Customer Churn Analysis Using SQL, Machine Learning & Power BI

Mrs. Bhavyashree S P[1], Shridhar S Kalasannavar[2], Sharanabasava Kandakur[3], Riteek Patil[4], Saicharan[5]
[1]*Assistant Professor, Department of Computer Science and Engineering Acharya Institute of Technology, Bengaluru Visvesvaraya Technological University, India*
[2,3,4,5]*Department of Computer Science and Engineering Acharya Institute of Technology, Bengaluru Visvesvaraya Technological University, India*

***Abstract*—Customer churn[1] is a very challenging in telecom companies, as the loss of customers directly impacts revenue and business growth. Predicting churn early allows organizations to improve customer satisfaction and implement effective retention[2] strategies. This mini project presents an end-to-end churn prediction system using SQL-based ETL processes, Power BI visual analytics, and a Machine Learning model. Raw telecom data is cleaned, transformed, and structured through an ETL pipeline. The processed data is then used to build interactive dashboards that highlight key demographic, service-related, and behavioral factors influencing churn. A Random Forest classifier is trained to categorize customers as churned or retained, helping identify high-risk individuals.**

By integrating data engineering, analytics, and predictive modeling, this system provides meaningful insights to help telecom companies make informed decisions and reduce customer loss.

Additionally, the project demonstrates how different technologies can work together to create a practical business solution. SQL[7] ensures reliable data preparation, Power BI enables clear visualization of trends, and the machine learning model adds predictive power.

I. INTRODUCTION

In the highly competitive telecom industry, customer retention[2] has become a growing challenge. With increasing service alternatives and competitive pricing, customers frequently switch from one provider to another, resulting in significant revenue loss. According to global telecom reports, companies lose millions each year due to customer churn[1], and retaining existing customers has proven to be far more cost-effective than acquiring new ones. This makes early churn[1] detection not just beneficial—but essential for long-term business success.

Telecom companies handle massive volumes of customer data daily, yet much of this information remains underutilized. Factors such as service dissatisfaction, billing issues, network problems, and poor customer engagement often contribute to churn. Without timely insight into these behavioral patterns, organizations struggle to intervene effectively. As customer expectations continue to rise, traditional methods of analyzing user behavior are no longer sufficient.

Recognizing the urgent need for a data-driven retention[2] strategy, our team developed a comprehensive Customer Churn[1] Prediction System that leverages SQL-based ETL processes, Power BI analytical dashboards, and Machine Learning algorithms. This integrated solution enables companies to not only understand why customers leave but also predict who is most likely to leave in the near future.

Our system begins with raw telecom data, which is cleaned, validated, and structured through an ETL pipeline. The refined dataset is then used to build interactive dashboard[3] that reveal churn-related patterns across demographics, services, and customer tenure. To strengthen decision-making, a Random Forest classification model is implemented to accurately identify high-risk customers based on historical trends.

II. RELATED WORK

In recent years, studies on Customer safety solutions have gained significant attention because of the increasing need for reliable and effective technologies. Churn related patterns over demographics, services, and customer tenure. To improve the decision-making, a Random Forest classification model is used to accurately find high-risk customers based on data.

Data Preprocessing & ETL-Approaches[12] Many research efforts emphasize the importance of high-quality preprocessing for churn prediction. Studies have highlighted techniques such as missing value imputation, outlier handling, normalization, and categorical encoding to enhance model performance. ETL-pipelines[13] are commonly used to structure large telecom datasets, enabling reliable downstream analysis. Researchers agree that clean, well-engineered data significantly improves both visualization accuracy and machine learning outcomes.

Statistical & Behavioral Analysis: Traditional churn[1] analysis often involves exploring customer behaviour using statistical methods and descriptive analytics. Several studies have used demographic segmentation, service usage patterns, billing trends, and complaint history to identify churn[1] signals. Visualization tools like dashboards and correlation analysis have been used to reveal how factors such as contract type, network issues, or monthly charges influence customer retention.

Machine Learning Models for Churn Prediction: Extensive research has been seen in various machine learning algorithms to classify churners and nonchurners. Commonly studied models include Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, Gradient Boosting, and Neural Networks. Among these, ensemble models especially Random Forest and XGBoost[9]—have consistently shown high accuracy due to their proeficiency to handle nonlinear data and reduce over training. Many studies also analyze model metrics such as precision, recall, F1score[6], and ROC curve[10] to compare performance.

