CSI5155 Fall 2019
Project Proposal

Jiacheng Hou
Email: jhou013@uottawa.ca
Student ID: 300125708

Yue Li
Email: yli2620@uottawa.ca
Student ID: 7064204

1. Project Title

Predict Unemployment Rate of Major Regions by Machine Learning Methods

2. Project Type

Application Oriented

3. Project Description

The unemployment rate is a direct indicator of local economy. However, by the nature of this problem, it requires large amount of samples which makes frequent updates unfeasible. Meanwhile, relevant departments urgently need the statistics in order to prevent potential issues. Although, the trend of unemployment rate has high co-variance with historical data. Thus, we believe it is a regression machine learning problem.

The Local Area Unemployment Dataset provides us the necessary historical records. The dataset provides us the total labor force, employment counts and unemployment counts for a month of a year. On top of that, records are also stored with the area name and type. We believe this already lowered noise of the data.

From there, we are planning to utilize the Python library "SciKit Learning" for further experiments. The library is known by its extensibility in the machine learning community. With its helps, we can build robust regression models and compare their pros and cons.

4. Dataset Description

Dataset URL:
https://data.edd.ca.gov/Labor-Force-and-Unemployment-Rates/Local-Area-Unemployment-Statistics-LAUS-/e6gw-gvii

Dataset Name:
Unemployment Dataset

Dataset Size:
151240 rows × 11 columns

Dataset Description:
The Local Area Unemployment Dataset stores 11 features and 151,240 records. Excluding the target feature unemployment rate, the 10 training features are Area Type, Area Name, Date, Year, Month, Seasonally Adjusted(Y/N), Status (Preliminary/Final), Labor Force, Employment counts, and Unemployment counts. The dataset maintains monthly updates by the U.S. federal government which indicates the credibility of this dataset.

Dataset Content:

1) Area type (Column name: area type, data type: string)
2) Area name (Column name: area name, data type: string)
3) Date (Column name: date, data type: string)
4) Year (Column name: year, data type: numeric)
5) Month (Column name: month, data type: string)
6) Seasonally Adjusted (Column name: seasonally adjusted, data type: string)
7) Status (Column name: status, data type: string)
8) Labor Force (Column name: labor force, data type: numeric)
9) Employment (Column name: employment, data type: numeric)
10) Unemployment (Column name: unemployment, data type: numeric)
11) Unemployment Rate (Column name: Unemployment, data type: numeric)

5. Project Statement

We hereby certify we will not use these dataset(s) and algorithm(s) in any other courses, past, present or future. We understand that submitting overlapping material for more than one course is not allowed. This work is also not part of my thesis or capstone project research. We understand that any such action on my part will lead to a grade of INC being awarded for CSI5155 and may lead to academic sanctions imposed by the faculty of Engineering.