Hello!

My name is Nicholas Rasmussen. I am a final year Master’s of Applied Science candidate in Chemical Engineering at the University of Waterloo. My research thesis is designing and modelling a hydrogel based photobioreactor with terrestrial and microgravity applications. My thesis was conducted in conjunction with a Canadian Space Agency design challenge jointly held with NASA. The Deep Space Food challenge aimed to develop novel technologies to address food needs in future long term space activities and exploration. On a team with my supervisors Valerie Ward and Nasser Abukdeir, we were first stage finalists in the competition.

I graduated in 2022 with a Bachelor of Applied Science, honours chemical engineering from the University of Waterloo. I finished my cooperative degree with 2 years of industry experience, a materials and manufacturing process specialization, and a management sciences option. I graduated with distinction on the Dean’s Honours List with a cumulative GPA of 91.10%. In my final year of study, my team was awarded the capstone design award for best overall project. Our project was the design of a solvent-based recycling process for multilayer plastic films lime plastic food pouches made form polyethylene and PET.

In this project I explored the IMBD dataset using different machine learning techniques to create a recommender system for users.

The first step in the project was to load and manipulate the data so that ultimately each movie had an average rating based off the dataset.

The first model I explored was k-nearest neighbours.

K-nearest neighbours classification is an unsupervised method to find similar datapoints based of the assumption that if two datapoints have similar feature vectors, then the datapoints are likely to have the same label.

This model is non-parametric. As such the time complexity of training is 0, and the space complexity is O(nd). However, the classification of a point is O(ndk) and O(nd) in space. where d is number of features, n number of datapoints, and k is the number of neighbours.

Regularization can help improve the performance of a model by reducing the influence of individual features on overall model. <br> <br>

Ridge regression (Tikhonov regularization, or L2 regulatization) uses a hyperparameter α to penalize the weight assigned in the model to any one feature. the loss function for this technique is:</p>

Lasso regression (L1 regularization) is another technique to reduce the weights of any individual feature. However,

lasso regression tends to favour sparse solutions allowing the feature space to be reduced for the model. the loss function for this model is:

</P>

Using Pandas, the Movie features and user features dataframes were merged to get the complete feature vectors for the model.

One hot encoding was used to convert categorical data to numerical data for the regression model

</p>

<p> The following Models presented were trained on the same train/test split. Seaborn was used to visual the results of the model. <br> <br>

The linear model achieved a mean squared error on the test set of 1.1954598871852038.

</p>

Ridge regression was tested next. The hyperparameter alpha was tested with grid search cross validation in the range of 1e-5 ro 90. The optimal alpha value was then used to train the model.

the mean squared error on the test set was 1.1954598871973383. The relative size of the parameter were visually represented in the following plot.

<p>

Finally lasso Regression was explored. The hyperparameter alpha was tested with grid search cross validation in the range of 1e-5 ro 90. The optimal alpha value was then used to train the model.

the mean squared error on the test set was 1.195466032036714. Though the model performance was similar to Ridge Regression, this model removed 3 features from the model. This is advantagious as less data must be collected for new predictions.

The relative size of the parameter were visually represented in the following plot.

</p>