Machine learning has become a core technique for predicting customer churn, and a wide range of models have been explored to understand which customers are likely to leave a service. Traditional algorithms like Logistic Regression are often used as a baseline because they provide simple, interpretable results. Decision Trees offer clear rule-based predictions, making it easy to understand why a customer is labeled as churn or nonchurn. More advanced models such as Support Vector Machines (SVM) capture complex boundaries between customer groups, while Neural Networks learn deeper patterns when larger datasets are available.

Precision focuses on how accurate the model's positive predictions are—meaning, out of all the customers predicted as "churners," how many true Are meaning, out of all the customers predicted as "Churners," how many true Recall[14], on the other hand, measures how many actual churners the model successfully detects. It tells us how well the model captures true positives without missing customers who are genuinely at risk of leaving.

While Precision and Recall[14] individually explain different aspects of model performance, the F1-Score[6] combines both into a single balanced metric. It is the harmonic mean of precision and recall[14], giving a more fair representation when the dataset is imbalanced—especially important in churn prediction where the number of churners is usually much lower than nonchurners. A high F1-Score[6] means the model is not only making accurate predictions but is also identifying churners effectively. This makes F1-score[6] a reliable indicator for real-world decision-making.

III.SYSTEM ARCHITECTURE

In addition to providing immediate safety features, the system emphasizes long-term usability through continuous monitoring and intelligent decision-making. The combination of sensor data, location tracking, and machine learning enables the device to learn user behaviour patterns over time, helping it differentiate between regular activity and genuine distress situations.

At the core of the system is an ESP32-based wearable device equipped with GPS for continuous location tracking, SOS triggering mechanisms, and biometric sensors that detect unusual physiological changes such as sudden movements or elevated heart rate. When the device senses danger—automatically or through manual input—it instantly sends alerts to emergency contacts. Features such as geofencing notify guardians when the user moves outside predefined safe zones, while a built-in shock generator acts as a self-defense tool to temporarily disable an attacker. The ESP32-CAM module further enhances safety by capturing images and enabling live video streaming, providing crucial evidence during critical moments.

A secure cloud platform supports the device, ensuring reliable data storage, remote access, and real-time analytics. Machine learning algorithms continuously analyze sensor data to find risk patterns and improve system responsiveness. Integrated communication modules such as GSM, GPS, and mobile notifications ensure that alerts are sent immediately through SMS, calls, and app notifications.

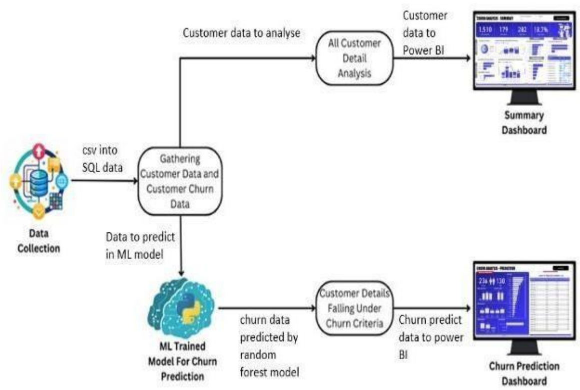

Figure-1: System Architecture

The diagram illustrates the complete workflow of a customer churn prediction system, starting from data collection and SQL[7] storage to machine learning–based churn prediction using a Random Forest model. The processed results are then visualized in Power BI through a summary dashboard and a churn prediction dashboard for decision-making.

IV. METHODOLOGY

The proposed methodology emphasizes seamless coordination between hardware components, communication modules, and software services to ensure swift and accurate emergency response. The device continuously monitors system health parameters such as battery status, sensor activity, and network connectivity to maintain optimal readiness.

The proposed methodology introduces a data- driven approach to understand and predict customer churn using ETL processing, Power BI analytics, and machine learning. It overcomes the limitations of existing systems by providing automated data cleaning, interactive dashboard[3], and predictive capabilities instead of static reports. The system first prepares and models data through SQL[7] Server ETL[5] and Power BI transformations. It then analyzes historical churn patterns and trains a Random Forest model to identify high-risk customers. Finally, the predictions are visualized in Power BI, giving businesses actionable insights to improve retention.

by loading all customer records into the system, followed by handling missing or incomplete values to ensure the dataset is clean.

