

# Phase 4 Project

Audrey, Slyvia, Stella, Petronila, George, Job



# Summary



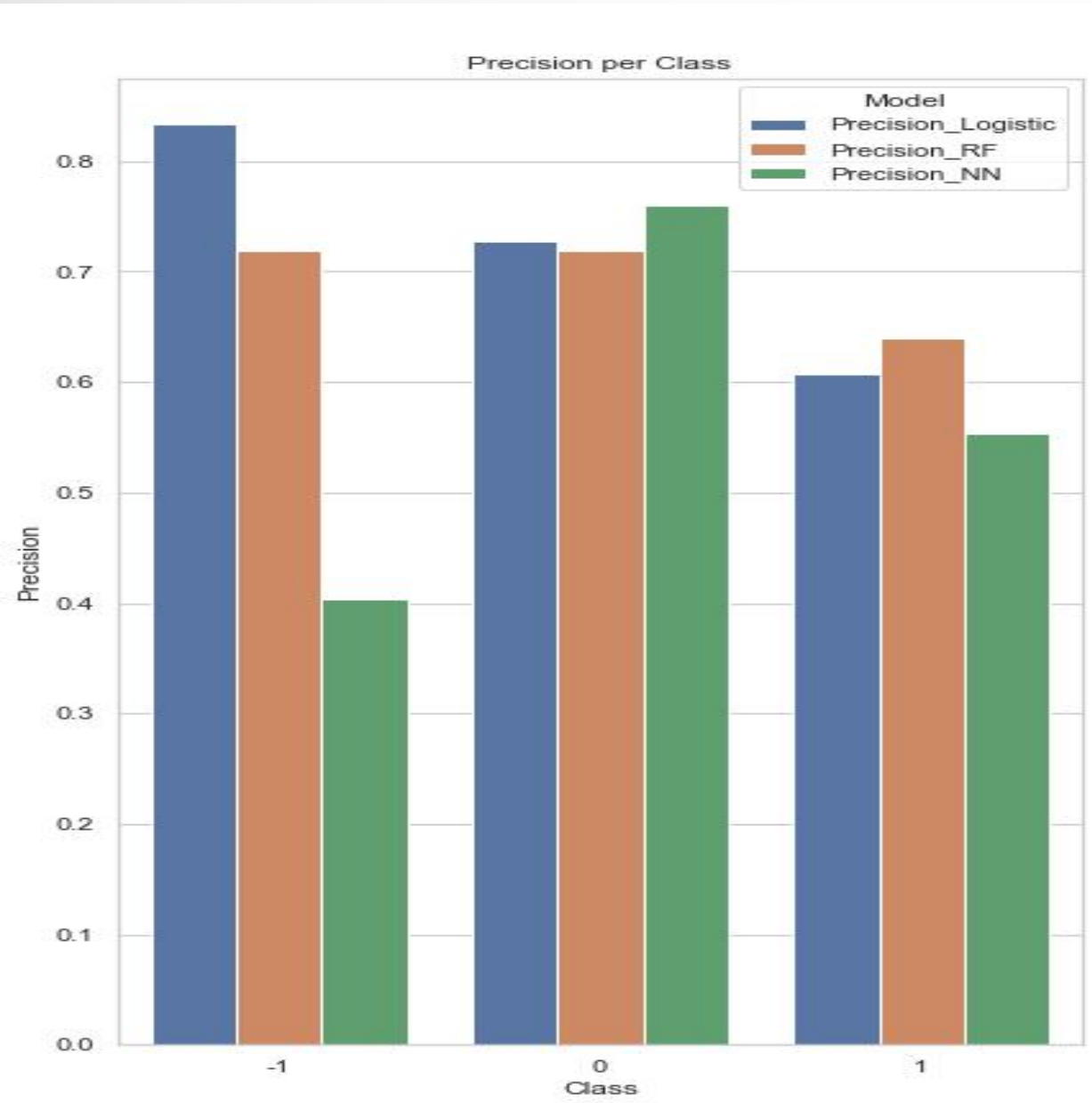
This project involves building a model to analyze tweets about Apple and Google products, identifying the underlying emotion(positive, negative, or neutral) to help companies monitor public sentiment efficiently without manually reviewing thousands of tweets.

