# SENTIMENT ANALYSIS

## 20.000

Different expression for different energy level

### Negative Energy Vibe

Totally dissatisfied with the service. Worst customer care ever

BRB | OOT | YOLO

### Neutral Energy Vibe

Good job but I will expect a lot more in the future

HPY | NICE | VIBE

### Positive Energy Vibe

Brilliant effort guys! Loved your work

HIGH | YOU | FLY

# Business Problem & Objectives

- Problem Statement:
- Companies struggle to analyze vast amounts of customer feedback on social media.

- Objectives:
- 1. Automate sentiment detection (positive, negative, neutral).
- 2. Enhance brand reputation management.
- 3. Optimize marketing strategies based on sentiment insights.

# Data Collection & Preprocessing

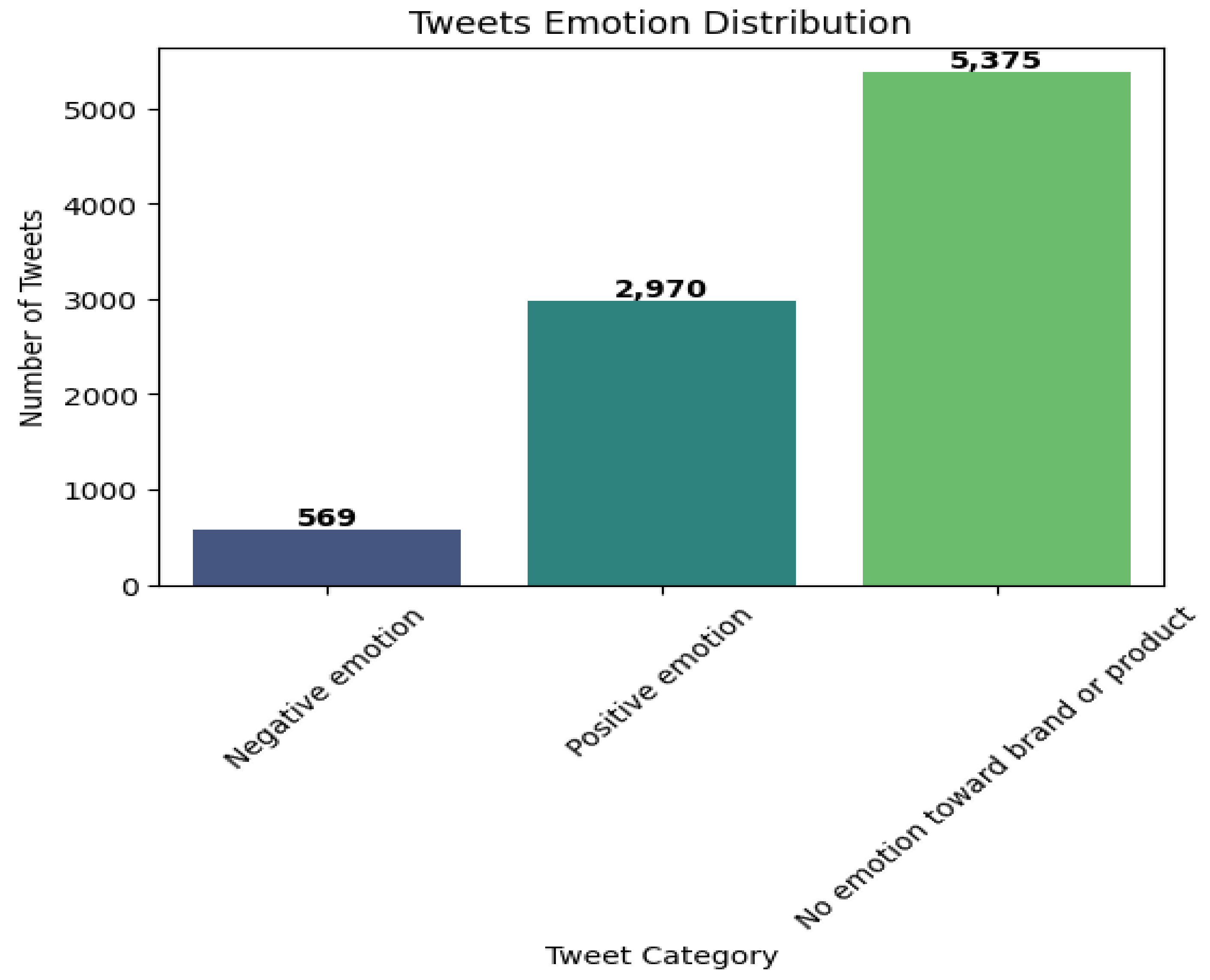
- Data Source: Tweets dataset with sentiment labels.

