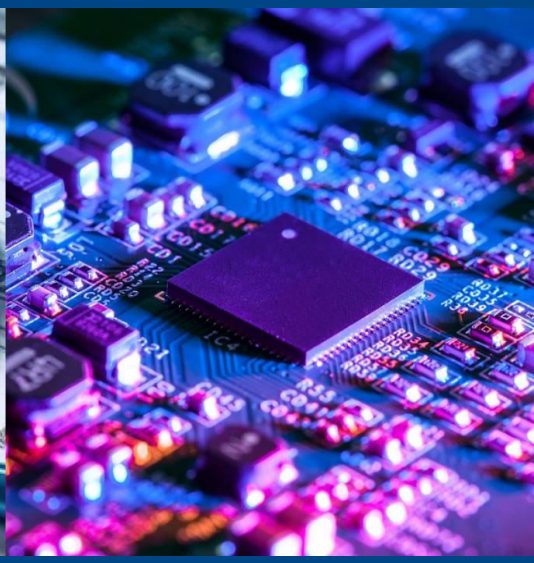
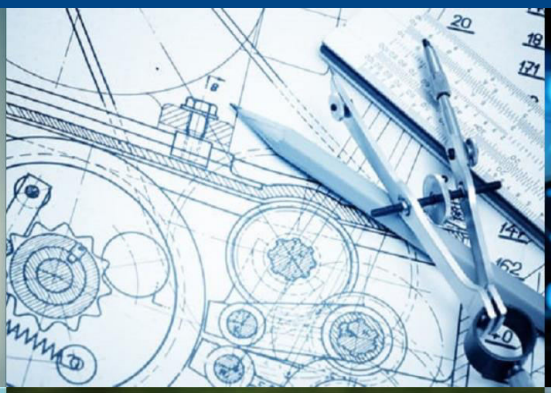
ISSN: 2582-7219

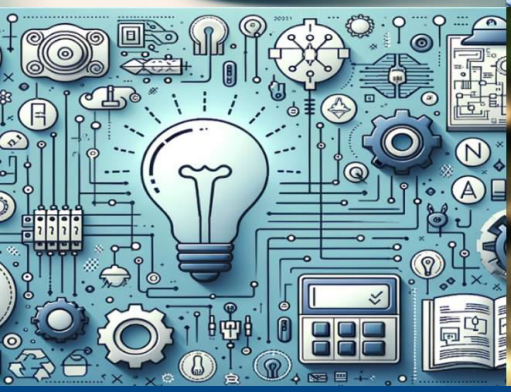




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 7.521

Volume 8, Issue 1, January 2025

© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Data-Driven Analysis and Forecasting of Customer Behavior

G.Renuka[1], M. Swathi[2], V.Gurumurthy[3]

Asst. Professor, Dept. of ECE, Anurag University, Hyderabad, Telangana, India[1, 2, 3]

ABSTRACT: This paper evaluates the use of predictive analytics for predicting client turnover in subscription-based services, with the objective of developing a predictive model to assist small and medium-sized enterprises in managing customer attrition in the face of digital disruption. Focuses on empirical consumer data to properly forecast purchasing trends and tailor marketing techniques. Four classifiers—Random Forest, Logistic Regression, Gradient Boosting, and XGBoost—are trained and evaluated using several performance indicators as part of a systematic approach that involves Kaggle data collection, preparation, and model selection. The results demonstrate excellent accuracy in estimating client attrition, the paper highlight the proactive use of predictive analytics to identify at-risk clients and implement targeted retention strategies.

I. INTRODUCTION

In today's competitive business landscape, subscription-based services face significant challenges in retaining customers. Client attrition, or customer churn, has become a critical issue for businesses that rely on recurring revenue models, such as SaaS providers, media streaming services, and e-commerce platforms. Understanding why customers leave and predicting future churn is essential for devising effective retention strategies. One of the most powerful ways to address this challenge is through the use of predictive analytics, a tool that allows companies to anticipate customer behavior based on historical data and trends.

