Université d'Ottawa | University of Ottawa

Detecting Suicide Intention on Social Media

A Multi-Modal Analysis on NLP techniques and Fairness vs Bias

Presented by: Jenifer Yu, Sabrina Cai, Kelvin Mock

2025-03-31

Faculté de génie | Faculty of Engineering

uOttawa.ca

uOttawa

Why This Topic

Datasets

Training Set

- Twitter Suicidal Data – Kaggle CSV
- Social Media Sentiment Analysis – Kaggle CSV

Validation Set

- Reddit SuicideWatch Posts – Web Scraping JSON

Test Set

- Depression Tweets – Kaggle JSON

Labels

- 0 = non suicidal
- 1 = suicidal

Data Preprocessing

Steps

1. Normalize Emojis
2. Normalize Symbols - @ # http
3. Normalize Punctuations
4. Convert to Lowercase
5. Lemmatize – with codes from Assignment 1
6. Tokenize Words – with codes from Assignment 1
7. Normalize Stopwords – using StopWords.txt for Assignment 1
8. Vectorize with `DistilBertTokenizer` – contextual meaning

- Extract Sensitive Attributes for bias analysis
- Annotate Social Media Sentiment Analysis with Google Gemini
  - DORIS Scale – 0 to 9 annotation (9 = suicidal)

Model Choices

Models

- Baseline: Simple ML
- Deep Learning based
- LLM-based Model
- Hybrid: Ensemble

SENTIMENT ANALYSIS
Discovering people opinions, emotions and feelings about a product or service

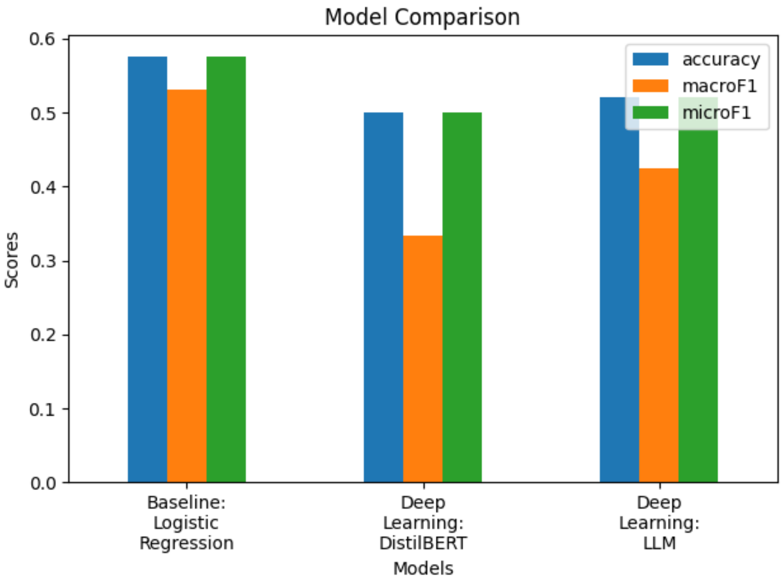

Figure 1: Model Comparison

Considerations

- In collaboration with *CSI5195 - Ethics in AI*
- Computation Limit

Evaluation Measures

Metrics

- Overall Accuracy
- Precision
- Recall
- F1-Score (Macro and Micro)

Class Imbalances

- Area-Under-the-Curve (AUC) from ROC curve
- Resampling – e.g., SMOTENN

Processes

- Hyper-Parameter Tuning: Grid Search or Randomized Search
- Cross-Validation
- Explainability: SHAP / LIME Model

Interim Results

- Fine-tuned a pre-trained DistilBERT model

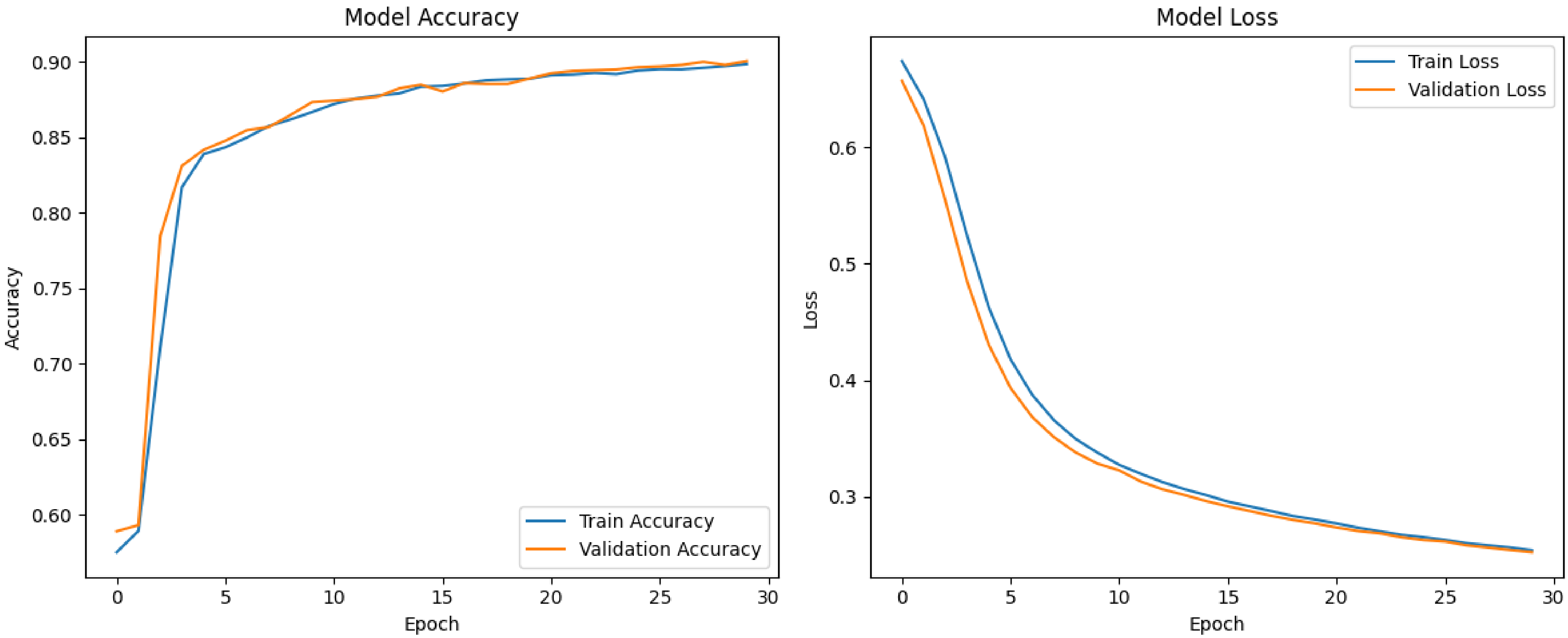

Figure 2: Fine-tuned DistilBERT Model's Performance

Thank you!

Any Questions?

Our Repo!
Feel free to comment here!