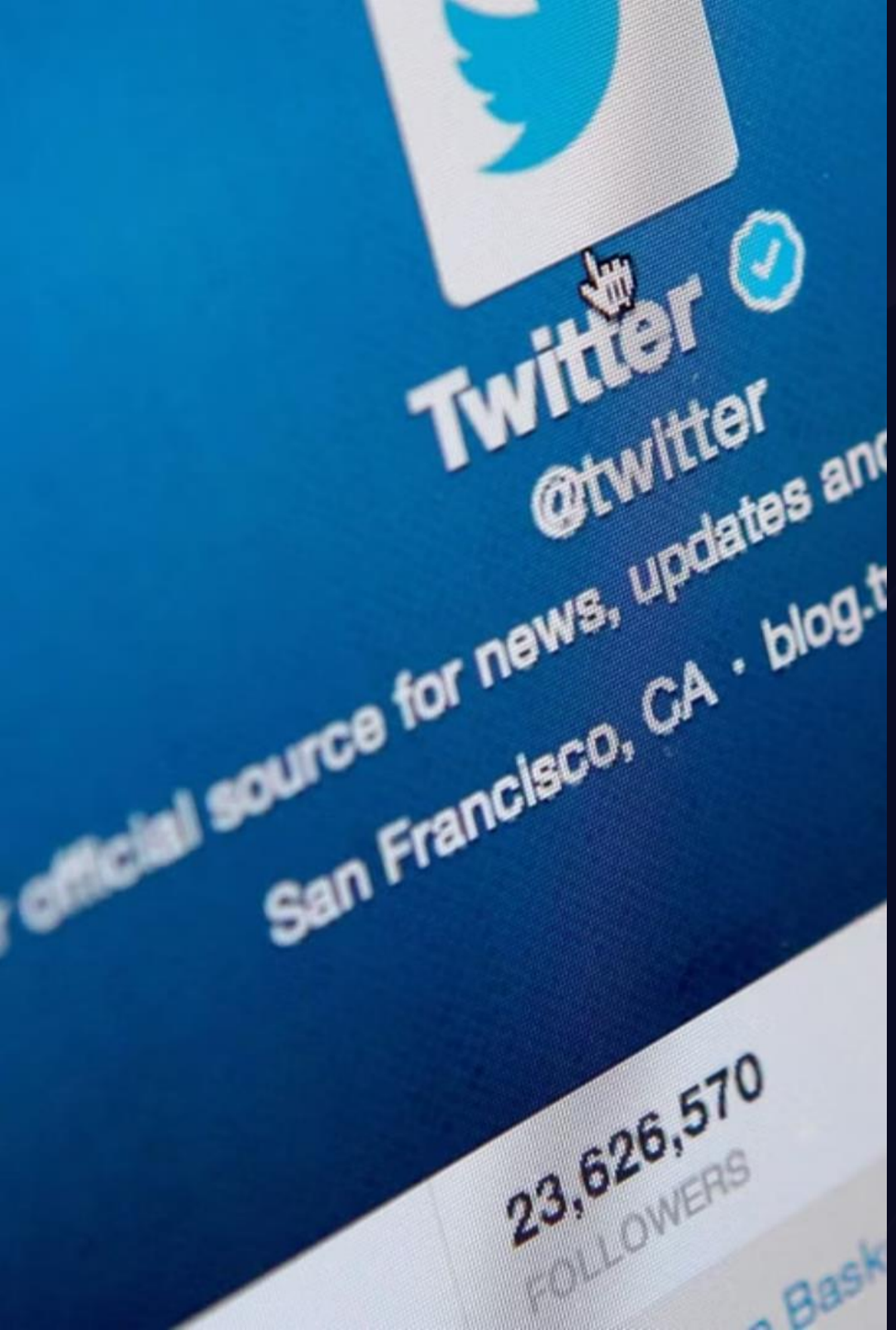


Twitter Sentiment Analysis: Unveiling Public Opinion

Discover how we leveraged machine learning to classify tweet sentiments, providing valuable insights for businesses and researchers