Predictive analytics involves the use of statistical techniques, machine learning algorithms, and data mining to identify patterns in data and predict future outcomes. In the context of customer churn, predictive analytics uses past customer behavior, transaction data, and demographic information to build models that can forecast the likelihood of a customer leaving. These models help businesses identify high-risk customers early, allowing them to take proactive steps to retain them.

The goal of this paper is to develop a predictive model to assist small and medium-sized enterprises (SMEs) in managing customer attrition. SMEs often lack the resources to invest in advanced analytics tools and may struggle to compete with larger organizations that have more robust data infrastructure. By using open-source tools and accessible data sources, this aims to demonstrate how predictive analytics can be applied even in resource-constrained environments, providing actionable insights that can improve customer retention.Four popular machine learning classifiers—Random Forest, Logistic Regression, Gradient Boosting, and XGBoost—are employed in this research to predict client turnover.

These algorithms were selected based on their proven effectiveness in classification tasks and their ability to handle complex, non-linear relationships in data. Random Forest is known for its robustness and ability to handle missing values, while XGBoost is widely recognized for its superior performance in structured data prediction. The research process begins with the collection of customer data from a publicly available Kaggle dataset. This data includes transactional information, demographic variables, and customer interaction details. After data preparation and cleaning, the selected classifiers are trained and evaluated using various performance metrics, including accuracy, precision, recall, and F1-score. These metrics help assess the model's ability to correctly identify customers who are likely to churn and differentiate them from those who are likely to stay.

The findings of this paper highlight the significant potential of predictive analytics for customer churn prediction. The models developed show promising accuracy in predicting client turnover. The results underscore the need for continuous refinement of predictive models and the integration of additional data sources, such as customer feedback and social media interactions, to enhance prediction accuracy. By doing so, businesses can improve their retention strategies and achieve long-term success in a digitally disrupted market.

© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fig1: Data Analysis by EDA

II. LITERATURE SURVEY

Previous studies have highlighted the significant role of big data analytics in understanding consumer behavior. One such study by Wilson and Davis (2022), titled "A Study on Consumer Behavior Using Big Data Analytics", explored the use of big data to analyze consumer behavior patterns. [1] The paper discusses various types of analytics, such as predictive and prescriptive analytics, employed to examine large data sets, and reviews several major models used for this purpose.

This study by Faiza and Taher presents a comprehensive analysis of e-commerce sales data using K-means clustering and the Silhouette method to optimize cluster configurations. [2] The paper highlights the significance of customer segmentation through the RFM model, combined with clustering techniques, to identify distinct consumer groups and enable targeted marketing efforts.

In the quest to effectively manage customer relationships, businesses must evaluate the costs and benefits associated with various alternative expenditures and investments.[3] By analyzing these factors, businesses can optimize resource allocation for marketing and sales activities over time. Forecasting the future behavior of customers is crucial for businesses, as it helps in developing effective marketing strategies and making informed decisions. [4]As a result, data mining and prediction tools have become essential for firms aiming to predict customer behavior and create targeted marketing programs.

III. METHODOLOGY

3.1 EXISTING METHOD

In the existing method, businesses often rely on dashboards for customer behavior analysis. These dashboards offer a visual interface to display real-time data and metrics, allowing businesses to monitor key performance indicators (KPIs), customer interactions, sales data, and other relevant metrics. While dashboards are effective for tracking historical and current data, they have a significant limitation—they do not provide future predictions or insights into how customer behaviour might change over time.

The existing approach primarily focuses on aggregating and visualizing customer data without leveraging predictive models that can forecast future behaviour. It typically lacks the ability to anticipate which customers might be at risk of churning, which customers are most likely to make a purchase, or how customer preferences might evolve. This is a critical gap because businesses today need to make data-driven decisions not only based on what has happened in the past but also by predicting what is likely to happen in the future.

