IJOGHO, SAMSON (Student)

Machine Learning

CIS4035-N

Table of Contents

[1. Abstract 2](#_Toc155454081)

[1. INTRODUCTION 2](#_Toc155454082)

[1.1. Statement of the problem: 2](#_Toc155454083)

[1.2. Project Goal: 3](#_Toc155454084)

[3.0. Literature review 3](#_Toc155454085)

[3.1. METHODOLOGY 4](#_Toc155454086)

[3.2. Dataset Description: 5](#_Toc155454087)

[4. Exploratory Data Analysis (EDA) 7](#_Toc155454088)

[4.1. Normalization 12](#_Toc155454089)

[5. Data pre-processing 12](#_Toc155454090)

[5.1 Modeling 13](#_Toc155454091)

[5.2. Using SMOTE to Balance Class Distribution 13](#_Toc155454092)

[5.3. Before and After SMOTE Class Distribution Comparison 14](#_Toc155454093)

[6. Machine Learning 15](#_Toc155454094)

[6.1. Data splitting for Testing and Training 15](#_Toc155454095)

[6.2. Classification 16](#_Toc155454096)

[7. Support Vector Machine (SVM): 16](#_Toc155454097)

[7.1. Training and Evaluation of SVM Models: 16](#_Toc155454098)

[7.2. Confusion Matrix and Accuracy Calculation: 16](#_Toc155454099)

[7.3. Conclusion 17](#_Toc155454100)

[8. K-Nearest Neighbor (KNN) 18](#_Toc155454101)

[8.1. Training and Evaluating the Model: 18](#_Toc155454102)

[8.2. Model Evaluation for Various k Values: 18](#_Toc155454103)

[8.3. Model Evaluation by Confusion Matrix 18](#_Toc155454104)

[8.4. Conclusion: 21](#_Toc155454105)

[9. Comparison of SVM and KNN Models 21](#_Toc155454106)

[Reference 22](#_Toc155454107)

[Appendices 23](#_Toc155454108)

[Figure 1:Figure 1 suggested moduls 6](#_Toc155553855)

[Figure 2 Dataset description 7](#_Toc155553856)

[Figure 3: distribution of bank costumer by churn 8](#_Toc155553857)

[Figure 4: distribution of gender by churn status. 9](#_Toc155553858)

[Figure 5: boxplot of age distribution. 9](#_Toc155553859)

[Figure 6: histogram of age distribution. 10](#_Toc155553860)

[Figure 7: churn distribution by credit score. 11](#_Toc155553861)

[Figure 8: histogram of churn by credict score 11](#_Toc155553862)

[Figure 9: boxplot of balance distribution. 12](#_Toc155553863)

[Figure 10: histogram of balance distribution 12](#_Toc155553864)

[Figure 11: feature correlation matrix. 13](#_Toc155553865)

[Figure 12: class distribution 16](#_Toc155553866)

[Figure 13: confusion matrix statistics 18](#_Toc155553867)

[Figure 14: svm linear grid graph 18](#_Toc155553868)

[Figure 15: KNN accuracy 19](#_Toc155553869)

[Figure 16: confusion matrix table(KNN.75) 20](#_Toc155553870)

[Figure 17: confusion matrix table(KNN.76) 21](#_Toc155553871)

[Figure 18: graph to show different values of K 22](#_Toc155553872)

[Figure 19: comparison table. 22](#_Toc155553873)

[Figure 20: comparison graph 23](#_Toc155553874)

# Abstract

Customer attrition is another term for customer churning. Currently, there are 1.5 million annual customer attrition, and the figure is rising. Customer retention is a challenge for the banking sector. For a number of reasons, including reduced charges for better financial services, bank branch locations, low interest rates, etc., customers may decide to transfer banks. Prediction algorithms are therefore used to identify clients who are most likely to leave in the future. Since it is less expensive to serve loyal customers than it is to lose a client, which causes the bank to lose money. Additionally, repeat customers bring in more advantages and new recommendations. In order to predict the chance of a client churn, machine learning models like support vector machine (SVM), K-nearest neighbour (KNN), and others are applied to the bank dataset in this research. Accuracy and recall are two performance indicators that are contrasted.

