**SAFEPASS:**

**AN IMPLEMENTATION OF AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) FOR CROWD FORECASTING APPLIED IN QUARANTINE PASS**

A Capstone Project Presented to the Graduate Program

College of Engineering and Technology

Pamantasan ng Lungsod ng Maynila

In Partial Fulfillment of the Requirements for the Degree

Master’s in Information Technology

­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

By

Joane Marie F. Llamera

Dr. Khatalyn E. Mata

Thesis Adviser

August 2021

**APPROVAL SHEET**

The capstone project hereto titled

**SAFEPASS:**

**AN IMPLEMENTATION OF AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) FOR CROWD FORECASTING APPLIED IN QUARANTINE PASS**

prepared and submitted by Joane Marie F. Llamera in partial fulfilment of the requirements for the degree of Master’s in Information Technology has been examined and is recommended for acceptance and approval for **Oral Examination**.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**DR. KHATALYN E. MATA**

Adviser

PANEL OF EXAMINERS

Approved by the Committee on Oral Examination

with a grade of \_\_\_\_\_\_\_\_\_ on \_\_\_\_\_\_\_.

**PROF. MANUEL L. OCAMPO**

Panel Chair

Chairman

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Panel Member Panel Member

Member Member

Accepted and approved in partial fulfilment of the requirements for the degree of

Master’s in Information Technology.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Denvert C. Pangayao Dr. Clydelle M. Rondaris

Director Dean

Graduate Program College of Engineering and Technology

**ACKNOWLEDGEMENTS**

**TABLE OF CONTENTS**

[APPROVAL SHEET i](#_Toc305755519)

[ABSTRACT ii](#_Toc305755520)

[ACKNOWLEDGEMENTS **Error! Bookmark not defined.**](#_Toc305755521)

[TABLE OF CONTENTS ii](#_Toc305755522)

[LIST OF FIGURES ii](#_Toc305755523)

[LIST OF TABLES ii](#_Toc305755524)

[INTRODUCTION 2](#_Toc305755525)

[1.1 Background 2](#_Toc305755526)

[1.2 Fonts **Error! Bookmark not defined.**](#_Toc305755527)

[1.3 Paragraphs and Line Spacing 2](#_Toc305755528)

[1.3.1 Section numbering **Error! Bookmark not defined.**](#_Toc305755529)

[1.3.2 Lists 2](#_Toc305755530)

[1.4 Significance of the Study 2](#_Toc305755531)

[1.5 Scope and Limitations **Error! Bookmark not defined.**](#_Toc305755532)

[REVIEW OF RELATED LITERATURE 2](#_Toc305755533)

[2.1 Referencing within the Text 2](#_Toc305755534)

[THEORETICAL FRAMEWORK 2](#_Toc305755535)

[3.1 Equations **Error! Bookmark not defined.**](#_Toc305755536)

[MATERIALS AND METHODOLOGY 2](#_Toc305755537)

[4.1 SI Units **Error! Bookmark not defined.**](#_Toc305755538)

[4.2 Figures **Error! Bookmark not defined.**](#_Toc305755539)

[RESULTS AND DISCUSSION **Error! Bookmark not defined.**](#_Toc305755540)

[5.1. Tables **Error! Bookmark not defined.**](#_Toc305755541)

[CONCLUSIONS AND RECOMMENDATIONS **Error! Bookmark not defined.**](#_Toc305755542)

[LIST OF REFERENCES 2](#_Toc305755543)

[APPENDIX A: GANTT CHART 2](#_Toc305755544)

[APPENDIX B: MATERIAL SAFETY DATA SHEETS (MSDS) **Error! Bookmark not defined.**](#_Toc305755545)

[A.1 Requirements for MSDS **Error! Bookmark not defined.**](#_Toc305755546)

**LIST OF FIGURES**

Figure 2.1 Sample Time Series Data 12

Figure 3.1 Positioning, alignment, data in a QR code 18

Figure 3.2 QR barcode pointing to a web page 18

Figure 3.3 QR barcode with five lines of text in it 18

Figure 3.4 IPO Diagram of Quarantine Pass Generation 23

Figure 3.5 IPO Diagram of Quarantine Pass Validation 24

Figure 3.6 IPO Diagram for Decision Support System 24

Figure 4.1 Prototype Model 26

Figure 4.2 Sample Quarantine Pass 28

Figure 4.3 Sample of Non- Series Plot 30

Figure 4.4 Sample Stationary Series Plot 30

Figure 4.5 Context Diagram 32

Figure 4.6 Data Flow Diagram 33

Figure 4.7 Use Case Diagram 34

Figure 4.8 Resident System Registration Flow 35

Figure 4.9 Essential Establishment System Registration Flow 36

Figure 4.10 Generate Quarantine Pass Flow 37

Figure 4.11 Scan Quarantine Pass Flow 38

Figure 4.12 Generate Statistical Information Flow 39

**LIST OF TABLES**

Table 4.1 Verbal Interpretation Reference on Weighted Mean 40

**Chapter One**

**INTRODUCTION**

* 1. **Background of the Study**

It has been more than a year since the novel coronavirus disease (COVID-19) outbreak was first reported in Wuhan, China, yet it remains a global issue. Although several vaccines are already available to help prevent people from getting seriously ill or dying from the virus, most countries are still struggling to battle the pandemic as the virus mutations continue to surface.

In the study entitled "Indirect Virus Transmission in Cluster of COVID-19 Cases, Wenzhou, China, 2020" by Cai, Jing et al. (2020), researchers had performed data analysis on COVID-19 cases associated with a shopping mall in Wenzhou. The results have indicated indirect transmission of the Covid-19 virus resulting from contamination of common objects, virus aerosolization in confined spaces, or spread of virus from asymptomatic infected individual.

In another study entitled "Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study" by Badr et al. (2020), results had shown a strong and statistically significant correlation between social distancing, quantified by mobility patterns, and reduction of COVID-19 case growth. With this, the researchers have emphasized the role of social distancing as an effective way to mitigate COVID-19 transmission and should remain part of personal and institutional response to the pandemic.

The World Health Organization have recommended maintaining an inter-personal distance of 1.5 or 2 m (about 6 feet) from each other to minimize the risk of contagion through the droplets that an individual usually disseminate through their nose and mouth.

