# Banking Sales Services Optimization

**Poonam Bala[1] Kirti Bhatia[2] Shalini Bhadola[3]**
[1]Student [2,3]Assistant Professor
[1,2,3]Sat Kabir Institute of Technology and Management, Haryana, India

*Abstract*— The project aims to optimize banking sales services using machine learning and deep learning models. Three focus areas were used to optimize the bank´s services. First is to predict the estimated salary of bank customers and offer them better services corresponding to their income. The second area is to predict customers who might leave the bank and focus more on them so that they will keep the bank´s services. The third area of research is to find a group of similar customers and classify them as valuable/non-valuable and loyal/non-loyal so that the bank can focus on valuable and loyal customers.To achieve this research result, different regression models such as Decision Tree Regression, Random Forest Regression, Polynomial and Support Vector Regression were used to predict the estimated salary of the customers. Several classification models such as Logistic, K-NN, Naive Bayes, Decision Trees Classifier, Random Forest Classifier, and SVM were used to predict the churn rate i.e. customers who might leave the bank. Later XGBoost model was used for better performance and faster execution to predict churn rate. A deep learning model - Artificial Neural Network has been used for better accuracy. To find a similar group of customers, unsupervised machine learning techniques such as K-Means and Hierarchical clustering were used. Several different evaluation techniques such as confusion matrix, R-Squared, Adjusted R-Squared and SSE were used to assess the performance of the models. K Cross-validation was used to confirm whether a model is overfitting or underfitting.

***Keywords:*** Banking Sales Optimization, Artificial Neural Network, Churning, Machine learning, Impact Factor

## I. INTRODUCTION

In recent years, Machine learning (ML) technologies are used tremendously to build effective sales automated solutions. There is a vast amount of sales strategies adopted by companies all over the world to optimize their services in different scenarios by using machine learning methods. Machine learning is the most attractive approach for the banking business to compare and optimize their services where they must pay more attention to several parameters while setting the offers. ML and deep learning models can help the financial organizations to understand the information better and detect prominent demands of the customers by applying a suitable algorithm for the given data. Nowadays, churning is the major issue facing companies all over the world, especially for banks, telecommunication, and insurance companies. Churn is defined as a process where customers exit the company due to dissatisfaction with the services provided or due to better offers provided by the other companies. This would be a considerable loss of that organization. To avoid churning and retain the existing clients rather than acquiring new ones is the main prediction task which can be done by some sophisticated techniques like decision tree, Random Forest (RF), logistic, naïve Bayes, SVM classification, Support Vector (SVM), XGBoost model and artificial neural network. By this, we can easily detect the dissatisfaction of the customers towards banks and make a way to improve the services. Here, we are introducing some vital prediction/forecasting techniques on the given data for predicting future sales by considering some parameters. Some major algorithms are used for the sales prediction for the optimization of the banking sales services department. The predictive models which are used in this work are parts of regression analysis, classification, and clustering.

## II. RELATED WORK:

Churn Prediction is a method of predicting customers who are likely to leave the organization or cancel their subscription with the organizations. Churn Prediction is used widely in different aspects of business and has a wide range of applications.

Churn Prediction in Banking Industry: Customers are the most valuable assets to banking organizations. In the current situation, most of the banks are digitalized and provide electronic banking services, which reduce the direct interactions with banking customers. In this situation, providing customer satisfaction is a challenging task for the banks and customers may probably churn. Thus predicting customer's churn is very important in the banking sector.

Related work in Banking Industry: U.F. Eze proposed a data mining model used to predict customers who likely to churn. The analysis was carryout using real-life customer records provided by a major Nigerian bank. K-Means was used to cluster customers. The results showed that the methods used could determine patterns in customer behaviours and help banks to identify likely churners and hence develop customer retention modalities (Eze, Onwuegbuchular and Diala, 2017).

Churn Prediction in Service Industry: The ratio of customers churning in the service industries such as media, entertainment, and gaming is very high. Customers may gradually leave the platforms due to loss of interest. Thus, predicting customer churn is very important to online service providers. By analyzing and understanding the customer's behaviours who are likely to churn and provide them with good services and keeping them entertained would be beneficial for the organizations.

Related work in Media and Entertainment Industry: A case study by Juzi entertainment tried to predict the churners from the Juzi platform. The dataset included ten thousand of their customer entries and built the churn model. Results showed that "gif and Beauty" of Juzi entertainment were more popular. The customers churning from these subscriptions were low. The results were then used in the improvement of products by Juzi entertainment.

## III. FUTURE WORK:

The Churn predication model for now can be used in any bank to understand the behavior patterns of churn customers

All rights reserved by *www.ijsrd.com*

and non customers in advance and also to predict the churn. In future same methodologies can be implemented to other financial institutes to other financial institutes or organizations such as sales organizations, retail organizations etc.

## IV. Research Methodology used in Banking Sales Services optimization:-

### A. Regression:

Regression analysis is a form of predictive modelling technique that investigates the relationship between a dependent (the variable we are trying to predict) and independent (the variable we use to predict the dependent) variables.

