CS550 Project Presentation

Dhiraj Gandhi - DDG74
Harsh Bhatt - HB371
Ramanathan Arunachalam - RA831
Sri Harsha Musunuri - SM2322

HYBRID MOVIE RECOMMENDATION ENGINE ENHANCED BY MICROBLOGGING DATA

Application:
% Personality Based Movie Recommendation System

Domain:
% Massive Data Mining

Motivation of the Idea

٭ Modern video-on-demand companies like Netflix, Amazon Prime Video, Hulu, DisneyPLUS & their state of the art recommendation systems.

٭ When we first read about sentiment analysis as a topic, it was really persuasive enough, to have a hands-on understanding and see if we can improve the traditional hybrid recommendation systems using sentiment score of a tweet about a movie.

٭ In the "information overload" age, making an informed choice is hard but recommendation systems are playing a great role in solving this problem. Hence, this project.

Contribution

% Sentiment Analysis of twitter tweets on movies in *MovieTweetings* Dataset

% Collaborative Filtering model using the User - Rating matrix.

% Content Based Filtering model using the movie Metadata and User - Rating matrix.

% Multiple Hybrid Recommendation engines (Hybrid 1, 2, 3, 1.1, 2.1, 3.1) employing permutations of the above three methods.

% Tried multi-layered Perceptron for Collaborative Filtering as experimentation and observed that the results not as promising as ALS because of less data in MovieTweetings dataset.

% TMDb dataset - This contains all the metadata (cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDB vote counts and vote averages.) of the movies that are present in it.

% Movie Tweetings Dataset comprises of three data files - users.dat, items.dat, ratings.dat
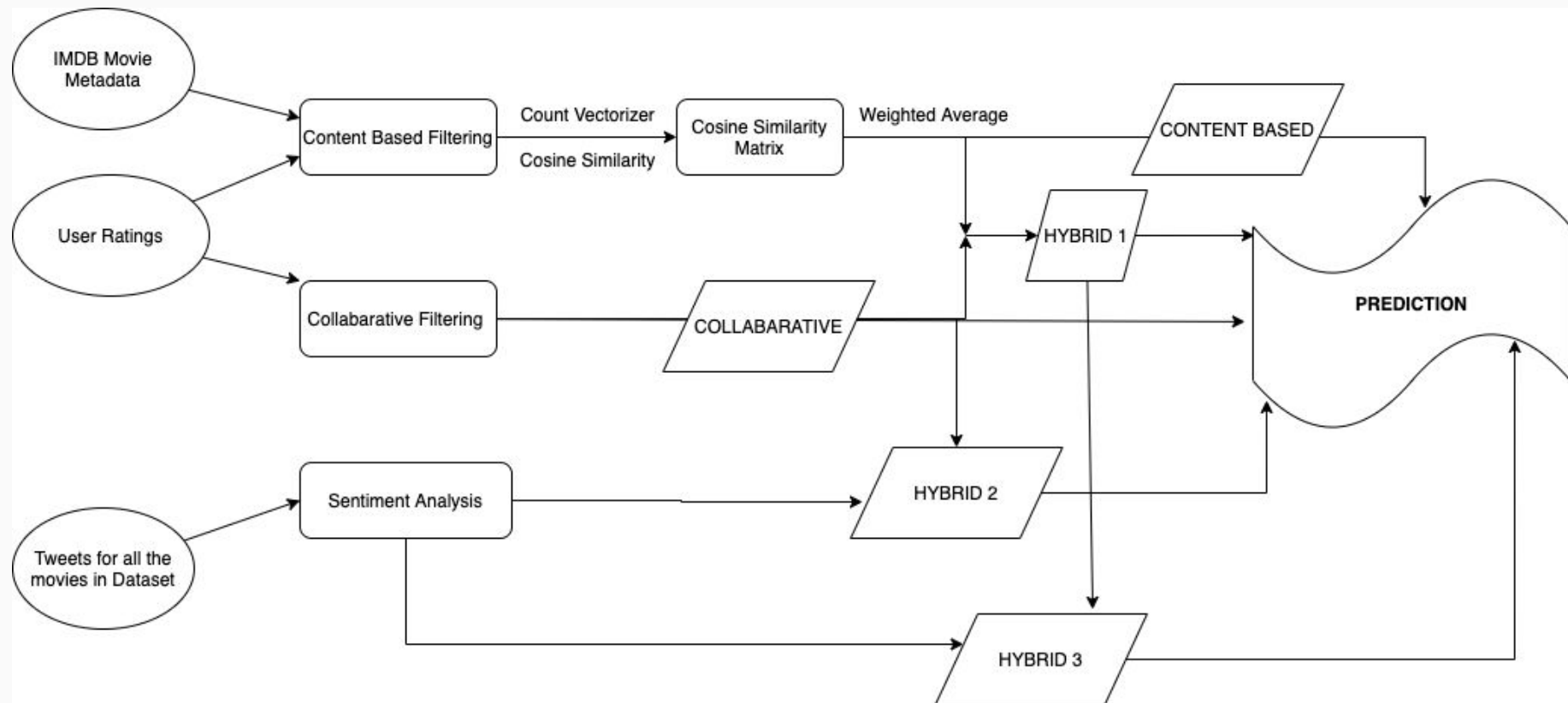
- / users.dat - consists of userid::twitter_id - Link
- / items.dat - consists of movie_id::movie_title(movie_year)::genre|genre|genre - Link
- / Ratings.dat - user_id::movie_id::rating::rating_timestamp - Link

