Project Status Report

Project Topic	Credit Card Fraud Detection using Machine Learning Techniques
Team Members	1. Manneyaa Jayasanker (MXJ180040) 2. Venkatesh Sankar (VXS200014) 3. Swathi Poseti (SXP190117) 4. Manasa M Bhat (MMB190005)
Technique/Algorithm you plan to implement	We will implement the Decision Tree algorithm for fraud detection. Also we will verify and compare the results with other algorithms such as Naive Bayes, KNN, Random Forest using MLlib library.
Dataset details, such as number of features, instances, data distribution	Dataset Link : https://www.kaggle.com/mlg-ulb/creditcardfraud Description : It contains 284, 807 transactions of European cardholders in a timespan of two days. Out of which 492 transactions correspond to fraud. Features: ● Time ● Amount ● V1, V2,..., V28 obtained with PCA dimensionality reduction. Since these are real time data of European cardholders the original features are not provided for confidentiality purposes. ● Class(Response variable) which takes value 1 if it is a fraud transaction and 0 otherwise.
Coding language / technique to be used	● Language - Python (Pyspark) ● Library - MLlib, Numpy , Pandas ● Framework - Apache Spark