*1) Simple Linear Regression:*

Simple linear regression studies the linear relation between the one independent and one dependent variable. The independent variable is also known as the predictor variable, and the dependent variable is also known as the response variable. The relationship can be expressed as = $(b0 + b1)$ Where y is a dependent variable, x is an independent variable, b1 is the slope of the regression line, and b0 is the intercept of the regression line.

*2) Multiple Linear Regression:*

Multiple regression is an extension of simple regression and allows us to consider more variables to get a better prediction. In multiple linear regression, we have two or more independent variables and one dependent variable. The mathematical term can be expressed as:

$y = b0 + b1\,x1 + b2\,x2 + \dots\dots\dots + bn$, Here $x1, x2$ $xn$ is an independent variable, y is a dependent variable, b0 is the y-intercept of the regression line, and b1 is the slope of the line.

*3) Polynomial Regression:*

Polynomial regression is like simple linear regression, the relationship between the independent variables and dependent variables is an nth degree.

$y = b0 + b1\,x1 + b2\,x^2 + \dots\dots\dots + bnx^n$ Where y is the predicted outcome, b1, b2. bn are the coefficients of regression, b0 is the y- intercept. The model is simply formed of simple linear regression with n predictors raised to the power of i where i = 1 to n.

*4) Support Vector Regression:*

SVR has a different regression compared to linear regression. In linear regression, we are trying to minimize the error between the predictor and data. In SVR, our goal is to make sure that errors do not exceed the threshold. The below figure shows that the blue line is the hyper Plane and the red line is the boundary line.

*5) Decision Tree Regression:*

Decision tree regression is like the decision tree classification; however, it uses a coefficient of determination or any similar metrics instead of cross-entropy or Gini impurity to determine splits. The parts of the decision tree regressor contain a root node, child or intermediate nodes and leaf nodes. Root nodes, child/intermediate nodes are the attributes or attribute values used to split the data and leaf nodes contain the target values.

*6) Random forest regression:*

Random forest regressor model is the bagging algorithm which forms an ensemble of regressors. Bagging is the average of predictions made by several estimators which are built individually on subsets of data. It reduces the variance with little or no effect on bias. Smaller models are combined to make a base model which is also known as an ensemble model where the base models perform higher than the small models.

### B. Classification:

Classification is an approach that is used to classify the customers concerning the exited rate (churn or not) and to predict the customer churn rate. Classification techniques are used to build the prediction model to predict customer behaviour through classifying churner and non-churner customers. The bank dataset used for this study consists of 10,000 observations with 14 attributes.

The main objective of the classification is to classify the customers into two classes, i.e. churn or non-churner and predicting the customer class whether they are going to churn or not by considering all the attributes of the dataset.

Classification Models/Techniques:

As discussed before, Classification is a technique used to predict the class of the customers (churner and non-churner). To implement the models, the dataset was divided into a training set and a test set. To build a model, a training set was used, and for churn prediction, the test set was used.

*1) Random Forest (RF):*

Random Forest is one of the best classification models used for the prediction of the customer churn class (Churner and non-churner). This algorithm gives essential information regarding those variables which are affecting the target variable (exited). This supervised technique consists of decision trees, and the outcome of the classifier is obtained by taking the average of the individual decision trees.

*2) Naive Bayes (Bayes):*

Naive Bayes classifier is an easy build model based on the Bayes theorem. It is one of the most often used classification techniques for massive datasets. The theorem, which is used in this model is written below.

Bayes Theorem:

$$P(A/B) = P(B/A)\,P(A)/P(B)$$

*3) Support Vector (SVM) Machines:*

Support Vector Machines (SVM) is a supervised learning algorithm that is used for the classification of the dataset. SVM constructs the hyperplane in an infinite-dimensional space to separate or distinguish two classes. The Support vectors are the coordinate points of all the observations of the dataset.

*4) K-Nearest Neighbours (KNN):*

K-Nearest Neighbours (KNN) is also one of the most used classification algorithms in machine learning. To build the churn prediction model, KNN is applied. We can use this technique without having prior information regarding the data distribution. It is used to classify the group of data points based on the similarity measure. It tries to find out if a point is in group A or B depends on one point on a grid.

All rights reserved by www.ijsrd.com

*5) Decision Tree Classification:*

A decision tree can be used for regression as well as a classification model, and it is a form of the tree structure of the dataset. It is a supervised machine learning algorithm used to split the customer between who is going to leave the bank or not, based on certain parameters like customer credit score, country, gender, and salary income.

*6) Logistic Regression:*

Logistic Regression is a supervised learning algorithm which is used for the classification issues, and it is based on the probability concept. To evaluate the above-mentioned classification models, confusion matrix and accuracy must be considered. To predict the customer churn, the results and analysis will be discussed in later sections. They are the most important parameter to evaluate the performance of classification models.

a) Confusion matrix:

The confusion matrix performs the models and information about the actual and predicted cases. It is a matrix of 2x2 in which rows represent the predicted values, and the columns represent the actual values obtained by the models. (Esther van der Knaap and Brian Ramsay, 2011)

Actual values

| | Positive(1) | Negative(0) | |
|---|---|---|---|
| Positive(1) | TP | FP<br>P | Redicted values |
| Negative(0) | FN | TN | |

Table 1: Confusion Matrix

b) Accuracy:

The accuracy of the classification model is termed as the indicator to predict the class of the dependent variable (i.e. customer exited or not) for future trends. It is observed by the number of correct classifications on the test set made by the classification models.

