Technical Report: Text Classification in Medical and Financial Domains

1. Consolidated Results Table

Pipeline	Domain	Accuracy	Precision (weighted)	Recall (weighted)	F1 (weighted)
Pipeline A	Medical	0.5891	0.5961	0.5891	0.5884
Pipeline B	Medical	0.5914	0.6007	0.5914	0.5908
Pipeline C	Medical	0.5619	0.5630	0.5619	0.5612
Pipeline A	Financial	0.7876	0.7911	0.7876	0.7674
Pipeline B	Financial	0.6737	0.6145	0.6737	0.6182
Pipeline C	Financial	0.8456	0.8434	0.8456	0.8441

---

2. Domain-Specific Discussion

Medical Domain

- **Pipeline A (TF-IDF)**:
  - Baseline method capturing basic word-level frequency patterns.
  - Performs reasonably well on common medical terms but struggles with **contextual dependencies**, abbreviations, and polysemous terms (e.g., “BP” for blood pressure vs. business process).
- **Pipeline B (Word2Vec)**:
  - Slightly better than TF-IDF. Captures **semantic similarity** between words, which helps with synonyms and medical acronyms.
  - Averaged embeddings may lose fine-grained context across long abstracts, limiting performance improvement.
- **Pipeline C (BioBERT)**:
  - Domain-specific Transformer captures **contextual meaning and long-range dependencies**.

    - Surprisingly, F1 is slightly lower than Pipeline B, potentially due to **limited training data** for classifier on high-dimensional embeddings or lack of fine-tuning.

Observation: Word embeddings provide a practical semantic boost over TF-IDF, while Transformers require careful handling or fine-tuning to fully exploit contextual understanding.

Financial Domain

- **Pipeline A (TF-IDF)**:
    - Strong baseline due to **repetitive domain-specific terms**.
    - Captures keywords indicating sentiment, e.g., “profit,” “loss,” “growth.”
- **Pipeline B (Word2Vec)**:
    - Weaker performance than TF-IDF. Averaged vectors may dilute subtle sentiment signals in financial text.
- **Pipeline C (FinBERT)**:
    - Best performance by a significant margin.
    - Captures **nuanced sentiment and market language** effectively (e.g., “increase capacity” vs. “zero pre-tax profit”), thanks to domain-specific pretraining.

Observation: Transformers excel in capturing subtle sentiment nuances, making them the preferred choice in financial NLP tasks.

---

3. General Conclusion and Recommendations

- **Trade-offs**:
    - **TF-IDF**: Fast, low memory usage, good for keyword-based classification.
    - **Word2Vec**: Slightly slower, semantic understanding improves performance on complex domains but may dilute document-level meaning.
    - **Transformers**: Highest computational cost, GPU required for efficiency, best for **contextual understanding and domain-specific subtleties**.
- **Recommendations**:

- **Medical domain** with limited data: Word2Vec is a good compromise between performance and resource usage. BioBERT can be used if fine-tuning or larger datasets are feasible.
- **Financial domain**: FinBERT is recommended due to its strong performance in sentiment classification, despite higher computational cost.
- For **projects with constraints** on data size or hardware: Start with TF-IDF or Word2Vec. Use Transformers only when **GPU resources and data size justify the cost**.