SP2 Presentation

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

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

21 May 2024

Background of the Study

Emotions

Emotions are an active part of daily life, and can influence human perceptions, opinions, and decision making.

Sentiment & Emotion Analysis

A wealth of textual data is present on the internet in many different forms. Sentiment analysis has been developed in large part to extract valuable insight from this freely accessible data.

Emotion Detection

Meant to expand upon the usual tasks of sentiment analysis such as determining polarity and subjectivity, emotion detection is concerned with associating a body of text with a range of emotions.

Approaches

Machine learning, and particularly deep learning-based approaches are currently the most accurate approaches to emotion detection. The use of lexicons and keyword lists are still popular for analysis of text.

Objectives

- Use emotion word intensities to identify the emotions expressed by a body of text.

Specific Objective 1

- Identify the relationship between the presence of emotion words in a body of text and the emotions expressed within said body of text

Specific Objective 2

- Determine the importance of emotion word intensities in relation to other features used to identify emotions in text.

Specific Objective 3

- Evaluate the performance of a machine learning model that uses emotion word intensities in classifying multi-labeled emotion data.

Dataset

XED Dataset

- Movie Subtitles
- Annotated according to text's association with emotions
- Multilabeled and Multilingual

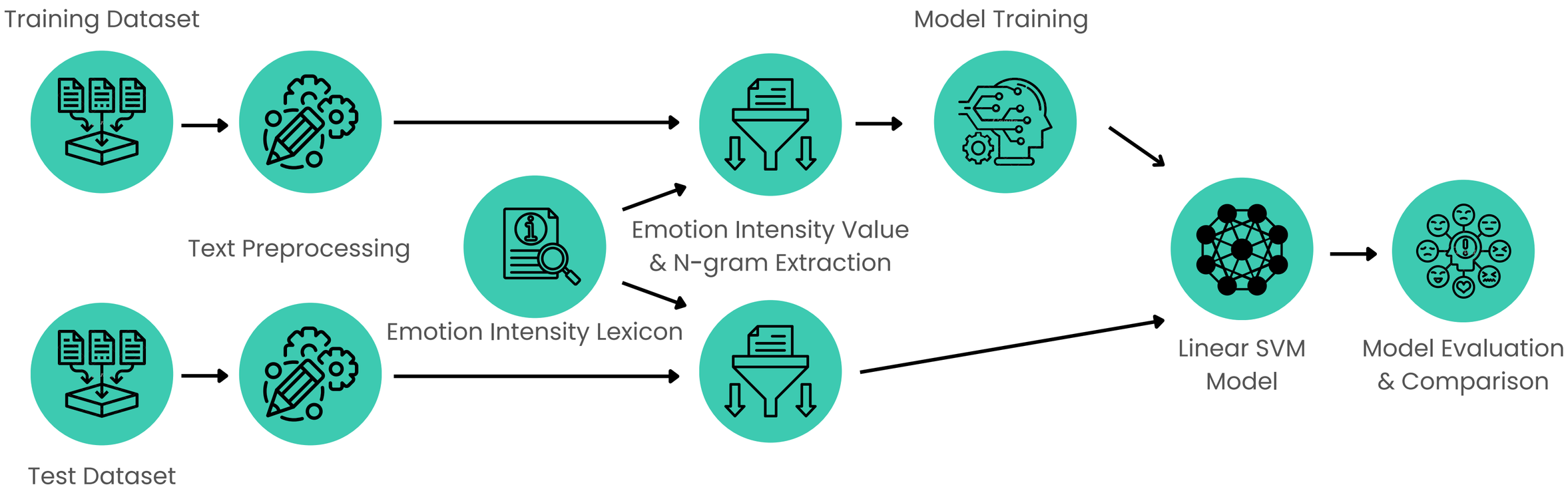
Text	Labels
Well, I really put my foot in my mouth this time but, I don't know, I like you two.	8, 3, 5

Emotion Values			
anger	0	joy	1
anticipation	0	sadness	0
disgust	1	surprise	0
fear	0	trust	1

