International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 08 | Aug 2026 www.irjet.net p-ISSN: 2395-0072

CUSTOMER CHURN PREDICTION SYSTEM USING GRADIENT BOOSTING CLASSIFICATION

V.PRAVEEN KUMAR[1], DR. G SRIRAM [2], DR. KONDAPALLI VENKATA RAMANA[3]

[1]*Department of Information Technology & Computer Applications, Andhra University, Visakhapatnam, Andhra Pradesh, India-530003*

[2]*Assistant Professor, Department of Computer Science & Systems Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India-530003*

[3]*Associate Professor, Department of Computer Science & Systems Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India-530003*

---***---

Abstract - *Customer churn prediction is a problem for Banks since it provides information about the customers who are likely to leave in the near future and give them enough time to convince them to stay with the firm. The current work proposes a system for customer prediction using Gradient Boosting Classification, which makes use of the socio- relational attributes of the customer to predict whether the customer will stay or not. Feature engineering and training were carried out to ensure that the system has accuracy. A web application was also developed so that the users could interact with the system and make predictions, Probabilities of customers going away. Visualizations WERE used to inform decision-making. The system was. Found to have an accuracy of 87.00% with precision of 79.28%, 48.89% recall and 60.49% F1 score which implies that the system can be used to reduce customer attrition.*

Keywords: Customer Churn Prediction, Gradient Boosting Classifier, XGBoost, Feature Engineering, Streamlit, Confusion Matrix, Predictive Analytics.

1. INTRODUCTION

Customer churn prediction is a problem of interest to organizations since it provides the information about the customers who are likely to switch to other entities in the near future. This information is useful to the companies because they can take measures to retain the customers by convincing them to stay with them instead of taking their services or products to other firms. The main idea behind customer churn prediction revolves around reducing the attrition rate of the existing customers since it's cheaper to retain than acquire new customers.

The current research presents a system based on Gradient Boosting Classification for predicting customer attrition. Using the attributes such as a credit score, demographics, location, gender, tenure and account balance among others to predict customer attrition. A web application was also developed to facilitate the prediction process of analyzing customer data and predicting their attrition.

2. LITERATURE REVIEW

Many researchers have used machine learning algorithms in the prediction of customer churn in the years. Some of them have reported that the logistic regression analysis, decision trees and random forest classification have performance when analyzing customers' data and predicting their attrition. In fact some works have demonstrated that combining algorithms can give better results because different algorithms have diverse patterns and structures which help in capturing complex relationships within the data.

Researchers have also demonstrated that Gradient Boosting Classification technique is better than others in modeling and predicting interactions.

In the work Gradient Boosting Classification is used to predict customer attrition. A system which facilitates analyzing and predicting customer attrition based on their characteristics is introduced. The system was built by carrying out data preprocessing followed by a training process. Finally a web application was developed for facilitating the prediction process by the end-users.

3. PROPOSED METHODOLOGY

The current work presents a system based on Gradient Boosting Classification for predicting customer attrition. The proposed system consists of the following steps: obtaining and preparing data, processing data training the model and prediction using a web-based interface.

3.1 Dataset

A sample customer dataset was obtained which contains socio- demography characteristics that help in predicting customer attrition. The data had ten thousand records with attributes such as: credit score, gender, age, tenure, balance, annual salary and many others.

The data was used in training the model to predict customer attrition.

3.2 Data Preprocessing

The data was preprocessed extensively so that it is in a format for training the model. In the step irrelevant information was deleted and categorical data was converted into numerical data. Secondly data was normalized to ensure that it has a range so that it is easy to process. Then the data was split into testing and training sets to train and evaluate the performance of the model.

3.3 Model Training

We used a Gradient Boosting Classification model to figure out if a customer is going to stop doing business with us. This model was trained using the customer churn data that we cleaned up earlier. We split this data into two parts: one for training and one for testing. The training part was bigger with 80 percent of the data and the testing part was smaller with 20 percent. We made sure that both parts had the number of customers who stopped doing business with us and those who did not.

