I. Research Background and Objectives:

During the COVID-19 pandemic, we seek to predict the unemployment rates in various cities of Canada using COVID-19 case and death statistics. The objective of our research is to analyze which model has a better predictive performance by comparing two commonly used machine learning methods — Random Forests (RF) and k-Nearest Neighbors (k-NN). In this process, we calculate and compare the Mean Squared Error (MSE) and Mean Absolute Error (MAE) for each method.

II. Data and Methods:

We use COVID-19 case, death, and unemployment rate data for various cities in Canada. The features include 'covid cases' and 'covid deaths' and the target variable is 'total unemployment rate(%)'.

We used the following two machine learning methods:

1. Random Forest (RF): RF is an ensemble learning method based on decision trees. We set the parameters as: number of trees = 100, minimum samples for a split = 2, minimum samples per leaf = 1.

2. k-Nearest Neighbors (k-NN): k-NN is an instance-based learning or a local approximation and simplification method. We set k = 5 and standardized the features.

III. Results:

Here are our predictive results for various cities in Canada:

1. Random Forest: Across the cities, the MSE values vary between 0.78 (Manitoba) to 17.48 (Quebec), and MAE values range from 0.57 (Nova Scotia) to 2.33 (Quebec).

一些文字和图案

描述已自动生成

2. k-Nearest Neighbors: Across the cities, the MSE values vary between 0.83 (Saskatchewan) to 18.78 (Quebec), and MAE values range from 0.80 (Saskatchewan) to 2.37 (Quebec).

一些文字和图案

描述已自动生成

IV. Discussion:

Although we attempted to use deep learning models for predictions, our data volume is relatively small and does not meet the deep learning model's demand for many samples. Hence, we had to resort to using traditional machine learning methods such as RF and k-NN.

From the results, the predictive performance of RF and k-NN varies across cities. Overall, RF performs slightly better than k-NN in most cities. However, in certain specific cities, such as Saskatchewan and Manitoba, k-NN outperforms RF.

Additionally, in the RF model, the importance of COVID-19 cases and deaths varies from city to city. For instance, in Alberta, the number of cases has a larger impact on predicting the unemployment rate than the number of deaths. This indicates that we need to consider the specific problem when deciding which features to use.

V. Conclusion:

Although we could not utilize deep learning models for prediction, this study provides preliminary insights into how COVID-19 case and death statistics impact unemployment rates by comparing RF and k-NN models. Our findings assist in selecting more appropriate models for similar problems in the future and offer directions to improve predictive performance.