Emotion Word Intensities

- Derived from NRC Emotion Intensity Lexicon
- Words scored according to their degree of association with one or more distinct emotion labels
- Higher value indicates a stronger association

Word	Emotion	Score
elated	joy	0.894
pleased	joy	0.725
sweet	joy	0.625
calm	joy	0.394
silence	joy	0.127



Methodology

Classification Model

- Linear SVM
- One-vs-rest strategy
- Unigrams, bigrams, trigrams as baseline features
- 90:10 stratified train-test split

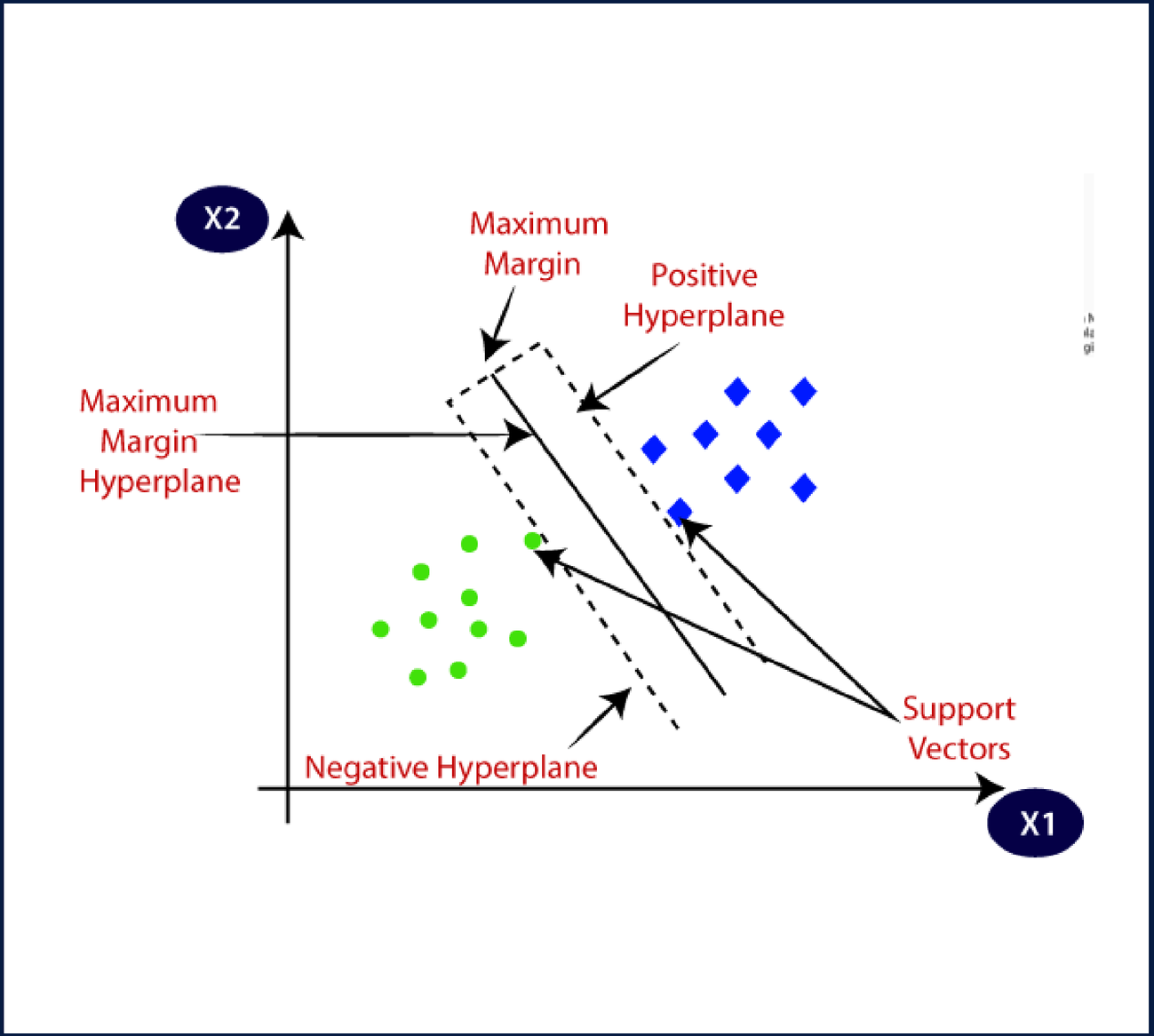


Image Source: https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/

Feature Extraction

Emotion Word Intensities

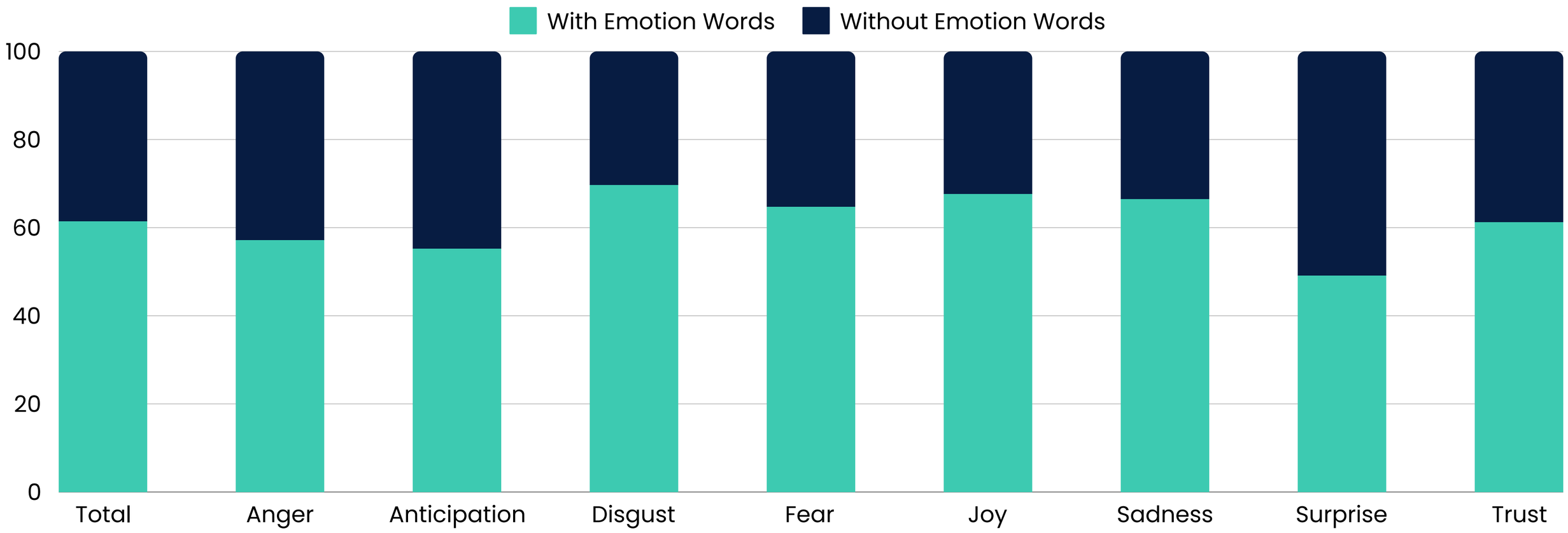
['we', 'must', 'protect', 'these', 'investments', 'so', 'the', 'area', 'can', 'continue', 'to', 'prosper', 'and', 'grow']

