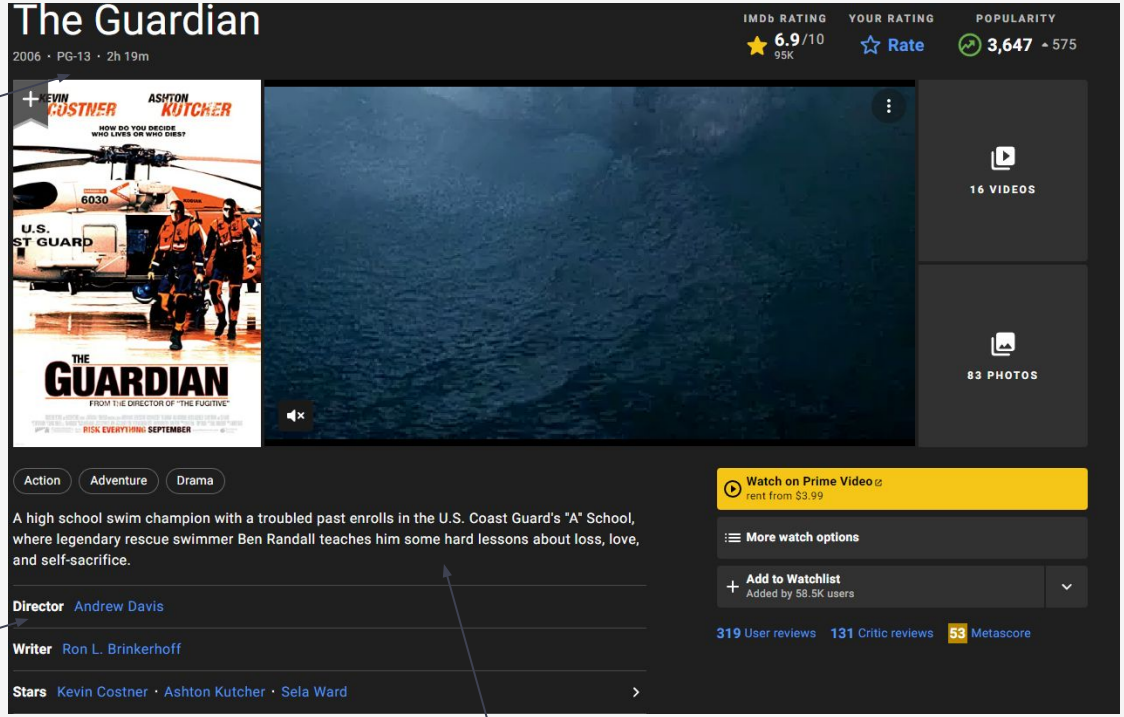
Movie Recommendation

by Grant Doan, Prayash Joshi, Reagan Orth, Ved Patel
(Group 6)

Data Processing - movie data scrape



rated

genre

director

synopsis

movie_synopsis1

A high school swim champion with a troubled past enrolls in the U.S. Coast Guard's "A" School, where legendary rescue swimmer Ben Randall teaches him some hard lessons about loss, love, and self-sacrifice.

similar?

An imprisoned drug kingpin offers a huge cash reward to anyone that can break him out of police custody, and only the L.A.P.D.'s Special Weapons and Tactics team can prevent it.

movie_synopsis2

Data Processing - review scraping

Webscraping

```
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
http://www.imdb.com/title/tt0406816/usercomments
```

