“Comparative Analysis of ARIMA and LSTM Models for Predicting Electricity Consumption, Electricity Price and Stock Prices: A Case Study of Victoria, Australia"

MSc Research Project
M.Sc. in Fintech

Chahine Firas
Student ID: X21233543

School of Computing
National College of Ireland

Supervisor: Pr. Brian Byrne

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	FIRAS CHAHINE
Student ID:	X21233543
Programme:	MSc in Fintech **Year:** 2022-2023
Module:	Research Project
Supervisor:	Pr. Brian Byrne
Submission Due Date:	14th August 2023
Project Title:	"Comparative Analysis of ARIMA and LSTM Models for Predicting Electricity Consumption, Electricity Price and Stock Prices: A Case Study of Victoria, Australia"

Word Count:6688 **Page Count 22**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Firas Chahine

Date: 14th August 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

"Comparative Analysis of ARIMA and LSTM Models for Predicting Electricity Consumption and Stock Prices: A Case Study of Victoria, Australia"

FIRAS CHAHINE
X21233543

Abstract

In an era when data-driven decision-making is becoming increasingly important, accurate prediction of complex events is critical. This research focuses on the predictive analytics phenomena by conducting a comparative analysis between the implementation of ARIMA and LSTM model on the S&P/ASX200 index on one hand, and the price and demand on the electricity in Victoria Australia on the other hand. The study focuses on the task of predicting the stock price and the demand on the electricity of Victoria, as well as the future stock market dynamics. The study is to investigate the usefulness of these techniques in forecasting both by using historical datasets encompassing on the stock prices and the power consumption data. The study findings provide useful insights that might possibly improve strategic decision-making in the energy and finance sectors, as well as showing strength and weaknesses of both algorithms methodologies used.

1 Introduction

1.1 Background

1.1.1 Stock Price Prediction

Predictive modelling is becoming increasingly crucial in a variety of of disciplines, allowing for precise predictions and informed decision-making. For a long time, several fields such as economics and computer science have investigated the study of predicting stock prices and stock market indexes. To solve this issue, a plethora of models an algorithm has emerged. However, accurate forecasting of stock indexes is almost impossible due to the vast and ever-changing set of variables that impact them. The significance of stock markets cannot be overlooked by any nation, as they a vital role in fostering economic growth. While stock markets can be turbulent, thy can also be profitable assets for investors if their behaviour becomes predictable. This predictability would benefit not just individual investors but also greatly contribute to the country's prosperity by enabling smarter investment decisions and creating a stable economy.

To put it differently, constructing a precise forecasting model that combines all the elements associated with human existence such as politics, diplomacy, security, economic considerations, and resources, all of which have a sporadic influence on the financial market, is a difficult undertaking (Vijh, 2020, p. 599). Nonetheless, advances in hardware and software in terms of computer technology and the implementation of performance-enhancing techniques like as machine learning have been used in predicting stock values, which results in better predictability (Nikou, 2019, p.172)

1.1.2 Energy Consumption and Price Prediction

Forecasting energy demand and price accurately is critical for effective power system management and planning. Many factors have influenced the growth of and increase of energy demand, such as population expansion, improved living standards, urbanization, technological breakthroughs, and industrialization (Ozturk, S). For government and monopoly electricity providers, predicting energy consumption accurately is a critical step to undertake, as overestimating energy consumption can result in an excess of energy produced, leading to difficulties and costs in storage. It can also lead to a significant waste of investment in surplus power facilities.

Underestimating demand, on the other side of the equation, could also be risky, as it might lead to the unfulfilled demand and insufficient spinning reverse planning, which could result in a riskier operating tactics. This has the potential to create system breakdowns, electricity cuts and power shortages especially in susceptible areas. As a result, accurate forecasting is critical to ensuring optimal resource allocation, cost effectiveness and dependable power energy.
Several incorporated exogenous variables influence the forecasting of energy consumption, such as the weather, calendar variables, economic variables and the inherent unpredictability in individual requirements. To overcome this challenges, multiple advanced prediction methods have been created and developed based on input data and historical data.

Linear Regression model, stochastic process models, auto regressive integrated moving average model, Long Short-Term modelling model and many other models have been developed and employed for similar issues. Artificial neural network-based models, on the other hand, have been developed as alternative forecasting methodologies in recent years, which have the ability to capture the non-linear relationships in power demand and demonstrating accurate and better results.

1.2 Purpose of the Study

Stock price forecasting is extremely difficult, especially when it comes to index prediction. This is partly due to the unpredictability of such data, which can exhibit a stochastic walk in the shortest amount of time. On the other hand, accurate prediction of energy price and demand is crucial for a variety of stakeholders, including energy providers, legislators, consumer and government suppliers. Understanding and anticipating energy consumption trends may lead to more efficient resource allocation, more efficient energy production planning and more effective demand response tactics.
The purpose of this study is to conduct a comprehensive comparison of multiple predictive modelling techniques in the context of stock prediction and electricity price and demand. The major goal is to assess and establish if these models are better at forecasting stock prices or energy usage trends. By doing this comparison research, we aim to establish which modelling techniques yield more accurate projections in each area.
To do this, the following study question was designed.:

1.2.1 Research Question

To what extent do predictive modeling such as ARIMA (autoregressive integrated moving average) and LSTM (Long-Short Term Memory) differ in their ability to predict the ASX 200 Australian stock index and energy consumption and price patterns in Victoria?

