

# Legal Document Classification using BERT and TF-IDF Models

**Team Members: Sara Samir Nasr - 22101068 | Shams Abdelhalim - 22101040**

## 1. Introduction

Legal document classification is an essential task in law-related NLP applications. It helps in automating the organization of documents like contracts, agreements, and corporate filings. This project explores and compares the effectiveness of two NLP techniques: a fine-tuned BERT model and a traditional TF-IDF + Logistic Regression approach, applied to the LEX\_GLUE (LEDGAR) dataset.

## 2. Related Work

Prior research has shown that transformer-based models, particularly BERT, perform exceptionally well in domain-specific classification tasks. However, traditional machine learning models like Logistic Regression with TF-IDF still offer strong baselines, especially in resource-limited settings.

## 3. Methodology

**Dataset:** We used the lex\_glue dataset, specifically the ledgar subset. It contains legal text snippets labeled with their document type.

**Model 1:** TF-IDF + Logistic Regression

- Vectorization: `TfidfVectorizer(max_features=3000)`
- Classifier: `LogisticRegression(max_iter=1000)`

**Model 2:** Pretrained BERT (bert-base-uncased)

- Tokenization: `AutoTokenizer.from_pretrained(...)`
- Classification: `TFBertModel` with a dense layer for classification.

**Tools:** Python, Hugging Face Transformers, Datasets, Scikit-learn, Matplotlib.

## 4. Experiments

We evaluated models on 1000 training and 200 test examples using accuracy, precision, recall, and F1

# Legal Document Classification using BERT and TF-IDF Models

score.

TF-IDF + Logistic Regression:

- Accuracy: 0.53
- Macro Precision: 0.41
- Macro Recall: 0.53
- Macro F1: 0.43

BERT Model:

- Accuracy: 0.67
- Macro Precision: 0.50
- Macro Recall: 0.56
- Macro F1: 0.50

## 5. Discussion

BERT outperforms TF-IDF + Logistic Regression across all metrics. It better captures contextual representations in legal texts, but is more resource-intensive. TF-IDF is faster but less powerful.

Challenges included:

- BERT memory limits on Google Colab
- Label inconsistencies in test vs. train
- Preprocessing of long legal texts

## 6. Conclusion

This project showed BERT's superior performance for classifying legal documents in the LEDGAR dataset. Future work could explore legal-specific BERT models or new datasets like scotus or ecthr\_a.