Tweet Sentiment Analysis Using Machine Learning and Neural Networks

Analysis & Recommendations about predicting tweet sentiment

By Charlotte Sun

Overview

What emotions do people express the most on Twitter?

Can we predict how people feel from just their words?

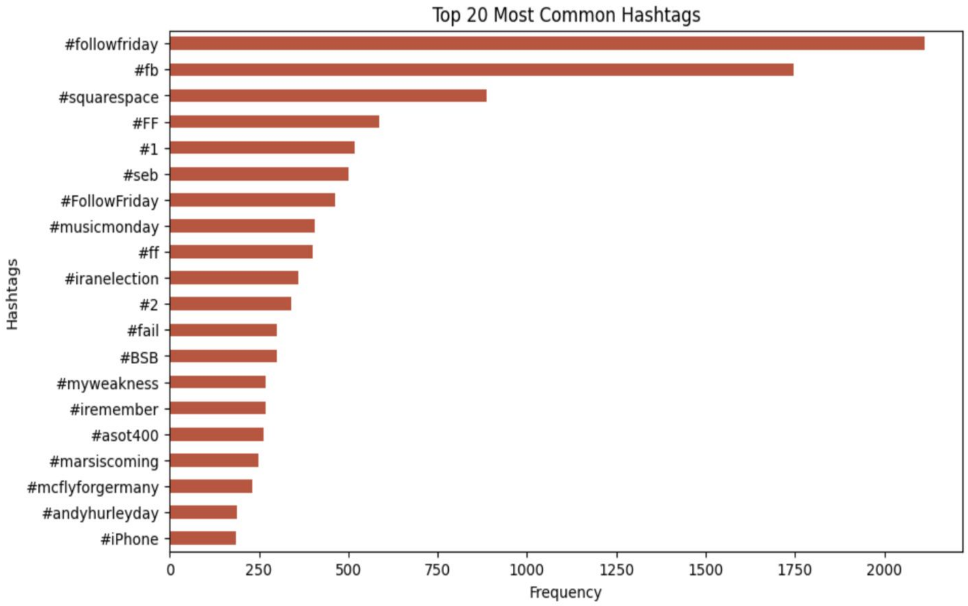
How do machine learning and neural networks help us understand social media sentiment?

"my whole body feels itchy and like its on fire" – negative example

"@Roy_Everitt ha- good job. that's right." - positive example

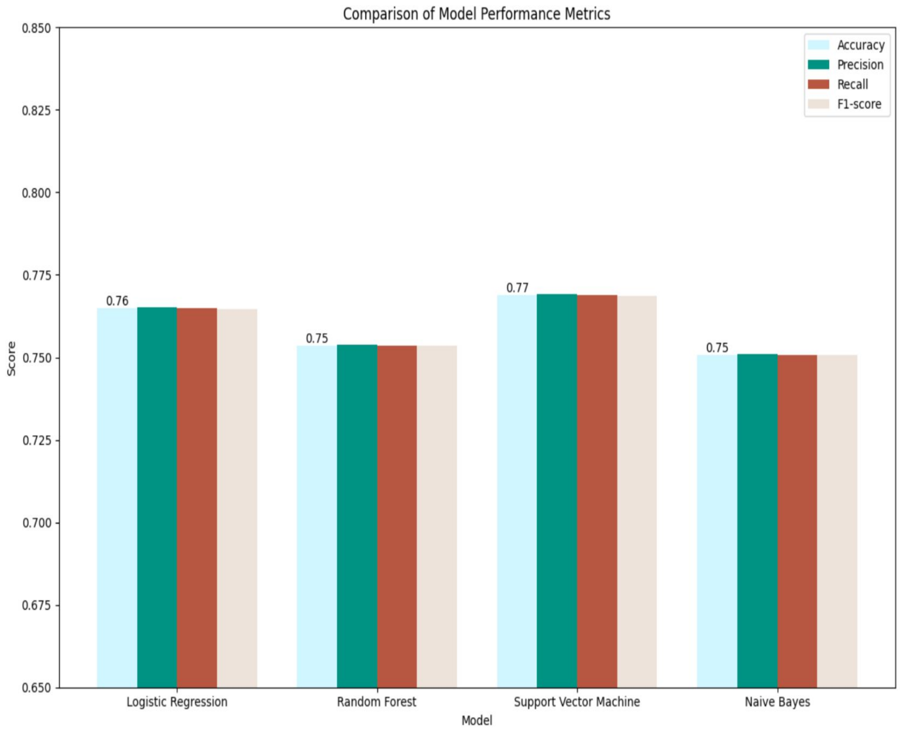
What are the most common hashtags used in the tweets?