We made sure the input information was standardized so the model could learn from it easily. We chose Gradient Boosting Classifier for training the model. The model was trained using the training data and a special code called random_state set to 42. This way the model gives us the results every time we use it. The model learned how the different pieces of information are related to each other and to whether a customer wants to stop doing business with us.

Then we checked how well the model works by looking at how accurate it's how precise it is how well it remembers things and its F1-score. After that we saved the model as a file called a pickle file and used it in our web application to make predictions, about whether a customer will stop doing business with us.

Here are the important things we used for training the model:

* Algorithm Used: Gradient Boosting Classifier
* Training Data Split: 80:20
* Feature Scaling: StandardScaler
* Model File: model.pkl
* Target Variable: Exited (1 means the customer stopped doing business with us and 0 means they did not)

3.4 Prediction and Web Application

In order to find out what customers might stop doing business with us a website was made that can predict customer loss. This website lets users enter details about a customer that could help decide if the customer will stop using the bank services. These details include credit score, location, age, account balance and how long the customer has been with the company.

The information is then put into a machine learning model called the Gradient Boosting Classification Algorithm to get a prediction about whether the customer will stop doing business with the company. Also the algorithm can quickly predict customer loss. Shows how accurate it is, in making that prediction.

The website also lets the user see graphs or charts that show what the algorithm has predicted. This helps in understanding the data and allows the company to come up with solutions to stop customers from leaving.

Technologies used:

Front End: HTML, CSS, Streamlit
Back End: Python
Model: Gradient Boosting Classifier
Output: Attrition Prediction, Accuracy Score, Visualization Plots

4. RESULTS AND DISCUSSION

The current work focuses on presenting the design and implementation of a system based on Gradient Boosting Classification for predicting customer attrition. The system was implemented using Python programming language and Streamlit web framework. The performance of the model was evaluated using classification performance metrics. The results showed that the model was able to capture customer attrition with an accuracy of 87%. This result demonstrates that the system can be used to predict the attrition of customers and help in minimizing the attrition rate in firms.

4.1 Performance Evaluation & Model Comparison

In an effort to test the effectiveness of the proposed Gradient Boosting Classifier, three other standard machine learning classification models, namely Logistic Regression, Random Forest and XGBoost, were trained and tested on the same dataset and same 80:20 train-test split. We used standard classification metrics like Accuracy, Precision, Recall, and F1-Score to evaluate the performance.

Table -1: Comparative Performance Metrics of Classification Models

Model	Accuracy	Precision
Logistic Regression	80.95%	0.6032
Random Forest	86.45%	0.7833
Gradient Boosting	**87.00%**	**0.7928**
XGBoost	86.65%	0.7869

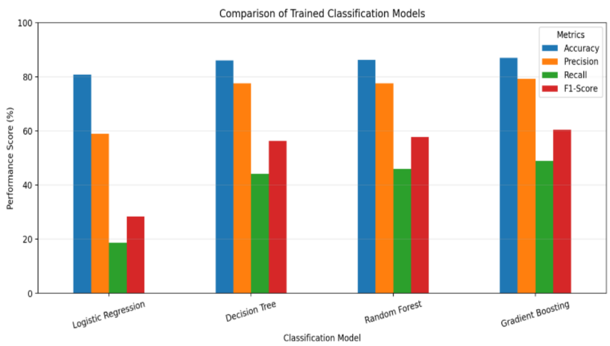

Fig-1: Comparison of All Four Models

Comparative analysis and model selection

The baseline performance of Logistic Regression was the worst among all the metrics especially in the Recall (18.67%) and F1-Score (28.50%) which indicates that the linear boundaries are not effective in capturing the complex non-linear relationships in the customer churn data.

Tree-Based Ensembles: Ensemble tree models (Random Forest, XGBoost and Gradient Boosting) showed significantly increased predictive power. Random Forest accuracy is 86.45, XGBoost accuracy is 86.65.