AT **Arun Totad**

HN **Hithesraja Navarethinam**

LR **Leekshitha Resoju**

PS **Preethi Sekar**

Professor Dr. Kevin Lee
STAT 5870
Big Data Analysis Using Python

Department of Statistics

Western Michigan University

Data Processed : 1.6 Million Tweets

1.6M

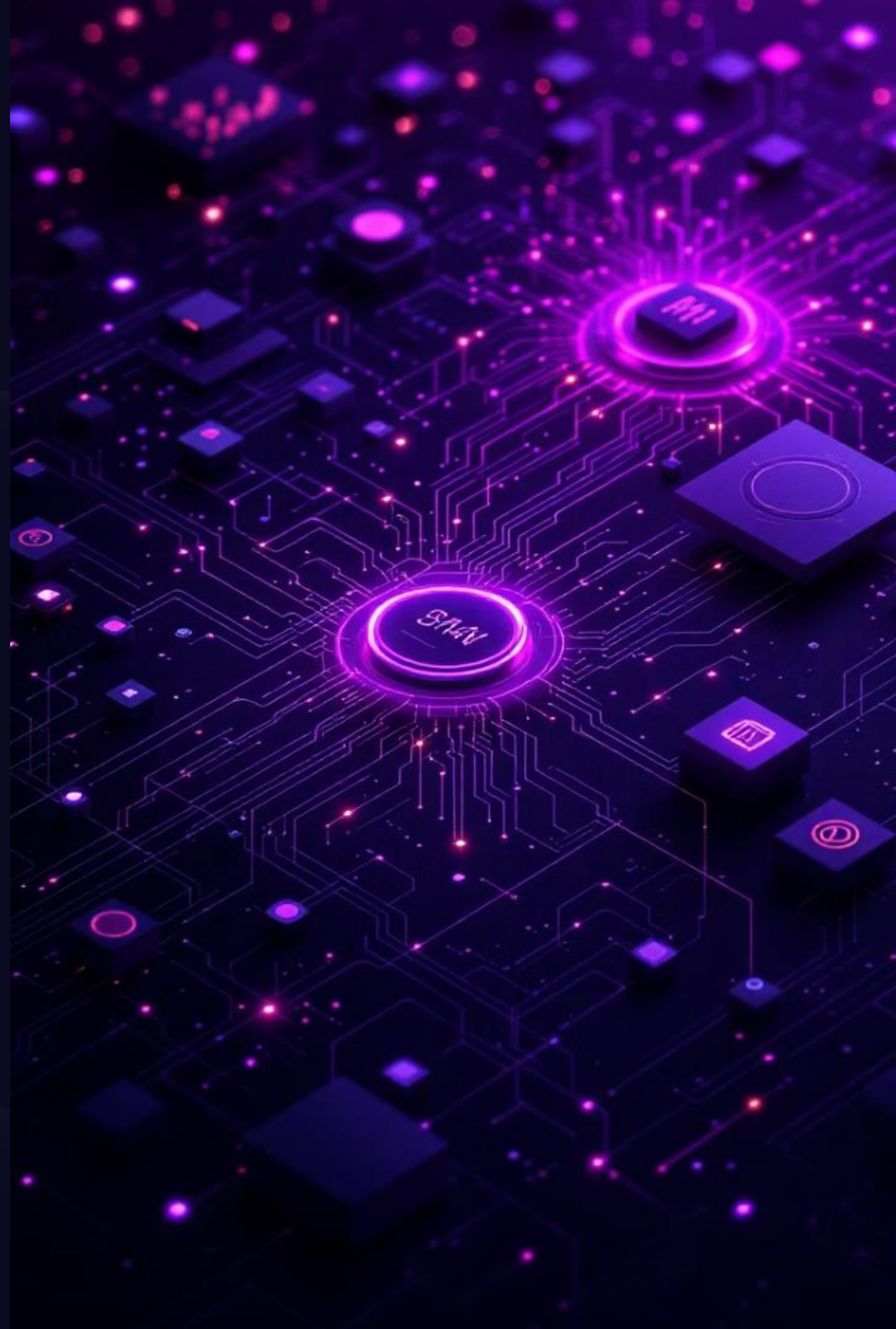
Tweets Analyzed

Our dataset contains a vast collection of user tweets, providing a robust foundation for analysis.

