Fake News Detection

Mridul Sharma

GROUP ID: 37

Problem Description & Motivation

The Target

Reduce the spread of misinformation

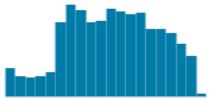
Do it automatically without manual intervention

Impact

Shaping public opinions (politics, health, events)

Goal

Develop an ML model to flag fake vs. real news

title title of news article	text body text of news article	subject subject of news article	date publish date of news article
17903 unique values	[empty] 3% AP News The regu... 0% Other (22851) 97%	News 39% politics 29% Other (7590) 32%	 2015-03-30 2018-02-18
Donald Trump Sends Out Embarrassing New Year's Eve Message; This is Disturbing	Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that. Instead, he had...	News	December 31, 2017
Drunk Bragging Trump Staffer Started Russian Collusion Investigation	House Intelligence Committee Chairman Devin Nunes is going to have a bad day. He s been under the as...	News	December 31, 2017
Sheriff David Clarke Becomes An Internet Joke For Threatening To Poke People 'In The Eye'	On Friday, it was revealed that former Milwaukee Sheriff David Clarke, who was being considered for ...	News	December 30, 2017
Trump Is So Obsessed He Even Has Obama's Name Coded Into His	On Christmas day, Donald Trump announced that he	News	December 29, 2017

Dataset Overview

Fake News detection dataset

- Dataset separated in two files:
  - Fake.csv (23502 fake news article)
  - True.csv (21417 true news article)
- Dataset columns:
  - Title: title of news article
  - Text: body text of news article
  - Subject: subject of news article
  - Date: publish date of news article

Data Preprocessing

Data Loading

- Fake.csv + True.csv

Cleaning

- Binary labels (0 = fake, 1 = real)
- Remove irrelevant information

Splitting

- Train (70%)
- Validation (15%)
- Test (15%)

- **Merging & Labeling:** Combine Fake.csv & True.csv into one dataset
- **Labels:** Add binary label column (0 = fake, 1 = real)
- **Cleaning:** Focus on text content (remove irrelevant info)
- **Text Only:** Use article text (and optionally title) as input features
- **Splitting:** Stratified split into Train (70%), Validation (15%), Test (15%)

