

ASTRAZENECA AI CHALLENGE

PROBLEM 1: JIGSAW UNINTENDED TOXICITY CLASSIFICATION

THE CRUX:

A main area of focus is machine learning models that can identify toxicity in online conversations, where toxicity is defined as anything *rude, disrespectful or otherwise likely to make someone leave a discussion*.

[Adapted from the PS given]

PROBLEM 1: JIGSAW UNINTENDED TOXICITY CLASSIFICATION

OBJECTIVES:

- A. The text of the individual comment is given in the `comment_text` column. Each comment in Train has a toxicity label (target), and models should predict the target toxicity for the Test data.
- B. This attribute (and all others) are fractional values which represent the fraction of human raters who believed the attribute applied to the given comment.
- C. For evaluation, test set examples with $\text{target} \geq 0.5$ will be considered to be in the positive class (toxic).
- D. Also Individual labels are specified which needs to be correctly identified.

[Adapted from the data description given]

PROBLEM 1: JIGSAW UNINTENDED TOXICITY CLASSIFICATION

WHAT HAVE WE DONE?

- A. We have loaded the data and done some preprocessing and cleaning.
- B. We have used TFIDF vectoriser which is used to convert sentences/words into a meaningful number. It tells us how important a particular word in a comment is with respect to the corpus. Here the corpus is the comments in the `comment_text` column.
- C. The output of interest here is target so Y is the target field.
- D. The input and output field are hence identified.
- E. Following this we have done a train-test split so that we can evaluate the algorithm at last.
- F. The train value is fed into a Logistic Regression model as the input for fitting.
- G. Then We use the predict command to predict the output for the test values.
- H. The predicted and actual values are compared to check the accuracy of the model.

PROBLEM 1: JIGSAW UNINTENDED TOXICITY CLASSIFICATION

WHAT HAVE WE DONE?

- A. We have loaded the data and done some preprocessing and cleaning.
- B. We have used TFIDF vectoriser which is used to convert sentences/words into a meaningful number. It tells us how important a particular word in a comment is with respect to the corpus. Here the corpus is the comments in the comment_text column.
- C. The output of interest here is target so Y is the target field.
- D. The input and output field are hence identified.
- E. Following this we have done a train-test split so that we can evaluate the algorithm at last.
- F. The train value is fed into a Logistic Regression model as the input for fitting.
- G. Then We use the predict command to predict the output for the test values.
- H. The predicted and actual values are compared to check the accuracy of the model.

PROBLEM 1: JIGSAW UNINTENDED TOXICITY CLASSIFICATION

Term Frequency (TF)

The number of times a word appears in a document divided by the total number of words in the document. Every document has its own term frequency.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

Inverse Data Frequency (IDF)

The log of the number of documents divided by the number of documents that contain the word w . Inverse data frequency determines the weight of rare words across all documents in the corpus.

$$idf(w) = \log\left(\frac{N}{df_t}\right)$$

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

[Adapted from towards data science]

PROBLEM 2: HOTEL RECOMMENDATION CHALLENGE

The challenge requires us to predict the hotel cluster given any destination ID. We have done the following in this regard:

- We have tried to understand the data type of each column.
- We have done missing value imputation.
- We have also converted the date and time based columns into date-time data type.
- We have done some visualisations that give insights about the various columns.
- We have developed a mechanism for predicting the hotel cluster given any destination ID.

THANK YOU!