2

Sentiment Labels

Each tweet is labeled as either positive (1) or negative (0).



Our Mission: Decoding Tweet Sentiments

1

Input Raw Tweets

We start with unprocessed tweets from the Sentiment140 dataset.

2

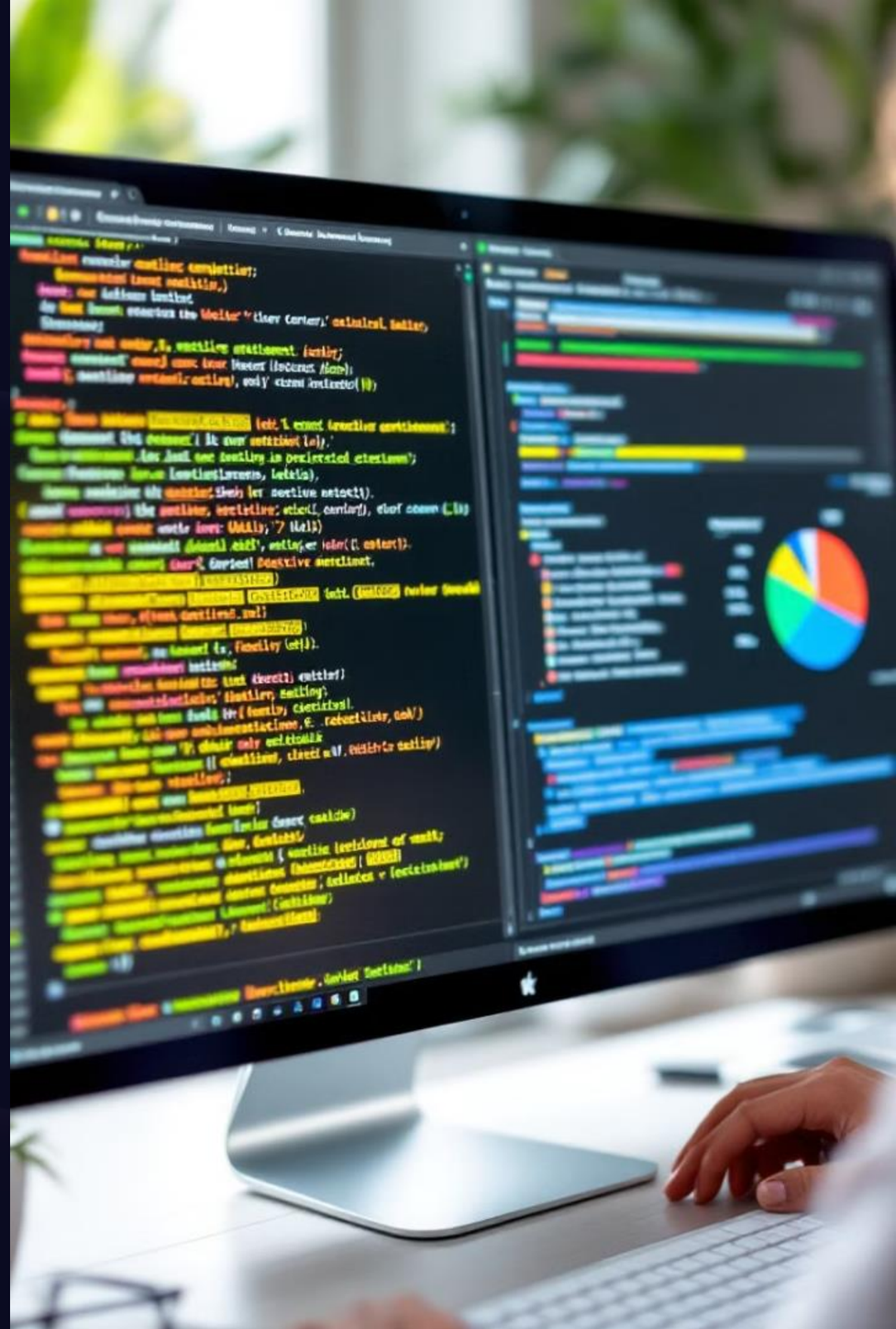
Process & Analyze

Our model processes and classifies the tweets based on sentiment.

