ADVANCED DBMS PRESENTATION ASSIGNMENT

4.1 BUSINESS COMPUTING

GROUP MEMBERS:

SCT222-0130/2020 – MERCY NGENDO

SCT222-0122/2020 – MICHAEL KAMOTHO

SCT222-0123/2019 – ESTHER MUTHONI

SCT222-0162/2020 – BRIAN OSEWE

SCT222-0153/2020 – VIVIEN SAWE

Introduction

We will analyze the outbreak of coronavirus across various regions, visualise them using charts and graphs, predict the number of upcoming cases for the next ten days using linear regression and svm(support vector machine learning ) model in python.

A time series file is a collection of observations of well defined data items obtained through repeated measurements over time

Three time series files obtained from who will be used:

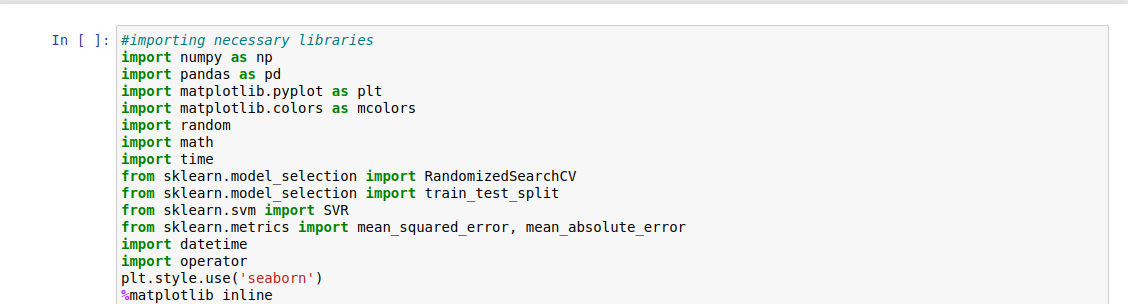
* confirmed.csv – Excel file detailing confirmed number of cases
* deaths.csv – Excel file detailing confirmed deaths
* recovered.csv – Excel file containing recovered patients

CSV Contents rows are:

* Province
* Country
* Latitude
* Longitude
* concurrent dates

Step One:

import all necessary libraries for the project

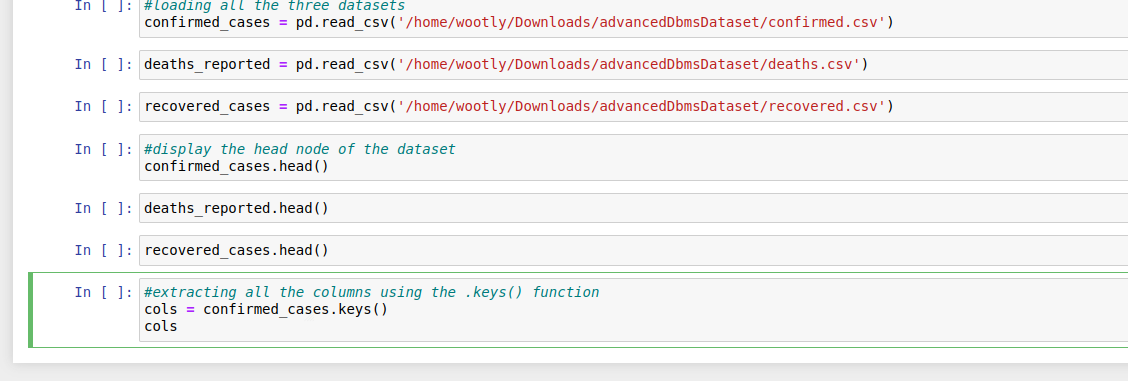


Load all three datasets

These include related csv files using panda read functions

Display the 5 top fields using the head function

Extract all the information from the dataset using the keys function



Pre- Processing tasks

Data Cleanup:

Removal of duplicates

handle missing data

Incorrect and inconsistent data – Looking for data that may be incorrect and incosistent which may in ling run affect the data prediction

Text and Date formatting:

convert text to columns

Format date and times

change text cases

Data Protection

Protect sheets and workbooks – Securing the data by limiting access to editing

Test results and Interpretations

SVM Prediction

Support Vector Machines (SVM) is a supervised machine learning algorithm used for classification and regression tasks. SVM aims to find a hyperplane in an N-dimensional space (N is the number of features) that distinctly classifies the data points

After visualising our data we can now begin to build a support vector machine algorithm for prediction. When building said algorithm we have to specify certain parameters such as:

Kernel

* This refers to the kernel type to be used in the algorithm
* must be one of linear, sigmoid, poly or rbf (default)

C

- is a regularization parameter

Gamma

This is the kernel coefficient of rbf polean sigmoid

Epsilon

This specifies epsilon tube where there are no penalties associated in the shrinking function

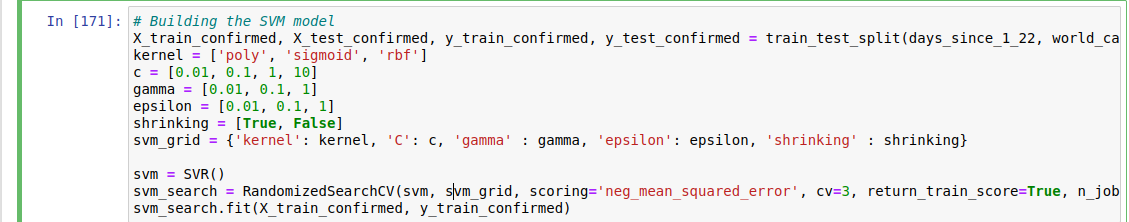
Shrinking

Boolean holder taking taking a an array of boolean values true or false

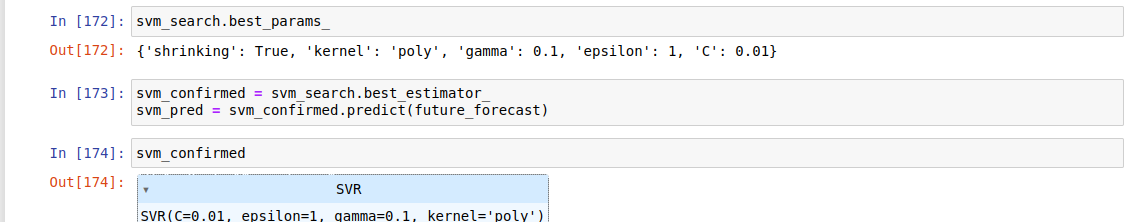
svm grid

This contains all the values passed as parameters to it

BUILDING AN SVM MODEL



Using the above build we can determine the best parameters for such a scenario I.e



Given the array of values we had set when defining the variables in the build we denote that the best parameters are as listed:

shrinking: True

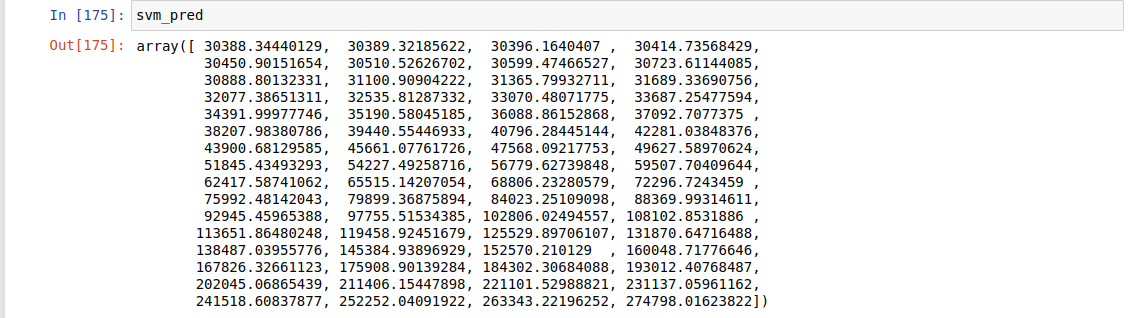
Kernel : Poly

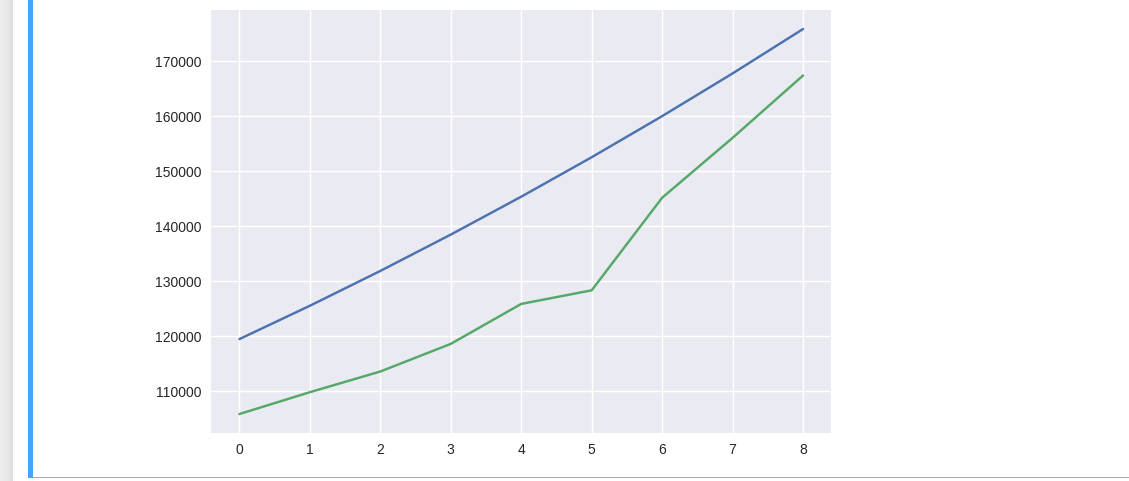
Gamma: 0.1

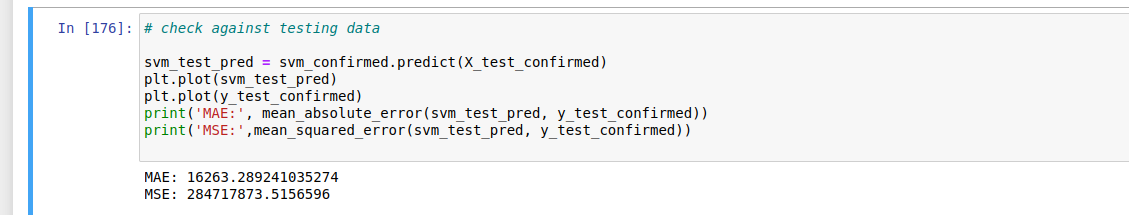
Epsilon: 1

C: 0.01

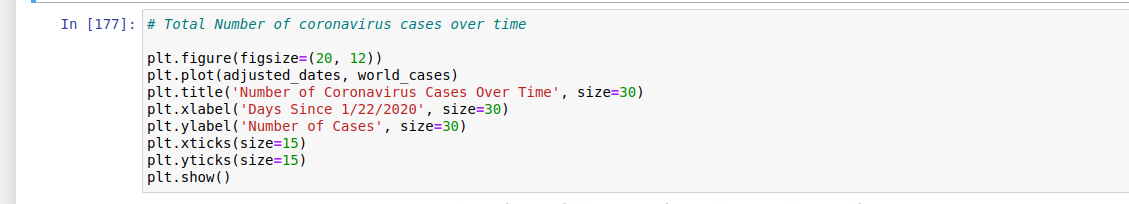
With the above set parameters we run the prediction algorithm and we get an assortment of values in an array form containing x and y coordinates on an x-y plane



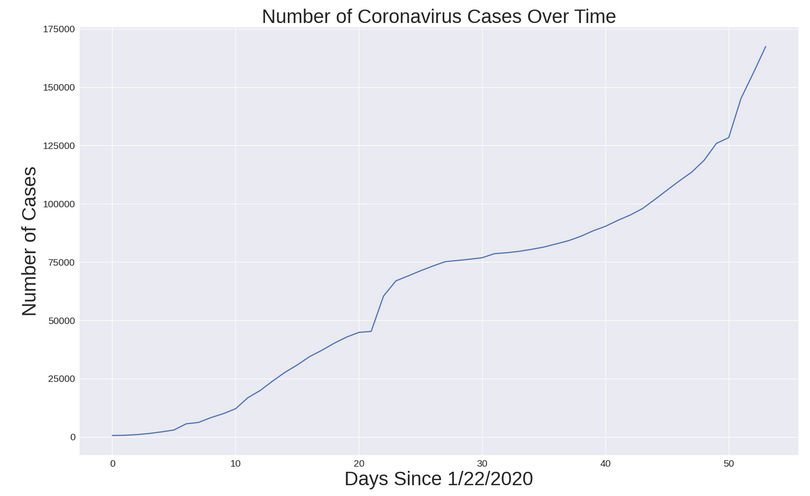
We then compare said values with the test data resulting in a similar trajectory as the test data as shown below in the graph:

For more context:

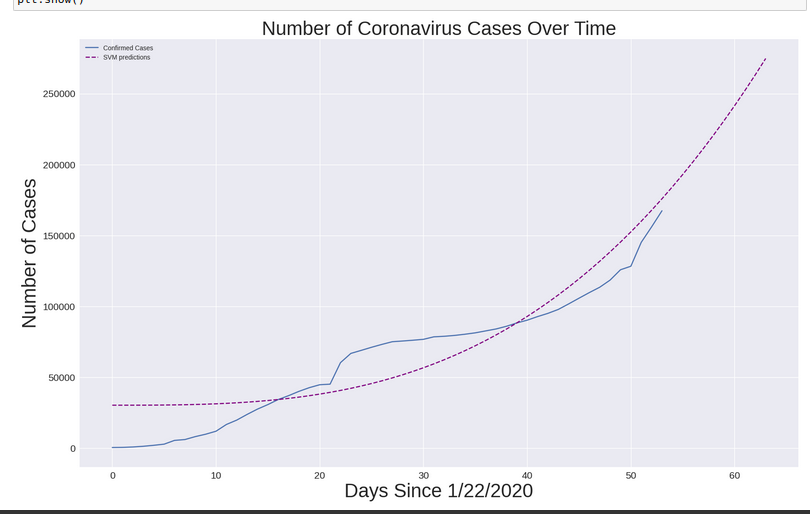
ultimately we then draw up the exiting covid rates as per the time series data



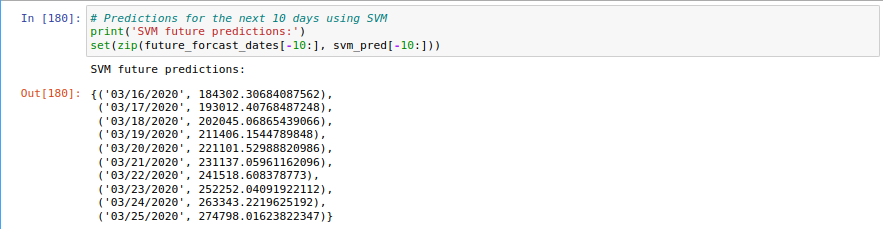
Below is the resultant chart:



For comparison we graph a chart showing the confirmed vs predicted cases for a more detailed visual of what the relation is between the two datasets



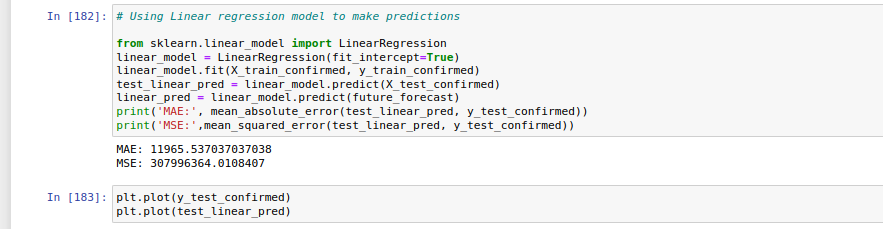
Having this as the baseline we can predict the outcome over the next ten days



Alternatively a linear Regression approach could be used to determine the outcome

