**DATA VISUALIZATION AND ALGORITHM FOR CANADA’S PROVINCES FOR INDUSTRIES**

**OBJECTIVE:**

* There are datasets of Canada’s Provinces (10). It contains various information in the area of Industries and Employment. Here, the prediction is to be done based on this given information for the employment. To determine the Industries associated with Employment through Data Visualization.
* Data Visualization: To predict the data there should be a graphical representation of the information. In Big Data the data visualization is the first step for prediction and analysis.

**STEPS:**

* The data of all provinces it taken on the terms of different industries.
* The prediction and graph are generated through the code in R Studio.
* The library which is installed is ggplot
* Ggplot: It is a library in R studio used when data is to be plotted and when data is to be represented in graphical format.
* The graph formed is a BAR Graph.
* There are two axis X and Y which shows number of people Employed and Number of Industries.

**DATA PREDICTED:**

Through different Canadian Provinces

* The provinces like Ontario, Alberta, British Columbia are highly populated and so the number of employed persons is more in comparison to other provinces.
* The provinces like Prince Edward Island province and Newfoundland and Labrador have less people employed.
* The most employed industry is the Services-Producing Industry and Finance, Insurance, real estate and leasing.
* The least are the Agriculture areas and Manufacturing.

**Algorithm for Machine learning:**

* 1) Here, the most effective algorithm that can be used for future prediction analysis is the KNN Algorithm K Nearest Neighbors Algorithm.
* 2) Through Algorithms in Machine Learning future data can be predicted.

**K Nearest Neighbors Algorithm:**

This algorithm depends on the Proximity.

Example:

* We have some data (Employees) so through this different aspect related to it (Industry, Areas, Provinces) such data can be used to predict other data related to it.
* So, it is called K Nearest Neighbors Algorithm.
* There are certain steps to perform this algorithm in Machine Learning. We need to define the K based on the neighbor’s data.
* There are also certain disadvantages of this algorithm is that it works slower when loads of data is there.