The accuracy of the classification model can be calculated based on a confusion matrix with the help of the formula which is given below.

Accuracy= TP+TN/TP + TN+FP+FN

### *C. Clustering:*

Customers churning are one of the major problems concerning financial institutions like banks. The cost of getting a new customer into the bank is costlier than retaining old customers. Predicting the customer's behaviours and clustering them well in advance before they churn would reduce the customer attrition rate. Customer clustering is a process of the grouping of customers into smaller groups with similar behavioural features. Customers are clustered based on different features of customers such as age, country, balance, estimated salary, tenure, etc. These clustering methods aim to cluster all the similar group of customers and provide various services to increase customer satisfaction and reduce the customer attrition rate.

*1) K-Means Clustering:*

To cluster the customers and segregate them into similar groups, the K-Means clustering model was implemented. K-means clustering is an iterative clustering method and divides the data into several clusters by minimizing an error function. It is a non- hierarchical clustering method that first involves the identification of several clusters the customers must be divided.To determine the number of clusters, the WCSS (minimizing the within-cluster sum of a square) concept is used.

*2) Hierarchical Clustering:*

Hierarchical clustering is a clustering algorithm that forms a hierarchy of clusters. The algorithm starts with each data points assigned to separate clusters. The two closest clusters at the bottom are merged until a single cluster is formed. The distances between two clusters in the data space is determined by the height at which two clusters are merged.

*3) Artificial Neural Network:*

Deep learning is the most exciting and powerful branch of Machine Learning. It helps a machine to do the same function as human neurons do. Deep learning getting lots of attention because it is giving the results out of the box that was not possible earlier.

An artificial neural network is used to find the prediction of the churning rate. Keras machine learning library has been used to implement this model. A function name "Sequential" is used to initiate the neural network, and the "Dense" function is used to add a hidden layer to train the model by adjusting the weights of the network. Later, the network has been compiled using "adam" optimizer and "binary_crossentropy" as loss function and the encoded input variables fitted into this network. This model has been run for 1000 epoch to get better performance

*4) Tools:*

The following tools have been used:

- Jupyter Notebook
- Python
- Keras Library
- Pandas Library
- Numpy library
- Sklearn library
- Matplotlib library

*5) Implementation:*

Implementation is done in python using several python packages and machine learning algorithms. Steps included in our implementation process are as follows: Data preparation: The dataset used for our analysis is taken from Kaggle in CSV format.It is a churn modelling dataset of online banking based out in Europe.The dataset has 10000 observations with 14 attributes (dependent and independent variables) as below:

All rights reserved by *www.ijsrd.com* 

```
data.head(10)
```

| Number | CustomerId | Surname | CreditScore | Geography | Gender | age | Tenure | Balance | NumOfProducts | HasCrCard | IsActiveMember | EstimatedSalary | Exited |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 15634602 | Hargrave | 619 | France | Female | 69 | 2 | 0.00 | 1 | 1 | 1 | 50441.60946 | 1 |
| 2 | 15647311 | Hill | 608 | Spain | Female | 52 | 1 | 83807.86 | 1 | 0 | 1 | 71089.41795 | 0 |
| 3 | 15619304 | Onio | 502 | France | Female | 41 | 8 | 159660.80 | 3 | 1 | 0 | 75408.22488 | 1 |
| 4 | 15701354 | Boni | 699 | France | Female | 29 | 1 | 0.00 | 2 | 0 | 0 | 71647.89894 | 0 |
| 5 | 15737888 | Mitchell | 850 | Spain | Female | 35 | 2 | 125510.82 | 1 | 1 | 1 | 45814.73657 | 0 |
| 6 | 15574012 | Chu | 645 | Spain | Male | 38 | 8 | 113755.78 | 2 | 1 | 0 | 68206.90508 | 1 |
| 7 | 15592531 | Bartlett | 822 | France | Male | 57 | 7 | 0.00 | 2 | 1 | 1 | 68206.90508 | 0 |
| 8 | 15656148 | Obinna | 376 | Germany | Female | 43 | 4 | 115046.74 | 4 | 1 | 0 | 30773.86497 | 1 |
| 9 | 15792365 | He | 501 | France | Male | 57 | 4 | 142051.07 | 2 | 0 | 1 | 40433.84016 | 0 |
| 10 | 15592389 | H? | 684 | France | Male | 60 | 2 | 134603.88 | 1 | 1 | 1 | 102245.78010 | 0 |

Table 1: Dataset Records

## V. DATA VISUALIZATION:

Dataset has been visualized to get insights into the data.

### *A. Exit Rate: Gender*

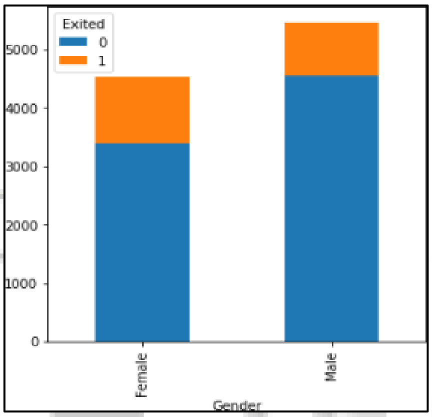



Fig. 1: Exit rate Gender

We can see from the above figure that females are more intend to leave the bank, but they are less in number compared to male customers.

