

Sentiment Analysis on YouTube Comments

AUTHORS

Kontos Christos
Efthymiadis Konstantinos



ARISTOTLE
UNIVERSITY
OF THESSALONIKI

SCHOOL OF
INFORMATICS
Data & Web
Science / MSc Program

INTRODUCTION

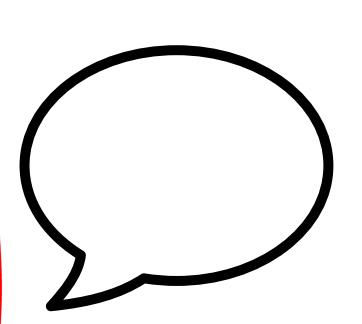
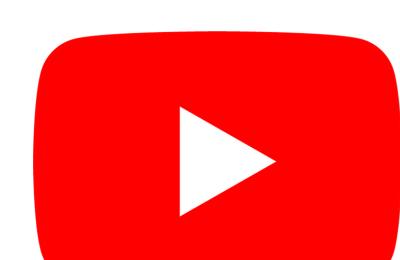
- Sentiment Analysis is a specialized field of Machine Learning that identifies and categorizes text as positive or negative.
- An application of Sentiment Analysis involves analyzing social media comments to track public perception.
- In this project, we utilized a multilingual YouTube comment dataset for training, evaluated their ability to predict sentiment accurately and tested their viability on a Greek dataset to assess real-world adaptability.