(https://data.world/crowdflower/brands-and-product-emotions)

- Preprocessing Steps:
- - Tokenization, Stopword Removal, Lemmatization
- - TF-IDF Vectorization & Word Embeddings (GloVe)

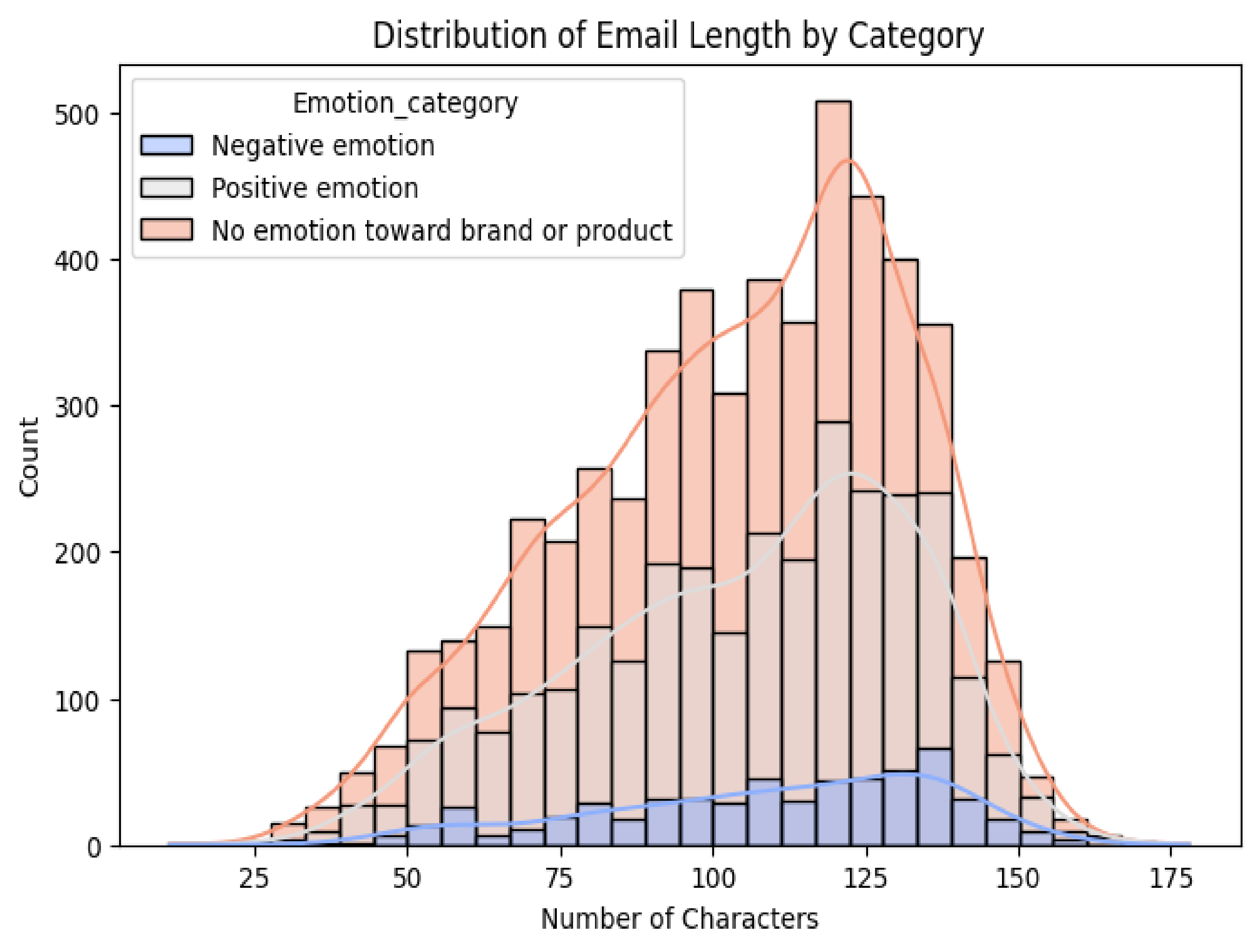
- Exploratory Analysis:
- - Word clouds, n-grams, sentiment distribution.

