Natural Language Processing

Assignment 2

Transformer models - Labelled Financial News Data

Ricardo Nunes Ribeiro - up202310095@edu.fe.up.pt

Matteo Tancredi - up202311682@edu.fe.up.pt

Tessa de Jong - up202311468@edu.fe.up.pt

1. **What data does the notebook focus on?**
    a. Labelled Financial News Data
    b. Economic times (India)

2. **What is the purpose of this notebook?**
    a. Sentiment analysis
    b. Categorizing as positive / negative
    c. To develop accurate sentiment analysis models tailored to the financial domain
    d. *Implement Hugging Face transformer models for Sentiment Analysis*

Model used from Hugging Face:

- mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis
- distilbert-base-uncased
- ProsusAI/finbert

    - 1 e 3 are already fine-tuned on financial news
    - And for sentiment analysis

DistilRoberta-financial-sentiment

- This model is a distilled version of the RoBERTa-base model
- The dataset consists of 4840 sentences from English language financial news categorised by sentiment

DistilBERT base model (uncased)

- Transformers model, smaller and faster than BERT
- Pretrained using BERT as teacher

ProsusAI/finbert

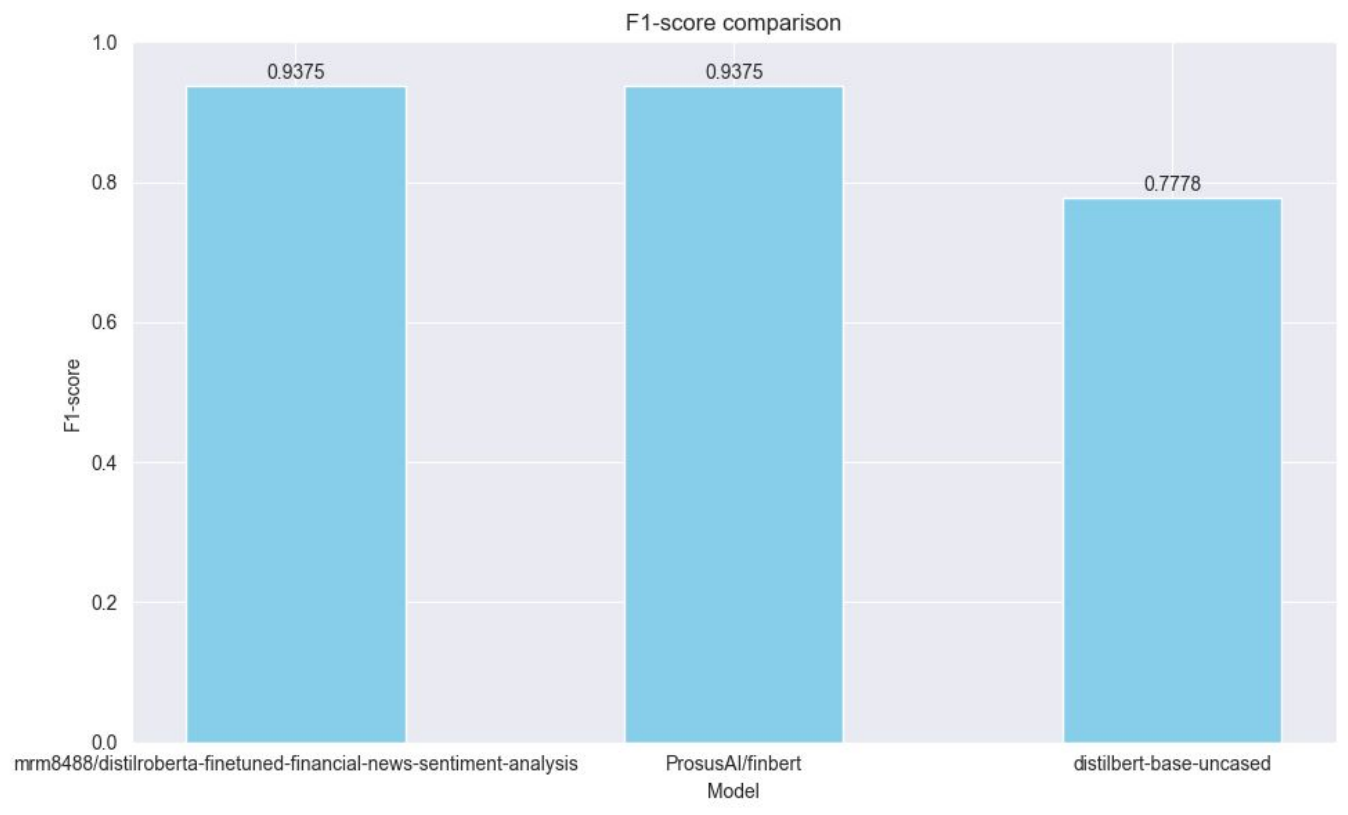
- Pre-trained NLP model to analyze sentiment of financial text.
- Built using BERT language model
- Uses large financial corpus and thereby fine-tuning it for financial sentiment classification.

