Customer Churn Prediction Project Report

1. Objective

The objective of this project is to predict customer churn using transactional, service, and behavioral data. The goal is to identify key factors that influence customer churn and build a machine learning model to proactively identify customers at risk.

2. Dataset Overview

- customers.csv: Contains customer demographics and ID.
- transactions.csv: Logs financial transactions per customer.
- service_tickets.csv: Logs number of service tickets and issues faced by customers.

3. Features Used

The final model uses the following features:
- avg_rating: Average customer feedback rating.
- total_spent: Total amount spent by the customer.
- num_tickets: Number of support tickets raised.
- num_issues: Number of issues logged.
- days_since_last_tx: Days since the last transaction.

4. Model and Methodology

A Logistic Regression model was chosen due to its simplicity, interpretability, and efficiency. The dataset was cleaned, merged, and preprocessed before being split into training and test sets. The model was trained using Scikit-learn's LogisticRegression.

5. Results and Evaluation

The model achieved an accuracy of approximately 86.5%. Precision and recall were evaluated to ensure the model performs well in identifying churn cases. Feature importance was also visualized to understand driving factors of churn.

Customer Churn Prediction Project Report

6. Key Insights

- avg_rating had a strong negative influence on churn (i.e., lower ratings lead to higher churn).
- days_since_last_tx had a strong positive correlation with churn.
- Customers with more service issues showed a slightly higher churn probability.

7. Future Improvements

- Incorporate time-series features from transaction logs.
- Use more advanced models like XGBoost or Random Forest.
- Deploy a real-time churn prediction dashboard using Streamlit.

8. Conclusion

This project provides a solid foundation for understanding customer churn using basic behavioral metrics. The model can be refined and expanded for production use in customer retention strategies.