# 2.0 INTRODUCTION

Customer attrition, also referred to as "customer churning," is the phenomenon wherein consumers discontinue doing business with a firm or sector. The term "churn rate" refers to the proportion of customers who discontinue doing business with a company or using its brand over a given time frame. The revenue of the business is directly impacted by customer attrition. Customer attrition is so undesirable. In order to prevent this, the customer's behavior is predicted in advance using the prediction model. Churn prediction requires the ability to anticipate when a particular customer is likely to leave the company; therefore, there is still a perfect chance to help the client by providing him with an added benefit. Technology is a major factor in all of these things being possible. Through data exploration and analysis, it is possible to predict if a customer will be dissatisfied or choose to stick with the company. Thus, data analysis is defined as a comprehensive review of data using data mining and machine learning techniques to find relevant information approaches [1 (Manas Rahman, V Kumar, 2020)]. Biomedical and other well-known fields have already tested machine learning, retail, and economic domains for credit evaluation, market basket analysis, and cancer diagnosis. This article, however, focuses on how machine learning models are used in the banking industry to predict customer attrition. Customer retention is a challenge for the banking sector. For a number of factors, including reduced charges for superior financial services, the location of bank branches, the calibre of digital tools, low interest rates, and more, customers may move to another bank. Prediction models are therefore employed to identify future client churn prospects. These models search for patterns in historical churner data by comparing current consumers. If there is a similarity, the current customers are then labelled as possible churners. Predicting customer churn is crucial, even if it means acquiring new clients at the expense of retaining current ones [2 (Dash, 2021)]. Additionally, maintaining a relationship with a Gaining a long-term customer is less expensive than losing one, which costs the bank money. Additionally, older clients provide more benefits and new recommendations. This research uses a variety of machine learning models., including Support vector machine (SVN), decis1ion trees (DT), K-nearest neighbour (KNN), etc., to a dataset in order to make predictions. The results are compared in terms of performance, including accuracy and recall.

## Statement of the problem:

Retail banks are very concerned about customer churn, also called customer attrition, because it has a direct effect on revenue, market share, and the long-term viability of the company. By recognizing the triggers and trends that lead to customers discontinuing their use of a bank's services, institutions can take proactive steps and modify their products to better satisfy their patrons. Banks may create focused customer retention strategies and acquire insights into the variables impacting customer attrition by utilizing advanced analytics and machine learning.

Importance Churn Prediction's in Retail Banking:

Protection of Revenue:

By anticipating customer attrition, banks may put retention plans in place that protect profitability and revenue streams from losing valuable clients.

Improved Client Experience

Banks may enhance customer satisfaction and loyalty by customizing their services to match consumer expectations and addressing the factors behind customer attrition.

## Project Goal:

The main goal of this project, "Customer Churn Prediction in Retail Banking" is to advance machine intelligence models that’s able to anticipate customer’s turnover in retail banking industry by utilizing data-driven insights. Through the examination of past customer data, the initiative seeks to pinpoint trends, elements, and actions suggestive of possible attrition. The analysis's findings will help comprehend the dynamics of customer attrition and offer banks practical suggestions for improving client retention tactics and lowering churn.

The dataset will be examined, conduct exploratory data analysis (EDA), pre-process the data, and use machine learning methods to create reliable churn prediction models in the sections that follow. By using an analytical and iterative process, with hope to make significant.

