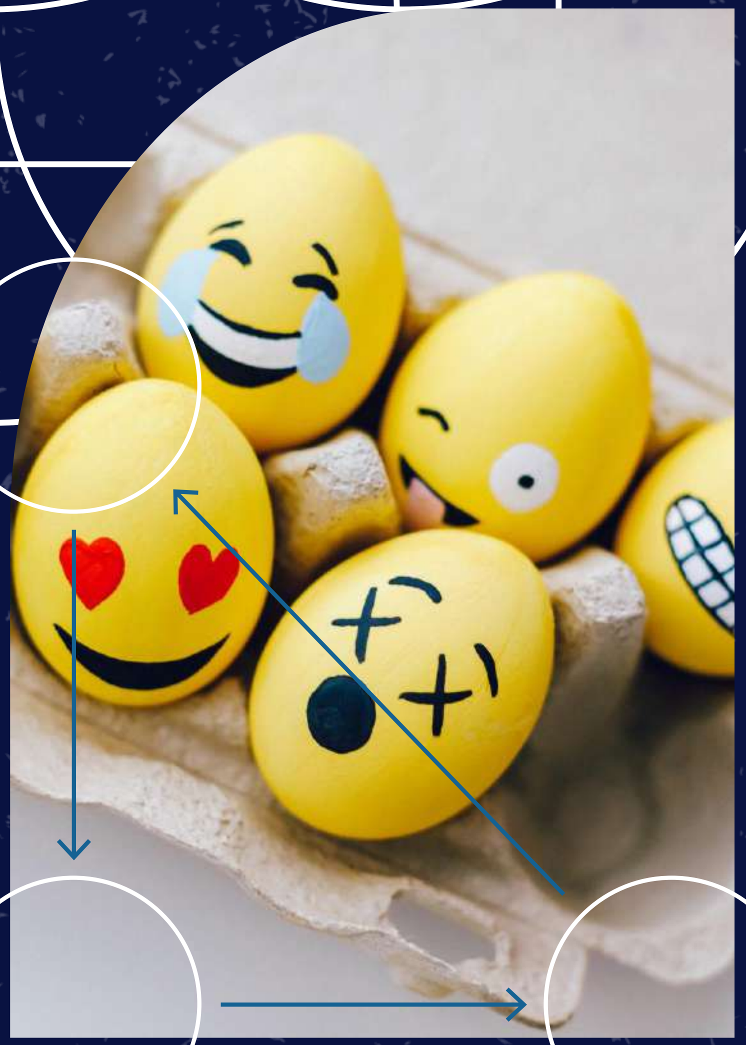
Emotion Insights Inc.

Understanding Emotions in Tweets



Presented by Group 4

Ronny Muthomi
Austin Mwasi
Faith Karimi
Mohammed Abdi
Mercy Kang'ethe

Understanding Emotions in Tweets

Social media platforms like Twitter are powerful mediums for users to express their opinions, emotions, and experiences. However, extracting meaningful insights from thousands of unstructured tweets can be challenging. Brands, governments, and researchers increasingly rely on sentiment analysis to understand public opinion and respond accordingly.

This project aims to build a machine learning model that can automatically classify the sentiment of a tweet Positive, Negative, No Emotion, or I Can't Tell based solely on its text content.

The core problem is the unstructured nature of tweet data, making it necessary to apply NLP preprocessing, vectorization, and robust classification techniques.

01 Insight into audience feelings

Brands can discern **real-time emotions** expressed by their customers in tweets, leading to better engagement strategies.

02 Enhanced customer connection

By understanding emotions, brands can create more **personalized marketing** campaigns that resonate with their audience's sentiments.

03 Improved brand reputation

Addressing customer concerns directly can significantly enhance a brand's **public image**, fostering trust and loyalty.

Objectives

01 Classify Tweet Sentiment

Build a machine learning model to classify tweets as Positive, Negative, No Emotion, or I Can't Tell using only their text content.