### *B. Exit Rate: Tenure*

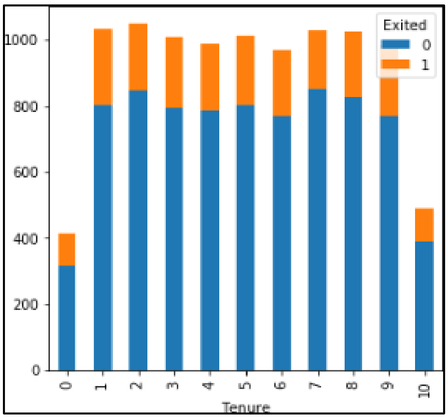



Fig. 2: Exit rate Tenure

We can see from the graph that customers who have a low or high number of years of experience are less intended to leave the bank.

### *C. Exit Rate: Geography*

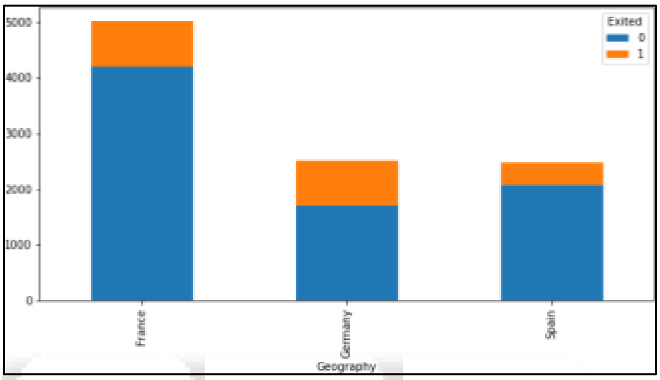



Fig. 3: Exit rate Geography-wise

We can see from the graph that customers from Germany are more intend to leave the bank, but France customers are more.

## VI. EXPERIMENTAL RESULTS:-

The different models from regression, classification, clustering and deep learning are evaluated as below:

### *A. Regression:-*

#### *1) Simple Linear regression*

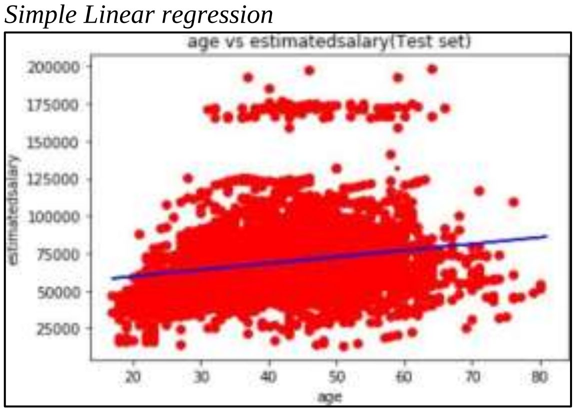



Fig. 1: Linear Regression for estimated salary

*2) Multiple Linear Regression*

| | | | |
|---|---|---|---|
| Dep. Variable: | y | R-squared: | 0.038 |
| Model: | OLS | Adj. R-squared: | 0.038 |
| Method: | Least Squares | F-statistic: | 398.6 |
| Date: | Fri, 20 Dec 2019 | Prob (F-statistic): | 5.41e-87 |
| Time: | 13:04:41 | Log-Likelihood: | -1.1563e+05 |
| No. Observations: | 10000 | AIC: | 2.313e+05 |
| Df Residuals: | 9998 | BIC: | 2.313e+05 |
| Df Model: | 1 | | |
| Covariance Type: | nonrobust | | |

