Reflective essay

Student's Name: Karl Shen

Professor’s Name: Zhichun Liu

Course: HUDK 4050

Date:12/21/2021

The field of machine learning is always emerging with renovations comes an increase in demand and significance. Predictions and classifications produced by machines and algorithms that can drive better judgments and smart actions in real-time without human interaction are one reason why I want to learn machine learning. Machine learning is getting a lot of traction and recognition as a technology that helps evaluate vast amounts of data and automates the jobs of data scientists. By involving automatic sets of generic approaches that have superseded traditional statistical techniques, machine learning has transformed the way data extraction and interpretation works. Before taking this class, I have little experience of data wrangling and data analytics using Python. In my spare time, I took online classes about machine learning and found machine learning topics quite interesting. My objective for this class is to learn Data mining in a more systematic way.

During this semester, I have learned techniques throughout the in-Data mining process pipeline. The first part is the data cleaning or data wrangling. In this section, I have applied multiple approaches, for example, I use dummy variables to transform categorical features. Also, I encountered Null values problems and learned how to validate them. Then, I have familiarized supervised and unsupervised machine learning algorithms. In supervised machine learning, I have learned linear regression in prediction and logistic regression in classification. In addition, I also tried several other algorithms by myself in some of my ACA projects such as Xgboost, ANN etc. I had a lot of fun training and validating the model and receive a great deal of satisfaction when I see improvement made by the models. For unsupervised machine learning, we have talked about k-mean clustering which is very popular algorithms when dealing objectives like clustering consumer data. Moreover, I have learned dimension reduction method like Principal Analysis which would help to reduce the number of features whenever it needs. Furthermore, we also talk about data visualizations. It gives us a clear idea of what the information means by giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets.

In my future study, I have planned to take more classes about data mining. I feel like I really enjoy working with data in terms of wrangling, modeling etc. Although, we have tasted what is data mining and many basic algorithms, there are more for me to explore. For example, computer vision area is one of my prefer area, I have watched videos on YouTube that use data mining techniques to colorized pictures and movies from the old time. I found this quite interesting that we can finally how people would look like and the views in the storytales etc. In conclusion, this course gives me a great systematic introduction about data mining and have set a great benchmark for me to move forward. It refines my skills in Data mining and give me confidence to study more complex theory and algorithms in the future.