



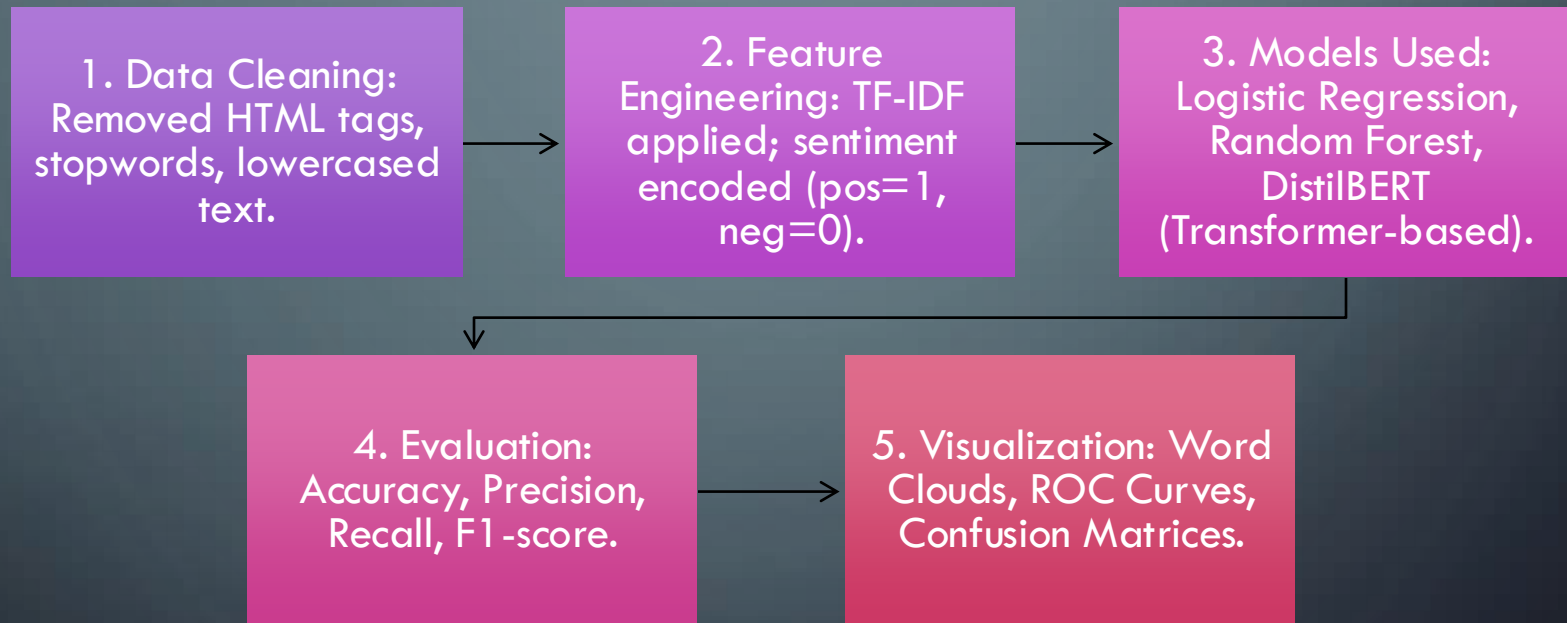
SENTIMENT ANALYSIS OF IMDB MOVIE REVIEWS

CARLOS ESCAMILLA,
TYLER HEFLIN,
ANDRES MELENDEZ,
DELILAH SLABAUGH
14 JULY 2025

BUSINESS PROBLEM / HYPOTHESIS

- Determine whether IMDb movie reviews express positive or negative sentiments using NLP and ML.
- Hypothesis: Classification accuracy can exceed 85% using classical ML techniques.