From these hashtags, we can see that Twitter users' primary interests on the platform revolve around social interaction (e.g., #followfriday), pop culture (e.g., #musicmonday), major global events (e.g., #iranelection), brand discussions (e.g., #iPhone), and emotional expression (e.g., #fail).



What models are used?

Evaluate the performance of four classifiers: **Logistic Regression, Random Forest, Support Vector Machine (SVM), and Naive Bayes**. The results, as shown in the bar plot, clearly indicate that **SVM** and **Logistic Regression** performed the best in terms of both accuracy and F1 score.



Best Result : SVM

- **ROC Curve**: The model shows good performance with an AUC of 0.85, meaning it effectively distinguishes between positive and negative sentiment.

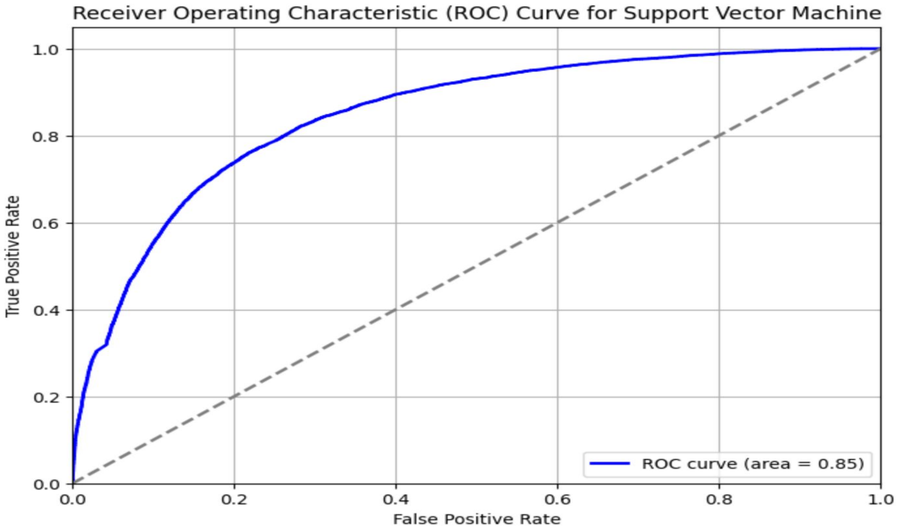
- **Confusion Matrix**: The model correctly predicted:

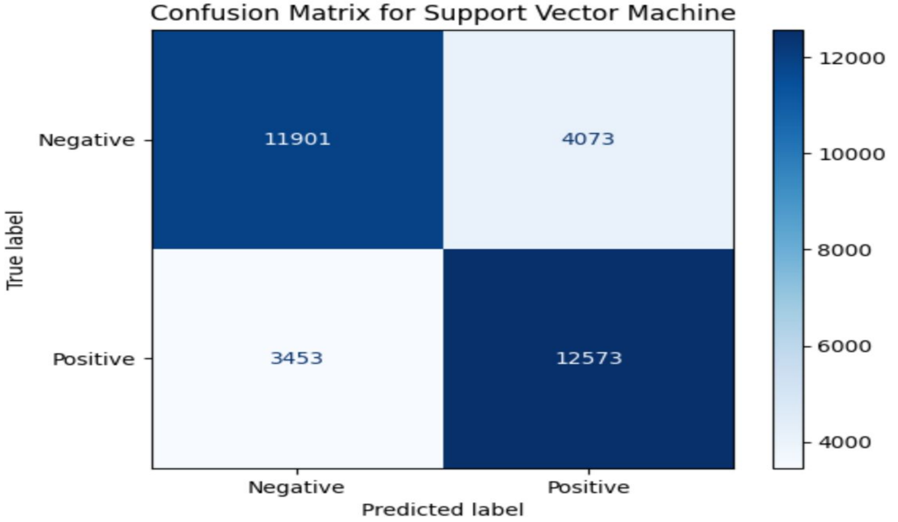
  - 11,901 negative cases

  - 12,573 positive cases

Final accuracy: 0.77

The SVM model performs well, balancing precision and recall.