Baseline Model: TF-IDF + Linear Classifier

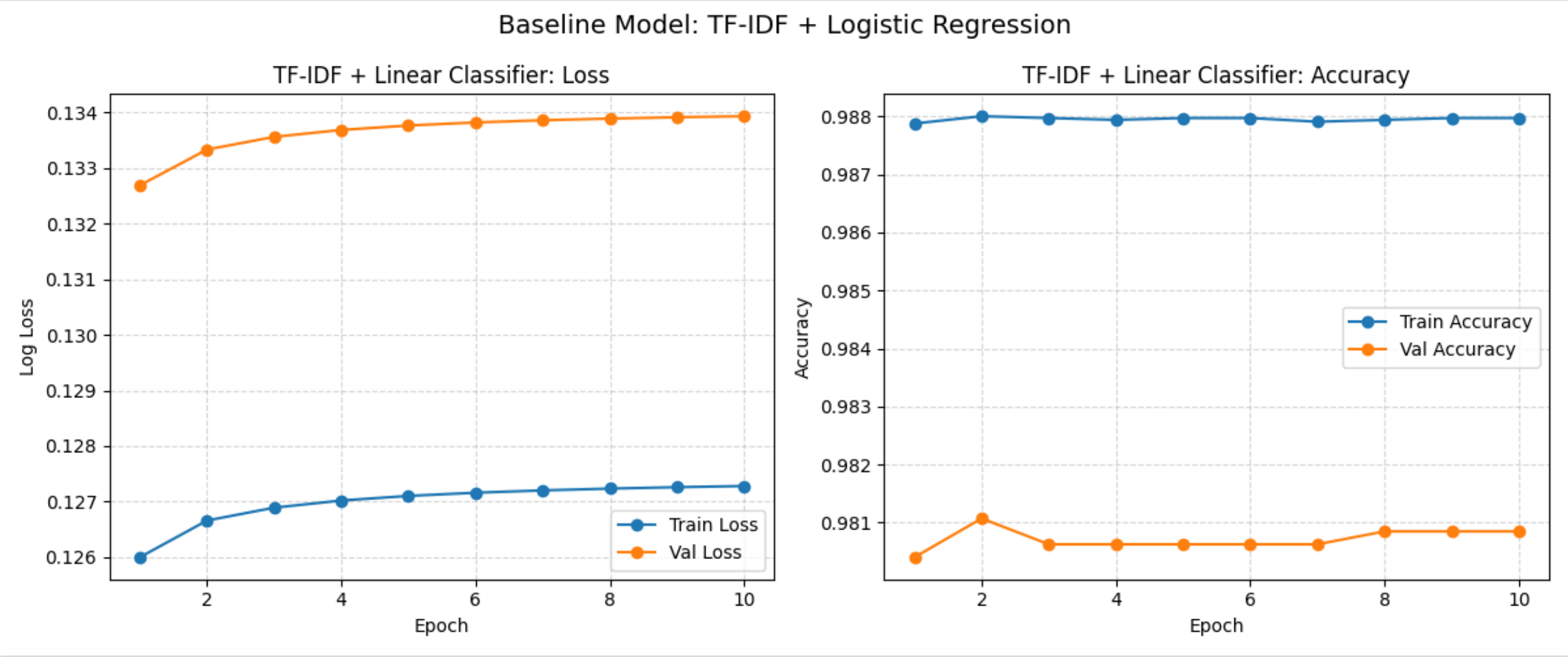
Approach: Classical machine learning baseline

Vectorization: Convert text to TF-IDF feature vectors

Classifier: Logistic Regression (or linear SVM) on TF-IDF features

Purpose: Simple, fast baseline without deep learning

Expectation: Reasonable accuracy, sets a benchmark for complex modls


Baseline Model: TF-IDF + Logistic Regression
TF-IDF + Linear Classifier: Loss
Log Loss
0.134
0.133
0.132
0.131
0.130
0.129
0.128
0.127
0.126
2
4
6
8
10
Epoch
Train Loss
Val Loss
TF-IDF + Linear Classifier: Accuracy
Accuracy
0.988
0.987
0.986
0.985
0.984
0.983
0.982
0.981
2
4
6
8
10
Epoch
Train Accuracy
Val Accuracy

