

Text Classification Project using NLP

Prepared by: Rabab

This project focuses on classifying text into positive or negative sentiment using NLP techniques and machine learning (Random Forest Classifier). The dataset contains 2600 labeled tweets.

Steps:

1. Import required libraries.
2. Upload the dataset.
3. Clean the text using tokenization and stemming.
4. Split the data into training and testing sets.
5. Transform text using TF-IDF.
6. Balance data using SMOTE.
7. Train RandomForestClassifier.
8. Print classification report.

Sample Output:

	precision	recall	f1-score	support
negative	0.77	0.49	0.60	180
positive	0.86	0.96	0.91	600
accuracy			0.85	780
macro avg	0.82	0.72	0.75	780
weighted avg	0.84	0.85	0.83	780

Conclusion:

The model achieved 85% accuracy after balancing the dataset using SMOTE. Random Forest proved to be effective in handling text classification after proper preprocessing.