

## PROJECT TOPIC

### A Data Mining Framework to Analyze Road Accident Data

**(DA) Group No.: 06**

**Project Group Members:**

1. Rajiv Kumar Yadav (H-64/161500434)
2. Shahaban Ali (H-70/161500496)
3. Anujay Jain (H-51/161500106)
4. Hemlata Saraswat (H-57/161500235)

**Project Supervisor:** Mr. Rahul Pradhan, Assistant Professor

**About the Project:**

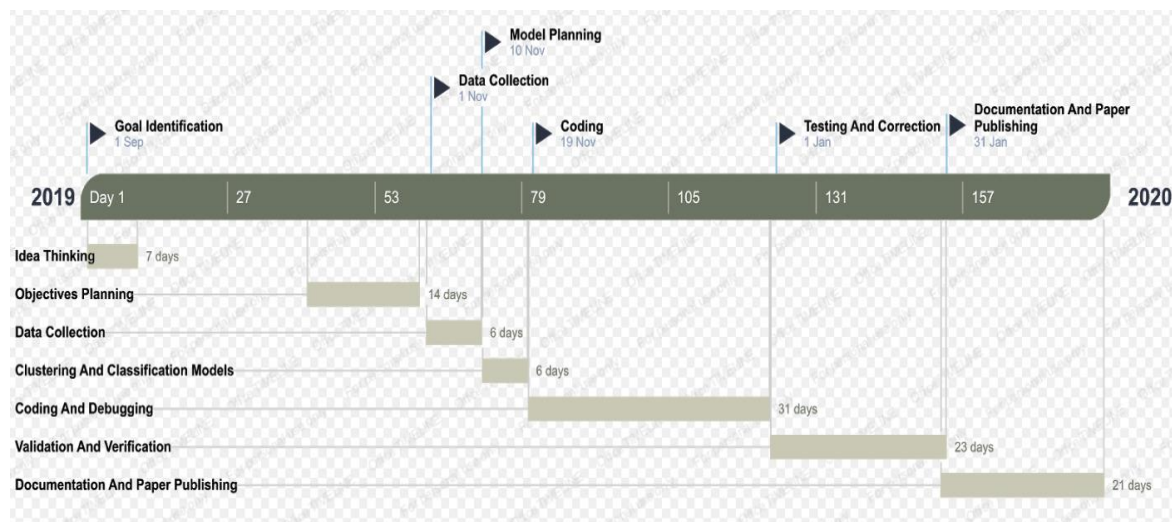
This is a research based data analysis project in which we try to analyze a large data set not capable of being analyzed by typical database or data analysis software like Excel.

To overcome this, we try to implement distributed processing using Hadoop and pipe the result with Apache Zeppelin to analyze and visualize the data set and generate a decision tree.

**Motivation:**

In the developed as well as developing countries, Infrastructure development is one of the major investments by the government, while safety of passengers on roads is of utmost importance. A road optimization during the construction or during maintenance phase, requires that the engineers analyze all the parameters that play a crucial role in ensuring safety for the passengers and preventing accidents. One of the key objectives in accident data analysis is to identify the main factors associated with road accidents.

**Project Planning:**



**Tools required:**

➤ **Hardware Requirements:**

- Processor Type: Intel Core I5-8250u CPU
- Processor Speed: 1.8 Ghz
- Hard Disk: 1 Tb
- Ram: 8 Gb

➤ **Software Requirements:**

- Operating System: Windows 10
- Front End: Apache Zeppelin
- Back End: Scala-Spark

**Signature of Project Supervisor:** \_\_\_\_\_