# Exploratory Data Analysis (EDA)




- No emotion sentiments : 5,375(Majority), followed by positive emotions92,970) and Negative emotions:569(least)

# Distribution of tweets by Emotion category

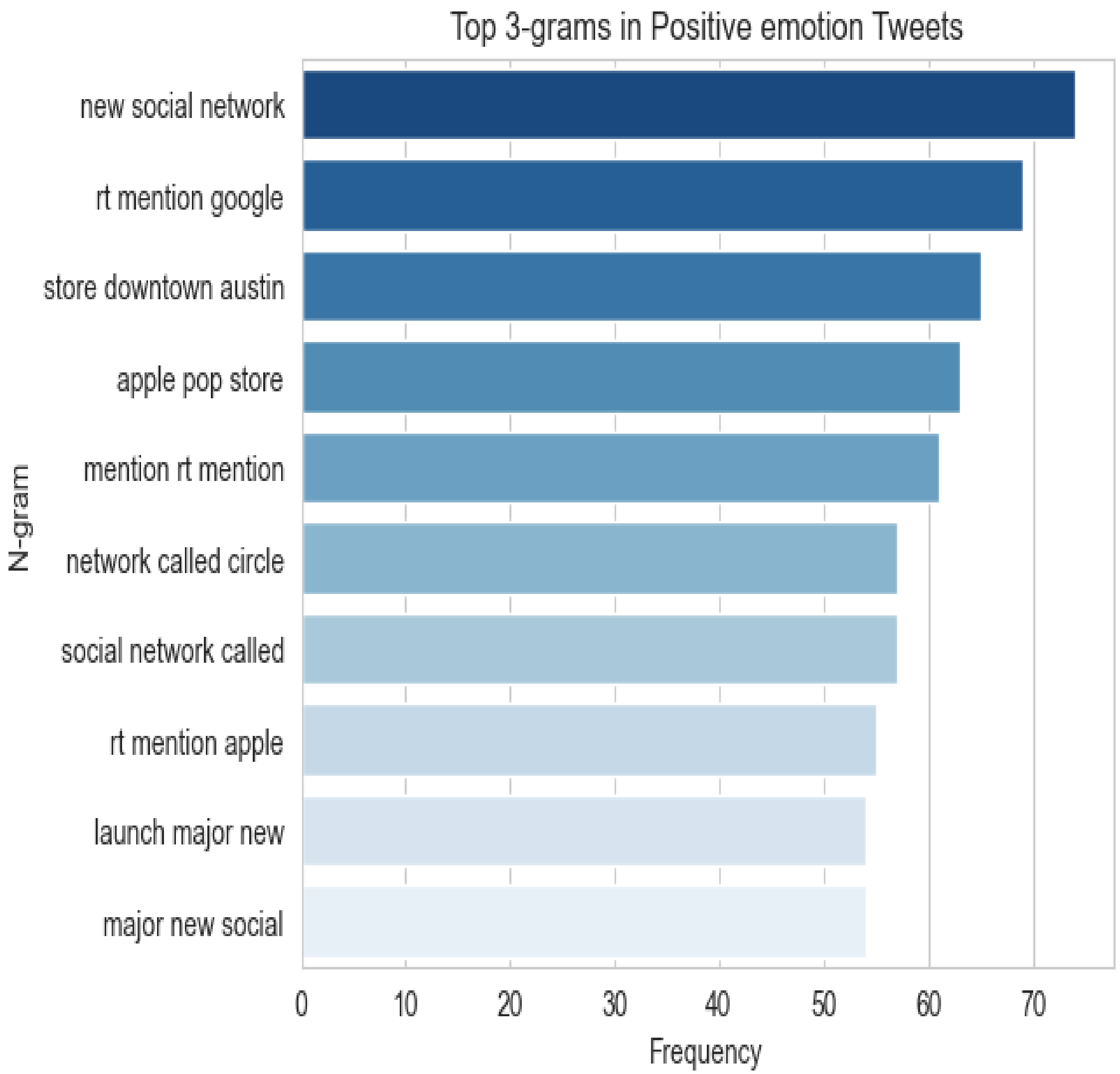


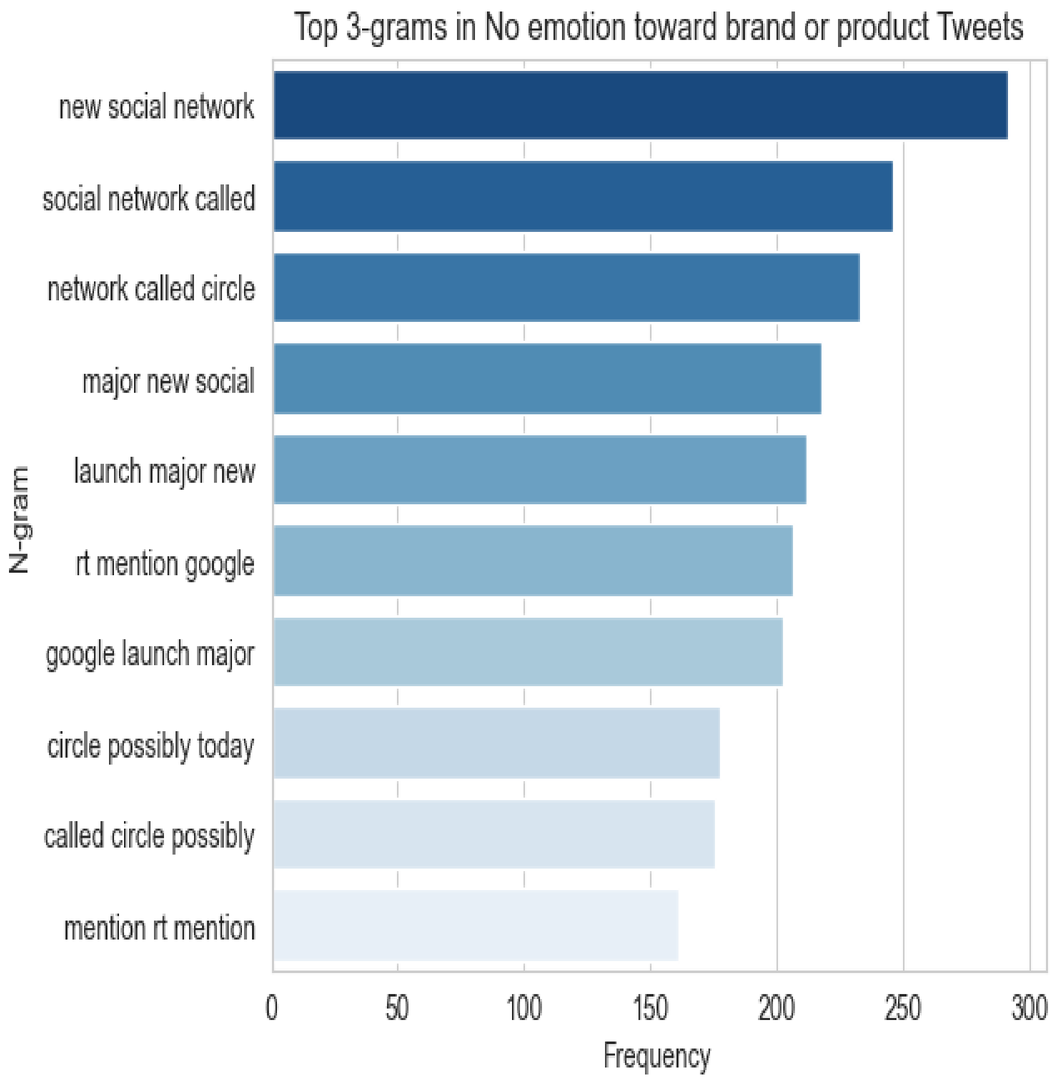

•**Tweet Length**: Most are **75–150 characters**.
•**Neutral Tweets**: Most frequent across all lengths.
•**Positive & Negative**: Similar distribution, but negatives are fewer.
•**Peak**: **"No emotion"** tweets peak at **125 characters**.
•**Distribution**: Roughly **normal**, cantered around **75–150 characters**.

