Predicting Customer Churn Using Machine Learning Techniques

Updated Abstract:
This project aims to predict customer churn using machine learning techniques. So far, I have identified the primary dataset and completed initial data cleaning. Feedback from peers suggested a more detailed evaluation plan, which has now been added. Future work includes feature engineering and model selection.

Updated Introduction:
Customer churn is a major problem in subscription-based businesses. The goal of this project is to use data mining techniques to accurately predict which customers are likely to leave the service.

Related Work:
Previous studies have used machine learning models such as logistic regression and random forests for churn prediction. Compared to previous work, my project will test multiple ensemble learning methods to improve accuracy.

Proposed Work:
The project will use the Telco Customer Churn dataset. Data preprocessing and exploratory data analysis have been started. Planned models include decision trees, random forests, and gradient boosting. Feature selection techniques and hyperparameter tuning will be applied to optimize models.

Evaluation Plan:
The model performance will be evaluated using accuracy, precision, recall, F1-score, and AUC-ROC curve. Cross-validation will be used to ensure robust results.

Timeline:

- Data cleaning and preprocessing: Completed
- Feature engineering: Week 1-2
- Model building: Week 3-4
- Evaluation and tuning: Week 5
- Final report preparation: Week 6

References:

- Kaggle Dataset: Telco Customer Churn
- Research Paper: "Predicting Customer Churn with Machine Learning: A Systematic Review"