Feb 22 Meeting Minutes, 2025 | CSC 442 Team Project Meeting

Attendees

- Anthony Brown
- Nikhil Vasudeva

Notes

- The team discussed how to handle the missing values in the **director** column of one of the datasets. We agreed that dropping the column entirely or removing the rows would worsen the scope of our analysis more than keeping the information and filling the missing entries with a sentinel value to indicate the value is unknown.
- There were too many cast members and genres to construct separate columns to represent the information.
- We addressed the multi-valued columns **listed_in**, and **cast** by extracting features from the columns such as first value or counts of entries.
- We discussed how to handle inconsistent formatting of film **duration** between TV shows and movies. We considered combining the film duration with the **runtime** column to make the data more consistent.
- Missing numerical features such as **imdb_votes** were imputed using the mean because there were few missing values in the columns
- Missing textual data such as **description** encouraged us to drop the associated rows because there were few instances and replacing the missing values would be difficult.

Action items

Task	Deadline	Person Responsible	Comments
Reformat and document notebooks for cleaning the IMDb notebook	February 25, 2025	Anthony Brown	
Reformat and document notebooks for	February 25, 2025	Nikhil Vasudeva	

cleaning the IMDb notebook			
Revise the **Homework 5b** report	February 25, 2025	All team members	
Upload project artifacts to Github	February 25, 2025	All team members	