Fig. 2: Multiple Linear Regression for estimated salary

*3) Polynomial Regression*

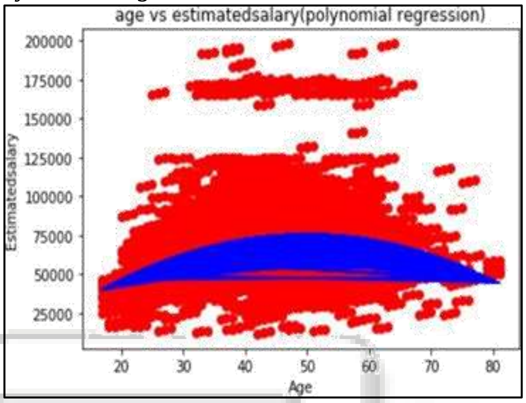



Fig. 3: polynomial regression for estimated salary

*4) Support Vector Regression*

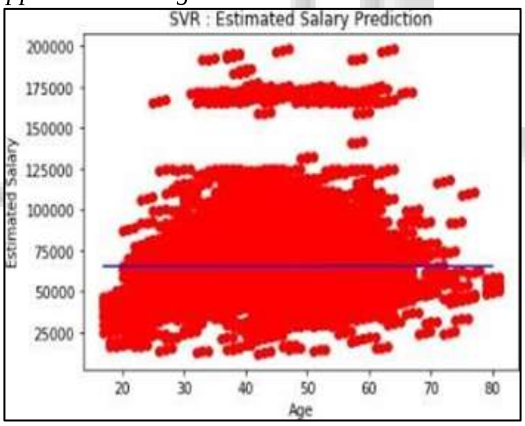
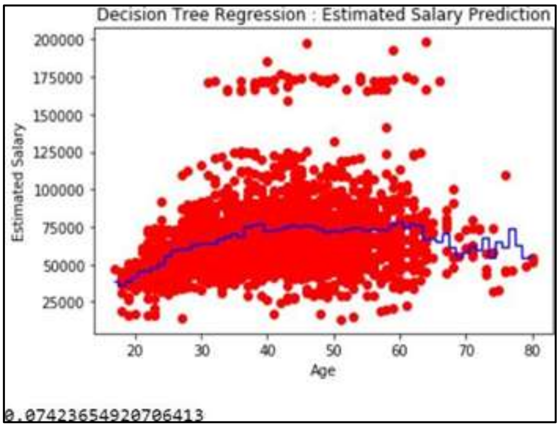



Fig. 4: Support Vector regression for estimated salary

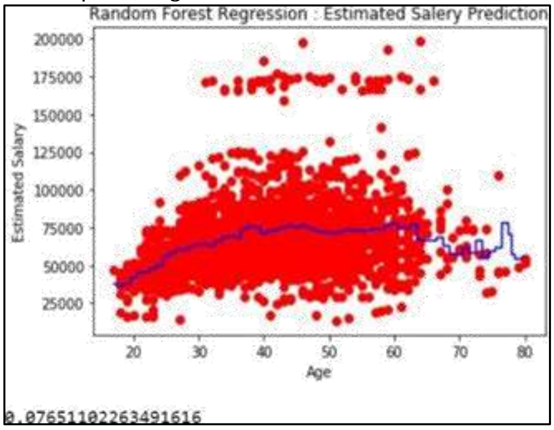
*5) Decision tree regression*




0.07423654920706413

Fig. 5: Decision tree regression for estimated salary

*6) Random forest regression*




=0.07651102263491616

Fig. 6: Random forest regression for estimated salary

*B. Classification*

*1) Random Forest (RF) Classification*

```
Out[34]: array([[1734,  257],
                [ 421,   88]], dtype=int64)

In [19]: def accuracy(confusion_matrix):
             diagonal_sum = confusion_matrix.trace()
             sum_of_all_elements = confusion_matrix.sum()
             return diagonal_sum / sum_of_all_elements
         accuracy(cm)

Out[19]: 0.7288
```

*2) Naive Bayes (Bayes):*

```
Out[3]: array([[1855,  136],
               [ 358,  151]], dtype=int64)

In [4]: def accuracy(confusion_matrix):
            diagonal_sum = confusion_matrix.trace()
            sum_of_all_elements = confusion_matrix.sum()
            return diagonal_sum / sum_of_all_elements
        accuracy(cm)

Out[4]: 0.8024
```

*3) Support Vector (SVM) machine:*

```
Out[2]: array([[1932,   59],
               [ 376,  133]], dtype=int64)

In [3]: def accuracy(confusion_matrix):
            diagonal_sum = confusion_matrix.trace()
            sum_of_all_elements = confusion_matrix.sum()
            return diagonal_sum / sum_of_all_elements
        accuracy(cm)

Out[3]: 0.826
```

*4) K-Nearest-Neighbours (KNN):*

```
Out[89]: array([[1787,  204],
                [ 436,   73]], dtype=int64)

In [104]: def accuracy(confusion_matrix):
              diagonal_sum = confusion_matrix.trace()
              sum_of_all_elements = confusion_matrix.sum()
              return diagonal_sum / sum_of_all_elements
          accuracy(cm)

Out[104]: 0.744
```

All rights reserved by *www.ijsrd.com* 

*5) Decision Tree Classification:*

```
Out[15]: array([[1651,  340],
                [ 284,  225]], dtype=int64)

In [16]: def accuracy(confusion_matrix):
             diagonal_sum = confusion_matrix.trace()
             sum_of_all_elements = confusion_matrix.sum()
             return diagonal_sum / sum_of_all_elements
         accuracy(cm)

Out[16]: 0.7504
```

*6) Logistic Regression:*

```
Out[53]: array([[1571,   24],
                [ 380,   25]], dtype=int64)

In [54]: def accuracy(confusion_matrix):
             diagonal_sum = confusion_matrix.trace()
             sum_of_all_elements = confusion_matrix.sum()
             return diagonal_sum / sum_of_all_elements
         accuracy(cm)

Out[54]: 0.798
```

*7) Comparison of Classification Models:*

| CLASSIFICATION MODELS | PREDICTION ACCURACY |
|---|---|
| Random Forest (RF) | 72.8% |
| Naive Bayes (Bayes) | 80.2% |
| Support Vector (SVM) | 82.6% |
| K-Nearest-Neighbors (KNN) | 74.4% |
| Decision Tree Classification | 75.04% |
| Logistic Regression | 79.8% |

Table 2: Performance Accuracy of Classification Models

Here, the results of the six-classification algorithm have been analyzed to find the best model. As we can see in the below Table 2 and graph, Support Vector (SVM) machine algorithm gives the best/highest accuracy, i.e. 82.6%, Naive byes gives 80.2% accuracy which is the second-highest followed by Logistic Regression, Decision Tree, KNN and the Random Forest(RF)classifier gives the lowest accuracy with 72.8%.

