University of Sunderland

**Faculty of Technology
School of Computer Science
PROM06 – Research Project**

SENTIMENT ANALYSIS IN SOCIAL MEDIA

Student Name: Husam Ahmed
Student Number : bh92nq
Supervisor: Becky Allen

OVERVIEW AND PROJECT INTRODUCTION

- **sentiment analysis** aims to extract the overall feel of a text using natural language processing.

- The research project **aims** to replicate previous research work and validate the points reached by conducting experiments and making recommendations.

SIGNIFICANCE AND OBJECTIVES

This study seeks to contribute to the growing body of research on sentiment analysis in social media by examining the accuracy and effectiveness of various classification algorithms that are used in sentiment analysis through the machine learning approach.

Building a machine learning sentiment analysis model that is capable of classifying texts' sentiment as either positive or negative with a 65–80% classification accuracy.

RESEARCH QUESTIONS

RESEARCH Q1:

What impact does the choice of the machine learning algorithm have on the performance metrics?

1– Random Forest.
2– SVM.
3– Logistic Regression.

RESEARCH Q2:

What impact does the choice of the feature extraction technique have on the performance metrics?
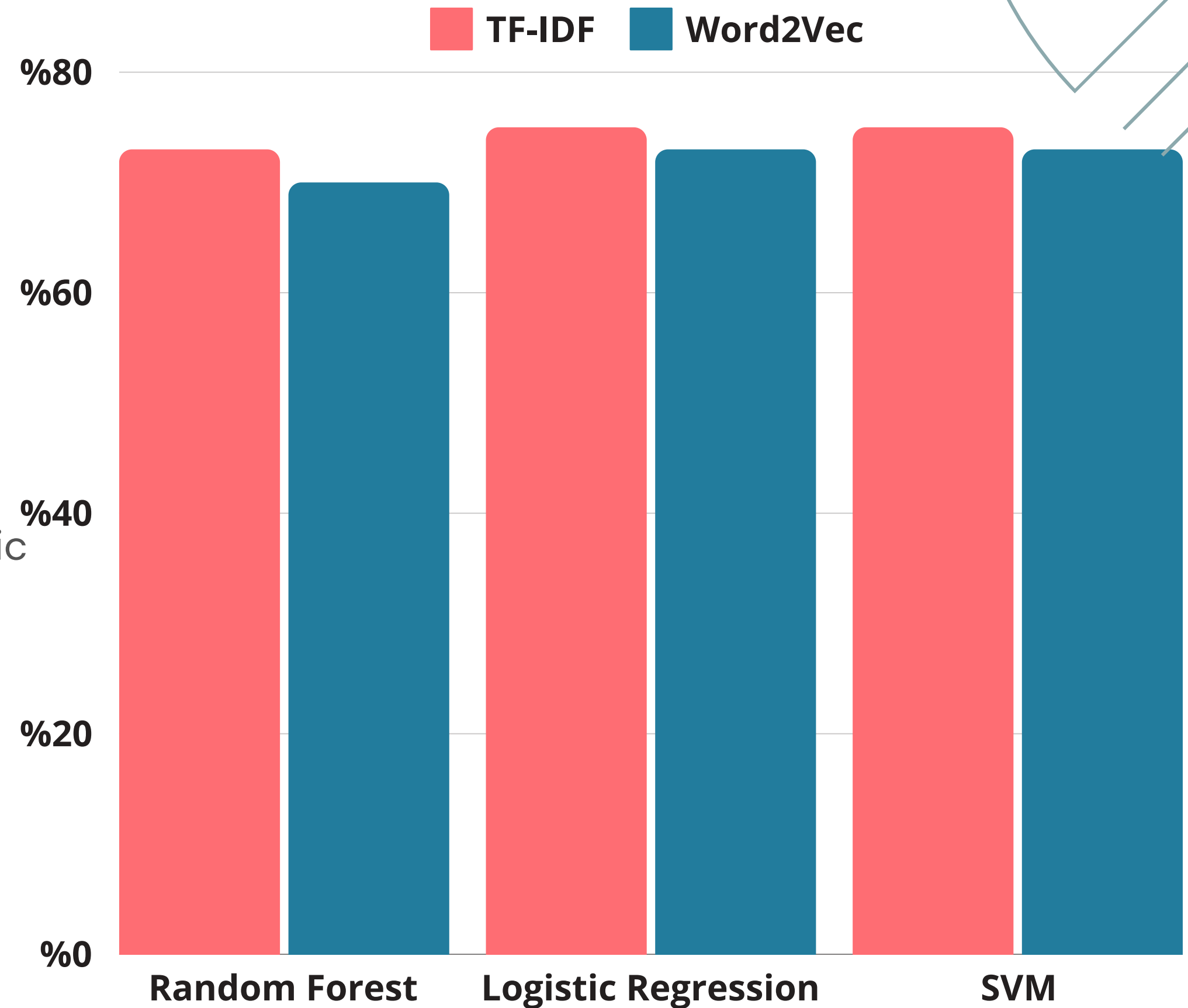
1– TF–IDF.
2– Word2vec.

