Harnessing the Power of AI in Investigating Trend of Lithium-ion Batteries Related Stock Market to Support the Sustainable Clean Energy Future

by Ahmad Kianrad

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Principle Supervisor Associate Professor. Eila Erfani, Co-Supervisor Professor. Amir H. Gandomi, External Supervisors Associate Professor. Amir Razmjou Chaharmahali, and Dr. Karim Hasani

University of Technology Sydney
Faculty of Engineering and Information Technology

July 2025

Certificate of Original Authorship

I, Ahmad Kianrad, declare that this thesis is submitted in fulfilment of the requirements for the award of PhD in the School of Computer Science, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship doi.org/10.82133/C42F-K220.

Signature: *Ahmad Kianrad* Production Note: Signature removed prior to publication.

Date: 25/07/2025

Acknowledgements

I would like to express my sincere gratitude to my principal supervisor, Associate Professor Eila Erfani, for her exceptional guidance, encouragement, and support throughout the course of my PhD. Her insights and feedback have been invaluable in shaping this research.

I am also deeply thankful to my co-supervisor, Professor Amir Hossein Gandomi, for his support and constructive suggestions. Special thanks to my external supervisors, Dr. Karim Hasani and Associate Professor Amir Razmjou Chaharmahali, whose expertise and perspectives greatly contributed to the depth and clarity of this work.

Last but not least, I want to thank my family for their unconditional love, patience, and encouragement throughout this journey. Their support has been the foundation of my progress and success.

List of Publications

- Kianrad, A., Arani, M.N., Hasani, K., Zargar, M, Erfani, E., Chaharmahali, A.R. Investigating the impact of company announcements on stock prices: an application of machine learning on Australian lithium market. Miner Econ 37, 163–172 (2024). https://doi.org/10.1007/s13563-024-00428-z

- Kianrad, A., Hasani, K., Erfani, E., Chaharmahali, A.R. & Gandomi, A.H., 2024. Twitter sentiment analysis and its impact on stock prices of Australian lithium-related companies: A support vector machine approach. In: Proceedings of the 2024 International Conference on Soft Computing and Machine Intelligence (ISCMI 2024).

- Kianrad, A., Hasani, K., Erfani, E., Chaharmahali, A.R. & Gandomi, A.H. Investigating the Impacts of X (Twitter) Data on Australian Lithium-Related Stock Market Trends, in *Journal of the Association for Information Systems.* (Under review).

Investigating the impact of company announcements on stock prices: an application of machine learning on Australian lithium market

Abstract

This paper investigates the effects of various types of announcements made by lithium producers on stock prices. We collected data from 40 lithium-producing companies listed on the world's largest stock exchanges, spanning from May 2020 to September 2022. To analyse the impact of announcements such as quoted and unquoted securities, market announcements, company reports, public meetings and presentations, financial announcements, and technical announcements on stock prices, we employed an extreme gradient boosting (XGBoost) model. Our results indicate that stock exchange market announcements and announcements about public meetings and presentations significantly influenced the stock prices of all eight large-cap companies studied. Announcements about public meetings and presentations were crucial predictors of stock prices for 73% of all companies analyzed. Additionally, positive financial announcements were key predictors for 70% of the companies. These findings suggest that investors should consider these predictors when making investment decisions in the lithium-related stock market. This study contributes to the existing literature by providing empirical evidence on the impact of different types of announcements made by lithium producers on stock prices. Furthermore, the XGBoost model used in this study can be applied to other industries and markets to analyze the impact of various types of announcements on stock prices.

Twitter Sentiment Analysis and Its Impact on Stock Prices of Australian Lithium-Related Companies: A Support Vector Machine Approach Abstract

Abstract

This paper aims to investigate the impact of social media contents, in particular tweets, on the stock price of lithium-related companies in Australia. To do so, first we use a sentiment analysis method to extract sentiments of companies’ tweets (neutral, positive, and negative). Second, we use the extracted sentiment as a feature in a Support Vector Machine (SVM) machine learning model to estimate the change direction of companies’ stock prices. The results of this research will provide insights to stakeholders, investors and business managers and help them make informed decisions.

Investigating the Impacts of X (Twitter) Data on Australian Lithium-Related Stock Market Trends

Abstract

This research investigates the complex connection between Twitter sentiment and the varying stock prices of Australian lithium companies. Highlighting the application of XGBoost algorithm, we apply an innovative method for modelling and analyzing market trends (from 2023 to 2024), emphasizing the significant influence of social media on stock market behavior. By employing appropriate machine learning techniques, we expose social media sentiment's critical role in affecting stock market dynamics. Demonstrating a relationship between Twitter sentiment and stock market results, our research offers valuable information for investors, analysts, and policymakers. It highlights the importance of using social media in making informed investment choices and policy decisions, contributing to ongoing discussions about the convergence of technology, social media, and the financial sector.

Table of Contents

List of Figures

List of Tables

Abstract

This thesis presents a comprehensive investigation into the factors influencing the stock market performance of lithium-related companies, with a particular focus on firms operating in Australia. Recognising the increasing significance of lithium in the global transition to clean energy and the dynamic nature of its market, this study explores how machine learning models can be leveraged to assess the impact of historical stock trends, formal corporate announcements, and social media content on stock price fluctuations. The research is conducted in three integrated parts, each examining a distinct data source to build a multi-faceted understanding of stock market behaviour in the lithium sector.

The first part of the study focuses on the analysis of historical stock price data for seven lithium-related companies listed in Australia. Drawing on one year of historical price records, this section applies a Support Vector Machine (SVM) model to determine how well past stock price patterns can predict future movements. The objective is to establish a baseline for stock price forecasting by understanding intrinsic trends within the companies' own trading history. This phase lays the analytical groundwork for integrating external data sources by identifying the degree to which internal historical patterns influence market behaviour.

The second part of the study shifts attention to external market signals in the form of formal announcements made by lithium producers. A broader dataset is employed, covering 40 lithium-producing companies listed on major global stock exchanges from May 2020 to September 2022. The announcements are categorised into six key types: quoted and unquoted securities, stock exchange market announcements, company reports, public meetings and presentations, financial announcements, and technical updates. An Extreme Gradient Boosting (XGBoost) model is used to assess the relative influence of each announcement type on stock price variation. The results indicate that stock exchange announcements and communications made during public meetings and presentations exert the strongest influence on stock prices, particularly for large-cap companies. Specifically, public meeting announcements emerged as significant predictors of stock price movement for 73% of all companies studied, while positive financial announcements were strong predictors for 70% of the firms. This part of the study contributes to the growing literature on financial market transparency by providing empirical evidence on how different forms of corporate disclosure impact investor behaviour and market valuation.

The third and final part of the study explores the influence of social media—specifically, Twitter—on the stock prices of lithium-related companies in Australia. In this phase, the focus is on unstructured textual data generated by companies themselves. Tweets published by these companies are analysed using the VADER sentiment analysis tool, which classifies each tweet into neutral, positive, or negative sentiment categories. The extracted sentiment scores are then used as features in an XGBoost model designed to analyse the directional change (increase or decrease) in stock prices. This section highlights the emerging importance of digital communication in shaping investor perceptions and market reactions. The results underscore the potential of social media sentiment as a supplementary influencer, offering timely insights for investors, analysts, and decision-makers.

Taken together, the three parts of this research demonstrate the value of using machine learning techniques to support data-driven decision-making in the lithium-related stock market. By integrating structured financial data, formal corporate disclosures, and unstructured social media content, the study provides a holistic framework for stock price analysis and investment analysis. Furthermore, the methodologies employed—SVM and XGBoost—show strong applicability across other sectors and markets, offering a replicable approach for future research.

This thesis contributes to the fields of Information Systems and Computer Science by developing a multi-source, machine learning–driven framework for analysing stock price movements in the lithium sector. Through the integration of structured historical data, formal corporate disclosures, and unstructured social media content, the study demonstrates how advanced computational techniques can support intelligent information processing and decision-making in volatile, data-rich environments.

The research applies and evaluates Support Vector Machines (SVM) and Extreme Gradient Boosting (XGBoost) models across diverse data types, offering methodological contributions in model design, feature engineering, and sentiment analysis. By incorporating natural language processing (NLP) techniques—specifically VADER sentiment analysis—within a financial context, the study illustrates how social media can be operationalised as a dynamic, real-time information system influencing market behaviour.

Moreover, this thesis highlights the role of corporate communications as formal information flows within socio-technical systems, providing new insights into how disclosure mechanisms and digital signals affect perception and valuation. The findings support the development of intelligent systems that combine internal and external data streams for adaptive forecasting and strategic decision-making.

This thesis contributes to Information Systems and Computer Science fields by advancing techniques in data-driven modelling, demonstrating cross-domain machine learning applications, and offering a replicable framework for integrating heterogeneous data sources to support market understanding and system behaviour analysis in complex, evolving domains.

Chapter 1 – Introduction

The Industrial Revolution marked a turning point in human history, propelling unprecedented technological and economic growth (Stearns, 2013). However, it also triggered a sharp rise in greenhouse gas (GHG) emissions, which have since become a significant contributor to climate change (IPCC, 2021). Since the latter half of the 20th century, global concerns over GHG emissions have intensified, highlighting the urgent need to transition towards a cleaner and more sustainable energy future (Stern, 2007; UNEP, 2019).

Central to this global shift is the development of renewable energy technologies and materials. Among these, lithium-ion batteries have emerged as a critical solution, offering an efficient and sustainable way to store electrical energy generated from renewable sources like solar and wind. Compared to traditional fossil fuels, lithium-ion batteries offer substantial environmental benefits, including longer operational lifespans, lower maintenance requirements, and significantly reduced GHG emissions over their lifecycle (Zubi et al., 2018). Moreover, advances in recycling and reusing these batteries can further reduce their environmental impact.

Monitoring trends in the lithium-ion battery market is essential for investors, policymakers, and industry leaders alike. Price analysis and trend forecasting enable stakeholders to allocate resources effectively, support strategic research and development, and foster innovation in clean energy technologies.

In this context, the application of data analytics and machine learning (ML) has become increasingly valuable. ML techniques are capable of processing large and complex datasets, uncovering hidden patterns, and adapting to dynamic market conditions in real time (Gururaj et al., 2019). These algorithms are particularly adept at identifying non-linear relationships and correlations in data, leading to increasingly accurate and reliable predictions (Yun et al., 2021).

This study employs machine learning algorithms to analyse future price fluctuations of lithium-related stocks based on current market data, with the aim of identifying high-impact explanatory features. The insights generated through this approach offer valuable guidance for investors, policymakers, and industry leaders striving to make data-driven decisions and drive innovation in the renewable energy sector.

Addressing climate change and environmental degradation requires a transition to low-carbon energy sources, which studies have shown are fundamental to long-term sustainability (Nishant et al., 2020). As a key mineral-based energy resource, lithium is vital to this transition. Its use spans from portable electronics and laptops to electric vehicles (EVs) and energy storage systems, making it an indispensable part of modern life and accelerating global demand.

Lithium's industrial significance has evolved over time. While it was once primarily used in glass and ceramics, demand has shifted towards battery production due to the growth of the electronics and renewable energy industries (Dessemond et al., 2019; Grey & Hall, 2020). In 2021, global lithium consumption reached 93,000 tons—a 33% increase from 2020—and is projected to grow ninefold by 2030 (Lithium, 2022).

Australia plays a pivotal role in this global market. With four major mineral lithium operations, the country accounted for 52% of global lithium production in 2021. Leading Australian Stock Exchange (ASX)-listed lithium producers include BHP, Wesfarmers, Pilbara Minerals, Mineral Resources, Allkem, and IGO—each with market capitalizations exceeding AUD 5 billion.The surge in lithium demand is closely linked to the adoption of EVs. In 2023 alone, Australia's EV market grew by 120%, with EVs comprising 8.4% of new car sales (Australian Vehicle Council, 2023). The International Energy Agency (IEA) anticipates that lithium demand will increase more than 40-fold as the global transition to sustainable energy accelerates.

In parallel, social media platforms—especially X (formerly Twitter)—are increasingly used by companies, executives, and stakeholders to communicate with the public. The sentiment expressed in these posts can have significant influence on investor perception and stock prices. Numerous studies have demonstrated the effectiveness of sentiment analysis tools in classifying social media content, thereby enabling its use as a influencer input in machine learning models. This study leverages such tools to examine how company-generated social media posts impact stock price trends for Australian lithium-related companies.

With Australia holding the world's second-largest lithium reserves and producing approximately 63% of the global supply (Geoscience Australia, 2018), the nation's lithium industry is of strategic global importance. Understanding the Australian lithium stock market through a data-driven lens is not only timely but essential for informed decision-making in the clean energy transition.

1.2 Problem Statement

The rising global demand for lithium-ion batteries driven by the shift toward clean energy and electric mobility has positioned lithium as a critical material in the global economy. Despite its strategic importance, the stock market performance of lithium-related companies, particularly those operating in Australia, remains highly volatile and difficult to analyse. This volatility stems from a complex interaction of internal factors (such as company-specific financial trends) and external influences (including corporate announcements and social media discourse).

Conventional statistical models are often limited in their ability to capture the nonlinear and multidimensional nature of stock market behaviour (Atsalakis & Valavanis, 2009; Patel et al., 2015). While the application of machine learning and artificial intelligence (AI) in financial forecasting has gained momentum (Feng et al., 2018; Sezer et al., 2020), there remains a notable research gap in applying these methods to lithium-related stocks, especially in the Australian context (Liu et al., 2022). Moreover, few studies have attempted to integrate heterogeneous data sources such as historical price data, formal corporate disclosures, and unstructured social media sentiment to understand market dynamics holistically (Zhang et al., 2020). There is a growing need for insights that can guide investors, policymakers, and industry leaders in making data-driven decisions and driving innovation in the renewable energy sector.

This study addresses these gaps by developing and evaluating a machine learning–driven framework that leverages Support Vector Machines (SVM), Extreme Gradient Boosting (XGBoost), and sentiment analysis techniques to model and investigate lithium stock performance. By incorporating both structured and unstructured data, this research contributes to the fields of Information Systems and Computer Science, offering novel insights into how AI-based models can support intelligent decision-making and market analysis in complex, data-rich environments.

1.3 Research Aim / Objective

This thesis aims to leverage AI – Machine Learning to analyse trends in the Australian lithium-related stock market by examining how key factors—including historical stock data, traditional company news and announcements, and social media content from X (formerly Twitter)—influence stock movements and investor sentiment.

To the best of our knowledge, while AI – Machine Learning has been applied in lithium supply chains, limited empirical studies exist on analysing lithium-related stock prices in Australia. This research considers factors such as:

- Historical stock data momentum,
- Company news and announcements,
- Social media data from X (formerly Twitter).

These data reflect corporate communications, public opinion, and investor sentiment. The study focuses on prominent Australian lithium companies and aims to provide actionable insights for investors, company managers, and policymakers.

While machine learning models such as SVM and XGBoost are commonly used for prediction, the core aim of this study is exploratory. These models were employed to investigate and analyse how historical data, traditional news, and social media interactions relate to stock prices, rather than to produce high-performing predictive models.

1.4 Research Questions

1. What is the trend of lithium-ion battery-related stock market in Australia?
   - Historical stock price data (data from November 2020 to November 2021) of seven lithium-related companies in Australia.

2. How and to what extent do traditional news data impact the trend of lithium-ion battery-related stock market in Australia?

   - Company-generated news and announcements (data from May 2020 to September 2022) of 40 lithium-related companies.

3. How and to what extent do social media data impact the trend of lithium-ion battery-related stock market in Australia?

   - Official social media posts from X (formerly Twitter) (data from July 2023 to July 2024) of five lithium-related companies.

The aim of this study is to investigate the effects of historical stock prices, traditional news, and social media posts on the stock market of lithium-related companies in Australia.

1.5 Significance / Rationale

This study addresses a gap in the literature concerning AI applications in the Australian lithium-related stock market. It contributes:

- **Theoretically**, by exploring the role of traditional news and social media data in stock market analysis.

- **Practically**, by developing AI tools that enable investors and stakeholders to make informed decisions.

By investigating how news categories and social media sentiments influence stock prices, this research helps reduce the risk associated with investing in green technologies and supports the broader global shift toward sustainable energy.

1.6 Scope and Delimitations

This study focuses on Australian lithium-related companies. It analyses:

- Historical stock price data,
- Company news and announcements,
- Official X (Twitter) posts by companies.

The study does **not** include:

- Posts by the general public,
- Investor commentary,
- Hashtags or influencer posts.

This deliberate narrowing of scope allows for clearer insights into how company-originated data affects stock performance.

1.7 Methodological Overview

This study adopts a quantitative approach, integrating multiple data sources relevant to the Australian lithium-related stock market. The data sources include:

- Historical stock price data (data from November 2020 to November 2021),
- Company-generated news and announcements (data from May 2020 to September 2022),
- Official social media posts from X (formerly Twitter) (data from July 2023 to July 2024).

To model and analyse stock price movements, the research employs the following machine learning techniques:

- Extreme Gradient Boosting (XGBoost),
- Support Vector Machines (SVM),
- Linear Regression (as a baseline model).

Social media content is further analysed using VADER sentiment analysis, which categorises posts into positive, negative, or neutral sentiment. These sentiment scores are incorporated as features in the machine learning models to assess their analysis value.

A detailed explanation of data preprocessing, feature engineering, model training, and evaluation procedures is provided in Chapter 3.

The following chapters elaborate on the literature review, research methodology, findings, and implications of the study. Chapter 2 reviews prior research on stock forecasting, lithium markets, and machine learning applications. Chapter 3 outlines the research design and data analytics approach, while Chapters 4 and 5 present the results and discussion, respectively.

Chapter 2 – Literature Review

This chapter provides the theoretical and contextual foundation for the thesis, which investigates how machine learning models can predict the stock market performance of lithium-related companies. By placing lithium within the broader context of the clean energy transition and examining both the industrial supply chain and financial markets, the literature review builds a bridge between the material importance of lithium and its influence on investor behaviour. Section 2.1 highlights the environmental and economic importance of lithium technologies, establishing why companies in this sector warrant financial scrutiny. Section 2.2 then explores the increasing role of data analytics within the lithium supply chain, setting the stage for Section 2.3, where the analysis modelling of stock price behaviour using structured and unstructured data is examined.

2.1 Energy Resources, Lithium Technologies, and Market Dynamics

The growing significance of lithium as a strategic resource is central to this thesis. As the global energy system transitions away from fossil fuels, lithium has become indispensable in supporting renewable technologies, particularly through its role in electric vehicle (EV) batteries and portable electronics. Understanding the market dynamics and environmental implications of lithium is essential for contextualizing why lithium-related companies attract increasing investor attention. This relevance forms the backdrop against which financial performance prediction—especially in volatile stock markets—becomes a critical focus of study.

Energy resources have a significant societal impact, and in today's world, both their benefits and limitations are subject to increasing scrutiny. Fossil fuels have historically powered industrial and societal growth, yet numerous studies underscore their adverse environmental effects due to high carbon emissions. Grant et al. (2020) advocate for the adoption of low-emission alternatives to mitigate environmental damage. Research by Colaka et al. (2012) supports the transition to renewable energy sources such as solar, wind, and geothermal.

Mineral-based energy sources, particularly lithium, have emerged as critical alternatives due to their role in reducing atmospheric carbon. Lithium is a key component in batteries used in everyday electronics such as smartphones and laptops, and it plays an increasingly vital role in electric vehicle (EV) technology. Lithium-ion batteries, the most widely used type, are found in devices ranging from personal electronics to EVs. These batteries vary in composition depending on the materials used, such as lithium-sulphur (Li-S), which shows promise for greater energy capacity and lower environmental impact.

Lithium batteries, particularly those used in EVs, rely on natural and often scarce raw materials. While lithium-ion batteries offer advantages like energy density and performance, they also present environmental challenges due to short lifespans and issues related to battery disposal and resource extraction. Garrido-Hidalgo et al. (2019) emphasized how different materials influence battery specifications in terms of density, safety, durability, cost, and recyclability.

EVs and hybrid vehicles predominantly rely on lithium-ion batteries, which use graphite carbon as the anode and varying materials in the cathode. This design flexibility enables manufacturers to optimize supply chains and reduce production costs. While lithium-ion battery costs are declining and performance is improving, lithium-sulphur batteries are gaining attention as a future alternative due to their lower environmental impact and high energy capacity.

2.2 Application of Data Analytics in the Lithium Supply Chain

To deepen this contextual foundation, Section 2.2 reviews how data-driven technologies are being integrated into the lithium supply chain. These innovations not only enhance production efficiency and transparency but also demonstrate the applicability of machine learning and AI in operational decision-making. This technical perspective complements the core aim of this thesis: to extend such data-driven methodologies into the financial domain. By understanding how AI supports optimization in the physical lifecycle of lithium products, we can better appreciate its potential in modelling stock price movements driven by market signals, corporate disclosures, and social sentiment.

