Nathan Wheelwright

12/13/21

CS 450

For this personal project, I used the provided “Occupancy” dataset. This dataset is used to train and test a model to detect if a room is occupied based on variables such as room temperature and humidity.

This dataset has a rather small number of features, so I set about to feature engineer more. I pulled the date feature into separate components such as minute, hour, day(of month), and day(of week). I also created a feature to show if it is a weekend or not. This feature engineering proved to be useful, as Fridays and Mondays tend to have more people, and the weekend has less people. Out of the standard features, the amount of light and CO2 in the room has a very strong correlation to whether the room is occupied.

One thing to note about this dataset is that it is quite unbalanced. The amount of time rooms were unoccupied far outweighed the times people were present. This would cause a problem when we got around to training and testing the models, so I over sampled the data. This means that I duplicated the some of the data to make the dataset more balanced when it comes to the occupied (target) feature.

The first model I created is an XGBoost model. After some fiddling with the hyper-parameters, Below are the results that I got from the test dataset. The results are pretty decent and are pretty balanced when it comes to false positives and false negatives.

Chart, treemap chart

Description automatically generated

'RMSE': 0.2625003060872891,

'Accuracy': 0.9310935893040795,

'F1': 0.8353808353808354,

'R2': 0.584200519113756

The second model I created is a standard TF Keras Neural Network. After some tweaking of the model, I got the results from the test dataset as seen below. The results in this model are significantly better than the XGBoost model above.

A screenshot of a computer

Description automatically generated with low confidence

'RMSE': 0.11313062056765052,

'Accuracy': 0.9872014626899783,

'F1': 0.9700854700854701,

'R2': 0.9227702456728701

Colab Link: <https://colab.research.google.com/gist/Natosphere/08f0ed43c8eb83eedb688df86376924a/nathanwheelwright_module7.ipynb>