DATASET

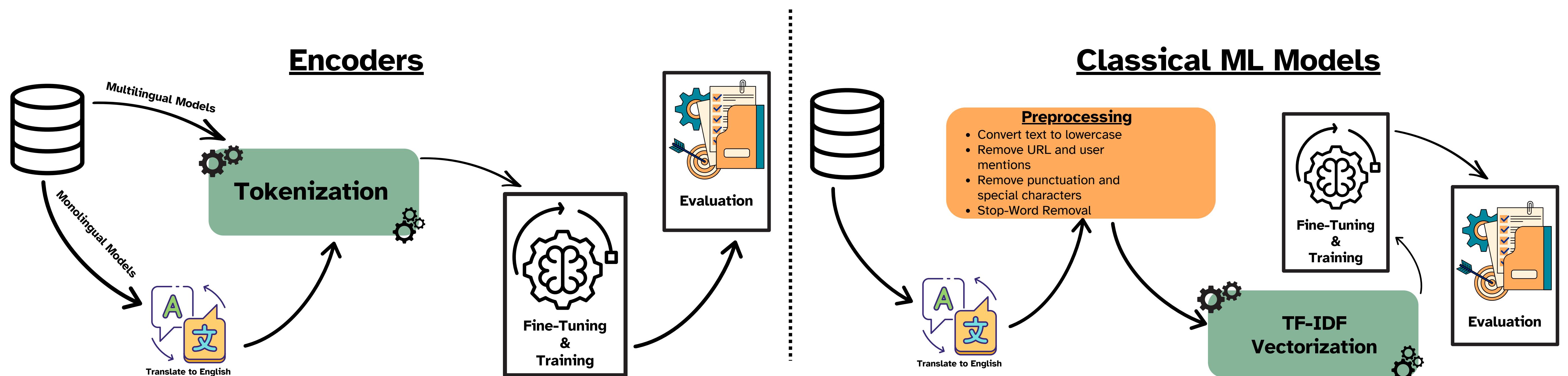
60000 Comments

Positive: 30000

Negative: 30000