1.2.2 Research Objectives

Objective: The objective of this study is to assess and compare the predictive capabilities of the ARIMA and LSTM models in their ability to forecast the ASX 200 Australian stock price index and energy consumption patterns. The project is to investigate the variations between two modeling methodologies approaches in their capacity to predict both stock market behavior and energy usage trends. The research intends to establish which model performs better in terms of prediction accuracy and dependability for both the ASX 200 stock index and energy usage patterns by performing a detailed comparison investigation.

1.2.3 Significance of the Study

The significance of these papers lies in their investigation and comparison of the performance of two newly advanced and sophisticated predictive modelling, highly used as well, which are ARIMA and LSTM, in forecasting the ASX 200 stock index and energy usage trends. Financial markets are inherently volatile and highly affected by numerous economic factors, such as interest rates, economic indicators and others. Understanding how various predictive models perform in forecasting stock market behaviour may be of particular importance for investors seeking to invest in the ASX 200- index, as this may aid them in developing an investment strategy by providing investors with a level of certainty to assist them in making a decision. In addition, forecasting energy price and consumption on the other hand can have a far-reaching consequence for investors, financial analysts, legislators and energy stakeholders.

Finally, this research would provide vital insights to the existing body of literature in this field of study, making it easier for future researchers interested in comparable issues to acquire relevant data.

1.2.4 Structure of this study

The first chapter of this article serves as an introduction, presenting several principles regarding the significance and credentials of the work. This chapter discusses the study's goals, objective and research question. Additionally, it highlights the relevance of the work and its contribution to the existing literature. Moving on to the second chapter, a comprehensive exploration of related works relevant to the topic and machine learning are undertaken. This investigation encompasses theoretical, empirical and conceptual frameworks that lay the groundwork for the study. The work's methodology is demonstrated and discussed in Chapter three. Subsequently, chapter four outlines the empirical approach employed in the study and the design specification of the research. This involves elucidating the data's nature and sources, as well as detailing the statistical and technological techniques chosen for analysis. Chapter five presents the implementation and the outcomes of the analysis. The sixth chapter delves deeper into the research findings, providing an extensive discussion. Jumping to Chapter 7 gives a discussion about the findings and the research objectives. Lastly, the concluding chapter synthesizes the findings by offering suggestions and recommending potential future directions in the field.

2 Related Work

2.1 Theoretical Review

2.1.1 Stock Market Prediction

The current spike in investor interest in the stock market has coincided with the development of research in this field. Notably, the introduction of machine learning has greatly aided the research process. Despite the growing interest, the existing literature on the S&P/ASX200 index is still minimal, leaving plenty of potential for additional research. The lack of current research creates an ideal environment for further investigation and inquiry into this specific subject.

Murthy et al. (2022) addressed in their study named “Predicting Stock Price with LSTM Networks” the difficulty of forecasting stock prices using LSTM. The researchers wanted to see how well LSTM networks could forecast stock values. It used historical data as input and trained the LSTM. The results showed that LSTM was more accurate in predicting the volatility of the S&P 500, meanwhile the Baseline was more accurate in turn of return prediction with a rate of 53%.

Another research conducted by McNally, Roche and Caton (2018) aimed to explore the viability of using machine learning techniques to forecast the price of Bitcoin. The machine learning used in this model consists of ARIMA, LSTM and Bayesian optimised recurrent neural networks. The results of the research showed that the machine learning can capture and leverage the intricate relationships within cryptocurrency data to make price forecasts. Researchers collected data from 3-year duration. The results showed that LSTM model displayed a capability to discern underlying patterns within the financial data of 52%, with an RMSE of 8%, meanwhile ARIMA, on the other hand, showed an accuracy of 50.05% and RMSE of 53.74%.

Khan and Singh (2022) investigated the objective of predicting the stock price of Suzuki using the ARIMA model. Data was gathered from the stock exchange and the results support the efficacy of employing established methodologies for the purpose of stock prediction.

Pandey, Singh, Hadiyuono and Mourya’s research (2023) takes a novel method to stock market analysis by combining classic ARIMA and current LSTM approaches. Researchers investigated how these machine learning might be used to improve the accuracy of stock price predictions. It includes a time span of six years. Researcher concluded that both model, LSTM and ARIMA, arrives at results that have the same level of accuracy. In contrast, researchers found that despites the better level of accuracy that ARIMA offer than LSTM, but ARIMA takes longer to process.

Adebiyi, Adewumi, and Ayo (2014) conducted a comparative examination of two popular predictive models for stock price prediction, ARIMA and ANN. This study involved a meticulous comparison of predicted values against actual values using the Nokia and Zenith Bank stocks. The finding of the study reveals that there is a slightly difference between the actual value of the stock, and the predicted value of the stock. Researchers concluded then that the ARIMA model is particularly well-suited for short-term predictions. However, they added that there are methodologies other than ARIMA that might offer superior accuracy for forecasting stock prices over longer period.

Samuel Olusegun, et al., (2019) analyze two Recurrent Neural Network based models. First one makes use of the LSTM, second one is the Gated Recurrent Unit for the purpose of stock market behaviour prediction. A comparative analysis has been made between these two models based on training on the same market dataset. Researcher concludes that the LSTM model have a greater accuracy in comparison to the Gated Recurrent Unit based model.

