive rewards, personalized benefits, and targeted promotions to incentivize credit card usage and improve customer retention.
4. Proactively retain customers: Monitor customer behavior and engagement patterns closely to identify early signs of churn. Implement proactive measures such as personalized communication, loyalty programs, and special offers to retain at-risk customers.
5. Continuously analyze and improve: Regularly analyze customer churn patterns and factors to stay updated with evolving trends and preferences. Refine machine learning models and strategies based on new insights and customer feedback, ensuring their effectiveness in

reducing churn.

IV. CONCLUSION

This research paper focuses on customer churn in the banking industry and utilizes machine learning techniques to predict churn and identify associated factors. The results indicate that ANN, XGBoost, Random Forest, and Support Vector Machines are the most effective machine learning methods for churn prediction. The analysis of churn predictions reveals noteworthy insights. Customers aged 50-60 demonstrate a higher churn rate compared to retained customers, and this trend is expected to continue for the 40-50 age group without preventive measures. Furthermore, customers with credit cards have a higher propensity for churn compared to those without, highlighting the significance of credit card-related strategies in mitigating churn.

These findings have critical implications for the banking industry. The high accuracy of machine learning techniques in churn prediction suggests their utility in identifying at-risk customers. Consequently, targeted marketing campaigns and customized service offerings can be developed to prevent churn among these individuals. Moreover, the insights into churn-contributing factors can guide enhancements to the overall customer experience. For instance, the bank could offer tailored incentives to customers aged 50-60 or those with credit cards. Additionally, adapting products and services to better align with the needs of these customer segments can foster greater satisfaction and retention.

In conclusion, the research underscores the power of machine learning in reducing customer churn in banking. By leveraging these techniques, banks can gain deeper insights into their customers, identify churn drivers, and implement targeted strategies that improve customer retention and overall business outcomes.

REFERENCES

[1] Investigation on Customer Churn Prediction Using Machine Learning Techniques,· K. Veningston, P. V. Venkateswara Rao, C. Selvan, M. Ronalda, vol 287. Springer, 2021, 978-981- 16-5348-3, https://doi.org/10.1007/978-981- 16-5348-3_8

[2] An effective hybrid learning system for telecommunication churn prediction, Ying Huang, Tahar Kechadi, Elsevier, 2021, https://doi.org/10.1016/j.eswa.2013.04.020

[3] Customer Churn Prediction Using Improved One-Class Support Vector Machine, Yu Zhao, Bing Li, Xiu Li, Wenhuang Liu, Shouju Ren, vol 3584. Springer, 2005, 978-3-540-31877-4, https://doi.org/10.1007/11527503_36

[4] The Application of the Locally Linear Model Tree on Customer Churn Prediction, Amineh Ghorbani, Fattaneh Taghiyareh, Caro Lucas, IEEE, 2009,978-1-4244-5330-6, 10.1109/SoCPaR.2009.97

[5] Customer Churn Prediction in Banking Industry Using Power Bi, Awe M. Oluwatoyin, Sanjay, Misra, John Wejin, Abhavya Gautam, Ranjan Kumar Behera, Ravin Ahuja, vol 421. Springer, 2022, 978-981-19-1142-2, https://doi.org/10.1007/978-981-19-1142- 2_60

[6] Machine Learning and Neural Network Models for Customer Churn Prediction in Banking and Telecom Sectors, Ketaki Patil, Shivraj Patil, Riya Danve, Ruchira Patil, Springer, 2022, 978-981- 16-7389-4, https://doi.org/10.1007/978-981- 16-7389-4_23.