As global demand for lithium-ion batteries continues to rise due to their widespread use in electric vehicles (EVs), portable electronics, and renewable energy storage, the lithium supply chain has become a critical focal point for innovation and optimization. The application of data analytics—including data mining, machine learning (ML), and artificial intelligence (AI)—has emerged as a transformative approach to address the challenges of efficiency, traceability, quality, and sustainability across the lithium battery life cycle.

2.2.1 Data Mining and Process Optimization

Schnell et al. (2019) utilized data mining models to predict battery capacity and lifespan by analysing production data, helping manufacturers reduce the rate of failure and increase product reliability. Similarly, Huber et al. (2016) leveraged image recognition and data analytics to detect physical defects within battery components, ensuring higher quality standards throughout the supply chain.

Lithium-ion battery production involves intricate processes such as electrode coating, calendaring, electrolyte filling, formation cycling, and cell assembly, each of which affects final battery performance and cost. Wessel et al. (2020) emphasized that these processes are data-rich and highly variable. By applying data mining across each step, manufacturers can develop a "digital twin" of the production line, enabling real-time monitoring and optimization.

2.2.2 Traceability and Lifecycle Monitoring

Wessel et al. (2020) proposed a model for full traceability, where each component in the supply chain is assigned a unique identifier and logged in a centralized system. This enables end-to-end data transparency for quality assurance, regulatory compliance, and sustainable recycling efforts.

2.2.3 Role of AI in Supply Chain Optimization

AI improves efficiency across multiple stages of the supply chain—from lithium exploration using satellite imaging to smart manufacturing and predictive maintenance. ML algorithms help forecast demand, optimize production, and predict equipment failures, while also supporting real-time quality control.

2.2.4 AI in Market Analysis and Strategic Planning

Market stakeholders now use AI to analyse not only historical stock data but also macroeconomic indicators and public sentiment. Jiao et al. (2020) found that media exposure can influence market volatility, while Liu et al. (2015) showed that social network data can help detect market clusters.

2.2.5 Challenges in Data-Driven Lithium Systems

Despite progress, challenges include the lack of data standardization, privacy concerns, cost barriers for SMEs, and the "black box" nature of many AI models. These gaps limit adoption and complicate compliance with emerging global regulatory frameworks.

To build on these insights, it is also necessary to examine how machine learning and AI techniques—so impactful in optimizing the physical and operational aspects of the lithium sector—are being leveraged in financial contexts. In particular, the investigation and analysis of stock prices for lithium and other resource-focused companies has become a growing area of research. This transition is essential as it highlights how predictive technologies extend from production lines to capital markets, forming a holistic view of value creation, investment potential, and market risk. However, while there is a growing body of research on stock market prediction using machine learning, few studies have specifically focused on lithium-related companies—particularly those operating in Australia, despite its status as a leading global lithium producer. Moreover, most existing models examine stock prices, corporate disclosures, or social media in isolation, rather than through an integrated lens. This thesis addresses these gaps directly, offering a unified framework that combines structured financial data, corporate announcements, and social media sentiment to analyse the stock market and provide insights for investors.

2.3 Literature Review on Stock Market Forecasting Using Machine Learning and Sentiment Analysis

To extend the application of machine learning and artificial intelligence beyond operational optimization in the lithium supply chain, it is equally important to explore how these technologies contribute to financial forecasting. As lithium becomes an increasingly strategic resource, not only for clean energy but also for national economic planning, the ability to predict stock price movements of lithium-related companies has gained critical importance. This section shifts focus from supply chain analytics to the predictive modelling of financial markets, highlighting how advanced analytical techniques—particularly those using ML and sentiment analysis—are shaping modern stock market forecasting.

The next section expands the focus from industrial systems to financial analytics, providing a structured review of stock market analysis models grounded in ML, AI, and sentiment analysis.

2.3.1 Introduction to Stock Market Prediction Challenges

Over the past decades, the stock market has become an increasingly significant aspect of national and individual economic interests. Investors and stakeholders have progressively turned to new technologies to mitigate risk and enhance profitability in stock trading. However, the volatile, non-linear, and dynamic nature of the stock market presents substantial forecasting challenges.

Fama (1995) advanced the Efficient Market Hypothesis (EMH), proposing that stock prices are inherently unpredictable due to their reflection of all available information. This theory, based on the notion of random walks, has faced increasing scrutiny with the evolution of predictive modeling through machine learning (ML) and artificial intelligence (AI).

2.3.2 Machine Learning Techniques in Stock Market Prediction

Tan et al. (2007) developed the Genetic Complementary Learning (GCL) model to simulate cognitive processes and improve market trend prediction. Similarly, Gupta R. et al. (2013) evaluated various ML algorithms using the Kappa statistic to identify models with the highest accuracy in stock forecasting.

Inthachot et al. (2016) used six years of historical data from Thailand's SET50 index, integrating Genetic Algorithms (GA) for feature selection and Artificial Neural Networks (ANN) for prediction. Their hybrid model significantly outperformed traditional methods. Likewise, Dey et al. (2016) applied Xtreme Gradient Boosting (XGBoost) to predict price directions over 60- and 90-day periods, confirming the superiority of ensemble learning models.

2.3.3 Social Media and Financial News Sentiment Analysis

Social media sentiment, particularly from platforms like Twitter (X), has emerged as a valuable data source for analysing. Bujari et al. (2017) introduced cashtags (e.g., $AAPL) as predictive features in modelling. Howells and Ertugan (2017) proposed a fuzzy logic approach to classify tweet sentiment beyond traditional binary labels, improving the handling of informal language and emojis. Devi and Bhaskaran (2015) integrated Twitter and financial news using ontology-based semantic analysis, improving forecasting accuracy. Ranco et al. (2015) found that tweet sentiment, especially during high-activity periods, significantly affects market movements. Smith and O'Hare (2022) compared traditional news and social media (tweets from CEOs) with stock price movements in pre-pandemic and pandemic period and showed that there is a weak relationship between Twitter sentiments and stock prices.

2.3.4 Comparative Studies and Hybrid Approaches

Nti et al. (2020) evaluated ensemble models such as stacking, blending, bagging, and boosting across four major stock exchanges, finding stacking and blending most effective. Saravanos and Kanavos (2024) combined regression techniques and deep neural networks (DNNs), showing how Twitter sentiment and historical data enhance prediction.

Rahman et al. (2017) utilized Support Vector Machines (SVM) to classify financial news sentiments in the Malaysian market. Liu et al. (2023) integrated FinBERT with SVM to analyze investor posts on Stocktwits, highlighting both the potential and challenges of sentiment integration.

2.3.5 Deep Learning and Advanced Neural Models

Multiple studies have highlighted the success of deep learning models such as LSTM, GRU, CNN, and hybrid frameworks:

- Sirimevan et al. (2019) integrated LSTM-RNNs and ensemble methods using Dow Jones data.
- Sedighi et al. (2019) used a hybrid of Artificial Bee Colony and Adaptive Neuro-Fuzzy Inference System with SVM.
- Rezaei et al. (2021) combined LSTM, CNN, and Empirical Mode Decomposition (EMD) techniques.

Jiang (2021) and Hu et al. (2021) provided comprehensive reviews of deep learning in stock market forecasting, emphasizing advantages and challenges like overfitting and interpretability.

2.3.6 Cross-Market and Global Influence Models

Kumar et al. (2023) studied the interdependencies between Asian stock markets using Granger Causality and Pearson's Correlation alongside LSTM. Yadav et al. (2024) developed a hybrid LSTM-GRU model, applied to Tata Motors and Honda Motors data, which outperformed conventional models.

Li and Bastos (2020) conducted a systematic review of 34 studies, identifying technical indicators frequently used as ML features and underlining interpretability issues.

2.3.7 Sentiment-Enriched Models and Case-Specific Applications

Jing et al. (2021) and Prachyachuwong & Vateekul (2021) illustrated how sentiment scores from investor forums and news headlines can improve model accuracy when combined with technical indicators. Valle-Cruz et al. (2021) analyzed Twitter sentiment during the COVID-19 pandemic, finding strong correlations with market movements.

De Oliveira Carosia et al. (2021) applied CNN to Brazilian news to guide investment strategies, while Li, Wu, and Wang (2020) used LSTM with sentiment vectors and technical features to improve accuracy in the Hong Kong market.

2.3.8 Model Optimization and Feature Engineering

Sharma et al. (2021) optimized ANN parameters using Genetic Algorithms (GA). Misra and Chaurasia (2020) demonstrated that raw input data with SVM outperformed other configurations, underscoring the role of feature selection and dataset composition.

Chung and Shin (2020) proposed a GA-optimized CNN hybrid model, while Yu and Yan (2020) used Phase Space Reconstruction (PSR) with DNNs to capture stock market complexity. Hoseinzade and Haratizadeh (2019) utilized CNNs with a wide range of input variables, including external stock data and currency exchange rates. Cao et al. (2019) merged complex network theory with SVM and KNN, highlighting the potential of interdisciplinary approaches. Basak et al. (2019) compared tree-based ML models and found XGBoost superior in forecasting price direction using technical indicators like MACD and RSI.

Despite the breadth of literature on stock market forecasting, several research gaps persist—particularly in the context of lithium-related markets. Many existing studies focus on broad indices or large-cap technology stocks, with limited attention paid to the resource sector, especially in regions like Australia. Moreover, few studies explicitly examine the combined impact of stock market data, corporate announcements, and social media sentiment on the financial performance of lithium companies. This research addresses these gaps by applying an XGBoost-based machine learning approach to integrate multiple data sources, including historical stock prices, corporate disclosures, and Twitter sentiment, to analyse the stock performance of lithium-related companies in the Australian market. By doing so, it provides a more nuanced and context-specific understanding of how AI-driven insights can support strategic decision-making for investors, analysts, and policymakers in the evolving clean energy economy.

Chapter 2 Summary – Literature Review

In summary, data analytics—particularly AI, ML, and data mining—are reshaping the lithium supply chain from end to end. Applications range from mine discovery and manufacturing optimization to lifecycle traceability and market forecasting. These technologies offer significant benefits in terms of efficiency, cost reduction, sustainability, and decision support. However, challenges related to data quality, system integration, and model transparency remain and warrant further research. This study contributes to this growing body of knowledge by applying AI techniques to explore the relationship between diverse data sources (e.g., stock prices, social media posts) and the performance of lithium-related markets in Australia.

This chapter establishes the interdisciplinary foundation for the thesis by situating lithium within the broader landscape of the clean energy transition and exploring its intersection with financial markets and data science. Section 2.1 highlights lithium's environmental and economic importance, emphasizing its role in battery technologies critical to electric vehicles and renewable energy systems. It underscores the increasing investor attention toward lithium-related companies due to their strategic relevance and market volatility. Section 2.2 shifts focus to the integration of data analytics in the lithium supply chain. It reviews how machine learning, artificial intelligence, and digital twin technologies are enhancing efficiency, traceability, and sustainability across the lithium lifecycle. These operational applications set the stage for the thesis's central focus—extending such data-driven methodologies to stock market analysis.

Section 2.3 reviews the growing body of literature on stock market prediction using machine learning and sentiment analysis. It examines a wide range of approaches, including SVM, XGBoost, LSTM, and ensemble models, and explores how social media sentiment, particularly from Twitter, complements traditional financial indicators. Despite extensive global research, the review identifies critical gaps: limited attention to lithium-focused companies, especially in Australia, and the lack of integrated models combining structured and unstructured data sources. By addressing these gaps, the chapter positions the thesis as a novel contribution that applies machine learning techniques to a multi-source dataset—historical stock data, corporate disclosures, and social media sentiment—to analyse the financial performance of lithium-related companies in Australia. This framework aims to support more informed, timely, and strategic decision-making in the resource-driven financial sector.

These insights from the literature provide the foundation for the methodological choices presented in Chapter 3, where machine learning techniques are applied to evaluate how these data sources influence lithium-related stock price behaviour.

Chapter 3 – Research Methodology

Building on the literature review in Chapter 2, this chapter translates theoretical foundations into an empirical framework using machine learning models tailored to the dataset and research objectives.

3.1 Introduction

This chapter outlines the research methodology developed to address all three research questions guiding this study. It presents the overall design, data sources, analytical techniques, and the rationale behind the selected machine learning models. The goal is to provide a transparent, structured, and replicable approach for how data was collected, processed, and analysed to generate meaningful insights into stock market behaviour within the Australian lithium sector.

Given the nature of the study focused on the financial dynamics of lithium-ion battery-related companies in Australia a quantitative, data-driven methodology was adopted. The stock market, particularly within emerging energy sectors, is characterised by high volatility, non-linearity, and sensitivity to a range of internal and external signals. These characteristics often limit the effectiveness of traditional statistical methods.

To address these complexities, this study applies machine learning (ML) techniques to model and analyse stock price trends. Support Vector Machines (SVM) are used as the core model due to their strong performance in handling high-dimensional and non-linear classification tasks. Additional techniques, such as XGBoost and sentiment analysis using VADER, are employed where appropriate to analyse structured and unstructured data.

The methodology is designed to answer the following research questions:

- What is the trend of lithium-ion battery-related stock market in Australia? Addressed using historical stock price data and ML-based trend forecasting.
- How and to what extent do traditional news data impact the trend of lithium-ion battery-related stock market in Australia? Investigated through corporate announcements and formal disclosures analysed via XGBoost.
- How and to what extent do social media data impact the trend of lithium-ion battery-related stock market in Australia? Explored using sentiment analysis of Twitter posts and their integration into predictive ML models.

While machine learning models are often applied for predictive tasks, the models selected in this study Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) are employed primarily as analytical tools. Their purpose is to examine the relationships between historical stock data, traditional news announcements, and social media activity, rather than to maximise predictive accuracy. This approach aligns with the broader aim of the research, which is to investigate and interpret how various information sources influence stock price movements in the Australian lithium market.

The following sections of this chapter detail the rationale for using machine learning, the process of data sourcing and preprocessing, the design and implementation of the models, and the evaluation procedures. A visual overview of the methodological workflow and a discussion of expected contributions and limitations conclude the chapter.

XGBoost was selected to analyse the effects of traditional news and social media interactions on lithium-related stock prices due to its suitability for structured financial data and its ability to model nonlinear relationships. The dataset used in this study consists of structured numerical features—such as stock indicators, transformed news categories, sentiment scores, and tweet engagement metrics—which align well with XGBoost's strengths in handling small-to-medium structured tabular data.

As stock market behaviour is influenced by complex, nonlinear interactions between sentiment, communication patterns, and historical indicators, a linear model would be insufficient. XGBoost's tree-based architecture enables it to capture these nonlinearities and interactions effectively, offering deeper analytical insight than linear regression or linear SVM.

The model also provides feature importance outputs, which align with the exploratory nature of this research by enabling interpretation of which news categories or social media engagement indicators most strongly relate to stock price movements.

Furthermore, XGBoost demonstrates robustness to missing or noisy data, which is essential when working with inconsistent social media activity or irregular corporate disclosures. Compared with alternatives such as Random Forest, deep learning models, or SVM regression, XGBoost offers a strong balance of performance, interpretability, and efficiency for this study's investigative aims.

3.2 Research Question 1: What is the trend of lithium-ion battery-related stock market activity in Australia?

To address this research question, a quantitative, data-driven methodology was employed using historical stock price data from a selection of Australian companies involved in the lithium industry. Due to the inherent volatility and non-linear behaviour of stock markets, particularly in emerging sectors like lithium, traditional statistical models often fall short in capturing complex market dynamics.

Therefore, this study applied machine learning (ML) techniques—specifically the Support Vector Machine (SVM) algorithm—to investigate the directional movement (upward or downward) of stock prices on the following trading day. The methodology comprises data collection, preprocessing, feature engineering, model training, and evaluation. The goal is to generate insights useful to investors, analysts, and policymakers navigating the Australian lithium market.

Figure 3.1 demonstrates the conceptual model and workflow of the stock market data analytics for this project.

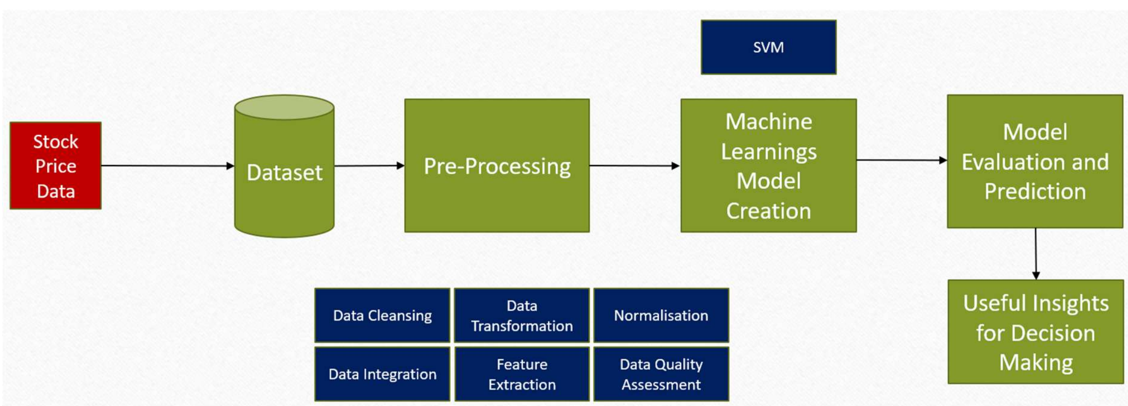


Figure 3.1: Conceptual model and workflow

3.2.1 Rationale for Machine Learning Approach

The Australian stock market, especially lithium-related stocks, is influenced by a wide range of volatile and interdependent factors, such as global demand for electric vehicles, commodity prices, investor sentiment, and regulatory changes. These factors exhibit non-linear interactions that limit the effectiveness of conventional analytical tools like linear regression or classical time-series models (e.g., ARIMA).

Machine learning algorithms, and particularly SVMs, have shown strong performance in financial analysis due to their capacity to detect patterns in complex, high-dimensional datasets. SVMs are robust to overfitting (when properly tuned), offer high generalisation ability, and are effective for classification tasks. Hence, the SVM model was selected to classify daily stock movements as either upward or downward.

3.2.2 Data Sources and Collection

The dataset includes historical stock price data from seven publicly listed Australian companies engaged in lithium exploration, extraction, or battery production. Selection criteria included market capitalisation, relevance to the lithium value chain, and consistent trading activity during the study period.

No.	Company Name	Ticker Symbol	Data Volume
1	Liontown Resources	LTR	249 rows
2	Mineral Resources	MIN	253 rows
3	Orocobre Limited	ORE (merged)	253 rows
4	Piedmont Lithium	PLL	245 rows
5	Pilbara Minerals	PLS	256 rows
6	Vulcan Energy Resources	VUL	151 rows
7	Wesfarmers	WES	253 rows

Table 3.1 Selected Companies for Analysis

3.2.3 Timeframe of Data

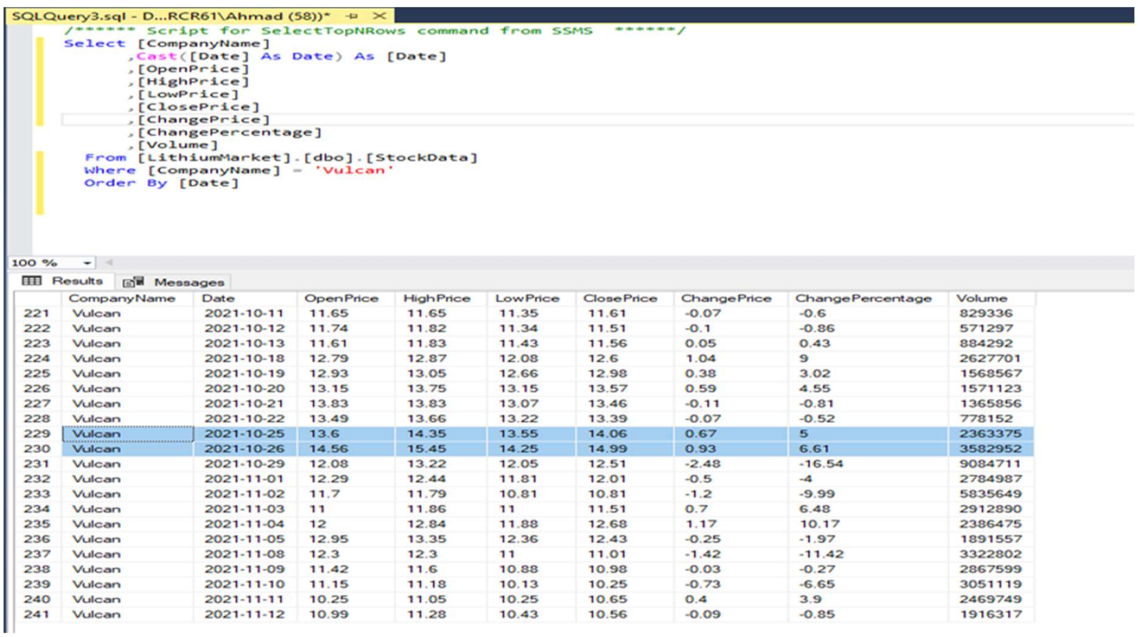
The analysis covers the period from November 2020 to November 2021. Data were sourced in CSV format from publicly available platforms such as Yahoo Finance, the Australian Securities Exchange (ASX), and the Alpha Vantage API.

