

Quality and Performance Excellence Based on the Big Data Management

Gye-Soo Kim

¹Business Department, Semyung University,

Abstract: Fierce competition, new technologies and the ability to gain and leverage data-driven insights in a timely manner are important. Big data is a source and source of decision making. Leading companies are trying to maintain an excellent quality level based on big data management and analysis results. It also focuses on developing new products based on big data analysis results. In this study, we will present the basic procedure of big data analysis and look at the case of companies that lead big data analysis results in management decision making in ecommerce area. The online shopping mall found hygiene factor and price factor are very significant factor for customer satisfaction at ecommerce due to Covid-19. This result is expected to present strategic implications to related industries.

Keywords: *Big Data, Excellent Quality, New Products, Big Data Analysis, Hygiene Factor, Covid-19.*

I. INTRODUCTION

Big Data is disrupting every aspect of business. Big Data is a game changer for the operation management and Total Quality Management of Business activities – not just the next new strategy but the tsunami disrupter that business person to know before your competition them over it.

Many people and organizations understand the power and importance of Big Data. But they fail to use it effectively. Capable companies can utilize and establish excellent strategy and action plan based on the data.

From retailer using Big Data to predict trends and customer behavior, public sector using Big Data to provide excellent service. As possible as the use of Big Data in cities, telecoms, sports, gambling, fashion manufacturing, research, motor racing, video gaming, and every business area in between-Leaders will realize that no matter what job they are industry you work in, Big Data will transform.

Today every organization needs ne cloud-based infrastructure and applications that can convert vast amount of data into predictive and analytical power through the use of advanced analytics, machine learning and AI[1].

Big data is called the 21st century crude oil and gold mine. It is emerging as a new growth engine and a new paradigm for various industries, countries, politics and academia. Accumulation of knowledge through data is a factor of production such as capital and labor, and serves as a new window for business. The countries that dominated the data in the corona crisis responded wisely. You need to know how the virus works and the path the infected person is going to survive on the battlefield. If you don't get data quickly and the update cycles are jagged, you will have to lag behind in times of crisis, and things will get worse.

COVID-19 pandemic has severely impacted the automotive sector, tourism industry, aviation industry, oil industry, construction industry, telecom sector, food industry, and healthcare industry [2]. Initial cases were reported in December 2019 and viewed as symptoms of pneumonia in the Wet Markets of Wuhan City in Hubei Province, China [3], and later named as COVID-19.

Since the outbreak of Corona Virus Diseases – 2019 (COVID-19), one of the intricate issues is to estimate the expected damage of human health and the global economy, and it becomes much more complicated while considering individual parts of the world. Frequent changes are noticed in developing the combating policies and plans because of high growth in cases and ambiguous data of infection rate, mortality rate, and other statistics [4].The number of confirmed coronavirus infections (Corona 19) around the world exceeded 20 million. According to the statistics website Waldometer, as of 00:00 on August 9, 2020, the number of cumulative corona19 confirmed cases in the world increased by 26,2524 from the previous day to 19,795,061. Considering the trend of increasing more than 200,000 people a day, it actually exceeded 20 million as of this day. The 20 million mark is only 223 days after China reported the outbreak of Corona 19 to the World Health Organization (WHO) on December 31 last year, mainly in Wuhan, Hubei Province. On June 28, 180 days after the outbreak of Corona 19, it took only 43 days to add 10 million people after surpassing 10 million. The number of confirmed cases by country was 5.15 million in the US, 3.01 million in Brazil, and 2.15 million in India. In the United States and Brazil, there are still more than 50,000 confirmed cases a day. India is spreading faster recently, with 65,156 confirmed cases coming out the day before and recording the highest record. Concerns are growing in Europe, the Middle East, and Southeast Asia as well.

As Corona 19 continues, the economic situation is deteriorating, and companies are busy preparing for the era of with Covid-19. Firms with ample funding are deeply researching customer needs and preparing for digital transformation. However, companies with insufficient funding power are getting more difficult as they follow past management methods. Corona 19 is accelerating the digital revolution. Big data-based decision making, the source of the digital revolution, is taking place.

The objectives of this research are the followings. First, we will propose the research model that quality linked performance excellence based on big data management. We will have to redefine big data and role of big data.