The core objective of the research conducted by Emioma (2020) is to employ a machine-learning algorithm to approximate the closing stock price within a given dataset, meaning enhancing the accuracy of stock predictions. This model uses the least-squares linear regression as machine learning. The findings revealed an error margin of approximately 1.4% in the predictions. Researcher concluded that the adjusted closing price is a model that cannot be designed in real world because the historical is deterministic, meanwhile the real time data can vary in the real world.

2.1.2 Electricity price and consumption prediction

Autoregressive Integrated Moving Average models, as well as Long Short-Term Memory can be used to predict energy use and price (Samuel Asuamah Yeboah and Manu, 2012).

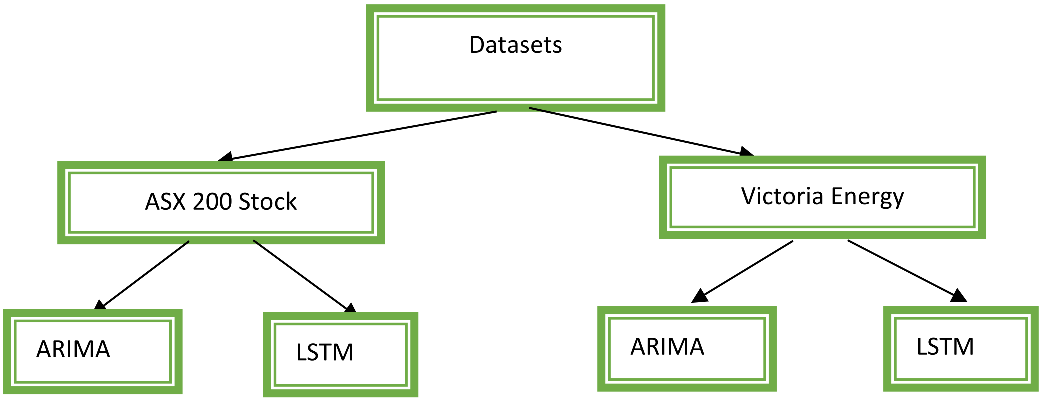
According to Ajith and Baikunth (2001), four step procedures for using ARIMA and LSTM in forecasting has been described. Model identification is involved, which patterns and components are provisionally recognised using techniques like graphs and autocorrelation functions. Following that, parameter estimation entails calculating coefficients using approaches such as least squared or maximum likelihood. Model diagnostics aid in determining the model's validity and encourage iterative refinement. Finally, forecast verification assures model improvement by assessing prediction validity and tracking model performance using various statistical approaches and confidence intervals, efficiently discovering anomalies (Samuel Asuamah Yeboah and Manu, 2012).

Wang and Meng(2012) conducted research in which they projected Hebei province's energy usage between 2009 and 2013. The data initially shows an increasing trend for energy usage and suggest that the usage of energy will continue to rise over the next five years. The research shows that the energy consumption for 2013 expect to be 2885.6 million tons, which means an annual increase of 2.8% from the starting date till 2013 (Wang and Meng, 2012).

Another study conducted by Ahmad and Latif (2011), where the demand for power in Selangor was forecasted. They used a dataset including weekly electrical generation statistics from a local substation from the start of 2009 till February 2011. The findings of this study shows that the predicted power generation increased over the period, and an increase in energy use in the next years is expected, emphasizing the importance of establishing effective measurements.

2.1.3 Conceptual Framework

The major goal of this study is to compare the prediction skills of two different models, ARIMA and LSTM, in projecting stock prices and energy usage. The study's goal is to assess each model's strengths and weaknesses in representing the complicated dynamics of these two areas. As a result, this work will be arranged in such a manner that readers seeking instruction on how to use predictive modelling machine learning in many fields can find it.



The analysis conducted in this study revolves around two main aspects. The first facet involved predicting the future stock prices of the ASX 200 index of Australia through the utilization of the ARIMA (Auto Regressive Integrated Moving Average) and LSTM (Long Short-Term Memory) model. The second facet is to predict the future energy consumption and price of Victoria Australia using the same two models. Next step will be comparing the results of the predictions and conclude whether it is more accurate to predict price and demand consumption of electricity or stock price.

3 Research Methodology and Specification

This study has a dual focus: one is to develop models for predicting the future values of the ASX 200 index and the other is to forecast power consumption and prices in Australia. The CRISP-DM approach was critical in extracting the data from the whole datasets. This technique provides an example for a process architecture for data mining. It creates a connection between mining and understanding the context. This approach provides clear direction for effective data mining, particularly by emphasising critical phases from the start, including a full description of the process.

The CRISP-DM Process Lifecycle

The CRISP-DM architecture is made up of six major stages, which are represented by arrows that show the critical and common interdependencies between them. The order of these stages is not set, and real-world projects frequently switch between them as needed (IBM,2021). This model is distinguished by its adaptability and flexibility, which allows for modification to unique project requirements. In addition, stage outcomes serve as a means to a goal typically resulting in the production for fresh commercial enquiries. Figure below shows the map of the CRISP-DM

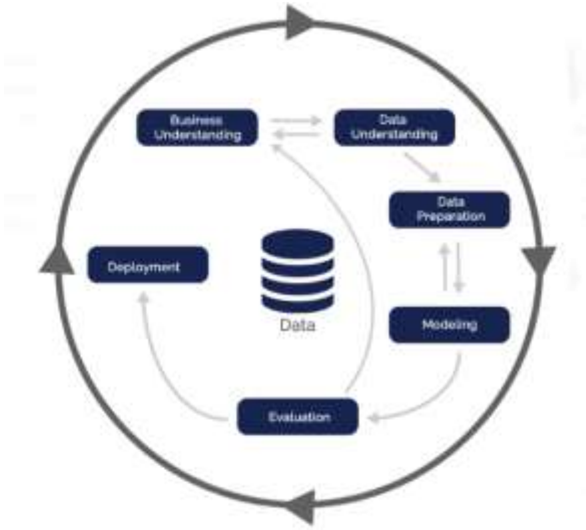


Figure 1: Crisp-DM diagram (Mafra, 2018)

i. Business Understanding:

The business understanding of the CRISP-DM process is dedicated to knowing the project's objective and requirements. The major goal of this research is to forecast the price of the ASX200 index as well as Australia's power consumption and pricing patterns. These goals include negotiating a slew of elements and variables related to stock price such as macro-economic variables and others. Similarly, power prices are affected by a large variety of complicated factors as well. While many factor are at play, a large reservoir of data is accessible for use in the modelling process.

ii. Data Understanding

the following step in the CRISP-DM process is Data Understanding, which focuses on acquiring a thorough understanding of the data at hand. This stage comprises selecting relevant data sources to produce the best possible results for the investigation. Researchers must analyse the quality of the data, select the best instruments for data collection, and develop a technique for acquiring the necessary information (Ingole, 2022). Once the data gathering and organisation phases are completed, analysts are free to develop their own hypotheses based on the data insights. This step is the cornerstone of the data mining process, allowing researchers to lay the groundwork for later analysis and model creation.

iii. Data Preparation

Following Data Understanding, the Data Preparation stage is critical in refining the obtained data for meaningful analysis. Data derived from a variety of raw statistics and resources may frequently lack consistency and contain flaws such as corruption and missing properties. This step is dedicated to fixing these issues, therefore improving data quality and resulting in an effective and cohesive dataset appropriate for modelling. Data cleaning to correct errors, data integration to merge disparate datasets, data transformation to improve structure, data reduction for manageable complexity, and data discretization to categorise continuous variables into discrete ranges are all critical steps in the process (Ingole, 2022). Researchers guarantee that the ensuing modelling phase is based on a solid foundation by carefully carrying out these data preparation operations.

iv. *Modelling*

The first stage in the Modelling phase is the building of machine learning models, which is a significant improvement. These models are intended to provide exact forecasts within specific time limits, removing the need for user intervention. These models produce real-time forecasts using automated procedures, allowing for quick decision-making (Ingole, 2022). The dataset is divided into two sections: training data, which accounts for 70% of the dataset, and testing data, which accounts for the remaining 30%. The model is built using training data, and its accuracy and precision are then tested using testing data. This method allows researchers to evaluate the model's performance in real-world events and determine its usefulness in making correct predictions. The Modelling step is critical in translating data-driven insights into practical recommendations.

1. ARIMA:

ARIMA, or Autoregressive Integrated Moving Average, is a statistical analysis approach that analyses time series data with the objective of improving data interpretation and anticipating future patterns (Hayes, 2019). The major objective is to forecast stock price in the future. To begin this procedure, the data must be transformed into a stable state. To do this, certain strategies must be used, such as taking logarithms in the case of non-linear trends or using the auto ARIMA method.
Box and Jenkins developed a model for forecasting values in 1960sm which is still the most often used approach for generating predictions in time-series analysis (Wang and Niu, 2009). Auto Regressive is the linear relationship of anticipated value at any given moment to its previous observations. The formula is:

$$xt = \delta + \emptyset xt{-}1 + \epsilon_t$$

Similarly, MA, indicating the moving average, means that at any point in time, the future value is the linear relationship of the past error terms.
The formula is written as follows:

$$xt = \mu + \theta \epsilon t{-}1 + \epsilon_t$$

The AR and MA involves p and q parameters, determining through autocorrelation and partial correlation functions based on AIC and BIC (Du, 2018)

2. LSTM:

Long Short-Term Memory (LSTM) networks are a form of Recurrent Neural Network (RNN) that excels at collecting and interpreting long-range relationships, particularly in the context of sequence prediction models. This method evaluates incoming data and chooses whether to keep or discard it depending on its relevance. LSTM is intended to strike a balance between keeping prior information for a set period of time and making educated decisions about what to forget and what to recall. Weighted connections are formed between layers, allowing for bidirectional data flow. This characteristic separates LSTM from other models, providing a significant benefit by contributing to constant weighting throughout the process.LSTM, an RNN type described in Hochreiter and Schmidhuber (1997), uses memory cells to capture long-term relationships in time-series data and is specified by basic equations. The equation of the LSTM is shown in the figure below:

$$i_t = W_i(x_t, h_{t-1}) + B_i$$
$$f_t = W_f(x_t, h_{t-1}) + B_f$$
$$g_t = W_g(x_t, h_{t-1}) + B_g$$
$$o_t = W_o(x_t, h_{t-1}) + B_o$$
$$c_t = sigmoid(f_t) \odot c_{t-1} + sigmoid(i_t) \odot tanh(g_t)$$
$$h_t = sigmoid(o_t) \odot tanh(c_t)$$

Figure (2): LSTM formulas (Hochreiter and Schmidhuber, 1997)

It assesses information, allowing passage based on signal strengths, balancing retention of earlier data and decision-making. This capacity to carry forward meaningful long-term dependencies, through bidirectional weight propagation, is essential for predicting cryptocurrency prices based on historical data (Brownlee, 2018).