Each dataset includes the following attributes:

- Date
- Open Price

- High Price
- Low Price
- Close Price
- Adjusted Close Price
- Volume Traded

An example of the raw dataset is shown in Figure 3.2, presenting structured stock data for Vulcan Energy Resources (VUL) retrieved from a SQL Server database. This data served as the foundation for feature engineering and model training.


SQLQuery3.sql - D...RCR61\Ahmad (58))*

```sql
/****** Script for SelectTopNRows command from SSMS  ******/
Select [CompanyName]
      ,Cast([Date] As Date) As [Date]
      ,[OpenPrice]
      ,[HighPrice]
      ,[LowPrice]
      ,[ClosePrice]
      ,[ChangePrice]
      ,[ChangePercentage]
      ,[Volume]
  From [LithiumMarket].[dbo].[StockData]
  Where [CompanyName] = 'Vulcan'
  Order By [Date]
```

100 %

Results | Messages

	CompanyName	Date	OpenPrice	HighPrice	LowPrice	ClosePrice	ChangePrice	ChangePercentage	Volume
221	Vulcan	2021-10-11	11.65	11.65	11.35	11.61	-0.07	-0.6	829336
222	Vulcan	2021-10-12	11.74	11.82	11.34	11.51	-0.1	-0.86	571297
223	Vulcan	2021-10-13	11.61	11.83	11.43	11.56	0.05	0.43	884292
224	Vulcan	2021-10-18	12.79	12.87	12.08	12.6	1.04	9	2627701
225	Vulcan	2021-10-19	12.93	13.05	12.66	12.98	0.38	3.02	1568567
226	Vulcan	2021-10-20	13.15	13.75	13.15	13.57	0.59	4.55	1571123
227	Vulcan	2021-10-21	13.83	13.83	13.07	13.46	-0.11	-0.81	1365856
228	Vulcan	2021-10-22	13.49	13.66	13.22	13.39	-0.07	-0.52	778152
229	Vulcan	2021-10-25	13.6	14.35	13.55	14.06	0.67	5	2363375
230	Vulcan	2021-10-26	14.56	15.45	14.25	14.99	0.93	6.61	3582952
231	Vulcan	2021-10-29	12.08	13.22	12.05	12.51	-2.48	-16.54	9084711
232	Vulcan	2021-11-01	12.29	12.44	11.81	12.01	-0.5	-4	2784987
233	Vulcan	2021-11-02	11.7	11.79	10.81	10.81	-1.2	-9.99	5835649
234	Vulcan	2021-11-03	11	11.86	11	11.51	0.7	6.48	2912890
235	Vulcan	2021-11-04	12	12.84	11.88	12.68	1.17	10.17	2386475
236	Vulcan	2021-11-05	12.95	13.35	12.36	12.43	-0.25	-1.97	1891557
237	Vulcan	2021-11-08	12.3	12.3	11	11.01	-1.42	-11.42	3322802
238	Vulcan	2021-11-09	11.42	11.6	10.88	10.98	-0.03	-0.27	2867599
239	Vulcan	2021-11-10	11.15	11.18	10.13	10.25	-0.73	-6.65	3051119
240	Vulcan	2021-11-11	10.25	11.05	10.25	10.65	0.4	3.9	2469749
241	Vulcan	2021-11-12	10.99	11.28	10.43	10.56	-0.09	-0.85	1916317

Figure 3.2 Sample raw data

This raw data was subjected to preprocessing and feature engineering to derive stock volatility and momentum—used as inputs in the SVM classification model.

3.2.4 Data Preprocessing and Feature Engineering

Several preprocessing steps were performed to prepare the data for model training:

- **Data Cleansing**: Removed missing values, duplicates, and anomalies such as zero or negative prices

- **Date Standardisation**: Ensured uniform date formats across all datasets
- **Feature Generation**:
    - *Price Movement Label*: Binary label (1 = price increase next day, 0 = no change or decrease)
    - *Daily Return*: (Close - Open) / Open
    - *Technical Indicators*: Moving Averages (MA5, MA10), RSI, trading volume changes

3.2.5 Machine Learning Model: Support Vector Machine (SVM)

SVM is a supervised learning algorithm designed for classification tasks. It finds an optimal hyperplane that separates data points into distinct classes—in this case, upward or downward movement of stock prices.

Introduced by Cortes and Vapnik (1995), SVM constructs a decision boundary that maximises the margin between data points of different classes. It supports both linear and non-linear classification through the use of kernel functions.

A comprehensive review by Ghoddusi et al. (2019), which surveyed over 130 energy economics studies, identified SVM as one of the most widely used tools for forecasting energy prices. This highlights the relevance and robustness of SVM in handling volatile, high-dimensional data environments such as the stock market.

3.2.6 Features for SVM Classification

Two derived features were created to enhance the model's predictive ability:

- Stock Price Volatility: This is an average over the past *n* days of percent change in the given stock's price per day.

$$\frac{\sum_{i=t-n+1}^{t} \frac{C_i - C_{i-1}}{C_{i-1}}}{n}$$

- Stock Momentum: This is an average of the given stock's momentum over the past *n* days. Each day is labelled 1 if closing price that day is higher than the day before, and -1 if the price is lower than the day before.

$$\frac{\sum_{i=t-n+1}^{t} y}{n}$$

3.2.7 Mathematical Formulation of SVM

If we have the below dataset, x_i is the feature for i-th sample and y_{i} is the label. The Label is between -1 and +1, also n is the number of samples in training dataset.

$$(x_1, y_1), (x_2, y_2), \ldots, (x_{\mathrm{n}}, y_{\mathrm{n}})$$

$$y_{\mathrm{i}} \in \{-1, +1\}$$

If the hyperplane (decision boundary) is defined as below, it creates the separating line between the two categories.

$$W.X + b = 0$$

In the above equation, W is the weight vector, X is the data point, and b is the hyperplane bias.

SVM classifies each data point on the one side of the hyperplane with a margin from the line.

$$y_{\mathrm{i}}\,(w \cdot x_{\mathrm{i}} + b) \geq 1$$

$$i \in \{1, n\}$$

The distance between hyperplane and the closest data points is called margin, and SVM's goal is to maximize the margin.

SVM uses below formula in prediction step to classify the data points. If $f(x)$ > 0, the model returns 1, and if $f(x)$ < 0, the model returns -1.

$$f(x) = sign(w \cdot x + b)$$

3.2.8 SVM Implementation

- Programming Environment: Python (scikit-learn, pandas, matplotlib, NumPy)
- Objective: Predict next day's stock price movement (up/down)
- Data Split: 70% training, 30% testing (stratified sampling)
- Evaluation Metrics:
    - Accuracy
    - Precision
    - Recall
    - F1 Score
    - Confusion Matrix
- Hyperparameter Tuning: Performed via grid search with cross-validation (parameters: kernel, C, gamma)

3.2.9 Code Implementation

The SVM model was implemented in Python using libraries such as pandas for data handling, NumPy for numerical operations, and scikit-learn for the machine learning pipeline. The workflow involved the following steps:

- Data Import: Transformed CSV files for each company were loaded individually. For illustration, the model shown in the code uses Pilbara Minerals' dataset.
- Feature Selection: The two key features used were stock_momentum and stock_volatility, which had been pre-calculated during feature engineering.
- Training and Testing Split: The dataset was manually sliced into training and testing sets. Specifically, the first 9 records were held out for testing, while the remaining data was used for training the model.
- Model Training: A linear SVM (LinearSVC) was trained with a high maximum iteration limit (max_iter=10000) to ensure convergence.
- Model Evaluation: Although not shown in this excerpt, the model's accuracy and classification metrics were later evaluated using confusion matrix and other indicators.

This approach allowed for a streamlined training process using two engineered features and served as a proof of concept for extending the model to multiple companies within the lithium sector.

- Programming Environment: Python (scikit-learn, pandas, matplotlib, NumPy)
- Objective: Predict next day's stock price movement (up/down)
- Data Split: 70% training, 30% testing (stratified sampling) to reduce the chance of overfitting, providing a reliable validation and balance the complexity of the model and capture the necessary patterns.
- Evaluation Metrics:
    - Accuracy
    - Precision
    - Recall
    - F1 Score
    - Confusion Matrix
- Hyperparameter Tuning: Performed via grid search with cross-validation (parameters: kernel, C, gamma)

Figure 3.3 displays a screenshot of the Python code implementation used in this study. The script demonstrates the use of pandas to load pre-processed CSV files, select the relevant engineered features (stock_momentum and stock_volatility), split the dataset, and apply a linear SVM model using the LinearSVC class from scikit-learn. This practical code was initially developed for Pilbara Minerals and subsequently adapted for other companies in the study.

```python
# pip install cvxopt
import numpy as np
import cvxopt
from sklearn.datasets import make_blobs
#from sklearn.datasets.samples_generator import make_blobs
from sklearn.model_selection import train_test_split
from matplotlib import pyplot as plt
from sklearn.svm import LinearSVC
from sklearn.metrics import confusion_matrix


#-------------------------Import Data--------------------------
import pandas as pd

df = pd.read_csv (r'C:\Data\Vulcan_TransformedData.csv')
#df = pd.read_csv (r'C:\Data\LiontownResources_TransformedData.csv')
#df = pd.read_csv (r'C:\Data\MineralResources_TransformedData.csv')
#df = pd.read_csv (r'C:\Data\OrocobreLimited_TransformedData.csv')
#df = pd.read_csv (r'C:\Data\PiedmontLithium_TransformedData.csv')
#df = pd.read_csv (r'C:\Data\PilbaraMinerals_TransformedData.csv')
#df = pd.read_csv (r'C:\Data\Wesfarmers_TransformedData.csv')

#--------------------------------------------------------------


print (df)
#df.shape[0]

yy_train = df['y']
XX_train = df[['stock_momentum', 'stock_volatility']]

yy_train = df['y']
#yy_train2 = yy_train.loc[9:152]
#XX_train2 = XX_train.loc[9:152, ['stock_momentum', 'stock_volatility']]

yy_train2 = yy_train.loc[9:df.shape[0]]
XX_train2 = XX_train.loc[9:df.shape[0], ['stock_momentum', 'stock_volatility']]

print (XX_train2)
XX_train2.shape
#XX_train2[0:142]


XX_train3 = XX_train2.to_numpy()
yy_train3 = yy_train2.to_numpy()


svc = LinearSVC(max_iter=10000)
svc.fit(XX_train3, yy_train3)
```

Figure 3.3 Screenshot of Python implementation for SVM training using LinearSVC, showing data import, feature selection, and model fitting steps.

```python
print(svc.fit(XX_train3,yy_train3).coef_)
print(svc.fit(XX_train3,yy_train3).intercept_)

plt.scatter(XX_train3[:, 0], XX_train3[:, 1], c=yy_train3, cmap='winter')


ax = plt.gca()
xlim = ax.get_xlim()
w = svc.coef_[0]
a = -w[0] / w[1]
xx = np.linspace(xlim[0], xlim[1])
yy = a * xx - svc.intercept_[0] / w[1]
plt.plot(xx, yy)
yy = a * xx - (svc.intercept_[0] - 1) / w[1]
plt.plot(xx, yy, 'k--')
yy = a * xx - (svc.intercept_[0] + 1) / w[1]
plt.plot(xx, yy, 'k--')

plt.title("Vulcan")
plt.xlabel("momentum")
plt.ylabel("volatility (%)")
```

Figure 3.4 Screenshot of Python implementation for SVM training using LinearSVC, showing plotting steps.

3.2.10 Output and Expected Contributions

The SVM model generates daily directional movements for the stock prices of selected lithium-related companies. This analysis can:

- Help investors make more informed trading decisions
- Support corporate managers in aligning strategic actions with market direction
- Provide policymakers with insights into how market trends relate to regulatory or technological developments

Additionally, the study offers a proof of concept for applying AI techniques to financial forecasting within resource-intensive sectors, which are often affected by external and geopolitical shocks.

3.3 Research Question 2: Impact of Traditional News on Stock Market Trends

Research Question 2: How and to what extent do traditional news data impact the trend of lithium-ion battery-related stock market activity in Australia?

To investigate this research question, this study explores the relationship between traditional news media and stock price fluctuations among lithium-related companies in Australia. In an era of rapid information dissemination, news coverage can significantly influence investor behaviour and market dynamics. Understanding the extent to which news content affects stock performance in the lithium sector is vital for investors, policymakers, and analysts aiming to interpret short-term market responses and anticipate directional movements.

A quantitative approach was employed by collecting and analysing traditional news data alongside historical stock price data for 40 publicly listed lithium-related companies in Australia. The news articles were categorized based on content themes such as policy updates, environmental impact, corporate performance, and market developments. These categories were then mapped to corresponding stock price movements on the following trading day to identify potential correlations and predictive patterns.

To assess the influence of news content on stock trends, two predictive models were developed and compared: Extreme Gradient Boosting (XGBoost) and Linear Regression. XGBoost is a robust ensemble learning algorithm known for its high accuracy and performance in complex, non-linear datasets. It is particularly effective for feature-rich, high-dimensional problems like sentiment or category-based news analysis. Linear Regression was also employed to offer a benchmark comparison and to assess the relationship under a simpler, interpretable framework.

The methodology included several stages: data collection and preprocessing, news classification and feature extraction, model development, and performance evaluation. Stock price movement was framed as a regression problem, aiming to determine whether certain categories or volumes of news articles had a measurable influence on price changes.

Figure 3.5 illustrates the end-to-end workflow of this phase of the study, including the integration of news data with stock price records and the modelling pipeline.

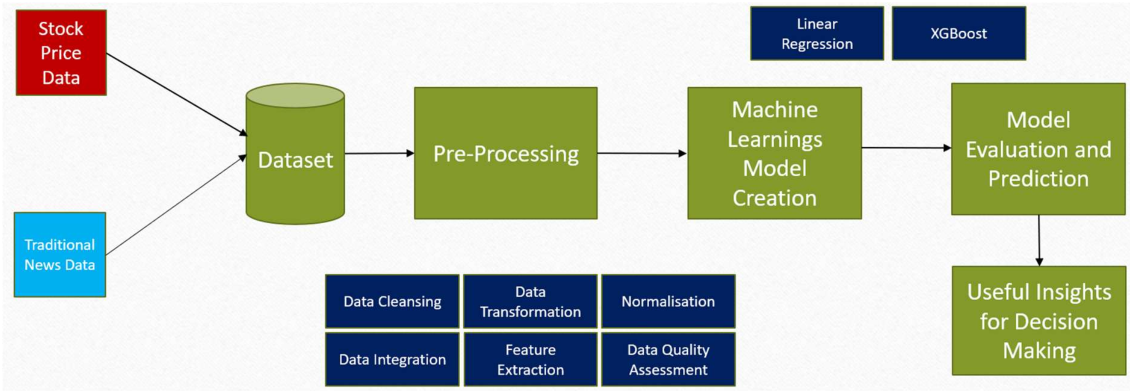


Figure 3.5 Workflow of News-Based Stock Market Prediction Using XGBoost and Linear Regression

3.3.1 Data Collection

Stock market closing price time series for 40 lithium-producing companies, from May 2020 to September 2022, were extracted from reliable websites such as www.marketindex.com.au and www.commsec.com.au. *Figure 3.6* illustrates their stock symbols and geographical distribution, revealing that nearly 80% of these companies operate in Western Australia.

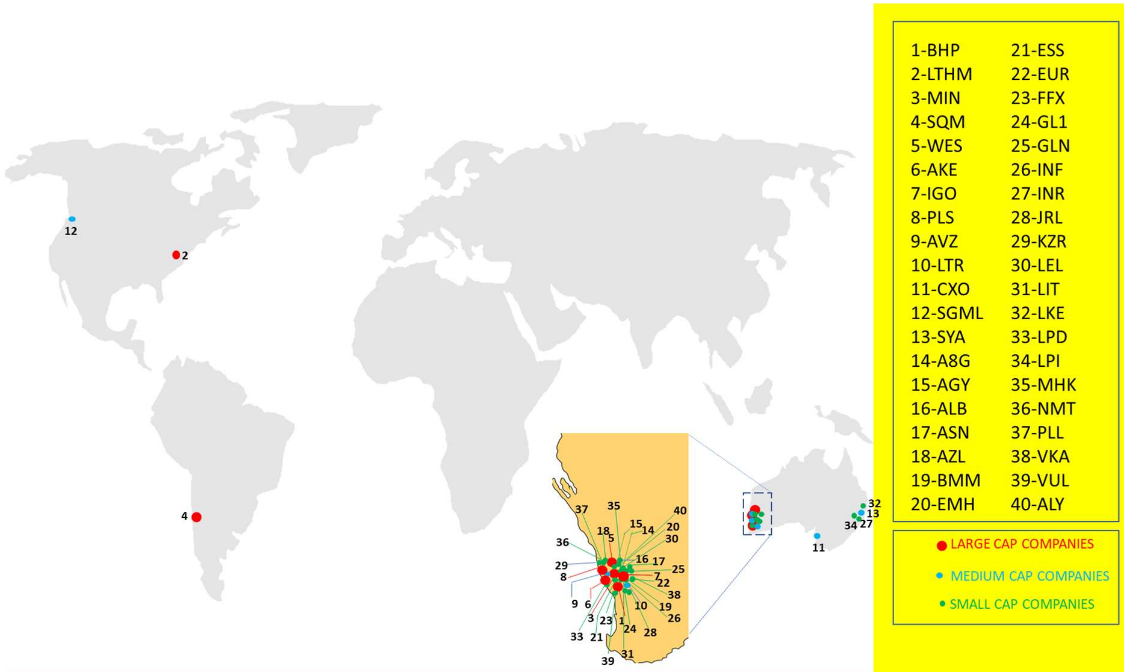


Figure 3.6 Geographical distribution of lithium-producing companies under study

Information about these companies, specifically their announcements, was extracted from their websites and stored in tabular form. Additional macroeconomic and contextual data were obtained from www.tradingeconomics.com for the same period.

The timeframe from May 2020 to September 2022 was chosen strategically. This period includes a significant surge in lithium carbonate prices, primarily driven by rising global demand in the electric vehicle (EV) industry. Analysing stock behaviour during this dynamic phase provides insight into how macroeconomic and industry-specific events influenced investor reactions and company performance.

This selection also ensures recency and relevance, with sufficient data points for robust statistical modelling. The large volume of announcements and market fluctuations during this timeframe enhances the generalizability and reliability of the model outputs. Moreover, this period aligns well with the literature, improving comparability across studies and contextual validity (Kianrad et al., 2024).

3.3.2 Feature Extraction and Variable Definition

Company announcements were categorized as quoted securities, unquoted securities, market announcements, company reports, positive financial announcements, and positive technical announcements. These were selected as explanatory variables affecting each company's average daily share price. For each trading day, a binary value of 1 was assigned if a given announcement type was released; otherwise, it was recorded as 0. This binary encoding allowed these announcements to serve as machine learning features.

The dependent variable was the daily **closing stock price**, not stock return. This decision aligns with findings by Firouz et al. (2021), who demonstrated that machine learning models tend to perform better when applied to price levels rather than returns, which are typically more volatile and less predictable due to their non-stationary nature.

To improve granularity, companies were also segmented by market capitalisation using ASX classifications:

- **Small-cap**: < AUD 2 billion (27 companies)
- **Medium-cap**: AUD 2–5 billion (5 companies)
- **Large-cap**: > AUD 5 billion (8 companies)

3.3.3 Features Defined in the Model

- **Feature A** – *Quoted Securities*: Includes applications for the quotation of securities and the release of escrowed shares.
- **Feature B** – *Unquoted Securities*: Involves the issuance of restricted shares (e.g., under employee incentive schemes) not traded on public markets.
- **Feature C** – *Market Announcements*: Covers events such as trading halts, index announcements, director notices, and merger updates.
- **Feature D** – *Company Reports*: Encompasses quarterly, half-yearly, and annual shareholder reports.
- **Feature E** – *Corporate Presentations*: Includes AGMs, investor webinars, feasibility studies, and related presentations.