Proposed Model Of all the algorithms tested, the Gradient Boosting Classifier provided the best overall performance:

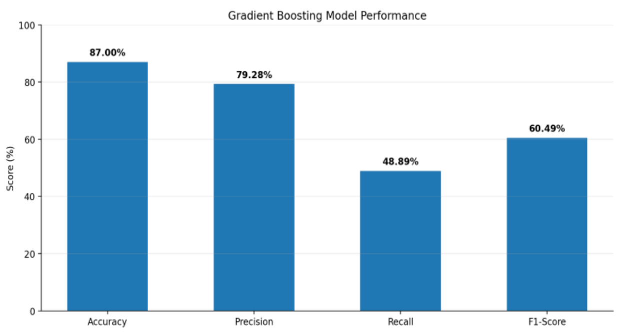

Fig-2: Gradient Boosting Classifier Model Performance Visualization

Accuracy (87.00%): It correctly classified 87% of the customer records in the held out test set.

Precision (79.28%): Nearly 80% of the customers labeled as "at-risk" are actual churners, lowering false alarms for retention teams.

Recall (48.89%) & F1-Score (60.49%): Provided the best balance between precision and recall among the benchmarked models.

4.2 Confusion Matrix

The confusion matrix gives us an idea of how well our classification is working by comparing what really happened with what we thought would happen. It shows us the number of times we got it right and the number of times we got it wrong.

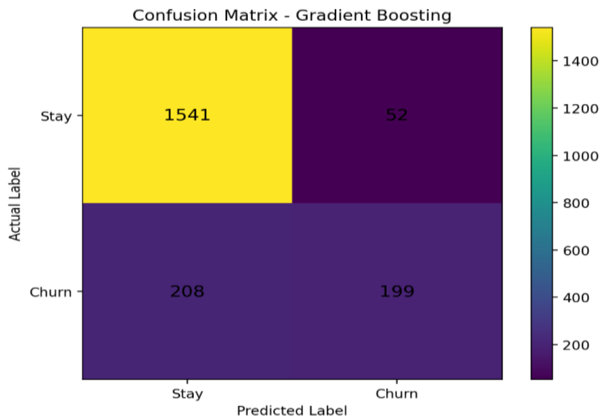

Fig -3: Confusion Matrix of the Proposed Model

The confusion matrix shows that the proposed model gets most of the customer records right which means our predictions are more accurate and we make mistakes with the proposed model. The proposed model does a job of classifying customer records correctly and that is why we see better results, with the proposed model.

4.3 Streamlit Application Results

The customer churn prediction system we made is now up and running as a web application using Streamlit. This application is easy to use. Lets users put in customer information to get predictions about whether they will leave or not in real time. When we get the prediction result the system also shows us the chance of the customer leaving and some interactive pictures, like charts that show how likely it is that the customer will leave and what the outcome of the prediction is. These pictures help us understand the results better and make it easier to decide what to do to keep our customers.

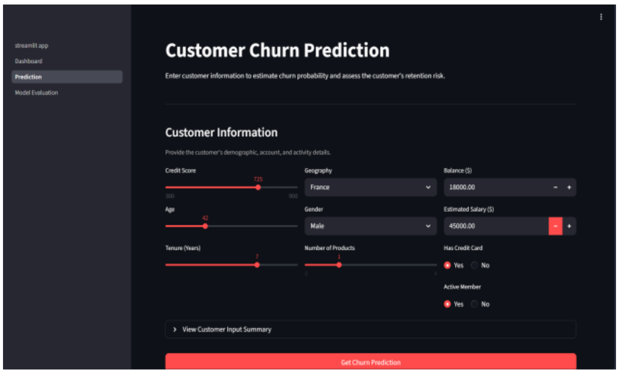

Fig -4: Streamlit Prediction Interface

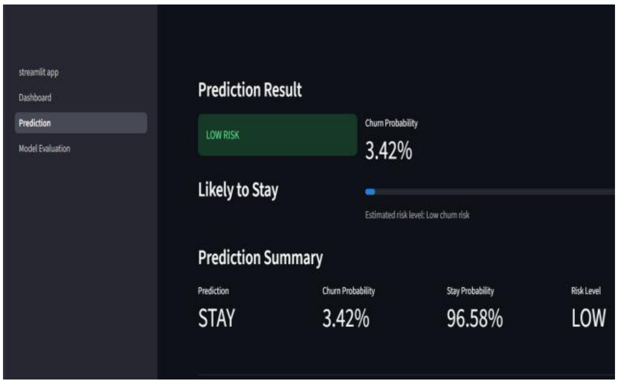

Fig -5: Low Risk Prediction Result Showing.

