**Customer churn prediction on Telecom Dataset using Linear and Non-Linear Regression Models**

**Abstract**

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customers to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main contribution of our work is to develop a churn prediction model which assists telecom service providers to predict customers who are most likely subject to churn. The models developed in this work uses both Linear and Non-linear Regression models like Linear Regression, Decission Trees, Random Forests, etc.

Since any dataset considered will be huge in volume in terms of number of data points and also number of attributes or dimensions, we aim to reduce the dimensionality and also to overcome the issues of attribute correlation on the dataset by Principal Component Analysis before we apply Regression Algorithms. We will also analyse the dataset with Exploratory Data Analysis (EDA) techniques and also perform a Model Evaluation to measure the goodness of the models trained. We also plan to compare the models built in terms of metrices like accuracy, recall, precision, F1-score.

**Key Words:** Linear Regression, Decission Trees, Random Forests, Principal Component Analysis

**References:**

1. S. I. Abba et al., "Modelling of Uncertain System: A comparison study of Linear and Non-Linear Approaches," 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Selangor, Malaysia, 2019, pp. 1-6, doi: 10.1109/I2CACIS.2019.8825085.
2. M. S. Acharya, A. Armaan and A. S. Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.
3. Ahmad, A.K., Jafar, A. &amp; Aljoumaa, K. “Customer churn prediction in telecom using machine learning in big data platform.” J Big Data 6, 28 (2019).

**Tools/Technologies:**

1. Python 3.x,

2. Python Libraries (numpy, pandas, scikit-learn, seaborn, matplotlib, etc),

3. Anaconda-Navigator,

4. Jupyter Notebook,

5. PyCharm