The process begins by loading preprocessed dataset into the analysis environment. Then, churn data distribution is visualized to observe how many customers are leaving and from which segments. A correlation analysis is performed to check how different features relate to churn. Based on these patterns, the main churn drivers—like contract type, tenure, or service issues—are identified. Finally, meaningful insights are extracted to guide decisions and improve customer retention.

By selecting suitable algorithms that fit the problem, such as Random Forest or Logistic Regression. The chosen models are then trained using the prepared dataset. After training, hyperparameters are tuned to improve accuracy and performance. The models are corrected using metrics like accuracy, precision, recall, and F1-score. And finally, the most efficient model is used for deployment and prediction tasks. By combining the cutting-edge technology with a user-friendly approach, this solution ensures scalability, efficiency, and reliability, making it a robust and innovative.

New customer data is then entered, which serves as the input for prediction. The model calculates the churn probability for each customer and identifies those who are at high risk of leaving. These high-risk customers are flagged for further attention. Finally, all prediction results are displayed in the dashboard.

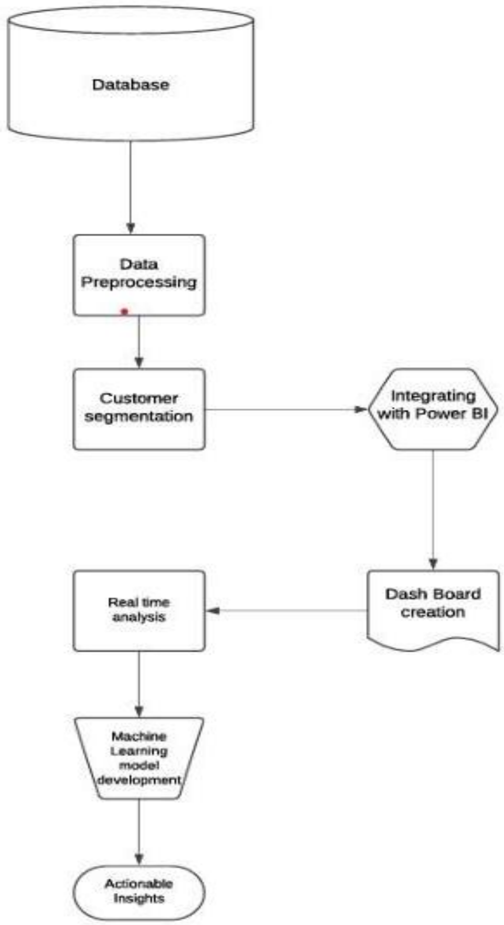

Figure-2: Data-Flow diagram

V. FEATURES

The Customer Churn Analysis system is designed to help telecom companies monitor, predict, and prevent customer attrition effectively. It continuously tracks customer behavior, including usage patterns, service interactions, and engagement levels, enabling the detection of early signs of churn.

$$\text{Recall} = \frac{TP}{TP+FN}$$

Recall measures how effectively a model identifies all actual positive cases. A higher recall means the model is missing fewer churners, making it especially important when failing to detect a positive instance has a high cost.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Precision indicates how accurate the model's positive predictions are. In churn prediction, it tells us what proportion of customers predicted as churners actually churned.

$$\text{F1-Score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

F1-score[6] is the important one here to find the accuracy of the model. It is 2 times of precision and recall by sum of two precision and recall.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

The system is also designed with a user-friendly interface and can integrate with existing CRM[8] platforms, ensuring smooth navigation, easy access to insights, and seamless execution of retention strategies.

VI. RESULTS

Figure-3. Churn analysis dashboard

The diagram illustrates the complete workflow of a customer churn prediction system, starting from data collection and SQL storage to machine learning–based churn prediction using a Random Forest model. The processed results are then visualized in Power BI through a summary dashboard and a churn prediction dashboard for decision-making

Customer Churn Analysis system is an integrated platform designed to enhance business decision-making by detecting at-risk customers and providing actionable insights to prevent churn. It typically incorporates a database management system (such as SQL Server or MySQL) for storing structured customer data, a data processing pipeline for cleaning and transforming raw information, and a machine learning module for predictive analysis.