Steps taken to use and fine-tune the models

1. Tokenize *combined_text* (*headline* + *synopsis* + *text*)
2. Create a dataset split into train, validation and test sub-datasets
3. Loading models
4. Fine-tuning the models
5. Predict test dataset
6. Evaluate the models

Model	Validation Loss	Training accuracy
distilroberta-finetuned-financial-news-sentiment-analysis	0.1981	0.9250
distilbert-base-uncased	0.6354	0.725
ProsusAI/finbert	0.4363	0.875

Fine-tuning F1-score results comparison



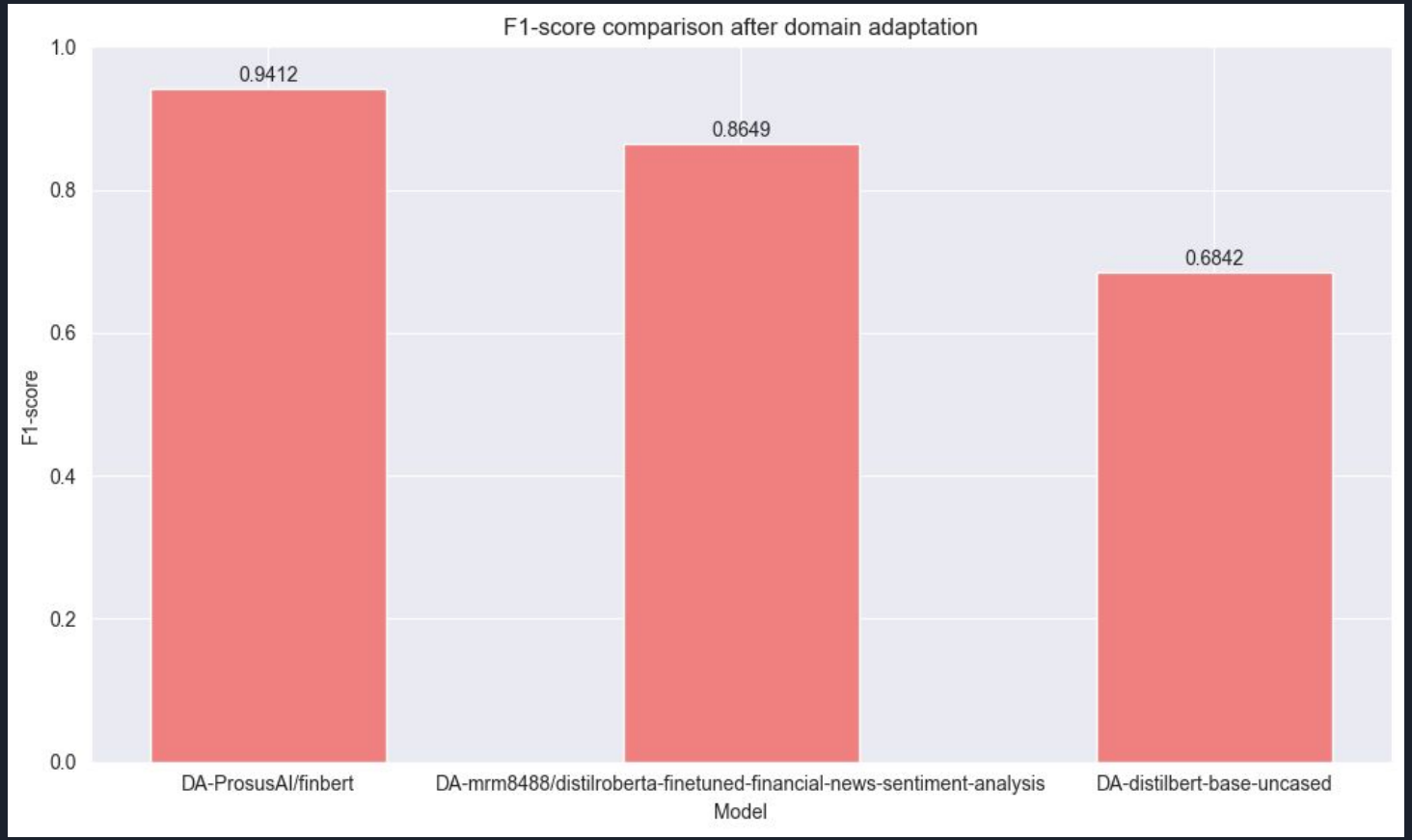
Domain Adaptation

1. perplexity
2. Fine-tuning models