02 Apply NLP Techniques

Use text preprocessing and TF-IDF vectorization to convert raw tweets into structured input for modeling.

03 Benchmark Multiple Models

Classifying emotions allows brands to base their strategies on solid data, enhancing their responsiveness to consumer sentiments.

04 Handle Class Imbalance

Ensure fair performance across all sentiment classes through evaluation metrics and model tuning.

05 Support Real-World Use Cases

Provide a scalable tool for social listening, public feedback analysis, and automated sentiment monitoring.

Classifying Emotions in Tweets

01 Enhanced Brand Engagement

Identifying emotional tones helps brands tailor their messaging and connect effectively with their audience.

02 Improved Customer Insights

Understanding emotions in tweets provides valuable feedback that can inform product development and marketing strategies.

03 Data-Driven Decision Making

Classifying emotions allows brands to base their strategies on solid data, enhancing their responsiveness to consumer sentiments.

CLEANING DATA

Remove irrelevant information to enhance the quality of analysis.

REMOVING DUPLICATES

Eliminate repeated entries to ensure unique data representation.

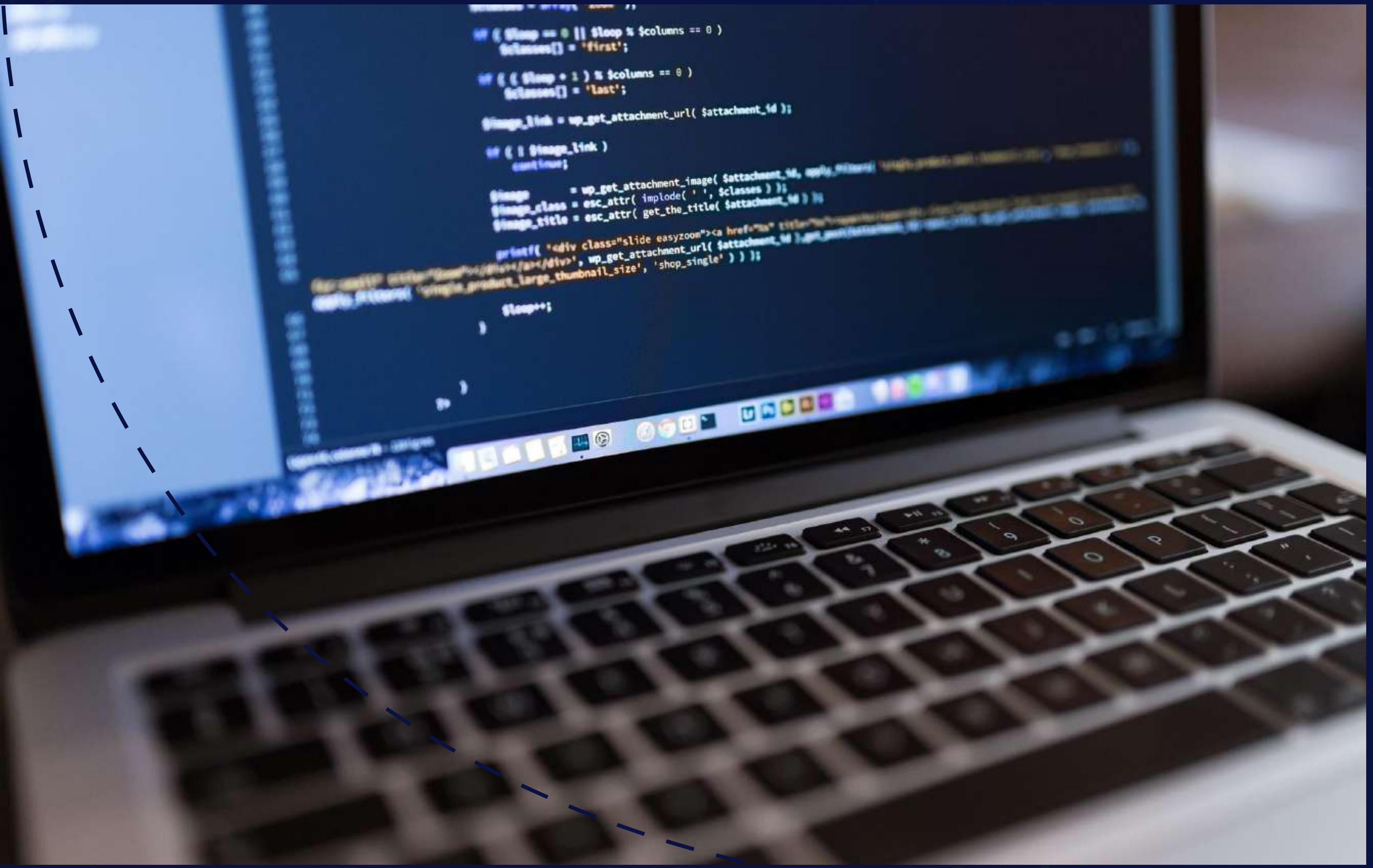
STANDARDIZING TEXT

Convert text to a consistent format for accurate processing.

FINAL REVIEW

Verify all steps to ensure data integrity before analysis.

Turning Text into Numbers



01 TF-IDF Method

TF-IDF measures word importance in a document, helping identify **relevant keywords** for understanding sentiment.