continue	
anticipation	0.414
trust	0.422

prosper	
anticipation	0.578
joy	0.562

grow	
anticipation	0.555
joy	0.242
trust	0.492

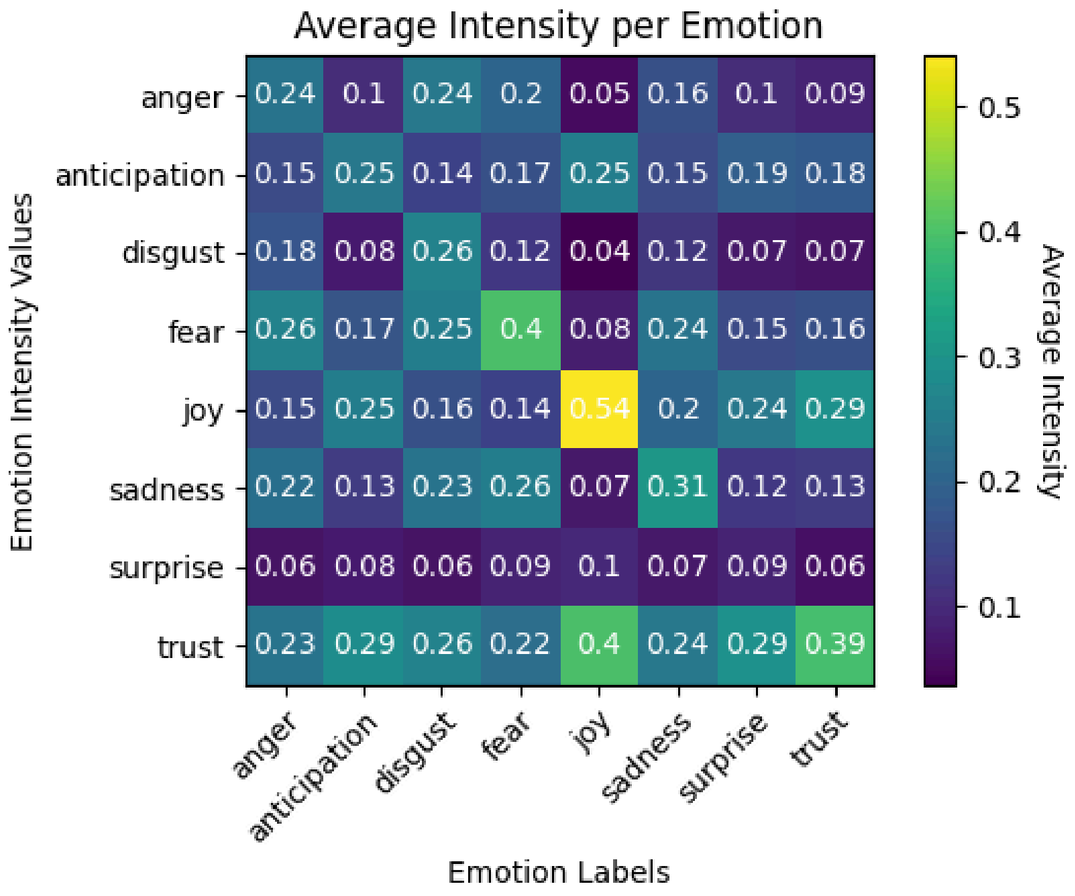
text total			
anger	0	joy	0.804
anticipation	1.547	sadness	0
disgust	0	surprise	0
fear	0	trust	0.914

