BIG DATA PROJECT REPORT - SSML

TEAM ID : BD2_084_111_177_376

1. **PROJECT TITLE CHOSEN** – Spark Streaming for Machine Learning
2. **DESIGN DETAILS** –
- The aim is to perform machine learning tasks on large data streams.
- We will be using the Enron Email Spam Detection dataset.
- We will use Naïve Bayes Classifier and Linear SVM.
- Naïve Bayes Classifier is a probabilistic classifier, which means it predicts on basis of probability of an object.
- Using Apache Spark for streaming.
3. **Surface level implementation details about each unit –**
4. **Reason behind design decisions-**
5. **Takeaway from the project -**