# 3.0. Literature review

Several studies have examined different aspects of churn prediction, employing diverse methodologies and datasets. An overview of key findings, methodology, and trends in retail banking customer churn prediction are given in this review of the literature. A study [3 (V Kumar, Manas Rahman, 2020)] the prediction of commercial bank client attrition using the SVM model. For this study, a consumer dataset from a Chinese commercial bank with 50,000 records was selected. After pre-processing records, there are 46,406 valid data records. A linear basis kernel function and a radial basis kernel function are the two types of SVM models that are selected. The prediction power of the classification models was greatly enhanced by the under-sampling approach. Because the actual commercial has uneven characteristics, the SVM model is unable to predict churners with any degree of reliability. The results indicate that random sampling technique combined with SVM model can greatly improve predictive power and help commercial banks anticipate churners more accurately. Nonetheless, the study's churner to non-churner ratio was fixed at 1:10. The maximum outcome in a 1:1 ratio is 80.84%. The main drawback of this work is this. The study suggested a model for churn analysis that can help telecom companies forecast which customers are most likely to leave. The technology uses machine learning algorithms on a large data platform. The effectiveness of the model is assessed using the standard measure known as the Area Under Curve (AUC). The telecom company Syriatel provided the study's dataset. Four alternative techniques have been used to test the model: Random Decision Tree was used to optimize the method hyper-parameters. Because the target class is uneven, the learning sample is rebalanced by obtaining a sample of data from both classes. The study began with oversampling, which was accomplished by multiplying the churn class by the other class. A random under-sampling method was also employed, which reduces the sample the broad class's size in relation to the second class. The maximum number of nodes and depth were optimised by the hyper-parameters of the Decision Tree method, which was utilised for training. The optimal number of trees was 200, as shown by the best outcomes in Random Forest and GBM. Furthermore, GBM beat DT and RF. It was discovered that the ideal AUC value for XGBOOST on 180 trees was 93.301%. When the models were tested by installing a new dataset for varying times and without any beneficial marketing engagement, XGBOOST also produced the best results, with an 89% AUC. The authors offer a novel method for improving churn prediction performance by utilising a data preparation treatment strategy The classifiers perform satisfactorily when redundant and irrelevant data is eliminated from the dataset through pre-processing. Another article provides a scientific investigation of the use of data mining to get information from repositories in the banking sector. Customers that utilise more financial services and products seem to be more committed, according to the research. Consequently, the bank ought to concentrate on offering solutions that satisfy the needs of its customers who utilise less than three products. The database used contains records on 1866 clients at the time of the study. The Alyuda NeuroInteligence software suite includes a neural network-based churn prediction method that serves as the foundation for this investigation. Data is divided into three sets: training, validation, and testing sets. Three categories of attributes are outlined in the data analysis step: the attributes to be measured, the attributes to be rejected, and the attributes that are required. The model chooses a number of hidden layers during the network design process. The results of network training show that the validation's CCR% is 93,959732The investigation revealed that the bank offers very well-tailored senior programmes and that there is little chance of competition because a significant share of its entire customer base (691/1886) consists of retirees. The primary drawback of this method is the neural network's very slow and labor-intensive processing speed.

## 3.1. METHODOLOGY

The goal of this effort is to use effective data mining techniques to predict client attrition in commercial banks as early as possible.

An illustration of the suggested model is provided diagrammatically in fig (1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Authors | Work Title | year | Methodology | Remarlk |
| Predictions by B. He, Y. Shi, Q. Wan, and X. Zha | Predictions by B. He, Y. Shi, Q. Wan, and X. Zha | Predictions by B. He, Y. Shi, Q. Wan, and X. Zha | Predictions by B. He, Y. Shi, Q. Wan, and X. Zha | Predictions by B. He, Y. Shi, Q. Wan, and X. Zha |
|  |  |  |  |  |

Figure suggested modules

## 3.2. Dataset Description:

The dataset, which was used to model churns, was downloaded from Kaggle. 10000 bank customers' details are included in the dataset, and An integer value that signifies whether a customer is still a customer or has left the bank is the target parameter. 2037 of these samples were negative class (exited), and 7963 were positive class (kept). When a client's bank account is closed, the intended variable shows the first binary flag, and also the when the customer is kept, it shows the binary flag 0. The dataset includes 18 features derived from customer information and transactions completed by customers.

RowNumber: Indicates the record (row) number; it has no bearing on the result.

CustomerId: Doesn't affect the customer's choice to quit the bank and is made up of random values.

Surname: A customer's decision to leave the bank is unaffected by their surname.

Credit Score: - It can impact customer churn because a customer with a higher credit score is less likely to leave the bank.

Geography: A customer's decision to quit a bank may be influenced by their location.

Gender: It's important to look into if a customer's decision to quit a bank is influenced by their gender.

Age: Obviously important, since older customers are less likely to leave their bank than younger ones.

Tenure: - The duration of a customer's bank account relationship. Older clients tend to be more devoted and less inclined to switch banks.

Balance: - This is another useful measure of customer attrition since customers who have larger balances are less likely to switch banks than those who have smaller balances.

