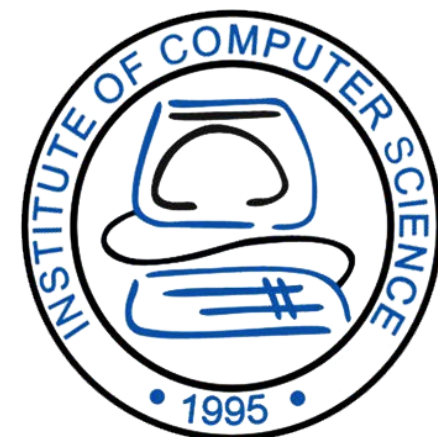# Using Emotion Word Intensities for Multi-label Emotion Detection of Text

Andrei Francis U. Enrique and Maria Art Antonette D. Clariño

Institute of Computer Science, University of the Philippines Los Baños
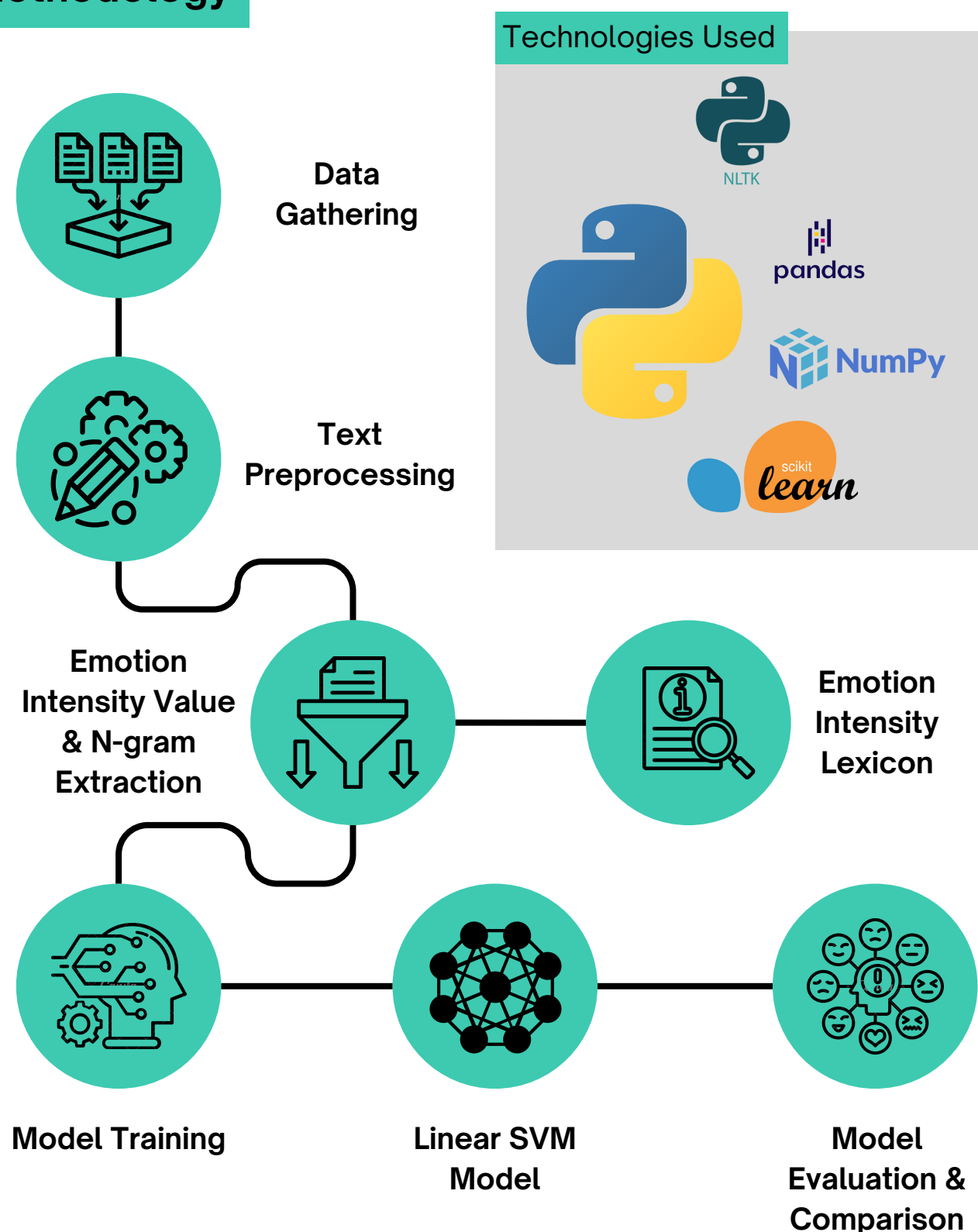
Abstract

Text-based emotion detection provides valuable insight into how emotions influence online behavior, in fields such as e-commerce and social media analysis. This study presents the use of **emotion word intensity values** derived from an emotion lexicon to create a linear **support vector machine classifier using the one-vs-rest strategy** for multi-labeled emotion data. The model was trained on a dataset of English subtitles and **compared to an n-gram-based classifier**. The results show a **significant improvement in classification performance** when emotion intensity values are included in the model. This suggests that the use of emotion word intensities can help classification models in identifying the emotions expressed within bodies of text.