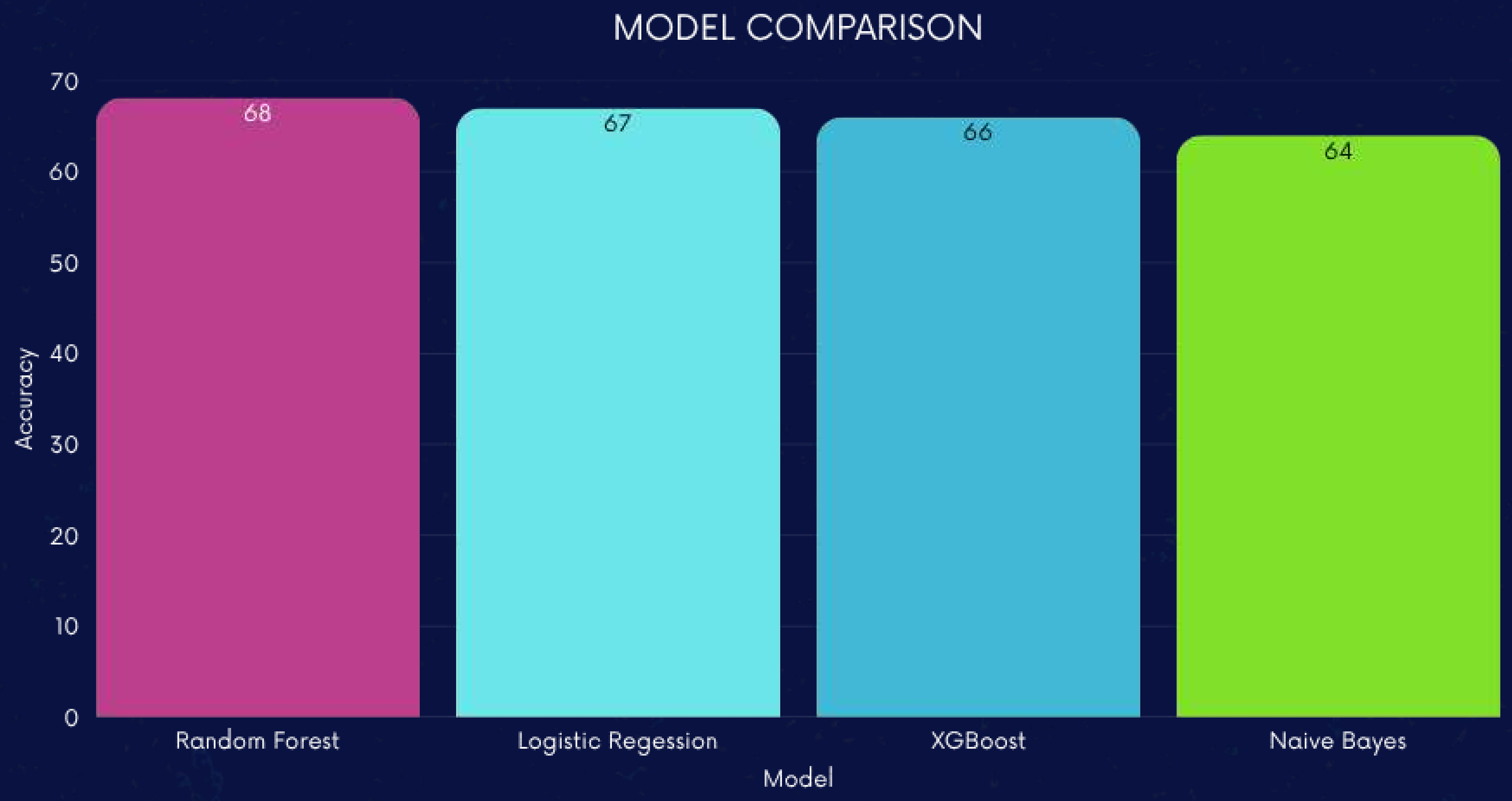
02 N-grams Analysis

N-grams capture sequences of words, enhancing context and meaning to improve sentiment classification accuracy.

03 Enhanced Data Representation

These techniques transform text into numerical format, making it easier for models to interpret and classify emotions effectively.

Why We Chose Random Forest



01 High Accuracy

Random Forest offers robust performance by effectively handling complex datasets, leading to reliable predictions in our sentiment analysis.

