Emotion Detector from text

Naman Sharma

Dataset Used - Kaggle Emotions dataset for NLP

- Structure:
  - 16,000 training samples
  - 2,000 test samples
  - 2,000 validation samples

- Format:
  - Each sample contains a sentence and its corresponding emotion label separated with a “;".

Data Preprocessing

- Text Cleaning:
  - Removing punctuation
  - Removing stopwords
  - Converting to lowercase

- Feature Extraction:
  - TF-IDF Vectorization to transform text into numerical features.

Model Selection

- Algorithms Used:
  - Logistic Regression
  - Naïve Bayes (MultinomialNB)
  - Support Vector Machine (SVM)

Model Training and Evaluation

- Training Process:
  - Splitting data into train and test sets
  - Training models on TF-IDF vectors

- Evaluation Metrics:
  - Accuracy
  - Classification Report (Precision, Recall, F1-score)

Results

- Best Model - SVM

Test Accuracy: 0.8795

Classification Report:

	precision	recall	f1-score	support
anger	0.88	0.87	0.87	275
fear	0.86	0.87	0.86	224
joy	0.88	0.93	0.91	695
love	0.77	0.71	0.74	159
sadness	0.93	0.90	0.92	581
surprise	0.73	0.61	0.66	66
accuracy			0.88	2000
macro avg	0.84	0.81	0.83	2000
weighted avg	0.88	0.88	0.88	2000

Conclusion

- SVM and logistic regression models performed well
- Imbalanced data may have led to less accurate predictions for underrepresented classes, affecting overall model performance.