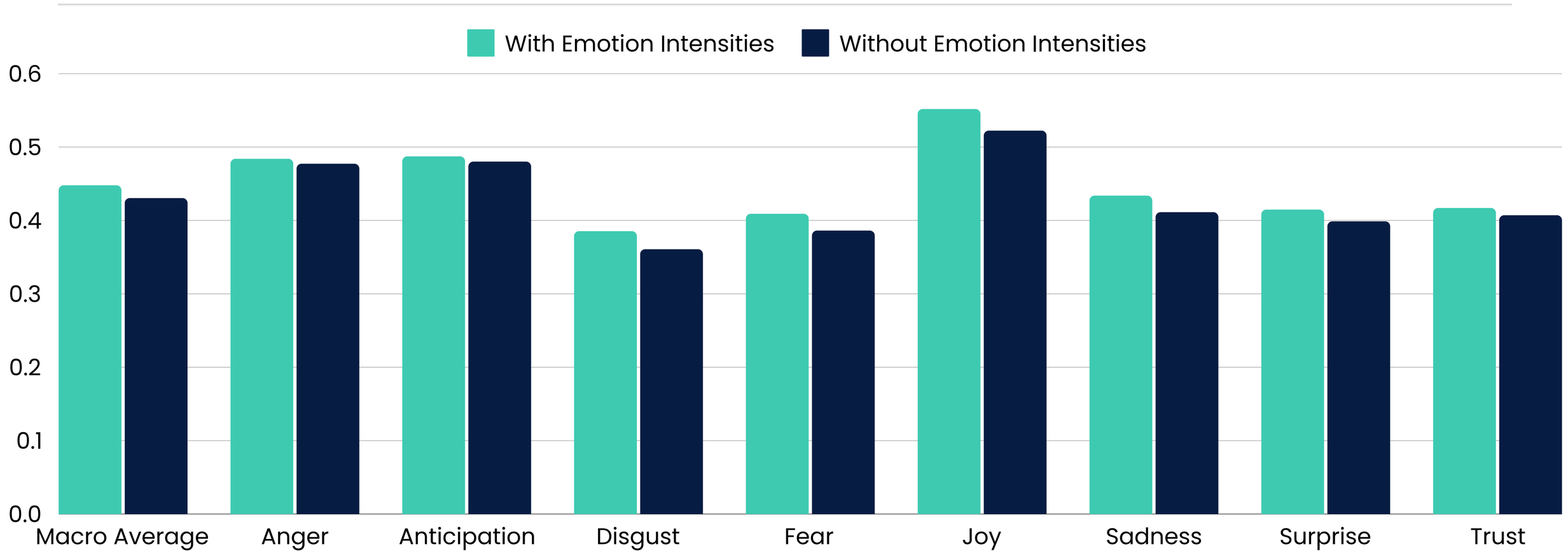


Percentage of Sentences with Emotion Words

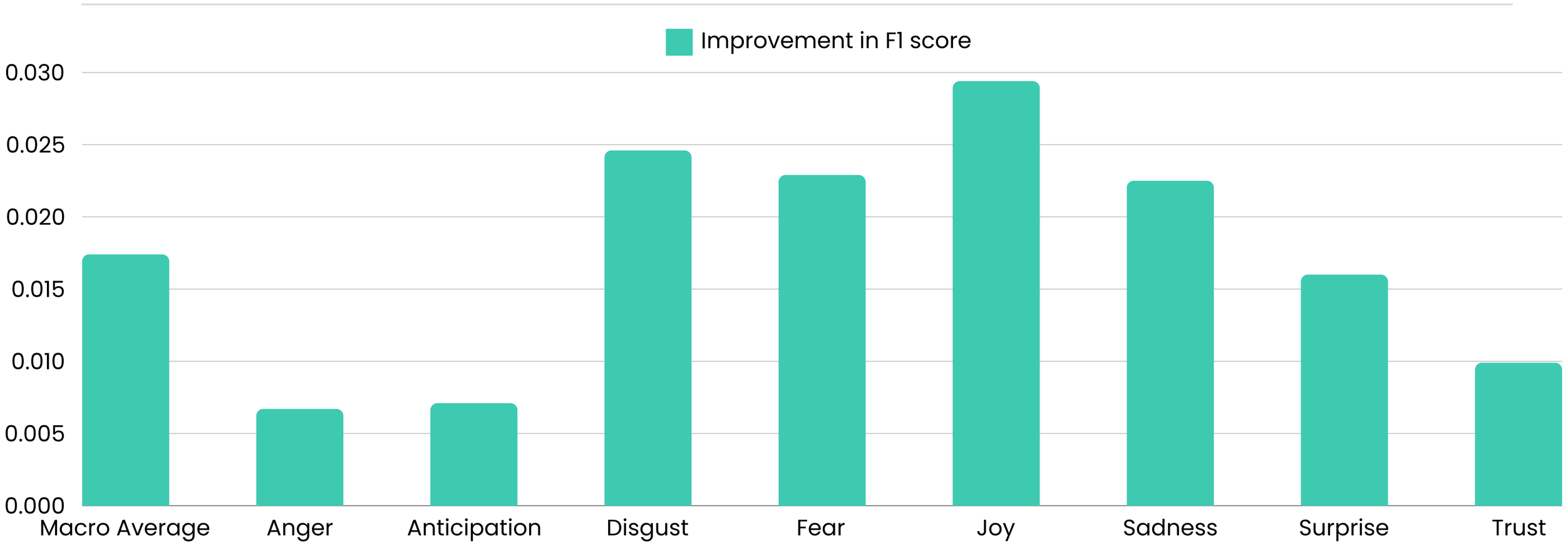
Average Emotion Intensity Values

- Intensity values for Joy and Trust appear frequently across all labels
- Intensity values for surprise are rarer
- Intensity values generally appear most frequently in their respective emotion label