Forecasting plays an important role during the pandemic and is highly important for effective governmental decision making, for managing supply chain resources, and for informing very difficult political decisions as, for example, imposing a lockdown or curfews.

The Inter-Agency Task Force for the Management of Emerging Infectious Diseases (IATF – EID) acts as the Philippine government's instrument to assess, monitor, contain, control, and prevent the spread and local transmission of COVID-19. They perform risk-level classifications of Provinces, Highly Urbanized Cities (HUCs), and Independent Component Cities (ICCs) upon which they implement localized lockdowns for identified critical areas.

Enhanced Community Quarantine is the most restrictive quarantine classification where strict home quarantine shall be observed in all households. People mobility will be limited to accessing essential goods, services and for work in offices or industries permitted to operate such as public and private hospitals, health, emergency and frontline services, essential goods manufacturers, and the likes. Localities under an Enhanced Community Quarantine (ECQ) are issued with quarantine pass allowing one person per household to buy essential goods or services.

The current implementation of quarantine pass is done manually on which several issues had been observed. An advisory from DILG Region V has been released on March 20, 2020, advising Provincial Directors to instruct Local Officials that the issuance of quarantine pass slips must be distributed on house-to-house basis in their areas of responsibility as releasing of Barangay pass in Barangay Halls defeats the intent of the Enhance Community Quarantine and Social Distancing. On a news article from Inquirer.net released on 23 March 2020, a barangay captain in Lanao del Sur has been arrested for “selling” passes that would allow people to leave their homes during the quarantine period.

By developing a mobile responsive application, these issues and the manual effort spent in generating and validating quarantine passes can be resolved. Augmented with decision support system, the proposed application will enable the government and residents to reduce make informed decisions aiming to reduce the transmission of COVID-19 virus.

* 1. **Statement of the Problem**

The current process of generating, issuing, and validating quarantine passes mostly involves face to face interaction defeating the purpose of the Enhanced Community Quarantine. Not only that it requires a lot of time and effort, but it is also prone to exploitations by those who wanted to get pass the quarantine guidelines.

The manual process does not have proper means to validate and authenticate the quarantine pass, thus, authorities are unable to apprehend any offenders. These loopholes may eventually result to an increase in the number of unauthorized persons outside of residence that forfeits the effectiveness of the quarantine guidelines.

Although one person per household is only allowed to go outside of residence to acquire essential goods, a high probability of having crowded establishment is also an issue of concern.

The policies alone set by IATF during Enhanced Community Quarantine is deemed to be less efficient if not implemented with appropriate tools and risks are not managed.

* 1. **Objective of the Study**

The main objective of this capstone project is to develop a mobile responsive web application that would automate the process of generating and validating quarantine pass and display statistical information for decision support system.

Specifically, this capstone project seeks to achieve the following objectives:

1. To provide digital quarantine passes for residents that will be validated at essential establishments by using Quick Response Code scheme.
2. To provide information to residents about the current crowd count of essential establishments.
3. To provide users with crowd forecast on essential establishments for upcoming days by using Autoregressive Integrated Moving Average (ARIMA) model.
   1. **Scope and Limitations**

To set boundaries on this capstone project, the proponent would focus on developing a mobile responsive web application which would automate the manual process of generation and validation of quarantine pass and embed a crowd forecasting functionality available to its users.

This capstone project will use dummy data for information of residents, households per barangay, barangay officials and essential establishments to give emphasize on how the automated system can provide usable crowd forecasts during the implementation of Enhanced Community Quarantine for better decision making and even risk management.

To limit the scope, the system to be developed will not cover other types of quarantine passes such as Special Quarantine Pass, Self-Employed Pass, 1 Week Transit Pass, 1 Day Transit Pass and the likes but only on Quarantine Pass intended for access to essential establishments.

* 1. **Significance of the Study**

Results obtained from this capstone project will benefit the following stakeholders:

**Local Government Officials.** This tool will remove the manual process of generating, issuing, and validating of quarantine passes during the implementation of Enhance Community Quarantine, enabling them to focus on other important activities. The application can act as an additional support system in ensuring that the guidelines of Enhanced Community Quarantine are implemented accordingly by providing crowd forecasts for decision support system.

**Residents.**  By having a mobile accessible application, individuals can easily obtain a quarantine pass without requiring face to face interaction. Using the application within their reach, they are provided with insightful data which allows them to decide on when is the best time to procure essentials goods minimizing their risk of exposure to congested areas.

**Future Proponents.**  This capstone project could serve as a good reference material for students who are to conduct study of the same nature.

* 1. **Definition of Terms**

**COVID-19**

Refers to the Coronavirus Disease 2019 which is caused by the virus known as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

**Community Quarantine**

Refers to the restriction of movement within, into, or out of the area of quarantine of individuals, large groups of people, or communities, designed to reduce the likelihood of transmission of an infectious disease among persons in and to persons outside the affected area.

**Enhanced Community Quarantine (ECQ)**

The enhanced community quarantine in Luzon is a series of stay-at-home orders and cordon sanitaire measures implemented by the Inter-Agency Task Force for the Management of Emerging Infectious Diseases (IATF-EID) on the island of Luzon and its associated islands. It is part of the COVID-19 community quarantines in the Philippines, a larger scale of COVID-19 containment measures with varying degrees of strictness. The "enhanced community quarantine" (ECQ) is the strictest of these measures and is effectively a total lockdown.

**Quarantine Pass**

A quarantine pass designates a household member as the representative of his or her household in procuring necessities. This disallows other family members to buy goods and conduct transactions outside their homes.

**Essential goods and services**

Covers health and social services to secure the safety and well-being of persons, such as but not limited to, food, water, medicine, medical devices, public utilities, energy, and others as may be determined by the IATF.

**Forecasting**

Forecasting is about predicting the future as accurately as possible, given all the information available, including historical data and knowledge of any future events that might impact the forecasts.

Forecasting should be an integral part of the decision-making activities of management, as it can play an important role in many areas of a company. Modern organizations require short-term, medium-term, and long-term forecasts, depending on the specific application.

**Time Series**

A time series is a sequence of data points that occur in successive order over some period of time. This can be contrasted with cross-sectional data, which captures a point-in-time..