3

Generate Insights

The results provide valuable insights into public opinion and brand perception.



Text Preprocessing: Cleaning the Data

1. **Tokenization**

   We convert all text to lowercase for consistency.

2. **Remove Noise**

   URLs, punctuation, and numbers are removed to focus on meaningful text.

3. **Lowercase Conversion**

   The text is split into individual words or tokens.

4. **Stopword Removal**

   Common words that don't carry sentiment are removed.

5. **Lemmatization**

   Words are reduced to their base form for better analysis.

Vectorization: TF-IDF in Action

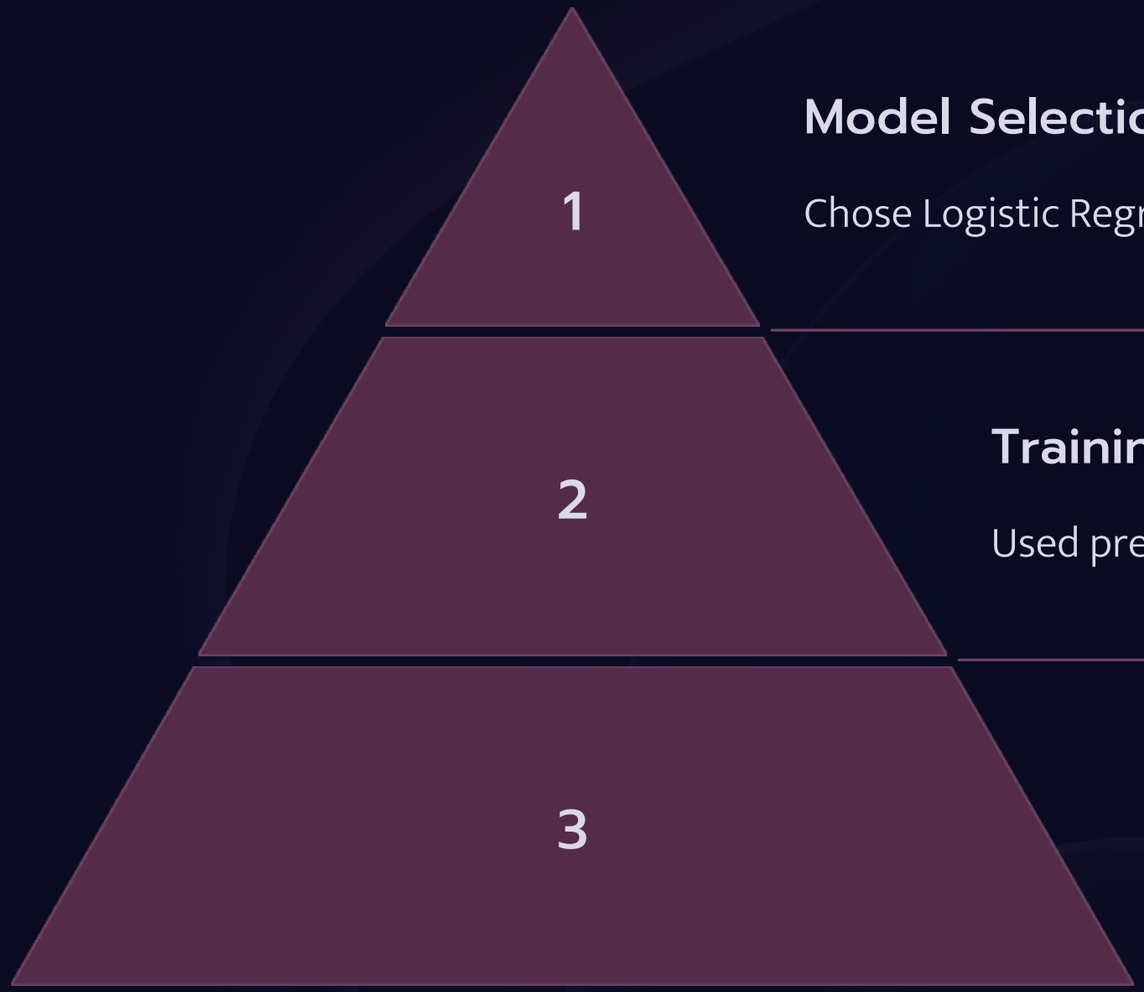
Term Frequency (TF)

Measures how frequently a term occurs in a document.