Upon detection of accuracy of churn, the system generates automated alerts for the management or customer retention team, allowing timely interventions such as targeted offers, personalized communication, or customer support. Advanced implementations integrate with business intelligence tools like Power BI, enabling real-time monitoring through interactive dashboard[3] and seamless visualization of trends, demographic patterns, and risk segments. The first and foremost objective of this project is to establish a fast, reliable, and efficient system for seeing future customer churn, supporting proactive retention strategies, and improving overall business performance using modern data analytics and machine. The system begins with comprehensive data integration, collecting information from multiple sources like customer profiles, transaction history, service usage logs, and complaint records. Raw data is cleaned, validated, and transformed using SQL- based ETL processes to ensure consistency, accuracy, and completeness. This structured dataset forms the foundation for predictive modeling and analytical visualization, enabling businesses to gain a unified view of customer.

These models analyze historical and current data to find patterns associated with the customer attrition. Key performance metrics like accuracy, precision, recall[14], and F1-score are used to evaluate model reliability. Complete workflow of a customer churn prediction system, starting from data collection and SQL storage to machine learning–based churn prediction using a Random Forest model. The processed results are then visualized in Power BI score are used to evaluate model reliability Interactive dashboards are designed using Power BI to provide real _time insights into customer behavior and churn trends. Users can filter and explore data by demographics, services, tenure, and engagement levels to understand the drivers of churn.

By visualizing these patterns, management can prioritize high-risk customers, design personalized retention campaigns, and monitor the effectiveness of implemented strategies.

Figure-4. churn prediction dashboard

The architecture is designed to be scalable and adaptable to growing datasets and expanding user requirements. Automated workflows ensure that new data is continuously processed and updated within the system without manual intervention.

The architecture is designed to be scalable and adaptable to growing datasets and expanding user requirements. Automated workflows ensure that new data is continuously processed and updated within the system without manual intervention.

By visualizing these patterns, management can prioritize highrisk customers, design personalized retention campaigns, and monitor the effectiveness of implemented strategies. The visual interface also simplifies decision-making by providing clear, actionable information in an intuitive format.

```
# Initialize the Random Forest Classifier
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)

# Train the model
rf_model.fit(X_train, y_train)
```

```
▾ RandomForestClassifier
▶ Parameters
```

Figure-5: Machine learning model

This image represents the machine learning model has loaded in the program successfully and also the data to train it.

Customer Churn Prediction and Analysis Dashboard, successfully leverages machine learning and Power BI to provide a comprehensive, data-driven solution for customer churn analysis. By integrating predictive models with interactive dashboards, the platform effectively empowers businesses to find at-risk customers and implement targeted retention strategies. The use of SQL Server, Python, and Power BI ensures that the system is both scalable and efficient, offering a seamless experience for stakeholder[4].

```
Confusion Matrix:
[[784  46]
 [133 239]]

Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.94      0.90       830
           1       0.84      0.64      0.73       372

    accuracy                           0.85      1202
   macro avg       0.85      0.79      0.81      1202
weighted avg       0.85      0.85      0.84      1202
```

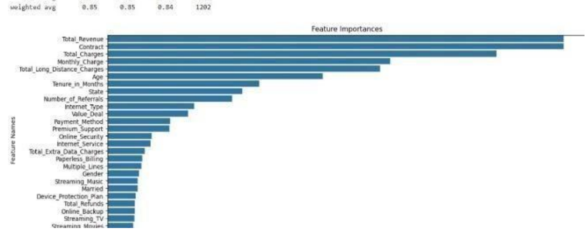

Figure-6 Accuracy of random forest model

The image presents the results of a Random Forest classifier[11] applied to a telecom customer churn dataset. At the top, a confusion matrix[19] show the model's performance in predicting churned versus retained customers, showing 784 true negative[16] (customers correctly predicted as retained), 46 false positives[17] (customers incorrectly predicted as churned), 133 false negatives[18] (customers incorrectly predicted as retained), and 239 true positive[15] (customers correctly predicted as churned). Below the matrix, a classification report highlights key evaluation metrics: the model achieves a precision of 0.85 for retained and 0.84 for churned customers, while recall[14] is 0.94 for retained and 0.64 for churned, indicating that the model identifies retained customers more reliably than churned ones. The overall prediction accuracy is 85%, and both macro and weighted averages show balanced performance across classes, with a weighted F1-score of 0.84. The bottom half of the image displays a feature importance chart, revealing the factors that most influence churn predictions. The top features include Total Revenue, Contract Type, Total Charges, Monthly Charge. Below the matrix, a classification report highlights key evaluation metrics: the model achieves a precision of 0.85 for retained and 0.84 for churned custom

The bottom half of the image displays a feature importance chart, revealing the factors that most influence churn predictions. The top features include Total Revenue, Contract Type, Total Charges, Monthly Charge, and Total Long-Distance Charges, while other and service-related attributes have lesser but notable importance.