**Time Series Forecasting**

Time series forecasting uses information regarding historical values and associated patterns to predict future activity. Most often, this relates to trend analysis, cyclical fluctuation analysis, and issues of seasonality. As with all forecasting methods, success is not guaranteed.

**Mean**

The arithmetic mean, more commonly known as “the average,” is the sum of a list of numbers divided by the number of items on the list. The mean is useful in determining the overall trend of a data set or providing a rapid snapshot of your data. Another advantage of the mean is that it’s very easy and quick to calculate.

**Regression**

Regression models the relationships between dependent and explanatory variables, which are usually charted on a scatterplot. The regression line also designates whether those relationships are strong or weak. Regression is commonly taught in high school or college statistics courses with applications for science or business in determining trends over time.

**Chapter Two**

**REVIEW OF RELATED LITERATURE**

This chapter presents the different research and other literatures form both foreign and local researchers, which have significant bearings on the variables included in the research. It focuses on several aspects that will help in the development of this study. The literatures of this study come from books, journals, articles, electronic materials such as PDF or E-Book, and other existing thesis and dissertations, foreign and local which are believed to be useful in the advancement of awareness concerning the study.

* 1. **Related Literature**

**Quick Response Code**

QR codes were first used in 1994 by Denso Wave, a Toyota subsidiary company in Japan. QR codes provided a quick and convenient approach to track vehicles during manufacturing process at automotive industry. After its successful implementation at Denso Wave, other industries attempted to adopt this technology too. Denso Wave has patented the QR code, but it’s open for worldwide use. In 2011 the QR code became commercial for the very first time through telecommunications industry (Walker, 2011).

In an article titled “Two-Level QR Code for Private Message Sharing and Document Authentication” (Tkachenko et. al, 2016), it was mentioned that the popularity of these codes is mainly due to the following features: they are robust to the copying process, easy to read by any device and any user, they have a high encoding capacity enhanced by error correction facilities, they have a small size and are robust to geometrical distortions. However, those undeniable advantages also have their counterparts:

1) Information encoded in a QR code is always accessible to everyone, even if it is ciphered and therefore is only legible to authorized users (the difference between “see” and “understand”).

2) It is impossible to distinguish an originally printed QR code from its copy due to their insensitivity to the Print-and-Scan (P&S) process.

QR code is a two-dimensional encoding of information and it is also called matrix code. This matrix code is machine-readable that consists of black and white squares. It can store information in the form of URL (Uniform Resource Locator), contact information, link to videos or photos, plain text and other types of content (International standard ISO/IEC 18004, 2000).

In a paper entitled “Exploring concept of QR Code and the benefits of using QR Code for companies” (Qianyu, 2014), discusses the QR code architecture. Each QR code symbol looks like a square pattern. This square pattern consists of two regions: encoding region and function patterns. The function patterns concentrate on the positioning where the encoding region represents the data encoding. The function pattern comprises finder patterns, timing patterns and alignment patterns. Three common structures on the three corners of QR code symbol are called finder patterns. Finder pattern is used for deciding the correct orientation of the symbol. Timing patterns are used by the decoder software to find the side of pattern. Alignment patterns are used in the case of image distortion to correctly decode the symbol by decoder software. The rest of the region i.e. other than function pattern is the encoded region where data code words and error correcting code words are stored.

Furthermore, the format information areas contain error correction level and mask pattern. The code version and error correction bits are stored in the version information areas. The QR code generation algorithm consists of information encoding using Reed-Solomon error correction code, information division on codewords, application of mask pattern, placement of codewords and function patterns into the QR code. The QR code recognition algorithm includes the scanning process, image binarization, geometrical correction and decoding algorithm (Tkachenko et. al, 2016)

Below are the key Characteristics of QR Code (Qianyu, 2014):

1) High Storage Capacity

A QR code symbol can store up to 7,089 characters of information, which is a huge amount as compared to 1-D barcode.

2) Encodable Character Set

* Numeric data (Digits 0-9)
* Alphanumeric data (upper case letters A-Z; Digits 0 - 9; nine other characters: space, : % \* + - / \_ $)
* Kanji characters

3) Small Printout Size

The information in QR code is stored in both horizontal and vertical directions. Due to this feature, for the same amount of data, space acquired by QR code is one fourth times less than the space acquired by 1-D barcode.

4) 360 Degree Reading

QR code is readable from any direction. This feature is provided by the finder patterns present at three corners of the symbol. The finder pattern helps to locate the QR code.

5) Capability of Restoring and Error Correction

If the part of code symbol is damaged or dirty, data can be recovered. The error detecting procedure can focus on the region of correct information. There are four levels of error correction of QR code that are L, M, Q and H. The level L has the weakest and level H has the strongest error correction capability.

**Time Series Forecasting**

Forecasting future values of an observed time series plays an important role in nearly all fields of science and engineering, such as economics, finance, business intelligence, meteorology, and telecommunication (Palit and Popovic, 2005).

A forecasting method is a procedure for computing forecasts from present and past values. As such it may simply be an algorithmic rule and need not depend on an underlying probability model. Alternatively, it may arise from identifying a particular model for the given data and finding optimal forecasts conditional on that model. Thus, the two terms ‘method’ and ‘model’ should be kept clearly distinct. It is unfortunate that the term ‘forecasting model’ is used rather loosely in the literature and is sometimes wrongly used to describe a forecasting method (Chatfield, 2000).

Chatfield on his book “Time-series Forecasting” (2000) broadly classified forecasting methods into three types:

(a) Judgmental forecasts based on subjective judgement, intuition, ‘inside’ commercial knowledge, and any other relevant information.

(b) Univariate methods where forecasts depend only on present and past values of the single series being forecasted, possibly augmented by a function of time such as a linear trend.

(c) Multivariate methods where forecasts of a given variable depend, at least partly, on values of one or more additional time series variables, called predictor or explanatory variables.

In an article titled “Financial time series forecasting with machine learning techniques: A survey” (Krollner et.al, 2010), authors discussed how different intervals were used in various literature. Figure 2.x below gives an overview of the different forecasting intervals used in the literature. The prediction periods are categorised into one day, one week, and one month ahead predictions. Publications using multiple or different time-frame are listed under ’Multiple / Others’. Most papers make one day ahead predictions e.g. predicting the next day’s closing price. However, being able to predict the stock index one day ahead does not necessarily mean that an investor can take advantage of this information in terms of trading profit, especially since the index itself cannot be traded.