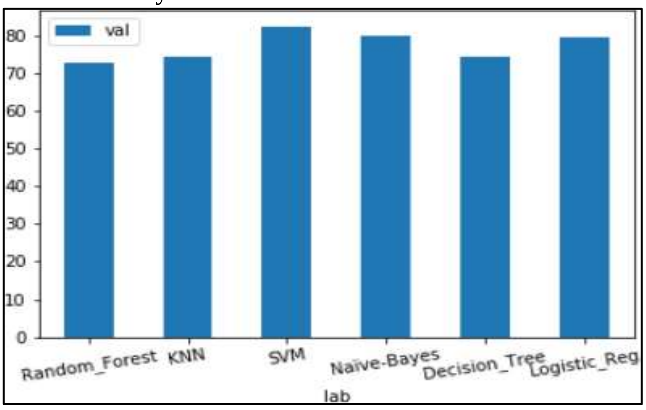



Fig. 7: Graph of Classification Model's comparison

*8) XGBoost Regression Model*

```
Out[8]: array([[1522,   73],
               [ 194,  211]], dtype=int64)

In [11]: from sklearn import metrics
         accuracy=metrics.accuracy_score(y_test, y_pred)
         accuracy

Out[11]: 0.8665
```

*C. Clustering*

*K-Means Clustering*

*1) K-means clustering on Age and Balance*

Clustering on age and balance was implemented to cluster the most, average and least valuable customers.

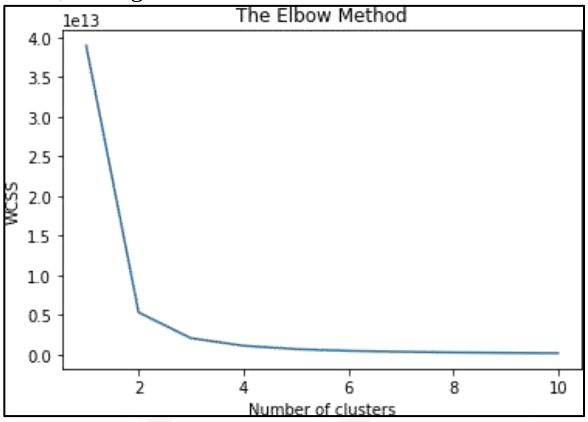



Fig. 8: WCSS for clustering on age and balance variables

We can see from the above graph that the optimal numbers of a cluster are 3.So setting the number of clusters to three seems to provide meaningful customer segregation as below:

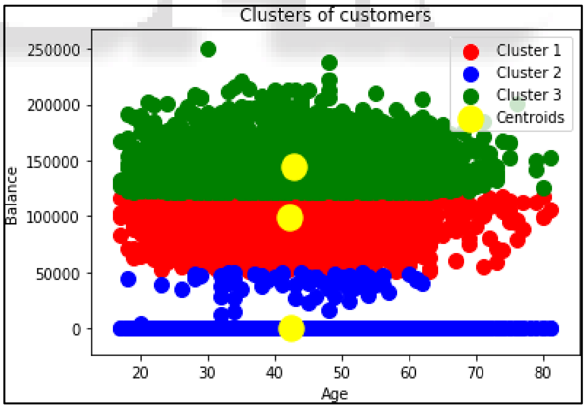



Fig. 9: Clustering on age and balance

Cluster 1:

- The priority level for the bank: Medium.
- A primary concern of people in this cluster: Services from banks to increase their balance.
- Our recommendations to prevent churn: Provide better interest rates.
  Offer attractive options to invest their balance money to increase their returns.

Cluster 2:

- The priority level for the bank: Low.
- The primary concern of people in this cluster: Protection of their existing balance with the bank.
- Our recommendations to prevent churn:

All rights reserved by *www.ijsrd.com* 

Provide better interest rates.
Strengthen current services with minimum investment.

Cluster 3:

– The priority level for the bank: High.
– The primary concern of people in this cluster: Attractive investment options to park their balance money.
– Our recommendations to prevent churn:
  Offer a range of attractive investment options with high returns.
  A customized and tailored approach to understanding individual concerns.