# Wordcloud Analysis

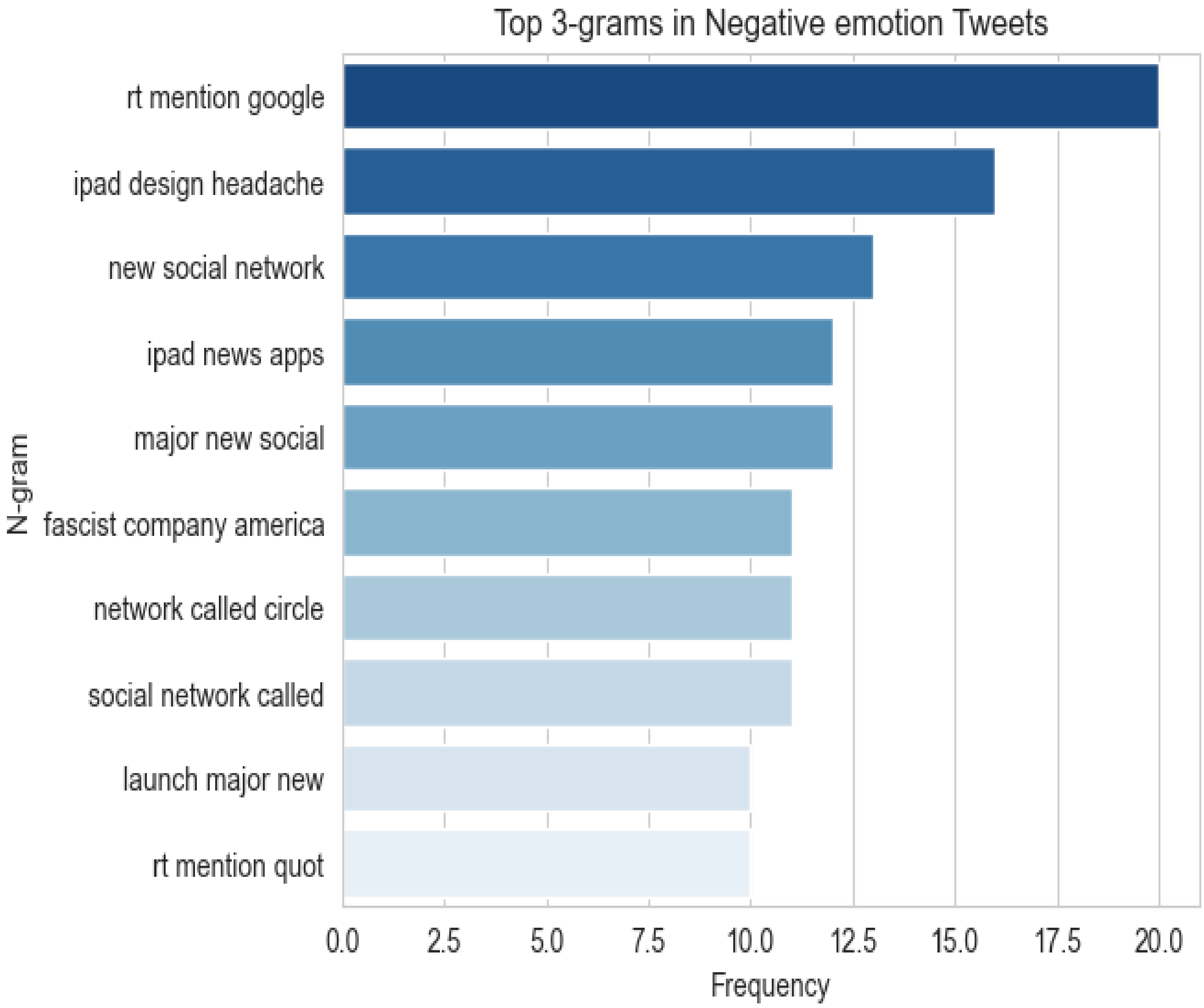

- **Top Words:** "SXSW," "mention," and "link" appear most across all sentiments.
- **Negative Tweets:** Words like "fail," "need," and "headache" show frustration.
- **Positive Tweets:** Terms like "great," "awesome," and "love" indicate enthusiasm.
- **Neutral Tweets:** Mostly event-related words, suggesting objective discussions

# N-grams Tweet category analysis



# Cont.




- **"RT mention"** dominates all sentiments.
- **Negative**: Frustration – "ipad design headache," "fascist company America."
- **Positive**: Excitement – "apple pop store," "store downtown Austin."
- **Neutral**: Info-driven – "new social network," "network called circle."
- **Trigrams**: Neutral = facts, Positive/Negative = emotions.

# Model Selection & Training

**Models Used:**

- Base Model – Logistic Regression for binary classification
- -Machine Learning: SVM, Naïve Bayes, KNN.
- - Deep Learning: **RNN (Recurrent Neural Network)** -LSTM with Pre-trained GloVe Embeddings.

**Training Process:**

- - Data split into Train & Test sets.
- - Hyperparameter tuning via GridSearchCV, random search

# Results & Evaluation

- **Metrics:** Accuracy, Confusion Matrix, F1-Score.

- **Best Performing Model:** [Model Name] with [Accuracy%].

**Insights:**

- - Positive tweets dominate, few neutral sentiments.
- - Misclassification mainly in no motion tweets vs. negative.

# RESULTS: BINARY CLASSIFICATION- LOGISTIC REGRESSION MODEL

Accuracy: 0.9090909090909091

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Negative emotion | 0.87 | 0.95 | 0.91 | 575 |
| Positive emotion | 0.95 | 0.87 | 0.91 | 613 |
| accuracy | | | 0.91 | 1188 |
| macro avg | 0.91 | 0.91 | 0.91 | 1188 |
| weighted avg | 0.91 | 0.91 | 0.91 | 1188 |

•**Accuracy**: 90.9%
•**Negative**: Precision 87%, Recall 95% (more false positives)
•**Positive**: Precision 95%, Recall 87% (more false negatives)
•**F1-score**: 0.91 → Strong overall classification