Text

Description automatically generated with medium confidence

**Figure 2.1 Sample Time Series Data**

Let *Y* = {*y*1, . . . , *yn*} denote a time series. Forecasting denotes the process of estimating the future values of *Y , yn+h*, where h denotes the forecasting horizon. Quantitative approaches to time series forecasting are split into two categories: univariate and multivariate. Univariate methods refer to approaches that model future observations of a time series according to its past observations. Multivariate approaches extend univariate ones by considering additional time series that are used as explanatory variables. We will focus on univariate approaches in this work. The forecasting horizon is another aspect to consider when addressing time series prediction problems. Forecasting methods usually focus on one step ahead forecasting, i.e., the prediction of the next value of a time series (*yn*+1). Sometimes one is interested in predicting many steps into the future. These tasks are often referred to as multi-step forecasting (Taieb et al., 2012).

In a paper titled “Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters” (Cerqueira et.al, 2019), the authors discussed common time series models:

1. Naive method, also known as the random walk forecast, predicts the future values of the time series according to the last known observation:

*yˆn+h = yn*

1. Seasonal naive model works similarly to the naive method. The difference is that the seasonal naive approach uses the previously known value from the same season of the intended forecast:

*yˆn+h = yn+h−m*

where *m* denotes the seasonal period.

1. ARMA (Auto-Regressive Moving Average) is one of the most commonly used methods to model univariate time series. ARMA(p,q) combines two components: AR(p), and MA(q). According to the AR(p) model, the value of a given time series, yn, can be estimated using a linear combination of the p past observations, together with an error term  and a constant term *c*.

A picture containing text, clock, watch

Description automatically generated

1. Exponential smoothing model is similar to the AR(p) model in the sense that it models the future values of time series using a linear combination of its past observations. Exponential smoothing methods produce weighted averages of the past values, where the weight decays exponentially as the observations are older.

*yn*+1 = *ynβ*0 + *yn*−1*β*1 + *yn*−2*β*2 + · · ·

There are three main strategies for Multi-step Time Series Forecasting: Recursive, Direct and DirRec. The Recursive strategy (Weigend and Gershenfeld, 1994) trains first a one-step model *f*

*yt*+1 = *f*(*yt*,...,*yt*−*n*+1) + *wt*+1,

with *t* ∈ {*n*, . . . , *N* − 1} and then uses it recursively for returning a multistep prediction. A well-known drawback of the recursive method is its sensitivity to the estimation error, since estimated values, instead of actual ones, are more and more used when we get further in the future.

The Direct strategy (Cheng et.al, 2006) learns independently *H* models *fh*

*yt*+*h* = *fh*(*yt*,...,*yt*−*n*+1) + *wt*+*h*,

with *t* ∈ {*n*, . . . , *N* − *H*} and *h* ∈ {1,...,*H*} and returns a multi-step forecast by concatenating the H predictions. Since the Direct strategy does not use any approximated values to compute the forecasts, it is not prone to any accumulation of errors. Notwithstanding, it has some weaknesses. First, since the *H* models are learned independently no statistical dependencies between the predictions ˆ*yN+h* is considered. Second direct methods often require higher functional complexity than iterated ones to model the stochastic dependency between two series values at two distant instants. Finally, this strategy demands a large computational time since the number of models to learn is equal to the size of the horizon.

The DirRec strategy (Sorjamaa and Lendasse, 2006) combines the architectures and the principles underlying the Direct and the Recursive strategies. DirRec computes the forecasts with different models for every horizon (like the Direct strategy) and, at each time step, it enlarges the set of inputs by adding variables corresponding to the forecasts of the previous step (like the Recursive strategy). However, note that unlike the two previous strategies, the embedding size *n* is not the same for all the horizons. In other terms, the DirRec strategy learns *H* models *fh* from the time series [*y*1,...,*yN* ] where

*yt*+*h* = *fh*(*yt*+*h*−1,...,*yt*−*n*+1) + *wt*+*h*,

with *t* ∈ {*n*, . . . , *N* − *H*} and *h* ∈ {1,...,*H*}.

**ARIMA**

Auto Regressive Integrated Moving Average (ARIMA) processes are a class of stochastic processes used to analyze time series (Box and Jenkins, 1994). The general scheme is as follows:

Step 0) A class of models is formulated assuming certain hypotheses.

Step 1) A model is identified for the observed data.

Step 2) The model parameters are estimated.

Step 3) If the hypotheses of the model are validated, go to Step 4, otherwise go to Step 1 to refine the model.

Step 4) The model is ready for forecasting.

A simple form of an AR model of order p, i.e., *AR(p)*, can be written as a linear process given by:

A picture containing watch, clock

Description automatically generated

Where *xt* is the stationary variable, *c* is constant, the terms in *φi* are autocorrelation coefficients at lags 1, 2,,*p* and *t* , the residuals, are the Gaussian white noise series with mean zero and variance *σ*2. An MA model of order q, i.e., *MA(q)*, can be written in the form:

A picture containing clock, watch

Description automatically generated

Where *μ* is the expectation of *xt* (usually assumed equal to zero), the *θi* terms are the weights applied to the current and prior values of a stochastic term in the time series, and *θ*0 = 1. We assume that *t* is a Gaussian white noise series with mean zero and variance *σ2* . We can combine these two models by adding them together and form an ARIMA model of order (*p, q*):

A picture containing text

Description automatically generated

Where *φi* = 0, *θi* = 0, and *σ*2 > 0. The parameters p and q are called the AR and MA orders, respectively. ARIMA forecasting, also known as Box and Jenkins forecasting, is capable of dealing with non-stationary time series data because of its “integrate” step. In fact, the “integrate” component involves differencing the time series to convert a non-stationary time series into a stationary. The general form of a ARIMA model is denoted as ARIMA(*p, d, q*) (Siami-Namini et.al, 2018).

**Chapter Three**

**THEORETICAL FRAMEWORK**

**Quarantine Pass Manual Process**

Upon issuance of IATF resolution declaring high-risk Provinces, Highly Urbanized Cities (HUCs), and Independent Component Cities (ICCs) under localized Enhanced Community Quarantine, the Local Government Units (LGUs) are advised to implement lockdown procedures limiting people mobility for essential activities only such as buying of food, groceries, medicines etc.

