

Classification of Cities and Regions of Saudi Arabia in 2021 in the spread of COVID-19 between regions with high, and low

Abstract

The goal of this project was to use the Classification of Cities and Regions of Saudi Arabia in 2021 in the spread of COVID-19 between regions with high, and low to help the Saudi Ministry of Health to take procedures regarding areas or cities where the virus spread is very high. Also, If a tourist plans to travel to one of the regions? It helps him/her make a decision about it.

I used a dataset from "the King Abdullah Center for Petroleum Studies and Research" website which included around 13 regions and 102 cities. My experimental is achieved promising results for this global crisis. For the visualization, which supports my results, I used the Jupiter notebook platform and GIS software with the IDW method.

Design

My study has a significant benefit for doctors and tourists in that it contributes to help the Saudi Ministry of Health to take procedures regarding areas or cities where the virus spread is very high. Also, If a tourist plans to travel to one of the regions? It helps him/her make a decision about it.

My contribution to this project is to give a reliable and useful analysis on the classification of level spread cases of COVID-19 around regions and cities between high, and low.

Data

I used a dataset from "the King Abdullah Center for Petroleum Studies and Research" website which included around 13 regions and 102 cities. A dataset contains **225.572k** rows and **11** columns. The sample about COVID-19 patients who live in Saudi Arabia. Also, I will predict as a target (High-low) in spread COVID-19 around Saudi Arabia cities and regions.

Features: **DC:** Daily or Cumulative rate of COVID-19 cases. **Indicator:** The status indicator is active, recovered, cases, mortalities in a particular state. **Date:** Date of case. **Day:** Day of case. **City:** A city of COVID-19 case. **Region:** A region of COVID-19 case. **CasesNumber:** Numbers of COVID-19 cases. **AvarageCases:** Average of numbers of COVID-19 cases. **WorkHoliday:** Status on a working day or holiday? , and **Country:** Saudi Arabia country.

Label: SpreedCovid19: COVID-19 spread case, Is high, or low?

Link of my dataset

<https://github.com/naz50/MVP/blob/main/NuhaDBooo.csv>

Algorithm

Feature Engineering

- 1- I split my data into X_train and X_test.
- 2- preprocessing steps - remove numbers in string, Null value, captial letters and punctuation
- 3- Converting categorical features to binary LebelEncoder variables.

- 4- I made StandardScaler to my data X_train and X_test.
- 5- I defined my model MLP Classifier and fit my datasets to it.
- 6- I made cross-validation by utilizing KFold 10 splits.
- 7- I tested and predicted datasets by using many report metrics.

Models

I applied many classification algorithms such as KNN, SVM, and MLP. In my experiment, dataset perform reliably and show promising results by using an MLP classifier. Also, via the experimental models, it is demonstrated that MLP performs the best in my dataset.

MLP is a multiple feed forward artificial neural network that maps input vectors to output vectors. The MLP can be defined as a directed graph with multiple node layers. Input layer is on the bottom, the output layer is on the top, and the others in the middle are the hidden layers.

Parameters:

- Alpha → float , L2 penalty (regularization term) parameter.
- max_iter → int ,Maximum number of iterations.

Model Evaluation and Selection

The entire training dataset of **225.572k** records was split into 80/20 train vs. test, and all scores reported below were calculated with 10-fold cross-validation on the training. Classification on the 20% test was limited to the very end. The official metric for data was classification rate (accuracy, and precision); however, class weights were included to improve performance against the F1 score. In addition, my model provides important aid in addressing the real-world crisis in the face of COVID-19, and this is the need to know the critical category classification "HIGH".

MLP classifier scores: 10 features (2 numeric) with class weights

- Accuracy 100%
- F1 1.00 micro, 1.00 macro
- Precision 1,00 micro, 1.00 macro
- Recall 1.00 micro, 1.00 macro

Tools

I used python libraries such as Numpy, Pandas, matplotlib, seaborn, sklearn, classification_report, and confusion_matrix.

Furthermore, I used the python platform is the Jupiter Notebook, and GIS software. In addition, I applied an MLPClassifier besides; I evaluated my model by utilizing some metrics such as precision, recall, and F1-score, ROC curve .etc.

Communication