Secondly, we will provide the management method of big data using practical example of big data management. We will analyze customer big data of online shopping companies that have achieved excellent management performance during the Corona 19 period to find out the purchasing decision factors of customers. Next, regression analysis will be conducted to find

*This work is a study supported by Semyung University Research Years in the 2020 school year.

out the influence of purchasing decision factors on customer satisfaction factors.

Thirdly, we will give some implications and action plans about big data management to quality manager who take care of quality division.

II. BIG DATA MANAGEMENT

The combination of cloud computing, sensors, Big Data, machine learning, and Artificial Intelligence (AI), mixed reality, and robotics foreshadows socioeconomic change rippled from the page of new trend[5]. In specific, three breakthrough-Big Data, massive computing power, and sophisticated algorithms-is accelerating AI. At astonishing rates, data is being gathered and made available thanks to the exponential growth of camera and sensors in our everyday life. AI needs data to learn.

It needs to be great at applied machine learning (ML). ML is a very rich form of data analytics that is foundational to artificial intelligence. The procedure of big data analysis is as follows.

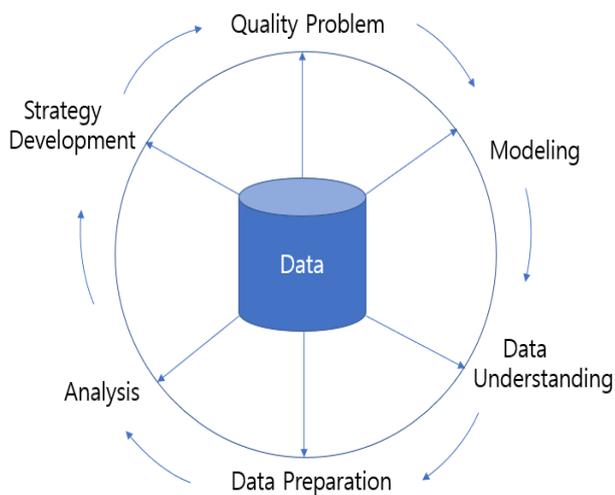


Fig. 1. Quality model based on big data

III. QUALITY PROBLEM

Quality problems should be observed from the external point of view, not from the internal point of view. What is the most uncomfortable thing for customers today? What do you want your customers to solve for your company? We have to constantly worry about it. This can be called the problem definition of customer perspective (PDOSP). Defining the problem internally can achieve efficiency, but loses the effectiveness of the external customer perspective. Organizations can define key issues as data uploaded by customers online, problems found in customer monitoring, and so on. In this study, we will analyze the big data related to customers of online shopping malls whose sales have increased sharply during the period of Corona 19 to determine the factors that determine online shopping malls in Korea.

IV. MODELING

Model is an abstraction of the reality. The model contains the direction of research and the goals the research seeks. Model can be represented by mathematical equations, pictures, and linguistic research hypotheses. The mathematical expression is a mathematical expression of the phenomenon. The pictorial model represents the abbreviation of the phenomenon, and the hypothesis expresses a mathematical

expression or a pictorial model as a verbal expression.

Modeling is the competency that Data Science should have as a career group in the 4th industrial revolution. It is most important to be able to describe the phenomenon in a mathematically correct or pictorial way. In order to be good at modeling, it is necessary to think from the customer's point of view, to take daily notes, and to practice mathematical thinking.

In this study, the purpose of this study is to investigate the effect of online purchasing decision factors on customer satisfaction in the Corona 19 situation. This is illustrated as follows.