To ensure the public’s access to essential goods quarantine passes are issued by respective LGUs to residents within their jurisdiction. Only one quarantine pass will be given per household.

The generation and issuance of quarantine pass is purely manual process done by LGU officials. Quarantine pass templates are printed out on a piece of paper and distributed either in the Barangay halls or delivered on house-to-house basis in their areas of responsibility.

At checkpoints/essential establishment, the quarantine pass will need to be presented to the checkpoint/barangay officials for manual verification.

**Quick Response Code**

QR Code is a two-dimensional symbol. It was invented in 1994 by Denso, one of major Toyota group companies, and approved as an ISO international standard (ISO/IEC18004) in June 2000. This two-dimensional symbol was initially intended for use in production control of automotive parts, but it has become widespread in other fields.

They encode either an e-mail address or a specific Web site URL, thus enabling smartphone users to directly access Web sites that are encoded by QR codes without typing, copying, or memorizing the Web site address, using their own device.

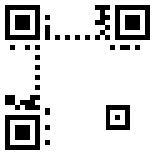
Next to being used for linking to share links, QR codes can contain other types of information:

* A QR code on a business card can contain an electronic version of the **contact information**. Scan the code and the reader application adds the contact to your address list.
* A QR code can contain **event information**. Scan the code on a poster for a concert and the app automatically adds its name, date, and location to the agenda on your smartphone or PC.
* A QR code can contain an **SMS** with a phone number and text. Scan the code and the scanning app lets you automatically participate in some contest to win fabulous prices.
* A QR code can contain an **e-mail message** with a subject and message text. That message can be a request for information so that in return you might get a reply email with additional information and attached files.
* A QR code can contain a **geographical location**. Scan the code on a poster advertising for a restaurant and its location becomes available to your navigation software, informing you how to get to that place.
* A QR code can contain **WIFI configuration data**. Scan the code and your Android device automatically configures itself to use the wireless access at the hotel.

## **Description of Quick Response bar codes**

The Japanese corporation Denso-Wave created the QR matrix code in 1994. It is an open standard for which no license fee has to be paid. The physical encoding of QR codes is nowadays in the hands of various standards bodies, including JIS and ISO (e.g. the ISO/IEC 18004:2006 standard). The standard for encoding URLs was established by NTT DoCoMo, the Japanese telecom company.

QR codes contain information in both the horizontal and vertical axis. Compared to ‘regular’ barcodes, this allows for much larger amounts of raw data to be embedded. These can be numeric, alphanumeric or binary data – of which up to 2953 bytes can be stored. Only a part of each QR bar code contains actual data, including error correction information. Below you see the above QR code with the URL data stripped away. As you can see quite a large area of the bar code is used for defining the data format and version as well as for positioning, alignment and timing purposes.

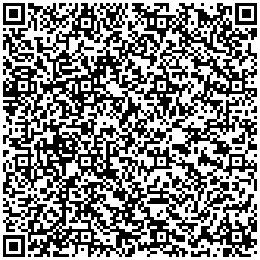
**[](https://www.prepressure.com/images/dict_qrcode_base.png)**

**Figure 3.1 Positioning, alignment, data in a QR code**

The more data need to be embedded, the larger the barcode becomes. Below is the QR code for this page. Since the URL is longer than that of the home page, the bar code has also grown. The barcode after it doesn’t contain a URL but the first 5 sentences of this page.

**[](https://www.prepressure.com/images/dict_qrcode_prepressure_qrpage.png)**

**Figure 3.2 QR barcode pointing to a web page**

**[](https://www.prepressure.com/images/dict_qrcode_prepressure_text.png)**

**Figure 3.3 QR barcode with five lines of text in it**

The smallest square dot or pixel element of a QR code is called a **module**. Like with other types of bar codes, it is recommended to have an empty area around the graphic, which makes it easier for devices to read the bar code. This **quiet area** is ideally 4 modules wide.

The minimum dimensions of a QR code depend upon the resolving power of the cameras that are used to scan the code. According to a Kaywa white paper, it is recommended to use a **minimum size** of 32 × 32 mm or 1.25 × 1.25 inches, excluding a quiet zone, for QR codes that contain a URL. This guarantees that all camera phones on the market can properly read the bar code. Changing the size to a width and height of 26 × 26 mm or roughly 1 square inch still covers 90% of the phones on the market. The latest camera models, which have improved macro capabilities, can however already deal with QR codes that are less than 10 mm (0.4″) wide and high.

The above rule applies to perfectly printed codes that the user has direct access to. Things change when using QR codes on a poster or billboard. The general consensus is that there is a direct relationship between the physical dimensions of a QR code and its scanning distance. That ratio is around 1/10, so if the reader is 50 centimeters removed from the code, the size of the QR code should be at least 5 centimeters. For a billboard viewable from 10 meters, the height of the code should be at least 1 meter.

For good reader accuracy, good contrast between the background and the bar color itself is very important. The bar code should have a dark color on a light background. You cannot go wrong by treating the QR code as line art, using black on white. If the background needs to be in color, make sure that it is a solid color, not a screened tint. Avoid using cyan or magenta but a 100% yellow background should work fine. Very light Pantone colors might also work, if the contrast with the bar code is high enough.

**Autoregressive Integrated Moving Average (ARIMA) Model**

Autoregressive Integrated Moving Average (ARIMA) model uses time-series data and statistical analysis to interpret the data and make future predictions. The ARIMA model aims to explain data by using time series data on its past values and uses linear regression to make predictions.

An ARIMA model can be understood by outlining each of its components as follows:

* Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
* Integrated (I): represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
* Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

**ARIMA Parameters**

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

* **p**: the number of lag observations in the model; also known as the lag order.
* **d**: the number of times that the raw observations are differenced; also known as the degree of differencing.
* **q**: the size of the moving average window; also known as the order of the moving average.

The parameters take the value of integers and must be defined for the model to work. They can also take a value of 0, implying that they will not be used in the model. In such a way, the ARIMA model can be turned into:

* ARMA model (no stationary data, d = 0)
* AR model (no moving averages or stationary data, just an autoregression on past values, d = 0, q = 0)
* MA model (a moving average model with no autoregression or stationary data, p = 0, d = 0)

Therefore, ARIMA models may be defined as:

* ARIMA(1, 0, 0) – known as the first-order autoregressive model
* ARIMA(0, 1, 0) – known as the random walk model
* ARIMA(1, 1, 0) – known as the differenced first-order autoregressive model, and so on.