*2) K-means clustering on Age and Credit score*

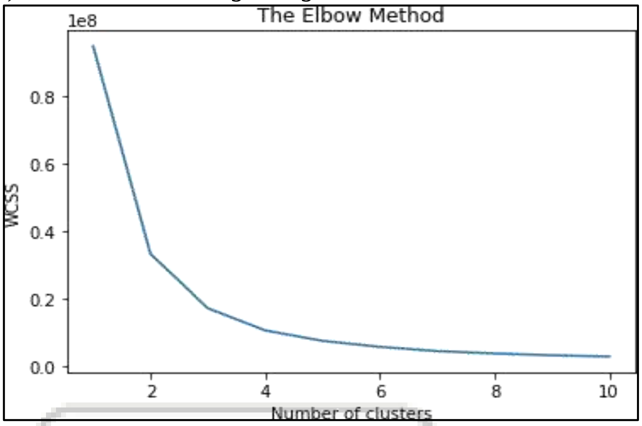



Fig. 10: WCSS for clustering on age and credit score variables

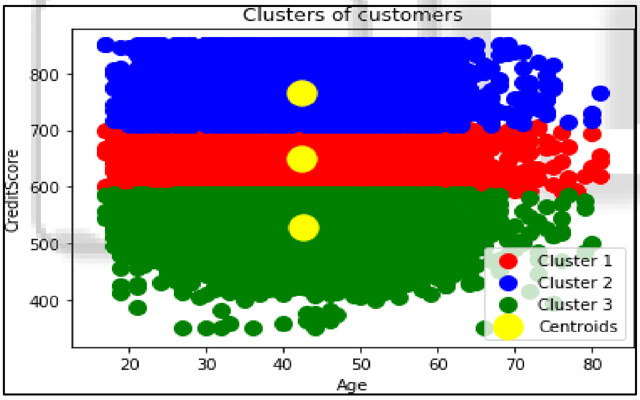



Fig. 11: Clustering on age and credit score

Cluster 1:

– The priority level for the bank: Medium.
– The primary concern of people in this cluster: Loans with low-interest rates and the opportunity to increase their credit score.
– Our recommendations to prevent churn:
  Offer loans with an average rate of interest (competitive to other banks).
  Provide credit card options for their daily transactions to offer them an opportunity to increase their credit score.

Cluster 2:

– The priority level for the bank: High.
– The primary concern of people in this cluster: Low-interest rates.
– Our recommendations to prevent churn:
  Low-interest loans and credit card options.
  Tie-ups with other companies to make credit cards more attractive (e.g., cashback offers).

Cluster 3:

– The priority level for the bank: Low.
– The primary concern of people in this cluster: Opportunity to increase credit score, relatively average interest rates.
– Our recommendations to prevent churn:
  Provide competitive interest rates (compared to competitors).
  Offer credit card options with a fine balance between the bank´s profitability and an acceptable interest rate for the clients in this cluster.

*3) K-means clustering on Age and Estimated salary*

Customers were segregated based on the age and their estimated salary to determine the most valued customers.

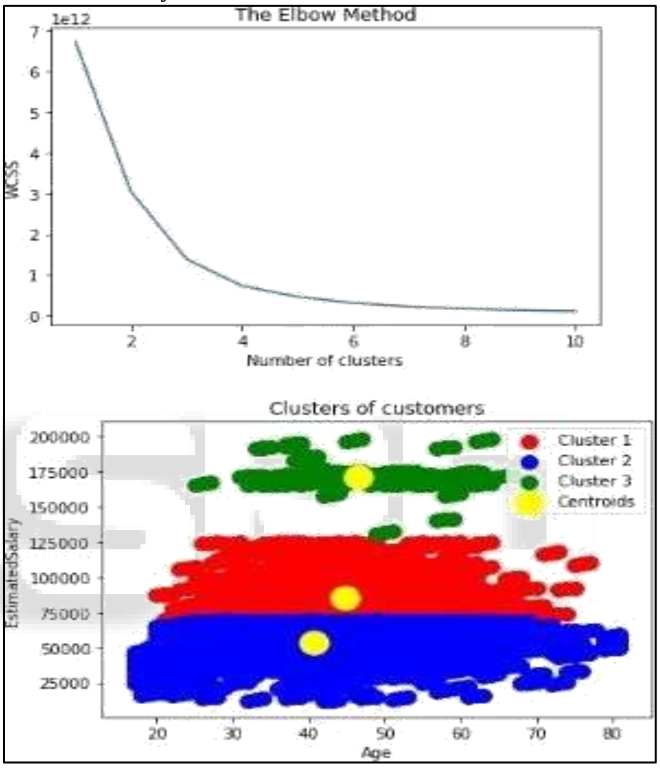



Fig. 12: Clustering on age and estimated salary

Cluster 1:

– The priority level for the bank: Medium to High.
– The primary concern of people in this cluster: Attractive options to invest their salary in.
– Our recommendations to prevent churn:
  Provide a range of investment options to utilize their savings.
  Offer attractive salary account options.

Cluster 2:

– The priority level for the bank: Low.
– The primary concern of people in this cluster: Protection of their existing salary. Our recommendations to prevent churn:
  Offer safe bank deposit options (savings bank account).
  Increase customer satisfaction value and to make sure to increase customers' interest in banks.

Cluster 3:

– The priority level for the bank: High.
– The primary concern of people in this cluster: Bank with customized relationship management.

All rights reserved by *www.ijsrd.com*

– Our recommendations to prevent churn:
  Provide personal relationship managers to offer an advisory service to the clients.
  Offer excellent salary account options to ensure that a majority of their salary is parked in our bank.

*4) Hierarchical clustering*


Fig. 13: Hierarchical clustering on age and credit score.

