INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1683-1687

CHURN PREDICTION USING MACHINE LEARNING IN BANKING SECTOR

Sonu Sharma[1], K. Ajith kumar[2], R. Bharath Kumar[3], MD. Imaduddin[4]

[1]Assistant Professor, Information Technology, ACE Engineering College, India.

[2,3,4]Information Technology, ACE Engineering College, India.

ABSTRACT

The number of service providers are increasing very rapidly in every business. In these days, there is no shortage of options for customers in the banking sector for choosing where to put their money. As a result, customer churn and engagement has become one of the top issues for most of the banks. A method to predicts the customer churn in a Bank, using Machine Learning techniques, which is a branch of Artificial Intelligence is proposed. The research promotes the exploration of the likelihood of churn by analyzing customer behavior by using Machine Learning Algorithms like KNN, SVM, Decision Tree, and Random Forest classifiers. Also, some feature selection methods have been done to find the more relevant features and to verify the system performance. The experimentation was conducted on the churn modeling dataset from Kaggle. The results are compared to find an appropriate model with higher precision and predictability. As a result, the use of the Random Forest model after oversampling is better compared to other models in terms of accuracy.

Keywords: Python, Machine Learning

1. INTRODUCTION

The market is very dynamic and highly competition now a days. It is because of the availability of a large number of service providers. The challenges of service providers are finding the changing customer behavior and their rising expectations. The raising aspirations of current generation consumers and their diverse demands for connectivity and innovative, personalized approaches are very distinct from previous generations of consumers. They are well educated and better informed of emerging approaches. Such advanced knowledge has changed their purchasing behavior, resulting in a trend of 'analysis-paralysis' over-analyzing the selling and purchase scenario, which ultimately helps them to improve their purchase decisions. Therefore, this is a big challenge for the new generation service providers to think of innovatively to fulfil and add values to the customers. Corporations need to recognize their consumers. Liu and Shih strengthen this argument by implying that increasing competitive pressures on organizations to develop innovative marketing approaches, to meet consumer expectations and enhance loyalty and retention. Canning argues that offering more to all is no longer a viable sales strategy, and a market environment that continues to become more competitive needs an agenda that emphasizes on the most effective use of marketing capital. Technology has been used to help businesses to retain a competitive edge.Data mining techniques are a commonly used information technology for the extraction of marketing expertise and further guidance for business decisions. It is very easy for customers to switch from one organization (Bank) to another for a better service quality or price rates. Organizations are convinced that recruiting new customers is far more expensive and harder than keeping existing clients. But delivering reliable service on time and in budget to customers while maintaining a good working partnership with them is another significant challenge for them. They need to consider consumers and their needs to resolve these challenges. Among these, one of their primary emphases will be on client churn. Customer churn takes place when clients or subscribers cease to engage incorporation with a company or service. For any organization, winning business from new clients means going via the sales pipeline, using their sales and marketing assets in the cycle

2. OBJECTIVES

Churn prediction using machine learning in the banking sector aims to enhance customer retention by identifying clients who are likely to leave the bank or reduce their engagement with its services. This proactive approach allows banks to engage at-risk customers with personalized offers, services, or incentives to maintain their business. By leveraging machine learning, banks can implement targeted marketing campaigns and loyalty programs tailored to the specific needs and preferences of these customers. Additionally, churn prediction helps banks optimize their resources by focusing retention efforts on the most at-risk segments, ultimately reducing customer acquisition costs and increasing overall profitability. Furthermore, understanding the factors contributing to customer churn enables banks to improve their services, address pain points, and enhance customer satisfaction, thereby fostering long-term loyalty and a stronger customer relationship.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1683-1687

3. METHODOLOGY

Churn prediction in the banking sector using machine learning involves a comprehensive approach starting with data collection, where customer demographic, behavioral, transactional, and external data are gathered. This data undergoes preprocessing, including cleaning, feature engineering, normalization, and encoding to ensure it is suitable for model training. Supervised learning models such as Logistic Regression, Decision Trees, Random Forests, Gradient Boosting Machines (GBM), and Support Vector Machines (SVM) are commonly employed, alongside neural networks accuracy. The model training phase involves splitting the dataset, using cross-validation, and evaluating performance with metrics like accuracy, precision, recall, F1-score, and ROC-AUC. Hyperparameter tuning optimizes model parameters. Post-training, feature importance and sensitivity analysis help interpret the model. Deployment involves integrating the model into production environments for real-time predictions and linking with CRM systems for proactive customer engagement