Custom Neural Network Model

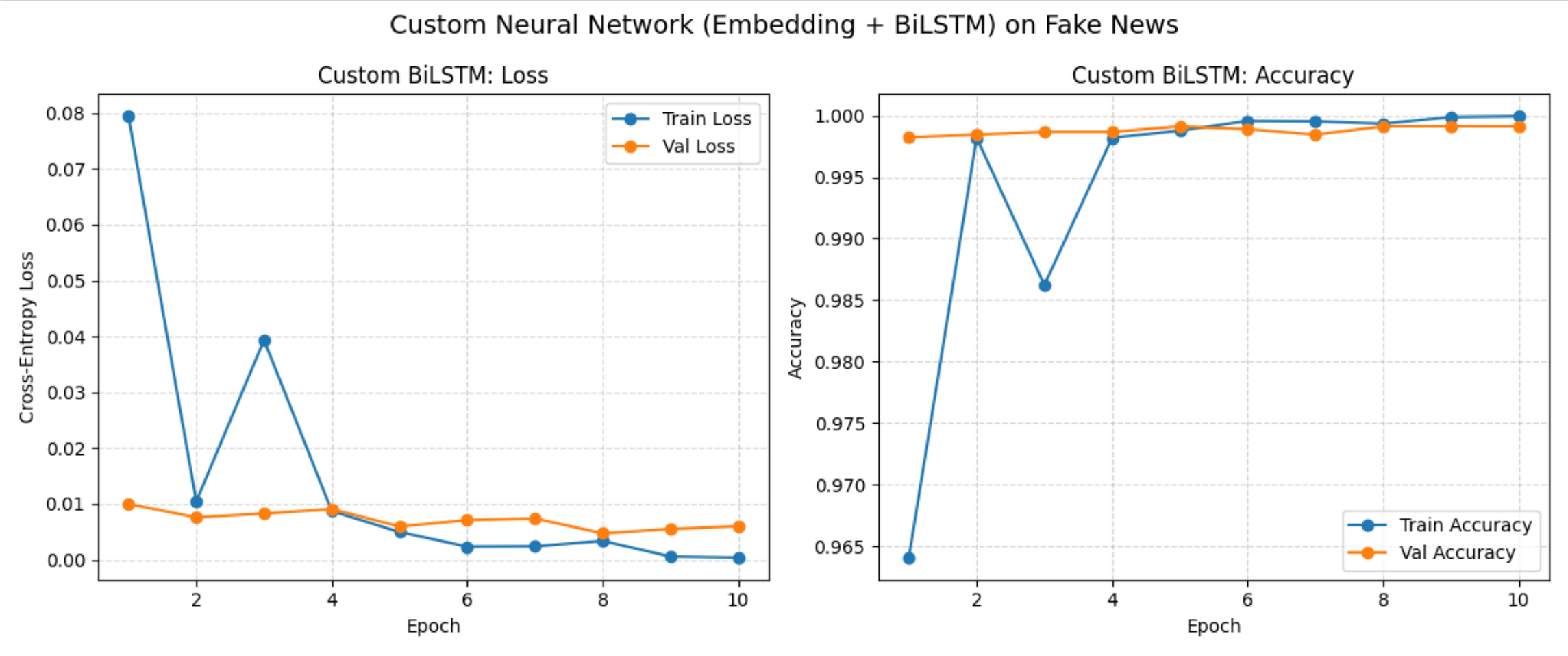
Architecture: Lightweight text classifier (embedding + encoder + dense output)

Example Design: Embedding layer → BiLSTM (or 1D CNN) → fully-connected layer

Parameters: Trained from scratch on our dataset (no external data)

