CHAPTER 4
ANALYSIS AND DESIGN

4.1. Naïve Bayes and Epidemiology Analysis and Relation

One of the key success to suppress epidemic are case-control. It is simply how the government and public alike to understand major and minor details pertaining the epidemic itself. A simple act such as refraining from going to public places and isolating oneself affect majorly to case development for the next day. And thus, the action we took today will affect tomorrow's result. For the reason of this description alone, the usage of Naive Bayes theorem to predict cases development is deemed suitable. Naive Bayes theorem rely heavily on prior knowledge related to said event. As epidemic are sudden and unpredictable, using one of Naive Bayes applications, Bayesian Inference is one of the advantages for forecasting using Naïve Bayes theorem. It is a method that is used for updating probability of a hypothesis as more information becomes available. Its updating function is important in analyzing dynamic sequence of a data.

Bayesian inference interpreted probabilities differently than other statistical inferences. Probabilities in Bayesian inferences are not depicted as frequency but it is an expectation of a reasonable events, or as a quantity of a personal belief. Thus, Bayesian method uses' Bayes' theorem to update and computed probabilities result after new data are acquired.

Analysis of data using Bayesian models shows structure and pattern in a series of number or graphs, it could uncover future probability of a pattern suggested by past information, and the patterns perceived by the model itself. The simplicity of Bayes' theorem for the public to understand and also its effectiveness in predicting probabilities of events are the sole reason, Naïve Bayes are chosen for the algorithm used in this paper. Unknowingly, humans always use Bayes' theorem in real life, by calculating the probabilities or choosing something influenced by the past information gathered correlating to the events of choosing itself.

The usage of Bayes' theorem in medical field are proven quite effective especially in events of counting the probability of an individual with certain symptoms to suffer certain diseases and also drug testing. In the genetics area, Bayes' could also be used to calculate the probabilities of an individual having a certain type of genotype.

4.2. Virus Spreading Illustration

One of the simplest way to describe virus spreading in one closed environment is explained in the image below based on the paper by Wang et al[10][1].

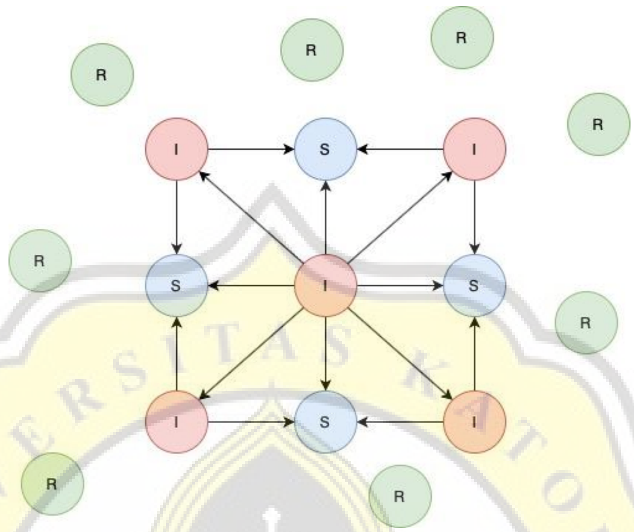

Figure 4-1 Example of Virus Spreading

In the image above, I represent Infected individual. When cases of epidemic spikes up, common mistakes that occurs in the first wave of epidemic are failure to recognize infected individuals and how the virus spreads. The spread of a virus is various, bodily fluid, air, skin-to-skin spreading. Identifying how the virus spread is the first thing to analyze when epidemic hits, thus proper precaution can be taken to suppress the number of infected individuals. S represent Susceptible individual. This correlates to what the virus attacks in human body, immune system, respiratory health, skin health. After analyzing the what the virus attacks, a number of people in a population can immediately grouped as susceptible. A susceptible individual has a high rate of being infected compared to other individual. And R, represent Recovered individual. A recovered individual has boosted immune system after infection, therefore the chance of recovered individual to be infected again are much slimmer than susceptible individual.

---

[1] W. Wang, Q.-H. Liu, L.-F. Zhong, M. Tang, H. Gao, and H. E. Stanley, "Predicting the epidemic threshold of the susceptible-infected-recovered model," *Sci Rep*, vol. 6, no. 1, p. 24676, Jul. 2016, doi: 10.1038/srep24676

4.3. Time Series Data of SARS 2003

Table 4-1 Time Frame Data

No	Date	Region	Number of Case	Number of Deaths	Number of Recovery
1	17/03/03	Hong Kong SAR, China	95	1	0
2	18/03/03	Hong Kong SAR, China	123	1	0
3	19/03/03	Hong Kong SAR, China	150	5	0
4	20/03/03	Hong Kong SAR, China	173	6	0
5	21/03/03	Hong Kong SAR, China	203	6	0
6	22/03/03	Hong Kong SAR, China	222	7	0
7	24/03/03	Hong Kong SAR, China	260	10	0
8	25/03/03	Hong Kong SAR, China	286	10	0
9	26/03/03	Hong Kong SAR, China	316	10	0
10	27/03/03	Hong Kong SAR, China	367	10	0
11	28/03/03	Hong Kong SAR, China	425	10	0
12	29/03/03	Hong Kong SAR, China	470	10	0
13	31/03/03	Hong Kong SAR, China	530	13	0
14	01/04/03	Hong Kong SAR, China	685	16	0
15	02/04/03	Hong Kong SAR, China	708	16	0
16	03/04/03	Hong Kong SAR, China	734	17	0
17	04/04/03	Hong Kong SAR, China	761	17	0
18	05/04/03	Hong Kong SAR, China	800	20	0
19	07/05/03	Hong Kong SAR, China	883	23	0
20	08/04/03	Hong Kong SAR, China	928	25	0