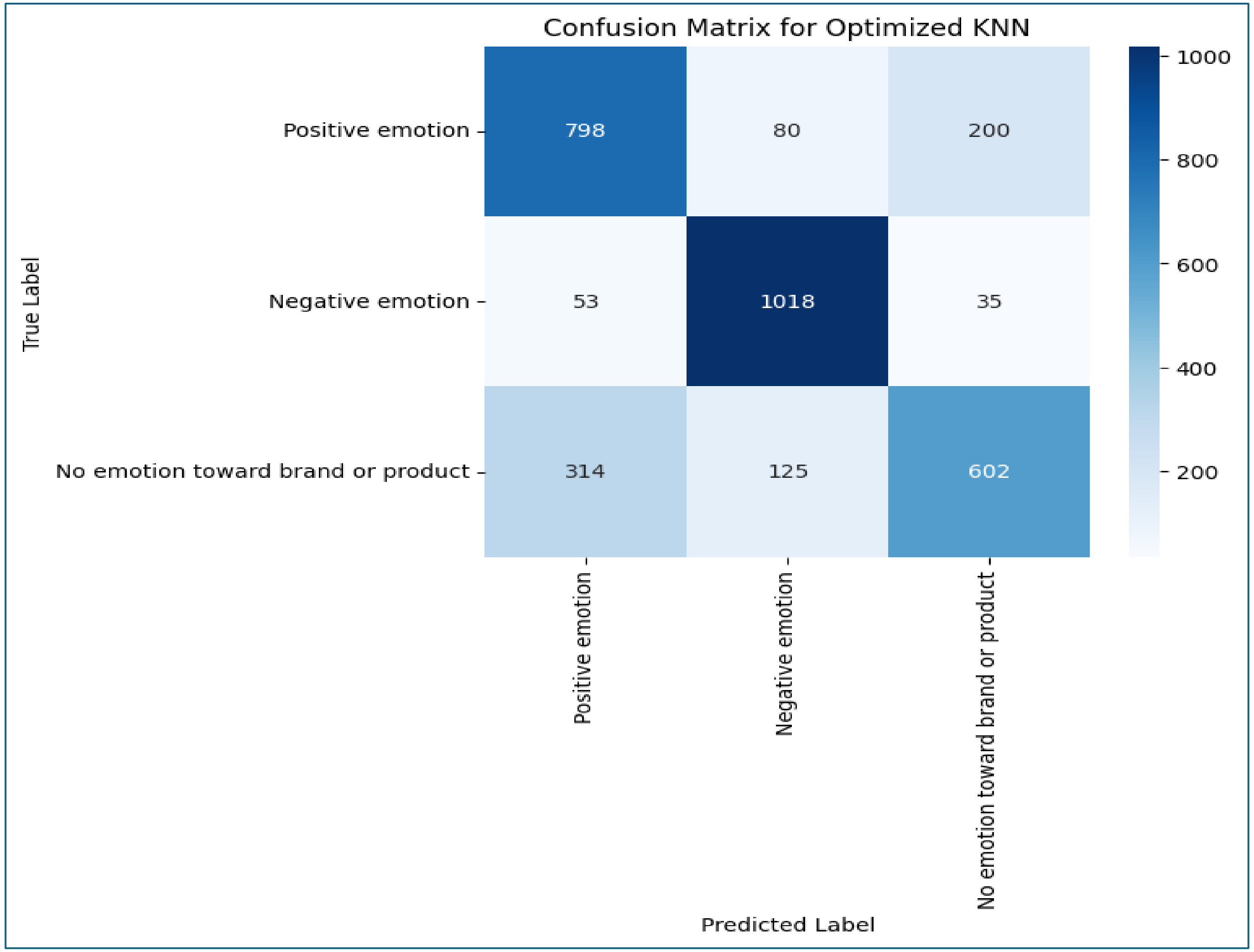
# OTHER CLASSIFIERS (SVS,KNN,NAÏVE BAYES)

•**SVC: 95% accuracy** – Best performer.
•**KNN: Low recall (35%) for Positive** → Poor balance.
•**Naïve Bayes: 90% accuracy** – Good, but SVC is better.
•**SVC ensures** strong generalization & class balance.

