



# IMDB

## MOVIE REVIEW

*Sentiment Analysis*

OMER S KOC

NORTHEASTERN UNIVERSITY  
SUPERVISED MACHINE LEARNING AND LEARNING THEORY

# PROBLEM & DATA SET



**Problem:** Classify movie reviews as positive or negative based on their content.

**Data:** Large Movie Review Dataset on Stanford AI Lab

- **Input:** Movie reviews
- **Output:** Sentiment labels (positive or negative)

**Goal:** Train a ML model that can accurately predict the sentiment of movie reviews that it has not seen before.

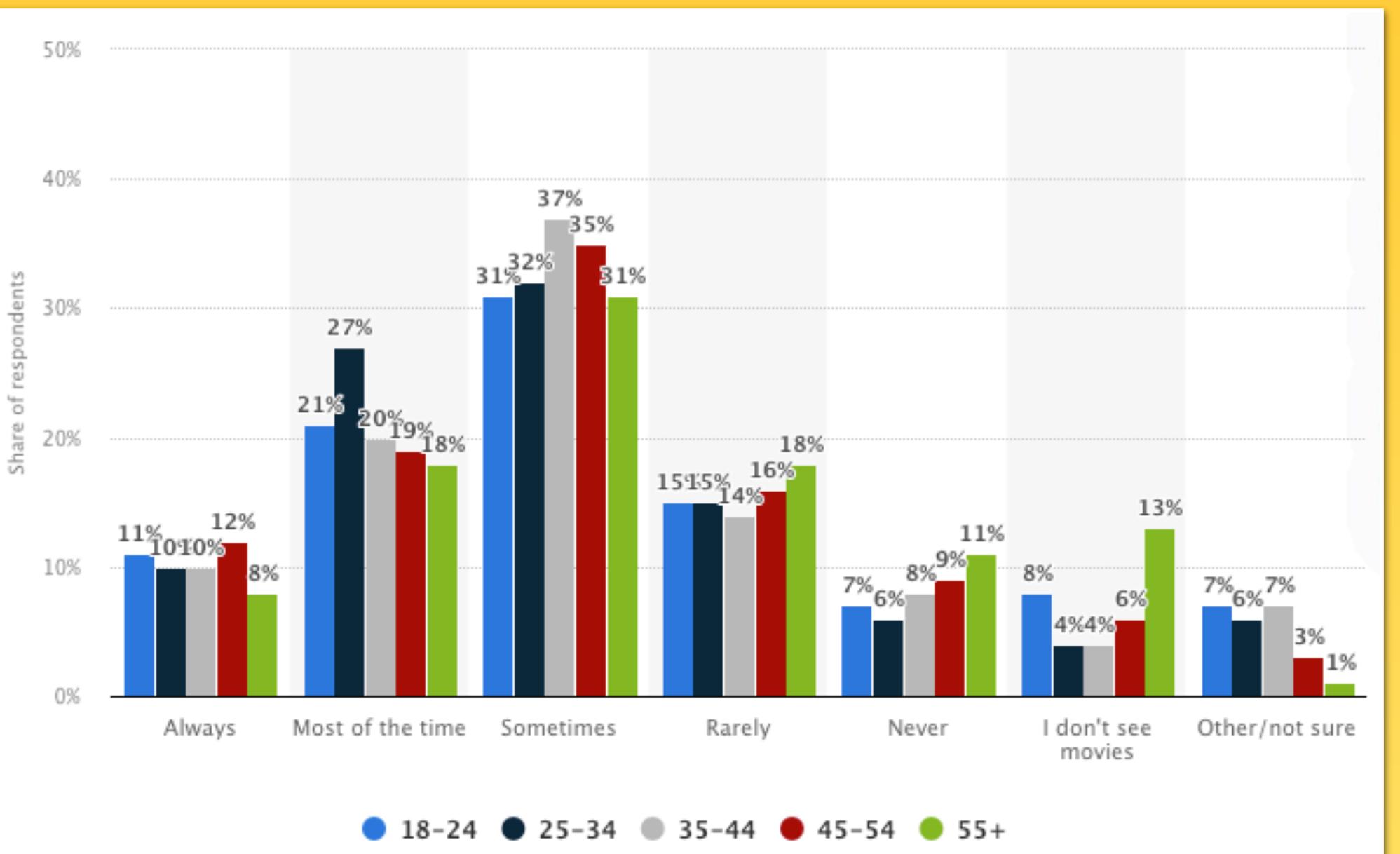


# PROBLEM & DATA SET



'Why did I choose this problem?'

According to a research, 60% of American people said that they looked at the review at least once before watching a movie.



# PROBLEM & DATA SET



## Sample Movie Review

A scene from the movie Titanic showing Jack and Rose dancing. The man is wearing a dark jacket over a white shirt, and the woman is wearing a dark jacket over a white top. They are both smiling and looking at each other.