- **Feature F** – *Financial Announcements*: Contains updates on funding, capital raises, agreements, grants, and contract news.
- **Feature G** – *Technical Announcements*: Relates to operational milestones like increased production, license approvals, or technical studies.

Model Feature	Short Definition	Detailed Definition
A	Quoted securities	Application for quotation of securities, release of securities from escrow, the number of securities to be quoted, issue of placement shares, cleansing notice
B	Unquoted securities	The issue, conversion, or payment up of unquoted equity securities
C	ASX Market announcements	Notification of cessation of securities, director's interest notice, response to ASX query letter, results of the announcement to the market, commencing trading on another market or stock exchange, trading halt, indices announcements, dispatch of the scheme booklet, updates on the proposed merger, disclosure document
D	Quarterly, half-yearly, and annual reports	Quarterly activities report, quarterly cashflow report, publishing half-yearly and yearly reports and accounts, annual reports to shareholders, financial reports
E	Meetings and presentations	Annual general meetings, investor and corporate presentations, webinars, presentations at the conference, the result of meetings, feasibility study presentations
F	Positive financial announcements	Raising placement, raising funds, product pricing, applying for grant funding, awarding funding, raising funds to accelerate lithium exploration, receiving firm commitments for investment via share placement, entering into binding agreements, declaration of partnerships, bonus option offers, completion of acquisitions, export shipment, placement, bonus option issue, exercise of options, purchase plan
G	Positive technical announcements	Increasing capacity, discovering high-grade Li, progress updates, commencing drilling, operational updates, identifying further lithium mineralization, exploration work updates, collection of first brine sample,

		granting of exploration and mining licenses, Environmental Impact Assessment approvals, completion of resource definition-focused drilling, commencement of site work, results of Li-battery tests, feasibility study results, metallurgical test work, survey works
P_{t-1}	Price First Lag	The closing stock price belonging to one day prior to the current day

Table 3.2 Overview of the Model Features

3.3.4 XGBoost: Mathematical Model and Explanation

Gradient Boosting (GB) is a widely used machine learning algorithm for analysing structured datasets. It operates as an ensemble method, where multiple 'weak learners'—typically decision trees—are sequentially trained, with each subsequent model correcting the errors of the previous ones. This is accomplished via gradient descent optimization. GB handles missing values, outliers, and categorical features well and can model non-linear relationships between features and target variables.

XGBoost was chosen for its ability to handle small to medium-sized structured datasets efficiently. Its capacity to manage noisy data, provide feature importance, and capture non-linear relationships made it ideal for analysing the impact of news categories and tweet-level interactions on stock price movement.

Karaca and Kozat (2025) introduced a gradient boosting model to combine capture the linear and non-linear patterns in solving the regression problems. Their results showed that their proposed soft gradient boosting approach performs better compared to boosting models.

Extreme Gradient Boosting (XGBoost) is an enhanced implementation of gradient boosting, introduced by Chen and Guestrin (2016). It builds upon the core principles of boosting by applying additive training strategies and regularization. As a highly scalable, supervised machine learning algorithm, XGBoost can be used for both classification and regression tasks and is well-suited for large datasets.

$$Objective\ (\theta) = L(\theta) + R(\theta)$$

Loss function, L(θ), shows the difference between the actual values and predictions. The purpose of Regularization function is to prevent the model from overfitting.

If we have a data set that has n training examples and m features,

$$D = \{(x_i\ , y_i)\}\ (i = 1, \ldots, n)$$

we can have an ensemble tree that uses K functions (K is the number of trees and f_k is the prediction of k[th] tree) to predict the labels.

$$\hat{y}_i = \emptyset(x_i) = \sum_{k=1}^{K} f_k(x_i), f_k \in F$$

The best combination of Loss and Regularization functions needed to minimize the Loss and avoid overfitting the model.

$$\sum_i L(y_i, \hat{y}_i) + \sum_k R(f_k)$$

In the below formula for Regularization function, T is number of leaves in the tree and W is the weight of the tree, parameter ϒ controls the number of leaves, and parameter λ controls the weight size:

$$R(f_k) = \Upsilon T + \frac{1}{2} \lambda \sum_{j=1}^{T} (W_j)^2 \;,\;\; L(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$$

3.3.5 Linear Regression: Mathematical Model and Explanation

Linear Regression models the relationship between a dependent variable (target) and one or more independent variables (features) using a linear equation:

$$y = A0 + A1x + E$$

In above equation, Y is dependent variable (stock price), x is dependent feature, A0 is intercept, A1 is coefficient (slope) and E is the error.

This model is useful for interpreting direct and proportional effects of features on the target variable. Although simpler than XGBoost, Linear Regression provides a useful benchmark and interpretability advantage.

3.3.6 Model Implementation and Accuracy Evaluation

The XGBoost model was implemented in Python using the xgboost and pandas libraries. The process involved loading structured news and price lag data from Excel, converting the dataset into a DMatrix format for efficient computation, defining model parameters, and using cross-validation to minimize the RMSE (root mean squared error). Finally, the model was trained and key insights such as feature importance and decision tree structures were visualised.

Figure 3.7 displays the Python code implementation for the XGBoost model used in this study. The script demonstrates the complete modelling workflow including data loading, feature assignment, parameter tuning, cross-validation, training, and graphical analysis.

```
#--------------------------------------------------
import xgboost as xgb
from sklearn.metrics import mean_squared_error
import pandas as pd
import numpy as np


# -- AGY
df = pd.read_excel(r'C:\_MyFiles\_PhD\Modelling_LithiumPrice_NewsShock\Data2\AGY.xlsx',sheet_name='Sheet1')
#print(df)
df.head()

#X = df[['X1', 'X2', 'X3','price_firstlag']]
X = df[['A', 'B', 'C', 'D', 'E', 'F', 'G' ,'Price_FirstLag']]
Y = df[['Price']]

data = X
#data.columns = boston.feature_names
data['PRICE'] = Y

data.describe()

XX, yy = data.iloc[:,:-1],data.iloc[:,-1]
data_dmatrix = xgb.DMatrix(data=XX,label=yy)

params = {"objective":"reg:linear",'colsample_bytree': 0.5,'learning_rate': 0.9,
                'max_depth': 5, 'alpha': 1}

cv_results = xgb.cv(dtrain=data_dmatrix, params=params, nfold=3,
                    num_boost_round=50,early_stopping_rounds=10,metrics="rmse", as_pandas=True, seed=123)

cv_results.head()

print((cv_results["test-rmse-mean"]).tail(1))
xg_reg = xgb.train(params=params, dtrain=data_dmatrix, num_boost_round=10)

import matplotlib.pyplot as plt

xgb.plot_tree(xg_reg,num_trees=0)
#plt.rcParams['figure.figsize'] = [50, 10]
plt.rcParams['figure.figsize'] = [10, 10]
plt.show()

xgb.plot_importance(xg_reg)
plt.rcParams['figure.figsize'] = [5, 5]
#plt.rcParams['figure.figsize'] = [1, 1]
plt.show()

#---------------
```

Figure 3.7 Python implementation of XGBoost regression model, including data loading, parameter tuning, cross-validation, and feature importance visualisation (AGY company).

```python
#---------------

# -- VUL
df = pd.read_excel(r'C:\_MyFiles\_PhD\Paper 2\Modelling_LithiumPrice_NewsShock\Data\VUL.xlsx',sheet_name='Sheet1')
#print(df)
df.head()

#X = df[['X1', 'X2', 'X3','price_firstlag']]
X = df[['A', 'B', 'C', 'D', 'E', 'F', 'G' ,'Price_FirstLag']]
Y = df[['Price']]

data = X
#data.columns = boston.feature_names
data['PRICE'] = Y

data.describe()

XX, yy = data.iloc[:,:-1],data.iloc[:,-1]
data_dmatrix = xgb.DMatrix(data=XX,label=yy)

params = {"objective":"reg:linear",'colsample_bytree': 0.5,'learning_rate': 0.9,
                'max_depth': 5, 'alpha': 1}

cv_results = xgb.cv(dtrain=data_dmatrix, params=params, nfold=3,
                    num_boost_round=50,early_stopping_rounds=10,metrics="rmse", as_pandas=True, seed=123)

cv_results.head()

print((cv_results["test-rmse-mean"]).tail(1))
xg_reg = xgb.train(params=params, dtrain=data_dmatrix, num_boost_round=10)

import matplotlib.pyplot as plt

xgb.plot_tree(xg_reg,num_trees=0)
#plt.rcParams['figure.figsize'] = [50, 10]
plt.rcParams['figure.figsize'] = [10, 10]
plt.show()

xgb.plot_importance(xg_reg)
plt.rcParams['figure.figsize'] = [5, 5]
#plt.rcParams['figure.figsize'] = [1, 1]
plt.show()

#---------------
```

Figure 3.8 Python implementation of XGBoost regression model, including data loading, parameter tuning, cross-validation, and feature importance visualisation (VUL company).

```python
import pandas as pd
import numpy as np

#-------------------------Import Data--------------------------

#df = pd.read_excel(r'C:\_MyFiles\_PhD\Paper 2\Modelling_LithiumPrice_NewsShock\Data\VUL.xlsx',sheet_name='Sheet1')
df = pd.read_excel(r'C:\_MyFiles\_PhD\Modelling_LithiumPrice_NewsShock\Data2\A8G.xlsx',sheet_name='Sheet1')


print(df)
df.head()

#X = df[['X1', 'X2', 'X3','price_firstlag']]
X = df[['A', 'B', 'C', 'D', 'E', 'F', 'G' ,'Price_FirstLag']]
Y = df[['Price']]

print(Y)
print(X)

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.4, random_state=101)

from sklearn.linear_model import LinearRegression
#Defining the model
model = LinearRegression()
#fitting the model or estimating the parameters of the model
model.fit(X_train,y_train)
print(model.intercept_)
print(model.coef_)

#model.get_params(True)

#predicting the price using X_test and the previous estimated parameters
predictions = model.predict(X_test)
print(predictions)
#sns.regplot(y_test,predictions)
#print(y_train)
print(y_test)
import seaborn as sns
sns.regplot(y_test,predictions)


#X = df[['Price_FirstLag']]

#from sklearn.model_selection import train_test_split
#X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.4, random_state=101)

import statsmodels.api as sm
#logit_model=sm.Logit(y_train,X_train)
Linear_model=sm.OLS(y_train,X_train)
#result=logit_model.fit()
result=Linear_model.fit()
print(result.summary())
```

Figure 3.9 Python implementation of Linear Regression model, including data loading, parameter tuning, prediction, and visualisation.

3.4 Research Question 3: Impact of Social Media Data on Lithium-Related Stock Market Trends

To address Research Question 3, this study combines historical stock price data with sentiment features derived from X (formerly Twitter) posts of five Australian lithium-related companies (AGY, 17 tweets, BMM, 27 tweets, Mineral 32 tweets, Vulcan, 70 tweets, and Wesfarmers, 15 tweets). The XGBoost machine learning model was employed to estimate the stock price on the next trading day using tweet-based features and stock price lag.

Figure 3.10 presents the conceptual model and workflow for integrating social media sentiment and stock market data.

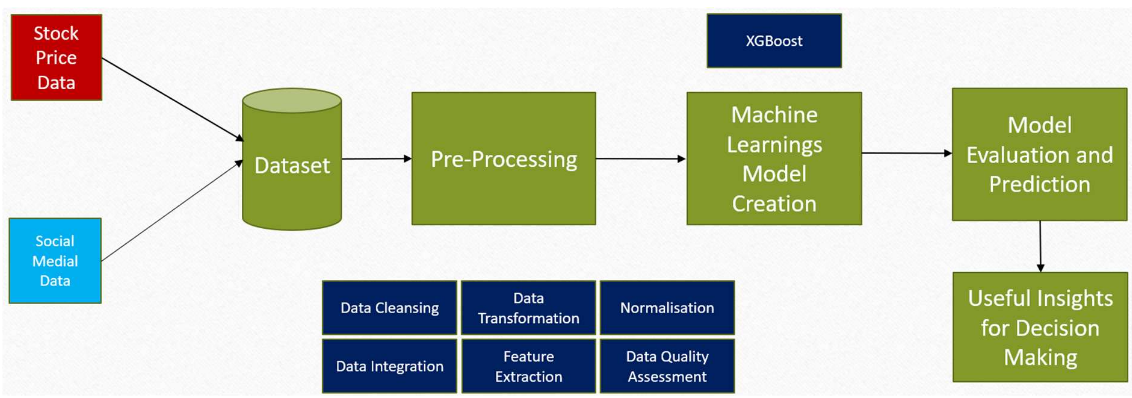


Figure 3.10 Conceptual Model for Stock Market Data Analytics

3.4.1 Data Definition

VADER (Valence Aware Dictionary and sEntiment Reasoner), a rule-based sentiment analysis tool designed for social media, was applied to label tweets as positive (1), neutral (0), or negative (-1). This approach is suitable for short and informal text such as tweets (Hutto & Gilbert, 2014).

The stock market closing prices were collected from www.commsec.com.au for the period of July 2023 to July 2024. Tweets were also collected for the same companies and period, and after preprocessing (data cleansing, removing the punctuations, …) each tweet was paired with interaction metrics such as likes, reshares, views, and comments. These features, along with the stock price from the previous trading day, were used as input variables in the XGBoost model.

Tweet's Date	Number of Likes	Number of Views	Number of Reshare	Number of Comments	Stock Price	Price Previous Trading Day	Price Next Trading Day	VADER
2024-05-02	1	247	0	0	66.48	66.01	68.31	0
2024-05-02	1	255	0	1	66.48	66.01	68.31	1
2024-05-02	1	172	0	0	66.48	66.01	68.31	0
2024-02-29	1	294	0	0	66.64	65.94	66.39	1
2024-02-15	4	1300	1	0	61.91	58.94	62.95	-1
2024-02-15	2	234	0	0	61.91	58.94	62.95	1
2024-02-15	0	167	0	0	61.91	58.94	62.95	-1
2024-02-15	1	191	0	0	61.91	58.94	62.95	1
2023-12-07	1	350	1	1	53.28	53.3	53.7	-1
2023-12-07	1	191	0	1	53.28	53.3	53.7	1
2023-10-26	4	366	0	2	50.37	50.82	50.33	1
2023-10-26	2	349	0	1	50.37	50.82	50.33	0
2023-10-26	2	249	0	1	50.37	50.82	50.33	1
2023-10-26	3	402	0	0	50.37	50.82	50.33	1
2023-08-25	4	842	1	10	51	49.42	52.1	-1

Table 3.3 Features

The objective of this analysis was not to achieve predictive accuracy but to explore the relationship between tweet interactions and sentiment and their influence on stock price movement.

In the next step, I used a XGBoost machine learning model with post's features (likes, reshares, views, comments, and sentiment label) and stock price of the previous trading day and investigated the impact of the features on the stock price.

3.4.2 Sentiment Analysis Methods Overview

Here is the overview of some sentiment analysis methods:

1. **VADER (Valence Aware Dictionary and sEntiment Reasoner)**

VADER is a lexicon-based, rule-based model introduced by Hutto and Gilbert (2014). It creates sentiment lexicons and assigns intensity scores to words or expressions. These scores are used to classify texts as positive, neutral, or negative. VADER is especially effective for analysing informal language, slang, abbreviations, punctuation, and emojis—making it particularly suitable for social media content like tweets.

VADER is a fast, lightweight model that does not require large computational resources or large training datasets. Its ability to handle short, informal texts gives it an edge over traditional models that struggle with brevity and informal phrasing. In this research, VADER was applied to label tweets of lithium-related companies, and the resulting sentiment labels (positive, neutral, or negative) were used as input features in the XGBoost model.

2. **TextBlob**

Introduced by Loria (2018), TextBlob is another lexicon-based NLP tool that classifies sentiment into positive, neutral, and negative. It also calculates polarity (ranging from -1 to 1) and subjectivity (0 to 1) scores. TextBlob is simple to use and works well for general sentiment analysis tasks such as product reviews or educational content. However, its performance is limited when handling informal language and slang.

3. **Afinn**

Afinn, developed by Nielsen (2011), is a lexicon-based sentiment analysis method for short texts and microblogs. Each word in a predefined lexicon is assigned a score from -5 to +5, and the total sentiment of a text is computed as the sum of its individual word scores. Afinn is fast and easy to implement but lacks deep linguistic parsing and nuance.

4. **BERT (Bidirectional Encoder Representations from Transformers)**

Introduced by Devlin et al. (2018), BERT is a transformer-based deep learning model that provides state-of-the-art performance in natural language understanding. It processes text bidirectionally, allowing it to capture context more effectively than traditional models. BERT is powerful for analysing large and complex documents, but it requires substantial computational resources and large amounts of labelled data for training.

While each method has its advantages, VADER was chosen for this study due to its effectiveness with short, informal texts and its minimal computational requirements. BERT, while accurate, was not suitable given the limited tweet data and the need for lightweight processing.

Several sentiment analysis techniques exist in the literature, each with unique strengths:

- **VADER** (Hutto et al. 2014): Lexicon and rule-based, ideal for short texts and social media content. Fast, simple, handles emojis and slang.
- **TextBlob** (Loria 2013): Lexicon-based, produces polarity and subjectivity scores. Easy to use but less effective on informal content.
- **Afinn** (Nielsen 2011): Assigns scores from -5 to +5 to words. Simple and interpretable but lacks nuanced text understanding.
- **BERT** (Devlin et al. 2018): Deep learning model for advanced sentiment analysis on large, complex datasets. High accuracy but resource-intensive.

Due to its suitability for informal, short-form content and small datasets, VADER was selected for this study.

