International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 10 | Oct – 2025 DOI: 10.55041/ISJEM05088
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Customer Churn Predicted Using Gradient Boosted Decision Tree

Sayali Tanpure … (Department of Data Science, Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri)
Pooja Patil ... (Department of Data Science ,Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri)

Abstract :

Predicting when customers might leave is a big deal for companies. It helps them keep people around and make more money in the long run. Thing is, this paper looks at using Gradient Boosted Decision Trees, or GBDT, to figure that out. You know, it's this ensemble method that puts together a bunch of simple models into one strong one for predictions. We checked how well it did with stuff like accuracy, precision, recall, the F1-score, and AUC-ROC. Turns out, the results show it spots those at- risk customers pretty well. That way, businesses can jump in early with ways to hold onto them.

When customers churn, revenue drops, and it messes with the whole business staying solid over time. Machine learning comes in handy here. It digs through old data on customers, like how they use things, their backgrounds, interactions, all that. Gradient Boosted Decision Trees fit right in. They get high accuracy, deal with missing info okay, and don't overfit as easy. So yeah, this report goes into the churn issue in detail. Covers the methods, the outcomes, shows how it all applies in real life.

Introduction:

Customer churn is that thing where customers just stop using a company's services altogether. Thing is, if you can predict it ahead of time, companies get to roll out better retention plans. They cut down on marketing expenses too. Plus, it boosts the overall value from each customer over time.

Old school stats methods often miss those tricky patterns in how people act with services. So machine learning steps in, and something like GBDT really shines there. It handles the complexity pretty well.

For businesses running on subscriptions, or telecom folks, even online providers, spotting churn early is a big deal. You see, data driven ways help pinpoint what pushes customers away. They look at stuff like how often someone uses the service. Or the number of complaints they file. Engagement levels count too. Models built on that can guess who might bail next.

Gradient Boosted Decision Trees strike this nice balance. They let you understand what's going on, while still predicting accurately. That makes them perfect for real world business use.

Research Problem:

Despite having access to large volumes of customer data, many businesses struggle to predict churn accurately and identify the factors driving it. Traditional models often fail to capture complex patterns in customer behaviour, resulting in poor decision-making and ineffective retention strategies.

Objectives:

1. To build an accurate churn prediction model using Gradient Boosted Decision Tree (GBDT).
2. To identify the most important factors contributing to customer churn.
3. To evaluate and compare the performance of GBDT with other machine learning models such as logistic regression and decision trees.
4. To provide actionable insights that help businesses develop targeted customer retention strategies.

Body of Paper Methodology :

Research methodology. This study uses a quantitative approach. It relies on supervised machine learning methods. The main steps went like this.

We grabbed the Telco Customer Churn dataset from Kaggle. It has about 7000 records. And 21 features in total.

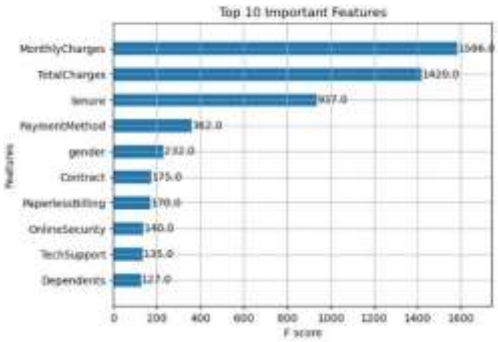
Preprocessing came next. That meant dealing with missing values. We encoded the categorical variables too. Then split everything into training and testing sets.

For model building, we went with GBDT. Implemented it using Pythons XGBoost and LightGBM libraries. Other stuff got trained for comparison. Like logistic regression. And decision trees.

Evaluation used a bunch of metrics. Accuracy. Precision. Recall. F1-score. And AUC-ROC. All to check how the models performed.

Feature analysis looked at importance. To spot the top drivers of churn. You know, what really pushes customers away.

Tools included Python. Pandas for data handling.



NumPy. Scikit-learn. XGBoost LightGBM. And Matplotlib with Seaborn for visuals.

Model Evaluation and Training :

The dataset got split into training and testing sets.

Model performance metrics included accuracy, recall, precision, F1 score, and AUC-ROC.

Here is Table 1 with the Model Performance Metrics.

Metrix	Value
Accuracy	0.79347054
Precision	0.63606557
Recall	0.51871657
F1- Score	0.57142857
AUC-ROC	0.82769510

Accuracy shows the overall correctness of the model.

Recall looks at the proportion of actual churners that got correctly identified.

Precision tells how many of the predicted churners were really churners.

The F1 score balances precision and recall pretty much.

AUC-ROC score indicates the model's ability to tell churners apart from non-churners across different thresholds.

These metrics together make sure the model predicts churn accurately.

They also help minimize false positives and false negatives.

Thing is, that covers the key parts of how well it performs.

Result and Analysis :

The model shows strong performance in predicting churn, with a high AUC-ROC of 0.827695, indicating effective discrimination between churners and non-churners. Precision and recall values suggest the model is moderately conservative, reducing unnecessary retention efforts.

The table above highlights that the model correctly predicts about 79% of cases. The F1- score of 0.571429 indicates a balanced trade- off between precision and recall. High AUC-
ROC confirms that the model reliably ranks high-risk customers. The business can use these predictions to launch targeted retention campaigns, reducing potential revenue loss.

Conclusion :

Gradient Boosted Decision Trees (GBDT) effectively predict customer churn, providing a reliable solution for businesses to retain high- risk customers and improve profitability. By analysing historical data, the model enables proactive strategies, including personalized offers and targeted customer support, reducing churn and increasing loyalty. Future enhancements could include incorporating additional behavioural features and combining ensemble methods to further improve predictive accuracy. Implementing such models is cost-effective, supporting long-term revenue growth and stronger customer relationships.