4. LITERATURE SURVEY

SVM-Based Churn Prediction in a Chinese Commercial Bank

A study focused on churn prediction using Support Vector Machine (SVM) models examined a dataset from a Chinese commercial bank, consisting of 50,000 customer records, which were pre-processed to yield 46,406 valid records. The study compared two types of SVM models: Linear SVM and SVM with Radial Basis Function (RBF) kernel. Due to the imbalanced nature of the dataset, with a significant disparity between churners and non-churners, the predictive accuracy was initially low. B. He, Y. Shi, Q. Wan, and X. Zhao in 2014

Neural Network-Based Churn Prediction in a Croatian Bank

Another research investigated the use of data mining techniques for churn prediction in a small Croatian bank, employing a dataset of 1,866 customer records. This study utilized a Neural Network within the Alyuda Neuro Intelligence software, dividing data into training, validation, and testing sets. It identified and categorized features into necessary, redundant, and target characteristics. A. Bilal Zeri in 2016

ML-Based Churn Prediction Using Syriatel Telecom Data

A different study proposed in 2017 a machine learning model for churn prediction using the 'churn modeling data' from Syriatel Telecom. This study implemented four machine learning methodologies: Decision Tree, Random Forest, Gradient Boosted Machine (GBM), and Extreme Gradient Boosting.

5. PROPOSED SYSTEM

Studies revealed that gaining new customers is 5 to 10 times costlier than keeping existing customers happy and loyal in today's competitive conditions, and that an average company loses 10 to 30 percent of customers annually. Many companies, being aware of this fact, are engaged in satisfying and retaining the customers. Especially in the subscription-oriented industries, such as telecommunications, banking, insurance, and in the fields of customer relationship management, etc, companies working with numerous customers, the revenues of the companies are provided by the payments made by these customers periodically. It is very important to be able to keep customers satisfied in order to be able to sustain this revenue with the least expenditure cost.

6. HARDWARE AND SOFTWARE REQUIREMENTS

6.1 HARDWARE REQUIREMENTS:

- System: Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz
- Hard Disk : 1 TB.
- Input Devices : Keyboard, Mouse
- Ram : 4 GB.

6.2 SOFTWARE REQUIREMENTS:

- Operating system : Windows XP/7/10.
- Coding Language : python
- Tool : VS Studio

7. PACKAGES USED

from flask import Flask, render template, request

import pickle

import numPy as np

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1683-1687

import sklearn

import matplotlib from sklearn.preprocessing

import StandardScaler

8. TECHNOLOGY DESCRIPTION

churn prediction system in Python involves several key steps, starting with data collection and preprocessing. Initially, the dataset is loaded using Pandas and preprocessed to handle missing values and encode categorical features. The data is then split into features (X) and the target variable (y), followed by a train-test split to create training and testing datasets. To address class imbalance, techniques such as resampling are applied to balance the number of churn and non-churn instances. The next step involves training various machine learning models, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting Machine (GBM), and Boost, using the Scikit-Learn and Boost libraries. Each model is trained on the balanced training dataset. Once the models are trained, their performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to determine their predictive power. This comprehensive approach ensures that the churn prediction system can effectively identify customers at risk of leaving, enabling proactive retention strategies.

9. SOURCE CODE

```
from flask import Flask, render_template, request
import pickle
import numpy as np
import sklearn
import matplotlib
from sklearn.preprocessing import StandardScaler
app = Flask(__name__)
model = pickle.load(open('Customer_Churn_Prediction.pkl', 'rb'))
@app.route('/', methods=['GET'])
def Home():
return render_template('index.html')
standard_to = StandardScaler()
@app.route('/predict',methods=['POST'])
def predict():
if request.method == 'POST':
CreditScore = int(request.form['CreditScore'])
Age = int(request.form['Age'])
Tenure = int(request.form['Tenure'])
Balance = float(request.form['Balance'])
NumOfProducts = int(request.form['NumOfProducts'])
HasCrCard = int(request.form['HasCrCard'])
IsActiveMember = int(request.form['IsActiveMember'])
EstimatedSalary = float(request.form['EstimatedSalary'])
Geography_Germany = request.form['Geography_Germany']
if(Geography_Germany == 'Germany'):
Geography_Germany = 1
Geography_Spain= 0
Geography_France = 0
elif(Geography_Germany == 'Spain'):
Geography_Germany = 0
Geography_Spain= 1
Geography_France = 0
else:
```