# Data and Methods



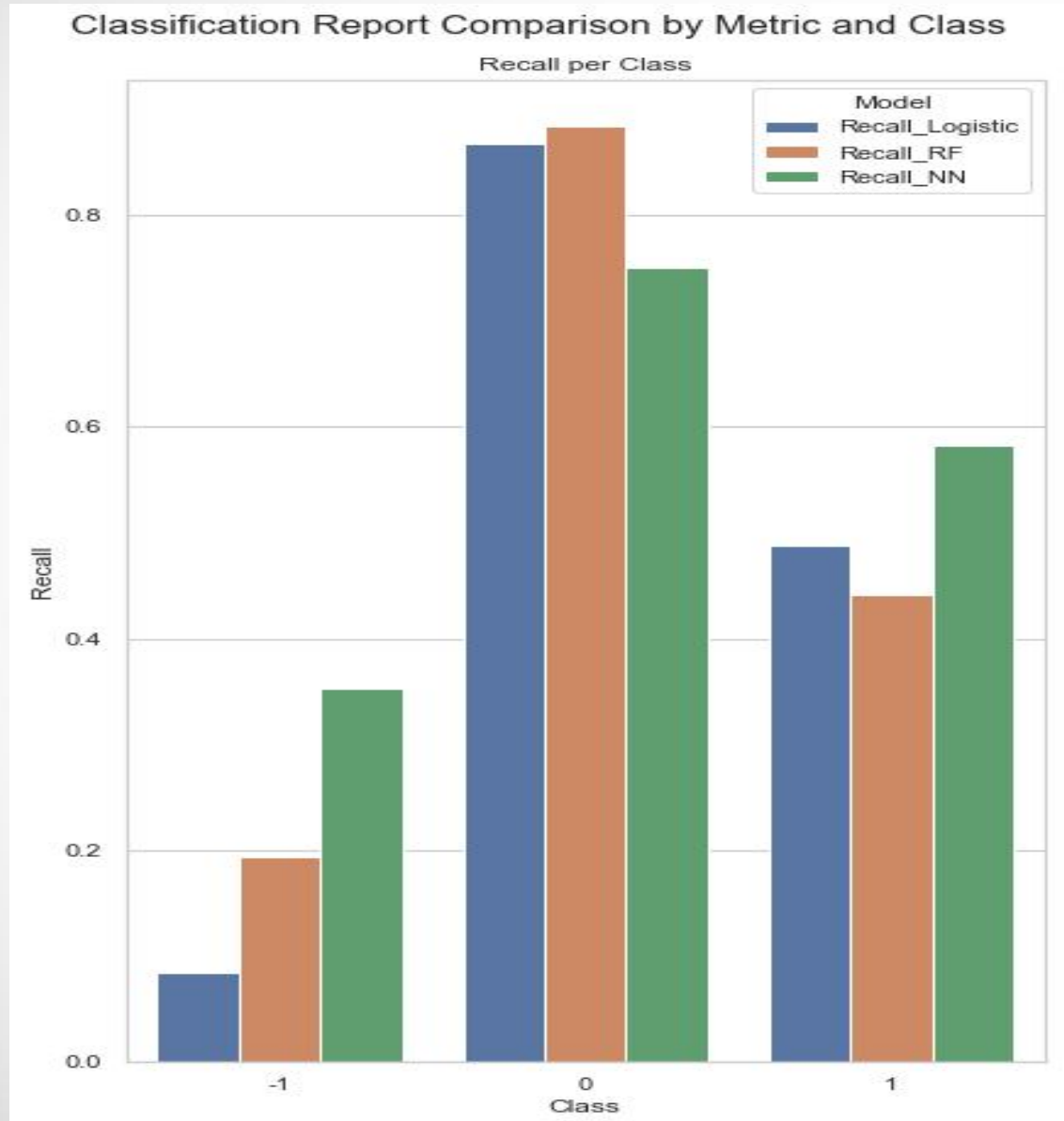
- The methods involved preprocessing the tweet text using NLP techniques, training Logistic Regression, Random Forest, and Neural Network models, and evaluating their performance using metrics such as precision, recall, and F1-score.