Epidemiology case used in this research are from SARS 2003 Epidemic in Hong Kong, China. According to the World Health Organization, an astounding number of 8.908 people globally become sick with SARS Virus during the 2003 period. Death tolls reached 774 people. The virus spreads by having close contact with the infected individual such as living in the same place or by direct contact with bodily fluid of an infected individual.

The first person reported with the SARS symptoms is in Asia and it periodically spreads to more than 2 dozen countries. Accumulative data on SARS 2003 Epidemic outbreak in Hong Kong are obtained and analyzed to prevent double data. Total cases in Hong Kong reached 1755 cases with 298 deaths and also 1433 recovered individual, with a total of 24 missing record data as to whether the individual recovered or died of the virus.

By the timeframe above, it is clear as how the virus spread quickly by the lack of precaution and the number of deaths also climbed up in the first months of the epidemic without record of recovered individual. Lack of precaution and public awareness of the epidemic causing the direct result of the high number of case which can be suppressed to a minimal.

4.4. Time Series Design

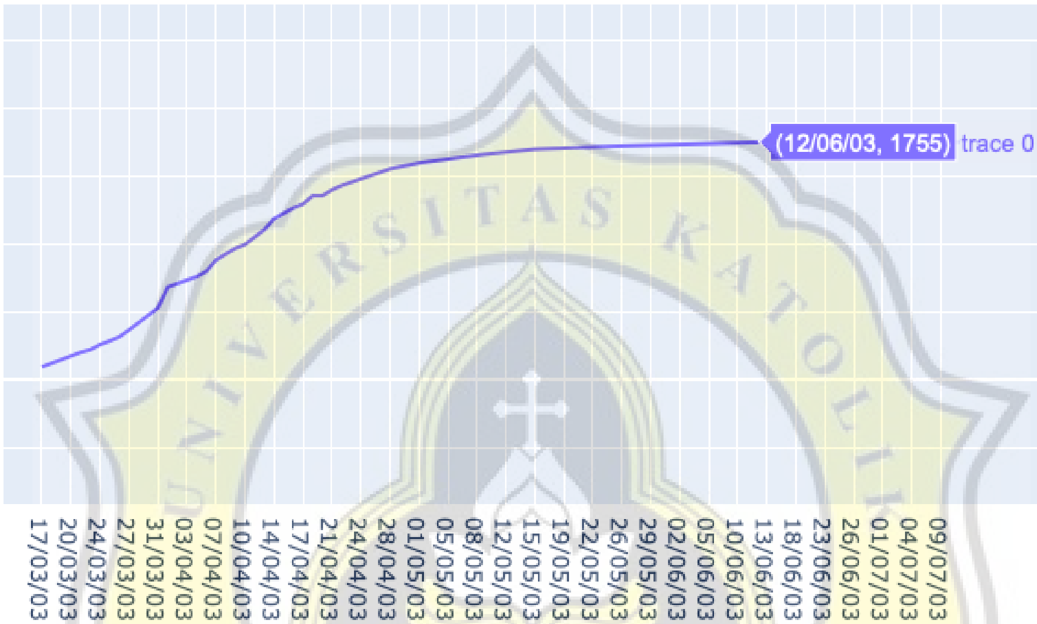

Figure 4-2 Time Series Graph on Total Cases

Based off the data accumulated from the Internet from various sources and also news pages[2], a graph has been created to show-case development of SARS-CoV-1 virus in Hong Kong. The last documented case of SARS-Cov-1 in Hong Kong was on June, 12^{th} 2003. In a span of 4 months, 1755 people were diagnosed with SARS-CoV-1 virus. These numbers could've been suppressed by government's support and also public's awareness of the virus symptoms, how to avoid it and how to contain it once infected. The rate of virus spreading was too high and government were not ready to handle epidemic such as these.