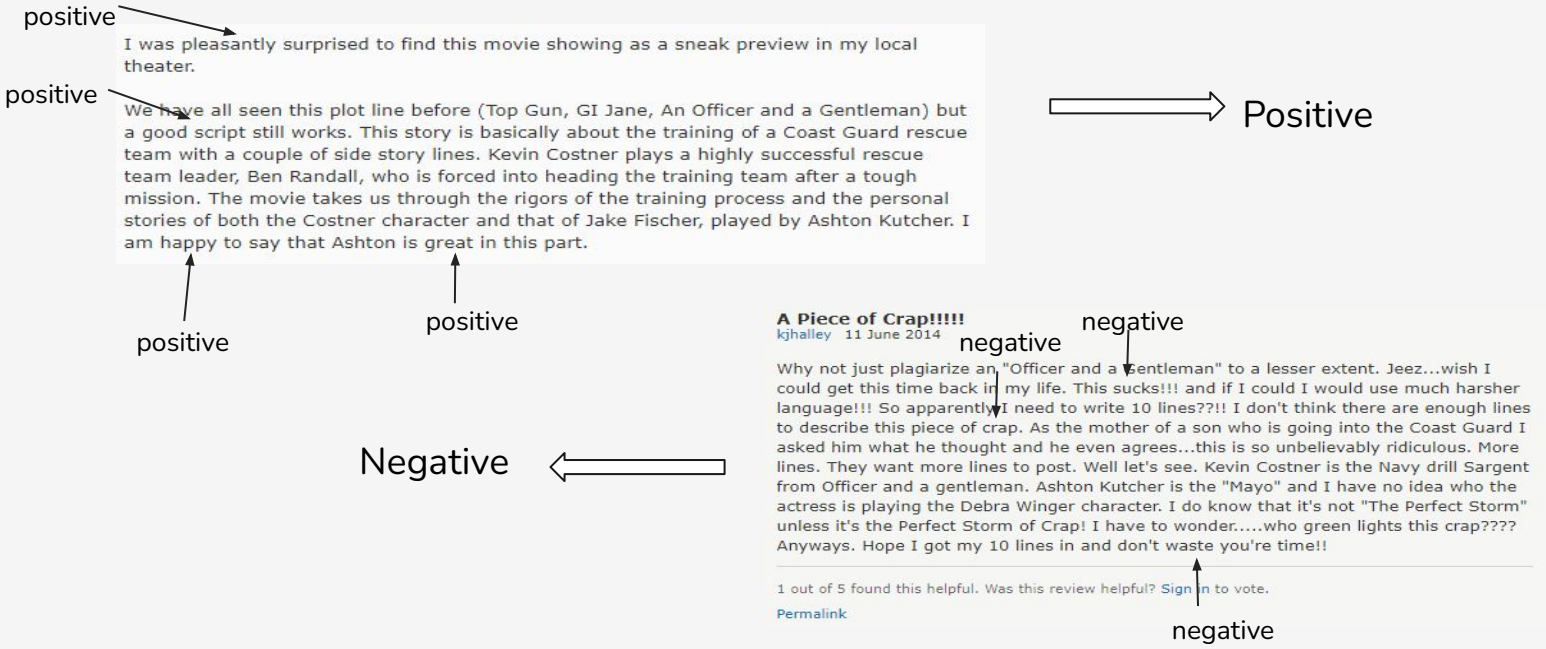
movie_id (unique)

rating

review

helpfulness_rating

Feature Engineering - sentiment analysis

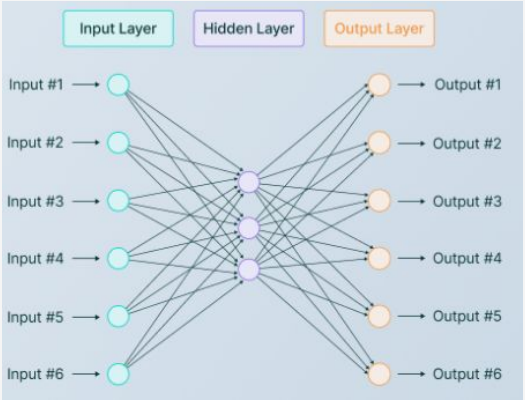


Application of Machine Learning

Two potential neural network architectures:

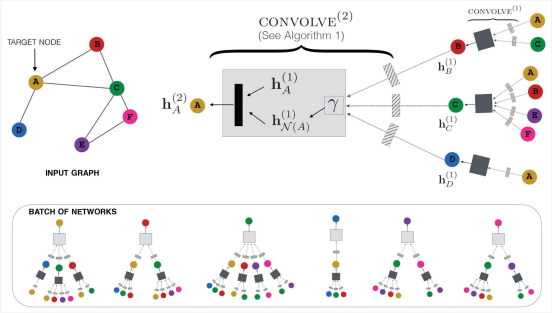
- LSTM/Recurrent Neural Network:
  - Great for natural language processing and speech recognition
  - "Forget gate" filters out unneeded information
  - LSTM rather than RNN due to precedence issue

