Emotion Intensity Predictor Project Report

BY-RISHI JAIN

CODALAB USERNAME- Rishi_Jain

Email- rj1016743@gmail.com ,rishijainai262003@gmail.com

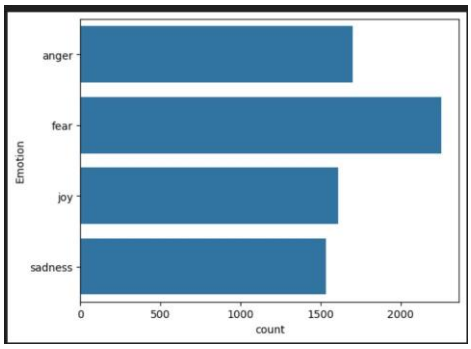
1. Introduction

This report presents an overview of the Emotion Intensity Predictor project, which aims to predict the intensity of emotions expressed in textual data using both machine learning and deep learning models.

2. Machine Learning Model

2.1 Data Visualization

This helps us understanding its characteristics and distribution



2.2 Data Cleaning

Text preprocessing techniques such as lowercase conversion, special character replacement, and URL removal are applied to prepare the data for modelling.

2.3 Sentence Embeddings using Sentence Transformers

It maps sentences & paragraphs to a 768-dimensional dense vector space and can be used for tasks like clustering or semantic search.

2.4 Model Used and Results

Three classifiers (SVM, XGB Classifier, Random Forest Classifier) are trained, and their accuracies are reported. SVM performs the best with an accuracy of 77.32%.

Test Cases:

Here 'sadness':0 ,'joy':1, 'anger':2, 'fear':3

```
st1 = ["I am very happy today"]
s1_emb = model1.encode(st1)
svm_classifier.predict(s1_emb)
```

```
array([1], dtype=int64)
```

```
st2 = ["I am very angry today"]
s2_emb = model1.encode(st2)
svm_classifier.predict(s2_emb)
```

```
array([2], dtype=int64)
```

```
st = ["I am not feeling good today"]
st_emb = model1.encode(st)
svm_classifier.predict(st_emb)
```

```
array([3], dtype=int64)
```

2.5 Classification Reports

This provides detailed performance metrics for each classifier, with SVM showing superior results compared to the other models.

```
Classification report using SVM:
              precision    recall  f1-score   support

           0       0.70      0.68      0.69       307
           1       0.85      0.81      0.83       322
           2       0.74      0.79      0.77       340
           3       0.79      0.79      0.79       451

    accuracy                           0.77      1420
   macro avg       0.77      0.77      0.77      1420
weighted avg       0.77      0.77      0.77      1420

Classification report using Random Forest Classifier:
              precision    recall  f1-score   support

           0       0.70      0.40      0.51       307
           1       0.76      0.71      0.74       322
           2       0.72      0.68      0.70       340
           3       0.61      0.84      0.71       451

    accuracy                           0.68      1420
   macro avg       0.70      0.66      0.66      1420
weighted avg       0.69      0.68      0.67      1420

Classification report using XGB Classifier:
...
    accuracy                           0.71      1420
   macro avg       0.72      0.71      0.71      1420
```

2.6 Comparison with Deep Learning Approach

Similar text preprocessing techniques used in the deep learning approach like using nltk for stopwords , tokenization and padding were applied to the machine learning model, but they yielded lower results.

```
Classification report using SVM:
              precision    recall  f1-score   support

           0       0.17      0.14      0.15       307
           1       0.24      0.18      0.20       322
           2       0.25      0.14      0.18       340
           3       0.30      0.49      0.37       451

    accuracy                           0.26      1420
   macro avg       0.24      0.24      0.22      1420
weighted avg       0.24      0.26      0.24      1420

Classification report using Random Forest Classifier:
              precision    recall  f1-score   support

           0       0.19      0.11      0.14       307
           1       0.35      0.27      0.30       322
           2       0.29      0.26      0.27       340
           3       0.32      0.49      0.39       451

    accuracy                           0.30      1420
   macro avg       0.29      0.28      0.28      1420
weighted avg       0.29      0.30      0.29      1420

Classification report using XGB Classifier:
...
    accuracy                           0.39      1420
   macro avg       0.38      0.38      0.38      1420
weighted avg       0.39      0.39      0.39      1420
```

3. Deep Learning Model

3.1 Data Visualization

Similar to Model1, dataset visualizations are presented to understand its characteristics.

3.2 Data Cleaning

Text preprocessing steps including URL removal and stop words removal are performed.

3.3 Tokenization and Padding with Tokenizer

Tokenization is the process of converting text into numerical sequences, while padding ensures uniform sequence lengths for input into neural networks.

3.4 Sequential Model Architecture

This Sequential model comprises four layers:

- Embedding Layer
- 2 LSTM Layer
- Dense Layer

3.5 Model Compilation and Results

The model is compiled using the Adam optimizer and sparse categorical cross entropy loss function, suitable for multi-class classification tasks. It achieves an accuracy of 84% on the test dataset.

3.6 Classification Report

A detailed classification report is provided, showcasing precision, recall, F1-score, and support metrics for each emotion class.

```
Classification report:
              precision    recall  f1-score   support

           0       0.79      0.80      0.79       307
           1       0.80      0.93      0.86       322
           2       0.85      0.83      0.84       340
           3       0.91      0.80      0.85       451

    accuracy                           0.84      1420
   macro avg       0.84      0.84      0.84      1420
weighted avg       0.84      0.84      0.84      1420
```

3.7 Comparison with Machine Learning Approach

Similar approaches such as sentence embedding using Sentence Transformers were attempted in the deep learning model, but they did not yield significant improvements in accuracy.

```
45/45 [==============================] - 2s 3ms/step
Classification report:
              precision    recall  f1-score   support

           0       0.65      0.64      0.64       307
           1       0.71      0.76      0.74       322
           2       0.72      0.72      0.72       340
           3       0.74      0.71      0.73       451

    accuracy                           0.71      1420
   macro avg       0.70      0.71      0.71      1420
weighted avg       0.71      0.71      0.71      1420
```

4. Conclusion

The Emotion Intensity Predictor project successfully develops machine learning and deep learning models for predicting the intensity of emotions in textual data. SVM performs well in the machine learning model, while the deep learning model achieves higher accuracy. These models offer valuable insights into emotional content, with potential applications in sentiment analysis, mental health monitoring, and customer feedback analysis.