Investigation of Saudi Arabia COVID-19 Cases (Mar 2020 - Oct 2021)

Sara Alfaraj

Abstract

The goal of this project was to investigate COVID-19 cases in Saudi Arabia. A key aspect of this study is to display the most common factors that lead to an increased COVID-19 number of cases, in addition, predict the number of cases on a certain date. First, I collect the data from The King Abdullah Petroleum Studies and Research Center. Then, apply feature engineering along with a random forest model to get the results.

Design

This project discusses the questions below:

- Are the different seasons and events (Summer, Winter, holidays, Eids...) affect the number of COVID-19 cases?
- Do the numbers of COVID-19 cases differ between big/main cities and small cities?
- What is the expected date of reaching zero cases?
- Is the COVID-19 cases indicators expected to rise again after returning to school (at 31 Oct 2021)?

Data

The main data in this study obtained from The King Abdullah Petroleum Studies and Research Center (KAPSARC)[1] data portal, which gathered by Ministry of Health in Saudi Arabia[2].

- Include these features:

  - Number of Cases (Daily / Cumulative)

  - Case Indicators (Active, Mortalities, Critical ...)

  - Number of Cases

  - Date

  - Events

  - City

  - Region

- Consists of 428346 rows related to 207 cities in Saudi Arabia.

- Unit of measure: Persons.

- Range of the dates: 1 Mar 2020 to 1 Oct 2021.

Algorithms

Feature Engineering

- Converting categorical features to binary dummy variables
- Selecting subsets of categorical features
- Standardize features by removing the mean and scaling to unit variance.
- During EDA; DataType converting, feature adding.

Models

- Logistic regression and random forest classifiers.
- The training dataset records was split into 80% for train and 20% for test.

Tools

- Numpy and Pandas for data manipulation
- Sklearn for modeling
- Matplotlib and Seaborn for plotting

Communication

Presentation slide, that have visuals and discussions

Sources

[1]

https://datasource.kapsarc.org/explore/dataset/saudi-arabia-coronavirus-disease-covid-19-situation/export/?disjunctive.daily_cumulative&disjunctive.indicator&disjunctive.event&disjunctive.city_en&disjunctive.region_en

[2]

https://covid19.moh.gov.sa