[7] Churn Prediction and Retention in Banking, Telecom and IT Sectors Using Machine Learning Techniques, Himani Jain, Garima Yadav, R. Manoov, Springer, 2020, 978-981-15-5243-4, https://doi.org/10.1007/978-981-15- 5243-4_12

[8] Prediction of Customer Status in Corporate Banking Using Neural Networks, Stanislaw Osowski, Lukasz Sierenski, IEEE, 2020, 978-1- 7281-6926-2,

10.1109/IJCNN48605.2020.9206693
[9] Predictive Model for Cutting Customers Migration from banks: Based on machine learning classification algorithms, Kahlid Alkhatib, Sayel Abualigah, IEEE, 2020, 978-1- 7281-6227-0, 10.1109/ICICS49469.2020.239544
[10] Banking sector reforms and customer switching intentions: evidence from the Ghanaian banking industry, Bright Senanu, Bedman Narteh, J Financ Serv Mark 28, 15–29, Springer,2023, https://doi.org/10.1057/s41264-021-00135-8
[11] A Comparative Study of Machine Learning Techniques for Credit Card Customer Churn Prediction, Anusmita Bose, K. T. Thomas, vol 141. Springer, 2022, 978-981-19-3035-5, https://doi.org/10.1007/978-981-19-3035- 5_23
[12] Churn Prediction using Ensemble Learning, Xing Wang, Khang Nguyen, Binh P. Nguyen, Association for Computing Machinery, 2020, Pages 56–60, 978-1-4503-7631-0, https://doi.org/10.1145/3380688.3380710
[13] Machine Learning Based Customer Churn Prediction In Banking, Manas Rahman, V Kumar, IEEE, 2020, 978-1-7281-6387-1, 10.1109/ICECA49313.2020.9297529
[14] Customer Churn Analysis and Prediction in Banking Industry using Machine Learning, Ishpreet Kaur, Jasleen Kaur, IEEE, 2020, 978-1- 7281-7132-6, 10.1109/PDGC50313.2020.9315761
[15] Customer Churn Prediction in Bank Based on Different Machine Learning Models, Xiaofeng Li, Zhongwei Chen, IEEE, 2022, 978-1-6654- 9271-3, 10.1109/ISPCEM57418.2022.00061
[16] Machine Learning based Prediction of Customer Churning in Banking Sector, Manoj Kumara N V, Bharath Kumar K K, Arun Chandra Mudhol, IEEE, 2022, 978-1-6654-8962-1, 10.1109/ICAISS55157.2022.10011126
[17] Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia, Muhamed Hassen Seid, Michael Melese Woldeyohannis, IEEE, 2022, 978-1-6654-5587- 9, 10.1109/ICT4DA56482.2022.9971224
[18] Customer Churn Prediction Using Machine Learning, Varsha Agarwal, Shwetkranti Taware, Suman Avdhesh Yadav, Durgaprasad Gangodkar, ALN Rao, V K Srivastav, IEEE, 2022, 978-1-6654-7657-7, 10.1109/ICTACS56270.2022.9988187
[19] Predict Churning Customers – An Explorative Study, Tomás Ferreira, Pedro Pita, Isabel Sofia Brito, IEEE, 2022, 2166-0727, 10.23919/CISTI54924.2022.9820260
[20] Customer Churn Prediction in the Iranian Banking Sector, Seyed Jamal Haddadi, Mohammad Ostad Mohammadi, Mojtaba Bahrami, Elham Khoeini, Mehdi Beygi, Mehrdad Haddad Khoshkar, IEEE, 2022, 978-1- 6654-6781-0, 10.1109/ICAPAI55158.2022.9801574
[21] Application of Machine Learning and Statistics in Banking Customer Churn Prediction, Animesh Shukla, IEEE, 2021, 978-1-7281-9687- 9, 10.1109/ICSCC51209.2021.9528258
[22] Enhanced Churn Prediction Model with Boosted Trees Algorithms in The Banking Sector, Noviyanti Tri Maretta Sagala, Syarifah Diana Permai, IEEE, 2021, 978-1-6654-4303-6, 10.1109/ICoDSA53588.2021.9617503
[23] A Machine Learning Approach To The Prediction Of Bank Customer Churn Problem, Ilham Huseyinov, Omobola Okocha, IEEE, 2022, 978-1-6654-5995-2, 10.1109/IISEC56263.2022.9998299
[24] Predicting Banking Customer Churn based on Artificial Neural Network, Amany Zaky,

