

# Movie Ratings Prediction

PRESENTED BY:

Tamasree Sinha  
Vilaysack  
Khonsavanh  
Maria DiPasquale  
Daniel Kletter



## Overview

Using a dataset of the last 10 years' worth of movies, we will revolutionize the audience experience by making it easier to decide what to watch!

## Data Sources

OMDb API = <http://www.omdbapi.com>

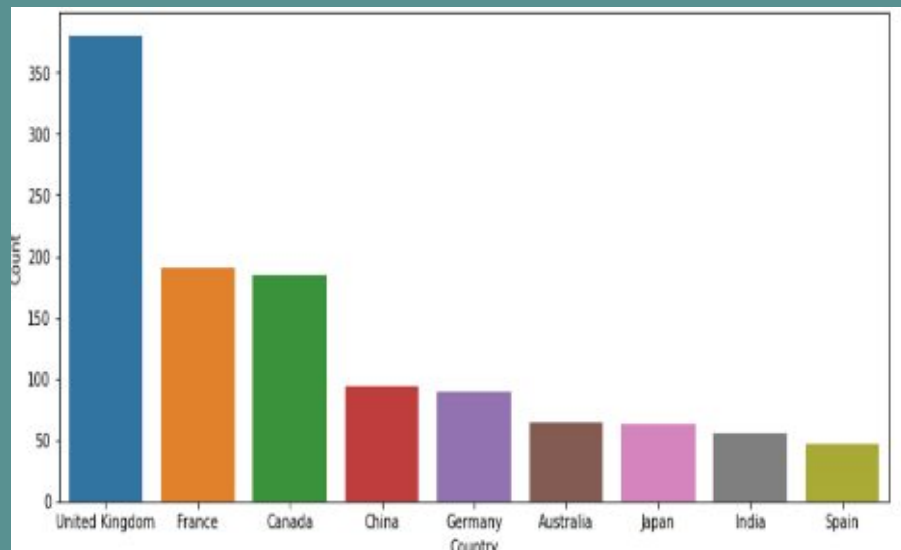
Kaggle = <https://www.kaggle.com/danielgrijalvas/movies>

## Objective

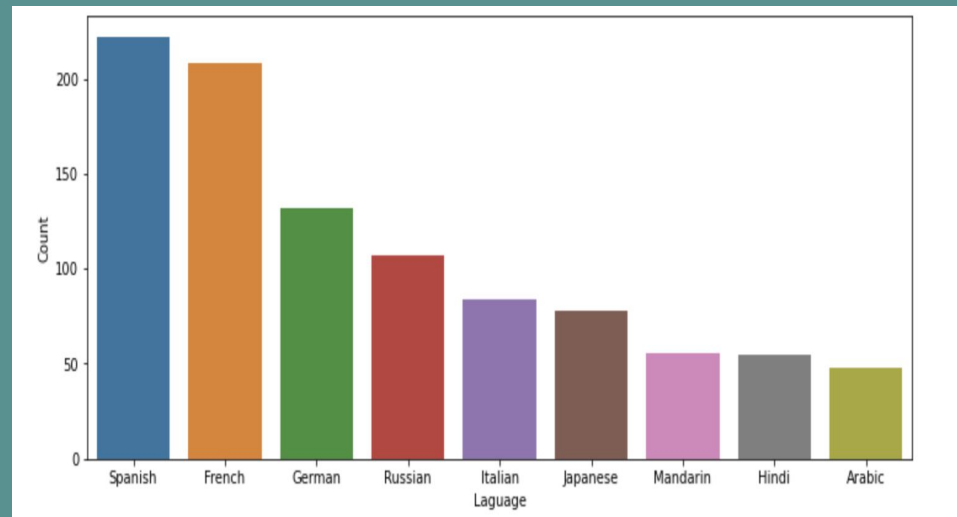
1. Train a supervised model with 70% confidence to predict ratings of proposed movies.
2. Create an historical data analysis with ten years of movie data.
3. Create an interactive web app with movie rating predictions and a dashboard of visualizations.

