CS 657 Project

VSB Power Line Fault Detection

Introduction:

From competition page:

Faults in electric transmission lines can lead to a destructive phenomenon called partial discharge. If left alone, partial discharges can damage equipment to the point that it stops functioning entirely.

The objective is to detect partial discharges so that repairs can be made before any lasting harm occurs.

DataSet:

The dataset is taken from Kaggle competition
https://www.kaggle.com/c/vsb-power-line-fault-detection/data.
It is stored in parquet format and the size of training data is around 4GB.

Each signal contains 800,000 measurements of a power line's voltage, taken over 20 milliseconds. Each column contains one signal 800,000 int8 measurements.

The training set is sufficiently large to be split into a suitable training and test set. The features are well suited for classification, as there are three integer features and one target output of either fault or no fault.

Proposed Models:

1. SVM
2. Naive Bayes
3. kNN

Team:

We will be doing this project in a group of 2. Following are the team members:

1. Ashutosh Bansal
2. Michael Harward