3.4.3 Comparison of Sentiment Analysis Techniques

Method	Type	Use Cases	Advantages	Disadvantages
VADER	Rule-based Lexicon	• Social media posts • Short and informal texts	• Fast • No need for training • Works well with small datasets • No need for extensive computational resource • Specifically designed for social media posts • Able to handle emojis and slangs (has specific emoji lexicon)	• Not suitable for long texts • Not suitable for other domains
TextBlob	Rule-based Lexicon	• General short texts • Product reviews • Educational sentiment analysis	• Easy to implement and use • Can show text polarity and subjectivity	• Not great for informal texts • Less effective with slangs and emojis (no specific emoji lexicons)
Afinn	Lexicon-based	• Short texts • SMS • Text scoring	• Simple implementation • Fast • Does not need a huge computational resource	• Not great in advanced sentiment analysis and parsing
BERT	Machine Learning - based	• Long complex texts sentiment analysis • NLP • Document analysis • Nuanced text analysis	• Suitable for large datasets • Transfer learning capability • Deep context understanding and processing	• Not great with small and medium size datasets • Requires large computational resources

Table 3.4 Sentiment Analysis Techniques Comparison

3.4.4 Sentiment Labelling and Stock Prediction Implementation

To label tweets and analyse their sentiment using VADER, the NLTK library and SentimentIntensityAnalyzer class were used. Figure 3.11 shows how tweets are evaluated for sentiment and assigned a label based on their compound score.

```
# Import the necessary libraries
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# Download the VADER lexicon
nltk.download('vader_lexicon')

# Initialize the VADER sentiment analyzer
analyzer = SentimentIntensityAnalyzer()

# Example tweet
tweet = "$AGY additional tenements expand Rincon Lithium Project landholdings to 8,606Ha, prc

# Analyze the sentiment of the tweet
sentiment_scores = analyzer.polarity_scores(tweet)

# Display the sentiment scores
print("Sentiment Scores:", sentiment_scores)

# Classify the tweet based on the compound score
if sentiment_scores['compound'] >= 0.05:
    sentiment = 'positive'
elif sentiment_scores['compound'] <= -0.05:
    sentiment = 'negative'
else:
    sentiment = 'neutral'

# Output the result
print(f"Tweet: {tweet}")
print(f"Sentiment Label: {sentiment}")

#-------------------------------------------------------

# Import the necessary libraries
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# Download the VADER lexicon
nltk.download('vader_lexicon')

# Initialize the VADER sentiment analyzer
analyzer = SentimentIntensityAnalyzer()
```

Figure 3.11 VADER Sentiment Analysis Code to Label Tweets

After assigning sentiment labels, the resulting features—along with tweet interactions like likes, views, reshares, and comments—were fed into an XGBoost model to predict the next trading day's stock price. Figures 3.11 and 3.12 show the Python implementation of the machine learning pipeline, from feature engineering to prediction and visualization.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from xgboost import XGBRegressor
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt
import seaborn as sns

# Load the data from the Excel file
data = pd.read_excel(r'C:\_MyFiles\_PhD\Sentiment Analysis\Data\Companies\Wesfarmers_2023-202

data = pd.read_excel(r'C:\_MyFiles\_PhD\Sentiment Analysis\Data\Companies\Wesfarmers_2023-202


print(data)

# Feature engineering
data['Is Shared'] = data['Nubmer of Reshare'].apply(lambda x: 1 if x > 0 else 0)
data['Is Liked'] = data['Number of Likes'].apply(lambda x: 1 if x > 0 else 0)
data['Is Viewed'] = data['Number of Views'].apply(lambda x: 1 if x > 61 else 0)
data['Is Commented'] = data['Number of Comments'].apply(lambda x: 1 if x > 0 else 0)
data['Negative Sentiment'] = data['VADER'].apply(lambda x: 1 if x == -1 else 0)

# Feature and target selection
#X = data[['Is Shared', 'Is Liked', 'Is Viewed', 'Is Commented', 'Negative Sentiment', 'Stock
#X = data[['Is Shared', 'Is Liked', 'Is Commented', 'Negative Sentiment', 'Stock Price']]
X = data[['Is Shared', 'Is Liked', 'Is Commented', 'Negative Sentiment']]

Y = data['Price_NextTradingDay']

print(X)

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.4, random_state=101)

# Define and train the XGBoost model
xgb_model = XGBRegressor(objective='reg:squarederror', random_state=101)
xgb_model.fit(X_train, y_train)

# Predictions
predictions = xgb_model.predict(X_test)
```

Figure 3.12 XGBoost Regression Pipeline with Tweet Interaction Features

```python
# Example tweet
tweet = "Exciting geochemical discoveries at Gorge Lithium have illuminated the promising pot

# Analyze the sentiment of the tweet
sentiment_scores = analyzer.polarity_scores(tweet)

# Display the sentiment scores
print("Sentiment Scores:", sentiment_scores)

# Classify the tweet based on the compound score
if sentiment_scores['compound'] >= 0.50:
    sentiment = 'positive'
elif sentiment_scores['compound'] <= -0.50:
    sentiment = 'negative'
else:
    sentiment = 'neutral'

