**Multi Intent Classifier**

The given dataset contains large number of Wikipedia comments which have been labeled by human raters for toxic behavior. The types of toxicity are:

* toxic
* severe\_toxic
* obscene
* threat
* insult
* identity\_hate

The goal of this project is to create a model which predicts a probability of each type of toxicity for each comment.

The steps involved are

* Data Exploration
* Text Preprocessing
* Create ML pipelines
* Evaluate Metrics
* Predictions on Test data

**Exploratory Analysis:**

The train dataset used consists of 159571 rows and 8 columns

The test dataset contains 153164 rows and 2 columns

**Distribution of length of comments from the dataset**

The average length of comments ranges between 0-500 and from the plot we can see there are some

Comments with length of 5000

Graphical user interface, application, Teams

Description automatically generated

**Count of comments in each label:**

Chart, bar chart

Description automatically generated

Here each comment can be assigned with multiple labels, so are there comments with multiple labels?

Graphical user interface, application, Teams

Description automatically generatedGraphical user interface, table

Description automatically generated

From the plot we observe that there are comments with all 6 labels and majority of the comments are not labelled

**Most common words from the comments per label:**

Text

Description automatically generatedText

Description automatically generated

**Text Preprocessing using NLTK:**

1. Cleaning to remove irrelevant items
2. Normalizing by converting to all lowercase and removing punctuation
3. Splitting text into words or tokens
4. Removing words that are too common, also known as stop words
5. Converting words into their dictionary forms, using stemming and lemmatization

**ML Pipelines:**

* Text Processing:
* Feature Extraction
* Modeling

**Evaluation Metrics:**

Average of the individual AUCs of each predicted column

**Predictions on Test data**

AUC score for Logistic Regression was slightly better than Naïve Bayes Algorithm. So, I have used Logistic Regression model for test data predictions

**Future Work:**

In this Notebook, I have used two algorithms (Logistic Regression and Naïve bayes) and compared the AUC score and chose the model based on the AUC score. The model can be further improved by trying the following methods

1. Hyperparameter tuning for Logistic Regression model
2. Use ensemble classification methods
3. Use pretrained Deep Learning models