METHODOLOGY



IMDB DATASET OVERVIEW

50,000 labeled
movie reviews from
Kaggle

Balanced dataset:
25,000 positive,
25,000 negative

Reviews are pre-
divided into training
and test sets

Average review
length: ~230 words

Challenges:
sarcasm, ambiguous
phrasing, class
overlap

MODELING STRATEGY & TUNING

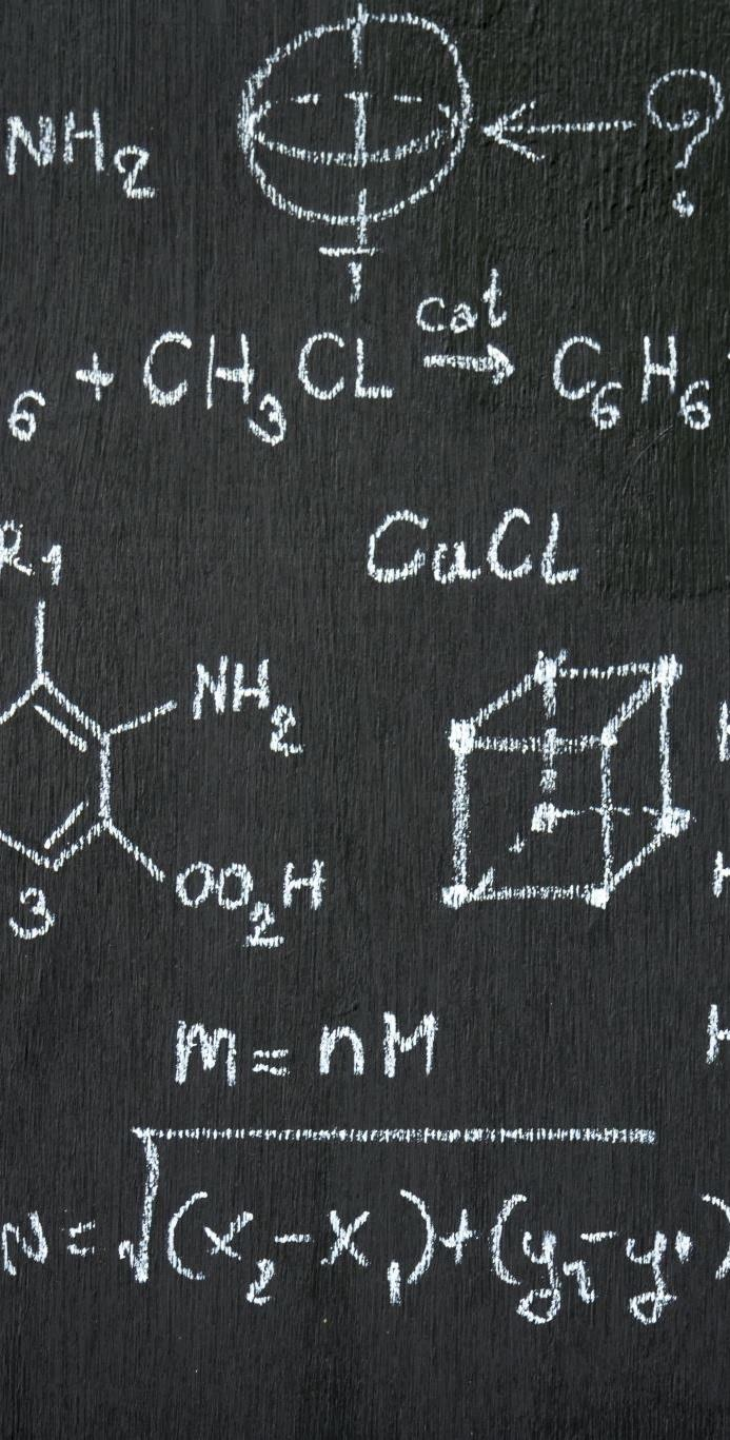
Used train/test split (80/20)

GridSearchCV for hyperparameter tuning

Avoided overfitting via cross-validation

DistilBERT used HuggingFace Transformers

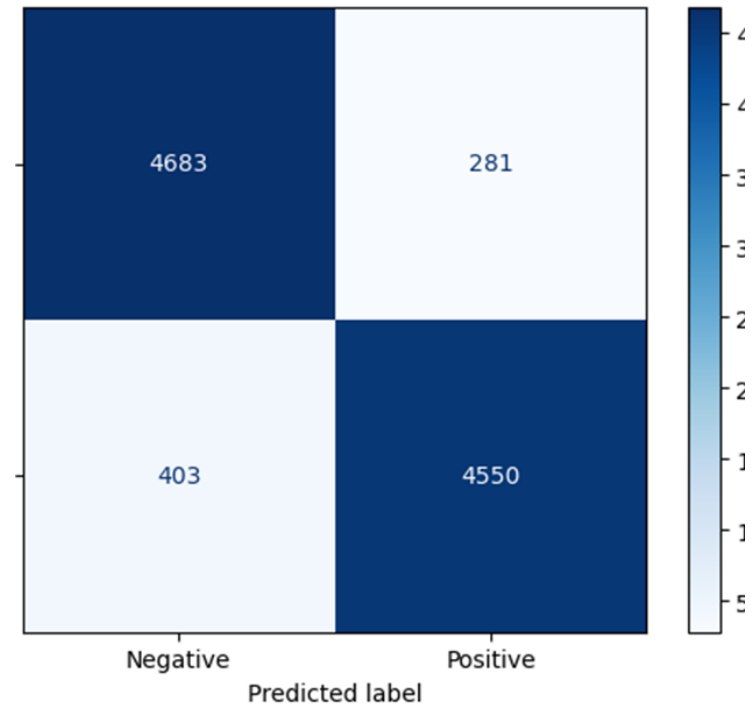
Model training: Logistic Regression (1 min), DistilBERT (~20 mins w/GPU)

