

# CHAPTER I

## INTRODUCTION

### 1.1 Research Background

Forecasting the time series data - a series of observations listed in the order of time - is an essential subject in the field of economics, business, and finance. Many people have been looking into forecasting the stock market, supply and demand, and even macroeconomic indicators from historical patterns or experiences [1, 2]. Traditionally, statistical models such as the Autoregressive Integrated Moving Average (ARIMA) have been frequently used to forecast time series data for its advantage of being easier to understand and compute mathematically.

Yet ARIMA is also known to suffer from certain conditions, such as the need for a certain number of data points to produce more accurate predictions [3], which are not always feasible in business data, to the assumed linearity and stationarity of the data [4] and therefore being unable to predict well with volatile time series data.

With the recent development and advancement in computational powers and improvements in advanced machine learning algorithms, deep learning method [5, 6] like the Long Short-Term Memory (LSTM) shows a lot of promise and potential to address the time series forecasting problems aforementioned [7, 8].

From the literature studies, there have been several studies that compared ARIMA with LSTM in time series forecasting. Despite several studies pointing out that deep learning-based methods are more flexible to handle both linear or

non-linear data, very few are analysing the effects of data quality conditions such as volatile and missing values that may be encountered in real life. In terms of performance, no research was found to have studied the impact of data conditions on the run-time performance of ARIMA and LSTM models. Moreover, there were inconsistent results where several research concluded with ARIMA being superior and others finding LSTM superior in the accuracy of forecasting the time series data.

Motivated by the need for a broad-spectrum time series forecasting model that can handle both linear or nonlinear data with reasonable accuracy and run-time performance, this research aims to identify a forecasting model that will satisfy such needs by comparing the ARIMA and LSTM modelling.

## **1.2 Problem Statements**

Considering the selected topic background and literature study, the research problems of this research are as follows:

- a. Is LSTM superior in terms of accuracy and run-time performance compared to ARIMA in short and long data points of volatile time series data?
- b. Is LSTM superior in terms of accuracy and run-time performance compared to ARIMA in the dataset that contains missing values?

## **1.3 Research Objectives**

Considering the selected topic background and research problem raised, the aims and objectives of this research are as follows:

- a. Investigate whether LSTM is superior to ARIMA in the case of short and long data points of volatile time series data.
- b. Investigate whether LSTM is superior to ARIMA in the case of the dataset with missing values.

#### **1.4 Benefits of the Research**

With the result of this thesis, the author hopes to establish a general guidance of when to use which forecasting model given specific data conditions, i.e., having a short or long number of data points in a volatile time series data, dataset containing missing values, and suitable performance metrics to be achieved.

#### **1.5 Thesis Structure**

This thesis is divided into five chapters, each discussing different goals and contents. The systematics are as follows:

Chapter I: Introduction. This chapter specifies the objectives, goals, and primary purposes of this thesis, also the methodology used to draft this thesis.

Chapter II: Literature Review. This section of the paper explains the theoretical foundations that are used and are related to the subject matter of the discussion described in the literature review and framework section to support this research paper to assist in providing solutions to the existing problems.

Chapter III: Modelling and Experiment. This section of the paper describes the research methodology and the research techniques, sampling techniques, data collection techniques, data measurement techniques, data analysis, and the framework used in this research.

Chapter IV: Result and Analysis. This section of the paper provides information regarding the type of research used by the author, the results of this research and discussions of the studied research results.

Chapter V: Conclusion and Future Work. This section of the paper provides the conclusions from the obtained research results, analytic tools, limitations of the current research, and suggestions for future research.

At the end of this thesis, the researchers' bibliography and curriculum vitae are attached.

