

Francesco Leopardi

12/15/2022

DS 210

Prof. Leonidas Kontothanassis

Using Kmeans for clustering

I initially thought of the idea of using kmeans clustering as I was interested in a challenge governments face nowadays, especially in Britain during the brexit elections. Through this exploration, I wanted to be able to determine whether more youth voters actually do vote for different parties than older people, and whether high youth voting turnouts would actually change the outcome of an election.

Through this exploration, instead, I learned much more than that. I had a good idea, before the project, about what data types, methods of defined structs, and the implementations of kmeans++ actually meant in rust. In addition, the methods of defined structs allowed me to construct vectors and iterate through variables that I had defined throughout my tests and functions. Through this project I was able to learn that my csv file had a datatype difference than what i had programmed my code to handle. I learned that iterating (or parsing) through string of csv data will output a data type called StringRecord for every row of the file. This posed the biggest challenge to my code. I tried using predefined structs, I tried creating my own struct and then implementing it for the function which would then read the file, and I also tried changing the data within the vector (within which the StringRecord was located) to an i32 first, but never

seemed to obtained the right Datatype, Vec<f64>. I then started over, iterating again over every StringRecord in the csv file and apply methods to it. This way, I thought, would allow me to convert the StringRecord to an f64. Unfortunately I was unable to do so, despite having worked some 30 hours in the past few days. The previous weeks helped me build the code, and draw a structure to it through tutorials and extensive research through documentations. I followed the schedule that I had predisposed in the project proposal, but ignored the issue of reading the file into a correct form.

This step was an important step in my Data Science career, as it allowed me to not only improve my knowledge for writing tests, functions, modules, structs, and kmeans, but I also learned how to obtain the necessary research skills to find the appropriate knowledge.

In conclusion, although my code wasn't able to run, I was able to learn and construct a code with an interesting problem, good tests, and modules, which given time, could run. I am proud of my tireless work throughout the past few weeks.