**Title:**  
Predicting Customer Churn in Telecommunications Using Machine Learning

**Abstract:**  
Customer churn is a critical issue in the telecommunications industry. This project aims to develop a predictive model that identifies customers likely to leave the service provider, enabling proactive customer retention strategies. Using machine learning techniques like decision trees, random forests, and gradient boosting, the project will analyze customer behavior data to predict churn with high accuracy.

**1. Introduction:**

Telecommunications companies face significant financial losses when customers leave for competitors. Being able to predict churn can help companies develop targeted marketing campaigns to retain customers. This project seeks to build a predictive model using historical customer usage data and demographics to identify customers at high risk of churn.

**2. Literature Review:**

Previous studies have utilized classification algorithms like Logistic Regression, Decision Trees, and ensemble methods such as Random Forests and XGBoost to predict churn. Studies show that combining customer demographics, service usage, and contract details can significantly improve prediction accuracy. Kaggle’s Telco Customer Churn dataset has been a benchmark for many academic and industry solutions.

**3. Project Methodology:**

* **Tools & Technologies:** Python (Pandas, scikit-learn, XGBoost, Matplotlib, Seaborn)
* **Dataset:** Telco Customer Churn dataset (public Kaggle dataset)
* **Tasks:**
  + Data cleaning and preprocessing
  + Exploratory Data Analysis (EDA)
  + Feature engineering
  + Model training (using Decision Trees, Random Forests, XGBoost)
  + Model evaluation and optimization
* **Techniques:** Classification, feature importance analysis, hyperparameter tuning

**4. Evaluation Plan:**

* **Metrics:** Accuracy, Precision, Recall, F1-Score, ROC-AUC
* **Validation Approach:** Cross-validation (k-fold) to avoid overfitting
* **Success Criteria:** Achieving an F1-Score greater than 0.85 on the validation set.

**5. Timeline:**

| **Phase** | **Timeframe** |
| --- | --- |
| Data Collection & Cleaning | 1 week |
| Exploratory Data Analysis (EDA) | 1 week |
| Model Building & Initial Evaluation | 2 weeks |
| Hyperparameter Tuning & Optimization | 1 week |
| Final Report & Presentation Preparation | 1 week |

**6. References:**

* Kaggle Telco Customer Churn Dataset: <https://www.kaggle.com/blastchar/telco-customer-churn>
* "A survey of churn prediction in telecommunications sector," Hossain et al., 2020
* Scikit-learn documentation: <https://scikit-learn.org/stable/>

**📑 Presentation Version (Slides Plan)**

**Slide 1:** Title Slide  
"Predicting Customer Churn in Telecommunications Using Machine Learning"

**Slide 2:** Project Abstract  
Brief description of the churn prediction problem and the project's goal.

**Slide 3:** Problem Introduction  
Why churn prediction matters in telecom.

**Slide 4:** Literature Review  
Key findings from previous works and datasets.

**Slide 5:** Project Approach  
Tools, datasets, tasks, and methods overview.

**Slide 6:** Techniques and Models  
Brief on algorithms: Decision Trees, Random Forest, XGBoost.

**Slide 7:** Evaluation Metrics  
Accuracy, Precision, Recall, F1-Score, ROC-AUC.

**Slide 8:** Timeline  
Table of phases and estimated time for each.

**Slide 9:** Expected Outcomes  
How success will be measured (F1-score target, impact).

**Slide 10:** References  
Dataset and research sources.