

# Insult Detection in Social Commentary

Main goal of this project is to detect if a comment or a post online is an insult or not using various machine learning techniques.

[Github Repository](#)

## Working Demo

- [Insult Detection](#) - Enter a test a sentence to tag it.
- [Insult Detection on Live Tweets](#) - Enter a query to search related to it.

## Getting Started

These instructions will get you a copy of the project up and running on your local machine for development and testing purposes.

## File Structure

- *root*
  - **data** - Contains the data sets for the project
  - **src**
    - *main.py* - Main entry point
    - *ensemble.py* - Ensembling code
    - *preprocess.py* - Helper preprocessing code
    - *features.py* - Helper feature extraction code
  - **interactive** - Contains the jupyter notebook for interactive project representation
  - **ppt** - The Presentation (ppt and pdf)
  - **misc** - Some miscellaneous files (Sample Output.txt)
  - **visualise** - Various graphs and curves for different Classification techniques used
  - *requirements.txt* - Requirements file for installed modules.
  - *README.md* - Readme file in Markdown format
  - *README.pdf* - Readme in portable document format

## Prerequisites

- Python 3.5+
- Following Python Modules
  - jupyter==1.0.0
  - scikit-learn==0.19.0
  - scipy==1.0.0
  - nltk==3.2.4
  - numpy==1.13.1

- pandas==0.20.3
- matplotlib==2.0.2
- virtualenv==15.1.0 [Optional]

Or you can directly install all the required modules along with dependencies using requirements.txt file.

```
pip install -r requirements.txt
```

## Installing

Follow the following steps to setup a virtual environment to run the project

### 1. Install Python 3.5.x

Refer the internet for installing python.

### 1. Setup virtual environment [Optional]

```
virtualenv -p python3 venv
```

### 1. Use the virtual env for further work [Optional]

```
# For Ubuntu/Linux
source venv/bin/activate

# For Windows - CommandPrompt
.\venv\Scripts\activate.bat

# For Windows - PowerShell
.\venv\Scripts\activate.ps1

# The CLI will have a (venv) at the beginning of every line from now on.
```

### 1. Installing the required modules

```
pip install -r requirements.txt
```

### 1. Run the main file for the project

```
cd src
python main.py
```

## Interactive Testing

To test the project and visualize the project more intuitively, try using our jupyter notebook. Note: Make sure to try the following with environment properly set up.

```
cd interactive
jupyter notebook
```

A browser tab will open with the notebooks listed. Try the *Presentation.ipynb* to use the project file. Then use the the notebook in a standard way.

## Running the tests

```
cd src
python main.py
```

The above should provide with all the usefull information neccessary including *Accuracy score*, *Confusion matrices*, *ROC Curves*, *Area Under Curve score*.

## Result Interpretation

- The **confusion matrix** helps represent the precision and recall of a classifier.
- The **accuracy score** gives the percentage of accurate predictions by the model.
- The **ROC AUC** of a classifier is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example, i.e.  
 $P(\text{score}(x+) > \text{score}(x-))$

## Train data and Test data

To test a custom set of data, some modifications in the code needs to be done, as the code in its natural form splits the train data in the train and test sets, therefore using a seperate file to test data requires minor configuration changes in the code. Although this can be done easily in the Jupyter notebook available in the package.

## The training and test data

Source [Kaggle](#)

## Built With

- [Jupyter](#) - Interactive computing
- [Scikit-Learn](#) - Machine Learning and Classification Library
- [NLTK](#) - Generic NLP tasks
- [spaCy](#) - Advanced and intuitive NLP tasks (dependency parsing)

## Authors

- [Chirag Khurana](#) - [Github](#)

- Shubham Goyal - [Github](#)
- Pallavi Rawat - [Github](#)

## Acknowledgments

- [Tanmoy Chakraborty](#) - Mentor / Instructor