NumOfProducts: - Indicates how many goods a consumer has bought using the bank.

HasCrCard: - Shows if a customer has a credit card. Because credit card customers are less likely to leave the bank, this column is much more important.

Salary Estimate: - Similar to balancing, employees are more inclined to leave the bank if their salary is lower than their salary.

Exited: Indicates whether or not the client departed the bank.

Complaint: - Regardless of whether the client has a grievance.

Customer satisfaction score: a rating given by the client for the resolution of their problem.

Cardtyped: - The type of card the customer is using.

Points Earned: a customer's point total obtained by using a credit card.Table below contains more information on these aspects.

A screenshot of a computer

Description automatically generated

Figure Dataset description

# 4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis is effective for deeply comprehending data and gaining significant knowledge through data visualisation. This project's exploratory data analysis comprises of univariate, bivariate, and correlation studies. Here are some graphs from the exploratory data analysis process:

**Distribution of Customer by Churn status**:

A chart of a distribution of bank customers

Description automatically generated

Figure : distribution of bank costumer by churn

The graph above depicts the distribution of the target variable; it is known that 20.4% of bank client's churn. It is concluded that the data on the target variable is unbalanced.

**Distribution of Gender by Churn status**:

A graph of different colored bars

Description automatically generated with medium confidence

Figure : distribution of gender by churn status.

The graph above depicts the target variable's distribution by 'Gender'. Female bank clients churn at a higher rate of 11.4% than males, who churn at a lower rate of 9%.

**Distribution of Age by Churn status**:

A diagram of a box plot

Description automatically generated

Figure : boxplot of age distribution.

A graph of age by churn status

Description automatically generated

Figure : histogram of age distribution.

The distribution of "Age" base on churn status is displayed in the above graph. Between the ages of thirty and fifty, there is a comparable age range among bank clients who have churned and those who have been kept. The majority of retained bank clients are in the 30-40 year old of age range.

**Distribution of Churn by Credit Score**:

A diagram of a number of red and blue squares

Description automatically generated

Figure : churn distribution by credit score.

A graph of a person with red and blue lines

Description automatically generated

Figure : histogram of churn by credict score

The "CreditScore" distribution by churn status is displayed in the above graphic. The range of 600 to 700 is the same for both retain and churn clients' credit scores.

**Distribution of Churn by Credit by account balance:**

A red and blue squares

Description automatically generated

Figure : boxplot of balance distribution.

A graph with red and blue bars

Description automatically generated

Figure : histogram of balance distribution

The distribution of "Balance" according to churn status is displayed in the graph above. Between 40,000 and 125,000 churn clients make up the remaining balance. Conversely, the remaining retain customer distribution ranges from 0 to 125,000.

**Feature correlation Matrix:**

A graph of a credit score

Description automatically generated with medium confidence

Figure : feature correlation matrix.

There aren't many features that have a high linear correlation with the target, as the graph illustrated. This indicates that most of the dataset is not correlations.

# 4.1. Normalization

Normalization plays an essential role to maximizing the efficiency and convergence of various algorithms, allowing for meaningful feature comparisons, and improving the general stability and interpretability of the models.

# 5. Data pre-processing

One crucial step in the data mining procedure is preprocessing the information. since they directly impact the success rate of the task. It needs to handle data that is unreliable, noisy, and irrelevant. Additionally, the data conversion as well if needed. In this study, these characteristics were used while determining the churn prediction.

* Unnecessary feature: Data features that is insignificance on the subject under discussion are deemed irrelevant. Retaining such features may occasionally have an impact on classifier performance. When looking at the characteristics, the churn data-set called Row number, Customer Id, Surname, and Geography have no influence on the forecast. As a result, these characteristics were deliberately overlooked in our study.
* Transformation: Data transformation is the process of transforming data from one form to another. Data quality is enhanced and programmes are shielded from potential hazards including null values, unwanted duplication, erroneous indexing, and incompatible formats by properly prepared and validated data.