The absence of predictive capabilities also limits businesses' ability to optimize their resource allocation and marketing efforts. By relying solely on dashboards, businesses might miss opportunities to target high-value customers or address potential issues before they impact customer retention. In contrast, machine learning models designed for future predictions, such as churn algorithms or customer lifetime value (CLV) models, can proactively guide business decisions to improve customer experience, retention, and revenue growth.

© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.2 PROPOSED METHOD AND WORKFLOW OF VISUALIZATION

The proposed method aims to predict customer churn in subscription-based organizations using secondary data from Kaggle, involves several key steps. The methodology will utilize data mining, machine learning algorithms, and statistical analysis to uncover patterns that can predict customer attrition.

1. **Data Preprocessing:**

- **Data Cleaning:** The dataset will be cleaned to handle missing values, outliers, and duplicates. Any irrelevant or redundant features will be removed.
- **Feature Engineering:** Relevant features such as customer support interactions, viewing preferences, payment methods, and subscription types will be extracted or transformed into suitable formats for analysis. This may include normalizing or standardizing numerical values and encoding categorical data.
- **Exploratory Data Analysis (EDA):** A detailed exploratory analysis will be conducted to understand the distribution of data, identify patterns, and detect relationships between variables. Visualization techniques such as correlation heatmaps, bar plots, and box plots will be used to gain insights.

2. **Customer Segmentation:**

- **Clustering:** A clustering algorithm like K-means or DBSCAN will be used to segment customers based on their behaviors, preferences, and subscription history. This segmentation will help in identifying different groups of customers, some of which may be more prone to churn than others.
- **RFM Model (Recency, Frequency, Monetary):** An RFM model may be applied to assess customer activity and loyalty, which will be useful for understanding which segments are most at risk for churn.

3. **Predictive Modelling:**

- **Model Selection:** A variety of machine learning models, such as Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, and Support Vector Machines (SVM), will be considered for predicting customer churn. The selection of the final model will be based on model accuracy, precision, recall, and F1-score.
- **Training and Testing:** The dataset will be split into training and testing subsets (e.g., 80- 20 split). The models will be trained on the training data and evaluated on the test set. Cross-validation techniques will be applied to ensure robust performance.
- **Hyperparameter Tuning:** Hyperparameters of the chosen models will be tuned using techniques like Grid Search or Random Search to optimize performance.

4. **Evaluation Metrics:**

- **Churn Prediction Metrics:** The effectiveness of the models will be assessed using metrics like accuracy, precision, recall, F1-score, and AUC-ROC curve. These metrics will help in evaluating the model's ability to predict both churn and non-churn customers effectively.
- **Model Comparison:** Different models will be compared to identify the best performer. If necessary, an ensemble method like stacking or voting will be applied to combine the predictions from multiple models for improved accuracy.

5. **Implementation of Results:**

- **Churn Prediction Deployment:** The final churn prediction model will be deployed for use in real-time decision-making. Businesses can use it to identify at-risk customers and target them with retention strategies.
- **Insight Generation:** Insights gained from the model will be analyzed and communicated to businesses in a way that helps optimize marketing and product strategies. For example, businesses could implement personalized offers, better customer support, or enhanced product features to reduce churn.

6. **Validation and Monitoring:**

- **Model Monitoring:** Post-deployment, the model will be continuously monitored to track its performance over time. Adjustments and retraining may be required as new data is collected and customer behavior evolves.
- **Model Interpretation:** Techniques like SHAP values or feature importance can be applied to interpret the model's predictions and provide actionable insights for the business.

© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This methodology integrates data analysis and machine learning tools to predict customer churn and offers actionable insights for businesses to improve retention strategies. The use of real-world Kaggle data strengthens the model's relevance and applicability to real-world scenarios.