The analysis provides valuable insights for targeted marketing campaigns, personalized offers, and loyalty programs, ensuring that retention efforts are focused on high-value or high-risk customers

VII. ADVANTAGES

The model shows a Python script where essential libraries such as pandas, NumPy, seaborn, scikitlearn, and joblib are imported. perform data by visualizing these patterns, management can prioritize high-risk customers, design personalized retention campaigns, and monitor the effectiveness of implemented strategies. preprocessing and machine learning tasks. The code reads customer churn data from an Excel file into a panda A Customer Churn Analysis project offers several dataframe significant advantages for businesses. It enables companies to identify at-risk customers who are likely to leave, allowing them to take proactive measures to retain these clients and reduce potential revenue loss. The analysis provides valuable insights for targeted marketing campaigns, personalized offers, and loyalty programs, ensuring that retention efforts are focused on high-value or high-risk customers. By understanding the key drivers of churn—such as Total Revenue, Contract Type, and Monthly Charges— businesses can improve their products, services, and customer support, improves the overall customer experience. Additionally, churn analysis supports datadriven decision-making, helps prioritize resources efficiently, and informs long-term business strategies. Overall, it strengthens competitive advantage by boosting customer retention, increasing brand loyalty, and maintaining market share.

By analyzing the key factors that influence churn—such as total revenue, contract type, and monthly charges—companies can enhance their products, services, and customer support, leading to an improved customer experience. This approach promotes data- driven decision-making, helps prioritize retention strategies, and supports long-term planning, ultimately strengthening brand loyalty and

VIII. CONCLUSION

This project focuses on analyzing customer behavior to predict and reduce churn in a business context. By leveraging machine learning techniques, such as Random Forest, the system identifies at-risk customers and provides actionable insights to improve retention strategies. Key factors influencing churn such as total revenue, contract type, and monthly charges are highlighted, enabling businesses to tailor personalized offers, loyalty programs, and targeted marketing campaigns.

The predictive model supports data-driven decision-making, helps prioritize resources effectively, and strengthens long-term business planning. By implementing such churn analysis solutions, companies can enhance customer satisfaction, increase retention, and maintain a competitive enhance overall business performance by improving customer satisfaction, reducing revenue loss, and maintaining a competitive edge. Future improvements could include integrating real-time analytics, expanding the dataset with additional customer behavior indicators, and deploying the sol by combining predictive modeling with strategic business actions, organizations can proactively retain valuable customers, optimize operational efficiency, and foster stronger, long-term relationship.

Key factors influencing churn—such as total revenue, contract type, and monthly charges—are highlighted, enabling businesses to design personalized offers, loyalty programs, and targeted marketing campaigns. The predictive model supports data-driven decision-making, helps keeps resources efficiently, and strengthens long-term business.

This project focuses on analyzing customer behavior to predict and reduce churn in a business environment. By leveraging machine learning techniques, such as Random Forest, the system identifies at-risk customers and provides actionable insights to improve retention strategies This project focuses on analyzing customer behavior to predict and reduce churn in a business environment The predictive model supports data-driven decision- making, helps allocate resources efficiently, and strengthens long-term business.

IX. FUTURE SCOPE

The future of customer churn analysis using advanced analytics and AI holds highest potential, giving improved business insights, real-time monitoring, and proactive retention strategies. AI-driven solutions will seamlessly integrate with customer relationship management systems and business intelligence platforms, ensuring instant detection of at-risk customers and timely interventions.

Advanced data collection and real-time analytics will play's crucial role in identifying churn patterns, automatically triggering personalized offers, loyalty programs, or engagement campaigns. With the adding of machine learning and predictive modeling, these systems will evolve to include predictive insights, forecasting potential churn trends and enabling businesses to act before customers disengage.