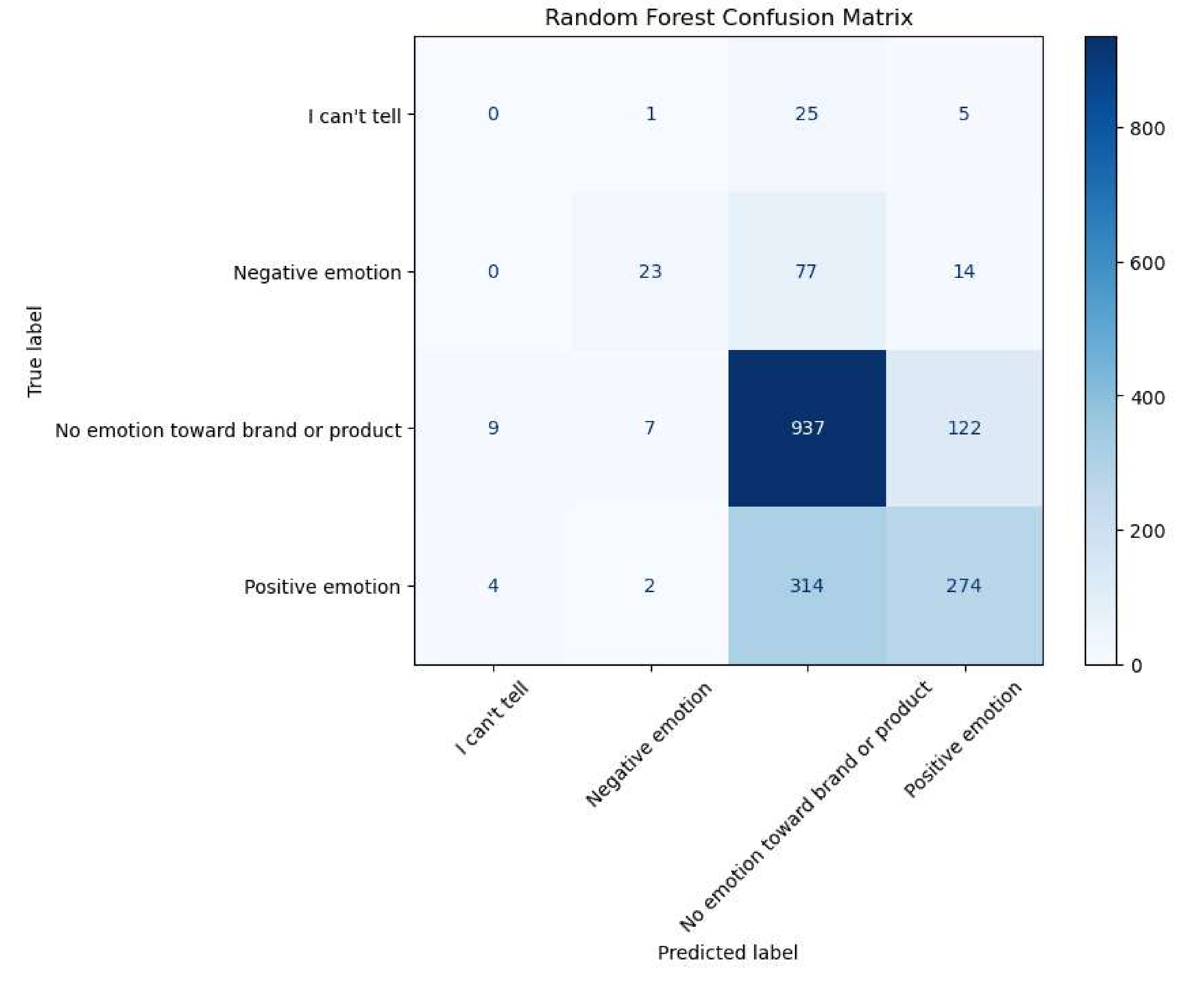
02 Handles Overfitting

Its ensemble approach reduces overfitting risks, enhancing generalization to new data and ensuring our model remains effective over time.

03 Versatile and Scalable

Random Forest easily adapts to different data sizes, making it suitable for our project's evolving needs and improving our understanding of tweet sentiments.

Model Performance: Successes and Challenges



01 High Accuracy Rate

Our model achieved a **high accuracy rate**, effectively classifying tweets into their respective emotion categories.

02 Robust Feature Extraction

Utilizing TF-IDF and N-grams resulted in **robust feature extraction**, enhancing the model's understanding of tweet sentiments.

03 Handling Ambiguity

One challenge faced was **handling ambiguity** in tweets, where sarcasm and nuanced emotions often led to misclassification.

04 Sentiment Classes Were Imbalanced

The dataset had more positive and neutral tweets compared to negative ones. This imbalance affected model recall for minority classes (like Negative), emphasizing the need for metrics like weighted F1-score.

Key Findings and Implications

01 Product Mention Influences Sentiment

Sentiment trends varied by product category, suggesting brand-specific perceptions. For example, one brand might receive more negative feedback than others.

02 Enhanced brand engagement

Understanding tweet sentiments allows brands to **craft targeted responses** that resonate with their audience.

03 Improved customer insights

Analyzing emotions in tweets provides deeper **insights into customer feelings**, helping brands adjust strategies.

04 Data-driven decision making

Leveraging sentiment analysis empowers brands to make **informed decisions** based on customer emotional trends.

Challenges Faced in the Project

01 **Data Quality Issues**

Inconsistent tweet formats and **noisy data** made cleaning difficult, affecting the accuracy of sentiment analysis.