# 5.1 Modeling

Upsampling and downsampling are procedures used in data processing to customize the class distribution of provided data. Because the data is significantly imbalanced 7962 positive class samples and 2038 negative class samples and the available data sample size is tiny, this study will employ the upsampling technique. Because if under sampling is selected, the size of the data will shrink to the point where there will be insufficient data to create the model. As a result, this study employs random oversampling by resampling the minority class (negative class). In order to prevent partiality in the model, this must be avoided. To overcome this, sampling must be performed so that the number of target classes approaches a balanced condition.

## 5.2. Using SMOTE to Balance Class Distribution

Synthetic Minority is used to Over-sampling Technique (SMOTE) to correct the imbalance in the target variable "Exited," which had an unbalanced distribution with a strong majority of a single class. SMOTE is a data augmentation approach that generates simulated instances of the minority class in order to balance class distribution.

The class distribution of the "Exited" variable was effectively balanced when we applied SMOTE to our dataset. The final distribution has 4076 instances for each class (Class 1 and Class 2). By minimizing the possible bias produced by uneven class proportions, this balanced distribution improves the effectiveness of our predictive modelling systems.

The following parameters were utilized in the SMOTE implementation:

Oversampling percentage (perc.over): 100%

Under-sampling percentage (perc.under): 200%

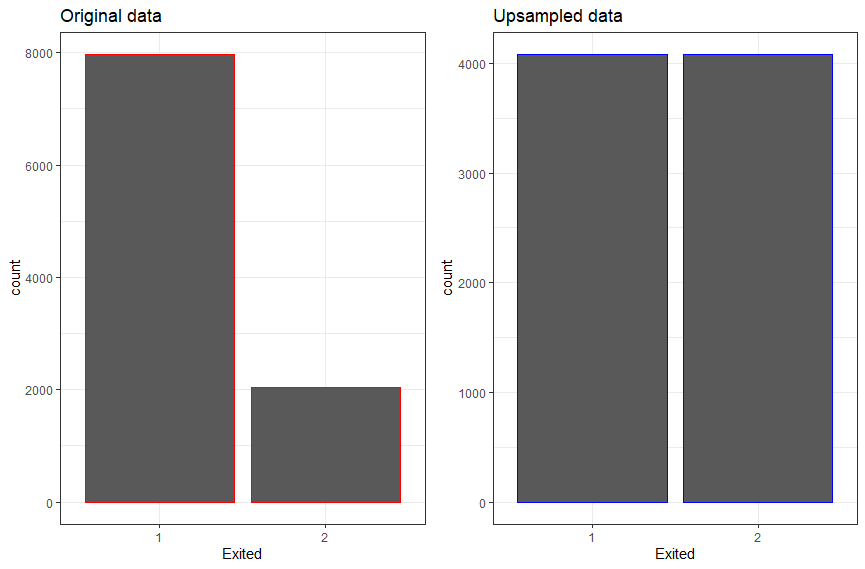
The number of nearest neighbors (k):5.

This method ensures that the minority group is oversampled by 100% and the majority group is under-sampled by 200%, resulting in a more equal representation of both groups.

SMOTE's balanced dataset is critical for boosting the model's capacity to generalize patterns in both classes, resulting in more trustworthy and robust predictions.

## 5.3. Before and After SMOTE Class Distribution Comparison

the class distribution before and after using SMOTE is shown visually below:

Figure : class distribution

Original data: The class distribution in the original dataset is depicted in the bar plot above, emphasising any potential imbalances between various classes. The counts of observations for each level of the "Exited" variable are shown by the red bars.

Upsampled data: On the other hand, the class distribution is shown in the plot above after the minority class was upsampled using SMOTE. The counts of observations for each level of the "Exited" variable in the upsampled data are represented by the blue bars.

# 6. Machine Learning

## 6.1. Data splitting for Testing and Training

The upsampled dataset was partioned into two portions to adequately evaluate the model's performance: a training set and a testing set. This division was done with great care to provide a representative distribution of the data.

5,706 observations make up the training dataset, which is represented by the term train.

There are 2,446 observations in the testing dataset, which is represented by test.

## 6.2. Classification

The preprocessed data was subjected to the classification techniques. SVM and KNN classifiers are used to compare the outcomes. Additionally, the outcomes of the two classifiers were compared using the two feature selection techniques (SVN and KNN).

# 7. Support Vector Machine (SVM):

The fundamental objective of support vector machine (SVM) is to develop an effective identifying hyperplane that properly organises data points and separates points of two classes by reducing the chance of misclassifying training samples and unknown test samples.