Customer analytics platforms will be more intuitive, efficient, and also user-friendly, integrating seamlessly into dashboards, mobile apps, and automated reporting tools. Innovations such as real-time data pipelines, cloudbased processing, and adaptive learning algorithms will enhance accuracy and scalability. Additionally, these solutions will become more accessible to businesses of all sizes, ensuring wider adoption and enabling companies to strengthen customer loyalty, reduce revenue loss, and drive sustainable growth.

REFERENCES

[1] S. Sharma *et al*., "Customer Churn Prediction in Telecom Industry Using Machine Learning Techniques," in *Proc. IEEE Int. Conf. on Computing, Communication and Automation (ICCCA)*, 2020.

[2] W. Wu *et al*., "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," *IEEE Access*, vol. 9, pp. 1–12, 2021.

[3] M. Tavakoli *et al*., "Customer Segmentation and Strategy Development Based on User Behaviour, RFM Model and Data Mining Techniques: A Case Study," in *Proc. IEEE Int. Conf. on e-Business Engineering (ICEBE)*, 2018.

[4] P. Poonsin and P. Boonma, "Efficient Segmentation of Customers Based on RFM Analysis," *Academic Publication*, 2025.

[5] R. De Alwis and E. Ekanayake, "Explainability, Risk Modeling, and Segmentation Based Customer Churn Analytics for Personalized Retention in E-Commerce," 2025.

[6] E. H. S. Addin, N. Admodisastro, S. N. S. M. Ashri, A. Kamaruddin, and Y. C. Chong, "Customer Mobile Behavioral Segmentation and Analysis in Telecom Using Machine Learning," *Applied Artificial Intelligence*, 2022.

[7] D. Banga, "Customer Segmentation for Marketing," in *Futuristic Trends in Management*, IIP Series, vol. 3, bk. 3, pt. 1, ch. 5, 2024.

[8] M. Imani, M. Joudaki, A. Beikmohammadi, and H. R. Arabnia, "Customer Churn Prediction: A Review of Recent Advances, Trends, and Challenges in Conventional Machine Learning and Deep Learning," *Preprints.org*, 2025.

[9] N. Gankidi, T. Tanzeela, S. Gundu, C. R. Prasad, M. V. Ahmed, and S. Yalabaka, "Customer Segmentation Using the Machine Learning," in *Proc. IEEE Int. Conf. on Intelligent Technologies (CONIT)*, 2022.

[10] R. Srinivasan, D. Rajeswari, and G. Elangovan, "Customer Churn Prediction Using Machine Learning Model Approaches," in *Proc. IEEE Int.*

Conf. on Electronics and Sustainable Communication Systems (ICECONF), 2023.

[11] S. Patel and M. Rathod, "Hybrid Machine Learning Framework for Customer Churn Prediction in Telecom Sector," in *Proc. IEEE Int. Conf. on Machine Learning and Applications (ICMLA)*, 2023.

[12] Kim *et al*., "Deep Customer Segmentation and Churn Modelling in Large-Scale Telecom Networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[13] Rodriguez and Singh, "Customer Churn Prediction Using Gradient Boosting, Random Forest and Explainable AI Techniques," *SN Applied Sciences*, Springer, 2022.

[14] C. Li and P. Hoang, "End-to-End AI Pipeline for Predicting Customer Churn in Subscription-Based Businesses," *arXiv Preprint*, 2025.

[15] R. Ahmed, M. Chowdhury, and S. Rahman, "Machine Learning–Based Customer Churn Segmentation and Churn Reduction Strategy for Telecom Operators," 2023.

KEY TERMINOLOGIES AND DEFINITIONS

1.Churn: Churn refers to the number of customers who stop using a company's service over a period of time. In telecom and subscription-based businesses, it represents customers who discontinue or switch to a competitor.

2.Retention: Gaining customers who leave a service during a specific timeframe. It helps companies understand how quickly they are losing customers and evaluate retention efforts.

3.Interactive Dashboard: An interactive dashboard is a visual analytics tool that allows users to explore data through filters, charts, slicers, and dynamic elements. It helps decision-makers analyse trends and gain insights in real time.

4.Stakeholders: Stakeholders are individuals or groups who are directly impacted by a project or business outcome. In churn analysis, stakeholders include managers, analysts, customer service teams, and executives who rely on insights to make decisions.

5.ETL (Extract, Transform, Load): ETL is a dataprocessing method involving extracting raw data from sources, transforming it into a clean and usable format, and loading it into a database or analytics system for further use.