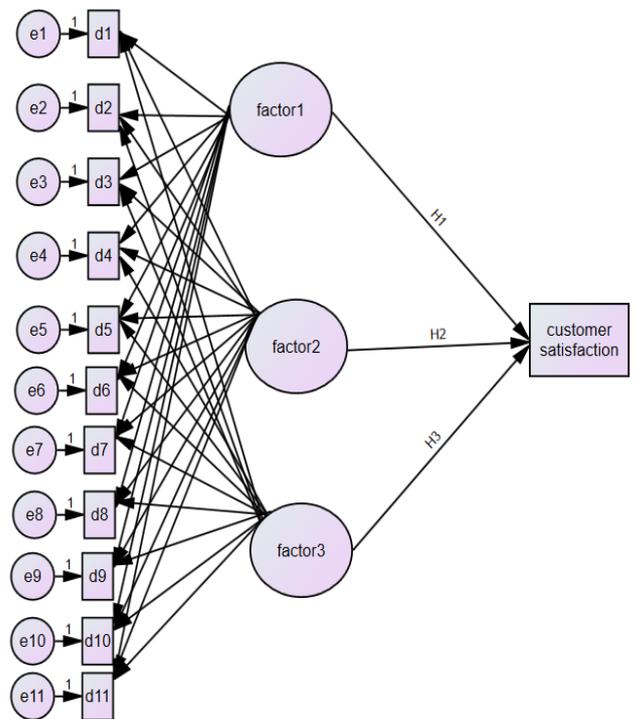


Fig. 2. Influence of purchasing decision factors on customer satisfaction

V. DATA UNDERSTANDING

The company's worries are how to manage the data accumulated every day. Various data can be secured in the manufacturing line and various information can be accumulated in exchange with customers. If you look closely at this vivid data, you can find clues to the problem. However, most organizations neglect the collection and management of big data that is the basis of decision making. From the customer database, 141 valid samples were extracted and analyzed. Data files include online purchasing patterns, product and service quality levels, and satisfaction with online shopping mall. As a result, it is often impossible to produce effective results whenever an issue arises, or when an issue emerges and is investigated and analyzed in an urgent condition. The saved file format is useful when importing with .csv (comma separated value).

For this study, a data file was obtained that surveyed the reactions to customers who use the A online shopping mall. Among the big data files, sensitive information about customers was not received during the delivery process.

VI. DATA PREPARATION

The process of understanding the data subject to analysis is the outpost of full-scale analysis. In relation to the research model set up above, it is necessary to decide which variable to

extract from the data and how to handle if there is a missing data. Variable extraction is not done by the researcher alone, but by research team members and colleagues. When there is a non-response value, determining whether to replace the non-response value with an average or removing a case with a non-response value from the analysis object can be said to be a process of preparing data. The data received from the online shopping mall was refined through a Python program.

Table 1: Data wrangling

```
import pandas as pd
import numpy as np
import seaborn as sns
import xlrd
from factor_analyzer import FactorAnalyzer
import matplotlib.pyplot as plt
df = pd.read_csv('C:/gskim/researchyear/researchyear/factordata.csv')
df.columns
df
df.describe()
#Some basic statistics of the numerical variables in the dataset.
df.shape
# Missing values1
print(df.isnull().values.any())
# Missing values2
print(df.isnull().sum())
# Replace missing value
print(df.isnull().sum())
```

VII. ANALYSIS

The method of analyzing big data should be suitable for the purpose of analysis. There are three main purposes of analyzing big data: description, control, and prediction. Description can present data in pictures, tabular form, or basic statistics. Control refers to the case where variables with low significance are excluded or deleted from decision making among variables included in the research model. Prediction refers to obtaining the value of the future thought by numerically substituting the expected value or organizational will into a mathematical or estimated equation. Machine learning and artificial intelligence are often used as predictions.

In the analysis process, the selection of statistical programs is also important. Recently, data science and young researchers' preferred programs are Python and R programs. Since these programs are non-commercial versions, they are very cost-effective and provide various libraries, so anyone with a little understanding of the algorithm can access them. In this analysis, a Python program was used.

The ability to visualize analytical results is also important. Numerical results can also provide persuasive power, but showing results in pictures can bring even greater appeal. In this study, exploratory factor analysis was conducted to find out the purchasing decision factors of customers. Then, a regression analysis was conducted to analyze the influence of factors on customer satisfaction. Prior to this, correlation analysis was conducted as follows to examine the relationship between variables.

Table 2: Analysis

```
# Compute correlations
corr = X.corr()
```

```
# Exclude duplicate correlations by masking upper right values
mask = np.zeros_like(corr, dtype=np.bool)
mask[np.triu_indices_from(mask)] = True

# Set background color / chart style
sns.set_style(style = 'white')

# Set up matplotlib figure
f, ax = plt.subplots(figsize=(11, 9))

# Add diverging colormap
cmap = sns.diverging_palette(10, 250, as_cmap=True)

# Draw correlation plot
sns.heatmap(corr, mask=mask, cmap=cmap, square=True, linewidths=.5, cbar_kws={"shrink": .5}, ax=ax)
```

The results of the correlation analysis are as follows.