Figure 5 shows a customer who's not likely to leave which means the customer churn prediction system thinks they will probably stay with us. The dashboard shows the chance of the customer leaving how risk there is and some pictures like bar and pie charts that help us see that the customer is likely to stay with our company.

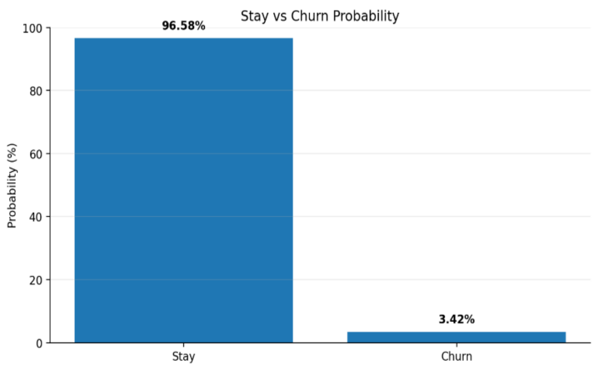

Fig -5(a): Bar graph of Low Risk Prediction Result Showing.

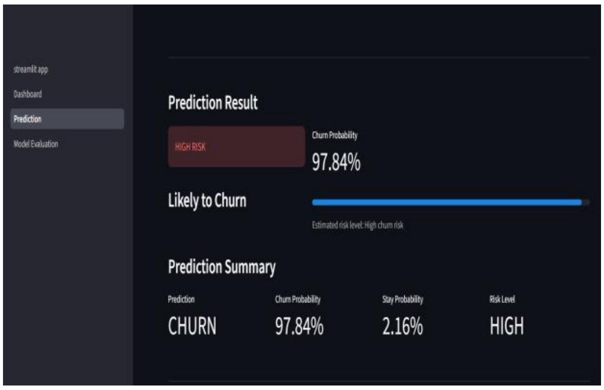

Fig -6: High Risk Prediction Result Showing

Figure 6 shows a customer who's likely to leave which means the customer churn prediction system thinks there is a high chance they will go. We use this information to try to keep these customers from leaving. The charts and pictures help the people making decisions understand what the prediction means and what they can do about it. The customer churn prediction system is really helpful, for understanding customer churn and the customer churn prediction system helps us to keep our customers.

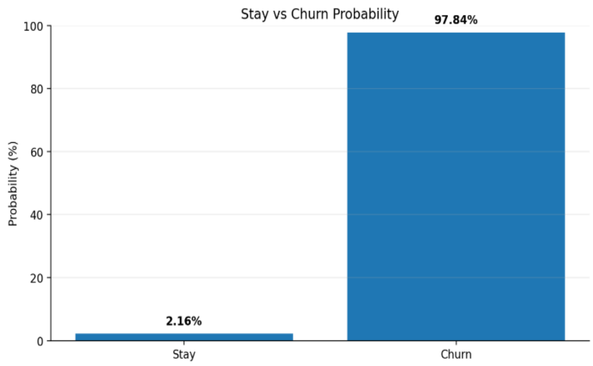

Fig -6(a): Bar graph of High Risk Prediction Result Showing

5. CONCLUSION

This paper introduced a Customer Churn Prediction System that uses the Gradient Boosting Classification algorithm. The system uses data preprocessing methods, such as feature selection, encoding and feature scaling to make predictions more accurate. The model that was trained was added to a web application built with Streamlit. This allows people to get predictions, about customer churn in time through a user-friendly interface. Testing showed that the model worked well and was able to find customers who might stop using the service. The system that was created can help companies make choices and take steps to keep their customers.