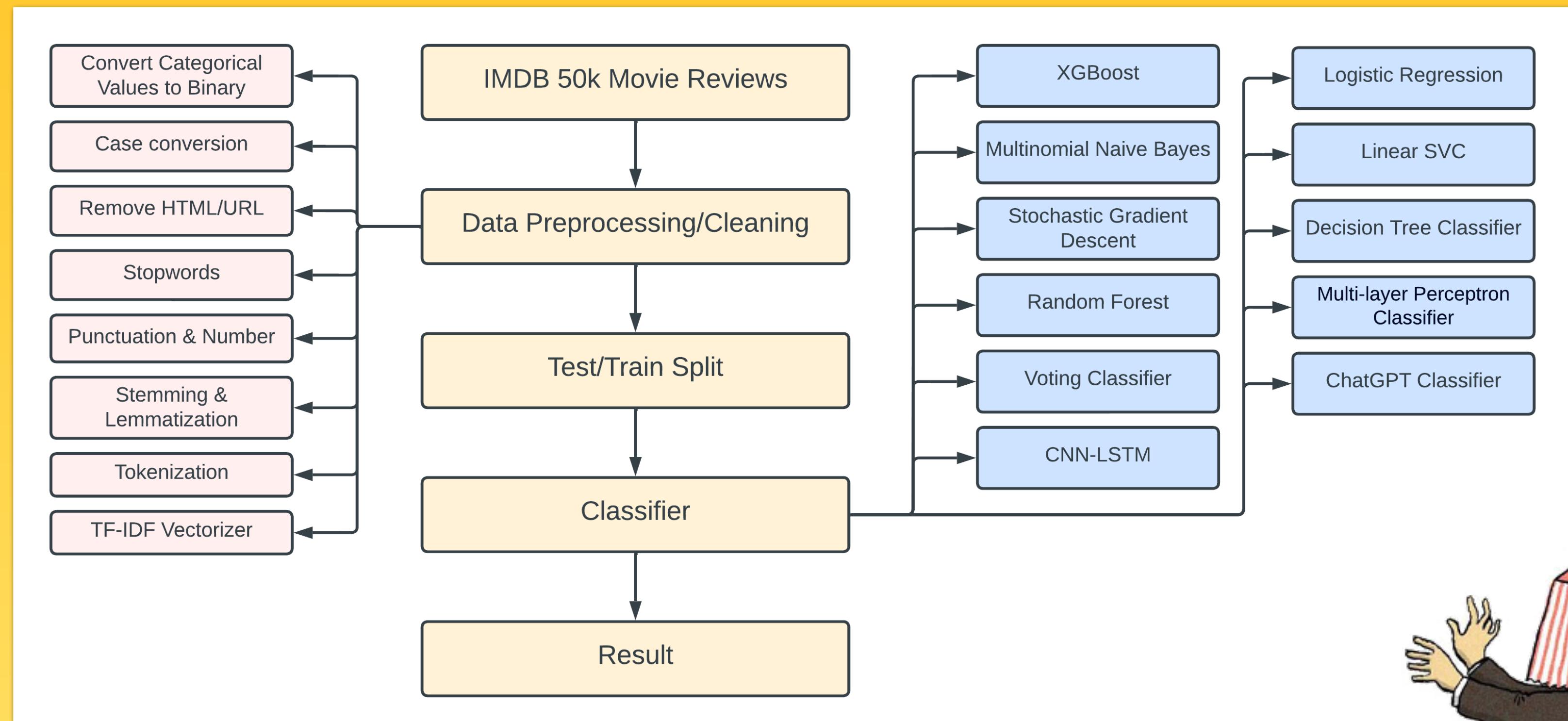
**★ 1/10 Foolish**

There is no way that a metal ship that large could have been sunk in by a measly iceberg, it's unrealistic. why are there even icebergs?



# PROBLEM FORMULATION

How did I approach the problem?



# DATA PREPROCESSING/CLEANING



'How does the review change?'

★ 9/10

**Perpetually Entertaining...**

Xstal 6 June 2022

If you like original gut wrenching laughter you will like this movie. If you are young or old then you will love this movie, hell even my mom liked it.  
<br /><br />  
Great Camp!!!

★ 9/10

**Perpetually Entertaining...**

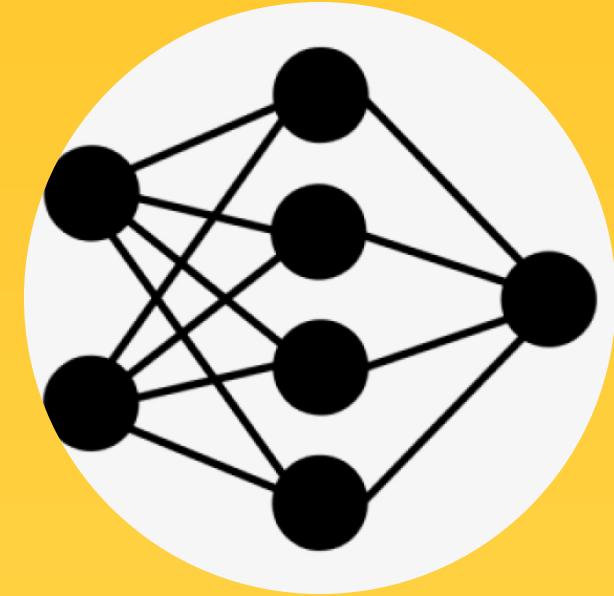
Xstal 6 June 2022

like origin gut wrench laughter like movi young old love movi hell even mom like it



# EVALUATION

Which model is good/bad and why?



CNN-LSTM



Decision Tree



Logistic R.



text-davinci-002



# REFERENCES



- [ai.stanford.edu/~amaas/data/sentiment/](http://ai.stanford.edu/~amaas/data/sentiment/)
- [kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews](https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews)
- Sentiment analysis of film reviews based on CNN-BLSTM (Lu Shang et al 2020)
- Natural Language Processing with Python (O'Reilly)