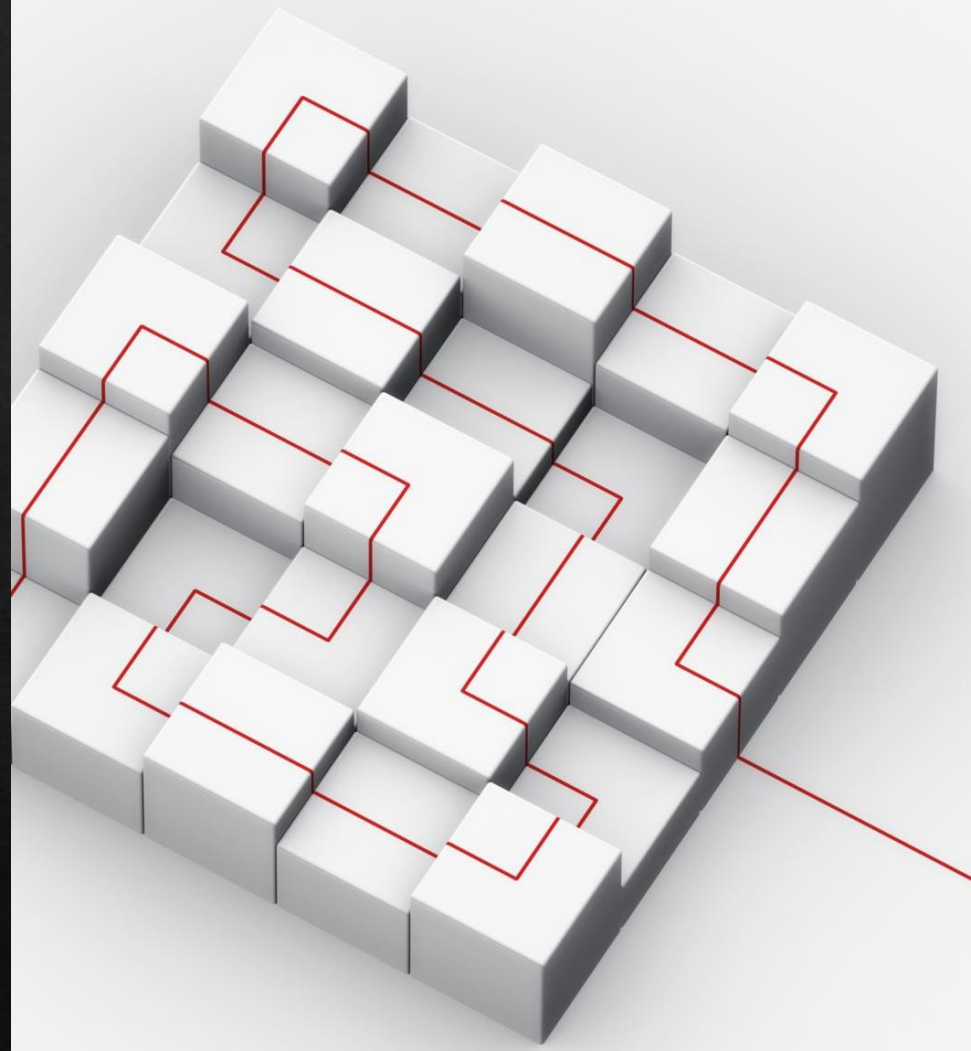
Rationale: Test a “pure” neural approach without pre-training

Expectation: Improve over baseline by learning feature representations (but smaller than Transformer)


Custom Neural Network (Embedding + BiLSTM) on Fake News
Custom BiLSTM: Loss
Train Loss
Val Loss
Cross-Entropy Loss
Epoch
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
2
4
6
8
10
Custom BiLSTM: Accuracy
Train Accuracy
Val Accuracy
Accuracy
Epoch
1.000
0.995
0.990
0.985
0.980
0.975
0.970
0.965

Pre-Trained Transformer Model (DistilBERT Fine-Tuned)