Additionally, the proposed methodology emphasizes the importance of data-driven decision- making in subscription-based businesses by leveraging predictive analytics to understand customer behavior. By segmenting customers and predicting churn, businesses can proactively design targeted marketing campaigns, personalized offers, and customer retention initiatives. Furthermore, continuous feedback loops and model updates ensure that the system adapts to changing customer trends and market conditions, making it a dynamic tool for long-term business success. This flow illustrates how raw data is transformed into meaningful insights through a structured data analysis pipeline. It begins with **Data Collection and Integration**, where data from multiple sources is gathered and combined. Next, **Data Pre-processing and Cleaning** ensures data quality by handling missing values, inconsistencies, and formatting issues.
Once cleaned, the data undergoes **Exploratory Data Analysis (EDA)** to identify patterns, trends, and correlations. The insights from EDA are then visualized and reported through **Automated Reporting and Visualization**, making data interpretation easier. Finally, these insights lead to Actionable Recommendations, enabling informed decision-making and strategic planning.

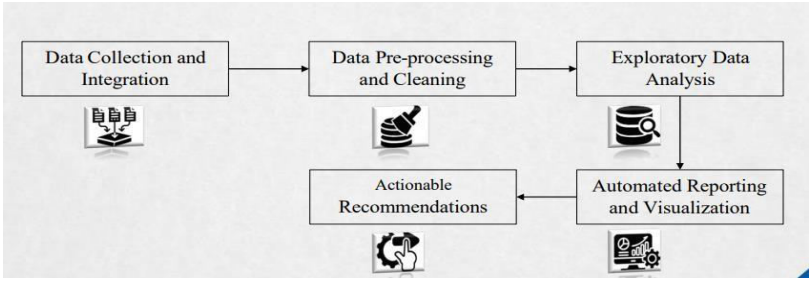


Fig: 2 Work Flow of Visualization

IV. TESTING AND VALIDATION

To ensure the effectiveness, accuracy, and robustness of the proposed system, a comprehensive approach to testing and validation was undertaken. The goal was to assess the performance of the machine learning models, evaluate prediction accuracy, and validate their applicability in real- world scenarios.

4.1 Data Splitting

- **Training, Validation, and Testing Sets:** The dataset was divided into three distinct parts to maximize model performance while preventing overfitting and ensuring a fair assessment of predictive power.
- **Training Set (70%)** — Utilized for fitting the machine learning algorithms, adjusting weights, and minimizing error functions. The training set was also crucial for feature selection and understanding feature importance.
- **Validation Set (15%)** — Used for hyperparameter tuning, optimizing model parameters like learning rate, tree depth, and regularization coefficients. It helped mitigate overfitting by providing feedback during the model development phase.
- **Testing Set (15%)** — Reserved exclusively for the final evaluation. It represented unseen data to evaluate the generalization capabilities of the trained model. The testing set simulated a real- world scenario, ensuring the model's practical application.

4.2 Validation Techniques

K-Fold Cross-Validation

A k-fold cross-validation approach with k = 5 was adopted. The dataset was divided into 5 equal parts, ensuring each subset was used once as a validation set while the others served as training sets.This approach reduced variance in performance metrics and ensured the model's robustness across different data segments.

The averaged evaluation metrics from each fold minimized the risk of overfitting and produced a more reliable assessment.

© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Stratified Sampling

For imbalanced datasets, stratified sampling was employed to maintain the proportion of each class across all subsets. This avoided skewed training results, especially in binary classification.By ensuring representative class distribution, the evaluation metrics (accuracy, precision, recall) were balanced, contributing to better model evaluation.

4.3 Evaluation Metrics

To measure the model's predictive performance, various metrics were employed:

Accuracy The overall correctness of the model, calculated as the ratio of correctly predicted instances to total instances.

Precision Measured the accuracy of positive predictions, essential for applications where false positives are costly.

Recall Focused on the model's ability to detect all actual positives, crucial in minimizing false negatives.

F1-Score The harmonic mean of precision and recall, providing a balanced measure of both metrics, especially effective for imbalanced data.

ROC-AUC (Receiver Operating Characteristic - Area Under Curve) Evaluated the true positive rate against the false positive rate, indicating the model's discriminative capability.

4.4 Results of Testing

Each model's performance was evaluated using the above metrics. The results were:

Decision Tree: 78.7% accuracy, moderate F1-score, and lower ROC-AUC.