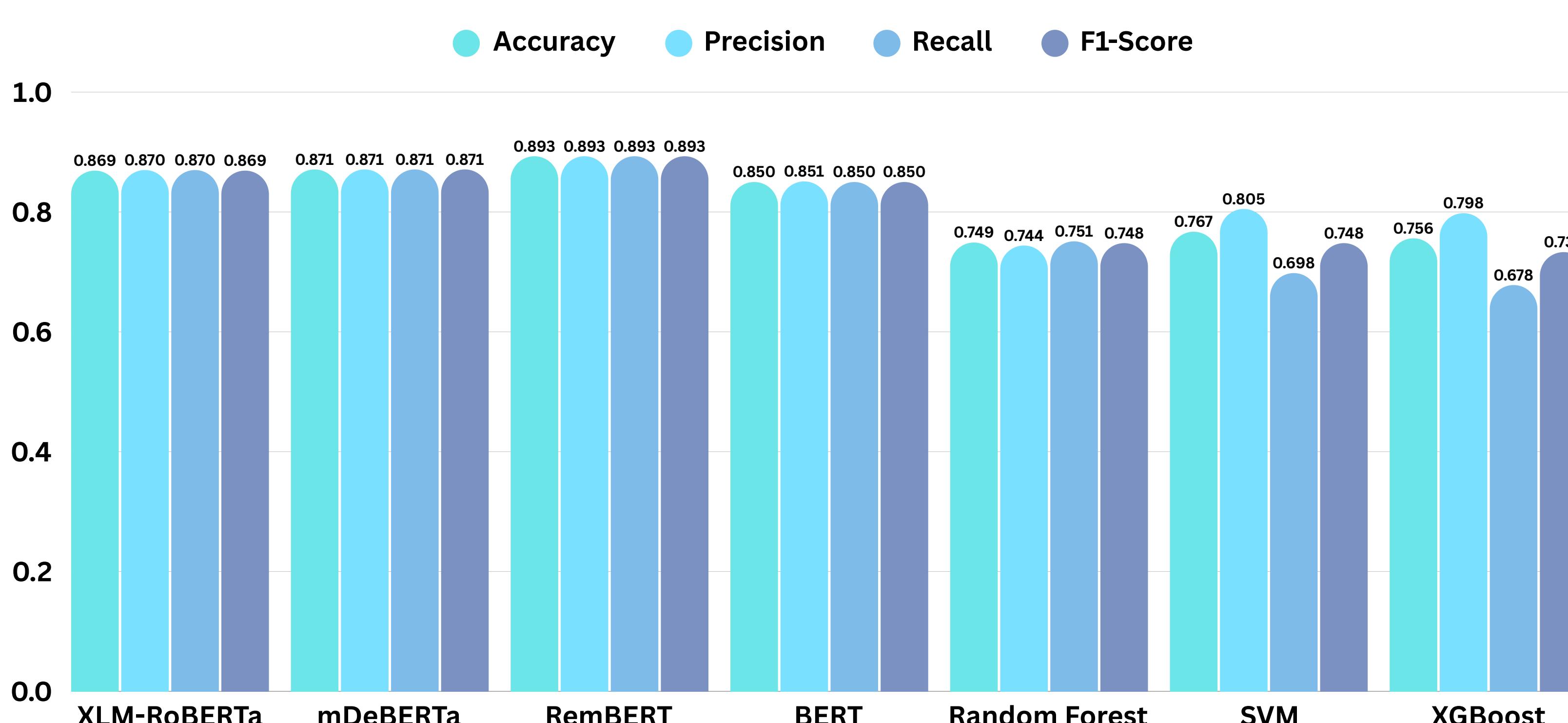


METHODOLOGY

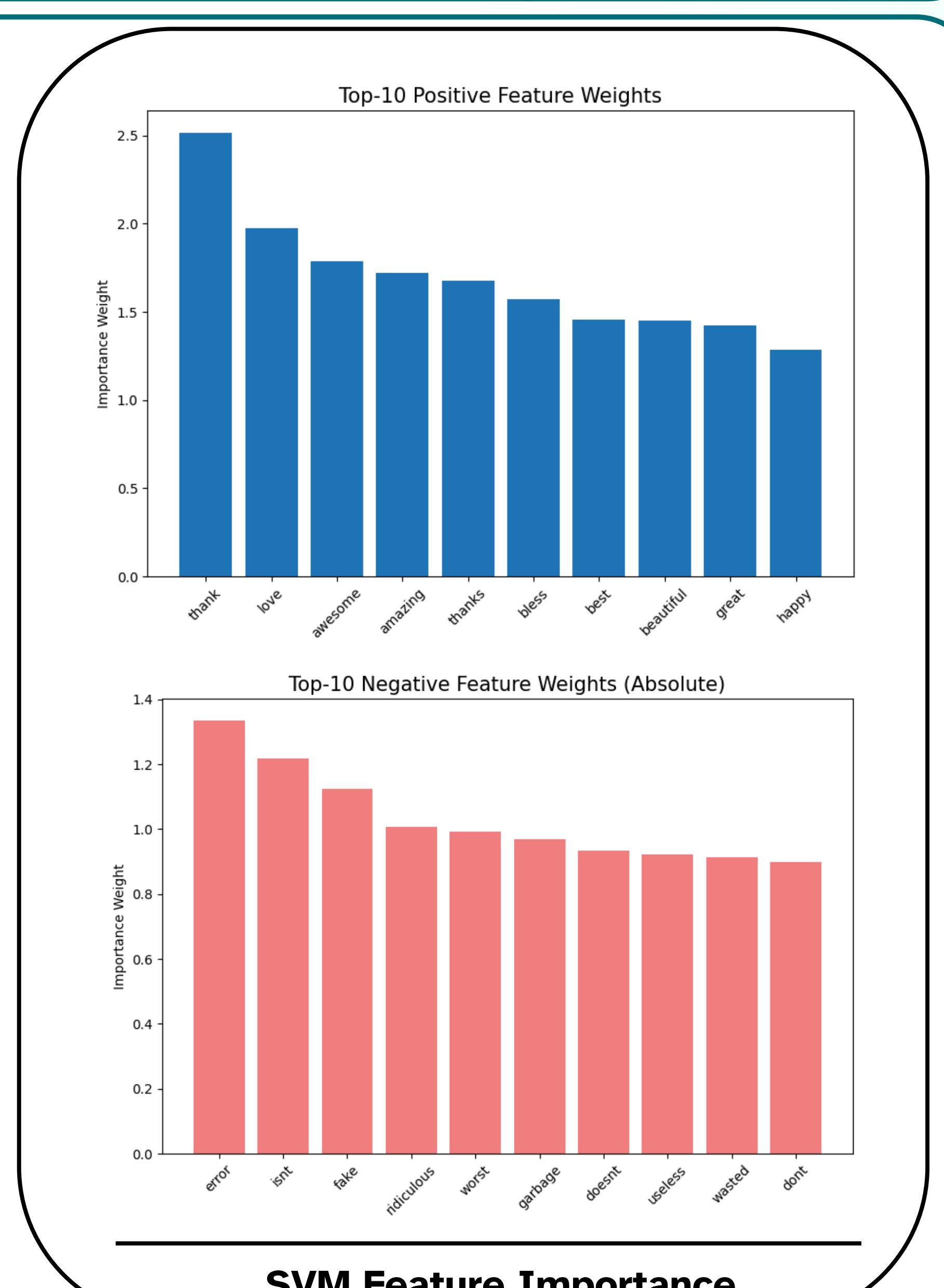
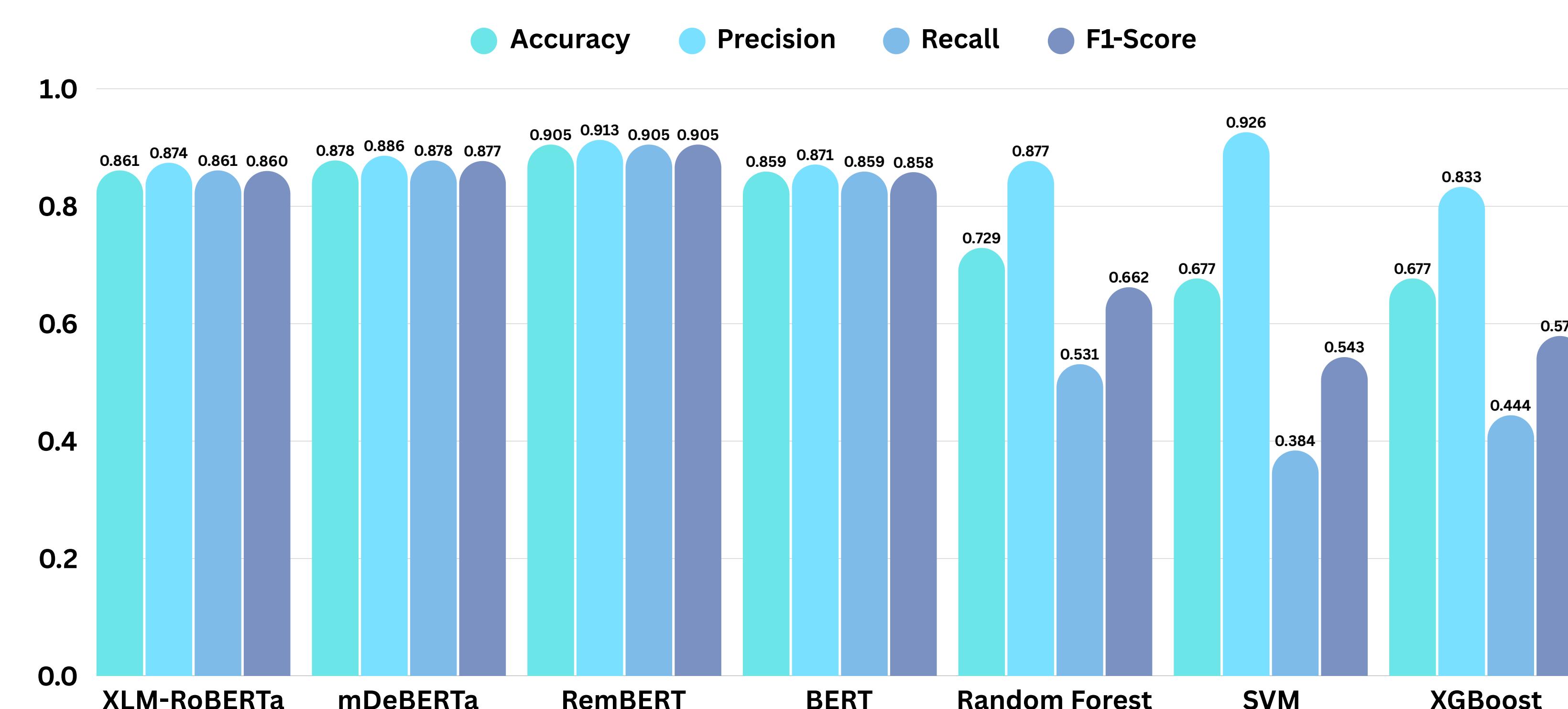


