Customer Churn Prediction Using Classical and Deep Learning Classification Models

MAST6100 – Classification Final Project
Individual Project

1. Introduction and Motivation

Customer churn prediction is a critical problem in subscription-based industries such as telecommunications, banking, and software-as-a-service platforms. Customer churn occurs when an individual discontinues their relationship with a service provider. From a business perspective, retaining existing customers is widely recognized as more cost-effective than acquiring new ones. As a result, understanding and predicting churn behavior has become a central focus for data-driven decision-making. Advances in machine learning have made it possible to leverage large volumes of customer data to detect patterns associated with churn. Traditional statistical models remain valuable due to their interpretability, while modern machine learning and deep learning models often provide improved predictive performance. This project aims to explore both perspectives through a comparative modeling approach.

2. Research Questions

This project seeks to answer several research questions related to customer churn prediction. First, it investigates whether customer churn can be accurately predicted using demographic, service usage, and billing data. Second, it compares the performance of a Generalized Linear Model with classical machine learning approaches such as Random Forests and Support Vector Machines. Third, it evaluates whether a deep learning model can improve predictive performance beyond traditional techniques. Finally, the project explores which factors appear to be most influential in driving churn behavior.

3. Dataset Description

The dataset used in this project is the Telco Customer Churn dataset provided by IBM and made publicly available on Kaggle. The dataset contains 7,043 observations, each representing a unique customer of a telecommunications company. The target variable is churn, which indicates whether a customer has discontinued the service. Predictor variables include customer demographics, contract details, subscribed services, and billing information. Examples include gender, senior citizen status, tenure, contract type, internet service, technical support, monthly charges, and total charges. This dataset is well-suited for classification tasks due to its size, diversity of predictors, and real-world relevance.

4. Data Preprocessing

Several preprocessing steps were required before model training. First, observations with missing values in the TotalCharges variable were removed. These missing values primarily occurred for customers with very short tenure. Next, categorical variables were converted into numerical representations using one-hot encoding. Numerical features such as tenure and monthly charges were standardized to ensure consistent scaling across models. The dataset was then split into training and testing sets using an 80/20 ratio. Stratified sampling was applied to preserve the original churn distribution in both sets. These preprocessing steps ensured that all models were trained on clean, comparable data.

5. Methodology

Four classification models were implemented in this project. Logistic regression was used as the baseline Generalized Linear Model. It models the log-odds of churn as a linear function of the predictors and provides interpretable coefficients. A Random Forest classifier was implemented to capture non-linear relationships and interactions among variables. Support Vector Machines with a radial basis function kernel were used to model complex decision boundaries in high-dimensional feature space. Finally, a deep neural network with multiple hidden layers was trained using the Keras framework to learn hierarchical feature representations.

6. Evaluation Metrics

Model performance was evaluated using a variety of classification metrics. Accuracy provides an overall measure of correctness, while precision and recall capture different aspects of classification performance. The F1-score balances precision and recall into a single metric. ROC-AUC was emphasized as the primary evaluation metric because it measures the ability of a model to discriminate between churners and non-churners across different thresholds. This metric is particularly useful in datasets with class imbalance.

7. Results

All models demonstrated meaningful predictive performance on the test dataset. Logistic regression served as a strong baseline, confirming that churn behavior can be partially explained through linear relationships. The Random Forest and Support Vector Machine models achieved improved performance by capturing non-linear effects. The deep neural network achieved the strongest overall performance, reflecting its ability to learn complex patterns from the data. However, this performance gain came at the cost of reduced interpretability compared to simpler models.

8. Feature Importance and Interpretation

Interpretation of model outputs revealed that contract type, tenure, and monthly charges were among the most influential predictors of churn. Customers on month-to-month contracts exhibited higher churn rates compared to those on longer-term contracts. Shorter tenure was also strongly associated with increased churn likelihood. Random Forest feature importance analysis supported these findings and highlighted the role of service-related variables such as technical support and internet service type.

9. Discussion

The results highlight the trade-off between interpretability and predictive performance. While deep learning achieved the best performance, logistic regression remains valuable for understanding the drivers of churn. In practical applications, organizations may choose to deploy multiple models to balance prediction accuracy with explainability. These findings align with existing literature on customer churn prediction, which emphasizes the importance of combining statistical and machine learning approaches.

10. Limitations and Future Work

This study has several limitations. The dataset represents a single telecommunications provider, which may limit generalizability. Additionally, some potentially important variables, such as customer satisfaction or service quality metrics, are not included. Future work could explore additional feature engineering, alternative deep learning architectures, or model explainability techniques such as SHAP values.

11. Conclusion

This project demonstrates that customer churn can be effectively predicted using classification models. While deep learning provided the strongest predictive performance, classical models such as logistic regression remain essential for interpretability. A combined modeling approach is recommended for real-world churn prediction systems.