# MULTICLASSIFIERS(KNN,SVC,NAIVE BAYES)

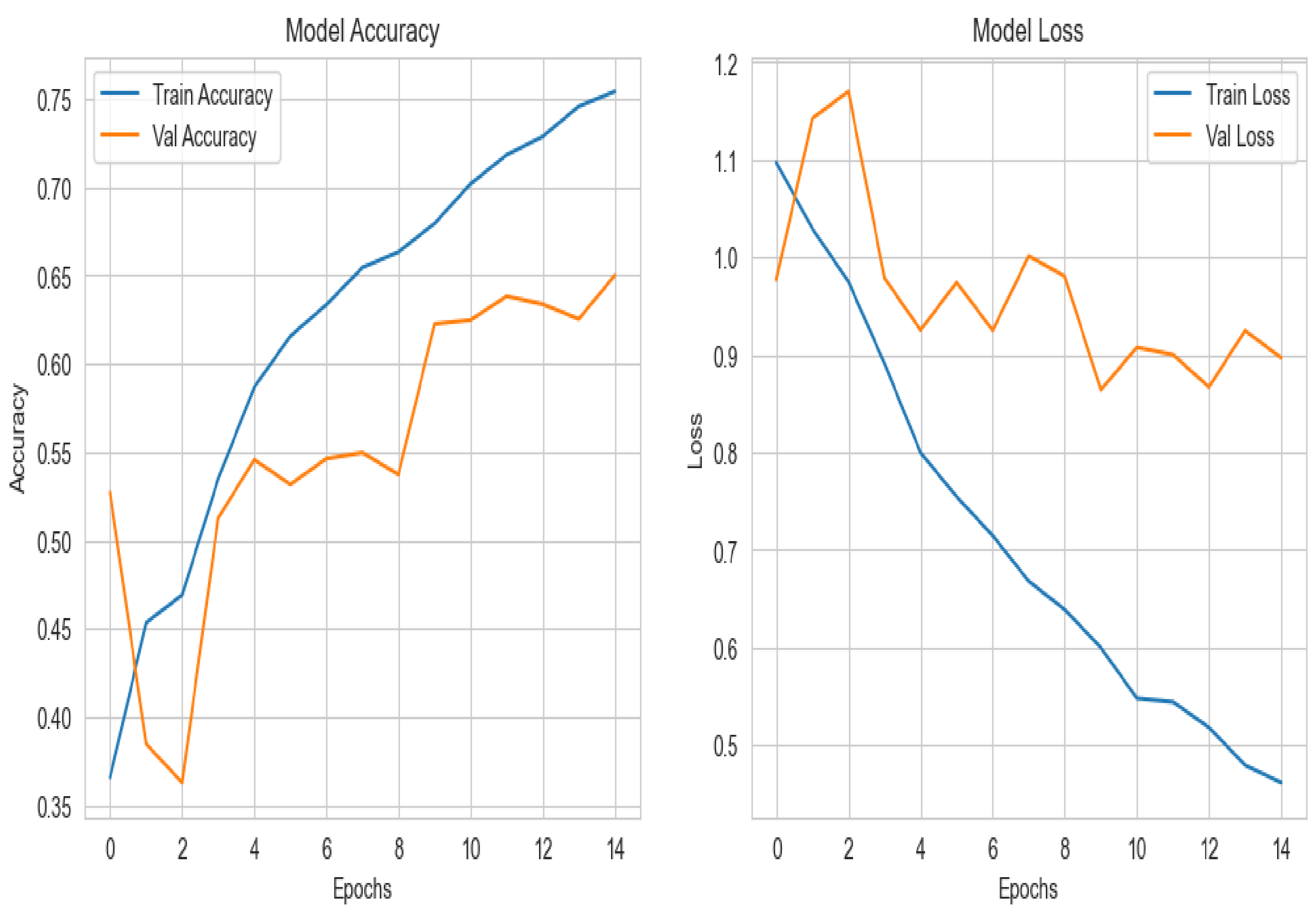
- **KNN: Best (74% accuracy)** → Outperformed SVC (57%) & Naïve Bayes (54%).
- **SVC struggled** with imbalance, even after SMOTE.
- **Naïve Bayes underperformed** due to feature independence issues.
- **KNN captures patterns better**, but tuning may improve results.