v. *Evaluation*

The study's final and most important stage includes determining the accuracy of the forecasts. These articles use two separate models, ARIMA and LSTM, and their respective performances will be evaluated in order to reach conclusions and answer the primary study topic. In parallel, the regression model will be evaluated using the following methodologies:

- T statistic test will be used to evaluate whether to accept or reject a hypothesis, resulting in a well-founded conclusion. In this example, the null hypothesis revolves around non-significance. The null hypothesis (Ho) is rejected if the p-value is less than the selected significance level. If p-value < significance level, Ho should be rejected.
- R squared coefficient: R squared coefficient: which is a percentage that indicates how well the model matched the data
- MAE, RMSE, MSE
    - MAE measures the average absolute differences between predicted and actual values.
    - RMSE calculates the square root of the average of squared prediction errors.
    - MSE computes the average of squared prediction errors.

4 Design Specification

The block diagram depicts the machine learning model's design architecture which illustrates the five critical processes needed in attaining accurate predicting results:

- Data Extraction: The project's dataset was taken from Kaggle.com and yahoo finance. It comprises information pertaining to fundamental characteristics of electricity consumption and stock index that are important in estimating the goal variable. The dataset has two parts: train and test.

- EDA methods: the exploratory data analysis, also known as EDA method, is carried out at this stage, where graphical representations are employed to visualize and comprehend the data. This assists in acquiring insights into the properties and patterns of the data.
- Data Preprocessing: Data preprocessing is an important step in ensuring data quality and eliminating discrepancies. This stage entails dealing with difficulties such as missing values in the dataset.
- Machine learning models implementation: Several machines learning models, including ARIMA, LSTM are used in this stage. Based on the collected attributes and the objective variable, these models are used to anticipate power use.
- Model Evaluation: The last stage is to assess the performance of the machine learning models. To measure how effectively each model forecasts electricity consumption and stock index closing prices, several assessment metrics such as Root Mean Square Value (RMSE), Mean Square Error (MSE) and Mean Absolute Error (MAE) are utilized.

By adhering to this planned structure and carrying out the five critical processes, the project delivers accurate anticipated outcomes, allowing for a better knowledge and analysis of stock market and power usage trends. The diagram below illustrates better the overflow of the project:

Figure 1: Overflow of the project

5 Implementation

In this step, we offered an outline of the numerous processes involved in building the models, with a particular emphasis on two models: ARIMA and LSTM. Our major goal in developing these models is to anticipate electricity consumption in Victoria Australia, as well as the national index price of Victoria, and compare whether these models perform better in predicting the stock prices or the electricity consumption.

1) Environmental steps.

The project has been executed using two different programming languages, R language and python, and the coding process was completed successfully on the Google Colab platform. In this context, all of the code required for building the models and executing data analysis has been accomplished.

System	Ram 8G
Processor	Intel I5
Speed	2.70 GHz
Software	Jupyter Notebook
Programming	Python and R
Python Libraries	Python Libraries
R Libraries	R Labraries

The preprocessing stages were rigorously followed in order to prepare the data modelling and ensure its appropriateness for analysis. Both the ARIMA and LSTM models have been implemented, with the help of numerous important libraries. The use of libraries such as pandas, numpy, matplotlib, auto-arima, create_lstm, statsmodels and more has aided in the creation and validation of the models. Various tests have also been performed to examine the regularity of the dataset and the performance of the models, confirming their efficacy in providing correct predictions.

2) Selection of data

The datasets for this study was gathered from Kaggle.com for the electricity consumption dataset, as well as yahoo finance for the index and includes critical records such as :

- Electricity Consumption: date, demand for electricity, RRP, demand positive RRP, RRP positive, demand neg RRP, RRP negative, minimum temperature, maximum temperature, solar exposure, rainfall, school day, holiday.
  For instance, the dataset encompasses records of the stock index from 1st January 2015 to 10th June 2020. Both the training and testing datasets together comprise 2107 rows and 14 columns, all available in CSV format.

- Stock index: Date, Open Price, Highest Price, Lower Price, Closing price.
  For instance, the dataset encompasses records of the stock index from 1st January 2015 to 31st December 2019. Both the training and testing datasets together comprise 1265 rows and 5 columns, all available in CSV format.

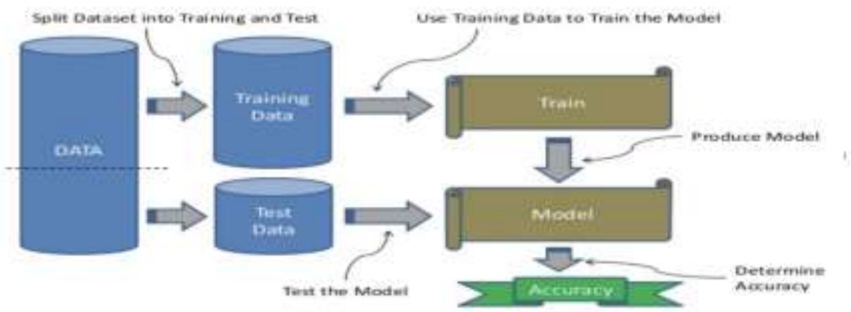


Figure (2)

Figure 2 above depicts the full operational approach that was used throughout the project. It starts with the importing of the dataset, then goes through extensive pre-processing to assure the data's quality and appropriateness for analysis. Using the pre-processed data, the required model ARIMA and LSTM is then implemented. Following that, the model's performance is tested and compared to other models using a variety of tests. Finally, depending on the evaluation findings, the best model is selected, and its representation is given, bringing the research to a close.