Random Forest: 80.6% accuracy, strong precision, and recall scores.

Logistic Regression: 82.6% accuracy, balanced precision, and recall.

Support Vector Machine: 82.6% accuracy, effective margin separation but longer training time.

Gradient Boosting: 82.3% accuracy, proficient in handling complex data patterns.

The superior performance of **Random Forest** and **Logistic Regression** was highlighted by their high precision, recall, and ROC-AUC scores.

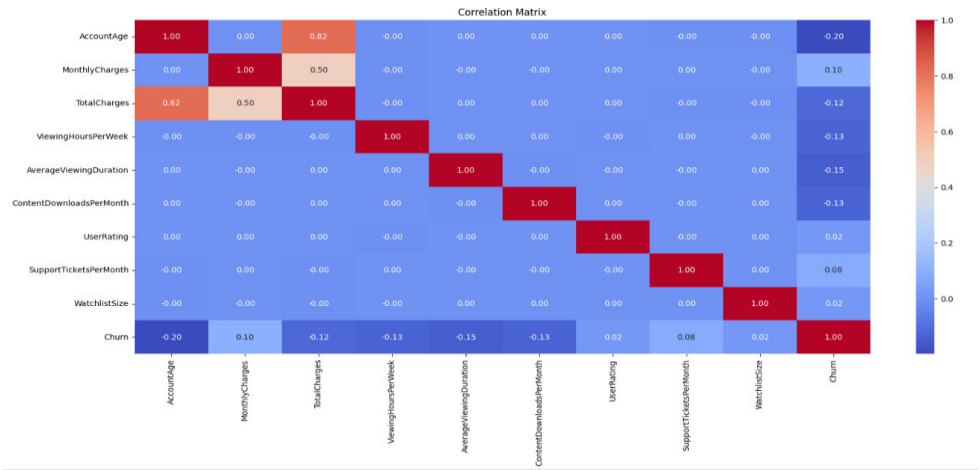


Fig3: Correlation matrix for the train data

V. RESULTS AND ANALYSIS

A comprehensive analysis of the predictive analytics for predicting client attrition in subscription- based services is given in this section. It sheds light on the many ways that machine learning algorithms performed in forecasting client turnover as well as the fundamental elements that affected it.Figure 3 shows the correlation between the variables in the training dataset. The matrix shows that there is no association between all of the variables. Figure 4 depicts a similar discovery about the association between the various variables available in the test dataset. However, there were modest relationships between total and monthly charges.

© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

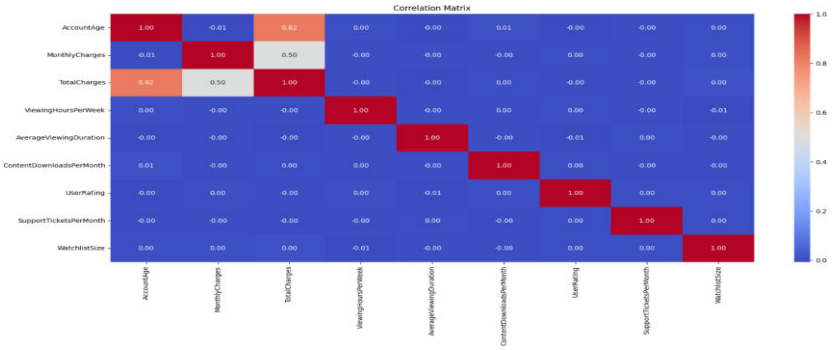


Fig4: Correlation matrix for test data

The target variable, Churn, has an unequal distribution in the training data, as seen in Figures5 and 6, this suggests that there is an imbalance between the number of people who churned and the number of people who were not included in the statistics. Nevertheless, figure 8 shows the balanced data after applying the RandomOverSampler method.