Inverse Document Frequency (IDF)

Reduces the weight of common words and increases the weight of rare words.

Building the Model: Logistic Regression



Model Selection

Chose Logistic Regression for its simplicity and interpretability.

Training

Used preprocessed and vectorized data to train the model.

Evaluation

Tested the model on a separate dataset to assess performance.

WESTERN MICHIGAN UNIVERSITY

Results: Impressive Performance

80% Accuracy

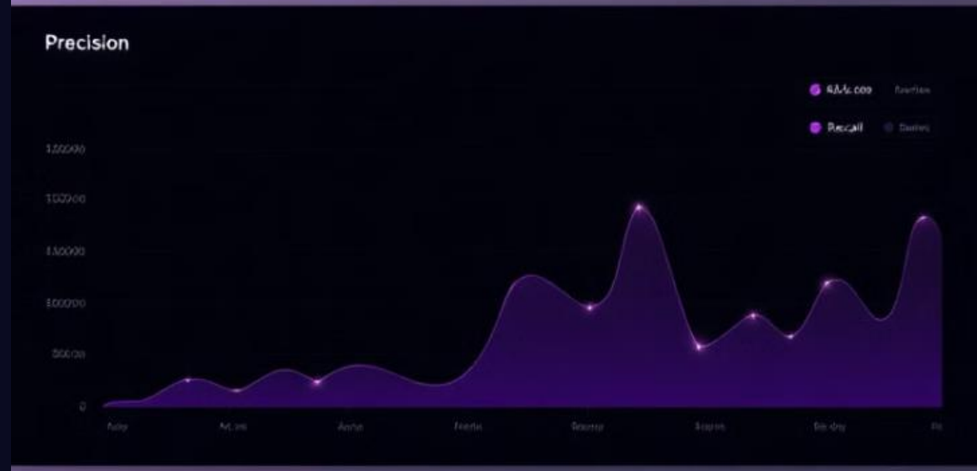
Our model achieved high accuracy in classifying tweet sentiments.

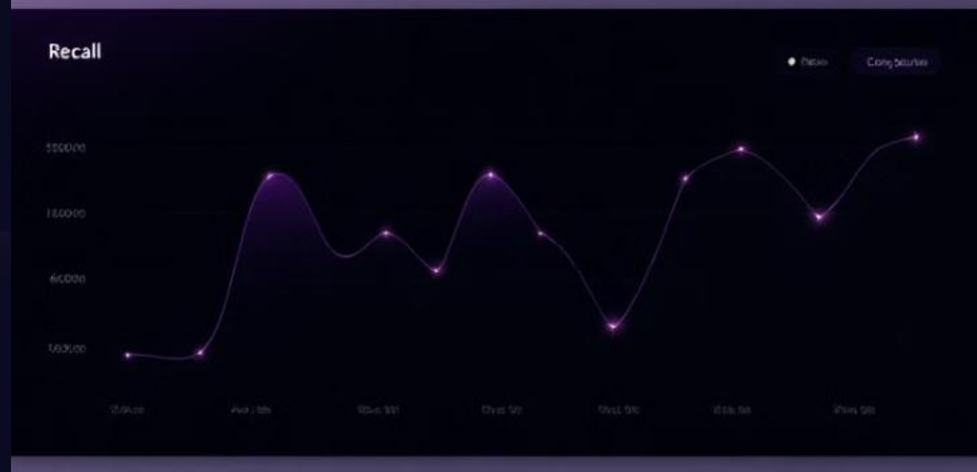
Balanced Performance

Similar precision and recall for both positive and negative classes.

Real-world Application

Successfully predicted sentiment on new, unseen tweets.





Conclusion: Unlocking Twitter's Sentiment

Effective Sentiment Classification

Our model successfully classifies tweet sentiments with high accuracy.

Wide-ranging Applications

From brand monitoring to market research, the possibilities are vast.

Future Enhancements

Exploring deep learning could further improve performance and insights.

WESTERN MICHIGAN UNIVERSITY