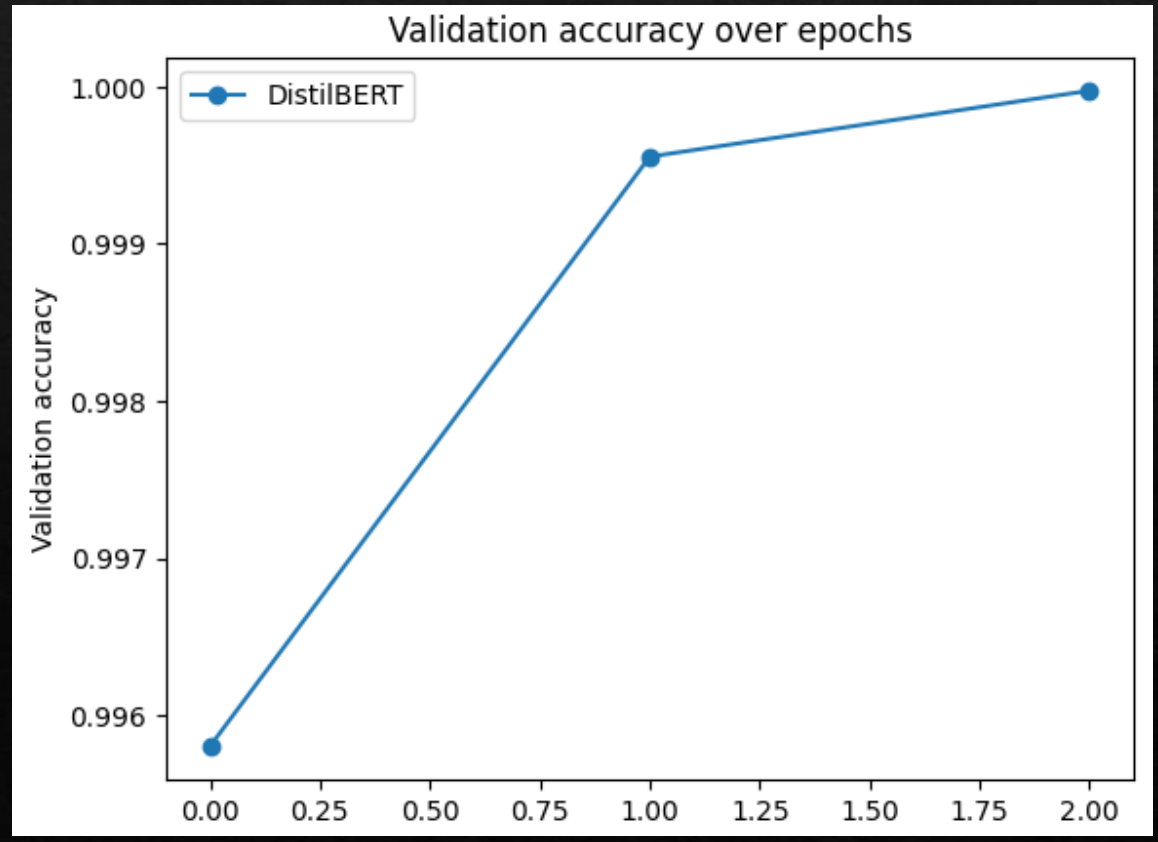
- **Model: DistilBERT** (Transformer-based language model) fine-tuned for fake news classification
- **Architecture:** 6-layer Transformer encoder (66M parameters) + classification head
- **Pre-training:** Already learned general language patterns on huge text corpora
- **Fine-tuning:** Additional training on our news dataset to adapt to fake vs real detection
- **Rationale:** Leverage transfer learning for improved accuracy with modest training effort
- **Expectation:** Highest performance, at cost of much larger model size and compute