# Output the result
print(f"Tweet: {tweet}")
print(f"Sentiment Label: {sentiment}")
```

Figure 3.13 Model Evaluation and Visualization of Predictions

3.5 Chapter Summary

This chapter presented the methodological framework for analysing the stock market performance of lithium-related companies in Australia using machine learning techniques and sentiment analysis. Three research questions were addressed through progressively complex models: Support Vector Machine (SVM) for historical price-based classification, XGBoost and Linear Regression for integrating traditional news features, and XGBoost with VADER-labelled sentiment for modelling social media influence. Each model was tailored to its data source, leveraging domain-specific features such as announcement categories, tweet interactions, and sentiment scores. By combining structured financial data with unstructured text from news and social media, this chapter lays the groundwork for evaluating the predictive power and practical implications of each approach. The following chapter will focus on the empirical results, model performance comparisons, and discussion of key findings. Ethics approval for this research was obtained and is provided in Appendix A.

Chapter 4 – Findings

4.1 Research Question 1

To investigate the trend of lithium-ion batteries related stock market in Australia, I used historical stock price data of seven lithium-related companies in Australia and utilised Support Vector Machines (SVM) model to estimate the direction of the stock price of the companies on the next trading day.

4.1.1 Results and Findings

Each point in Figure 4.1 to 4.7 represents three pieces of information about each companies' stock on a specified trading day

- The x-coordinate of a point shows the average of the company stock's price gap over 9 days leading to the trading day. A positive momentum means most of these 9 days, the stock experienced an increasing-price trend.
- The y-coordinate of a point is a measure of the magnitude of changes in price over 9 days. It is possible that a stock experiences a positive momentum across 9 days but a negative overall price change.
- The colour of a point depicts the sign of the price gap. Green point means a positive gap, meaning an increasing direction in price, and blue point means a negative price change.

What I estimated is a line (or a curve or a hyperplane) that puts the data points on two sides of it. An ideally estimated line is a line that perfectly separates the blue points and green points, meaning this line estimates the direction of change in the stock price. In the next step after estimation, we decomposed our data into two subsets, called a training subset and a test dataset, and tried to forecast the direction of change in stock price in the test dataset using the explanatory variables, which are stock momentum and stock volatility, of the training dataset. The results of all companies show that the green and blue points are not fully separated by the line, whish show that SVM model could not predict the direction of the price using only historical momentum and volatility.

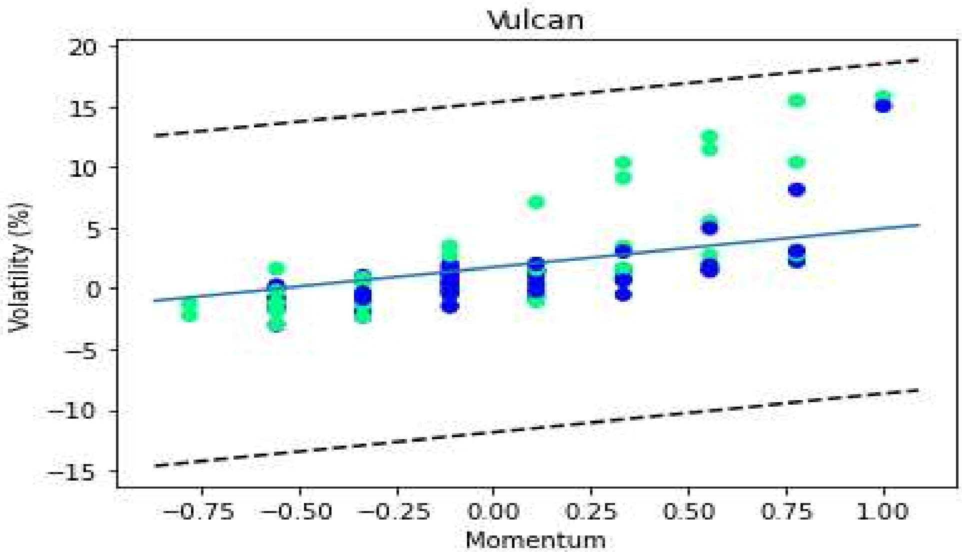


Figure 4.1 Estimated Results for Vulcan Company

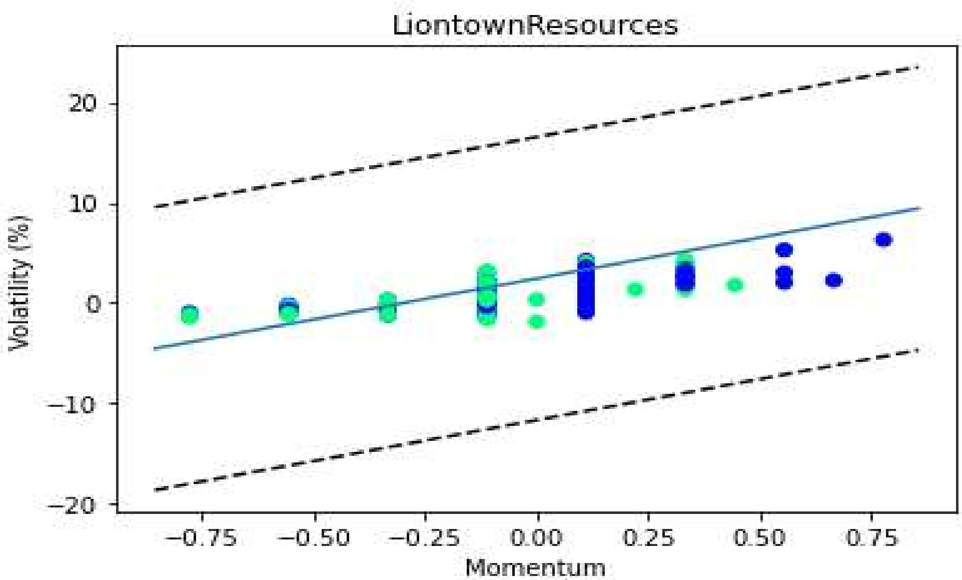


Figure 4.2 Estimated Results for Liontown Resources Company

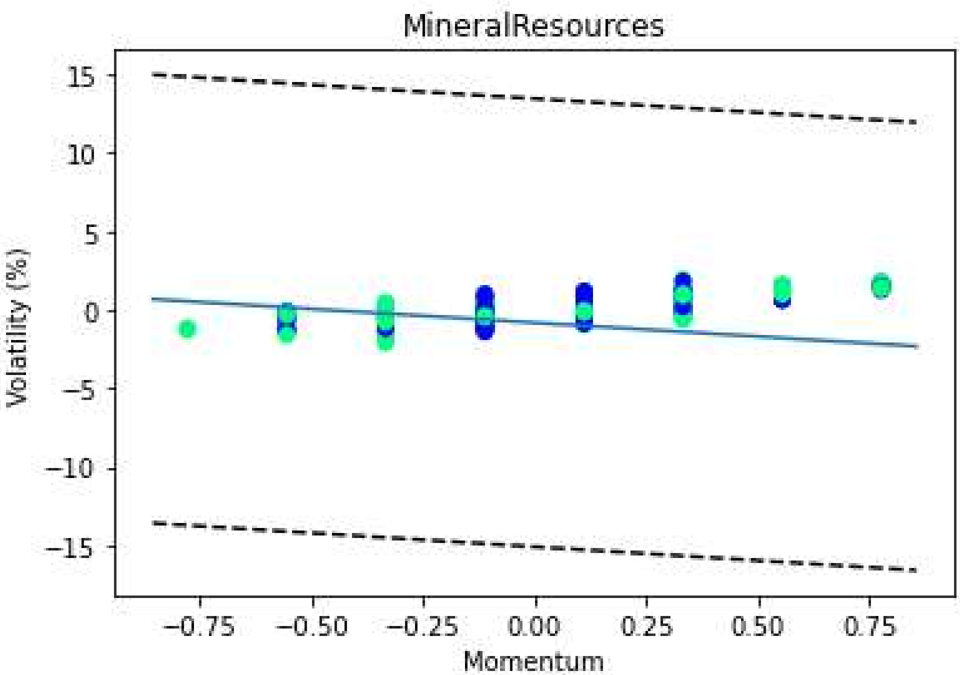


Figure 4.3 Estimated Results for Mineral Resources Company

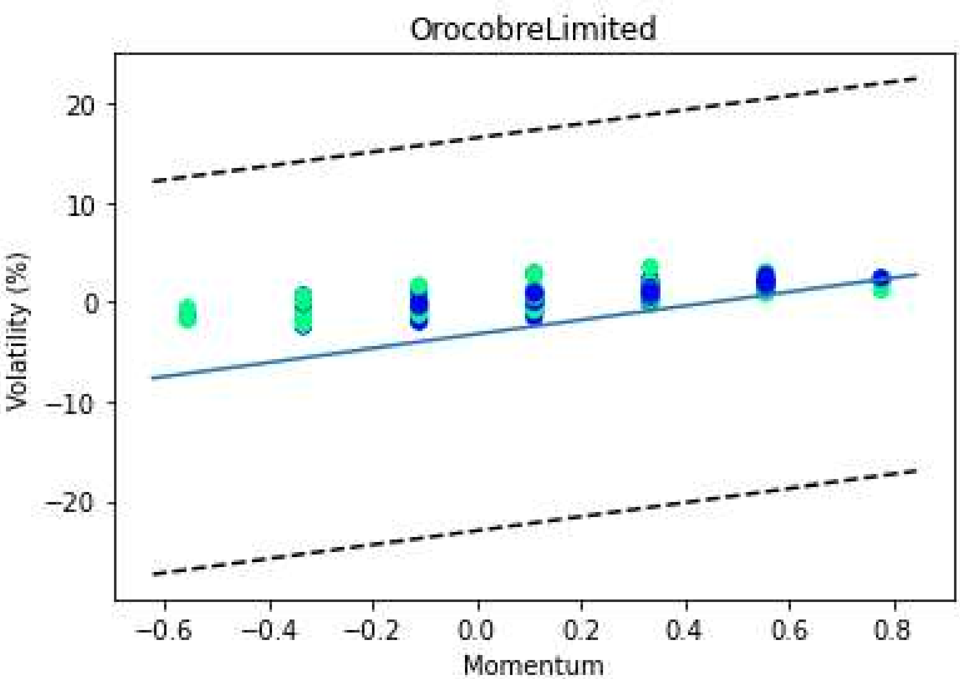


Figure 4.4 Estimated Results for Orocobre Limited Company

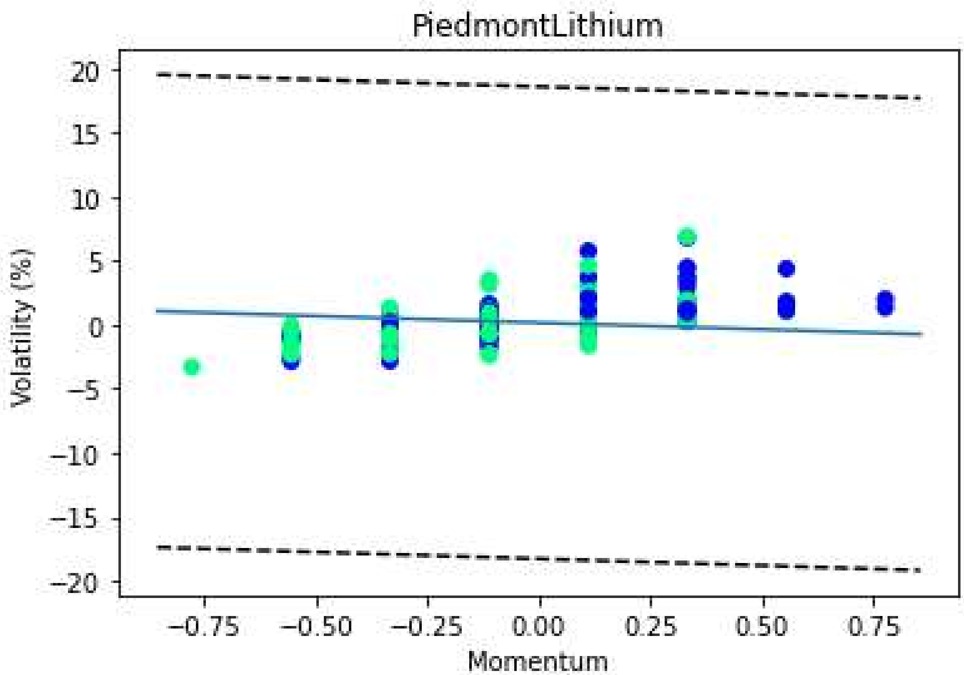


Figure 4.5 Estimated Results for Piedmont Lithium Company

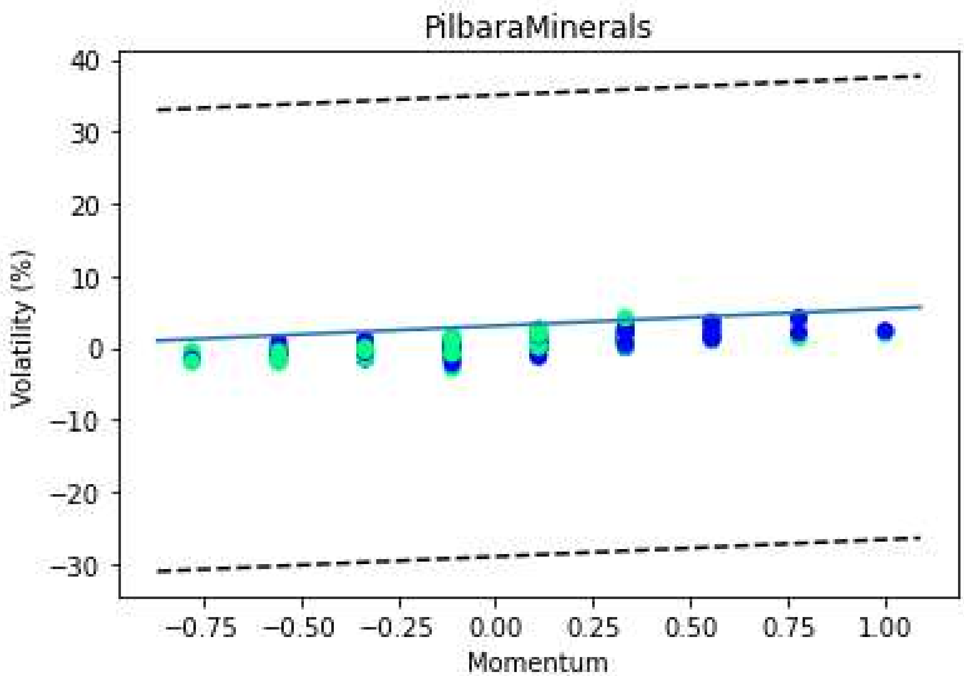


Figure 4.6 Estimated Results for Pilbara Minerals Company

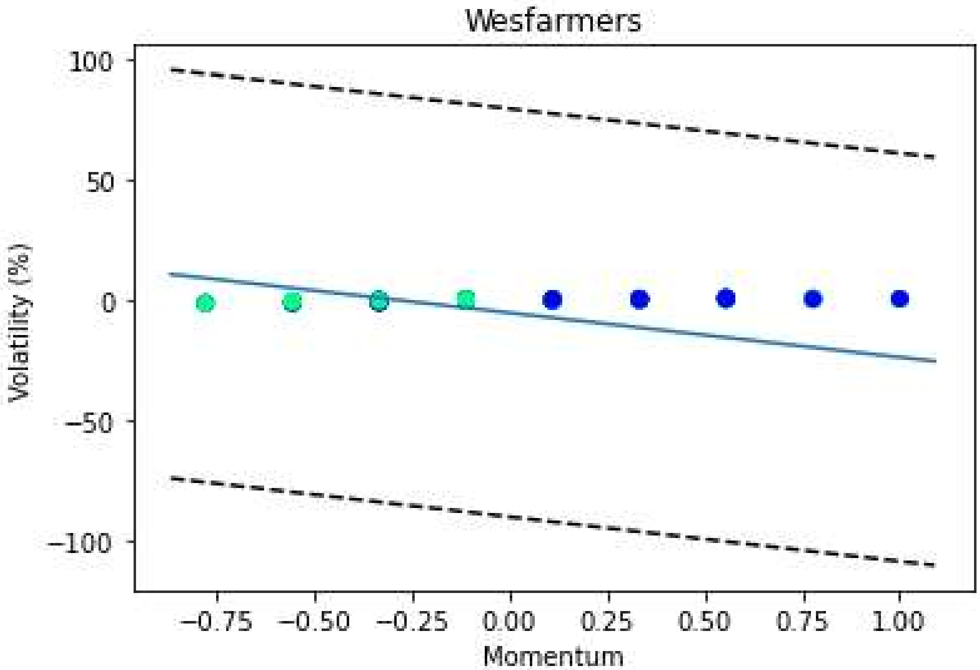


Figure 4.7 Estimated Results for Wesfarmers Company

4.2 Research Question 2

4.2.1 Research Question 2: XGBoost Results

Having a dataset of stock price data of 40 lithium-related companies and their traditional news categories, I applied the XGBoost ML algorithm on the data outlined above, then using feature importance estimation, examined the impact of features A to G on the stock market closing price of each of the 40 companies (see Table 4.1) (Kianrad et al., 2024).

	Number of Companies by Feature Importance			
	Small Cap	Mid Cap	Large Cap	Total
Feature				
A	18	3	5	26
B	18	3	4	25
C	14	3	8	25
D	16	3	7	26
E	16	5	8	29
F	16	5	7	28
G	14	4	7	25
	Number of Companies			
	27	5	8	40

Table 4.1 Number of companies by feature importance and market capitalization

As illustrated in Table 4.1, it's common for multiple features to be significant for a single company simultaneously. For instance, Feature A is important for 18 out of the 27 total small-cap companies. Similarly, Feature B is also crucial for 18 small-cap companies, indicating a notable overlap among the small-cap companies for which Features A and B are relevant. To provide a clearer understanding, the numerical data in Table 4.2 can also be expressed as percentages, as demonstrated in Table 4.2 (Kianrad et al., 2024).

	Number of Companies by Feature Importance			
	Small Cap	Mid Cap	Large Cap	Total
Feature				
A	18/27 (66.67%)	3/5 (60.00%)	5/8 (62.50%)	26/40 (65.00%)
B	18/27 (66.67%)	3/5 (60.00%)	4/8 (50.00%)	25/40 (62.50%)
C	14/27 (51.85%)	3/5 (60.00%)	8/8 (100.00%)	25/40 (62.50%)
D	16/27 (59.26%)	3/5 (60.00%)	7/8 (87.50%)	26/40 (65.00%)
E	16/27 (59.26%)	5/5 (100.00%)	8/8 (100.00%)	29/40 (72.50%)
F	16/27 (59.26%)	5/5 (100.00%)	7/8 (87.50%)	28/40 (70.00%)
G	14/27 (51.85%)	4/5 (80.00%)	7/8 (87.50%)	25/40 (62.50%)

Table 4.2 Feature Importance (%)

In Table 4.2, the percentage calculated based on the number of companies in the specific category affected by the feature divided by the total number of companies in the category.

For example, feature A was significant in determining the closing stock price of 18 small cap companies, out of 27 small cap companies studied. Feature B was significant in determining the closing stock price of 25 companies, out of 40 total companies studied, so the percentage is 15/40 = 62.50%.

In Figure 4.8, the colours of the bars denote different features (i.e., types of announcements). The y-axis represents the percentage of companies for which each feature was significant in determining their closing stock prices. The x-axis categorizes the companies based on their market capitalization. For instance, Feature F, depicted in green, is crucial for 87.5% of large-cap companies (7 out of 8 companies). However, when considering the entire sample of companies (small, mid, and large cap), Feature F is important for 70% (28 out of 40 companies) (Kianrad et al., 2024).

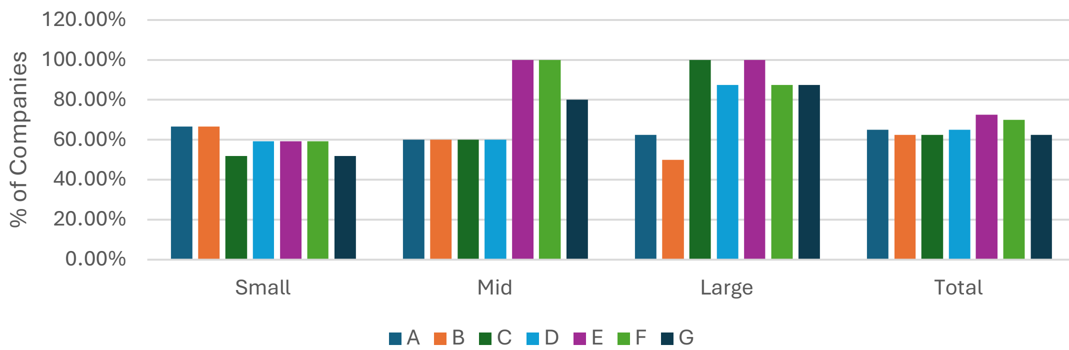


Figure 4.8 Feature Importance (%) Across Different Company Market Capitalization
A (Quoted Securities)
B (Unquoted Securities)
C (ASX Market Announcements)
D (Quarterly, Half-Yearly, and Annual Reports)
E (Meetings and Presentations)
F (Positive Financial Announcements)
G (Positive Technical Announcements)

As depicted in Figure 4.8, for small-cap companies, Features A (Quoted Securities) and B (Unquoted Securities) are the most significant factors affecting share prices, each with an importance rating of 67%. This indicates that the prices of 18 out of the 27 small-cap companies studied are influenced by Features A and B. Additionally, Features D (Quarterly, Half-Yearly, and Annual Reports), E (Meetings and Presentations), and F (Positive Financial Announcements) each have an equal impact on the share prices of small-cap companies, with an importance of 60%. Features C (ASX Market Announcements) and G (Positive Technical Announcements) have the least effect, influencing the stock prices of nearly half of the small-cap companies. Notably, positive financial announcements (Feature F) are more influential than positive technical announcements (Feature G) for small-cap companies.

In the case of medium-cap companies, as shown in Figure 4.8, Features E and F are the most impactful, with an importance rating of 100%. This suggests that both features are critical in determining the stock prices of all 5 medium-cap companies studied. Feature G follows with an 80% importance rating. Features A, B, C, and D have similar but less significant contributions compared to other factors. Similar to small-cap companies, the impact of positive financial announcements (Feature F) is greater than that of positive technical announcements (Feature G) for medium-cap companies (Kianrad et al., 2024).

In Figure 4.8, it is evident that for large-cap companies, Features C (ASX Market Announcements) and E (Meetings and Presentations) are the most crucial. All eight large-cap companies analysed are significantly affected by these factors. It's noteworthy that the impacts of Features D (Quarterly, Half-Yearly, and Annual Reports), F (Positive Financial Announcements), and G (Positive Technical Announcements) on the stock prices of these companies are similar. In contrast to small-cap companies, Features A (Quoted Securities) and B (Unquoted Securities) have the least impact on the stock prices of large-cap companies.

The real-world implications of Features A and B on the stock prices of large-cap companies can be observed in the following instances. On March 2, SQM reported its 2021 earnings, showing a substantial increase in net income. However, SQM's stock price did not experience a significant rise until May 18, coinciding with the announcement of its Q1 earnings report and the approval of an interim dividend. Similarly, Livent's announcement of increasing its stake in Nemaska Lithium and its Q1 results, followed by its collaboration news with Lilium, led to notable increases in its share price (Kianrad et al., 2024).

In a broader perspective, Feature E (Meetings and Presentations) was a key predictor of stock prices for the largest proportion of companies analysed (73%), as shown in Figure 2. Feature F (Positive Financial Announcements) was the second most significant predictor, affecting 70% of the companies (Kianrad et al., 2024).

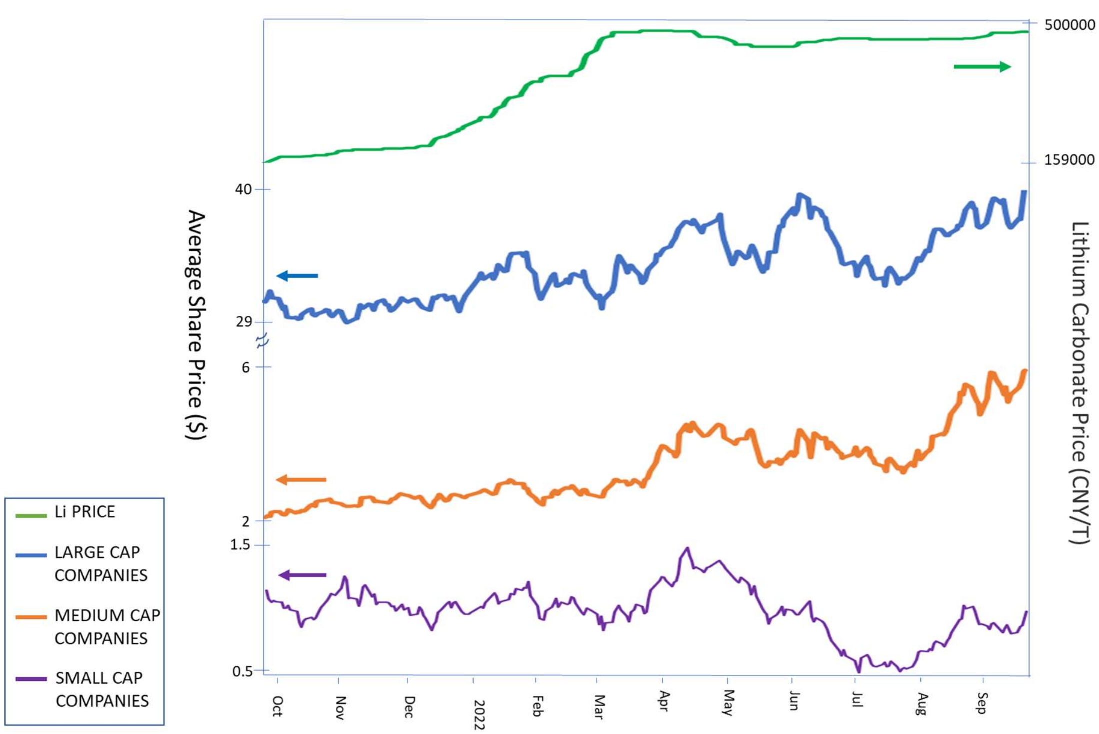


Figure 4.9 Changes in Lithium price and the average share price of companies

Figure 4.9 presents the trends in lithium prices alongside the average share price of a selected sample of 40 lithium companies from May 2020 to September 2022. During this period, particularly in 2022, the price of lithium carbonate saw a significant increase, reaching record highs. This surge was primarily driven by the escalating demand from the electric vehicle (EV) industry. In January 2022, a notable rise in lithium carbonate prices had a consequential impact on the average share prices of these companies, which peaked until April 2022 before undergoing a decline. However, despite these fluctuations, lithium prices remained elevated throughout the first half of 2022, exceeding market predictions. This was largely due to a renewed demand surge from China, contributing to an over 123% increase in lithium prices between January and September 2022 (Kianrad et al., 2024).

The IEA Critical Minerals Market Review 2023 underscores that the sales of EVs and the consumer trend towards larger vehicles (which require larger batteries) were key factors in the tripling of lithium demand between 2017 and 2022.

In 2021, despite the challenges of the COVID-19 pandemic and supply chain bottlenecks, sales of electric vehicles (EVs) nearly doubled to 6.6 million, as reported by the International Energy Agency (IEA). The momentum continued into the first quarter of 2022, with EV sales surging by 75% compared to the same period in 2021, reaching a remarkable 2 million units. This upward trend in EV sales persisted throughout the first half of 2022. The IEA now anticipates that EVs will constitute approximately 13% of the total market share, signalling a positive outlook for the lithium market, which is crucial for EV battery production.

During the third quarter, demand for lithium-battery chemistry remained strong, especially in China, contributing to the surge in lithium prices. This demand drove lithium prices to record highs in 2021, with the upward trend persisting well into 2022 (Kianrad et al., 2024).

4.2.2 Research Question 2: Linear Regression Results

Company Name	Confidence Level > 90%	Market_Cap_Category	MarketCap_Quantity
BHP	E (94%), G (93%)	Large	189.74
WES	D (100%)	Large	49.93
SQM	C (100%)	Large	24.53
PLS	F (94%)	Large	14.55
MIN		Large	13.6
IGO		Large	11.76
AKE		Large	8.97
LTHM		Large	5.61
LTR		Medium	4.09
AVZ		Medium	2.75
CXO		Medium	2.46
SGML	C (90%)	Medium	2.37
SYA	F (94%)	Medium	1.95
LKE	C (95%)	Small	1.37
INR	A (99%), B (95%)	Small	1.13
VUL	C (99%), G (99%)	Small	1.04
AGY	F(91%)	Small	0.72429
NMT	F (94%), G (90%)	Small	0.59101
ASN	B (100%), C (100%), G (100%)	Small	0.53831
PLL	F (94%)	Small	0.49064
GLN	G (100%)	Small	0.46295
GL1	B (100%), F (99%)	Small	0.3904
ALB	D (94%),	Small	0.254
FFX	A (98%), F (96%)	Small	0.23625
LPI		Small	0.23276
AZL		Small	0.19279
JRL	B (100%), C (99%), G (94%)	Small	0.13255
EUR	B (100%), C (100%), E (95%)	Small	0.12998
LPD	B (93%), C (97%), E (92%)	Small	0.12506
ESS	A (98%)	Small	0.11582
EMH	F (100%), G (100%)	Small	0.09096
INF	C (99%), G (97%)	Small	0.08751
LEL	A (100%)	Small	0.06737
LIT		Small	0.05862
ALY	D (100%), F (99%)	Small	0.03871
KZR	B (100%), F (100%), G (100%)	Small	0.02902
BMM	B (100%), F (96%)	Small	0.01183

MHK	C (93%)	Small	0.01056
A8G		Small	0.00968
VKA	B (100%), C (100%)	Small	0.0082

Table 4.3 Companies, market capitalisation categories and quantities, and confidence levels of different features

Table 4.3 shows the companies studied, their market capitalisation at the time of study, their market capitalisation category and the features that had confidence level of over 90%.

Market capitalisation category of each company category is determined by its market capitalisation. Companies with market capitalisation over 5 billion are categorised as Large, companies with market capitalisation between 2 billion and 5 billion are categorised as Medium, and companies with market capitalisation less than 2 billion are categorised as small.

For example, features E and G have confidence level of 94% and 93% for BHP company (Large), while features C and G have confidence level of 99% for VUL company (small).

Table 4.4 below shows the summary of feature importance across different company market capitalisations. As Table 4.4 shows, feature A is important in 4 out of 27 small companies, while feature B is important in 9 out of 27 small companies, and so on. The Total column shows the total amount of importance for each feature across small, medium, and large companies. Each feature can be important across different companies, so the total amount of importance of each feature can be more than 40 (number of companies studied).

> 90%	Small	Mid	Large	Total
A (Quoted Securities)	4	0	0	4
B (Unquoted Securities)	9	0	0	9
C (ASX Market Announcements)	9	1	1	11
D (Quarterly, Half-Yearly, and Annual Reports)	2	0	1	3
E (Meetings and Presentations)	2	0	1	3
F (Positive Financial Announcements)	9	1	1	11
G (Positive Technical Announcements)	8	0	1	9
Number of Companies	27	5	8	40

Table 4.4 Feature Importance Across Different Company Market Capitalisations.

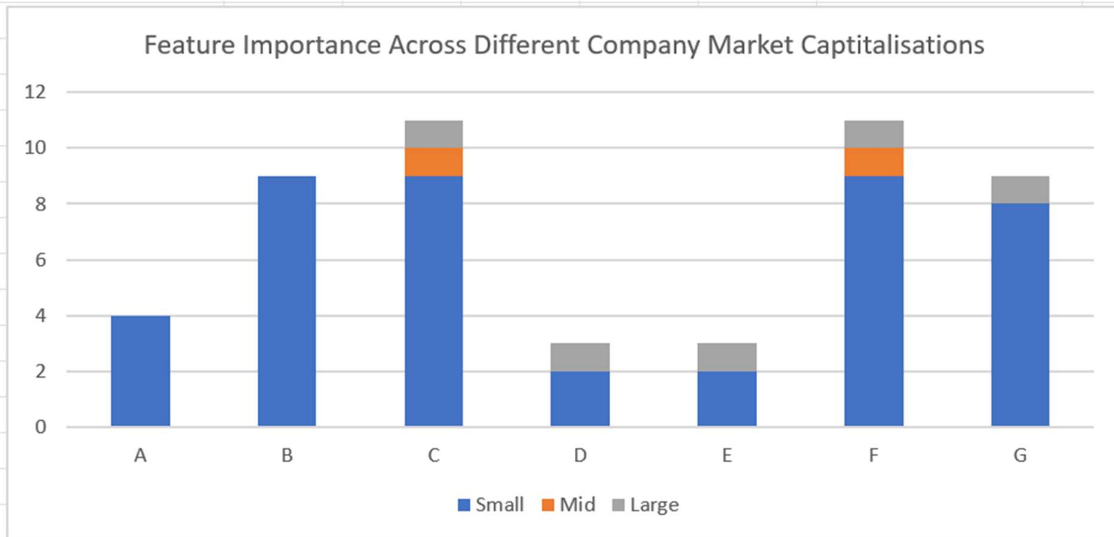


Figure 4.10 Feature Importance Across Different Company Market Capitalisations

Figure 4.10 shows the feature importance across companies with different market capitalisation. For example, feature A is important for 4 small companies, feature C is important for 9 small companies, one medium company, and one large company.

> 90%	Small	Mid	Large	Total
A	14.81%	0.00%	0.00%	10.00%
B	33.33%	0.00%	0.00%	22.50%
C	33.33%	20.00%	12.50%	27.50%
D	7.41%	0.00%	12.50%	7.50%
E	7.41%	0.00%	12.50%	7.50%
F	33.33%	20.00%	12.50%	27.50%
G	29.63%	0.00%	12.50%	22.50%
Number of Companies	27	5	8	40

Table 4.5 Feature Importance Across Different Market Capitalisations (Percentage)

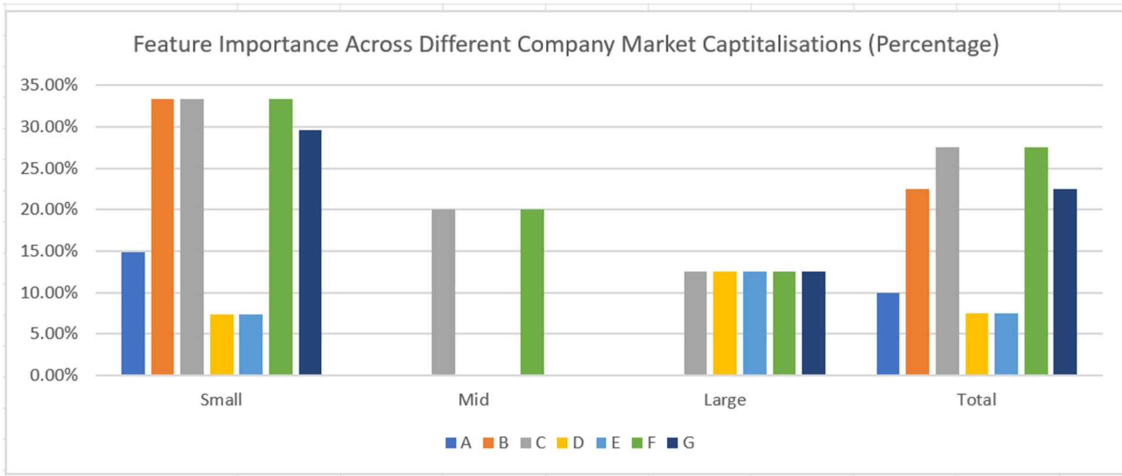


Figure 4.11 Feature Importance Across Different Market Capitalisations (Percentage)

Figure 4.11 shows the percentage of feature importance across different market capitalisation groups. For example, features C and F are important in 20% of mid-cap companies, while they are important in more than 25% (27.50%) of total companies studied. Features B, C, and F are important in 33.33% of small-cap companies, while feature B is not important for mid-cap and large-cap companies. Feature G is important in 29.63% of small-cap companies and 12.50% of large-cap companies.

In Figures 4.12 to 4.51, if P-Value (P> |t|) is less than 0.1, then the feature has a statistically significant impact on the stock price of the next trading day.

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.5260	0.698	-0.754	0.452	-1.903	0.851
B	0.2449	0.399	0.613	0.541	-0.543	1.033
C	-0.0621	0.184	-0.337	0.736	-0.426	0.301
D	-0.1500	0.298	-0.503	0.616	-0.739	0.439
E	0.6229	0.325	1.918	0.057	-0.018	1.264
F	0.1170	0.487	0.240	0.810	-0.844	1.078
G	0.4282	0.233	1.842	0.067	-0.031	0.887
Price_FirstLag	0.9973	0.002	487.407	0.000	0.993	1.001

Figure 4.12 BHP Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0	0	nan	nan	0	0
B	9.973e-16	1.67e-15	0.596	0.552	-2.3e-15	4.3e-15
C	0.2517	0.156	1.616	0.108	-0.055	0.559
D	-1.2712	0.369	-3.442	0.001	-2.000	-0.543
E	0.0704	0.270	0.261	0.794	-0.461	0.602
F	-0.1303	0.237	-0.549	0.583	-0.598	0.337
G	0.0365	0.246	0.148	0.882	-0.448	0.521
Price_FirstLag	0.9987	0.001	1001.495	0.000	0.997	1.001

Figure 4.13 WES Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0	0	nan	nan	0	0
B	-2.294e-15	8.62e-15	-0.266	0.790	-1.93e-14	1.47e-14
C	5.4535	1.928	2.829	0.005	1.644	9.263
D	1.1854	1.713	0.692	0.490	-2.201	4.572
E	-0.7168	0.818	-0.876	0.382	-2.333	0.900
F	-2.6998	2.457	-1.099	0.274	-7.555	2.156
G	-2.2938	2.221	-1.033	0.303	-6.683	2.095
Price_FirstLag	1.0008	0.003	315.427	0.000	0.995	1.007

Figure 4.14 SQM Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0068	0.026	-0.256	0.798	-0.059	0.045
B	-0.0494	0.105	-0.471	0.638	-0.256	0.157
C	0.0459	0.024	1.905	0.058	-0.002	0.093
D	0.0124	0.030	0.415	0.678	-0.046	0.071
E	0.0255	0.033	0.780	0.436	-0.039	0.090
F	-0.0511	0.027	-1.889	0.060	-0.104	0.002
G	0.0072	0.034	0.208	0.836	-0.061	0.075
Price_FirstLag	0.9998	0.003	326.740	0.000	0.994	1.006

Figure 4.15 PLS Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.1469	1.012	-0.145	0.885	-2.143	1.849
B	-1.21e-15	2.89e-15	-0.419	0.676	-6.91e-15	4.49e-15
C	0.1922	0.346	0.555	0.579	-0.490	0.875
D	-0.5281	0.590	-0.896	0.372	-1.691	0.635
E	-0.8111	0.623	-1.303	0.194	-2.039	0.417
F	-0.3064	0.446	-0.687	0.493	-1.186	0.573
G	0.7073	0.672	1.053	0.294	-0.617	2.032
Price_FirstLag	1.0004	0.003	370.550	0.000	0.995	1.006

Figure 4.16 MIN Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0309	0.199	-0.155	0.877	-0.423	0.361
B	-0.0644	0.166	-0.388	0.698	-0.391	0.262
C	0.0936	0.109	0.857	0.392	-0.122	0.309
D	-0.0109	0.112	-0.098	0.922	-0.231	0.209
E	0.0278	0.064	0.431	0.667	-0.099	0.155
F	0.0543	0.086	0.631	0.529	-0.115	0.224
G	-0.0305	0.044	-0.694	0.488	-0.117	0.056
Price_FirstLag	1.0026	0.002	485.118	0.000	0.998	1.007

Figure 4.17 IGO Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0725	0.201	0.361	0.719	-0.325	0.470
B	0.2798	0.300	0.933	0.352	-0.313	0.873
C	0.0076	0.093	0.082	0.935	-0.176	0.191
D	-0.0173	0.155	-0.111	0.912	-0.325	0.290
E	0.0576	0.169	0.340	0.734	-0.277	0.393
F	0.4937	0.327	1.511	0.133	-0.152	1.140