3. Predict and evaluate the models

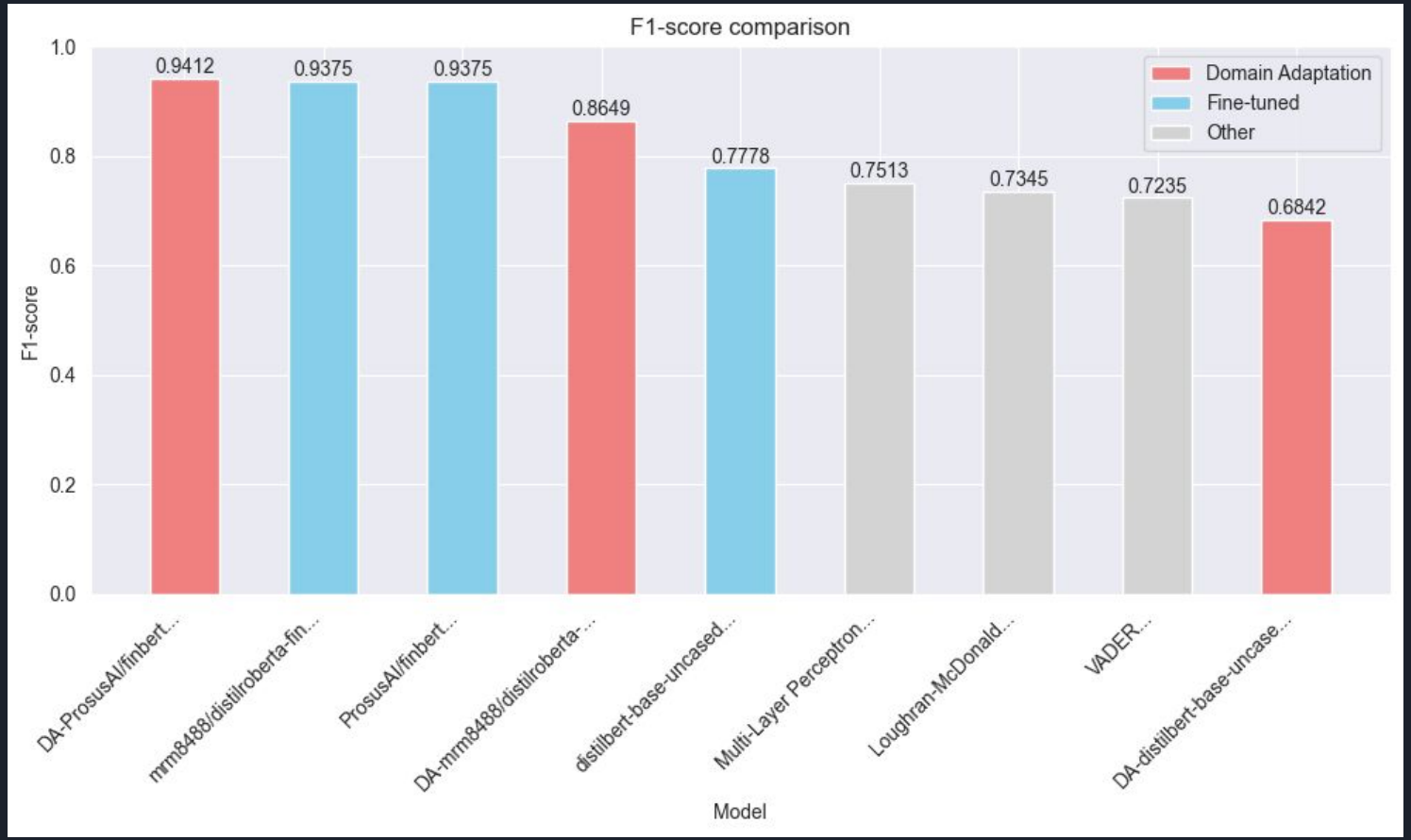
Model	Before training	After training
distilroberta-finetuned-financial-news-sentiment-analysis	18945522.29	393.99
distilbert-base-uncased	14.32	7.55
ProsusAI/finbert	55414.04	165.91

Perplexity values before and after training

F1-score results comparison after domain adaptation



Full F1-score comparison - Assignments 1 and 2



Thank you for your attention

Ricardo Nunes Ribeiro - up202310095@edu.fe.up.pt

Matteo Tancredi - up202311682@edu.fe.up.pt

Tessa de Jong - up202311468@edu.fe.up.pt