Accuracy and Confusion Matrix

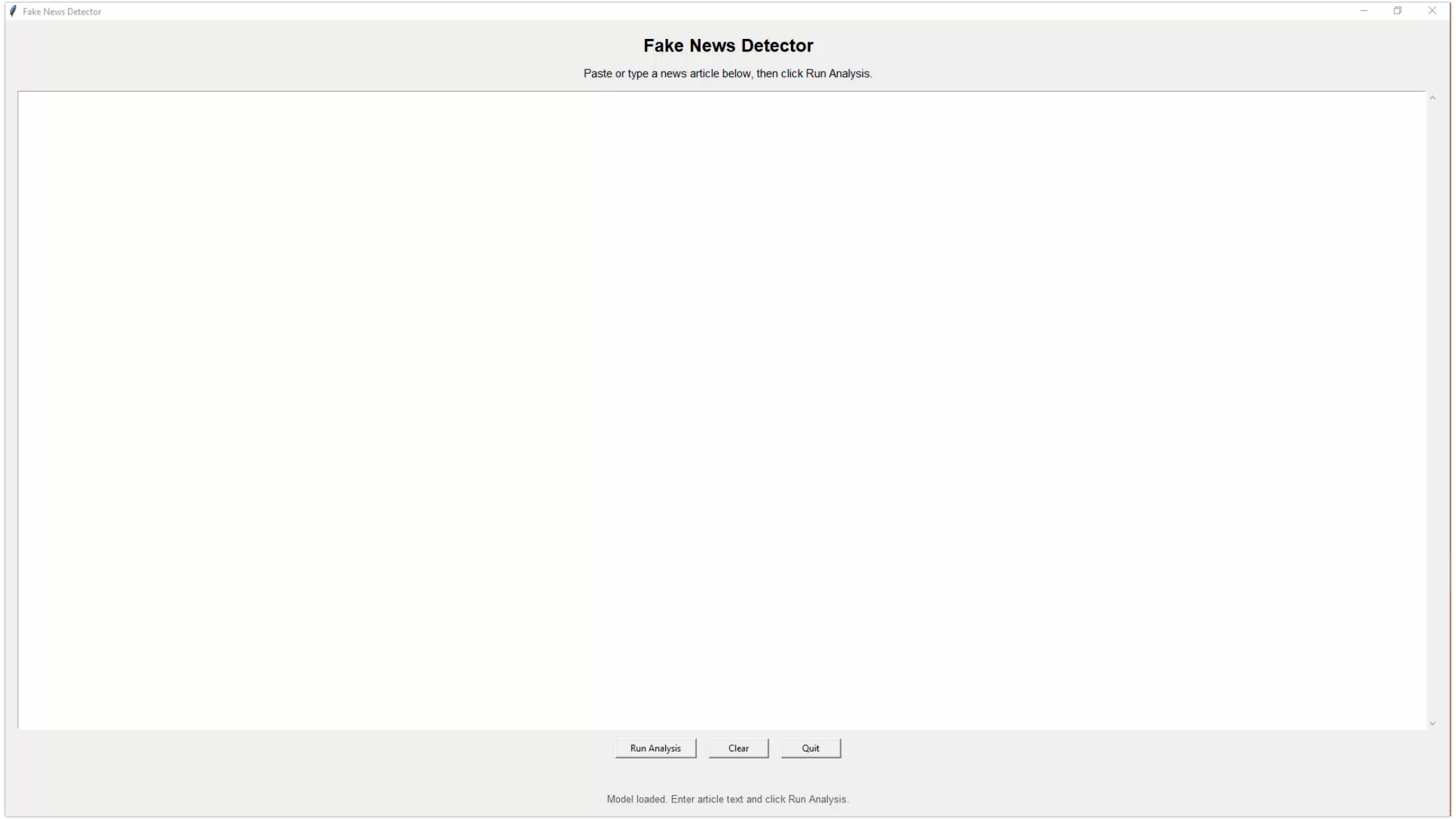
	Precision	Recall	F1-Score	Support
Fake	0.9997	1	0.9999	3523
Real	1	0.9997	0.9998	3212
Accuracy			0.9999	6735

3523	0
1	3211



Model Comparison & Results

- **Accuracy (Test Set):** Baseline ≈ *98%*, Small NN ≈ *99%*, DistilBERT ≈ *99.999%+*
- **Precision/Recall/F1:** Higher for more complex models (nearly perfect for DistilBERT)
- **Model Size:** Baseline ~0.1M params; Small NN a few M; DistilBERT ~66M params
- **Training Time:** Baseline (less than a minute), Small NN (two or three minutes), DistilBERT (~10 minutes, ~3 epochs fine-tuning)
- **Performance vs Complexity:** Larger models = better metrics, but more resources required


Fake News Detector
Fake News Detector
Paste or type a news article below, then click Run Analysis.
Run Analysis
Clear
Quit
Model loaded. Enter article text and click Run Analysis.

Summary of Findings & Discussion

- **Baseline vs Advanced:** Simple model was decent but missed nuanced fakes
- **Neural Network:** Learned from data, improved accuracy (captured context)
- **Transformer:** Significantly best results, caught almost all fakes reliably
- **Trade-offs:**
  - Simplicity (baseline): fast, interpretable, low resource
  - Power (Transformers): high accuracy, but large and resource-intensive
- **Misclassifications:** Rare errors (ex. a real article flagged as fake)