MODEL EVALUATION:

ACCURACY

TRAINING TIME

it was found through testing that Logistic Regression accompanied with TF–IDF yield the highest text classification accuracy (75%) and the significantly shortest training time (0,03 minute).

RESEARCH OUTCOMES

6 sentiment analysis prototypes have been developed. Logistic Regression accompanied with TF-IDF yields the highest text classification accuracy (75%) and the significantly shortest training time (0,03 minute).

Ethical and legal considerations which have been a key element of this research project, they safeguard user privacy, mitigate biases, and ensure fair data usage, aligning with regulations like GDPR.

The **research project encountered constraints** in data utilization, employing only 100,000 data points out of the available 6 million.

Computational limitations restricted the thorough exploration of resource-intensive techniques such as SVM and Word2Vec.
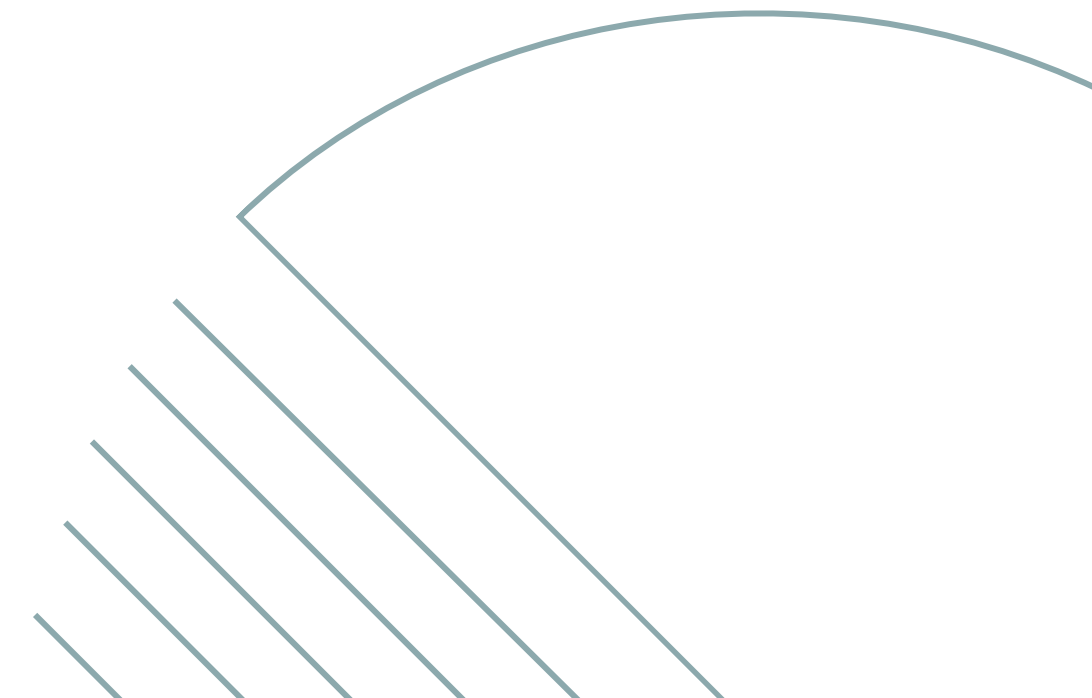
CONCLUSION

Sentiment analysis empowers decision-making and strategy formulation by deciphering human behavior in diverse contexts.

Replicating previous work has validated the points reached and put trust in what has been discovered and that is one of the objectives of this research.

The research project was able to achieve all the objectives that have been set by developing a prototype sentiment analysis system with a classification accuracy of 75%.

Computational constraints hindered a comprehensive exploration of resource-intensive techniques like SVM and Word2Vec, future researchers with access to greater resources are encouraged to explore the capabilities of these techniques.