

### ### Report Structure

#### #### Preprocessing and TF-IDF Vectorization

- **\*\*Lemmatization and Text Cleaning\*\***: Applied lemmatization to normalize the text, along with lowercasing, removing URLs, emojis, punctuation, and stopwords.
- **\*\*TF-IDF Vectorization\*\***: Utilized TfidfVectorizer with a maximum of 5000 features to transform the cleaned text data into TF-IDF vectors.

#### #### Model Training and Hyperparameter Details

##### 1. **\*\*Logistic Regression\*\***

- Best Hyperparameters:
  - C: 10
  - solver: 'newton-cg'
  - max\_iter: 1000
- Classification Report:
  - Accuracy: 0.91
  - Precision: 0.91
  - Recall: 0.92
  - F1-Score: 0.91

##### 2. **\*\*K-Nearest Neighbors\*\***

- Best Hyperparameters:
  - n\_neighbors: 9
  - weights: 'distance'
  - algorithm: 'auto'
- Classification Report:
  - Accuracy: 0.89
  - Precision: 0.93
  - Recall: 0.85
  - F1-Score: 0.89

### 3. **Support Vector Machine (SVM)**

- Best Hyperparameters:

- C: 10

- kernel: 'rbf'

- gamma: 'scale'

- Classification Report:

- Accuracy: 0.92

- Precision: 0.91

- Recall: 0.94

- F1-Score: 0.93

### 4. **K-Means Clustering**

- Optimal Number of Clusters: 8

- Silhouette Score for Optimal Clusters: 0.0093

### 5. **Neural Networks**

- Best Hyperparameters:

- Hidden Layer Size: 100

- Alpha: 0.00096

- Activation: 'relu' for hidden layer, 'sigmoid' for output layer

- Optimizer: 'adam'

- Loss: 'binary\_crossentropy'

- Accuracy: 0.91

### 6. **FastText**

- Best Hyperparameters:

- Vector Size: 150

- Window: 7

- Epochs: 15

- Classification Metrics using Logistic Regression on FastText Embeddings:

- Accuracy: 0.82

- Precision: 0.81
- Recall: 0.84
- F1-Score: 0.83

#### #### Model Saving and Sharing

- Used joblib to save trained models and the TF-IDF vectorizer.
- Shared via Google Drive: [Model Files](<https://drive.google.com/drive/folders/1fQZwv8IbAKm3eafTXB83e8FGQdAxvOpn?usp=sharing>)