RESULTS

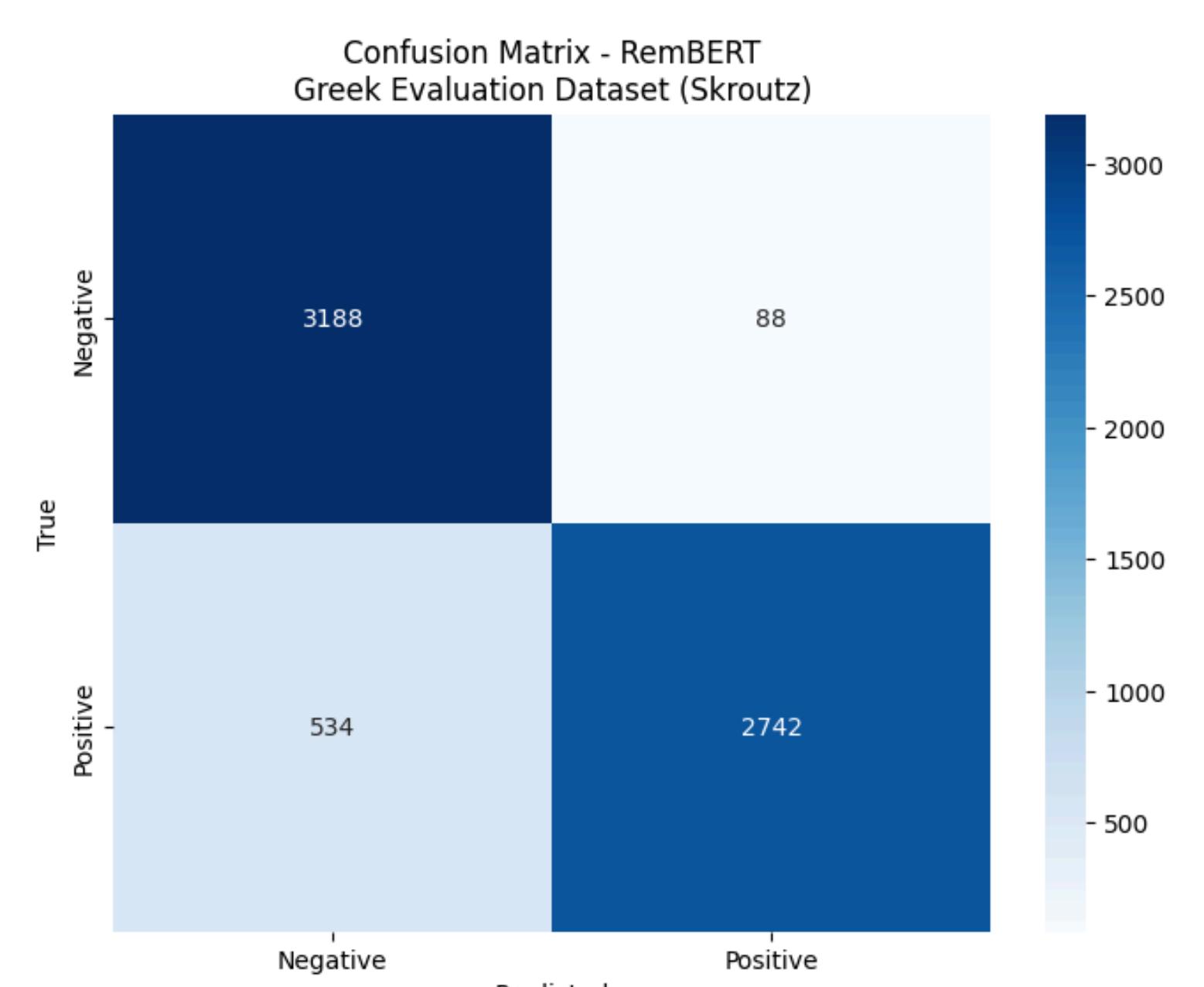
Evaluation Metrics for YouTube Comments



Evaluation Metrics for Greek Dataset Comments (Skroutz)



SVM Feature Importance



- Encoder models, specifically RemBERT significantly outperformed classical ML baselines in multilingual sentiment classification
- The high accuracy maintained on the Skroutz dataset demonstrates the models' strong adaptability to real-world data

CONCLUSION

CONTACT INFO

ckontosb@csd.auth.gr
kefthyme@csd.auth.gr