6 Results and Explanation

After all of the implementation was accomplished, we moved on to the assessment phase to evaluate all of the models used in our research and study. To achieve a complete review, we used some parameters such as MSE,MAE,RMSE, Rsquared

- RSME
  It is a popular way of assessing the performance of the models. The average size of the variations between anticipated and actual values in a dataset is measured by RMSE. It indicates how closely the model's predictions match the real data, with lower RMSE values suggesting higher predictive accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}$$

- MSE
  It is an estimator that measure the average of square errors.

$$\mathbf{MSE} = \frac{1}{n} \Sigma (Y_i - \hat{Y}_i)^2$$

- MAE
  It is a statistic used to evaluate the performance of regression models, especially time series forecasting models.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

- R squared
  It is the coefficient of determination.

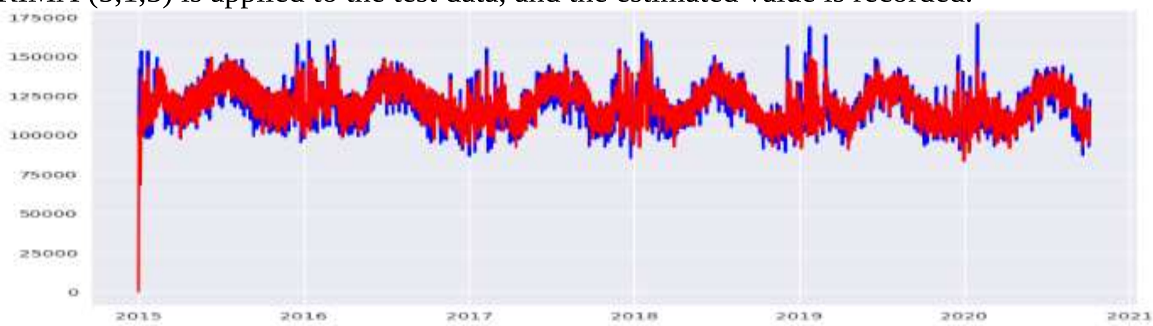
$$R^2 = \left(\frac{1}{n-1} \frac{\sum (x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} \right)^2$$

6.1 Experiment 1

<u>ARIMA for electricity Demand</u>

We used the ARIMA model to anticipate electrical usage in our study. We initially performed the Augmented Dickey Fuller Unit test to check that the data was suitable for modelling. This

test returns an array of values. In addition to that, we calculated the Mean Absolute Error (MAE) and other parameters to assess the test's efficiency.
The Dickey Fuller Unit test was run on electricity demand, the RRP, both demand positive RRP and negative demand RRP and both RRP positive and RRP negative. We concentrated on the p-value, aiming for a value less than 0.05. The p-value for all the parameters were less than 0.05, suggesting that our data sets are stationary and that we may proceed with the ARIMA modelling procedure.
In the initial phase of our experimentation, our objective was to identify the optimal ARIMA model for our dataset. After reviewing the findings of the stepwise fit, it was clear that ARIMA(5,1,3) was chosen as the best model for the electricity consumption dataset. The ARIMA (5,1,3) is applied to the test data, and the estimated value is recorded.



ARIMA	
Performance metrics	
Mean Squared Error	71394306
Mean Absolute Error	6082.29
Root Mean Square Error	8449.515
R squared	62%
Mean Absolute Percentage Error	5.11%

Table 1: Results of the ARIMA on Demand Prediction

The auto.arima() function which was installed from the package pmdarima shows that the best ARIMA model fitted for this dataset was the ARIMA(5,1,3)

```
Best ARIMA Model (p, d, q): (5, 1, 3)
Seasonal Order (P, D, Q, s): (0, 0, 0, 0)
```

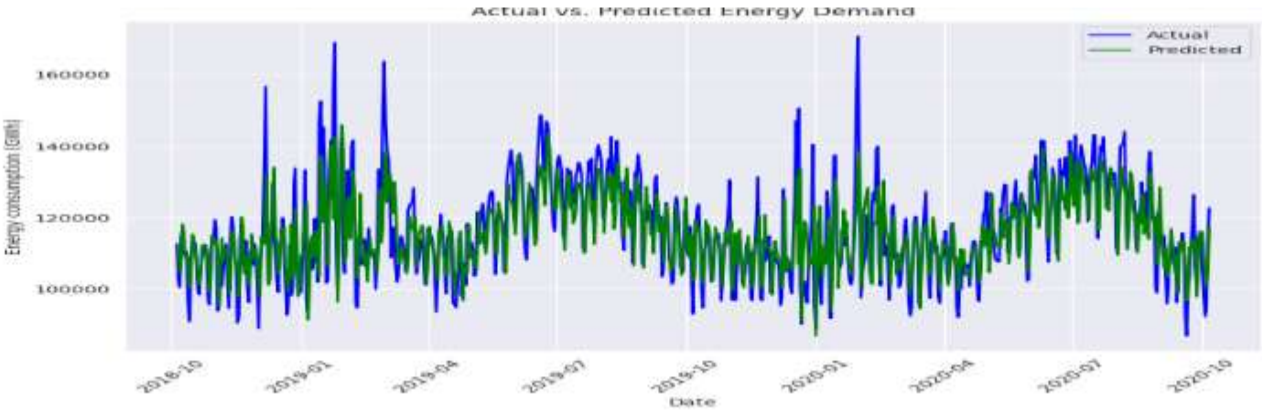
6.2 Experiment 2

<u>LSTM for Electricity Demand</u>

In our second experiment, we employed the Artificial Neural Network, also known as ANN (LSTM) to address this task.

