**Predict the spreading of Coronavirus**

1. **Introduction :**

The outbreak of Covid-19 is developing into a major international crisis, and it's starting to influence important aspects of daily life.

For example:

* Travel: Bans have been placed on hotspot countries, corporate travel has been reduced, and flight fares have dropped.
* Supply chains: International manufacturing operations have often had to throttle back production and many goods solely produced in China have been halted altogether.
* Grocery stores: In highly affected areas, people are starting to stock up on essential goods.

A strong model that predicts how the virus could spread across different countries and regions may be able to help mitigation efforts. The goal of this task is to build a model that predicts the progression of the virus throughout 2020.

1. **Data Source & Information:**

<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>

1. **Basic Methodology :**

There will be a well-defined process to build a model that can deliver decent results.

The Project main focus will be at following points:

1. Accuracy - Model should perform well on the real data. It should be generalized over time and it can be applied to other scenarios.
2. Data Preparation - A good data analysis prior to feeding it into the model. It should include useful visualizations.
3. Documentation - code, and notebook, and additional data sources will be well documented so a reader can understand how it was done. Sources should be clearly cited. A high quality analysis should be concise and clear at each step so the rationale is easy to follow and the process is reproducible.
4. **Deliverables :**

I will be Submitting a notebook that implements the full lifecycle of data preparation, model creation and evaluation.

The Model will produce a table in the following format for all future days

* ObservationDate: Observation date in mm/dd/yyyy
* Province/State: Province or State
* Country/Region: Country or region
* Confirmed: Cumulative number of confirmed cases
* Deaths: Cumulative number of deaths cases
* Recovered: Cumulative number of recovered cases

The notebook will be well documented and contain:

* All steps to prepare the data, including references to external data sources
* Training of model
* The table mentioned above
* An evaluation of a table against the real data. Let's keep it simple and measure Mean Absolute Error.