6. FUTURE SCOPE

The system they are talking about can be made better by using advanced machine learning and deep learning algorithms. This will help it make accurate predictions.We can do work on this system in the future. For example we can make it work with real-time data and train the models automatically.We can also put the system on cloud platforms so it can be used for applications.The system can also use information, about what the customers do and their past transactions. This will help it predict when a customer is going to leave accurately.The system can also be used to give customers personalized recommendations and help the business keep its customers. This will help the business make decisions.

ACKNOWLEDGEMENT

The author feels very thankful to Dr. Kondapalli Venkata Ramana Associate Professor Department of Computer Science & Systems Engineering Project Guide Department of Information Technology & Computer Applications Andhra University for advice and ongoing support and for giving useful suggestions during the making of this research work. The author also says thanks to the teachers, in the Department for giving the needed help and resources. The author also wants to express appreciation to family and friends for their never-ending support and encouragement that helped a lot in finishing this work successfully.

REFERENCES

1) J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001. [Online]. Available: https://doi.org/10.1214/aos/1013203451

2) L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

3) F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://jmlr.org/papers/v12/pedregosa11a.htmlF. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

4) A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022. [Online]. Available: https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/S. Raschka and V. Mirjalili, Python Machine Learning, 3rd ed. Birmingham, UK: Packt Publishing, 2019.

5)] S. Raschka and V. Mirjalili, Python Machine Learning, 3rd ed. Birmingham, UK: Packt Publishing, 2019. [Online]. Available: https://www.packtpub.com/product/python-machine-learning-third-edition/9781789955750

6) Kaggle, "Churn Modelling Dataset." [Online]. Available: https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling

7) T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009. [Online]. Available: https://doi.org/10.1007/978-0-387-84858-7

8) G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning: With Applications in R. New York, NY, USA: Springer, 2013. [Online]. Available: https://www.statlearning.com/

9) C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006. [Online]. Available: https://link.springer.com/book/9780387310732

10) B. Huang, M. T. Kechadi, and B. Buckley, “Customer churn prediction in telecommunications,” Expert Systems with Applications, vol. 39, no. 1, pp. 1414–1425, 2012. [Online]. Available: https://doi.org/10.1016/j.eswa.2011.08.024

11) A. K. Ahmad, A. Jafar, and K. Aljoumaa, “Customer churn prediction in telecom using machine learning and social network analysis in big data platform,” Journal of Big Data, vol. 6, Art. no. 28, 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0191-6

12) M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen, “Social network analytics for churn prediction in telco: Model building, evaluation and network architecture,” Expert Systems with Applications, vol. 85, pp. 204–220, 2017. [Online]. Available: https://doi.org/10.1016/j.eswa.2017.05.028

13) Ö. G. Ali and U. Arıtürk, “Dynamic churn prediction framework with more effective use of rare event data: The case of private banking,” Expert Systems with Applications, vol. 41, no. 17, pp. 7889–7903, 2014. [Online]. Available: https://doi.org/10.1016/j.eswa.2014.06.034

14) K. A. Amuda and A. B. Adeyemo, “Customers churn prediction in financial institution using artificial neural network,” 2019. [Online]. Available: https://arxiv.org/abs/1912.11346

15) J. Maan and H. Maan, “Customer churn prediction model using explainable machine learning,” 2023. [Online]. Available: https://arxiv.org/abs/2303.00960

16) R. Genuer, J.-M. Poggi, and C. Tuleau, "Random forests: Some methodological insights," 2008. [Online]. Available: https://arxiv.org/abs/0811.3619

17) Python Software Foundation, "Python 3 Documentation," [Online]. Available: https://docs.python.org/3/

18) Kaggle, "Churn Modelling Dataset," [Online]. Available: https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling

19) Scikit-learn Developers, "Gradient Boosting Classifier," Scikit-learn Documentation, [Online]. Available: https://scikit-learn.org/stable/modules/ensemble.html

20) Scikit-learn Developers, "StandardScaler," Scikit-learn Documentation, [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

21) Streamlit Inc., "Streamlit Documentation," [Online]. Available: https://docs.streamlit.io/