With Emotion Intensities
Without Emotion Intensities
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Macro Average
Anger
Anticipation
Disgust
Fear
Joy
Sadness
Surprise
Trust

F1 Scores



Score Improvement per Class

Feature Importance

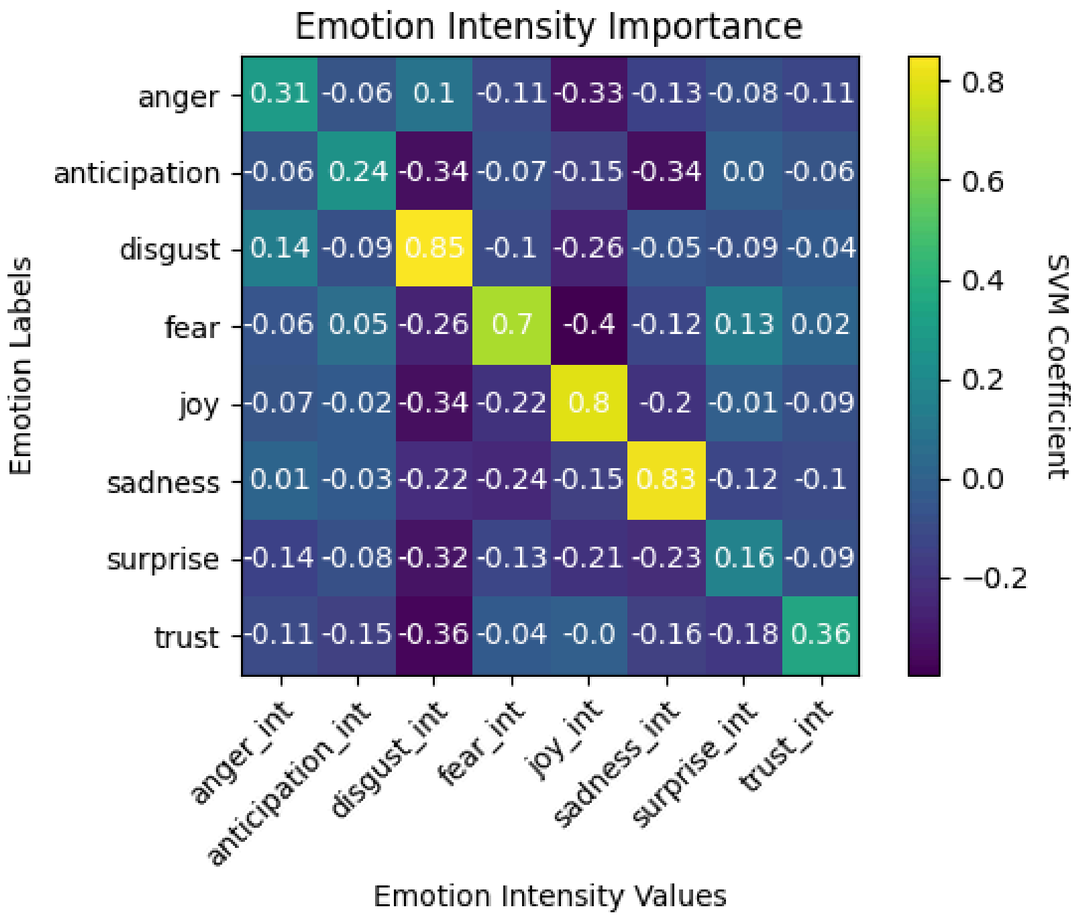
N-grams

Table 7: Five highest positive coefficient n-grams for each label (model with emotion intensities)

