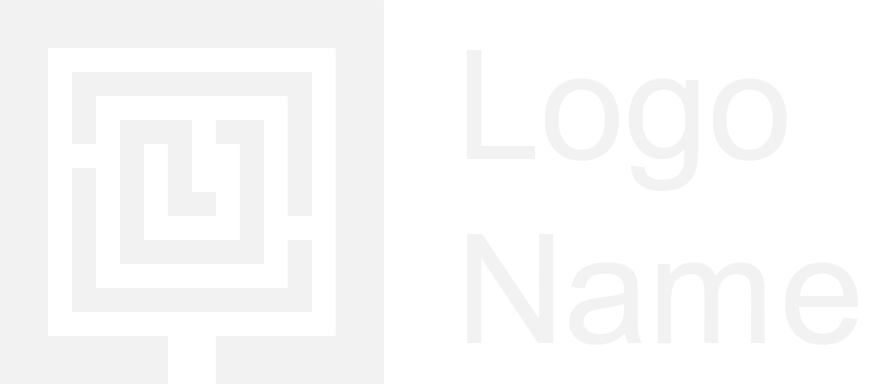


|  |  |  |
| --- | --- | --- |
| Great Learning  Capstone Project  Np;[  2018 | | |
| Project: Natural Language Processing – Chatbot Interface  Prepared by:  Ankita Singh  Chethan RH  Rahul Mashru  Dr. Raghu Gogada | |
| JUly 2021 |



# Problem Statement

Across different plants in varied industries one of the biggest risks is the potential accidents that can occur and directly/indirectly impact the worker(s) deployed on the particular job/work.

While several safety measures are normally in place to help the worker(s) avoid any potential accident, the risk of accidents cannot be eliminated 100%; and hence the need for having proactive insights on the risk levels so that the workers can take the necessary measures to avoid graveness and severity of the accident to the best possible extent.

Through the medium of a chatbot (also known as Bot, Artificial Conversational Entity, Conversational Agent), the objective is to enable the user to type the particular industrial incident description (i.e., essentially free text box), either by the Machine Learning or Deep learning algorithms, by predicting the level of potential severity (kindly refer data dictionary for more details), so that the workers deployed at the industrial site can take the necessary safety precautions and completely and /or to a large extent eliminate any catastrophic event that could cause loss of livelihood and lives.

A chatbot is like a computer program that facilitates conversation with a human to help them achieve a specific goal, and it usually replicates any chat GUI, except it is a computer that responds/helps the human with their query/request. The more comprehensively the chatbots are trained or robustly modelled, the thinner the lines and difference between how a human and a robot would respond.

The key advantages of a chatbot are:

1. They are available 24/7
2. They can easily be scaled to support multiple queries
3. The help automate certain tasks improving efficiency and allowing the company to deploy their resources in more key activities

Broadly speaking there are two kinds of chatbots:

1. Rule Based Chatbots - Based on simple rules, cannot handle complex queries
2. Self-Learning Chatbots – Bots that rely of Artificial intelligence (AI) and Machine Learning (ML) technologies to converse with users.

To elaborate further on category 2 (i.e., Self-Learning Chatbots), there are few key terminologies that are pertinent to understand:

* Natural Language Processing (NLP): A subset of AI, that helps computers read and understand natural human language.
* Natural Language Understanding (NLU): It is a sub-branch of NLP, which refers to the aspects around the Machine’s ability to interpret / understand the text that is written to it.
* Natural Language Generation (NLG): This is contrary to the NLU topic, which helps the machine respond in a language that the user can understand based on what it has understood.

About the Dataset (incl. Data Dictionary)

The Dataset contains industrial accidents that occurred across 3 countries, one of them being Brazil, and across 12 different plants, with the focus of understanding why employees still suffer from injuries and fatal accidents.

There are 425 records, with each record representing an occurrence of an accident.

|  |  |  |
| --- | --- | --- |
| # | Column | Description |
| 1 | Data | Timestamp or time/date information |
| 2 | Countries | Country where the accident occurred (anonymized) |
| 3 | Local | City in the specific country, where the plant is located (anonymized) |
| 4 | Industry Sector | Sector in which the plant belongs to |

|  |  |  |
| --- | --- | --- |
| # | Column | Description |
| 5 | Accident Level | Severity of the accident level, that actually happened, from Grade 1 to Grade 6 (1 being least severe and 6 being extremely severe) |
| 6 | Potential Accident Level | Severity of the accident level, that could potentially happen, from Grade 1 to Grade 6 (1 being least severe and 6 being extremely severe) |
| 7 | Gender | Sex of the person / worker / employee who encountered the accident i.e., Male or Female |
| 8 | Employee or Third party | If the worker is an employee or a third party |
| 9 | Critical Risk | Broad category of risk involved in the accident |
| 10 | Description | Detailed Description of how the accident happened |

# Summary of the Approach to EDA and Pre-Processing

# To be filled by Raghu

# Model Building Approach

# To be filled by Ankita / Chetan

# Results