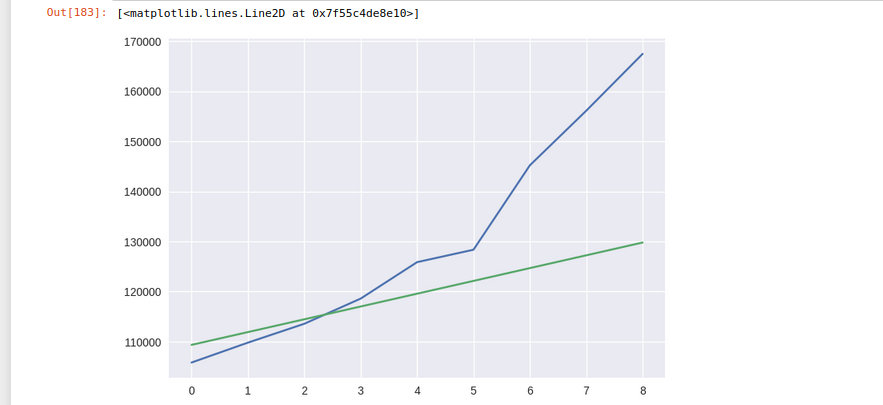
Linear Regression

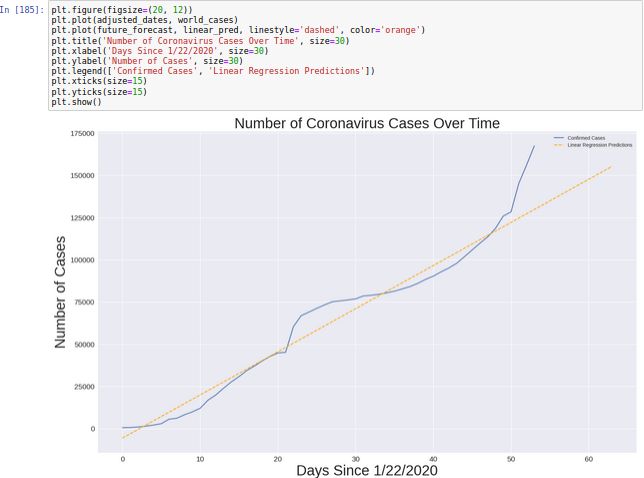
Linear regression is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. The simplest form is simple linear regression, which deals with the relationship between two variables, while multiple linear regression deals with two or more predictors.



MAE: Mean absolute Error

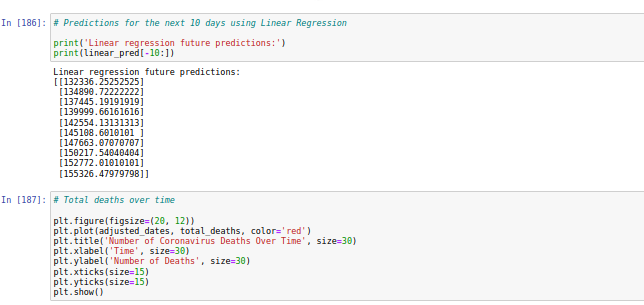
MSE: Mean Squared Error

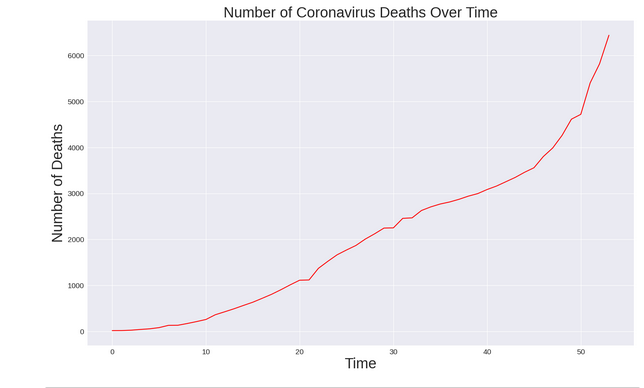




After establishing the build for linear regression we can now use the algo to determine the outcomes for the next ten days varying.

Below is a list of resultant x y values for a graph estimated for the next ten days based on current trajectories





Predictive models for COVID-19 have various potential applications in managing and mitigating the impact of the pandemic. These are some key areas where COVID-19 predictors can be applied:

1. **Resource Allocation:**
   * **Hospital Beds and Ventilators:** Predictive models can help forecast the demand for hospital beds, ICU beds, and ventilators, enabling healthcare systems to allocate resources effectively.
2. **Vaccine Distribution:**
   * **Optimizing Vaccine Distribution:** Predictive models can aid in planning and optimizing the distribution of vaccines by identifying areas with higher predicted infection rates or vulnerable populations.
3. **Public Health Interventions:**
   * **Targeted Interventions:** Predictors can guide the implementation of targeted public health interventions such as lockdowns, social distancing measures, and travel restrictions based on the predicted spread of the virus.
4. **Testing Strategies:**
   * **Optimizing Testing Strategies:** Models can help design efficient testing strategies by predicting areas with high infection rates, allowing for targeted testing and contact tracing.
5. **Trend Analysis:**
   * **Monitoring Trends:** Predictive models provide insights into the future trajectory of the pandemic, helping authorities and healthcare professionals anticipate surges, declines, or changes in the spread of the virus.
6. **Occupancy Planning:**
   * **Workplaces and Public Spaces:** Businesses and public spaces can use predictive models to plan for occupancy levels, implementing measures to maintain social distancing and reduce the risk of transmission.
7. **Supply Chain Management:**
   * **Medical Supplies:** Models can assist in managing the supply chain for medical equipment, ensuring that there are adequate supplies of personal protective equipment (PPE), testing kits, and other essential items.
8. **Public Communication:**
   * **Risk Communication:** Predictive models can inform public communication strategies, helping authorities convey the potential risks and necessary precautions to the public based on predicted trends.
9. **Research Prioritization:**
   * **Research Focus:** Predictive models can guide research priorities by identifying areas or populations that are likely to be most affected, helping researchers focus on critical areas of study.