Great Learning

AIML

CAPSTONE PROJECT

NATURAL LANGUAGE PROCESSING

CHATBOT INTERFACE

PROBLEM STATEMENT

- **DOMAIN:** Industrial safety. NLP based Chatbot.

- **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 [for your reference only]

- **PROJECT OBJECTIVE:**

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

- **PROJECT TASK:** [Score: 100 points]

  1. **Milestone 1:** [Score: 40 points]
     - **Input**: Context and Dataset
     - **Process**:
       - Step 1: Import the data [3 points]
       - Step 2: Data cleansing [5 points]
       - Step 3: Data preprocessing (NLP Preprocessing techniques) [7 points]
       - Step 4: Data preparation - Cleansed data in .xlsx or .csv file [5 points]
       - Step 5: Design train and test basic machine learning classifiers [10 Points]
       - Step 6: Interim report [10 points]
     - **Submission**: Interim report, Jupyter Notebook with all the steps in Milestone-1
  2. **Milestone 2:** [Score: 60 points]
     - **Input**: Preprocessed output from Milestone-1
     - **Process**:
       - Step 1: Design, train and test Neural networks classifiers [5 points]
       - Step 2: Design, train and test RNN or LSTM classifiers [10 points]
       - Step 3: Choose the best performing classifier and pickle it. [5 points]
       - Step 4: Final Report [40 Points]
     - **Submission**: Final report, Jupyter Notebook with all the steps in Milestone-1 and Milestone-2
  3. **Milestone 3:** [Optional]
     - **Process**:
       - Step 1: Design a clickable UI based chatbot interface
     - **Submission**: Final report, Jupyter Notebook with the addition of clickable UI based interface

- Hints:
  - Please refer to the blog to understand the basic designing and functioning of chatbots: https://www.mygreatlearning.com/blog/basics-of-building-an-artificial-intelligence-chatbot/
  - To make GUI as a desk app you can use TKINTER library.
  - To make web service GUI you can use FLASK or DJANGO library.

POINTS TO REMEMBER

1. A maximum of 100 points will be awarded for this project
2. Project to be submitted within 6 weeks of date of release. Late submission will be accepted under genuine situation. Score will be given as per the below formula:

   If the current score is greater than 40 then the final score will be capped at 40.

   Else the current score will be awarded.
3. Any form of plagiarism is strictly prohibited. No score will be awarded in this case.

HAPPY LEARNING!