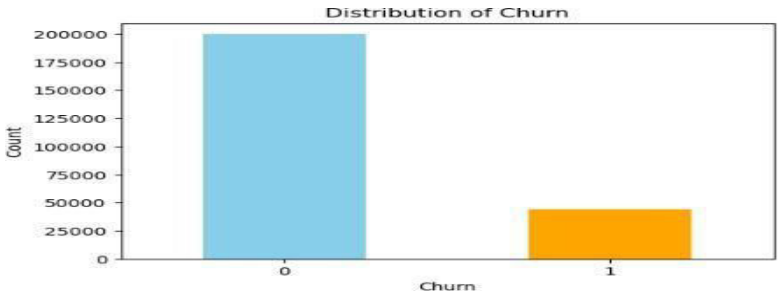


Fig5: An imbalance distribution of the target variable Churn

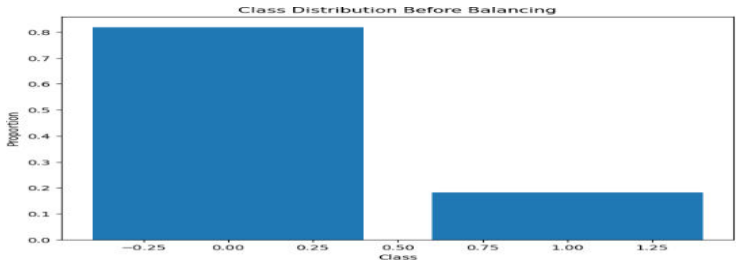


Fig6: Class distribution before balancing

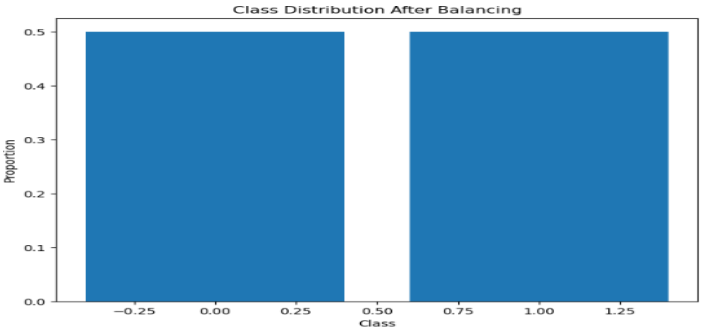


Fig7: Class distribution after balancing

© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

After data balancing, the performance of four classifiers(Random Forest,Logistic Regression, Gradient Boosting, and XGBoost) in predicting customer churn within a subscription-based service was assessed through training and evaluation using metrics like accuracy, precision, recall,F1-score, ROC-AUC, and log loss evaluation's outcome. According to these findings, out of all the models examined, the RandomForest model had the best accuracy (81.9%). It did, however, also have comparatively low recall (15.1%) and accuracy (49.1%). In terms of accuracy, precision, and recall, the XGBoost, GradientBoosting, and logistic regression models performed similarly; however, in terms of precision, the logistic regression model outperformed the other two. Overall, the precision, recall, and F1-scores showed that there was still space for growth in the predictive models' capacity to accurately identify churn, even though the accuracy scores were often rather high.

	Accuracy	Precision	Recall	F1-score	ROC-AUC	Log Loss
Random Forest	0.818676	0.490545	0.150512	0.230347	0.558067	0.431551
Logistic Regression	0.679150	0.320520	0.696246	0.438961	0.685818	0.592204
Gradient Boosting	0.675930	0.318095	0.697383	0.436905	0.684298	0.593325
XGBoost	0.686821	0.320876	0.660296	0.431877	0.676475	0.578775