This diagram shows the architecture of my TextCNN model:

1. Input Layer: Converts text sequences into embeddings.

2. Conv1D Layer: Uses 64 filters and L2 regularization to detect features.

3. GlobalMaxPooling1D Layer: Reduces dimensionality by capturing key features.

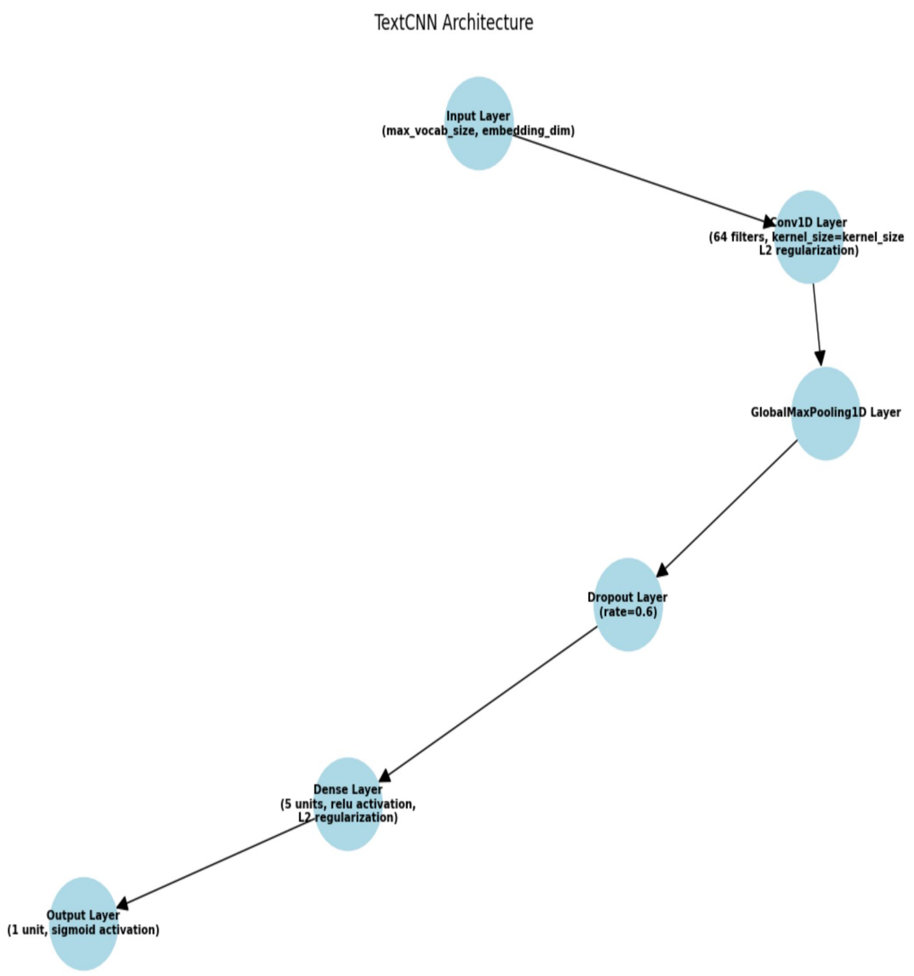
4. Dropout Layer: 60% dropout rate to prevent overfitting.

5. Dense Layer: 5 units, with ReLU activation and L2 regularization.

6. Output Layer: A single unit with sigmoid activation for binary classification.

The model is designed to balance feature extraction and regularization to prevent overfitting.