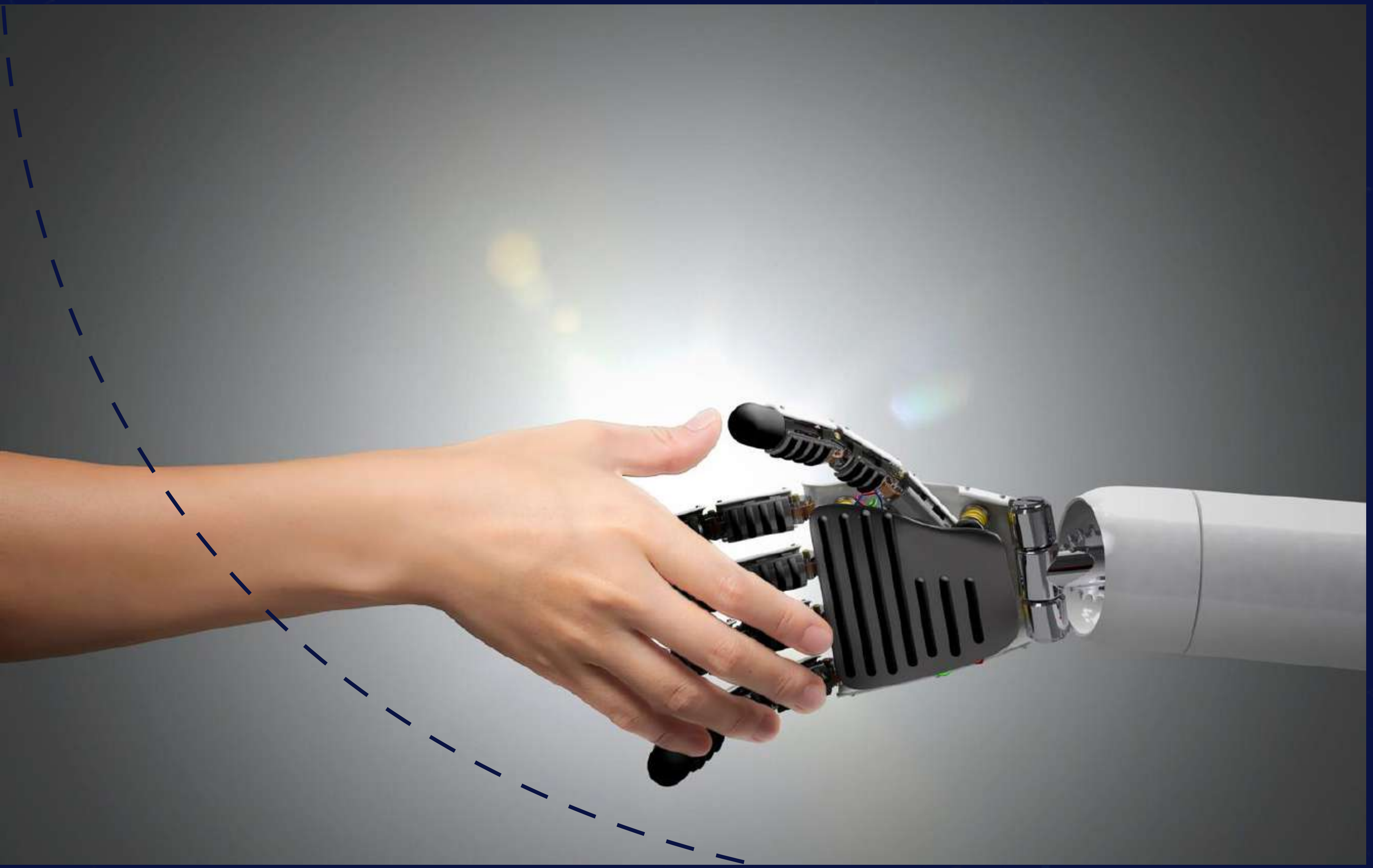
02 **Emotional Ambiguity**

Tweets often express **mixed emotions**, complicating the classification into distinct emotional categories.

03 **Resource Limitations**

Limited computational resources hindered the processing speed and scalability of our sentiment analysis model.

Future Improvements and Directions



01 **Enhanced Emotion Categories**

Expanding our classification to include more **nuanced emotions** for better sentiment analysis.

02 **Advanced Data Techniques**

Implementing cutting-edge algorithms and **data preprocessing** methods for improved accuracy and insights.

03 **Real-time Analytics**

Developing a system for **real-time sentiment analysis** to help brands respond to customer emotions instantly.

Contacts

Email:
ronny.muthomi@student.moringaschool.co
m
faith.karimi@student.moringaschool.com
Github:
https://github.com/RonnyMuthomi/Group4_NLP_Project