Once the parameters (p, d, q) have been defined, the ARIMA model aims to estimate the coefficients α and θ, which is the result of using previous data points to forecast values.

**Limitations of the ARIMA Model**

Although ARIMA models can be highly accurate and reliable under the appropriate conditions and data availability, one of the key limitations of the model is that the parameters (p, d, q) need to be manually defined; therefore, finding the most accurate fit can be a long trial-and-error process.

Similarly, the model depends highly on the reliability of historical data and the differencing of the data. It is important to ensure that data was collected accurately and over a long period of time so that the model provides accurate results and forecasts.

**Box- Jenkins Methodology**

The Box-Jenkins approach to modelling ARIMA processes was described in a highly influential book by statisticians George Box and Gwilym Jenkins in 1970. An ARIMA process is a mathematical model used for forecasting. Box-Jenkins modelling involves identifying an appropriate ARIMA process, fitting it to the data, and then using the fitted model for forecasting. One of the attractive features of the Box-Jenkins approach to forecasting is that ARIMA processes are a very rich class of possible models and it is usually possible to find a process which provides an adequate description to the data.

The original Box-Jenkins modelling procedure involved an iterative three-stage process of model selection, parameter estimation and model checking. Recent explanations of the process (e.g., Makridakis, Wheelwright and Hyndman, 1998) often add a preliminary stage of data preparation and a final stage of model application (or forecasting).

1. Data preparation involves transformations and differencing. Transformations of the data (such as square roots or logarithms) can help stabilize the variance in a series where the variation changes with the level. This often happens with business and economic data. Then the data are differenced until there are no obvious patterns such as trend or seasonality left in the data. “Differencing” means taking the difference between consecutive observations, or between observations a year apart. The differenced data are often easier to model than the original data.
2. Model selection in the Box-Jenkins framework uses various graphs based on the transformed and differenced data to try to identify potential ARIMA processes which might provide a good fit to the data. Later developments have led to other model selection tools such as Akaike’s Information Criterion.
3. Parameter estimation means finding the values of the model coefficients which provide the best fit to the data. There are sophisticated computational algorithms designed to do this.
4. Model checking involves testing the assumptions of the model to identify any areas where the model is inadequate. If the model is found to be inadequate, it is necessary to go back to Step 2 and try to identify a better model.
5. Forecasting is what the whole procedure is designed to accomplish. Once the model has been selected, estimated and checked, it is usually a straightforward task to compute forecasts. Of course, this is done by computer.

Although originally designed for modelling time series with ARIMA processes, the underlying strategy of Box and Jenkins is applicable to a wide variety of statistical modelling situations. It provides a convenient framework which allows an analyst to think about the data, and to find an appropriate statistical model which can be used to help answer relevant questions about the data.

**Conceptual Framework**

This section aims to demonstrate the overview of the final product of this capstone project. An I-P-O (Input-Process-Output) model will be used as the conceptual schema of the system It identifies relevant variables, inputs, mappings, and other components and how they will interact with each other. This includes all the underlying concepts and their associated mappings based on the system’s use.

Diagram

Description automatically generated

**Figure 3.4 IPO Diagram of Quarantine Pass Generation**

Diagram

Description automatically generated

**Figure 3.5 IPO Diagram of Quarantine Pass Validation**

Diagram

Description automatically generated

**Figure 3.6 IPO Diagram for Decision Support System**

The users of the proposed application are the residents who need to acquire a quarantine pass for access and procurement of essential goods or services identified by the IATF guidelines, local government officials implementing the quarantine policy, and essential establishments personnel allowed to operate during the quarantine period. They will have access to the same mobile responsive web application but will have role segregation to define their level of access.

The diagram on figure 3.4 and figure 3.5 illustrates the first component of the system where the system provides an automated process of generating and validating quarantine passes. Once a resident has been granted with quarantine pass, they will be able to download the generated QR code and present it to checkpoints/essential establishments whenever they go outside of residence. On the other end, when an essential establishment personnel scans the QR code, the system will display the information of the authorized quarantine pass holder along with the status of the quarantine pass. This action will then record the timestamp of scan in/scan out of the quarantine pass holder.

Diagram on figure 3.6 demonstrates how the data captured from quarantine pass holder’s mobility history can be utilized in generating insightful data regarding the current crowd count on a specific essential establishment and generate a one week forecast.

**Chapter Four**

**METHODOLOGY**

This proponent of this capstone project used prototype method in delivering the objectives of this project.

Diagram

Description automatically generated

**Figure 4.1 Prototype Model**

Figure 4.1 shows the Prototype Model used by the proponent as a guide in developing the project entitled “SafePass: An Implementation of Autoregressive Integrated Moving Average (ARIMA) for Crowd Forecasting Applied in Quarantine Pass” which is a Systems Development Methodology (SDM) within which a paradigm output (or an early approximation of a final system or product) is constructed, tested, and then reworked.

Using this model, it will enhance the usability, design quality of the proposed application and it will also make the development process more cost-efficient since the development cycle becomes shorter.

The phases of the prototype model involve the following steps:

* 1. **Requirements Modeling**

This first step involves understanding the basics product requirements. The proponent of this capstone project possess firsthand experience on how the Quarantine Pass is obtained and utilized when going outside of residence. Apart from the proponent’s personal knowledge, it is also in this stage that the proponent had requested permission from local government authorities and other parties to conduct the study and where all relevant data and information were examined.

To build a logical model of the application, preliminary investigation was conducted via interviews with the involved parties. With these interviews, the researcher was able to identify the transactions involved and analyzed them against the proposed solution. This collected information had enabled the researcher to identify critical decisions geared toward implementing the proposed application.

In Molino IV, a barangay in the city of Bacoor, Cavite, the process of implementing home quarantine passes is done manually by designated LGU or barangay officials. Quarantine Pass templates are printed on a paper with the details of their barangay and some blank fields to be manually populated by the individual who will utilize it. The dissemination process of quarantine pass is for the barangay officials to deliver on house-to-house basis. In case the issued quarantine pass was lost, the resident should request for a replacement from their respective Barangay Hall.