(I) We imported essential libraries for data manipulation and neural network building.
(II) We defined an LSTM model through the 'create_lstm_model' function, sets a window size of 12 for input sequences, preprocesses data by normalizing it and prepares input-output pairs for the LSTM model.

(III) We split the data, compiles and trains the model and predict the energy consumption on the test sets,
(IV) Evaluated the model's performance using some metrics.
(V) Visualize the actual vs predicted plot.



LSTM	
Performance metrics	
Mean Squared Error	91272873
Mean Absolute Error	7249.4671
Root Mean Square Error	9553.68
R squared	51%
Mean Absolute Percentage Error	6.3399%

Table 1: Results of the LSTM on Demand Prediction

6.3 Experiment 3:

ARIMA for Electricity Price

In our third Experiment, ARIMA was applied on the electricity price dataset, which is the price of the electricity in Victoria Australia. The first step was to test the stationarity of the data, and since the p value was less than 0.05, the null hypothesis is rejected and we can proceed to the ARIMA model. In the first phase of the experiment, the objective is to determine the ideal and optimal ARIMA model for the dataset. It was clear that ARIMA(1,1,1) was chosen as the best model for the dataset. After applying the ARIMA(1,1,1) to the test data, the estimated value is recorded.

Mean Absolute Error	24.642774336
Mean Squared Error	15257.081
Root Mean Squared Error	123.519561
R squared	10.02%
Mean Absolute Percentage Error	37.61%

6.4 Experiment 4:

LSTM for Electricity Price:

Same as Experiment 2 ii), the ANN was employed with the same steps, and the LSTM model was created using the 'create_LSTM_model'. The results are as follows:

Mean Absolute Error	37.33804157
Mean Squared Error	41970.89
Root Mean Squared Error	204.8679
R squared	2.25%
Mean Absolute Percentage Error	46.3941%

