

Daily Cumulative Confirmed Cases Prediction of Covid-19 in Italy

Xiaofu Geng
Anqi Xiong
Lulu Zha



Contents

01

Background

02

Dataset

03

Approach

04

Results

05

Limitations

01

Background

Background

December 2019, following Wuhan reports of a cluster of cases of new viral pneumonia, this epidemic of a novel coronavirus illness has become a global threat. 11th February, the World Health Organization(WHO) named it as COVID-19.



02

Dataset

Dataset

The original data set used in this project is provided by DingXiangYuan, which is collected by COVID-19 Infection Data Real-time Crawler. By plotting some graphs, we could get a general knowledge of the data set.

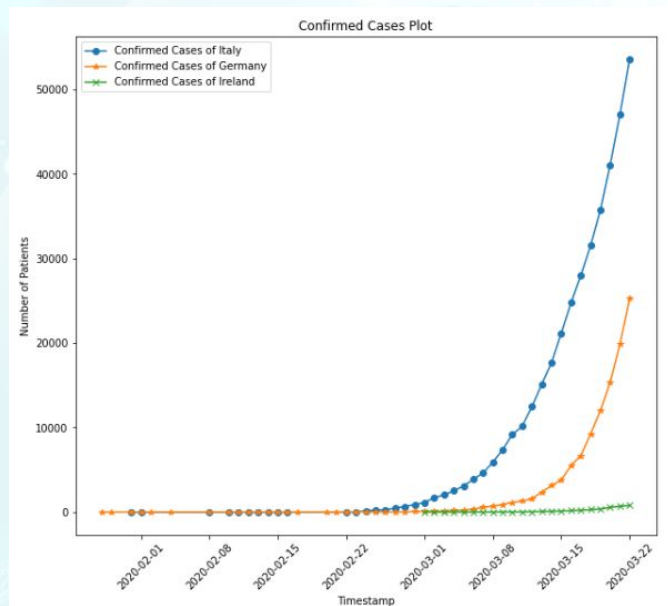


Fig.1

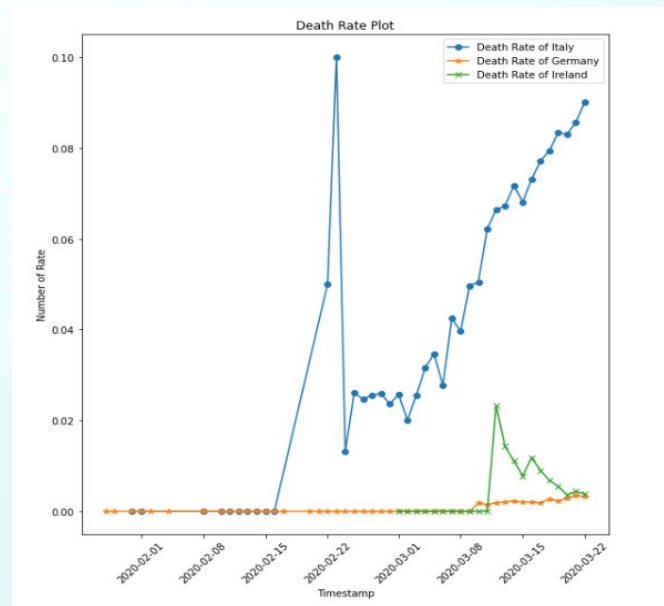


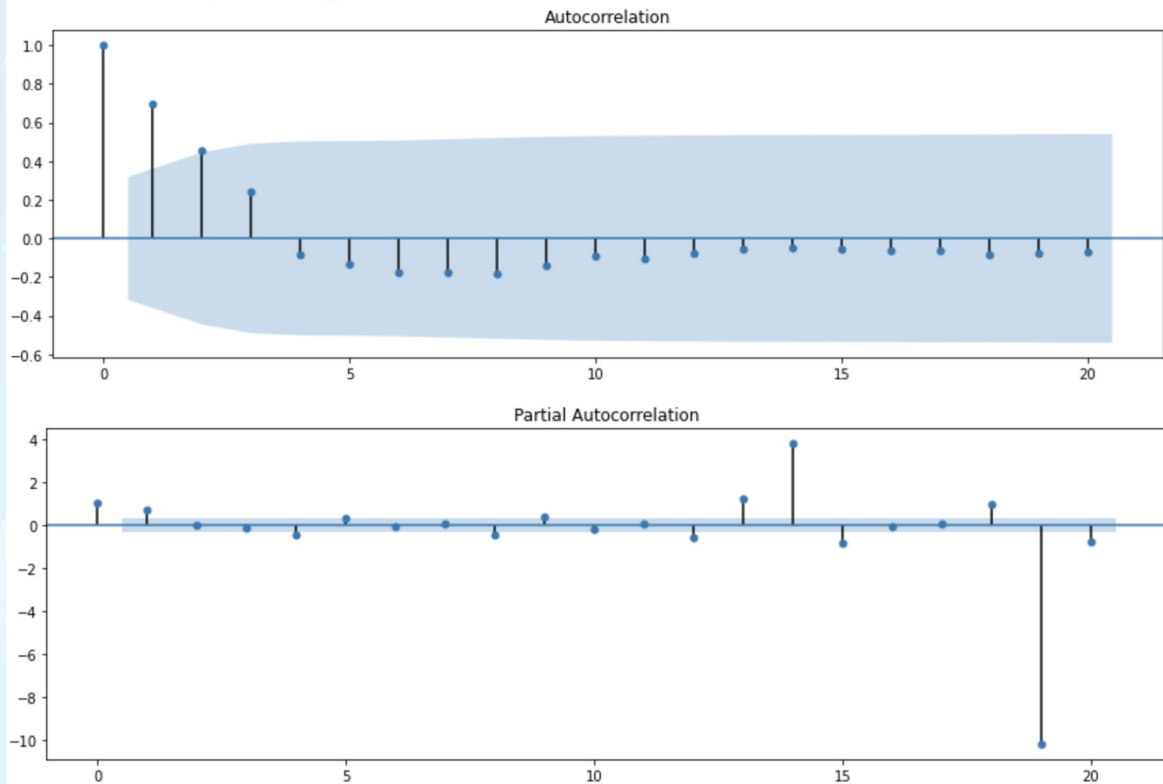
Fig.2

03

Approch



Model Selection



- ARIMA was chosen in this task
- ARIMA(1,2,2) were estimated by ACF and PACF
- AIC and BIC were used to help model selection

Model Evaluation - Residuals

- No autocorrelation between the residual values (Figure 1)
- The mean of the residual is 0.0117 - very close to 0
- Normally distributed (Figure 2)