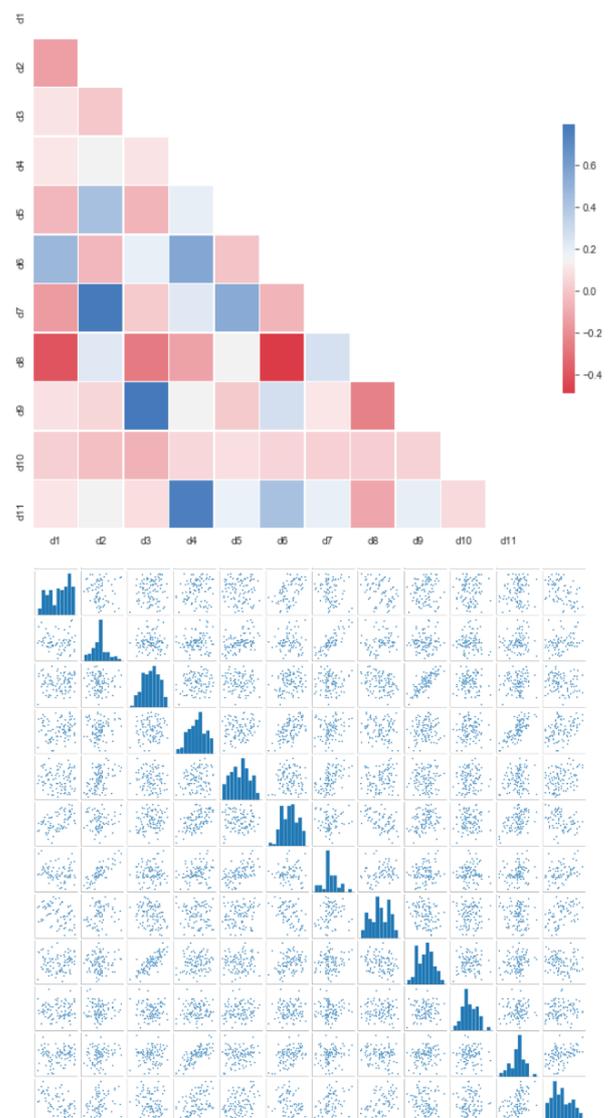


Fig. 3: Correlation Analysis

From the figure, it can be seen that the variables d4, d7, and d11 are roughly related. The variables d2, d5, d7, and d8 are roughly related. And d3, d9 are significantly related. These related variables are very likely to be grouped as one factor in factor analysis. Exploratory factor analysis was conducted using Python in the following procedure.

Table 3 Factor Analysis

```
X = df[["d1", "d2","d3","d4", "d5","d6","d7","d8","d9","d10","d11"]]
y_use = df["y1"]
fa = FactorAnalyzer(rotation='varimax', n_factors=3)
fa.fit(X)
fa_lo = fa.loadings_
print(fa_lo)
fa.get_communalities()
from factor_analyzer.factor_analyzer import calculate_bartlett_sphericity
chi_square_value,p_value=calculate_bartlett_sphericity(X)
chi_square_value, p_value
from factor_analyzer.factor_analyzer import calculate_kmo
kmo_all,kmo_model=calculate_kmo(X)
kmo_model
import numpy as np
fa_df = pd.DataFrame(fa_lo)
X_fa = np.dot(X, fa_df)
X_fa
X_fa_df = pd.DataFrame(X_fa)
X_fa_df.columns=['hygienefactor','productfactor','pricefactor']
X_fa_df.head()
ds_fa = pd.concat([df, X_fa_df], axis=1)
ds_fa.head()
```

	id	c1	c2	c3	c4	c5	d1	d2	d3	d4	...	y1	y2	y3	y4	y5	y6	y7	hygienefactor	productfactor	pricefactor
0	1	2	0	1	1	1	8.5	3.9	2.5	5.9	...	5.1	3.7	8.2	8.0	8.4	65.1	1	13.177272	14.829940	7.262280
1	2	3	1	0	0	0	8.2	2.7	5.1	7.2	...	4.3	4.9	5.7	6.5	7.5	67.1	0	7.709562	17.470771	11.257371
2	3	3	0	1	1	1	9.2	3.4	5.6	5.6	...	4.0	4.5	8.9	8.4	9.0	72.1	1	11.395724	18.129866	12.679089
3	4	1	1	1	1	0	6.4	3.3	7.0	3.7	...	4.1	3.0	4.8	6.0	7.2	40.1	0	11.987811	11.105173	12.248963
4	5	2	0	1	0	1	9.0	3.4	5.2	4.6	...	3.5	3.5	7.1	6.6	9.0	57.1	0	9.260060	14.176765	11.146976

Fig. 4. Result of Factor Analysis

After factor analysis, the names of the factors were named as hygiene factors, product factors, and price factors considering of the nature of the variables. Because of Covid-19, customers mainly use online shopping, and in particular, it can be seen that hygiene factors, product factors, and price factors are the main reasons for using online shopping malls.