## 7.1. Training and Evaluation of SVM Models:

This section describes how a Support Vector Machine (SVM) with a linear kernel is trained and evaluated for a classification task that involves the target variable "Exited." The analysis includes training the model, making predictions on the testing set, and thoroughly assessing the results using performance metrics and confusion matrices. The dataset was split into training and testing sets, with 70% of the data designated for training, in order to start the SVM model.

## 7.2. Confusion Matrix and Accuracy Calculation:

To evaluate the performance of the model, a confusion matrix was created. The percentage of correctly identified occurrences, or total accuracy, was determined to be 99.88%, see below confusion matrix and statistical table and svm linear grid graph.

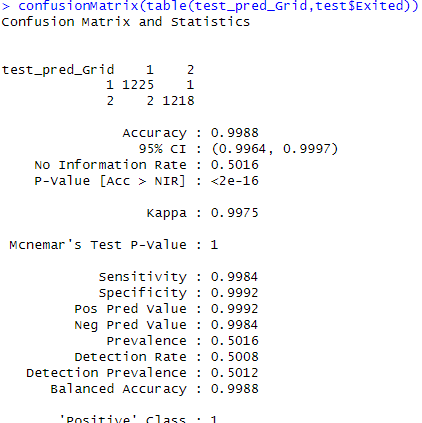


Figure : confusion matrix statistics

A graph with a line going up

Description automatically generated

Figure : svm linear grid graph

## 7.3. Conclusion

In summary, the SVM model demonstrated exceptional accuracy in categorising instances of the "Exited" variable, especially when paired with a linear kernel. The cost parameter's importance was further underscored during the optimisation process, as the final model operated at its best with a C value of 0.01. The foundation for the confident deployment and implementation of the SVM model to produce precise predictions in real-world scenarios is laid by this thorough evaluation.

# 8. K-Nearest Neighbor (KNN)

KNN works by detecting the k-closest samples from an existing dataset and classifying the new sample in the most similar class when a new unknown sample arises. In other words, the classification algorithm identifies the test sample group using the k training examples closest to the test sample and assigns it to the class with the highest likelihood.

In this part, we investigate how to classify occurrences of the "Exited" variable using the k-Nearest Neighbours (KNN) method. Model evaluation, training, and a comparative analysis with two distinct values of k are all included in the analysis.

## 8.1. Training and Evaluating the Model:

The KNN model was trained with 10-fold cross-validation and the train function. For this, the dataset (data2) was utilised, and "Exited" was the target variable.

## 8.2. Model Evaluation for Various k Values:

Two models, k = 75 and k = 76, were developed with varying k values. After each model's accuracy was assessed, it was found that both had a high accuracy of roughly 99.88% as seen below.

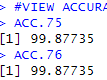


Figure : KNN accuracy

## 

## 8.3. Model Evaluation by Confusion Matrix

For this section, we examine a thorough model evaluation for two k-Nearest Neighbours (KNN) models with distinct k values (k = 75 and k = 76) utilising confusion matrices. The confusion matrix looks False negatives, false positives, real positives, and true negatives have been found to give comprehensive insights into how well the models are performing.

See below confusion matrix for both K=75, K=76 and a graph that shows different values of K.

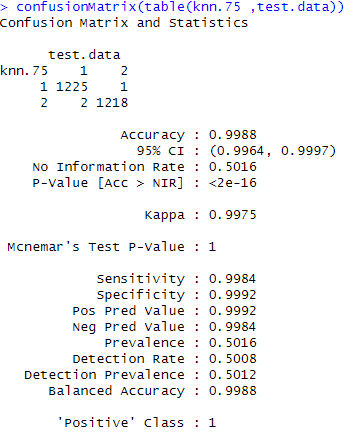


Figure : confusion matrix table(KNN.75)

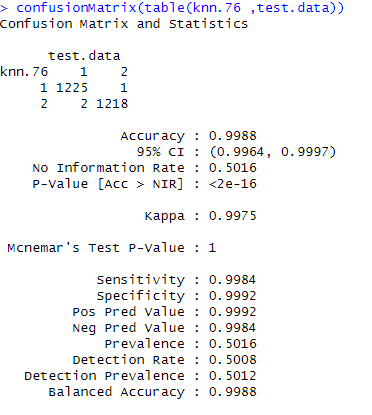


Figure : confusion matrix table(KNN.76)

