

IMDB_Sentiment_Presentation

Search

Delilah Slabaugh

DS

FileHomeInsertDrawDesignTransitionsAnimationsSlide ShowReviewViewHelp

CommentsCatch upPresentViewingShare

1

SENTIMENT ANALYSIS OF IMDB MOVIE REVIEWS

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

2

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.

3

METHODOLOGY

1. Data Cleaning: Remove HTML tags, punctuation, lowercase text.

2. Feature Engineering: TF-IDF vectorization, word embeddings (e.g., word2vec).

3. Model Selection: Logistic Regression, Random Forest, Support Vector Machines, Neural Networks.

4. Evaluation: Accuracy, Precision, Recall, F1 score.

5. Visualization: Word Clouds, ROC Curves, Confusion Matrices.

4

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, ambiguity, phrase overlap

5

MODELING STRATEGY & TUNING

Used train/test split (80/20)

FUTURE IMPROVEMENTS & 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

New add-in from your admin

An add-in was installed to add new capabilities and improve your experience with Office.

Try itNot now

Slide 13 of 13

English (U.S.)

Give Feedback to Microsoft

Notes

83%