- Graph Neural Network:
  - Great at modeling relationships
  - Better memory footprint
  - Can model groups/genres of films



Application of Machine Learning - GNN

- Use a Graph Neural Network to learn high-dimensional mappings for movies
  - Large, undirected, unweighted graph
  - Movies are nodes
  - Edges represent similarities
  - Similar movies will be close together



- Use user ratings and/or reviews to check user preferences
  - Check movies most similar to movies the user already likes
  - Check descriptive words in any user reviews provided
  - Combine movie preferences

Evaluation Metric

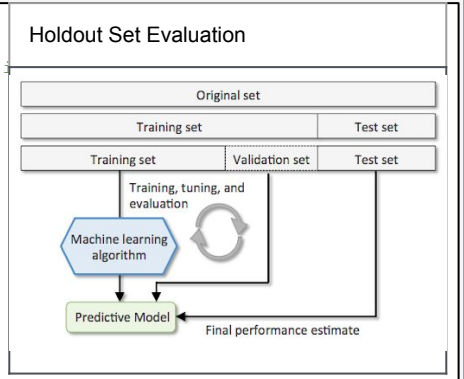
- **Milestone 4:** Holdout Set Evaluation

Capstone Timeline

- Define Problem
- Assess Possible Approaches
- Setting up Environment
- **Phase 1: Classification**
  - **Evaluation Metrics:**
  - **Accuracy**
  - **Precision**
  - **Recall**
  - •
  - •
- Phase 2: Recommendation
  - •
  - •
  - •
- Tweaks to recommender model
- Present and Host Results

Approach 1: Utilizing Sklearn.model & Sklearn.metrics

```
# Loading the dataset ...
# Splitting data into training and testing sets, splitting training data i
training and validation sets
X_trainval, X_test, y_trainval, y_test =
train_test_split(X, y, test_size=0.2, random_state=42)
X_train, X_val, y_train, y_val =
train_test_split(X_trainval, y_trainval, test_size=0.25, random_state=42)
# Preprocessing ...
# LTSM model ...
model = Sequential()
model.add(..........)
# ... LTSM Model building cont'd
# Training the model
model.fit(X_train, train_data['rating'], epochs=10, batch_size=32, validation_data=(X_val, val_data['rating']))
# Evaluating the model
y_pred = model.predict(X_test)
y_pred = np.round(y_pred).astype(int)
accuracy = accuracy_score(test_data['rating'], y_pred)
precision = precision_score(test_data['rating'], y_pred)
recall = recall_score(test_data['rating'], y_pred)
```



Discussion

Any Questions?

References

1. Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18). Association for Computing Machinery, New York, NY, USA, 974–983. https://doi.org/10.1145/3219819.3219890

2. Shin, Hyunseok, and Sejong Oh. “Feature-Weighted Sampling for Proper Evaluation of Classification Models.” Applied Sciences, vol. 11, no. 5, 2021, p. 2039., https://doi.org/10.3390/app11052039.

3. Son, Jieun, and Seoung Bum Kim. “Content-Based Filtering for Recommendation Systems Using Multiattribute Networks.” Expert Systems with Applications, vol. 89, 2017, pp. 404–412., https://doi.org/10.1016/j.eswa.2017.08.008.

4. Sunilkumar, Chaurasia Neha. “A Review of Movie Recommendation System: Limitations, Survey and Challenges.” ELCVIA Electronic Letters on Computer Vision and Image Analysis, vol. 19, no. 3, 2020, p. 18., https://doi.org/10.5565/rev/elcvia.1232.

5. Steffan Rendle, Li Zhang ,and Yehunda Koran. “On the Difficulty of Evaluating Baselines: A Study on Recommender Systems.” Cornell University. https://doi.org/10.5565/rev/elcvia.1232.