In the 2015 census done by Philippine Statistics Authority, the estimated population in Barangay Molino IV, Bacoor Cavite is 51, 362 and the average number of people per household is 4. Given this data, there would be around 12,841 quarantine passes to be issued to each household on a manual basis requiring a lot of time and effort.

At each designated checkpoints or basic commodity establishments, residents are then required to present their quarantine pass allowing them access to the establishment. The probability of having crowd congestion in a specific establishment is high since the residents do not have visibility on the actual number of quarantine pass holders that are inside an essential establishment on a specific period. Additionally, instances of long queue in grocery stores are mostly observed as only 50% capacity is allowed inside the establishments.

It is for these reasons that the researcher opted to design and develop an application that would not only automate the manual and time-consuming processes but also to expand in utilizing the gathered data on quarantine pass holder’s mobility activity to display useful information and statistics for decision making and risk management.

Graphical user interface, application

Description automatically generated

**Figure 4.2 Sample Quarantine Pass**

In compliance with the Data Privacy Act of 2021, the data to be used in the study such as the Resident Personal Information, Household Demographics, and Essential Establishment details will be dummy data.

Below are the functional requirements grouped by specific role:

* Resident
* Register and Login – Register to gain access to the system.
* Generate Quarantine Pass – Generate a digital quarantine pass by providing personal information to be validated by the system.
* View Statistical Information – Access to the dashboard on actual crowd count and forecast per registered essential establishment.
* Essential Establishment Personnel
* Register and Login – Register to gain access to the system.
* Scan Quarantine Pass – Scan quarantine passes for individuals entering and leaving the essential establishment premises.
* View Statistical Information - Access to the dashboard on the establishment’s actual crowd count and forecast.

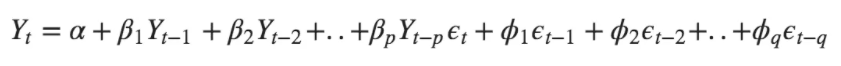
In generating the digital quarantine pass, QR code scheme will be implemented. This would allow residents to download the QR code that will be scanned at essential establishment premises upon entry and exit.

Date time stamps of the entry and exit activity will be logged in the system and will be analyzed to generate crowd demographics. ARIMA model will be used to analyze the time series data on SafePass.

**Forecasting using Auto Regressive Integrated Moving Average (ARIMA) Model**

A time series is a sequence where a metric is recorded over regular time intervals. Depending on the frequency, a time series can be of yearly (ex: annual budget), quarterly (ex: expenses), monthly (ex: air traffic), weekly (ex: sales qty), daily (ex: weather), hourly (ex: stock price), minutes (ex: inbound calls in a call canter) and even seconds wise (ex: web traffic).

A popular and widely used statistical method for time series forecasting is the ARIMA model.



**ARIMA model in words:**

Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)

The first step is to observe at the plots of the autocorrelations and partial autocorrelations of the Mobility Time Series data. A series with a trend will have an autocorrelation pattern like the following:

Chart

Description automatically generated

**Figure 4.3 Sample of Non- Series Plot**

If the systems data series contains large autocorrelations, then the system will need to perform differencing where the original series, xt, must be transformed to become stationary around its mean and its variance. The autocorrelation plots might appear as follows:

Chart

Description automatically generated

**Figure 4.4 Sample Stationary Series Plot**

Differencing usually reduces the number of large autocorrelations considerably. If the differenced series still does not appear stationary, we would have to difference it again.

It is often useful to determine the magnitude of a large autocorrelation and partial autocorrelation coefficient. An autocorrelation must be at least 2 / N , in absolute value to be statistically significant. The following list gives some common values of significant autocorrelations for various sample sizes. Note that even though an autocorrelation is statistically significant, it may not be large enough to worry about.

Table

Description automatically generated

If at first instance the data series is stationary, no differencing will be needed and that is, d=0.

The next step is to select values of d and then p and q in the ARIMA(*p,d,q*) model.

The value of *p* is determined from the partial autocorrelations of the appropriately differenced series. If the partial autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of *p*. If the partial autocorrelations do not cut off, you either have a moving average model (*p*=0) or an ARIMA model with positive *p* and *q*.

The value of *q* is found from the autocorrelations of the appropriately differenced series. If the autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of *q*. If the autocorrelations do not cut off, you either have an autoregressive model (i=0) or an ARIMA model with a positive *p* and *q*.

Once a model has been fit, the final step is the diagnostic checking of the model. The checking is carried out by studying the autocorrelation plots of the residuals to see if further structure (large correlation values) can be found. If all the autocorrelations and partial autocorrelations are small, the model is considered adequate, and forecasts are generated.

If some of the autocorrelations are large, the values of p and/or q are adjusted, and the model is re-estimated.

This process of checking the residuals and adjusting the values of p and q continues until the resulting residuals contain no additional structure. Once a suitable model is selected, the program may be used to generate forecasts and associated probability limits.

* 1. **Quick Design**

At this stage the initial prototype is developed, where the very basic requirements are showcased, and user interfaces are provided. This stage would provide a high-level view of the application to the client.

* + 1. **Context Diagram**

Diagram, venn diagram

Description automatically generated

**Figure 4.5 Context Diagram**

The context diagram in Figure 4.5 illustrates the authorized users input to the system and the expected output information. The target users are the residents and essential establishment personnel. The expected output of the proposed application is: Digital Quarantine Pass (QR Code), Quarantine Holder Information and Quarantine Pass Status, Crowd Count per Establishment and Crowd Forecast per Establishment.

* + 1. **Data Flow Diagram**

Diagram

Description automatically generated

**Figure 4.6 Data Flow Diagram**

The researcher used the Data Flow Diagram, which is a dramatic representation of the information flow within a system that shows how information enters the system and leaves the system, what changes the information and where it is stored (Kendall, 2005).

* + 1. **Use Case Diagram**

Diagram

Description automatically generated

**Figure 4.7 Use Case Diagram**

The development of the proposed application does not solely depend on how the current system is manually implemented. It also contains a workflow process that was identified by the proponent as a necessary step to improve the system flow and accomplish the target results. The components of the proposed application “SafePass: An Implementation of Autoregressive Integrated Moving Average (ARIMA) for Crowd Forecasting Applied in Quarantine Pass”, is illustrated in Figure 4.7 with the use of Use Case Diagram. It describes its user, processes, and the relations between the system components that give the overall behavior of the application.