VIII. QUALITY PERFORMANCE

Quality is the user's suitability for use or excellence in experience. Quality is the user's suitability for use or excellence in experience. Excellent quality is the source of customer satisfaction and management performance (Kim 2020). Customers who have experienced excellent products and services will continue to use them and recommend them to others or upload a positive word of mouth on SNS.

Collection and continuous management of big data are important, but it is more important to analyze big data to find meaningful points. New analytical techniques of big data will

continue to emerge and develop in the future. And the demand for human resources in the fields related to big data will surge.

To make good use of big data, collaboration between different fields is also required. In Amazon, when forecasting demand and revenue, one department predicts based on big data techniques such as machine learning, and the other competes with the accuracy of forecasting at regular intervals by deriving prediction results based on economic models. So far, the big data technique is producing better results. However, global IT companies such as Amazon and Google are steadily raising economic research teams. If the existing studies such as economics strengthen the understanding and use of the results of analyzing big data, the value of utilizing big data will become greater in the future. In order to find out the influence of the hygiene factors, product factors, and price factors, which are the purchasing decision factors of online shopping malls, on the degree of customer satisfaction, a regression analysis was performed as follows.

Table 4: Regression Analysis

```
import statsmodels.formula.api as sm
result = sm.ols(formula='y_use ~ hygienefactor + productfactor + pricefactor', data=ds_fa).fit()
print(result.summary())
```

OLS Regression Results

Dep. Variable:	y_use	R-squared:	0.420
Model:	OLS	Adj. R-squared:	0.402
Method:	Least Squares	F-statistic:	23.20
Date:	Sun, 09 Aug 2020	Prob (F-statistic):	2.22e-11

Time:	12:49:25	Log-Likelihood:	-132.86
No. Observations:	100	AIC:	273.7
Df Residuals:	96	BIC:	284.1
Df Model:	3		
Covariance Type:	nonrobust		

coef	std err	t	P> t	[0.025	0.975]	
Intercept	2.0130	0.766	2.628	0.010	0.493	3.533
hygienefactor	0.2895	0.040	7.168	0.000	0.209	0.370
productfactor	0.0448	0.040	1.127	0.262	-0.034	0.124
pricefactor	-0.1282	0.047	-2.740	0.007	-0.221	-0.035

Omnibus:	0.543	Durbin-Watson:	1.834
Prob(Omnibus):	0.762	Jarque-Bera (JB):	0.667
Skew:	-0.044	Prob(JB):	0.716
Kurtosis:	2.610	Cond. No.	179.

Fig. 5. Result of Regression Analysis

As a result of regression analysis, the explanatory power of the estimated regression equation is 40.2%. And it was found that hygiene factors and price factors had a significant effect on customer satisfaction at $\alpha=0.05$ level. In particular, the price factor was found to have a negative and significant effect on customer satisfaction. This can be inferred that customers are dissatisfied when the price level is high compared to the quality. Based on this, online shopping mall companies are constantly striving to complete an assortment of excellent product lines. In addition, in order to reduce the price as much as possible, it is making efforts not only to reduce costs but also to systematically manage the supply chain based on ICT.

In the unprecedented situation of covid-19, customers prefer online shopping with high non-face-to-face contact. It can be seen that the hygiene factor is considered when shopping online and the high price level acts as a dissatisfaction factor. A box plot was conducted to find out the degree of satisfaction according to sex.

Table 5: T-test

```
import pandas as pd
import math
import numpy
import plotly.express
import plotly.offline
import scipy
import statsmodels.api
import statsmodels.formula.api
import patsy.contrasts
plotly.offline.init_notebook_mode(connected = True)
# boxplots of y by group
plotly.offline.iplot(
plotly.express.box(df, x = "c5", y = "y1")
)
```

As a result of the analysis, it can be seen that women are more satisfied with online shopping than men. This suggests that online shopping companies need to focus on targeting female customers. It is important to organize data-based quality problems and reflect them in strategies. Strategy is differentiation and uniqueness that distinguishes it from competitors. It is necessary to establish a differentiated strategy,

share it with members of the organization, and act quickly so that customers can experience it.