# Results



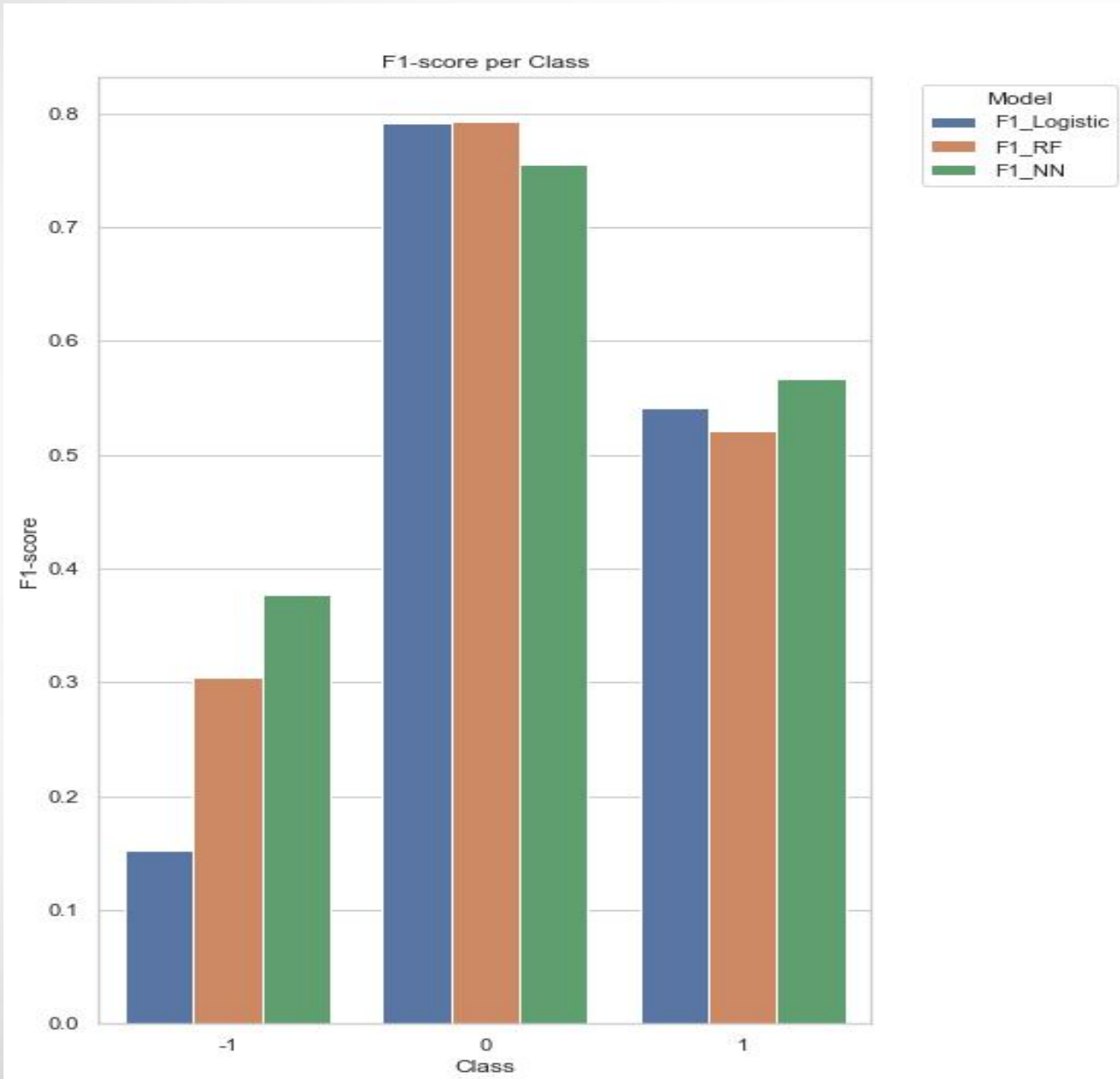
- No single model outperforms the others across all classes. Instead, Logistic Regression is most precise for class -1, the Neural Network for class 0, and Random Forest for class 1.