## VII. ARTIFICIAL NEURAL NETWORK

```
array([[1485,  110],
       [ 237,  168]])

[ ] def accuracy(confusion_matrix):
        diagonal_sum = confusion_matrix.trace()
        sum_of_all_elements = confusion_matrix.sum()
        return diagonal_sum / sum_of_all_elements
    accuracy(cm)

0.8265
```

## VIII. CONCLUSION

In regression, after applying every possible model for testing the accuracy of the estimated salary of the customers, we concluded that random forest is working more efficiently when compared to other models on this dataset. But further improvement is required in data pre- processing (a better dataset) before fitting the model for getting better results for the R squared score.

In classification, several models were implemented and compared to classify the customers into two classes (churner or non-churner) and also to predict whether the customer will stop using the services of the bank and switch to some other bank. The usage of models helps us to predict churning of customers correctly and identifying the attributes which lead to retaining them. Finally, we can conclude that the XGBoost classifier model would be the best algorithm for the churn prediction model.

In Deep Learning Model, an artificial neural network is performing well for 1000 epochs but not as much as the XGBoost model is performing. A higher number of epochs could help model for a better result but the execution time will grow up exponentially.

Clustering was carried out to cluster a similar group of customers based on different features such as age, balance, credit score, etc. The main aim of clustering was to cluster a similar group of customers and to investigate most, average and least valuable customers. By doing this, banks can identify different classes of customers and provide benefits and services to increase customer satisfaction and to retain customers who may probably churn.

## REFERENCES

[1] Neural designer. Churn prevention (2020) [online] Available at: https://www.neuraldesigner.com/solutions/churn-prevention .
[2] Neuroshell. The R-squared formula, (2008). Availableat:http://www.wardsystems.com/manuals/neuroshell2/index.html?rsquared.htm.
[3] Medium. (2020).Simple and Multilinear Regression. [online] Available at:https://towardsdatascience.com/simple-and-multiple-linear- regression-in-python- c928425168f9.
[4] Medium.(2020).Support vector regression.[online] Available at: https://medium.com/coinmonks/support-vector-regression-or-svr-8eb3acf6d0ff.
[5] Abhay Kumar.Data Science and Artificial intelligence.Random forest regression [image] Available at:https://acadgild.com/blog/random-forest-algorithm-regression .
[6] Oyeniyi, A. and Adeyemo, A. (2020). Customer Churn Analysis In Banking Sector Using Data Mining Techniques. [online] Semanticscholar.org. Available at: https://www.semanticscholar.org/paper/Customer-Churn-Analysis-In-Banking-Sector- Using- Oyeniyi-Adeyemo/0fd0067775eef2aa52547229c17e06128fdc0633 .
[7] Anon, (2020). [online] Availableat:https://www.academia.edu/21524961/Data_mining_application_in_banking_se ctor_wit
[8] Ieeexplore.ieee.org. (2020). Data mining application in the banking sector with clustering and classification methods - IEEE Conference Publication. [online] Available at: https://ieeexplore.ieee.org/document/7093731 [Accessed 16 Jan. 2020].
[9] Medium. (2020). Naive Bayes ( Bayes) Classifier. [online] Available at:https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c .
[10] DataFlair Team. “Support Vector (SVM) Machines Tutorial - Learn to Implement SVM in Python DataFlair.” DataFlair, 4 Aug. 2017, data-flair.training/blogs/svm- support-vector-machine-tutorial/.
[11] GeeksforGeeks. (2020). K-Nearest Neighbours - GeeksforGeeks. [online] Available at: https://www.geeksforgeeks.org/k-nearest-neighbours/ [Accessed 4 Jan. 2020].
[12] Your Bibliography: Anon, (2020). [image] Available at: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 [Accessed 4 Jan. 2020].
[13] Medium. (2020). Artificial Neural Network. [online] Available at: https://towardsdatascience.com/introduction-to-artificial-

neural-networks-ann- 1aea15775ef9 [Accessed 4 Jan. 2020].
[14] XGBoost Model. [online] Available at: https://xgboost.readthedocs.io/en/latest/tutorials/model.html [Accessed 5 Jan. 2020].
[15] Keramati, Abbas & Ghaneei, Hajar & Mirmohammadi, Seyed. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. Financial Innovation.
[16] Anil Kumar D, Ravi V (2008) Predicting credit card customer churn in banks using data mining. Int J Data Anal TechStrateg.
[17] Han J, Kamber M, Pei J (2012) DATA MINING Concepts and Techniques, 3rd edn. Morgan Kaufmann, USA.
[18] Huang B, Kechadi MT, Buckley B (2012) Customer churn prediction in telecommunications. Expert Syst Appl 39:1414–1425.
[19] Keramati A, Ardabili SMS (2011) Churn analysis for an Iranian mobile operator.
[20] Keramati A, Jafari-Marandi R, Aliannejadi M, Ahmadian I, Mozzafari M, Abbasi U (2014) Improved churn prediction intelecommunication I ndustry using data mining techniques. Appl Soft Comput 24:994–1012.
[21] Kumbhar VM (2011) Factors affecting the customer satisfaction in e-banking: some evidences from Indian banks