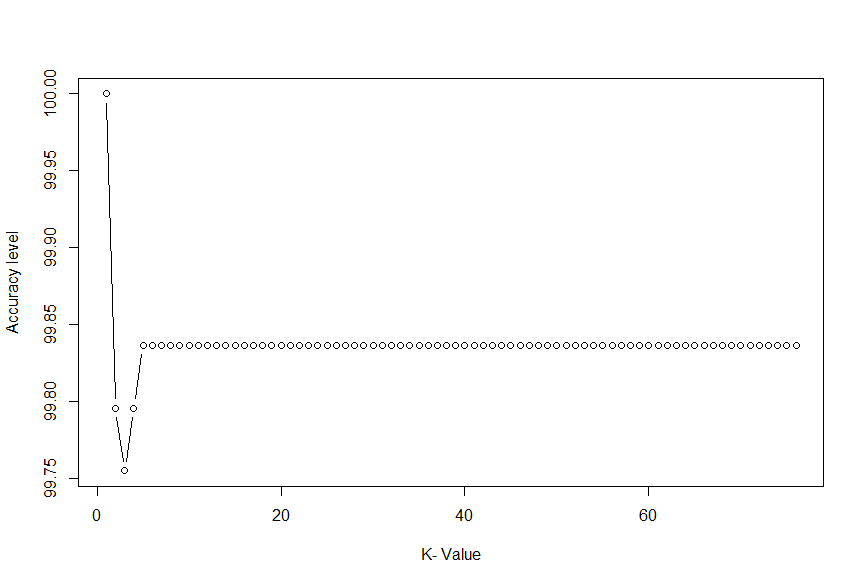


Figure : graph to show different values of K

## 8.4. Conclusion:

determining the perfect k value is aided by the visual depiction of accuracy across a range of k values. The accuracy and confusion matrix in the particular evaluation of the model with k = 1 shed light on how well the KNN algorithm functions with this dataset.

# 9. Comparison of SVM and KNN Models

Support Vector Machine (SVM) and k-Nearest Neighbours (KNN), two classification methods, are compared in this section. The "Exited" variable was used to train and assess the models, while accuracy and kappa values were used to compare the models' performance measures.

|  |  |  |
| --- | --- | --- |
| Metrix | SVM | KNN |
| Accuracy | Mean: 99.79% | Mean: 99.01% |
|  | Max: 99.95% | Max: 99.48% |
|  |  |  |
| Kappa | Mean: 99.58% | Mean: 98.03 |
|  | Max: 99.90% | 98.96% |

Figure : comparison table.

Both the SVM and KNN models work admirably, with high accuracy and kappa values. In terms of accuracy and kappa, the SVM model surpasses the KNN model on average. It is crucial to note, however, that the performance of these models may vary based on the peculiarities of the dataset and the task at hand. More fine-tuning and experimentation may be required to find the best model for the given categorization assignment.