* + 1. **System Flowcharts**

Diagram

Description automatically generated

**Figure 4.8 Resident System Registration Flow**

Figure 4.8 shows the flow of resident registration wherein they need to provide their personal information, residency information and attach a valid ID for verification purposes. Personal information includes First Name, Middle Name, Last Name, Date of Birth and Contact Number. Residency information such as Barangay Name and Street/Subdivision/Purok/Zone are needed to validate if the user belongs to the correct Barangay base on the resident’s database in the system. Lastly, the user must provide an identification card and upload an image for further system validation.

Diagram

Description automatically generated

**Figure 4.9 Essential Establishment System Registration Flow**

The illustration on Figure 4.9 defines the process flow when an essential establishment registers in the application. Information such as Registered Business Name, Registered Business Owner Name, Registered Business Address/Barangay, number of people allowed in the premises shall be provided together with Business Registration Certificate and will be validated against the essential establishment’s database which contains a list of permitted establishments to operate during Enhanced Community Quarantine.

A picture containing text, hanger

Description automatically generated

**Figure 4.10 Generate Quarantine Pass Flow**

The process flow for the generation of quarantine pass is illustrated in Figure 4.10 wherein the system will check on what household the user belongs to and validate if the quarantine pass allotted for their household was already granted. If no other member of the same household has not requested for a quarantine pass, then the system will generate a QR Code and update the database that a quarantine pass was already granted. The user will have the ability to download the QR Code or view it on the system every time they login.

Diagram

Description automatically generated

**Figure 4.11 Scan Quarantine Pass Flow**

Figure 4.11 displays the system flow when a QR Code is scanned in an establishment. Resident with valid QR Code should present the QR code to the essential establishment personnel using SafePass application upon entry/exit. The system will then validate the QR code validity and should display the quarantine pass holder’s Full Name, Barangay, and the uploaded Image. This action will record the entry/exit date and time stamp of the holder and the establishment name in the system database.

Diagram

Description automatically generated

**Figure 4.12 Generate Statistical Information Flow**

In the process flow presented in Figure 4.12, a user will have the ability to view the current crowd count in an essential establishment registered in SafePass application. The crowd count information will provide details if a certain establishment is already reaching their allowed capacity which will help the residents decide whether to proceed or not. By using the historical data in the system, a weeklong crowd forecast can be generated which will aid the residents in planning when is the best time and day of the week should they go to the establishment to avoid crowd exposure. The same statistical information can also be utilized by the establishment in risk management and planning.

* 1. **Building Prototype**

At this stage, system requirements and other components necessary to develop the proposed application will be identified.

* 1. **User Evaluation**

A survey questionnaire has been defined by Calderon & Gonzalez as simply a set of questions, which when answered properly by a required number of properly selected respondents, will supply the necessary information to complete a research study. The capstone project will utilize a checklist type of survey questionnaire in which the respondents will be able to answer faster and easier at their convenience. The survey questionnaire will be based on the criteria evaluation on the system as guided by the ISO 9126-1. The survey questionnaire will comprise the following criteria: Functionality; Efficiency; Reliability; Usability and Portability.

**Table 4.1**

**Verbal Interpretation Reference on Weighted Mean**

|  |  |  |
| --- | --- | --- |
| **Scale** | **Interpretation** | **Descriptive Equivalent** |
| 1 | 1.00-1.49 | Needs Improvement (NI) |
| 2 | 1.50-2.49 | Fair (F) |
| 3 | 2.50-3.49 | Satisfactory(S) |
| 4 | 3.50-4.49 | Very Satisfactory (VS) |
| 5 | 4.50-5.00 | Excellent (E) |

**4.5 Refining Prototype**

In this stage, the end user will evaluate the prototype. Dissatisfaction with the prototype at this level will result to a revision based on the given requirements. The new prototype will be re-evaluated, and the process continued until such time that the requirements identified by the end-user were met. Revisions will be done based on the user’s comments and suggestions during the evaluation of the developed application.

**4.6 Engineer Product**

The last stage of this approach will conclude with the confirmation and approval of the application by the end-users. This will also be referred to as the user acceptance phase. It is also in this phase that the proponent will be able to appraise the overall performance of the final system, using the predetermined indices or indicators such as functionality, efficiency, reliability, usability, and portability.

**LIST OF REFERENCES**

Qianyu, Ji (2014). *Exploring concept of QR Code and the benefits of using QR Code for companies*

International standard ISO/IEC 18004 (2000). *Information technology Automatic identification and data capture techniques Bar code symbology QR Code*

Chatfield, Chris (2000). *Time-series Forecasting*

Palit, Ajoy K.; Popovic, Dobrivoje (2005). *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*

Walker, J.D (2011). *QR Codes Are the Next Marketing Wave*

Tkachenko, Iuliia; Puech, William; Destruel, Christophe; Strauss, Olivier; Gaudin, Jean-Marc; Guichard, Christian (2016). *Two-Level QR Code for Private Message Sharing and Document Authentication*

Krollner, Bjoern; Vanstone, Bruce; Finnie, Gavin; (2010). *Financial time series forecasting with machine learning techniques: A survey*

Taieb, Souhaib Ben; Bontempi, Gianluca; Atiya, Amir F; Sorjamaa, Antti (2012). *A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition.*

Cerqueira, Vitor; Torgo, Luis; Soares, Carlos (2019). *Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters*

Weigend, Andreas .S. and Gershenfeld, Neil A. (1994) *Time Series Prediction: forecasting the future and understanding the past*

Cheng, Haibin; Tan, Pang-Ning; Gao, Jing; Scripps, Jerry (2006). *Multistep-Ahead Time Series Prediction*

Sorjamaa, Antti and Lendasse, Amaury (2006). *Time series prediction using dirrec strategy*

Box, George E.P.; Jenkins, Gwilym; Reinsel, Gregory C. (1994). *Time Series Analysis Forecasting and Control*

Siami-Namini, Sima; Tavakoli, Neda; Siami Namin, Akbar (2018). *A Comparison of ARIMA and LSTM in Forecasting Time Series*

**APPENDIX A:**