Jasmine Brown

Springboard

Machine Learning

17 July 2019

Machine Learning

The plot and models are now created for the understanding of the data. Because my independent variable is a numerical value. The best model to use was liner regression to explore the relationship between the dependent variables with the independent variables. Total shipped is the dependent variable We used the rest of the variables as independent variable to view how total shipped is changed. We can also write the relationship in a linear equation.

Before the linear model was created, we had created a testing and a training data set to run the model on 80 percent and 20 percent to get the best outcome. There were multiple models created with each variable and then all variables to look at the r squared RMSE and the p value. The control set was running on the model. The control set randomly and chooses the best models.

The P value give for each independent variable lets us know how likely that the coefficient is significant. The P-value for each coefficient is best if the number is low. There are stars beside each p value for the variable if it is significant. In my data for the model with the independent variable’s year, rank, platform, and critic score. The variable critic score, rank and the platform Wii are significant. The rank has two stars which means it has a probability less than .001.

The lower the p-vale the shows how good the model. The lover the p-value the more accurate your model maybe. On the model with my training data set that the control set picked my p value is 0.0004145 Which means that the model is likely to be accurate. There is other deciding factor such as the r squared value. When looking at your linear equation on a graph it forms a line. The r^2 vales show how close the independent variable is to the line. The closer this number is to one the better. The r^2 value on the same model used earlier is 0.56 so that means that the independent variable is halfway there. The Root mean squared error is an average of the variance. This number is based on the units of your independent variable when deciding its fate. It can range from 0 to any number. The smaller the number the better the model. My RMSE is 5.38 this this low compared to the average number of total shipped, there for this model was halfway successful.