# TUNED BEST MODEL(KNN)- CONFUSION MATRIX



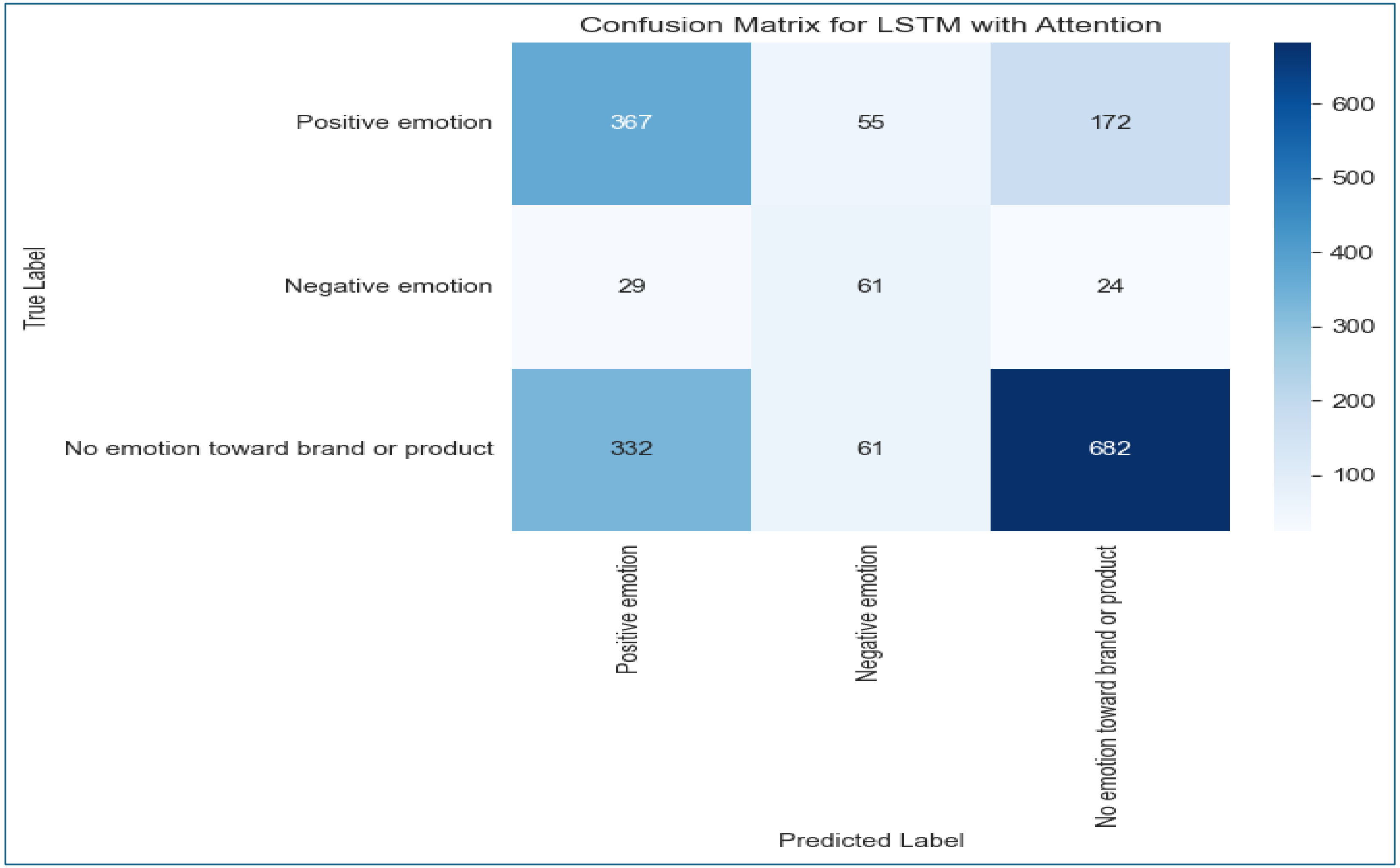

- **Optimized KNN: 75% accuracy** (↑ from 74%).
- **Best params:** n_neighbors=3, metric='euclidean', weights='distance'.
- **Negative emotions: Best classified (87% F1-score).**

# DEEP LEARNING FOR SENTIMENT ANALYSIS




- **Accuracy**: Improved from **33.9%** → **75.3%** (epoch 15).
- **Validation Accuracy**: Peaked at **65.0%**.
- **Learning Rate**: Adjustments at **epochs 5 & 14** boosted performance.
- **Class Performance**:
- **Positive**: Good recall (**62%**), moderate precision (**50%**).
- **Negative**: Weak detection (**34% precision, 54% recall**) – needs improvement.
- **No Emotion**: High precision (**78%**), lower recall (**63%**).

# Confusion matrix



# Conclusions

- The automation of sentiment classification reduces manual effort and enables large-scale tweet analysis.
- The optimized KNN model, achieving 75% accuracy, outperforms the deep learning model at 62% accuracy, with well-balanced F1-scores across all classes.
- Given its superior performance, KNN is the preferred choice for this classification task.

# Recommendations

- ✅ Adopt KNN for Sentiment Analysis – Achieves 75% accuracy, outperforming deep learning.
- ✅ Monitor Negative Sentiment in Real Time – 92% recall enables proactive brand management.
- ✅ Leverage Sentiment Insights for Marketing – Optimize ads and messaging based on trends.
- ✅ Automate Sentiment Analysis – Reduces manual effort and enhances decision-making.
- ✅ Next Steps – Deploy KNN, re-evaluate models, and expand data sources.

# END

THANK YOU