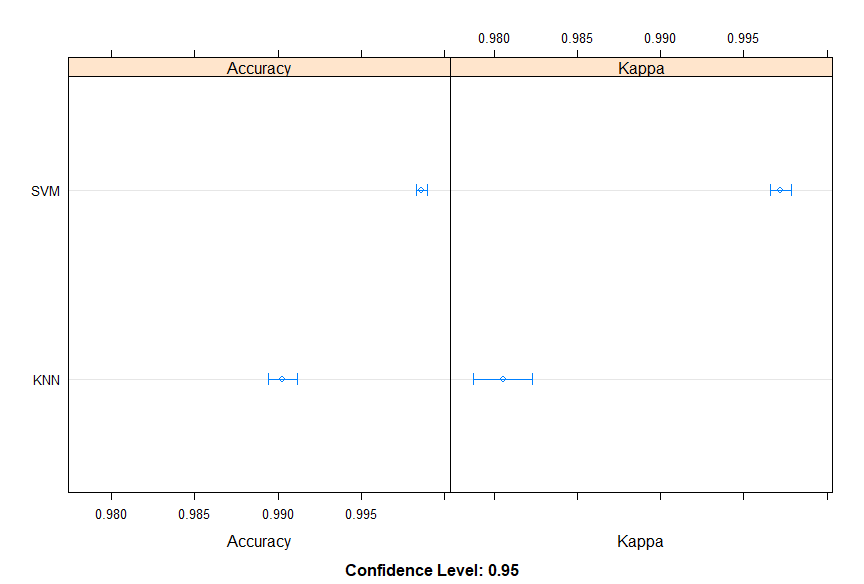


Figure : comparison graph

# Reference

# Bibliography

Dash, P. K., 2021. Predictive oc customer Churn in Telecom: A Machine Learning Perspective. *Computational Intelligence and Machine Learning,* p. 9.

M. A. Hassonah, A. Rodan, A. -K. Al-Tamimi and J. Alsakran, october 2021. Prediction of customer churn in telecom industry: A machine lear. *computational Intellegence and Machibe Learning,* p. 9.

Manas Rahman, V Kumar, 2020. Machine learning based customer churn prediction in banking. *Fourth International Conference on electronic, communication and aerospace technology(ICECA-2020),* p. 6.

**https://www.kaggle.com/datasets/bank-customer-churn**

# Appendices

1. Data description: The dataset, which was used to model churns, was downloaded from Kaggle. 10000 bank customers' details are included in the dataset, and a binary variable that indicates whether a customer is still a customer or has left the bank is the target parameter. 2037 of these samples were negative class (exited), and 7963 were positive class (kept).
2. Exploratory Data Analysis: Exploratory Data Analysis is effective for deeply comprehending data and gaining significant knowledge through data visualization.

Smote balance technique: The following parameters were utilized in the SMOTE implementation:

Oversampling percentage (perc.over): 100%

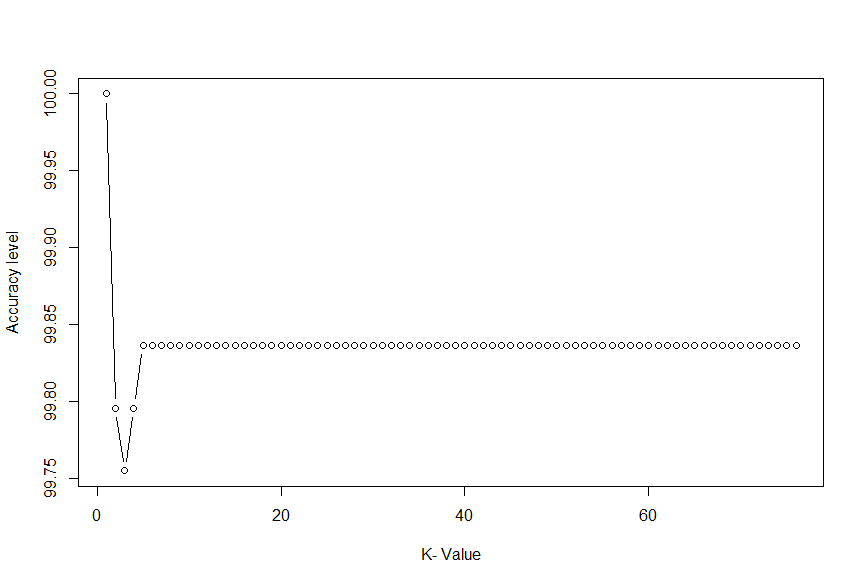
Under-sampling percentage (perc.under): 200%

The number of nearest neighbors (k):5.

1. This method ensures that the minority group is oversampled by 100% and the majority group is under-sampled by 200%, resulting in a more equal representation of both groups.

Data splitting: The upsampled dataset was partioned into two portions to adequately evaluate the model's performance: a training set and a testing set. This division was done with great care to provide a representative distribution of the data. 5,706 observations make up the training dataset, which is represented by the term train.

There are 2,446 observations in the testing dataset, which is represented by test.

1. Visualization of model performance: 
2. Model performance: The SVM model, especially with a linear kernel, performed admirably in categorising instances of the "Exited" variable. The optimisation process emphasised the importance of the cost parameter even further, with the final model obtaining optimal performance at a C value of 0.01.