

## Dependencies and Versions Required to Run the Code:

To compile and run the provided classification pipeline, the following libraries and packages must be installed:

### Python version:

- Python 3.7+

### Required libraries:

- pandas
- numpy
- matplotlib
- re
- math
- scikit-learn
  - sklearn.feature\_extraction.text.TfidfVectorizer
  - sklearn.model\_selection.train\_test\_split, GridSearchCV
  - sklearn.metrics (accuracy, precision, recall, F1, ROC, AUC)
  - sklearn.ensemble.RandomForestClassifier, VotingClassifier
  - sklearn.linear\_model.LogisticRegression
  - sklearn.naive\_bayes.GaussianNB
- imbalanced-learn (for SMOTE)
  - imblearn.over\_sampling.SMOTE
- nltk (for stopwords removal)
  - Requires downloading stopwords:

```
python
```

```
import nltk
```

```
nltk.download('stopwords')
```