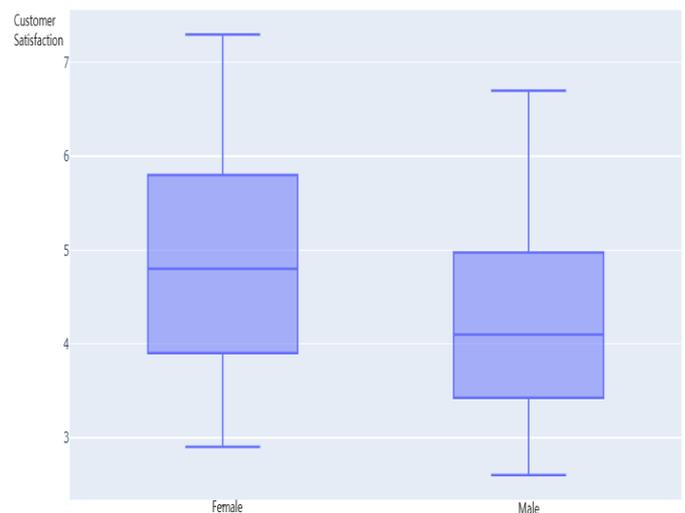


Fig. 6. Customer satisfaction by gender

CONCLUSION AND IMPLICATIONS

In Korea, covid-19 and the digital revolution are accelerating online shopping. Covid-19, unprecedented in history, is advancing the digital revolution. In the future, there will be more companies seeking digitization for rapid crisis management. The core of the future management strategy is digitization, and new technologies such as big data, the Internet of Things (IoT), blockchain, and robots will be used more.

Organizations need to be great at applied machine learning for keep excellent quality. Machine learning is a very rich form of data analytics that foundational to artificial intelligence. It's important to be dedicated to creating great products, serving customers, and earning profits for our investor-but it's not sufficient. As a result of analyzing purchasing decision factors based on data of online shopping customers, it was found that customers place hygiene factors most importantly. This was found to be due to a special situation called Covid-19. As the price factor appears to act as a factor of dissatisfaction, it can be inferred that customers actively consider cost-performance as well.

Organizations need a sophisticated understanding of how to do two things at once-discern the internet of someone searching the Web and then match that intent with accurate knowledge gained from crawling the Web, ingesting and understanding information.

Big data is the fuel for decision making. In the future, competitiveness will vary depending on how effectively accumulated data is used(Kim 2020). We no longer lived in a PC-Centric world. Computing was becoming more ubiquitous. Intelligence was becoming more ambient, meaning computers could observe, collect data, and turn feedback into insights. Organizing data is important in understanding data, and I insist on organizing and storing data in a relational database with matrices as a way to manage data. If you have a habit of managing with a relational database, you can easily analyze it anytime and it is easy to call it regardless of which program you use. If you systematically manage and analyze the data, you can see quality problems and discover critical to quality from a customer perspective.

*This work is a study supported by Semyung University Research Years in the 2020 school year.

References

- [1] Satya Nadella(2017), Hit Reflesh, Willam Collins.
- [2] Chamola, Vinay , VikasHassija, VatsalGupta, and MohsenGuizani . 2020. “A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing Its Impact.” *IEEE* Access 8: 90225–0265. [Crossref], [Web of Science ®], [Google Scholar]
- [3] Rothan, HussinA. , and Siddappa N.Byrareddy . 2020. “The Epidemiology and Pathogenesis of Coronavirus Disease (COVID-19) Outbreak.” *Journal of Autoimmunity* 109: 102433. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]
- [4] Prichet, Elissa. 2020. *Why a Coronavirus Vaccine Takes Over a Year to Produce – and why That is Incredibly Fast* . World Economic Forum. April 3. <https://www.weforum.org/agenda/2020/04/why-a-coronavirus-vaccine-takes-over-a-year-to-produce-and-why-that-is-incredibly-fast/> . [Google Scholar]
- [5] Kim, Gye, Soo. 2020., “The Effect of Quality Management and Big Data Management on Customer Satisfaction in Korea’s Public Sector,
- [6] *Sustainability* 20, 12(13), 5474; <https://doi.org/10.3390/su12135474> - 07 Jul 2020.