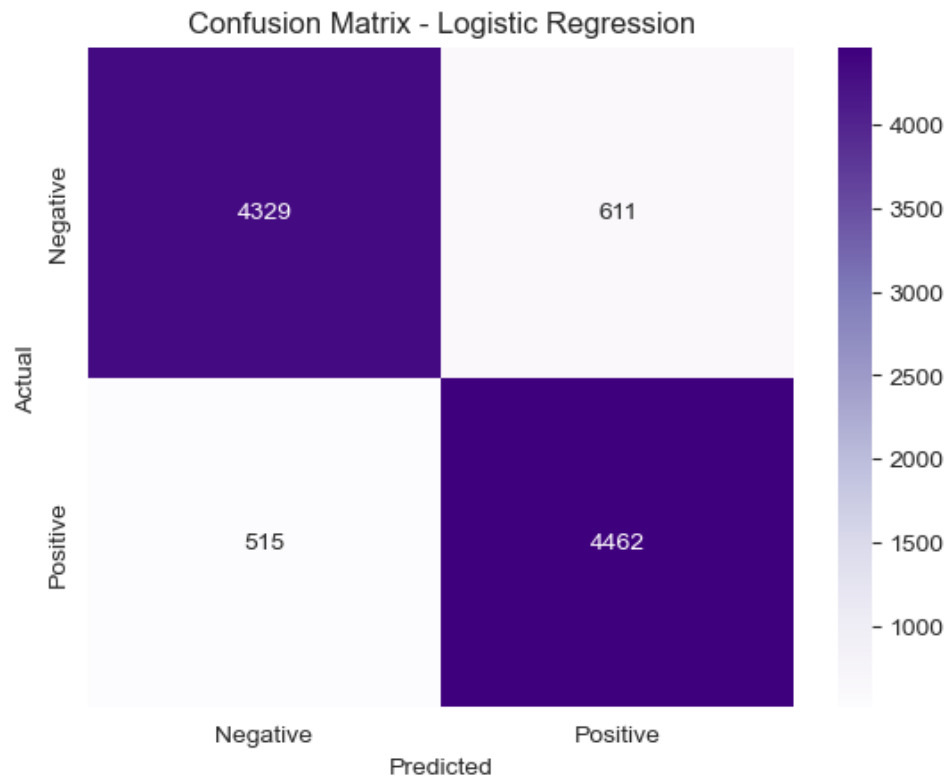


RESULTS SUMMARY

- **DistilBERT (Transformer-based):** Accuracy = 91%
 - **Logistic Regression:** Accuracy = 88%, AUC = 0.96
 - **Random Forest:** Accuracy = 84%
- DistilBERT outperformed other models overall.
- Strong separation between classes demonstrated via ROC and Confusion Matrix.