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1683-1687

```
Geography Germany = 0
Geography_Spain= 0
Geography_France = 1
Gender_Male = request.form['Gender_Male']
if(Gender_Male == 'Male'):
Gender_Male = 1
Gender_Female = 0
else:
Gender_Male = 0
Gender_Female = 1
prediction = model.predict([[CreditScore,Age,Tenure,Balance,NumOfProducts,HasCrCard,IsActiveMember,EstimatedSalary,Geography_Germany,Geography_Spain,Gender_Male]])
if prediction==1:
return render_template('index.html',prediction_text="The Customer will leave the bank")
else:
return render_template('index.html',prediction_text="The Customer will not leave the bank")
if __name__=="__main__":
app.run(debug=True)
```

10. RESULTS

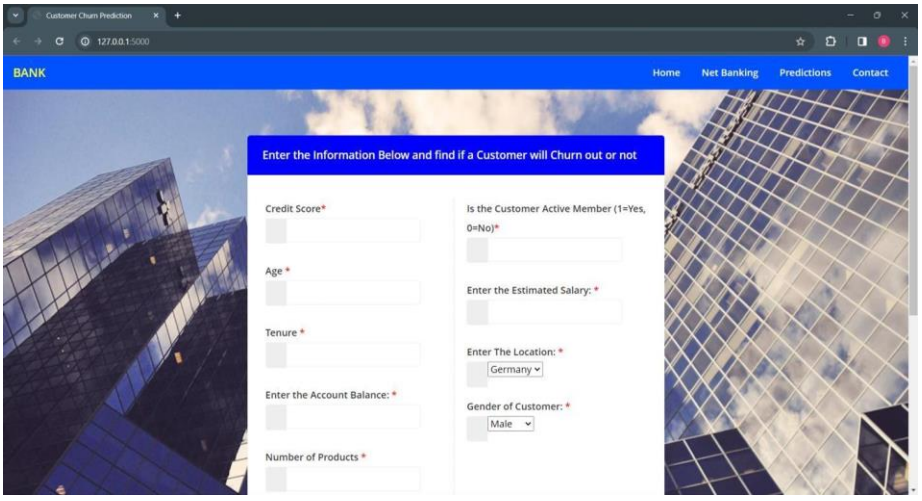

Fig No 10.1 Redirected web page

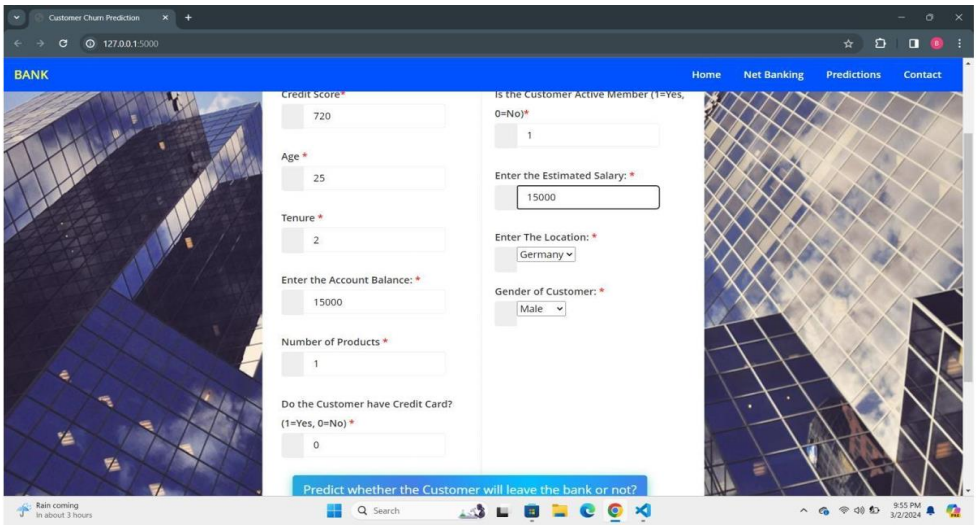

Fig No 10.2 providing customer details as input

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1683-1687

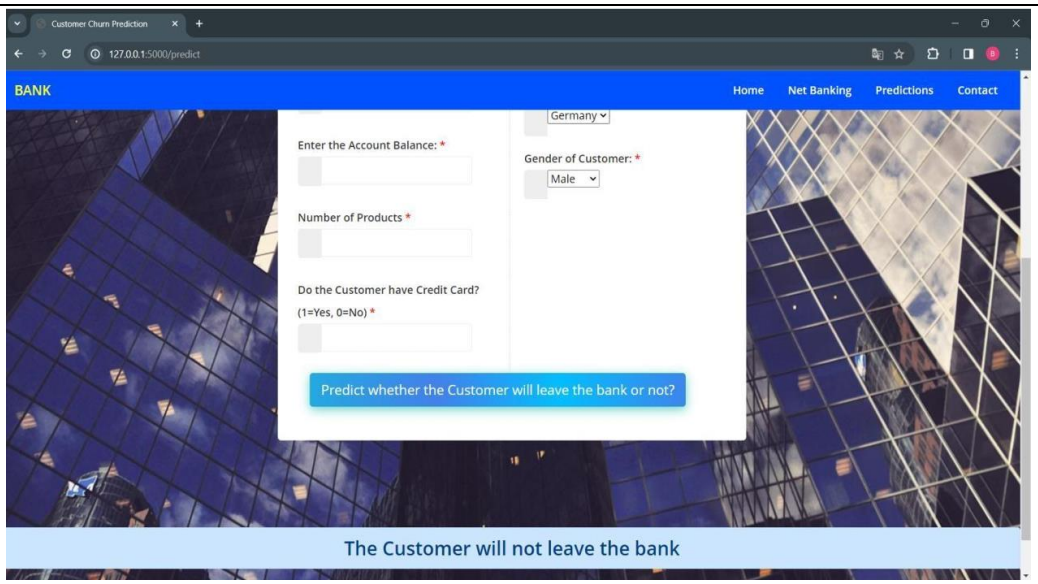

Fig No 10.3 Output (Weather the customer will churn or not)

11. CONCLUSION

This study helps to predict churn among bank customers with relative success. However, there is scope for improvement in the future. Due to the sensitive nature of banking data, access to large datasets is restricted. Access to more data points would enhance the generalizability of predictions. Additionally, having access to more granular data would contribute to improved forecasts. The current attributes are more specific to a customer's profile than their recency (the metrics that record behavior immediately before churning). Prospective researchers can derive these metrics, which will help track a shift in customer behavior just before they churn, and thus better identify churn patterns. There is also an opportunity for prospective researchers to improve our app by automating the model-training process, incorporating new features and data points, and generate updated models. This would help build a feedback loop in the models, ensuring greater veracity of model prediction by changing patterns and increasing the dataset. In addition, they can go further by incorporating additional prediction algorithms that can be integrated into the visualization app for comparative analysis and better churn management.

