**DAB402 CAPSTONE PROJECT**

PREDICT WORKFORCE REQUIREMENT USING MACHINE LEARNING ALGORITHMS AND ANALYSIS

INTRODUCTION:

Due to the covid19 pandemic, risk and uncertainty are now common and unavoidable in every profession and industry, making it impossible for industry recruiters to maintain control over human resource management and planning. To overcome this problem, human resource department demands forecasting to anticipate the amount and quality of workers required soon. It is a worthwhile investment as an accurate prediction helps to meet the business requirements effectively with the best results.

The objective to carry out this project is to predict Labor Workforce among Canada, about its demands and shortages using various methods like LSTM, traditional machine learning methods and statistical models. We will be using analytical tools such as Microsoft Excel, Python and Tableau. This takes us to the next stage of our project work which is identifying appropriate dataset and assessing its quality, fitness, and usability for our proposed solution to the business problem mentioned above. We have gone through few possible data sources and chose appropriate datasets. The steps carried out for our data assessment as well as choosing an appropriate dataset are listed below.

Git-hub link: - https://github.com/abhaykorat/DAB\_402\_Capstone\_Group9.git

DATA ASSESSMENT

DATA SOURCE(S) AND GATHERING:

The first and the foremost step for our project is to choose an appropriate dataset. We reviewed some possible data sources which are mentioned below.

CEIC: <https://www.ceicdata.com/en/indicator/canada/labour-force-participation-rate>

Undergone through this website which contains Canada’s Labour Force Year-wise Participation Rate from Jan 1976.The information was not exactly that we were in search of. It does not contain information like industry type, sex and also, we found that this dataset is paid to download it. This made us to look for another dataset.

Trading Economics: <https://tradingeconomics.com/canada/unemployment-rate>

While downloading the dataset is asked for some add-ins or API and for that too we need to pay for it. So again we carried to find another data source suitable for our project.

Statistics Canada: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002301>

From this website, we found the data source with proper information like industry type, age group and gender which can help us to achieve our objective of the project. We can find out which industry sector will demand more workforce in the future. The dataset is huge and so we were in benefit that we can train more samples in creating machine learning algorithms. It was also free of cost to download it which was a plus point. So, we decided to consider this dataset for the project.

DATASET FEATURES/ATTRIBUTES, QUALITY, FITNESS, USABILITY & MISSING VALUES:

The chosen dataset contained useful features like labour force characteristics (part-time, full-time, unemployment and employment), NAICS (industry type), Sex, Age group; also, the dataset is large so there will be more training samples which will allow us to obtain better accuracy results of the machine learning model and would also help us in predicting which industry will have more future in near future.

Graphical user interface, text, application, email

Description automatically generated

DATASET TERMS OF USE:

The dataset is taken from an open data source i.e., Statistics Canada. We have gone through its terms and conditions of use to ensure all the ethical guidelines which includes Consent, Clarity, Consistency, Control and Consequences are followed till the end. Data is only accessible for academic and knowledge purposes and will require permission for the further extensive use.

DATASET REPRODUCIBILITY:

With the help of proper usability of the datasets and information mentioned above, it ensures that result of our analysis can be reproduced if followed till the end.

DATASET STORAGE AND HANDLING:

We do not need any special storage for the dataset as chosen dataset does not exceed 180MB and so it is not a major concern.

REFERENCES:

{1}. Canada Labour Force Participation Rate, 1976 – 2022 | CEIC Data. (2022). Retrieved 30 May 2022, from <https://www.ceicdata.com/en/indicator/canada/labour-force-participation-rate>

{2}. Canada Unemployment Rate - April 2022 Data - 1966-2021 Historical - May Forecast. (2022). Retrieved 30 May 2022, from <https://tradingeconomics.com/canada/unemployment-rate>

{3}. Labour force characteristics by industry, annual. (2022). Retrieved 30 May 2022, from <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002301>

{4}. (2022). Retrieved 30 May 2022, from <https://www.statcan.gc.ca/en/reference/terms-conditions?MM=1>