Anger		Anticipation		Disgust		Fear		Joy		Sadness		Surprise		Trust	
n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient
'fuck'	0.60	'ill'	0.59	'stupid'	0.51	'troubl'	0.41	'nice'	0.73	'sorri'	0.68	'thought'	0.43	'worri'	0.46
'shut'	0.37	'readi'	0.38	'drunk'	0.39	'uh'	0.36	'great'	0.59	'poor'	0.58	'dont know'	0.37	'ill'	0.44
'stop'	0.36	'tonight'	0.36	'bullshit'	0.32	'kill'	0.35	'thank'	0.48	'cant'	0.49	'oh'	0.35	'thank'	0.39
'goddamn'	0.35	'let'	0.34	'mother'	0.29	'god'	0.33	'fine'	0.46	'never'	0.46	'whi'	0.35	'dont worri'	0.36
'bullshit'	0.35	'hell'	0.33	'peopl'	0.27	'pleas'	0.33	'love'	0.31	'didnt'	0.34	'surpris'	0.35	'help'	0.33

Table 8: Five highest positive coefficient n-grams for each label (model without emotion intensities)

Anger		Anticipation		Disgust		Fear		Joy		Sadness		Surprise		Trust	
n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient	n-gram	coefficient
'fuck'	0.62	'readi'	0.46	'stupid'	0.51	'kill'	0.55	'love'	0.77	'sorri'	0.66	'thought'	0.43	'thank'	0.40
'hate'	0.39	'tonight'	0.36	'bastard'	0.47	'danger'	0.50	'nice'	0.70	'die'	0.56	'surpris'	0.39	'worri'	0.38
'damn'	0.39	'come'	0.36	'lie'	0.45	'scare'	0.50	'beauti'	0.60	'poor'	0.56	'dont know'	0.38	'promis'	0.35
'shut'	0.38	'let'	0.35	'fool'	0.45	'troubl'	0.41	'wonder'	0.59	'cant'	0.48	'oh'	0.36	'dont worri'	0.32
'stop'	0.38	'wait'	0.35	'shit'	0.42	'afraid'	0.40	'great'	0.59	'cri'	0.44	'what'	0.36	'right'	0.32



Feature Importance

Emotion Intensities

Statistical Significance

Ten-fold stratified t-test

- t-value - 2.262
- p-value - 0.05

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

Conclusions

Emotion Word Intensities

The addition of emotion word intensities was able to significantly improve the performance of the classification model.

Feature Importance

Intensity values for disgust, joy, sadness, and fear displayed very high correlation with their respective classes.

Classification Model

The Linear SVM classifier performed well at identifying sentences with joy; more work can be done in improving the identification of other classes.

Recommendations

Emotion Intensity Lexicon

An expanded emotion intensity lexicon can improve the performance of the classifcation model.

Grammar-based modifiers for Emotion Intensities

Derived emotion intensity values may be further influenced by positive adjectives or negative words.

Other Features of Interest

- Emojis
- Special characters
- Figures of Speech

References

- Alzoubi, H. M., & Aziz, R. (2021). Does emotional intelligence contribute to quality of strategic decisions? The mediating role of open innovation. Journal of Open Innovation: Technology, Market, and Complexity, 7(2), 130.
- Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Vora, D., & Pappas, I. (2022). A Review on Text-Based Emotion Detection--Techniques, Applications, Datasets, and Future Directions. arXiv preprint arXiv:2205.03235.
- Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. International Journal of Research in Marketing, 36(1), 20-38
- Mohammad, S. M. (2017). Word affect intensities. arXiv preprint arXiv:1704.08798.
- Öhman, E., Pàmies, M., Kajava, K., & Tiedemann, J. (2020). Xed: A multilingual dataset for sentiment analysis and emotion detection. arXiv preprint arXiv:2011.01612.

Thank you for Listening

Andrei Francis Enrique

21 May 2024