# Results



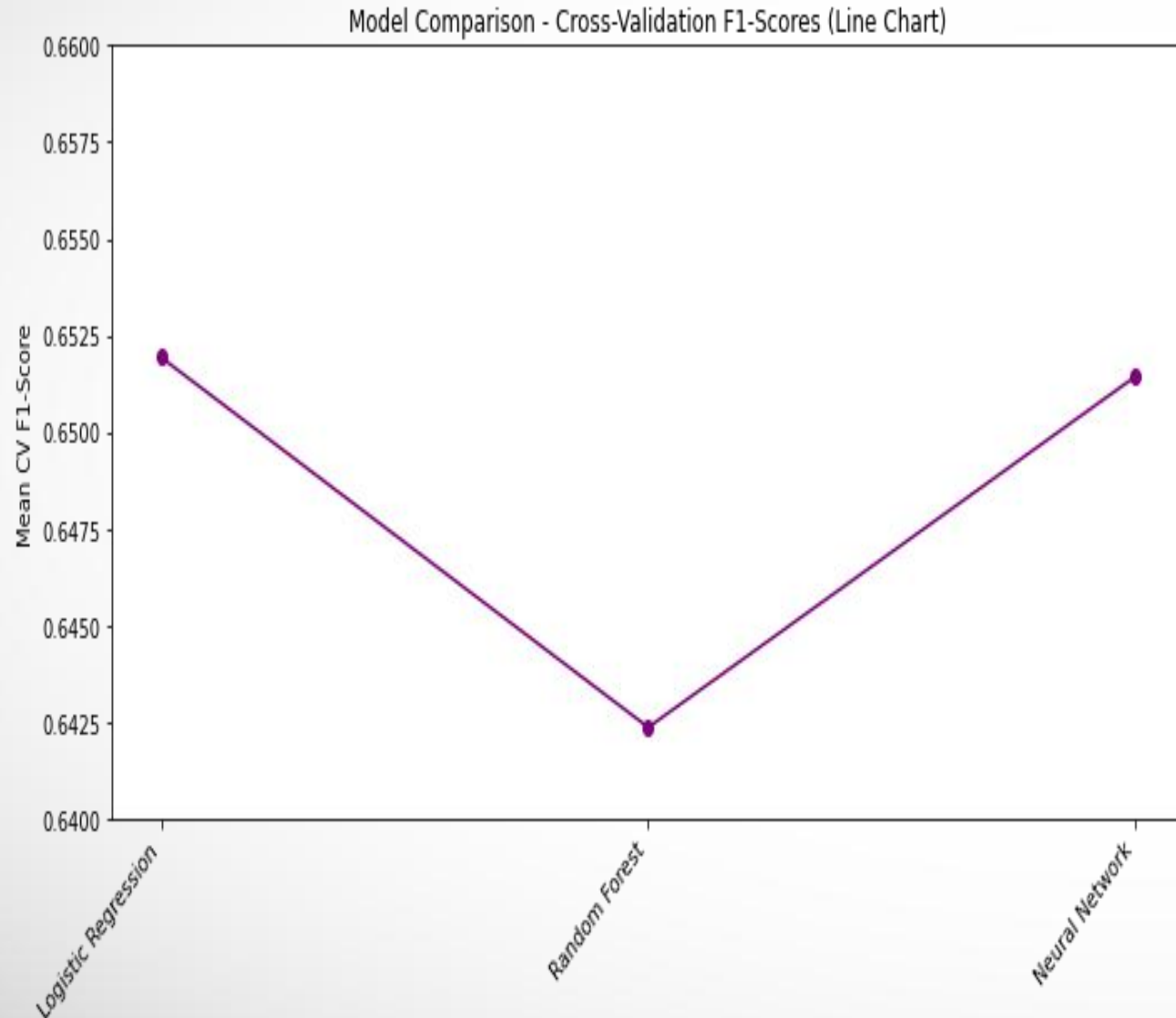
- The recall chart indicates the Neural Network achieved peak performance in classes -1 and 1, whereas the Random Forest was most effective in class 0.

# Results



- Based on F1-score, the Neural Network is superior in classes -1 and 1, while Logistic Regression and Random Forest show the top results in class 0.

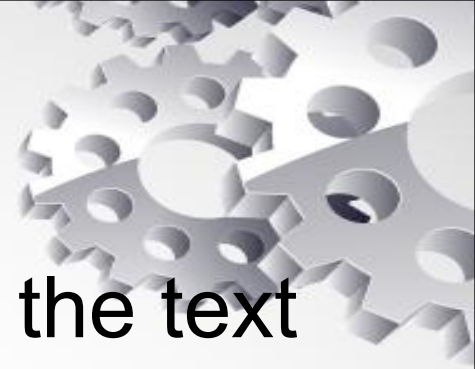
# Results



- Based on the mean cross-validation F1-scores, the Logistic Regression and Neural Network models performed slightly better on average than the Random Forest model.

# Conclusion

- Sentiment can be accurately predicted using only the text of tweets, as demonstrated by the strong performance of the models, particularly Logistic Regression and Neural Networks.
- Logistic Regression performed best overall, followed closely by the Neural Network. Both outperformed the Random Forest model in terms of F1-score.





# Recommendations

- Continuously retrain the model with new tweets and expand to include emojis, hashtags, and user metadata to capture richer sentiment signals.
- Regularly analyze sentiment data to identify recurring complaints or praise. This can guide product development teams on areas of improvement or features that users love.
- Host the trained model in a cloud environment and integrate it with Twitter's API to continuously collect and analyze tweets.

