A logo with purple letters

Description automatically generated

**CSC 340 Machine Learning**

**Final Project Evaluation**

**BD15034_**

**Name:** **Project Title:**

**Possible / Earned**

|  |  |
| --- | --- |
| 20 |  |
| 20 |  |
| 20 |  |
| 40 |  |
| Total | /100 |

**Phase I – Data Preparation and Preprocessing**

Data Collection - Type/Count: [How is the data stored and what file types]

The data is being stored in a CSV file. This file shows all the possible games made between 1998 to 2024! The earliest release date I found in this data set is March 5th which I am surprised by.

Data Cleaning: [How data will be handled or converted if changed from original]

Within the dataset, there was a lot of cleaning that needed to be done. Looking at the user and metric scores, some of them had null values. If one of them had null values, I used the other attribute’s value to replace. For the age rating, I put RP which stands for “Rating Pending” if there was null values. If there was no values with the title, developer, and publisher, I ended up deleting it since it ended up having no information with the rest of the attributes. Other than that, If there was no critic score or user score, I also deleted the field because it would be hard to predict for a recommendation system. I deleted about 400 fields but I still have about 13.7k fields. A screenshot of a computer

Description automatically generated

Data Visualization: [What parts of the data will prove useful]

I feel like the most useful attributes would be critic/user score, genre, and release date. These four attributes would be very helpful on deciding what games would be recommended depending on the games I’ve played in the past.

A screenshot of a graph

Description automatically generatedA chart with many squares

Description automatically generated with medium confidence

Data Acquired and Approved: (Yes - 20)  (Partial - %)  (No - 0)

**Phase II – Model Development and Training**

Model Selection: [What algorithm or machine learning framework is utilized, detail]

The main algorithm I use within this recommendation system is called Cosine similarity. This algorithm measures the similarity between two vectors. Those two vectors are within one inner product space. Instead of looking at measurements within distance, it looks at the angles. The angles are then used by plugging in cosine functions. The range between the two are -1 to 1, like correlations. For this project, I made it positive by adding 1 to the return value. So, I will be getting ranges between zero and two. Two represent being very similar to each other while zero having the opposite similarity.

A screenshot of a computer program

Description automatically generated

Training: [What are the results of your training data set, adjust factors as needed]

For my training data set, it is very hard for me to train the data because it is mostly put in a game, and it would give me a few games back. I can change what it is going to look at within the data set. The algorithm looks at one attribute. I can change the attribute to focus on publisher rather than genre. What I noticed is that the Genre plays a very important role compared to other attributes

Results Delivered:  (Yes - 20)  (Partial - %)  (No - 0))

**Phase III – Model Evaluation and Deployment**

Evaluation Metrics: [What metric(s) are used to display effectiveness]

The only way to tell how it display effectiveness is based on the similarity score.

A screenshot of a computer

Description automatically generated

Deployment/Replication: [Where is this project deployed, accessible to replicate]

The code and excel sheet will be on GitHub and in the submission post.

Monitoring/Maintenance: [Does re-training the dataset change the outcome]

Re-running the same game would give you the same results with the same similarity factor. The only time it would be changed is if you add games to your least favorite list. If you put a game on that list, it would not

Results Delivered:  (Yes - 20)  (Partial - %)  (No - 0)

**Final Deliverable**

Project Completion (Phases):  (Yes - 40)  (Partial - %)  (No - 0)

**Comments:**