6.5 Experiment 5

ARIMA for ASX index

In our third experiment, we applied the ARIMA model on the ASX index, the national index of Australia. Same as Experiment 1 (IX.i)) the first step was testing the stationarity of the data using ADF Test. The test showed a p-value of 0.75, which indicates a non-stationarity of the data. After using the difference data method and bringing back the data into a stationarity data with a p value of 0, the null hypothesis is now rejected and the best ARIMA model is fitted.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  884
Model:               SARIMAX(0, 1, 0)   Log Likelihood               -4626.682
Date:                Sat, 12 Aug 2023   AIC                           9255.363
Time:                        10:31:33   BIC                           9260.147
Sample:                             0   HQIC                          9257.192
                                - 884
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2      2083.0411     74.056     28.128      0.000    1937.895    2228.187
===================================================================================
Ljung-Box (L1) (Q):                   0.06   Jarque-Bera (JB):               119.96
Prob(Q):                              0.81   Prob(JB):                         0.00
Heteroskedasticity (H):               0.36   Skew:                            -0.40
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.61
===================================================================================
```

ARIMA	
Performance metrics	
Mean Squared Error	2084.97
Mean Absolute Error	34.3413
Root Mean Square Error	45.6615
R squared	98.34%

Table 1: Results of the ARIMA on ASX index

6.6 Experiment 6

LSTM for ASX200

Same as Experiment 2 ii), the ANN was employed with the same steps, and the LSTM model was created using the 'create_LSTM_model'. The results are as follows:

LSTM	
Performance metrics	
Mean Squared Error	4393.7689
Mean Absolute Error	49.8786
Root Mean Square Error	66.285510
R squared	96.28%

Table 1: Results of the LSTM on ASX index

7 Discussion

The goal of the study article is to comprehend the influence of ARIMA and LSTM on the prediction of the price of the ASX 200 Australian index, in addition to the demand and the price of the electricity in Australia, Victoria. Two machine learning algorithms are chosen to achieve the goal. hile ARIMA is widely recognized for its robust and efficient forecasting, the LSTM model stands out in its ability to predict sequences. This is due to the critical aspect of employing a “ remember and forget” gated design and carrying over necessary historical information for correct long-term dependence.

A suitable hardware of core i5-7200U CPU @2.50GHz 8.00GB RAM is utilized in the design specification template. To fit the models efficiently, a software Google Collab, accompanied with multiple libraries and packages such as pandas, pmdarima, numy, matplotlib.pyplot, seaborn, statsmodels.api, sklearn, dicky fuller and seasonal was installed. Data is gathered on a daily basis for five years; two different datasets were utilized:

- ASX 200 index: data consisting of 5 years starting 2015 ending 2020 were gathered on yahoo finance.
- Electricity Consumption dataset: data consisting of 5 years and half were gathered on Kaggle.com.

The preprocess and cleaning of the data was tested and validated separately, and a well-organized visualization of the data was implemented in the procedure. Different valuable parameters were used also to measure the accuracy and the results of machine learning, such as MSE, R squared, MAE, MAPE and RMSE.

8 Conclusion and Future Work

It is well known that machine learning techniques may be applied to real-world challenges. This research looks into how well ARIMA and LSTM machine learning models forecast the price of the ASX index of Australia, as well as the price and the demand of the Electricity in Australia Victoria. Despite their differing designs, both models were properly developed and executed. For training and validating the models, researcher have used two different datasets spanning five years, first one is a dataset related to the electricity demand and price in Australia Victoria, and the second dataset is the S&P/ASX 200 index. The data are normalized, outliers and missing values are removed, and the dataset is split into a train and test for validation. Performance assessment employs consistent metrics: Mean Error, MAE,RMSE,MAPE and Rsquared. The ARIMA model, based on the S&P/ASX200 and electricity dataset, outperforms the LSTM model in terms of accuracy based on lower values for MSE and MAE and RMSE. Additionally, the ARIMA model demonstrates a higher R squared value indicating a better fit to the data.

From these rankings, the Index Forecasting using the ARIMA model provides the most accurate predictions, followed by the Demand Price Forecasting with the ARIMA model. The Electricity Price Forecasting using ARIMA performs better than the LSTM.

References

Aprilianto, M., Hasibuan, M. N. S., & Harahap, S. Z. (2021). Forecasting health sector stock prices using ARIMAX method. Journal of Physics: Conference Series, 1815(1), 012077.

Brownlee, J. (2018). *How to Develop LSTM Models for Time Series Forecasting*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting [Accessed 13 Aug. 2023].

C. C. Emioma and S. O. Edeki (2020). Stock price prediction using machine learning on leastsquares linear regression basis, International Conference on Recent Trends in Applied Research (ICoRTAR)

Du, Y. (2018) 'Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network', in 2018 Chinese Control And Decision Conference (CCDC), Chinese Control And Decision Conference (CCDC), Shenyang, China, 9-11 June 2018, pp. 2854–2857. doi: 10.1109/CCDC.2018.8407611.

Eric (2022). Data Science Life Cycle: CRISP-DM and OSEMN frameworks. [online] data rundown. Available at: https://datarundown.com/data-science-life-cycle/#2-7 [Accessed 13 Apr. 2023].

Hayes, A. (2019). Autoregressive Integrated Moving Average (ARIMA). [online] Investopedia. Available at: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp [Accessed 13 Apr. 2023].

IBM (2021). CRISP-DM Help Overview. [online] www.ibm.com. Available at: https://www.ibm.com/docs/it/spss-modeler/saas?topic=dm-crisp-help-overview [Accessed 13 Apr. 2023].

Ingole, M. (2022). Data Science Process Framework (CRISP-DM). [online] www.linkedin.com. Available at: https://www.linkedin.com/pulse/data-science-process-framework-crisp-dm-mayur-ingole?trk=pulse-article [Accessed 13 Apr. 2023].

Khan, K. and Singh, D. (2022) 'Stock Price Forecasting of Maruti Suzuki using Auto Regressive Integrated Moving Average (ARIMA) Model', *Review of Professional Management*, 20(1), pp. 86–95. doi:10.1177/09728686221099285.

McNally, S., Roche, J. and Caton, S. (2018) 'Predicting the Price of Bitcoin Using Machine Learning', 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Parallel, Distributed and Network-based Processing (PDP), 2018 26th Euromicro International Conference on, PDP, pp. 339–343. doi:10.1109/PDP2018.2018.00060.

Mahla Nikou, Gholamreza Mansourfar and Jamshid Bagherzadeh (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithm, Intelligent Systems in Accounting Finance & Management 26:164-174.

Mehar Vijha, Deeksha Chandolab, Vinay Anand Tikkiwalb and Arun Kumarc, Stock Closing Price Prediction using Machine Learning Techniques (2020). International Conference on Computational Intelligence and Data Science (ICCIDS) 167: 599-606

Murthy, A. *et al.* (2022) 'Predicting Stock Price with LSTM Networks', *2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Mobile Networks and Wireless Communications (ICMNWC), 2022 IEEE 2nd International Conference on*, pp. 1–7. doi:10.1109/ICMNWC56175.2022.10031935.

Ozturk, S.; Ozturk, F. Forecasting energy consumption of Turkey by Arima model. J. Asian Sci. Res. 2018, 8, 52–60.

Paredes-Garcia, W.J. *et al.* (2019) 'Price Forecasting and Span Commercialization Opportunities for Mexican Agricultural Products', *Agronomy*, 9(12), p. 1AY. doi:10.3390/agronomy9120826.

Pandey, A. *et al.* (2023) 'Using ARIMA and LSTM to Implement Stock Market Analysis', *2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Artificial Intelligence and Smart Communication (AISC), 2023 International Conference on*, pp. 935–940. doi:10.1109/AISC56616.2023.10085405.

Samuel Olusegun, O., Pius Adewale, O., Maredi, M. & Jiliana, A., 2019. Stock market behaviour prediction using stackedLSTM network, Delhi,India:IEEE

Segal, T. (2022). Fundamental Analysis Definition. [online] Investopedia. Available at: https://www.investopedia.com/terms/f/fundamentalanalysis.asp [Accessed 11 Apr. 2023].

Wang, W. and Niu, Z. (2009) 'Time Series Analysis of NASDAQ Composite Based on Seasonal ARIMA Model', in 2009 International Conference on Management and Service Science. Wuhan, China, 20-22 Sept. 2009, pp. 1–4, IEEE Xplore. doi: 10.1109/ICMSS.2009.5300866.

Yee Han, L. (2019). SAAS Berkeley. [online] saas.berkeley.edu. Available at: https://saas.berkeley.edu/rp/arima [Accessed 12 Apr. 2023].