# 10 Countries with most Movie production (Besides USA) & 10 most used languages in Movies (Besides English)

Countries

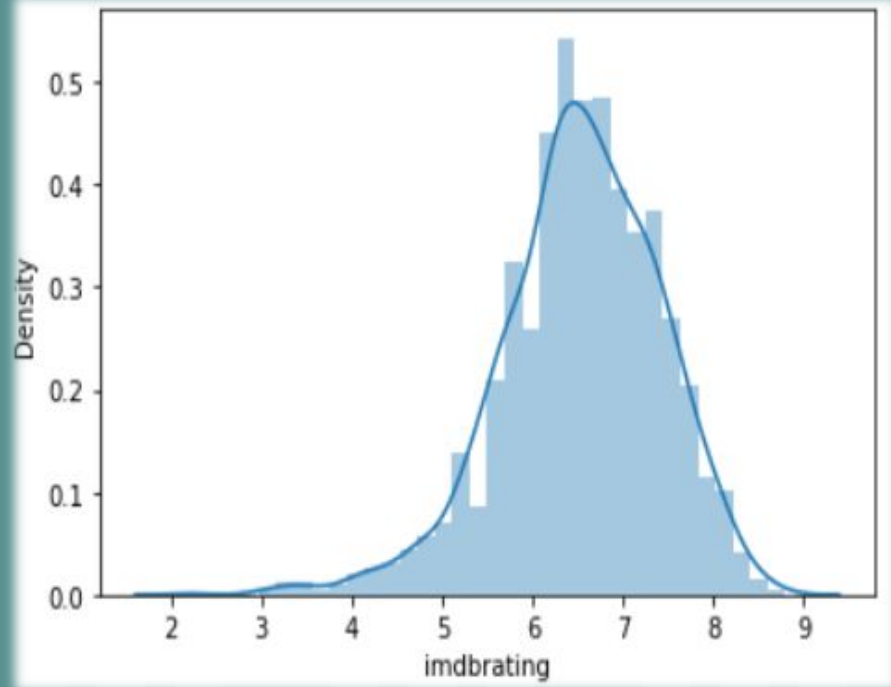


Languages



# Density Plot of IMDB RATINGS

Maximum no. of movies are  
rated by IMDB in the range  
5-8.



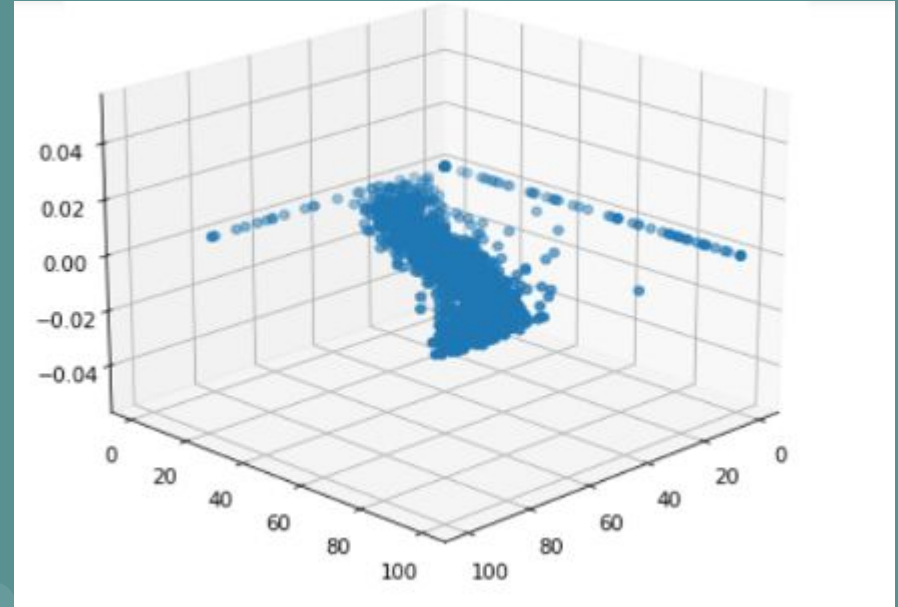
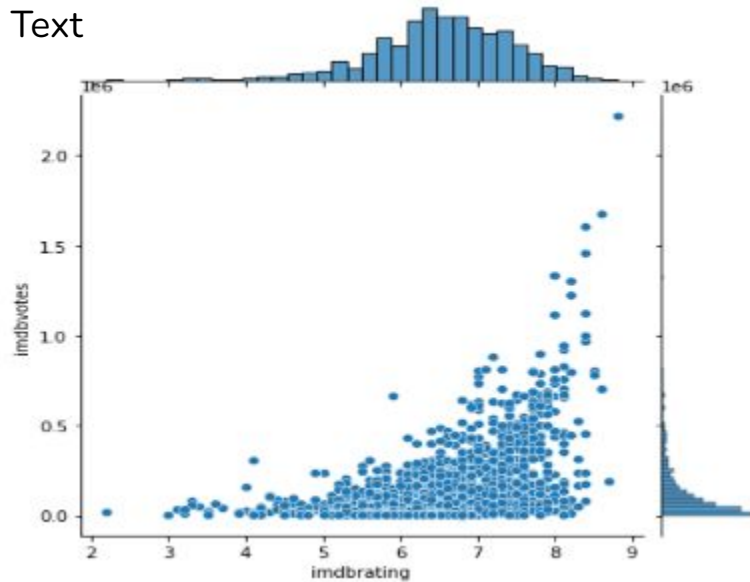
# Correlation Chart Between IMDB Rating and other features