Textcnn Neural Network



Tune the model

Final accuracy: 0.77

Major methods to tune:

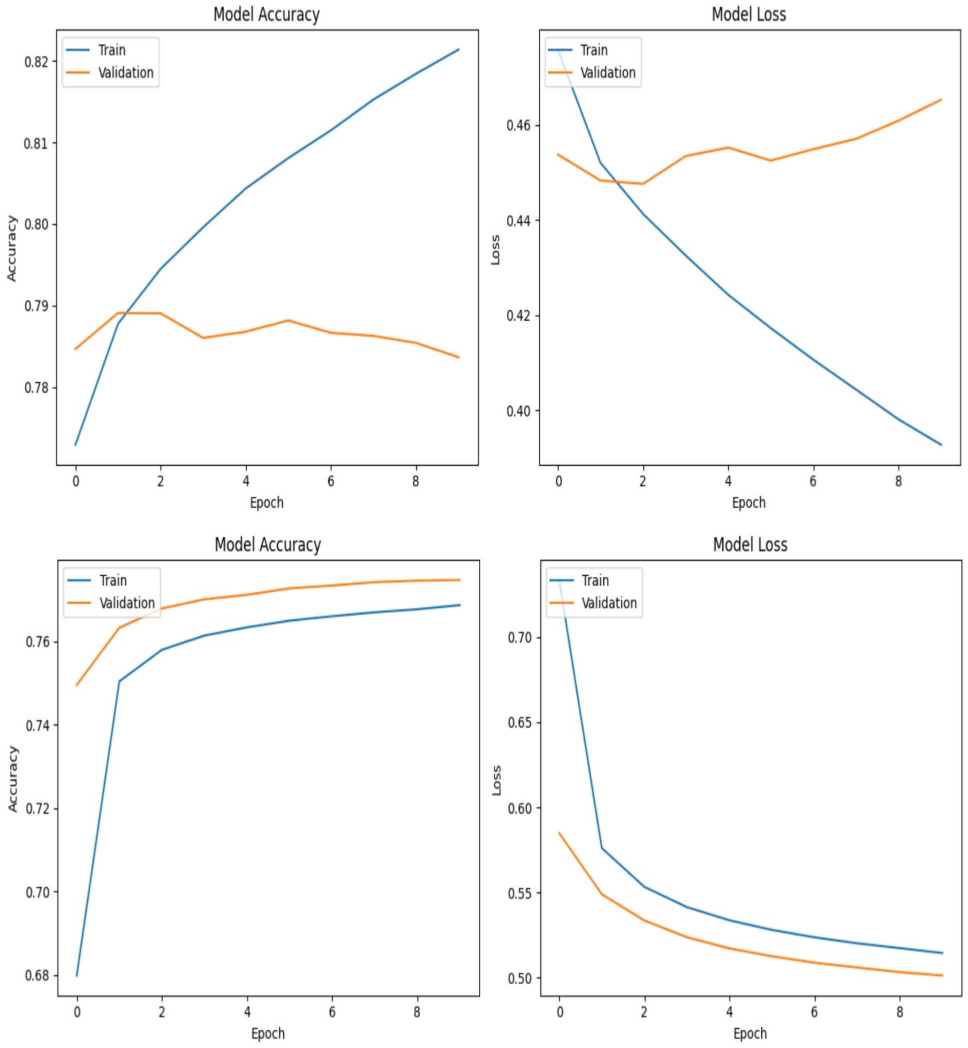
Dropout: Increased dropout rates to 0.6 to prevent overfitting.

L2 Regularization: Applied L2 regularization in the convolutional and dense layers to reduce model complexity.

Reduce Learning Rate on Plateau: Used the ReduceLROnPlateau method to dynamically lower the learning rate when validation loss stopped improving.

Batch Size: Tuned batch sizes, which helped stabilize training.

Early Stopping: Employed early stopping to halt training when validation performance no longer improved.



Summary & Recommendation

Summary: In this project, sentiment analysis was conducted using both machine learning and deep learning techniques. Support Vector Machines (SVM) proved to be the best-performing classical model with an accuracy of 77%. The TextCNN neural network was fine-tuned with dropout and regularization to improve generalization and reduce overfitting.

Business Application: The model can help businesses analyze customer sentiment on social media in real time. By understanding customer opinions, companies can improve their products and services and make more informed marketing decisions.