manipalexecutive education

CAPSTONE PROJECT

PGAIDL_21_JUN_1

Submission Date : 8th Feb 2022

PROBLEM STATEMENT

• DOMAIN: Industrial safety.

• CONTEXT: The database comes from one of the biggest industry in Brazil and in the world. It is an urgent need for industries/companies around the globe to understand why employees still suffer some injuries/accidents in plants. Sometimes they also die in such environment.

• DATA DESCRIPTION: This The database is basically records of accidents from 12 different plants in 03 different countries which every line in the data is an occurrence of an accident. Columns description:

▸ Data: timestamp or time/date information

▸ Countries: which country the accident occurred (anonymised)

▸ Local: the city where the manufacturing plant is located (anonymised)

▸ Industry sector: which sector the plant belongs to

▸ Accident level: from I to VI, it registers how severe was the accident (I means not severe but VI means very severe)

▸ Potential Accident Level: Depending on the Accident Level, the database also registers how severe the accident could have been (due to other factors involved in the accident)

▸ Genre: if the person is male of female

▸ Employee or Third Party: if the injured person is an employee or a third party

Critical Risk: some description of the risk involved in the accident

▸ Description: Detailed description of how the accident happened.

Link to download the dataset:

https://www.kaggle.com/ihmstefanini/industrial-safety-and-health-analytics-database

PROJECT OBJECTIVE: Design a ML/DL based utility which can help the professionals to highlight the safety risk as per the incident description.

manipalexecutive education

PROJECT TASK: [Duration: 3 weeks, Score: 100 points]

1. Milestone 1: [Duration: 1 weeks, Score: 40 points]
   - ▸ Step 1: Import the data
   - ▸ Step 2: Data cleansing
   - ▸ Step 3: Data preprocessing
   - ▸ Step 4: Data preparation to be used for AIML model learning
   - ▸ Step 5: Summary of Exploratory Data Analysis

   - ▸ Output: Clean data as .xlsx or .csv file to be used for AIML model learning

2. Milestone 2: [Duration: 2 weeks, Score: 60 points]
   - ▸ Input: Output of milestone 1 (This should be the clean xls or csv file)
   - ▸ Step 1: NLP pre processing
   - ▸ Step 2: Design, train and test machine learning classifiers
   - ▸ Step 3: Design, train and test Neural networks classifiers
   - ▸ Step 4: Design, train and test RNN or LSTM classifiers
   - ▸ Step 5: Choose the best performing model classifier and pickle it.
   - ▸ Output: Pickled model to be used for future prediction.