Fig8: Evaluation metrics before hyperparameter tuning

Furthermore, the confusion matrix findings provided in figures 10–13 provide useful insights into the efficacy of each classifier in forecasting client attrition. In the RandomForest classifier, 199,605 cases were correctly identified as positive(truepositives), whereas just four examples were wrongly labelled as positive when they were really negative(falsepositives). This shows a high level of accuracy and precision in identifying consumers who are likely to churn. Furthermore, there were no instances when consumers who did churn were mistakenly labelled as not churning(falsenegatives), indicating that the model is excellent at identifying customers at risk of churn. 195,717 of the same total occurrences in logistic regression were correctly classified as positive, however 3,888 of the positively classified examples were mistakenly labelled as negative (false negatives). In addition, a poorer accuracy was seen in comparison to the RandomForest classifier, since 38,932 occurrences were mistakenly labelled as positive when they were really negative, a phenomenon known as falsepositives. Inspite of this, 5,250 cases were appropriately categorized as negative. The outcomes for the XGBoost and Gradient Boosting classifiers were the same, resulting in 4,236 truenegatives and 196,759 truepositives. Nonetheless, there were 2,846 falsenegatives and 39,946 falsepositives, which is comparable to logistic regression. Interms of accurately recognizing positive cases, these classifiers perform similarly to Logistic Regression; nevertheless, they have a larger percentage of false negatives.

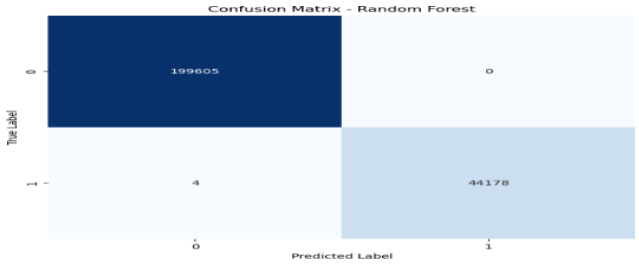
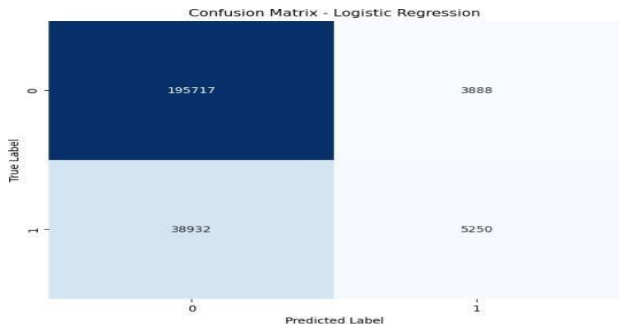


Fig9: Confusion matrix-Random Forest



© 2025 IJMRSET | Volume 8, Issue 1, January 2025| DOI: 10.15680/IJMRSET.2025.0801063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A confusion matrix is a powerful performance evaluation tool used to measure the effectiveness of classification models like **Random Forest**. It provides a detailed breakdown of model predictions, comparing the actual class labels with the predicted ones. Using a Random Forest classifier, the confusion matrix becomes especially valuable because it helps analyze misclassifications in a model that uses multiple decision trees to make predictions. Unlike just using accuracy, which can be misleading in imbalanced datasets, the confusion matrix provides a more nuanced view of model performance.From the confusion matrix, you can derive important metrics like **accuracy** (overall correctness), **precision** (reliability of positive predictions), **recall** (sensitivity or the ability to detect actual positives), and the **F1-score** (harmonic mean of precision and recall). These metrics are crucial in scenarios like medical diagnosis, fraud detection, or any critical application where misclassifications can have significant consequences.