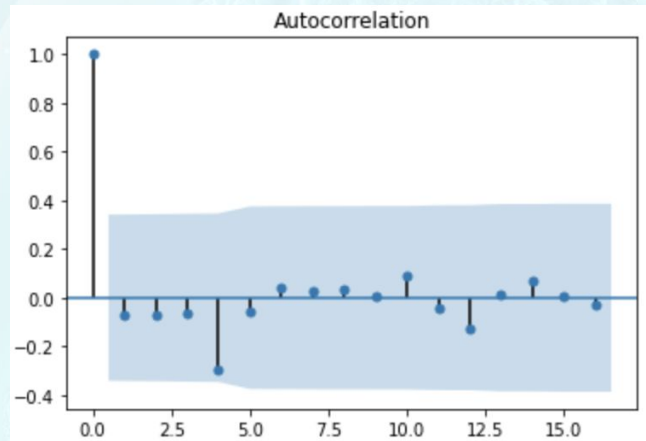


Figure 1

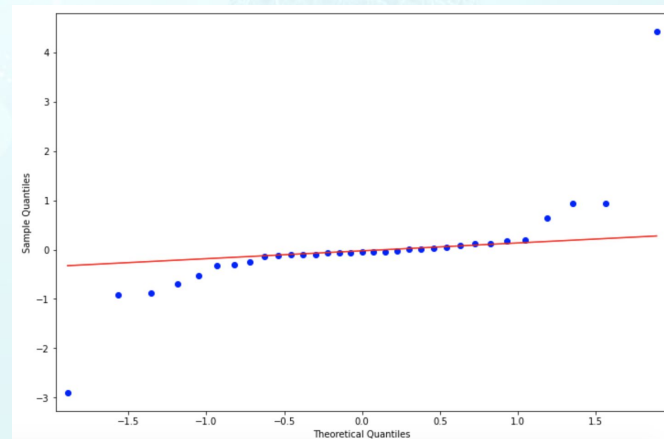
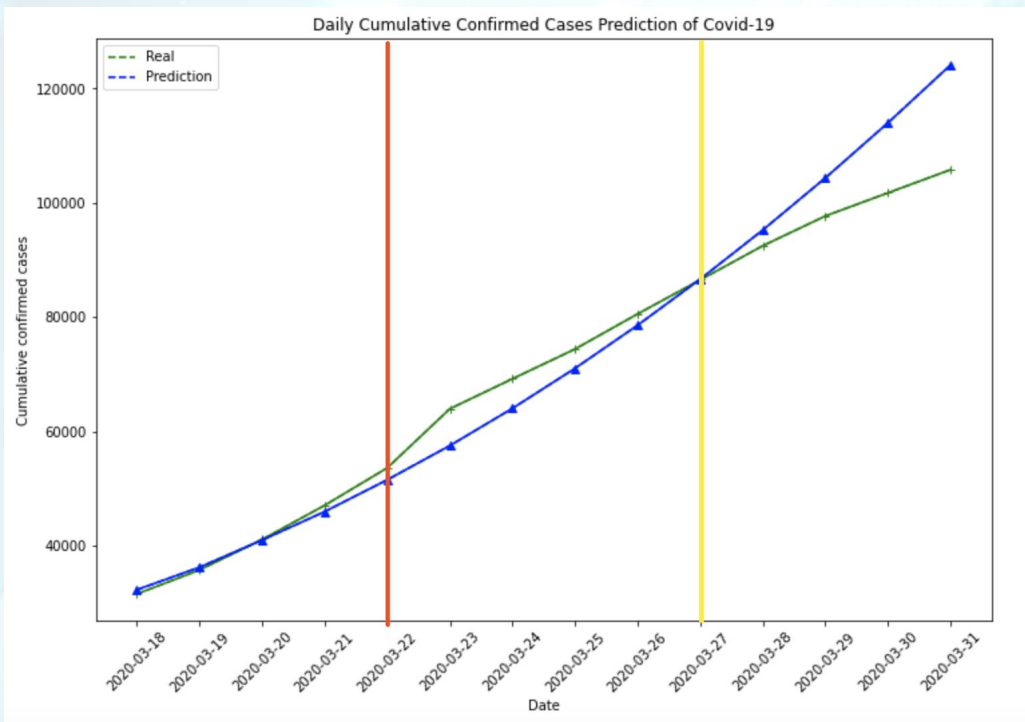


Figure 2

04

Results



- Outperformed in testset
- Performed well within 5 days
- Deviation gradually becomes large out of 5 days

05

Limitations



Limitations

- Limited by the size of dataset
- More external objective factors should be taken into account (lockdown, capacity of healthcare resources, etc...)



Task Division

Members	tasks
Anqi, Xiaofu	Data pre-processing
Anqi, Xiaofu, Lulu	Data visualization
Anqi, Lulu	ARIMA build and result analysis
Xiaofu	Random Forest model build but failed

The background is a soft-focus composition of various blue and green spheres of different sizes, some with a textured, bumpy surface. These are set against a light blue backdrop with wispy, ethereal patterns and numerous small, out-of-focus light spots (bokeh) in shades of blue and green. The overall effect is dreamlike and scientific.

Question?