BE COMPS
Batch B
BDA LAB

Hemal Mamtora (28)
Meghan Lendhe (25)
Aman Kalaskar (19)

Project Abstract
Stroke Prediction using Apache Spark

Stroke is the second leading cause of death worldwide and remains an important health burden both for the individuals and for the national healthcare systems. Potentially modifiable risk factors for stroke include hypertension, cardiac disease, diabetes, and dysregulation of glucose metabolism, atrial fibrillation, and lifestyle factors. The primary factors are age, gender, work conditions, bmi, etc. Data for stroke prediction is available here : https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data. Using big data techniques, one can successfully process and predict the probability whether a person will get a stroke or not.

Features:
1. Predict whether a person would have a stroke or not based on data such as age, gender, bmi, work-life, health history, etc.
2. Machine learning based solution for prediction.

Technologies:
1. Spark
2. Cassandra
3. pySpark Machine learning algorithms : Decision Tree, Random Forests, etc