- Observations: 1. Imdb Votes and IMDB ratings are not strongly correlated.
2. Budgets and IMDB ratings are not strongly correlated .
3. Budget and Box Office amount is correlated.
4. Ratings from different sources are not always reflects same kind of feedback for same movie.

	imdbvotes	boxoffice	budget	gross	metascore	imdbrating	internet_movie_database_rating	rotten_tomato_rating	metacritic_rating
imdbvotes	1.000000	0.680483	0.551815	0.642377	0.313106	0.451652	0.451652	0.287468	0.313106
boxoffice	0.680483	1.000000	0.722155	0.911370	0.169221	0.206332	0.206332	0.146010	0.169221
budget	0.551815	0.722155	1.000000	0.784985	0.064433	0.092982	0.092982	0.005359	0.064433
gross	0.642377	0.911370	0.784985	1.000000	0.136701	0.207146	0.207146	0.131744	0.136701
metascore	0.313106	0.169221	0.064433	0.136701	1.000000	0.544940	0.544940	0.789362	1.000000
imdbrating	0.451652	0.206332	0.092982	0.207146	0.544940	1.000000	1.000000	0.662932	0.544940
ie_database_rating	0.451652	0.206332	0.092982	0.207146	0.544940	1.000000	1.000000	0.662932	0.544940
tten_tomato_rating	0.287468	0.146010	0.005359	0.131744	0.789362	0.662932	0.662932	1.000000	0.789362
metacritic_rating	0.313106	0.169221	0.064433	0.136701	1.000000	0.544940	0.544940	0.789362	1.000000

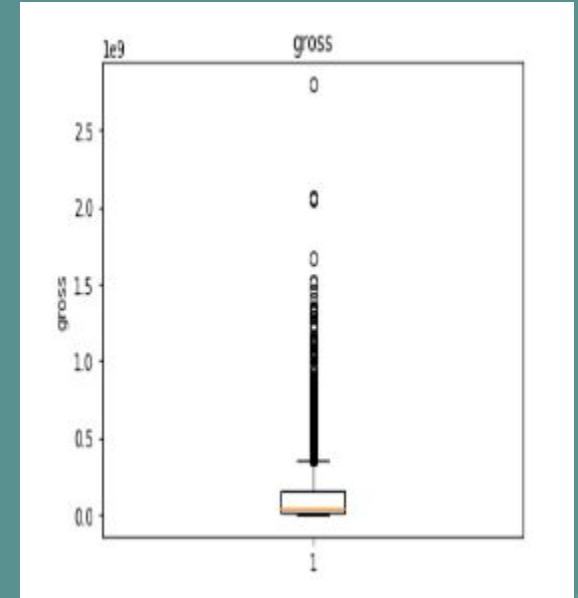
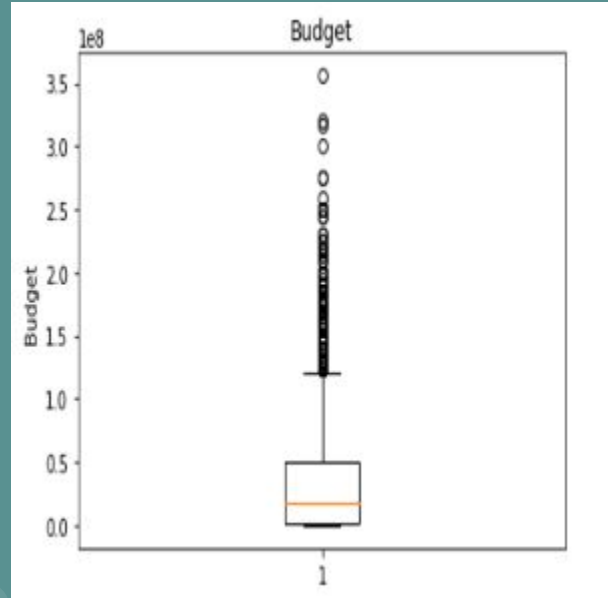
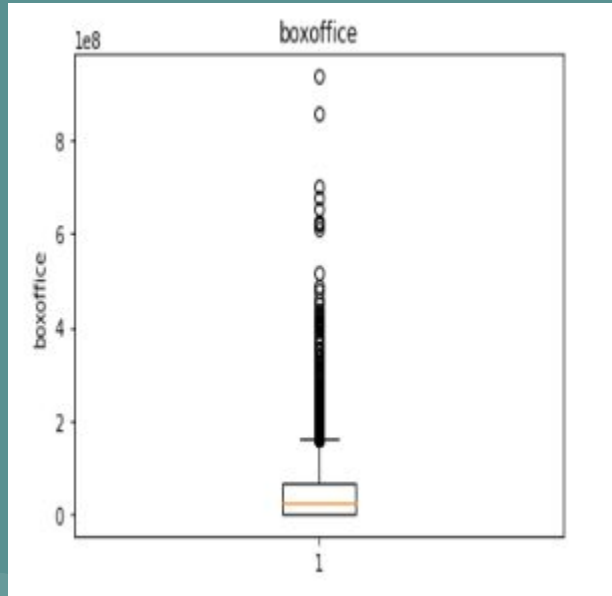
# Density,scatter plot between IMDb ratings and Votes and data Clusters