Shimaa Ouf, Mohamed Roushdy, IEEE, 2022, 978-1- 6654-9973-6, 10.1109/ICCI54321.2022.9756072
[25] Machine Learning Techniques for Predicting Customer Churn in A Credit Card Company, Victor Chang, Xianghua Gao, Karl Hall, Emmanuel Uchenna, IEEE, 2022, 978-1-6654- 5455-1, 10.1109/IIoTBDSC57192.2022.00045
[26] Application of Machine Learning in Customer Churn Prediction, Soumi De, P Prabu, Joy Paulose, IEEE, 2021, 978-1-6654-2691-6, 10.1109/i-PACT52855.2021.9696771
[27] Predictive Modeling to Investigate and Forecast Customer Behaviour in the Banking Sector, Amira Marouani, Andrea Tick, IEEE, 2023, 979-8-3503-1986-6, 10.1109/SAMI58000.2023.10044499
[28] Analysis of ensemble classifiers for bank churn prediction, Ankita Bansal, Sarabjot Singh, Yashonam Jain, Ayush Verma, IEEE, 2022, 978- 1-6654-6200-6, 10.1109/ICCCIS56430.2022.10037623
[29] Interpretable Machine Learning for Predicting Customer Churn in Retail Banking, Sudi Murindanyi, Ben Wycliff Mugalu, Joyce Nakatumba-Nabende, Ggaliwango Marvin, IEEE, 2023,979-8-3503-9728-4, DOI:10.1109/ICOEI56765.2023.10125859
[30] Machine Learning to Develop Credit Card Customer Churn Prediction, Dana AL-Najjar, Nadia Al-Rousan, Hazem AL-Najjar, 2022, J. Theor. Appl. Electron. Commer. Res. 2022, 17(4), 1529-1542, https://doi.org/10.3390/jtaer17040077
[31] Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector, Edvaldo Domingos, Blessing Ojeme, Olawande Daramola, Computation 2021, 9(3), 34, https://doi.org/10.3390/computation9030034
[32] Customer Churn Prediction in Banking Industry Using K-Means and Support Vector Machine Algorithms, Abdulsalam Sulaiman Olaniyi, Arowolo Micheal Olaolu, Bilkisu Jimada- Ojuolape, Saheed Yakub Kayode, International Journal of Multidisciplinary Sciences and Advanced Technology, Vol 1 No 1 (2020), 48– 54
[33] Propension to customer churn in a financial institution: a machine learning approach, Renato Alexandre de Lima Lemos, Thiago Christiano Silva, Benjamin Miranda Tabak, Neural Computing and Applications volume 34, pages11751–11768 (2022), https://doi.org/10.1007/s00521-022- 07067-x
[34] An Ensemble Model for Predicting Retail Banking Churn in the Youth Segment of Customers, Vijayakumar Bharathi S, Dhanya Pramod, Ramakrishnan Raman, Data 2022, 7(5), 61, https://doi.org/10.3390/data7050061
[35] U M Gopal Krishna (2020). “Problem and Supporting System of Aged People in Sugali Tribes Chittor District, Andhrapradesh”, in International Journal of Scientific & Technology Research (IJSTR), Volume-09, Issue-2, pp 2277-8616, Impact Factor 7.466, February- 2020.
[36] U M Gopal Krishna (2022). “Employee Stress And Its Impact On Employee Performance Among Women Employees In Mysore District”, Nat. Volatiles & Essent. Oils,, Volume-8 Issue-4, January 2021.
[37] U M Gopal Krishna (2021). “An Analysis of the Lack of Large Scale Entrepreneurship in IT Industries of India”. International Journal of Aquatic Science, Volume-12, Issue-01, pp 14- 27, July - 2021.
[38] U M Gopal Krishna (2022). “TRASE Model for Performance & Workload Management – Academic Scenario” Turkish Journal of Physiotherapy and Rehabilitation, Volume-32, Issue-

03, ISSN 2651-4451 | e-ISSN 2651-446X, pp 39622 - 39631, January – 2022.