



Spoiled Tomatillos Project

Daniel Chen

Jay Lok

Matt Morgan

Veronica Shei

Abel Shin

Team Felix

January 20, 2018

Overview of the Problem

The client, a stealth startup, is looking to enter the movie industry through a social recommendation platform product called Spoiled Tomatillos. The main client asking for this product is the professors and the teaching staff of CS 4500. CS 4500 is a upper level undergraduate course taught at Northeastern University by professors Jose Annunziato, Nathaniel Derbinsky and Michael Weintraub. These three individuals represent the main members of the Spoiled Tomatillos Board. The client expects to see a huge adaptation of the product, fast expansion of the user base, and to aim for acquisition and profit by the end of the fiscal year 2019.

As part of the Spoiled Tomatillos team, we are an internal software development team that was tasked with creating and shipping this product. At a high level, the main features we expect to achieve is a social recommendation system for movies that bridges the chasm between services like Netflix or Amazon to Facebook. The platform's system should comprise of these primary functions:

- Users should be able to create accounts
- Users should be able to search for movies, plotlines, cast, awards, and showtimes at nearby theaters
- Users should be able to identify friends on the platform
- Users should be able to recommend movies to friends
- Users should see three types of suggested movies based on average critic rating, average rating on the site, and via User-User Collaborative Filtering that utilizes advanced Unsupervised Machine Learning technology

The client wants to fund and develop the Spoiled Tomatillos product because it sees a need in the social recommendation technology industry and believes there is potential for monetization and high profit. Because there is no technology that currently bridges the gap between modern digital platforms and social media, the Spoiled Tomatillos Board believes that this product would give them a big competitive advantage in the movie recommendation industry. Additionally, the three different types of movie suggestion algorithms that Spoiled Tomatillos will use gives them a big advantage over other similar movie suggestion websites like IMDb. We want to especially highlight the User-User Collaborative Filtering technology which is technology that has not yet gained high penetration in this industry. Depending on how well the Spoiled Tomatillos team does, the client is hoping that we can increase user adoption and the strength of our platform by linking Spoiled Tomatillos with affiliate codes on movie tickets and digital content providers like Netflix, Amazon Prime Video, iTunes, and more. Depending on how good the automation algorithm is, Spoiled Tomatillos may even be able to infiltrate music platforms or book platforms. Once the Spoiled Tomatillos Board decides on the ethical requirements for the product, we may potentially be able to sell all the user data that we have collected on our platform to interested companies, such as different movie production companies like DreamWorks or movie platforms like Netflix or Amazon Prime Video. Ultimately, the timeline for the Spoiled Tomatillos product is to make a Phase 1 prototype, build value via a user base, and improve the social recommendations algorithm through consumer usage, and aim for acquisition in the 2019 fiscal year, and profit.

Background

While there is no legacy implementations specific to the project a lot of the features that currently live in services like Netflix or Amazon and Facebook will have to be taken into account when we implement the Spoiled Tomatillos product. However, the product will be completely new and not based off of any existing code. Thus, there is no current legacy system, process, and no stakeholders that are maintaining or operating this system.

This group will be developing an entirely new system from scratch. We will need to establish a process to determine the requirements and the features, and what is needed to develop, test, maintain, and operate our system. Since there is no known mainstream or widely used copy of the system the only way to buy this function would be to buy services from corporations that implement these specific features. For instance, buying IMDb would provide us with movie information, ratings and a database with customer reviews. We could then incorporate this information into our main product by adding a social recommendation element so that users could identify friends and suggest movies to them.

The operators of the system will be the system administrators and the developers charged with keeping the service up and running. We will need to make sure servers to the websites are up and that the backend processing is accurate and returns valid results. We will need to ensure that there is an appropriate amount of computational power for the recommendation engine as well as the data storage units. We will also need to operate the system by inputting movie data into the database and monitoring reviews or the database for any inaccuracies. Whoever is managing the database will also be in charge of the best way to represent each movie and the best way to accurately and efficiently store the information. Another thing to take into consideration is that in some cases movie companies may complain that their film is inaccurately portrayed or represented on the site and this may require manual intervention for us to more fairly display their work. From a user perspective, consumers will also be operating the system

by creating accounts, adding friends and prodding them with exciting new movies in the system. While there are many other stakeholders, the above represent an overview of the main stakeholders that will be operating the system.

Scope

After completion, Spoiled Tomatillos should be a fully functioning web-based social recommendation system for movies. Users will be able to access this system from their accounts or as an anonymous user. These users can then share their movie experiences or opinions with friends from around the world. The site will allow users to search movies to find more information about plot, cast, production, awards, and theatre times for that specific film. Users will be also able to rate movies and these ratings will contribute to future recommendations and automatically form playlists.

There are many requirements for the system which we will go over below. One of the main requirements is developing a user management or account management service. We will need to develop a system for users to create and update their accounts. We will also need to develop an administrative system to possibly delete or edit accounts for users in order to provide services like password reset or banning users from the service. Permissions will have to be managed so that users can only see their own information as well as public information. Admin accounts will also have special privileges so that they can view all parts of the system.

From the system's social aspect, we will need to implement services that allow users to connect with their friends. Friends should be able to "prod" each other with movie recommendations. The more the user and their friends use the system, the more accurate their recommendations will be.

These recommendations will come from three algorithms. The first algorithm will come from average critic rating. This algorithm will pull from external, publicly available sources to aggregate critical reception. The next algorithm will be based on average rating on the site. This algorithm should become more accurate as more users join the site and rate movies, providing a larger dataset.

The final algorithm will be the most interesting, and likely the most complex. This algorithm, the User-User Collaborative Filtering, will recommend films using Unsupervised Machine Learning. This means that in order to find recommendations for User A, the algorithm will find users that have similar tastes to User A and recommend movies that these similar users enjoyed the most.

This "Collaborative Filtering Algorithm" will likely prove to be challenging to get right considering how many factors need to be considered. Some tough decisions will have to be made regarding how many "similar users" will be pulled, whether friends will be weighted differently, and if user specific "prods" will have any effect on the algorithm. We anticipate having to use machine learning technology and statistical methods to find the optimal weights for all these factors. Although none of us are data scientists, several of us are interested in the field and should be able to create an effective algorithm.

Another challenge will be movie data storage. There are hundreds of thousands of movies available to pull information from that are currently available on Netflix and Amazon, so we will have to be smart with how we store this data to ensure that we do not overextend our resources. We have a limited amount of AWS credits so having a description, cast list, and production information for every movie on the internet in a SQL database is not very realistic. We will have

to make decisions on how many movies to include in the system and how we store this information as we continue the development process.

As with any project, time and resources will be a limiting factor. We are all full-time students, so time management will be crucial for a project of this scope. Since we all have different schedules, meeting will take more coordination than in a traditional office environment. We also have a limited amount of AWS credit, so we must be conservative with how much data we host.

For this project we've assumed that the professors will act as the startup client. We've also assumed the startup's ability to gain a significant user base, as the accuracy behind the recommendation algorithms will rely on user supplied data.