Objectives

The general objective of this study is to utilize emotion word intensities to implement a classification model that can identify the emotions expressed by a body of text. The specific objectives are the following:

- Identify the relationship between the presence of emotion words in a body of text and the emotions expressed within said body of text
- Determine the importance of emotion word intensities in relation to other features used in classification models to identify the emotions expressed in text
- Evaluate the performance of the machine learning model in classifying multi-labeled emotion data using metrics such as per-label F1 score and macro F1 score, and test these metrics against a baseline model for statistical significance

Methodology

Technologies Used

- NLTK
- pandas
- NumPy
- scikit learn

- Data Gathering
- Text Preprocessing
- Emotion Intensity Value & N-gram Extraction
- Emotion Intensity Lexicon
- Model Training
- Linear SVM Model
- Model Evaluation & Comparison

Results

The inclusion of emotion word intensities generally improved the classification performance of the linear SVM model. Furthermore, emotion intensity values for disgust, fear, joy, and sadness showed very high feature importance for their respective emotion labels.
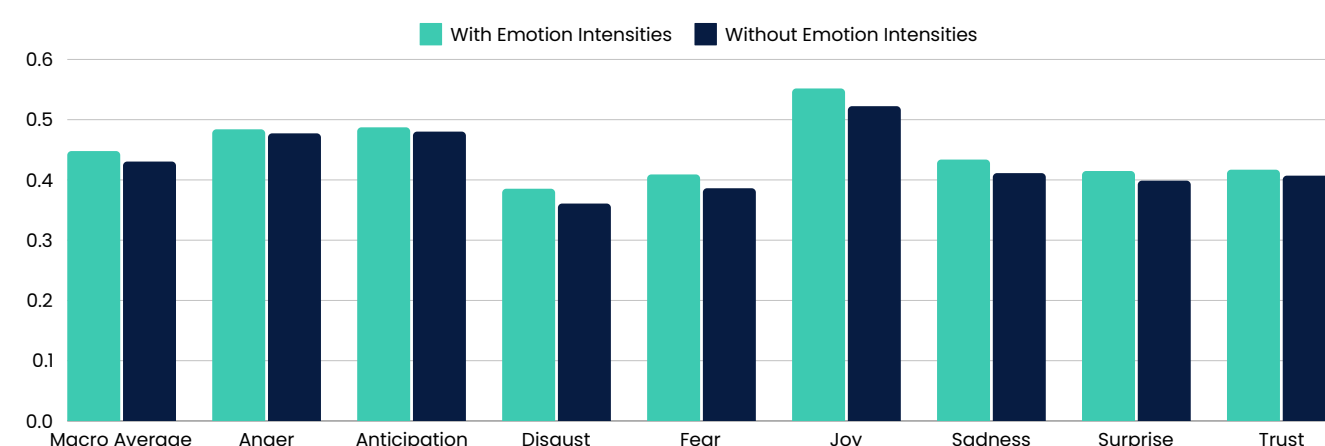

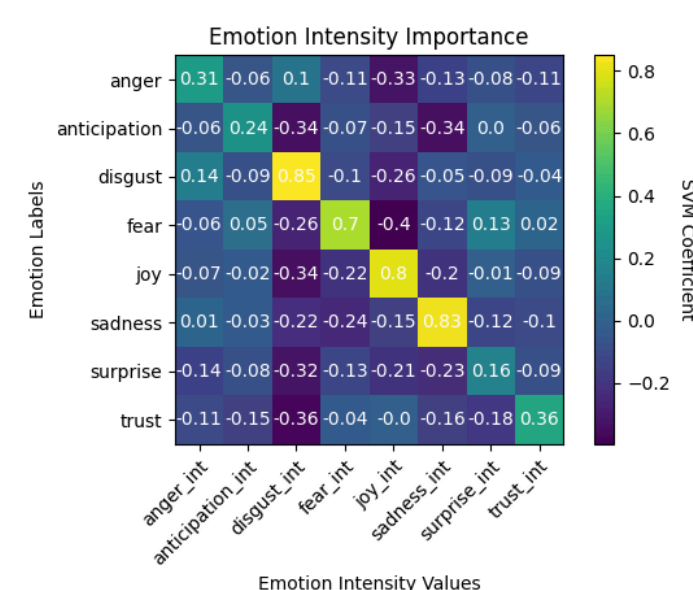**Fig 1: Comparison of F1 scores for each class**

Fig 2: Feature Importance of Emotion Intensity Values for each class

Performing a ten-fold stratified t-test showed that the improvement in performance observed in all classes except for anger and trust is statistically significant. The statistically significant classes are highlighted in **bold**:

Class	T-statistic	p-value
Macro F1	6.7505	8.3601e-5
Anger	1.9331	0.0852
Anticipation	2.6758	0.0254
Disgust	4.1034	0.0027
Fear	3.0154	0.0146
Joy	4.1379	0.0025
Sadness	4.0615	0.0028
Surprise	2.7542	0.0223
Trust	2.2045	0.0549

Table 1: T-test results for F1 scores of each class

Conclusion

The results show that emotion word intensities can be a valuable tool in identifying the emotions present in a body of text. In particular, words expressing disgust, fear, joy, and sadness were found to have high correlation with their respective emotion labels. Further improvements in classification performance can be made through an expanded lexicon or the extraction of additional important features.

About the Author

Andrei Francis U. Enrique is a BS Computer Science student from the University of the Philippines Los Baños, residing in Mandaluyong City. He has a keen interest in understanding how systems work, especially in the context of software and video game mechanics.