Text

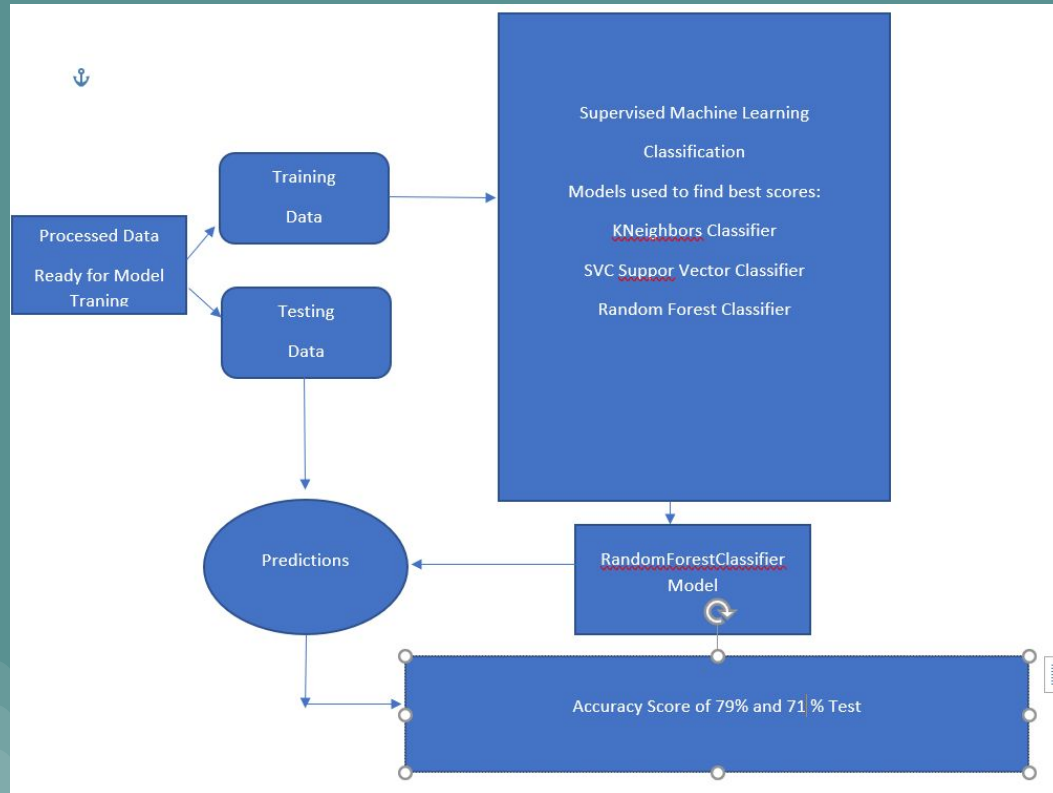


# Data Analysis

Box Office, Budget and gross Amounts



# Model





# Word Cloud

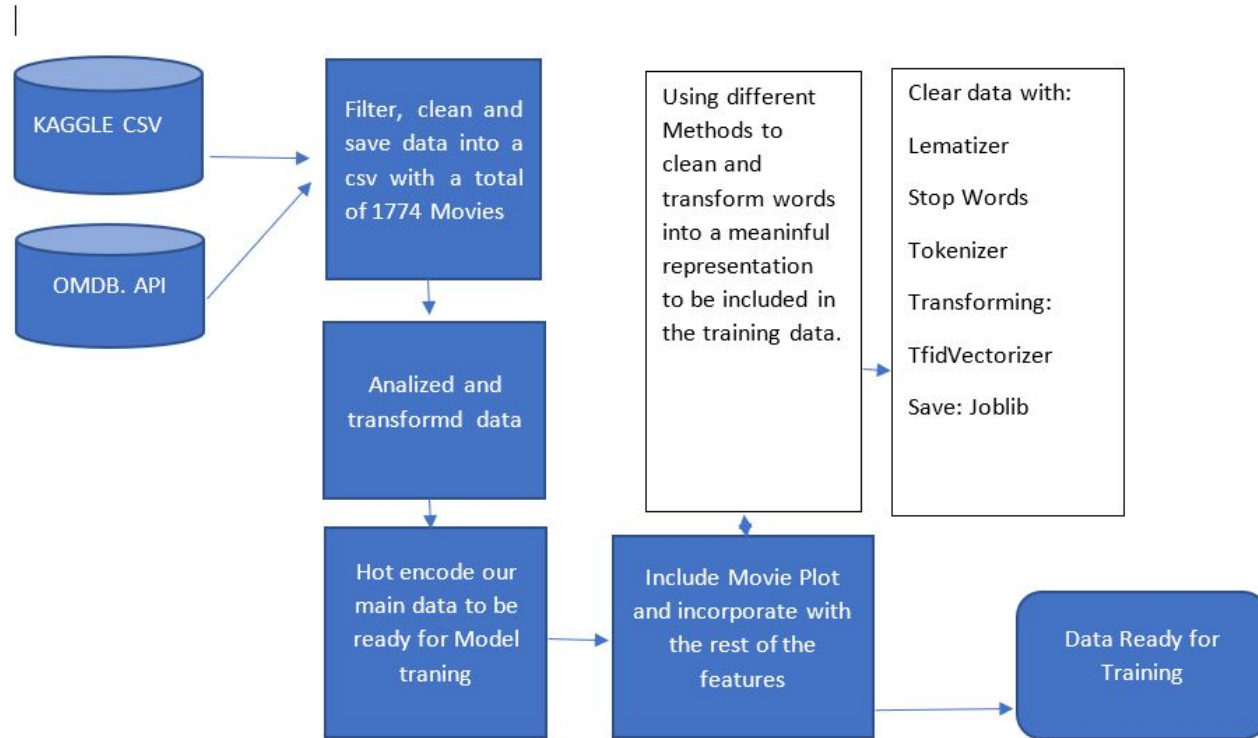
# Movie Title



## Movie Plot



# Data Preprocessing



# Front-end & deployment

Let us predict if your movie will be successful.

Runtime (mins):

175

Budget (USD): 0

Rated:

R  
PG-13  
PG  
G

Director:

other\_director  
Ridley Scott  
Robert Schwentke  
Steven Spielberg

Writer:

Luc Besson  
Nicholas Stoller  
Jane Goldman  
Allan Loeb

Actors:

Mark Wahlberg  
Michael Fassbender  
Liam Neeson

**Numerical features:** Runtime and Budget

**Categorical features:** Rated, Director, Writer, Actors, Language, Country, Star and Genre

**Text feature:** Plot

Plot:

After graduating high school, two best friends go on a road trip to discover themselves.

I'm Feeling Lucky!

# Model Operation

## Flask Application

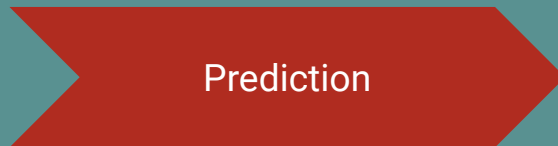
WordNetLemmatizer

nltk.tokenizer

Vectorizer (70 keywords)



**2D Numerical  
array (304)**



Prediction

## Result



- Interpret
  - Display

# Thank you