12. FUTURE SCOPE

As machine learning algorithms continue to evolve, more advanced models may be developed to improve the accuracy of churn prediction. Deep learning techniques, ensemble models, and other sophisticated approaches could be explored to enhance the predictive capabilities. The banking sector generates massive amounts of data daily. Integrating big data analytics with machine learning can provide a more comprehensive understanding of customer behaviour. This may involve incorporating non-traditional data sources such as social media activity, transaction history, and customer service interactions. Moving towards real time prediction of customer churn enables banks to take immediate corrective actions. By leveraging real-time data streams and advanced analytics, banks can identify potential churn indicators promptly and implement targeted interventions to retain customers. Machine learning can enable banks to personalize their interactions with customers based on their behaviour and preferences.

13. REFERENCES

[1] B. He, Y. Shi, Q. Wan, and X. Zhao, "Prediction of Customer Attrition of Commercial Banks based on SVM Model," Procedia Compute. Sci., vol. 31, 2014, Accessed: Sep. 11, 2022.

[2] Churn Analysis of a Telecom Company," Analytics Vidhya, Jan. 27, 2022.

[3] Celia and U. Osmanoğlu, "Comparing to Techniques Used in Customer Churn Analysis," Jul. 2019.

[4] A. K. Ahmad, A. Jafar, and K. Algoma, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 1, p. 28, 2019

[5] T. J. Swamidason, "Survey of data mining algorithm's for intelligent computing system," Journal of Trends in Computer Science and Smart Technology, vol. 01, pp. 14–23, 09 2019. 7. A. Bilal Zoric,

[6] Predicting customer churn in banking industry using ´ neural networks," Interdisciplinary Description of Complex Systems: INDECS, vol. 14, no. 2, pp. 116–124, 2016.