G	0.0210	0.241	0.087	0.931	-0.455	0.497
Price_FirstLag	1.0003	0.004	273.576	0.000	0.993	1.007

Figure 4.18 AKE Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0	0	nan	nan	0	0
B	0	0	nan	nan	0	0
C	-0.4652	1.231	-0.378	0.706	-2.899	1.968
D	3.822e-16	2.6e-16	1.472	0.143	-1.31e-16	8.95e-16
E	0.4407	0.510	0.864	0.389	-0.567	1.449
F	0	0	nan	nan	0	0
G	1.5850	1.127	1.407	0.162	-0.642	3.812
Price_FirstLag	0.9921	0.004	275.553	0.000	0.985	0.999

Figure 4.19 LHTM Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0062	0.023	0.269	0.788	-0.039	0.051
B	0	0	nan	nan	0	0
C	-0.0038	0.020	-0.191	0.849	-0.043	0.035
D	0.0274	0.032	0.863	0.389	-0.035	0.090
E	-0.0015	0.019	-0.079	0.937	-0.039	0.036
F	0.0013	0.023	0.058	0.954	-0.044	0.047
G	0.0074	0.016	0.458	0.647	-0.024	0.039
Price_FirstLag	0.9973	0.004	257.962	0.000	0.990	1.005

Figure 4.20 LTR Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0096	0.010	-0.974	0.331	-0.029	0.010
B	-0.0110	0.033	-0.332	0.740	-0.076	0.054
C	0.0038	0.007	0.586	0.559	-0.009	0.017
D	0.0004	0.010	0.036	0.972	-0.020	0.021
E	-0.0056	0.010	-0.539	0.590	-0.026	0.015
F	0.0129	0.013	1.009	0.314	-0.012	0.038
G	-7.03e-05	0.007	-0.010	0.992	-0.014	0.014
Price_FirstLag	0.9966	0.004	241.623	0.000	0.988	1.005

Figure 4.21 AVZ Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0069	0.011	0.610	0.542	-0.015	0.029
B	-0.0064	0.030	-0.211	0.833	-0.066	0.053
C	0.0160	0.012	1.347	0.179	-0.007	0.039
D	0.0097	0.017	0.558	0.577	-0.025	0.044
E	0.0112	0.013	0.833	0.406	-0.015	0.038
F	-0.0069	0.012	-0.557	0.578	-0.031	0.018
G	0.0034	0.009	0.362	0.718	-0.015	0.022
Price_FirstLag	1.0021	0.004	240.310	0.000	0.994	1.010

Figure 4.22 CXO Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0	0	nan	nan	0	0
B	0	0	nan	nan	0	0
C	1.493e-16	9.07e-17	1.646	0.102	-3e-17	3.29e-16
D	1.1613	0.778	1.493	0.138	-0.376	2.698
E	0.0770	0.306	0.251	0.802	-0.529	0.682
F	-0.1284	0.306	-0.419	0.676	-0.734	0.477
G	-0.2763	0.254	-1.086	0.279	-0.779	0.226
Price_FirstLag	1.0055	0.004	234.021	0.000	0.997	1.014

Figure 4.23 SGML Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0044	0.005	0.951	0.343	-0.005	0.013
B	0.0079	0.024	0.329	0.743	-0.039	0.055
C	-0.0068	0.005	-1.421	0.157	-0.016	0.003
D	0.0030	0.008	0.376	0.707	-0.013	0.019
E	-0.0002	0.011	-0.021	0.984	-0.022	0.021
F	0.0040	0.007	0.618	0.537	-0.009	0.017
G	-0.0048	0.004	-1.102	0.272	-0.013	0.004
Price_FirstLag	0.9779	0.010	97.053	0.000	0.958	0.998

Figure 4.24 SYA Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0036	0.016	0.231	0.818	-0.027	0.034
B	0.0298	0.027	1.090	0.277	-0.024	0.084
C	-0.0342	0.018	-1.934	0.055	-0.069	0.001
D	0.0068	0.031	0.222	0.825	-0.054	0.068
E	0.0133	0.026	0.503	0.616	-0.039	0.066
F	0.0064	0.027	0.236	0.814	-0.047	0.060
G	-0.0101	0.019	-0.528	0.598	-0.048	0.027
Price_FirstLag	0.9978	0.006	170.283	0.000	0.986	1.009

Figure 4.25 LKE Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0134	0.008	-1.631	0.104	-0.030	0.003
B	0.0305	0.015	1.975	0.050	5.36e-05	0.061
C	0.0025	0.008	0.318	0.751	-0.013	0.018
D	0.0081	0.010	0.799	0.425	-0.012	0.028
E	-0.0006	0.010	-0.061	0.951	-0.021	0.019
F	0.0102	0.009	1.177	0.240	-0.007	0.027
G	-0.0037	0.009	-0.402	0.688	-0.022	0.014
Price_FirstLag	0.9993	0.003	290.007	0.000	0.993	1.006

Figure 4.26 INR Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.1007	0.113	-0.893	0.373	-0.323	0.121
B	-0.2103	0.546	-0.385	0.700	-1.286	0.865
C	0.3261	0.125	2.603	0.010	0.079	0.573
D	0.2074	0.226	0.917	0.360	-0.238	0.653
E	-0.1702	0.164	-1.035	0.302	-0.494	0.154
F	0.0361	0.111	0.324	0.746	-0.184	0.256
G	0.2824	0.113	2.501	0.013	0.060	0.505
Price_FirstLag	0.9929	0.004	230.139	0.000	0.984	1.001

Figure 4.27 VUL Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0024	0.004	0.623	0.534	-0.005	0.010
B	0.0206	0.018	1.125	0.262	-0.016	0.057
C	0.0062	0.006	1.054	0.293	-0.005	0.018
D	0.0080	0.005	1.508	0.133	-0.002	0.019
E	0.0020	0.013	0.156	0.877	-0.023	0.027
F	-0.0147	0.009	-1.693	0.092	-0.032	0.002
G	0.0051	0.004	1.414	0.159	-0.002	0.012
Price_FirstLag	0.9961	0.005	213.793	0.000	0.987	1.005

Figure 4.28 AGY Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0166	0.041	0.406	0.685	-0.064	0.097
B	0.0605	0.069	0.877	0.382	-0.076	0.197
C	0.0183	0.023	0.809	0.420	-0.026	0.063
D	0.0059	0.020	0.297	0.767	-0.033	0.045
E	-0.0199	0.021	-0.939	0.349	-0.062	0.022
F	0.0385	0.020	1.884	0.061	-0.002	0.079
G	0.0233	0.014	1.667	0.097	-0.004	0.051
Price_FirstLag	0.9900	0.004	232.227	0.000	0.982	0.998

Figure 4.29 NMT Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0033	0.003	1.068	0.287	-0.003	0.009
B	-4.377e-17	1.62e-18	-27.044	0.000	-4.7e-17	-4.06e-17
C	0.0101	0.004	2.696	0.008	0.003	0.017
D	0.0003	0.004	0.077	0.939	-0.008	0.008
E	-0.0022	0.003	-0.648	0.518	-0.009	0.004
F	-0.0025	0.004	-0.579	0.564	-0.011	0.006
G	0.0048	0.002	3.120	0.002	0.002	0.008
Price_FirstLag	0.9946	0.006	160.094	0.000	0.982	1.007

Figure 4.30 ASN Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0	0	nan	nan	0	0
B	-0.0287	0.028	-1.014	0.312	-0.085	0.027
C	0.0186	0.013	1.440	0.152	-0.007	0.044
D	0.0083	0.018	0.452	0.652	-0.028	0.044
E	0.0290	0.026	1.121	0.264	-0.022	0.080
F	0.0248	0.013	1.880	0.062	-0.001	0.051
G	0.0099	0.011	0.904	0.367	-0.012	0.031
Price_FirstLag	0.9941	0.004	280.925	0.000	0.987	1.001

Figure 4.31 PLL Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0137	0.012	-1.190	0.236	-0.037	0.009
B	0.0348	0.043	0.814	0.417	-0.049	0.119
C	-0.0161	0.015	-1.083	0.280	-0.045	0.013
D	-0.0091	0.023	-0.398	0.691	-0.054	0.036
E	0.0300	0.024	1.224	0.222	-0.018	0.078
F	-0.0181	0.035	-0.523	0.601	-0.086	0.050
G	0.0573	0.013	4.270	0.000	0.031	0.084
Price_FirstLag	0.9998	0.003	289.540	0.000	0.993	1.007

Figure 4.32 GLN Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0251	0.042	0.599	0.550	-0.057	0.108
B	3.387e-16	2.1e-17	16.108	0.000	2.97e-16	3.8e-16
C	-0.0029	0.028	-0.103	0.918	-0.058	0.052
D	-0.0628	0.046	-1.361	0.175	-0.154	0.028
E	0.0113	0.022	0.513	0.609	-0.032	0.055
F	0.0967	0.039	2.480	0.014	0.020	0.174
G	-0.0193	0.021	-0.901	0.369	-0.062	0.023
Price_FirstLag	1.0017	0.004	223.874	0.000	0.993	1.011

Figure 4.33 GL1 Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0	0	nan	nan	0	0
B	0	0	nan	nan	0	0
C	-1.034e-15	6.71e-16	-1.543	0.125	-2.36e-15	2.91e-16
D	7.9021	4.187	1.887	0.061	-0.373	16.177
E	-1.9664	2.414	-0.815	0.417	-6.738	2.805
F	2.6931	3.822	0.705	0.482	-4.861	10.247
G	-0.2210	2.787	-0.079	0.937	-5.729	5.287
Price_FirstLag	0.9994	0.003	323.991	0.000	0.993	1.005

Figure 4.34 ALB Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0139	0.006	2.272	0.024	0.002	0.026
B	-0.0006	0.020	-0.031	0.975	-0.039	0.038
C	-0.0047	0.006	-0.763	0.446	-0.017	0.007
D	0.0116	0.011	1.034	0.302	-0.011	0.034
E	0.0012	0.009	0.137	0.891	-0.017	0.019
F	0.0174	0.009	2.023	0.044	0.000	0.034
G	0.0099	0.007	1.524	0.129	-0.003	0.023
Price_FirstLag	0.9947	0.004	230.406	0.000	0.986	1.003

Figure 4.35 FFX Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0067	0.019	-0.361	0.719	-0.043	0.030
B	-0.0461	0.038	-1.215	0.226	-0.121	0.029
C	-0.0015	0.009	-0.175	0.861	-0.019	0.015
D	0.0164	0.012	1.417	0.158	-0.006	0.039
E	-0.0057	0.012	-0.479	0.632	-0.029	0.018
F	-0.0040	0.013	-0.318	0.751	-0.029	0.021
G	0.0096	0.012	0.828	0.409	-0.013	0.032
Price_FirstLag	0.9916	0.005	212.559	0.000	0.982	1.001

Figure 4.36 LPI Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0019	0.002	0.860	0.391	-0.002	0.006
B	-0.0015	0.009	-0.169	0.866	-0.019	0.016
C	0.0042	0.003	1.271	0.206	-0.002	0.011
D	0.0005	0.004	0.118	0.906	-0.008	0.009
E	-0.0010	0.003	-0.311	0.757	-0.008	0.006
F	0.0004	0.004	0.128	0.898	-0.006	0.007
G	0.0006	0.002	0.271	0.786	-0.003	0.005
Price_FirstLag	0.9957	0.006	158.750	0.000	0.983	1.008

Figure 4.37 AZL Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0150	0.044	-0.342	0.733	-0.101	0.071
B	1.249e-16	7.53e-18	16.586	0.000	1.1e-16	1.4e-16
C	0.1139	0.045	2.510	0.013	0.024	0.203
D	0.0574	0.047	1.231	0.220	-0.035	0.149
E	0.0615	0.045	1.365	0.174	-0.027	0.150
F	0.0121	0.054	0.225	0.822	-0.094	0.119
G	0.0601	0.032	1.882	0.061	-0.003	0.123
Price_FirstLag	0.9962	0.004	277.541	0.000	0.989	1.003

Figure 4.38 JRL Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0001	0.001	-0.104	0.917	-0.003	0.002
B	-0.0117	0.004	-2.827	0.005	-0.020	-0.004
C	0.0055	0.002	3.581	0.000	0.002	0.009
D	0.0017	0.002	0.737	0.462	-0.003	0.006
E	0.0065	0.003	2.020	0.045	0.000	0.013
F	-0.0005	0.002	-0.319	0.750	-0.004	0.003
G	0.0011	0.002	0.632	0.528	-0.002	0.005
Price_FirstLag	0.9975	0.005	212.833	0.000	0.988	1.007

Figure 4.39 EUR Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0006	0.000	-1.599	0.111	-0.001	0.000
B	0.0029	0.002	1.826	0.069	-0.000	0.006
C	0.0013	0.001	2.198	0.029	0.000	0.003
D	-0.0005	0.001	-0.913	0.363	-0.001	0.001
E	-0.0010	0.001	-1.744	0.083	-0.002	0.000
F	0.0008	0.001	1.222	0.223	-0.000	0.002
G	0.0005	0.000	1.013	0.312	-0.000	0.001
Price_FirstLag	1.0027	0.004	235.272	0.000	0.994	1.011

Figure 4.40 LPD Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0181	0.008	2.285	0.023	0.002	0.034
B	-0.0244	0.027	-0.903	0.368	-0.078	0.029
C	-0.0031	0.010	-0.315	0.753	-0.023	0.016
D	0.0088	0.009	0.984	0.326	-0.009	0.027
E	0.0044	0.008	0.573	0.568	-0.011	0.019
F	0.0060	0.015	0.400	0.690	-0.023	0.035
G	-0.0031	0.005	-0.680	0.497	-0.012	0.006
Price_FirstLag	0.9972	0.005	187.673	0.000	0.987	1.008

Figure 4.41 ESS Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0136	0.022	0.632	0.528	-0.029	0.056
B	0.0030	0.035	0.085	0.932	-0.066	0.072
C	0.0238	0.023	1.034	0.302	-0.022	0.069
D	0.0207	0.020	1.016	0.311	-0.019	0.061
E	-0.0080	0.027	-0.296	0.768	-0.061	0.045
F	0.0990	0.035	2.850	0.005	0.030	0.168
G	0.1191	0.030	3.916	0.000	0.059	0.179
Price_FirstLag	0.9988	0.004	278.489	0.000	0.992	1.006

Figure 4.42 EMH Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0009	0.003	0.331	0.741	-0.004	0.006
B	0.0036	0.007	0.539	0.590	-0.009	0.017
C	-0.0061	0.004	-1.727	0.086	-0.013	0.001
D	-0.0018	0.004	-0.425	0.671	-0.010	0.007
E	-0.0065	0.004	-1.792	0.075	-0.014	0.001
F	0.0006	0.005	0.114	0.909	-0.009	0.010
G	0.0064	0.003	2.158	0.032	0.001	0.012
Price_FirstLag	0.9926	0.006	167.772	0.000	0.981	1.004

Figure 4.43 INF Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.3179	0.063	5.021	0.000	0.193	0.443
B	-0.0308	0.059	-0.527	0.599	-0.146	0.085
C	-0.0010	0.059	-0.017	0.987	-0.117	0.115
D	0.0258	0.024	1.054	0.293	-0.022	0.074
E	-0.0092	0.024	-0.377	0.706	-0.057	0.039
F	0.0045	0.066	0.069	0.945	-0.125	0.134
G	0.0188	0.018	1.041	0.299	-0.017	0.054
Price_FirstLag	1.0010	0.005	193.020	0.000	0.991	1.011

Figure 4.44 LEL Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0009	0.001	-0.859	0.392	-0.003	0.001
B	-0.0037	0.004	-0.845	0.399	-0.012	0.005
C	0.0009	0.001	0.685	0.494	-0.002	0.003
D	0.0007	0.002	0.363	0.717	-0.003	0.005
E	-0.0002	0.003	-0.057	0.955	-0.006	0.006
F	-0.0015	0.001	-1.562	0.120	-0.003	0.000
G	0.0005	0.001	0.757	0.450	-0.001	0.002
Price_FirstLag	1.0021	0.003	310.655	0.000	0.996	1.008

Figure 4.45 LIT Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0016	0.001	-1.609	0.110	-0.004	0.000
B	4.312e-05	0.002	0.026	0.979	-0.003	0.003
C	0.0003	0.001	0.617	0.538	-0.001	0.001
D	-0.0025	0.001	-3.719	0.000	-0.004	-0.001
E	0.0004	0.001	0.498	0.619	-0.001	0.002
F	-0.0032	0.001	-2.583	0.011	-0.006	-0.001
G	0.0003	0.000	0.683	0.495	-0.001	0.001
Price_FirstLag	0.9967	0.008	118.742	0.000	0.980	1.013

Figure 4.46 ALY Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0009	0.008	0.113	0.911	-0.015	0.016
B	-1.026e-16	6.65e-19	-154.336	0.000	-1.04e-16	-1.01e-16
C	0.0164	0.011	1.529	0.128	-0.005	0.037
D	0.0005	0.010	0.052	0.959	-0.019	0.020
E	-0.0031	0.007	-0.478	0.633	-0.016	0.010
F	3.492e-18	1.28e-18	2.733	0.007	9.73e-19	6.01e-18
G	0.0152	0.005	3.096	0.002	0.006	0.025
Price_FirstLag	0.9988	0.006	175.684	0.000	0.988	1.010

Figure 4.47 KZR Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0	0	nan	nan	0	0
B	0.1162	0.035	3.307	0.001	0.047	0.186
C	-0.0020	0.016	-0.124	0.901	-0.034	0.030
D	-0.0036	0.013	-0.280	0.780	-0.029	0.022
E	0.0109	0.016	0.676	0.500	-0.021	0.043
F	-0.0979	0.047	-2.095	0.038	-0.190	-0.006
G	-0.0142	0.010	-1.374	0.171	-0.035	0.006
Price_FirstLag	0.9869	0.006	168.654	0.000	0.975	0.998

Figure 4..48 BMM Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0454	0.043	1.063	0.289	-0.039	0.130
B	0.0494	0.124	0.398	0.691	-0.195	0.294
C	0.1090	0.059	1.847	0.066	-0.007	0.225
D	0.0252	0.044	0.566	0.572	-0.063	0.113
E	-0.0165	0.048	-0.341	0.733	-0.112	0.079
F	1.268e-19	5.99e-18	0.021	0.983	-1.17e-17	1.19e-17
G	0.0285	0.026	1.106	0.270	-0.022	0.079
Price_FirstLag	0.9415	0.034	27.625	0.000	0.874	1.009

Figure 4.49 MHK Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	-0.0133	0.060	-0.220	0.826	-0.132	0.106
B	-0.0056	0.113	-0.049	0.961	-0.229	0.218
C	0.0039	0.084	0.047	0.963	-0.162	0.170
D	0.0042	0.054	0.079	0.938	-0.102	0.110
E	0.0222	0.086	0.259	0.796	-0.147	0.192
F	-0.0156	0.062	-0.253	0.801	-0.137	0.106
G	0.0252	0.041	0.615	0.539	-0.056	0.106
Price_FirstLag	0.9900	0.022	45.898	0.000	0.947	1.033

Figure 4.50 A8G Company (Linear Regression Model Results)

	coef	std err	t	P>|t|	[0.025	0.975]
A	0.0007	0.001	0.785	0.433	-0.001	0.002
B	-2.405e-15	1.89e-17	-127.287	0.000	-2.44e-15	-2.37e-15
C	0.0022	0.001	3.631	0.000	0.001	0.003
D	-0.0003	0.001	-0.477	0.634	-0.001	0.001
E	-0.0003	0.001	-0.269	0.788	-0.003	0.002
F	0.0008	0.001	0.857	0.393	-0.001	0.003
G	-7.018e-05	0.000	-0.176	0.860	-0.001	0.001
Price_FirstLag	0.9947	0.008	126.522	0.000	0.979	1.010

Figure 4.51 VKA Company (Linear Regression Model Results)

4.3 Research Question 3

In this study, the aim was to show the impact of features of the tweets (number of likes, number of views, number of comments, number of reshares and sentiment label) posted by Australian lithium-related companies on their stock prices.

I used VADER, a rule-based sentiment analysis method, introduced by Hutto, C. J., & Gilbert, E. (2014) to classify company tweets into three categories of positive, negative, and neutral. Hutto, C. J., & Gilbert, E. (2014) combined quantitative and qualitative methods and algorithms to introduce a model to analyse the sentiment of the texts. They compared their model to eleven sentiment analysis methods and VADER provided more accurate results. VADER builds a list of lexicon features and uses grammatical and syntactical rules to analyse the text sentiment and assign a label to the text.

The computed sentiment label then will be used as a feature in a XGBoost machine learning model. Finally, we add some additional features such as price lag, number of likes, number of reshares, number of views, and number of comments and run our XGBoost model to estimate the impact of the incorporated features on the change direction of stock price of the company.

Below screenshots show the result of the XGBoost model run on each company's data.

	coef	std err	t	P>|t|	[0.025	0.975]
Is Shared	0.0495	0.024	2.057	0.074	-0.006	0.105
Is Liked	0.0495	0.024	2.057	0.074	-0.006	0.105
Is Commented	0.0349	0.051	0.688	0.511	-0.082	0.152
Negative Sentiment	0	0	nan	nan	0	0

Figure 4.52 AGY Company (XGBoost Model Results)

As we can see in the results, Is Shared and Is Liked features have statistically significant impact on the stock price data of the company on the next trading day. This shows tweets with more like and shares had more impact on the stock price of the company, compared to the tweets with more comments. The results also show that the Negative Sentiment feature has no statistically significant impact on the stock price of the company in the selected data period. For this company, the comments of the tweets have some impact on the stock price of the company on the next trading day, but the impact is not statistically significant as the Confidence Level is less than 90%.

	coef	std err	t	P>|t|	[0.025	0.975]
Is Shared	0.0148	0.035	0.421	0.680	-0.061	0.091
Is Liked	0.1403	0.035	4.055	0.001	0.066	0.215
Is Commented	0.0166	0.033	0.504	0.622	-0.054	0.088
Negative Sentiment	0	0	nan	nan	0	0

Figure 4.53 BMM Company (XGBoost Model Results)

Figure 4.53 demonstrates the results of our XGBoost model on BMM company's stock data. As our model shows, the Is Liked feature (which indicates the number of likes of each tweet) has a statistically significant impact on the stock price of the company's next trading day, with Confidence Level of 99%. The other features, Is Shared (number of reshares of each tweet) and Is Commented (number of comments of each tweet) have Confidence Level of 68% and 62%, but they don't impact the stock price of the next trading day significantly. While the sentiment label of the tweet (Negative Sentiment which shows if the tweet sentiment's label is negative) does not impact the stock price of the company on the next trading day.

	coef	std err	t	P>|t|	[0.025	0.975]
Is Shared	4.7461	3.113	1.525	0.147	-1.852	11.345
Is Liked	64.9635	1.994	32.584	0.000	60.737	69.190
Is Commented	0.8661	3.113	0.278	0.784	-5.732	7.465
Negative Sentiment	0	0	nan	nan	0	0

Figure 4.54 Mineral Company (XGBoost Model Results)

Figure 4.54 shows the result of our XGBoost model on Mineral company's stock price data. The model shows Is Liked feature (number of likes of the tweet) has a statistically significant impact on the stock price of the next trading day. The Confidence Level of the Is Liked feature is 100%, and the estimated coefficient for this variable (64.9) is larger than the coefficient of the other variables. This shows that Is Liked feature has the largest impact on the output (stock price) compared to Is Shared (number of the tweet's reshare) and Is Commented (number of the comments of each tweet). Although the Is Shared feature's P-Value is 0.14, which shows it affects the stock price of the next trading day, but it has not been considered statistically significant as its confidence level is less than 90%. The Is Commented feature has the Confidence Level of 78%, which shows that it does not have much impact on the next trading day's stock price. Similar to the BMM company, the model shows no relationship between the negative sentiment of the tweet and the stock price of the next trading day.

	coef	std err	t	P>|t|	[0.025	0.975]
Is Shared	0.0474	0.492	0.096	0.924	-0.949	1.044
Is Liked	3.0325	0.445	6.819	0.000	2.132	3.933
Is Commented	0.3768	0.292	1.288	0.205	-0.215	0.969
Negative Sentiment	-0.3274	0.476	-0.688	0.496	-1.291	0.636

Figure 4.55 Vulcan Company (XGBoost Model Results)

Figure 4.55 illustrates the results of the XGBoost model on Vulcan company's stock price data. Similar to BMM and Mineral companies, the Is Liked feature that indicates the number of likes of each tweet has confidence level of 100% which means it impacts the stock price of the next trading's data significantly. Is Commented (the number of comments of each tweet) shows the p-value of 0.2, which shows its impact on the output (next trading day's stock price). As the confidence level is less than 90%, it has not been considered as statistically significant. The Negative Sentiment feature which shows the tweet's label has the P-value of 0.4, which means while it affects the stock price of the company on the next trading day, its effect is not significant. Is Shared feature shows the highest P-value among other feature, with 0.9 which means it has less impact on the stock price compared to the other features.

	coef	std err	t	P>|t|	[0.025	0.975]
const	48.9525	3.634	13.472	0.000	39.612	58.293
Is Shared	-17.8775	4.510	-3.964	0.011	-29.471	-6.284
Is Liked	17.0775	4.182	4.084	0.010	6.328	27.827
Is Commented	-8.8500	4.627	-1.913	0.114	-20.745	3.045