6.F1-Score: The F1-score is a metric that combines precision and recall into a single value. It represents the balance between identifying true positives and reducing false predictions, especially useful when data is imbalanced.

7.SQL (Structured Query Language): SQL is a programming language used to store, retrieve, and manage data in relational databases. It helps in performing operations like filtering, joining, and aggregating data for analysis.

8.CRM (Customer Relationship Management):
CRM refers to systems and tools used by businesses to manage customer interactions, store customer data, and improve relationships. It supports customer service, marketing, and retention efforts.

9.XGBoost: XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm known for high accuracy and efficiency. It uses boosting techniques to build strong predictive models, making it effective for churn prediction.

10.ROC Curve (Receiver Operating Characteristic Curve): The ROC curve is a graphical representation of a model's ability to distinguish between positive and negative classes. It plots the true positive rate against the false positive rate at various thresholds.

11.Random Forest Classifier: A Random Forest classifier is an ensemble machine learning algorithm that builds multiple decision trees and combines their outcomes. It improves accuracy and reduces overfitting, making it reliable for churn prediction.

12.ETL Approaches: ETL approaches refer to methods and strategies used to extract raw data, clean and standardize it, and then load it into a target system. These approaches ensure that data is accurate and consistent for analysis.

13.ETL Pipelines: An ETL pipeline is an automated workflow that continuously extracts data from sources, transforms it, and loads it into a data warehouse or database. It ensures smooth and timely data flow for analytics and reporting.

14.Recall: Recall measures how well a model identifies all actual positive cases. In churn prediction, it represents how many true churners the model successfully detects.

15.True Positive (TP): A true positive occurs when the model correctly predicts a positive case. In churn prediction, this means the system correctly identified a customer who actually churned.

16.True Negative (TN): A true negative happens when the model correctly predicts a negative case. For churn analysis, this means the model accurately identified a customer who did not churn.

17.False Positive (FP): A false positive occurs when the model predicts a positive outcome incorrectly. In churn prediction, it means the model marked a customer as churned even though they actually stayed.

18.False Negative (FN): A false negative is when the model fails to identify a true positive case. In churn prediction, this means a customer churned, but the model predicted they would stay.

19.Confusion Matrix: A confusion matrix is a table that shows the performance of a classification model by comparing predicted values with actual outcomes. It displays true positives, true negatives, false positives, and false negatives, helping evaluate accuracy, precision, recall, and other metrics.

Page 2 of 12 - Integrity Overview

9% Overall Similarity

The combined total of all matches, including over lapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

43Not Cited or Quoted 13%
Matches with neither in-text citation nor quotation marks

2Missing Quotations 1%
Matches that are still very similar to source material

0Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

10% Internet sources

8% Publications

4% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.
A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Page 3 of 12 - Integrity Overview

Match Groups

43Not Cited or Quoted 13%
Matches with neither in-text citation nor quotation marks

2Missing Quotations 1%
Matches that are still very similar to source material

0Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

10% Internet sources

8% Publications

4% Submitted works (Student Papers)

Page 4 of 12 - Integrity Overview

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
ijirt.org		2%
2	Publication	
Umamaheswari M, Raguraman B, Monica A. "Design Thinking Approaches to Ind..		1%

3	Publication	
	Udara Yedukondalu, V Vijayasri Bolisetty. "Advancing Innovation through AI and...	1%
4	Publication	
	Huijian Dong. "Data Analytics in Finance", CRC Press, 2025	<1%
5	Publication	
	Arvind Dagur, Dhirendra Kumar Shukla, Nazarov Fayzullo Makhmadiyarovich, Ak...	<1%
6	Student papers	
	Queen's College	<1%
7	Internet	
	ijsrcseit.com	<1%
8	Internet	
	link.springer.com	<1%
9	Student papers	
	Nilai University	<1%
10	Publication	
	"Innovations in Data Engineering: Sustainability for Societal and Industrial Impac...	<1%
11	Internet	
	www.irjmets.com	<1%
12	Internet	
	crm.teknobgt.com	<1%
13	Internet	
	www.embraercommercialaviation.com	<1%

13 Internet

www.embraercommercialaviation.com <1%

14 Publication

Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Re.. <1%

15 Publication

Suneeta Satpathy, Álvaro Rocha, Sachi Nandan Mohanty, Tanupriya Choudhury. "... <1%