Eric Swanson

Cs303 hw08

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generatedA close up of a map

Description automatically generated The Naive Bayes algorithm was used in this assignment to predict income from census data. The benefits of using a gaussian NB classifier are that it is fast and works with high dimensional data. It works by assuming attributes are conditionally independent, meaning there is no correlation between them. This algorithm was not particularly ideal for the dataset in question, given that there was a correlation albeit a weak one between some attributes of the dataset, however this was not an ineffective algorithm to use either. I was able to achieve a prediction accuracy rate of 82% by predicting on the education attribute. Using KNIME, I found the most effective attribute to train on was education. This aligns with my expectations from the principle component analysis performed with a scatter matrix. This accuracy is in part due to confounding factors of the relationship between education and income not recognized in the census data. The results of this predictive model were output to a data file using Python.