**Slide 1**

Hello everyone my name is Joseph Chan, welcome to my CISC 372 final project presentation.

**Slide 2**

The topic that the project is oriented around is “ESRB” ratings. “ESRB” Ratings are used to show what age range a video game is meant for. These ratings are E for everyone, ET for everyone ten plus, T for teen, and M for Mature. The goal is to build a model that can predict the “ESRB” ratings of a game. There is a Kaggle dataset that contains the attributes of a game and the ratings. This was used to train the models for this project. The purpose is to create a model that new video game developers can use to predict the rating of their game.

**Slide 3**

There were five models used in this project. The neural network was a 7 layer model with four output nodes that predict the classes. The XGBoost classifier used random grid search to look for optimal parameters and was the model that took the longest. The Random Forest model was a simple to design model and it had excellent results. The Birch Clustering method performed the worst as it had an accuracy of 20%. The K-Nearest Neighbour model used 3 neighbours and performed decently. The K-Nearest Neighbour model is quite simple and not many parameters are available to adjust. All of these models went through 5-15 changes before the optimal parameters were found.

**Slide 4**

As seen in the accuracy table, the neural network had the best performance out of the five models. Another interesting point is that even though XGBoost takes minutes while the others are instant, it's performance is not better. The confusion matrix shown belongs to the Random Forest model. The bad predictions in the confusion matrix are the ones where the prediction is less strict than the true label.

**Slide 5**

There were a few lessons learned over the process of this experiment. One of the things is when a model is overfitting. I have found that just increasing the amount of times a model is trained or epochs doesn’t actually increase the accuracy. After a certain point it just starts overfitting the model. Another importance is the quality of parameters, a good parameter is what leads to model performance. Keeping the model small and optimizing the parameters and then increasing the size is how to build a good model. The voting ensemble method was learnt later on in the method but, not enough time was left to implement the methods. The model's predictions would have been compiled as votes,the majority vote becomes the prediction used.

**Slide 6**

In conclusion the model with the highest accuracy is the neural network model. However, it's good to note that the neural network and random forest model accuracy will vary every time it is run. By looking at the confusion matrix we can tell which models have the most bad predictions, which are the ones where the true label is more strict. The model that had the least of these predictions was the random forest model. In conclusion if the model were to be used for games put on the market I would suggest the random forest method as misclassifying a game as a less strict game could be bad. If the model was just used to see what rating your game would get for personal information I would suggest the neural network since the accuracy is higher.