---

[2] https://www.kaggle.com/imdevskp/sars-outbreak-2003-complete-dataset
https://www.who.int/csr/sars/country/en/

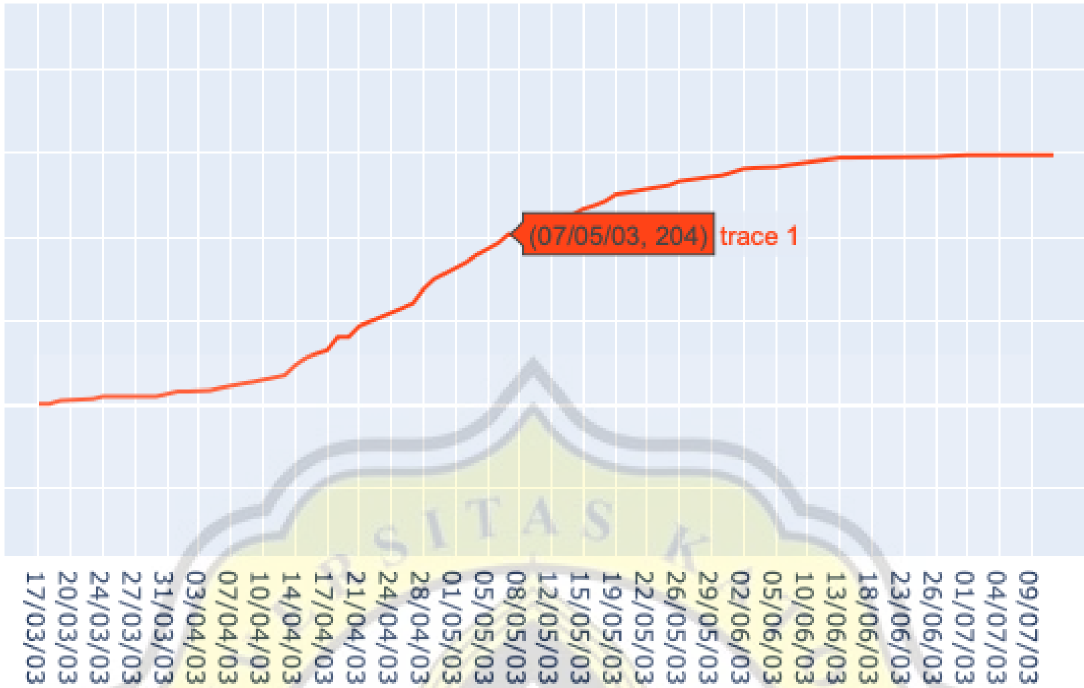

Figure 4-3 Time Series Graph on Total Deaths

This graph on the other hands, shows the rate of deaths by the virus. It shows an exponential growth on the second month since the virus started. The aforementioned graphs the growth of the total case and the total deaths of the SARS-Cov-1 virus in Hong Kong without any kind of containment measures, thus the growing numbers. Eventually, the rate of virus spreading stopped due to containment being held on locations where cluster of cases happened, such as Amoy Gardens and also the hospitals around it, but these containments could've been done faster if only the government were well aware of the epidemic cases.

SARS-Cov-1 had grown from an outbreak to an epidemic, alike as the SARS-Cov-2 that happened on 19th December 2019 [3]. A sudden rise of case number of a disease is called an outbreak, and the difference from an outbreak to an epidemic is the number of cases and also the people affected, epidemic is when an infectious disease spreads rapidly to many people. If an epidemic isn't well contained, it could expand and become a pandemic. A pandemic, like-wise an epidemic,

[3]https://apps.who.int/iris/bitstream/handle/10665/332197/WHO-2019-nCoV-FAQ-Virus_origin-2020.1-eng.pdf

is a global outbreak, it affects wider area geographically and infects much more people than epidemic.

SARS-Cov-1 and SARS-Cov-2 development were not much different, so as SARS-CoV-1 had happened, governments of every nation in the world could've prepared for a sudden outbreak of a certain infectious diseases.

A collaboration between national health department and local public health center could prevent these diseases to spread rapidly, after these collaborations have been established, local health center could educate smaller group of society in radius to said local health center to help prevent the spread of the virus. Clear information and communication are one of the key to prevent the spikes of another virus outbreak.

"For fast epidemics, the estimation is subject to over-fitting due to the limited number of data points available, which also limits our choice of models for the epidemic curve" [4]

As quoted from the paper by Junling Ma, simpler models are used for fast forecasting due limited data points and late data accumulation, simpler models are managed as an emergency to suppress cases growth for the next day. Time and speed are two major points once an outbreak occur, and simpler models helped government and public alike to be aware of the virus spreading rate and how to contain it.

[4] Junling Ma,"Estimating epidemic exponential growth rate and basic reproduction number," vol. 05, pp. 129-141, 2020, doi: 10.1016/j.idm.2019.12.009.

4.5. Naïve Bayes Function

$$\widehat{\Upsilon}_{t+1} = \Upsilon_t$$

Equation 4-1 Simple Naive Formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Equation 4-2 Naive Bayes Formula

On the first equation, the result of t (time) + 1 prediction is equals to the result of t. Meaning tomorrow's prediction equals to today's result Naïve Bayes formula is the likeliness of an occurrence based on the prior probability of a class.

By examining both Naive time series formula and Naive Bayes formula above, predicting time series data can be quite simple to understand and to apply as well. The naïve time series formula simply set all forecasts to be the value of the last observation. And the Bayes theorem calculates the prediction of posterior probability given the value of predictor on a given class.

Both formulas are extremely simple yet surprisingly effective used as a benchmark for forecasting. These formulas can be developed to achieve higher accuracy results.