% Sentiment Score Dataset - Twitter's movie tweets analysed and given a sentiment score for the movies of MovieTweetings Dataset.

% We used web scraping techniques to get the IMDb data related to MovieTweetings Dataset.

% Content-Based Filtering:

/ Content based filtering trains all the item set to find similarities between items and then rate a movie to the user based on similar movies which user has liked before.

/ Useful when a new user enters the system and has no ratings (cold start problem).

/ Recommends movies based on particular movie and it's metadata (like genre, director, actor)

% Our Approach:

/ Calculated similarity between movies using count Vectorization and cosine similarity using features - genres, cast, country and keywords.

/ Rating of a movie for a user is calculated as a weighted average of movies the user has rated where weights are the cosine similarities.

Section 3: Recommendation Systems

% Collaborative Filtering:

/ Collaborative filtering uses the opinion of other users of same type to recommend an item.

/ Useful when a new user enters the system and has no ratings (cold start problem).

/ Recommends movies based on particular movie and it's metadata (like genre, director, actor)

% Our Approach:

/ We are using ALS - Alternating Least Squares algorithm that was used in during the Netflix Prize challenge.

/ We start by choosing a random number of latent factors N and optimise it by reducing the loss function value.

/ After attaining best RMSE possible out of the tried Regularization param, Rank combinations. We choose the trained model as best model and predict ratings using that.
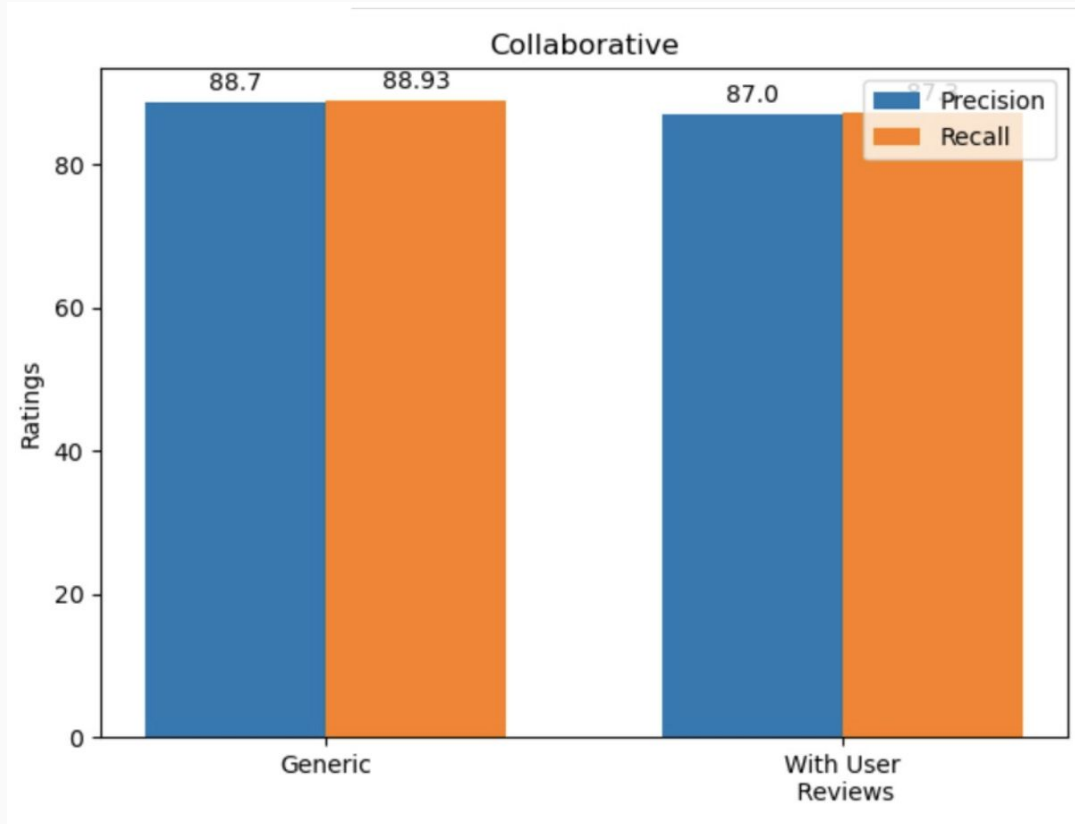
% Top 10 Recommendation for a user that mentions his favourite movies via Collaborative Filtering:

```
Recommendations for the movies:
Shutter Island|

1: Rejected (2000)
2: City of Life (2009)
3: The Wind of Al Amal (2013)
4: The Woodcarver (2012)
5: The Green Girl (2014)
6: Through My Eyes (2012)
7: Super Bowl XLIX (2015)
8: Urge (2015)
9: A Billion Lives (2016)
10: Nowhere to Hide (2016)
```
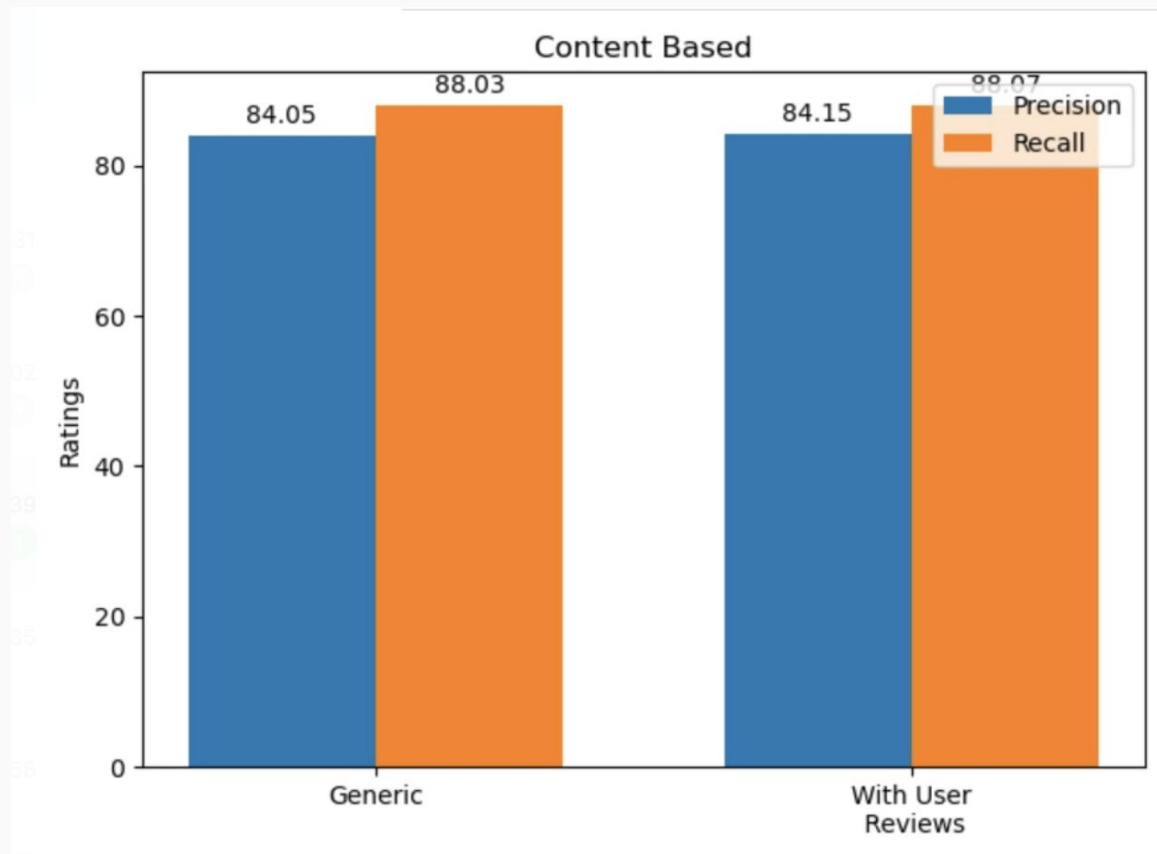
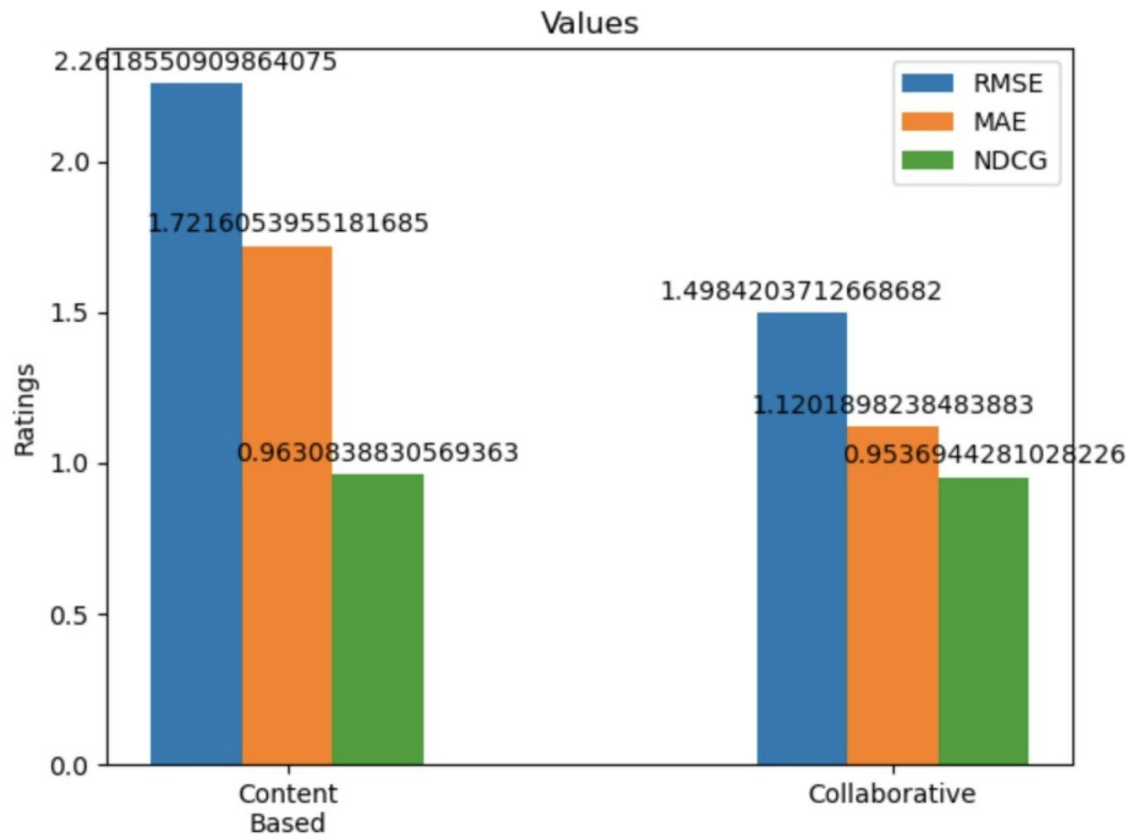# Section 5:References

