Ashley Nguyen

Github - https://github.com/ashleynguyen04/DS4002/tree/main/CS3

Lights, Camera, Reaction! – Analyzing IMDb Movie Reviews Sentiment Across Genres

Introduction: It's Friday evening on the Downtown Mall, and you and your friend are standing in front of the Violent Crown, paralyzed by too many options. The trailers all grab your attention. Multiple films look promising. Action or thriller? Comedy or horror? What genre are people currently loving? Your friend pulls out their phone and opens IMDb, scrolling through the reviews to see what real people are saying. There are way too many reviews for you to read, but that's when a lightbulb in your brain turns on. Even if they are just words, movie reviews aren't just options, they're data too. Hundreds and thousands of rows of data. You wonder to yourself, what if I could uncover some sort of pattern in how audiences respond to films across genres and over time?

Deliverables: In this case study, you will be given two datasets: thousands of user-generated IMDb movie reviews and the official IMDb metadata containing movie titles, genres, and release years. Your goal is to merge these datasets and create a Multinomial Naïve Bayes model trained on this merged dataset to reveal how sentiment differs across genres, decades, and rating intensities. Your purpose is to translate messy, real-world text data into a compelling story about how real-life audiences experience movies.