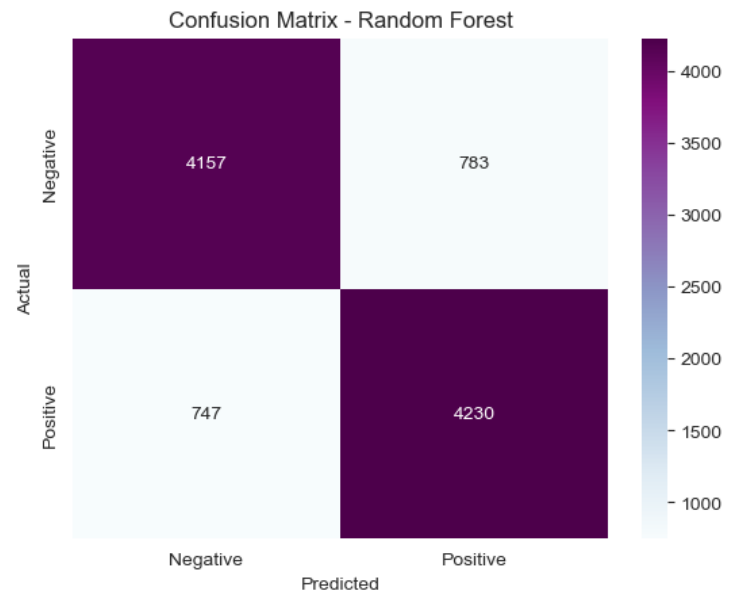
CONFUSION MATRIX – DISTILBERT

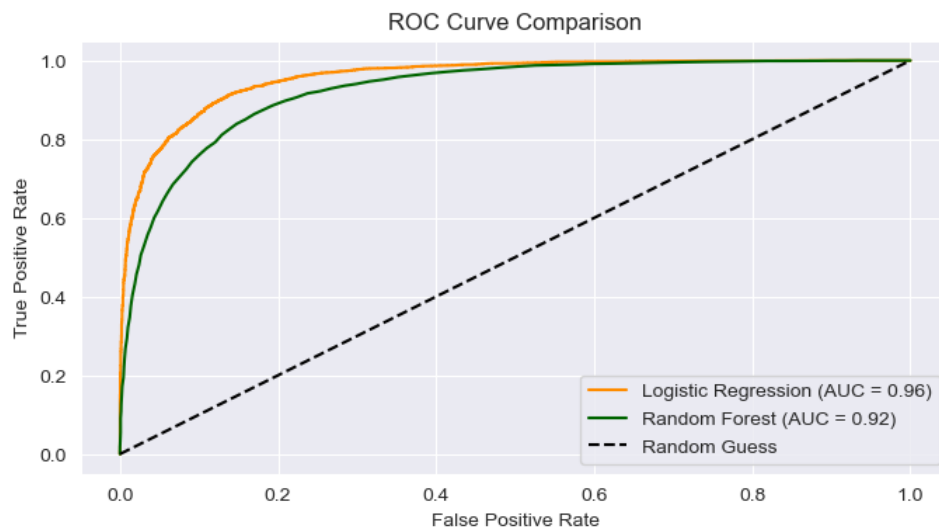




CONFUSION MATRIX – LOGISTIC REGRESSION

CONFUSION MATRIX – RANDOM FOREST





ROC CURVE COMPARISON

A dense word cloud where words are arranged based on their frequency or importance. The largest words, occupying the most space, are "character", "time", "scene", "movie", "story", "way", "show", "think", "director", "life", "see", "must", "bad", "play", "watch", "funny", "girl", "really", "little", "love", "lot", "someone", "looking", "another", "end", "look", "part", "dialogue", "played", "better", "anything", "original", "interesting", "great", "well", "got", "something", "much", "people", "actor", "plot", "point", "reason", "know", "day", "though", "man", "enough", "maybe", "with", "start", "woman", "least", "friend", "role", "music", "never", "good", "script", "still", "going", "life", "help", "want", "horning", "claw", "coldly", "expect", "ally", "time", "light", "more", "id", "take", "kind", "guy", "set", "need", "fan", "make", "fact", "kid", "wonder", "bit", "thing", "thought", "probably", "sure", "moment", "color", "se", "believe", "come", "shot", "find", "performance", "two", "said", "first", "go", "work", "made", "turn", "father", "put", "simply", "many", "stage", "stupid", "miss", "ser", "actually", "nearly", "happy", "stare", "sit", "even". Smaller words include "comedy", "female", "male", "script", "still", "going", "life", "help", "want", "horning", "claw", "coldly", "expect", "ally", "time", "light", "more", "id", "take", "kind", "guy", "set", "need", "fan", "make", "fact", "kid", "wonder", "bit", "thing", "thought", "probably", "sure", "moment", "color", "se", "believe", "come", "shot", "find", "performance", "two", "said", "first", "go", "work", "made", "turn", "father", "put", "simply", "many", "stage", "stupid", "miss", "ser", "actually", "nearly", "happy", "stare", "sit", "even".

WORD CLOUDS – SENTIMENT WORDS

CONCLUSION

- ✓ Hypothesis confirmed with 91% accuracy using DistilBERT.
- ✓ Classical ML + TF-IDF = effective and interpretable.
- ✓ Strong model generalization with room for deep learning extensions.

Q&A – DISCUSSION POINTS



- Why was Logistic Regression more effective than Random Forest?

- How does TF-IDF compare to Word Embeddings?

- What are some ways to handle neutral sentiments?

- What ethical considerations exist in sentiment analysis?

APPLICATIONS OF SENTIMENT ANALYSIS



Streaming services:
personalized
recommendations



Brands: product
feedback and
customer satisfaction



Media platforms:
detect toxic or
biased content



Business intelligence:
trend analysis from
reviews

FUTURE IMPROVEMENTS & NEXT STEPS



Add a neutral class for 3-way classification



Explore more advanced transformer models (e.g., RoBERTa, BERTweet)



Perform domain-specific fine-tuning (e.g., movie vs. product reviews)



Deploy as a Streamlit dashboard or API for real-time use