[1] Jiangqin Wu Yin Zhang Liang Zhang Chenxing Yang, Baogang Wei. 2009. CARES:a ranking-oriented CADAL recommender system. (2009), pages 203–212.

[2] William Cohen Chumki Basu, Haym Hirsh. 1998. Recommendation as Classification: Using Social and Content-Based Information in Recommendation.(1998).

[3] Altigran S da Silva Marcos Andre Gonc alves Cristiano Nascimento, AlbertoHF Laender. 2011. A source independent framework for research paper recommendation. (2011), pages 297–306.

[4] Brian Oki Douglas B. Terry David Goldberg, David Arthur Nichols. 1992. Using Collaborative System to Weave an Information Tapestry. (1992).

[5] Joep J.M. Kierkels Thierry Pun Mohammad Soleymani, Guillaume Chanel. 2008.Aﬀective ranking of movie scenes using physiological signals and content analy-sis. (2008), pages 32–39.

[6] Kanjar De Partha Pratim Roy Sudhanshu Kumar, Shirsendu S. Halder. 2018.Movie Recommendation System using Sentiment Analysis from Microblogging Data. (2018).

[7] Wei Xu Yibo Wang, Mingming Wang. 2018. A Sentiment-Enhanced HybridRecommender System for Movie Recommendation: A Big Data Analytics Frame-work. (2018)

[8] Yunhong Zhou, Dennis Wilkinson, Robert Schreiber and Rong Pan. Large-scale Parallel Collaborative Filtering for the Netflix Prize. (2009)

"Computers are able to see, hear and learn. Welcome to the future"

Dave Waters

Questions ?

Thanks!