Taylor Yavari

Hobert Bush

GT Data Analytics

November 1, 2021

Regression Analysis

Personally, I think both Logistic Regression and the Random Forrest have their place in this homework. The reason being both are good at analyzing and ultimately predicting using large data.

With that being said, the Random Forrest Model was better at predicting than the Logistic Regression when the data was not scaled. Once the Data was scaled, the Logistic Model became the better model. I believe this comes down to the amount of complexity each model can handle. As variables increase, I think the decision trees in the Random Forrest Model become very convoluted and ultimately less accurate, which is why we see the Logistic Model being 16% more accurate than the Random Forrest once the Data was Scaled.

Although the Logistic Model scored higher than the Random Forrest, I do not necessarily believe that makes the Logistic Model the best possible model for this data. After Scaling, the accuracy level was only 66.03%. This value would not make me comfortable implementing in a business environment.