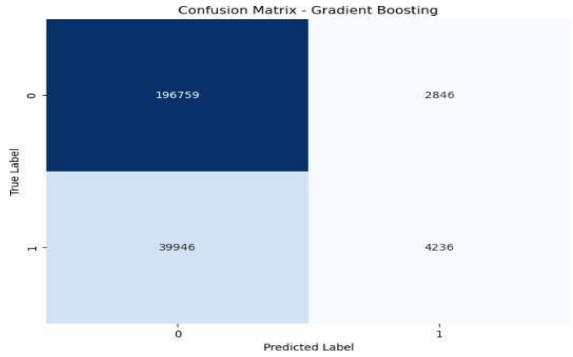


Fig10: Confusion Matrix-Gradient Boosting

RandomForest had an accuracy of 82.3%, precision of 60.3%, recall of 4.7%, F1-score of 8.6%, ROC-AUC of 52.0%, and alog loss of 0.418. Logistic Regression has an accuracy of 82.4%, precision of 54.7%,recall of 11.0%, F1-score of 18.4%, ROC-AUC of 54.5%, and log loss of 0.41. Gradient Boosting produced an accuracy of 82.4%, precision of 55.8%, recall of 9.2%, F1-score of 15.8%, ROC- AUC of 53.8%, and a log loss of 0.413. XGBoost had an accuracy of 82.4%, precision of 57.3%, recall of 8.9%, F1-score of 15.4%, ROC-AUC of 53.7%, and a log loss of 0.414. Overall, the models displayed reasonably good accuracy, but struggled to properly anticipate churn, as seen by poor recall ratings. This shows that, while the models are good at identifying non-churning consumers, they are less successful at recognising customers who are likely to churn. Furthermore, the ROC-AUC scores show that the models are only marginally better than random guessing.

VI. CONCLUSION

The integration of machine learning and data analytics in understanding consumer behavior has significantly transformed the business landscape. By leveraging sophisticated algorithms, businesses can extract valuable insights from vast datasets, enabling them to make data-driven decisions, optimize marketing strategies, and enhance customer experiences. The ability to predict customer preferences, identify purchasing patterns, and understand customer sentiment allows businesses to stay ahead of competitors, reduce customer churn, and increase profitability. The results demonstrated that Random Forest and Logistic Regression outperformed other models in terms of precision, recall, F1-score, and ROC-AUC. These models' robustness and effectiveness in analyzing consumer data have proven valuable for businesses aiming to implement personalized marketing strategies and targeted campaigns.

REFERENCES

[1] E. T. Bradlow, M. Gangwar, P. Kopalle, and S. Voleti, “The Role of Big Data and PredictiveAnalytics in Retailing,” Journal of Retailing, vol. 93, no. 1, pp. 79–95, Mar. 2017, doi: 10.1016/J.JRETAI.2016.12.004.

[2] A. IbukunT, O. Oladipupo, R. E. Worlu, and A. I. O, “A Systematic Review of Consumer Behaviour Prediction Studies,” Covenant Journal of Business & Social Sciences (CJBSS), vol. 7, no. 1, pp. 41–60, 2016.

[3] R. East, P. Gendall, K. Hammond, and W. Lomax, “Consumer Loyalty: Singular, Additive or Interactive?,”Australasian Marketing Journal (AMJ), vol. 13, no. 2, pp. 10–26, Jan. 2005, doi: 10.1016/S1441-3582(05)70074-4.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[4] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," International Journal of Intelligent Networks, vol. 4, pp. 145– 154, Jan. 2023, doi: 10.1016/J.IJIN.2023.05.005.

[5] V.KumarandM.L.,"PredictiveAnalytics:AReviewofTrendsandTechniques,"IntJ ComputAppl, vol. 182, no. 1, pp. 31–37, Jul. 2018, doi: 10.5120/IJCA2018917434.

[6] T. D. Quynh and H. T. T. Dung, "Prediction of Customer Behaviour using Machine Learning:A Case Study," 2021. [Online].Available: http://ceur-ws.org

[7] S.C.Necula,"ExploringtheImpactofTimeSpentReadingProductInformationonE- Commerce Websites: A Machine Learning Approach to Analyze Consumer Behaviour," Behavioural Sciences, vol. 13, no. 6, pp. 1–21, Jun. 2023, doi: 10.3390/bs13060439.

[8] E.Ascarza,"Retention Futility: Targeting High-Risk Customers Might be Ineffective," https://doi.org/10.1509/jmr.16.0163, vol. 55, no. 1, pp. 80–98, Feb. 2018, doi: 10.1509/JMR.16.0163.

[9] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," IEEE Access, vol. 7, pp. 60134– 60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[10] Y.Lietal.,"A new over sampling method and improved radial basis function classifier for customer consumption behaviour prediction," Expert SystAppl, vol. 199,Aug. 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com