greatlearning
Power Ahead

AIML

CAPSTONE PROJECT

NATURAL LANGUAGE PROCESSING

CHATBOT INTERFACE

©Great Learning. Proprietary content. All Rights Reserved. Unauthorised use or distribution prohibited

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)
  - Gender: 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://drive.google.com/file/d/1_GmrRP1S2OIa02KlfOBNkYa8uxazGbfE/view?usp=sharing,

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

- **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:** [Duration: 6 weeks, Score: 100 points]

  1. **Milestone 1:** [Duration: 2 weeks, Score: 20 points]
     - **Input**: Interim report
     - **Process**:
       - Step 1: Import the data
       - Step 2: Data cleansing
       - Step 3: Data preprocessing
       - Step 4: Data preparation to be used for AIML model learning
     - **Output**: Clean data as .xlsx or .csv file to be used for AIML model learning
  2. **Milestone 2:** [Duration: 2 weeks, Score: 20 points]
     - **Input**: Output of milestone 1
     - **Process**:
       - 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
     - **Submission**: Interim report
  3. **Milestone 3:** [Duration: 2 weeks, Score: 60 points]
     - **Input**: Pickled model from milestone 2
     - **Process**: [15 points]
       - Step 1: Design a clickable UI which can automate tasks performed under milestone 1 [5 points]
       - Step 2: Design a clickable UI which can automate tasks performed under milestone 2 [5 points]
       - Step 3: Design a clickable UI based chatbot interface [5 points]
     - **Output**: Clickable UI based chatbot interface which accepts text as input and replies back with relevant answers.
     - **Submission**: Final report [45 points]

- 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.

©Great Learning. Proprietary content. All Rights Reserved. Unauthorised use or distribution prohibited