Foundations of Big Data Analytics Unit 5

David Finger

Colorado Technical University

Abstract

Describe the data pipeline and how codes could be used in the future in regards to predictive algorithms.

Data pipelines are used in programs to determine the order in which processes are going to happen. Data pipeline programs are configured to take the output from the last process and use it in the next process in the line until the program ends. This process is very useful in the prospect of descriptive reports and predictive algorithms.

Descriptive reports are created by sending any database information into a processing program. These programs will take all the data and scrub away anything that the users have set as “unimportant”. Using the data that they receive as the output; companies can use that data to inference how things are going to turn out in the future and see how the products that they produce are doing overall.

Predictive algorithms go a few steps further than just reporting what the data provides. They take the ending result from the data report and use that information to predict, based on a few different methods, what the most likely outcome is going to be. The program that I created uses KNN, also known as K-nearest neighbors, to predict the likelihood that any chosen demographic resided in a certain zip code. This was done by filtering all the zip codes to display the highest number of people that fit into different demographics to see where they resided when compared to each other.

Codes like this can be used almost anyway to obtain a lot of information and by doing so create influence in that area. This information is how a lot of companies promote certain products in certain areas and different products in others. Using this information in such ways allows for companies to try to understand whether it is due to location or demographic that appeals to its customer bases. Taking that information, companies can then alter their advertisements in other areas that have similar demographics.

Resources

*Double-Pipelined Join Algorithm—Javatpoint*. (n.d.). Www.Javatpoint.Com. Retrieved November 19, 2020, from <https://www.javatpoint.com/double-pipelined-join-algorithm>

Pipeline (computing). (2020). In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Pipeline_(computing)&oldid=970450786>