Negative Sentiment	13.9975	3.634	3.852	0.012	4.657	23.338

Figure 4.56 Wesfarmers Company (XGBoost Model Results)

As figure 4.56 shows, Is Shared, Is Liked and Negative Sentiment features have statistically significant impact of the next trading day's stock price of the Wesfarmers company. All

Negative Sentiment (the negative label of the tweet), Is Liked (number of likes of each tweet), and Is Shared (number of reshares of the tweet) are having the confidence level of 99%, so they are considered as statistically significant features.

Although Is Commented feature (number of comments on each tweet) shows the confidence level of 89% and has impact on the stock price of the company on the next trading day, it has not been considered as statistically significant as its confidence level is less than 90%.

Here's the summary of the results for all companies studied:

Company	Is Shared (P > |t|)	Is Shared (coef)	Is Liked (P > |t|)	Is Liked (coef)	Is Commented (P > |t|)	Is Commented (coef)	Negative Sentiment (P > |t|)	Negative Sentiment (coef)
AGY	0.074	0.495	0.074	0.495	0.511	0.349	nan	0
BMM	0.68	0.0148	0.001	0.1403	0.622	0.0166	nan	0
Mineral	0.147	4.74	0	64.9635	0.784	0.8661	nan	0
Vulcan	0.924	0.0474	0	3.0325	0.205	0.3768	0.496	-0.3274
Wesfarmers	0.011	-17.8775	0.01	17.0775	0.114	-8.85	0.012	13.9975

Figure 4.57 Summary of the results for all companies studied

In this chapter the findings of the three research questions included, started from the historical stock prices of lithium-related companies and their effects on the companies' stock market. Then I investigated the effects of the company news and announcements on their stock market. In the next section, I analysed the effects of company tweets and tweets' features on the stock market of companies studied.

Chapter 5: Discussion and Conclusion

This chapter presents a comprehensive synthesis of the research findings derived from the three central research questions, each of which explored how distinct types of data—historical stock trends, formal corporate announcements, and social media signals—can be leveraged to model and predict stock market performance in the lithium sector. The study lies at the intersection of machine learning, financial analytics, and sustainable investment, with a particular emphasis on Australian lithium-related companies, which play a critical role in the global clean energy transition.

By employing advanced machine learning models—namely, Support Vector Machines (SVM) and Extreme Gradient Boosting (XGBoost)—and integrating multiple heterogeneous data sources, this research develops a robust, data-driven framework for understanding market behaviour in high-growth, resource-intensive sectors. This chapter outlines the study's theoretical, practical, and methodological contributions, addresses key gaps in the existing literature, highlights critical insights, and suggests directions for future inquiry

5.1 Theoretical Contributions

This thesis advances theoretical understanding in multiple dimensions, particularly in how diverse and complex data sources can be operationalised through machine learning for financial analysis.

For RQ1, the application of SVM demonstrated the predictive value of historical stock indicators such as momentum and volatility. The model, applied to companies like Vulcan and Pilbara Minerals, achieved approximately 49% accuracy in forecasting next-day stock direction. While modest, this performance reflects the inherent volatility in commodity markets and reinforces the utility of linear classification models as effective baselines in financial modelling (Inthachot et al. 2016 and Dey et al. 2016).

For RQ2, XGBoost models applied to over two years of corporate announcements revealed how market capitalisation mediates the influence of different communication types. Results showed that investor reactions were highly dependent on both the content and context of disclosures, with large-cap firms responding more strongly to formal events like public meetings and financial reports, while small-caps reacted more to security-related notices. This nuanced differentiation contributes to financial communication theory by empirically validating that "one-size-fits-all" investor communication strategies are suboptimal (Dey et al. 2016).

For RQ3, the integration of social media sentiment—using VADER—and user engagement metrics showed that digital investor sentiment has measurable effects on market behaviour, particularly when measured through engagement indicators such as likes, shares, and views. While sentiment polarity alone had limited predictive strength, the strong performance of interaction-based features provides a new lens for understanding behavioural finance and market microdynamics (Smith and O'Hare 2022).

Taken together, these findings contribute to a growing body of knowledge on data fusion in financial forecasting, demonstrating how machine learning can unify structured and unstructured signals into a cohesive analytical model. By integrating historical prices, corporate announcements, and social media engagement, this study uncovers previously unrecognised patterns in stock price changes—patterns that are not detectable when data sources are analysed in isolation. This directly advances theoretical understanding by showing that lithium stock performance is shaped by a multilayered information ecosystem rather than by single-factor influences.

5.1.1 Addressing Literature Gaps

Beyond these theoretical contributions, the research directly addresses several pressing gaps in current academic literature:

1. **Lack of sector-specific analysis**: Most existing stock market studies focus on broad indices or major tech and industrial sectors. Very few have examined commodity-specific stocks—particularly lithium—which is increasingly central to global decarbonisation efforts. This thesis offers a focused investigation of lithium-related companies, responding to the need for more tailored financial insights into green economy sectors.

2. **Underrepresentation of the Australian market**: Despite Australia's leading role in global lithium production, most academic studies emphasise the U.S. or Chinese markets. This research contributes critical Australia-specific findings, reflecting the local market structure, investor behaviour, and regulatory environment.

3. **Limited integration of multi-modal data**: The majority of prior studies rely on single-source data, typically historical prices or news sentiment. By integrating time-series

stock data, structured announcement disclosures, and unstructured social media content, this thesis creates a multi-dimensional model that mirrors the complexity of modern financial ecosystems.

4. **Neglect of firm-generated digital communications**: While social media-based finance research often focuses on retail investor sentiment or public forums, few studies have examined company-published content. By analysing tweets posted by lithium companies themselves, the study reveals how corporate digital narratives affect investor perception and market outcomes.

5. **Minimal application of AI/ML to sustainability-aligned sectors**: Although machine learning is widely used in financial modelling, its application to clean energy-related domains remains limited. This study contributes to this emerging space by applying explainable AI methods to a sustainability-critical sector, thereby bridging machine learning innovation with ESG-aligned investment analysis.

By explicitly targeting these research gaps, this study strengthens the interdisciplinary dialogue between information systems, computer science, finance, and sustainability. Moreover, it provides a scalable analytical blueprint for future investigations into other high-impact resource sectors such as cobalt, nickel, and hydrogen.

5.2 Practical Contributions

The practical relevance of this research is evident across various stakeholder groups, including investors, financial analysts, corporate communicators, and ESG portfolio managers.

- **For Investors and Traders**: The SVM model provides an accessible and computationally efficient tool to forecast short-term price movements based on internal market signals. While not a standalone solution, it offers a baseline for hybrid models in high-frequency trading contexts.

- **For Portfolio Managers**: The XGBoost model demonstrated the importance of tailoring monitoring systems to firm size and announcement type. Large-cap firms warrant more attention during financial briefings and public meetings, while monitoring of small-cap firms may benefit from focusing on security disclosures and listing changes.

- **For Financial Analysts and Data Scientists**: The research introduces a framework for blending structured and unstructured data sources, increasing transparency in feature importance, and supporting explainability in algorithmic trading models.
- **For ESG-Focused Funds**: Given the environmental relevance of lithium, the study's insights can directly inform ESG investment screening and decision-making, especially in evaluating how well companies communicate their performance and sustainability strategy.

5.3 Methodological Enhancements

This study contributes several methodological innovations:

- Developed a rich, domain-specific dataset integrating price history, structured announcements, and social media content from over 40 companies.
- Employed both linear (SVM) and non-linear (XGBoost) algorithms to explore the relative strengths of different modelling strategies.
- Applied VADER sentiment analysis to corporate tweets while incorporating social engagement metrics, a novel combination that extends standard textual analysis.
- Used feature importance interpretation and comparative accuracy to validate model transparency and relevance, supporting the trustworthiness of AI in financial contexts.

These methodological choices contribute to reproducible, adaptable, and ethically sound applications of machine learning in financial research.

The primary objective of this study is the investigation and analysis rather than prediction of the effects of various data sources, including historical data, traditional news, and social media, on the stock prices of lithium-related companies in Australia. In addressing the second research question, the study examined how different categories of company announcements (such as quoted securities, unquoted securities, meetings and presentations, and others) affect the stock performance of small-cap, medium-cap, and large-cap companies. An XGBoost model was employed to conduct feature importance analysis, identifying which categories of announcements exerted the most significant influence on stock price fluctuations across various market capitalisations. For the third research question, input variables such as tweet likes, reshares, comments, and sentiment labels were analysed in relation to stock prices on the subsequent trading day. The aim of this investigation was not to develop a predictive model with high statistical performance (e.g., high R-squared), but rather to explore whether interactions and sentiment expressed through official company tweets contribute to observable stock price changes. The methodological innovation of this study is in feature engineering, constructing domain-specific variables (e.g. stock momentum, stock volatility, news categories, tweets' sentiment, …), application design, and multi-layer investigation of Australian lithium-related companies stock market. The machine learning models used in this thesis served the analytical purpose effectively, enabling the identification of key influencing features across heterogeneous data sources. These findings contribute toward the broader objective of providing actionable insights to inform future research and support the decision-making processes of stakeholders, investors, financial analysts, and policymakers.

5.3.1 Model Selection Rationale

XGBoost model was selected for investigating the effects of traditional news data (news articles) and tweets interactions and sentiment on the stock price of lithium-related companies because of the flowing reasons:

- Small and Medium Sized Structured Tabular Dataset

As the stock price data and the input datasets are structured and tabular, I needed to use a method that can handle these datasets very well. XGBoost performs well in handling structured and tabular datasets and identifies the relationship between the news article categories and tweets interactions and sentiment and stock price. Since tweets interactions (likes, comments, shares), and sentiment scores are numeric and structured, and news categories transformed to numeric data, XGBoost performs well in learning from this type of features and analyse this data. It also works very well with small to medium size datasets. Deep learning models such as neural networks need large datasets and longer training time, and they might not perform very well with small datasets.

- Nonlinear Relationships in data

Stock prices are affected by nonlinear combinations of tweets sentiments and interactions, also by news articles categories. XGBoost performs well in identifying non-linear relationships between news articles categories features, tweets interactions (features like likes, comments, shares), and sentiment score and stock price. XGBoost uses decision trees to model the complex interactions, while linear models such as Linear Regression or SVM with linear kernels assume and linear relationships in dataset.

- Feature Importance

XGBoost provides feature importance scores which show the influence of different features (historical news categories, tweets interactions, tweets sentiment) on the stock price data. This helps the interpretability and analysis of the results and aligns with purpose of this study which is investigation (and not prediction) of the effects of various data sources on stock price.

- Handling Missing or Noisy Data

Social media data and traditional news data can be incomplete as some companies don't share posts regularly or don't share news about themselves often. To analyse this kind of data, the machine learning model needs to be less sensitive to missing and noisy data. XGBoost handles incomplete data very well compared to basic decision trees and deep learning models.

- Performance and Interpretability

SVM in powerful tool for classification but not great in interpretability in regression so it is not a great option in this use case. Linear regression has limited ability in capturing complex relationships in data and it assumes linear relationships, while the relationships in this study's data in not always linear.

Random Forest is usually slower and need more processing power, also not easy to tune compared to XGBoost. Deep learning models including Neural Network models require large datasets. Compared to above models, XGBoost offers a trade-off in performance, interpretability and finding the complex non-linear relationships between data.

5.4 Key Insights and Implications

The triangulated findings from RQ1–RQ3 yield several meaningful implications:

- Momentum and volatility remain key technical indicators, supporting their continued use in predictive models.
- Investor reactions are contingent on the nature and perceived credibility of information sources, with firm size influencing which types of disclosures matter most.

- Engagement-based social media metrics outperform sentiment polarity in predictive strength—suggesting a behavioural shift in how market participants process digital signals.
- Explainable models with clear feature importance are essential for gaining stakeholder trust, particularly in ESG-aligned and compliance-sensitive sectors.

These insights are not only academically valuable but also operationally significant for those building intelligent investment systems and communication strategies in resource-intensive industries.

One key insight is the disproportionate impact of public meeting and presentation announcements on the stock prices of large-cap companies. This suggests that institutional investors and media outlets are more likely to respond to corporate communications from firms with established market reputations. As such, large firms may need to strategically manage their investor relations activities to shape market expectations. Conversely, smaller firms may benefit more from operational or technical disclosures that demonstrate growth potential.

5.5 Limitations

While this research offers significant insights, it is not without limitations:

- The social media dataset (company tweets) was limited by varying levels of corporate activity; some companies rarely post, affecting consistency and breadth.
- VADER, while effective, may not fully capture financial nuances or sarcasm present in corporate communications.
- All models were trained and tested retrospectively. Real-time deployment would further validate their utility and performance.
- The study is geographically focused on Australia; extrapolation to global markets requires further testing and regional calibration.

5.6 Future Research Directions

To expand on the current work, future research may consider:

- Leveraging transformer-based NLP models (e.g., FinBERT, RoBERTa) for higher-fidelity sentiment and contextual analysis.

- Incorporating public sentiment from forums (e.g., Reddit, StockTwits) and macroeconomic indicators (e.g., lithium futures, interest rates).
- Exploring causal inference and reinforcement learning approaches for adaptive financial modelling.
- Developing interactive, real-time dashboards for portfolio managers, integrating digital sentiment and traditional price signals.
- Applying the same multi-source framework to other ESG-aligned commodities, including cobalt, hydrogen, and rare earth metals.

Future research could explore how the nature of different industries such as those involving Nickel and Cobalt, which are also essential for battery production modulates the influence of news and social sentiment on stock prices. Examining sector-specific characteristics may reveal how regulatory frameworks, public awareness, and investor behaviour uniquely shape market reactions in various clean energy segments.

Conclusion

This research demonstrates that a multi-source, machine learning–driven approach can significantly improve the accuracy, transparency, and relevance of stock price prediction models particularly in fast-evolving, sustainability-critical sectors such as lithium. By integrating historical price data, corporate disclosures, and real-time digital sentiment, the study builds a comprehensive model for understanding investor behaviour and market dynamics.

The findings show that the traditional news data and social media data have impact on the stock prices of lithium-related companies in Australia. Different news categories have different impacts on the stock prices of the companies in small, medium and large categories. Tweets sentiments also have impact on the stock market of lithium-related companies studied.

The findings contribute meaningful theoretical advancements, practical tools, and methodological blueprints that bridge the gap between artificial intelligence, financial modelling, and sustainable investment. In doing so, this thesis lays the foundation for future interdisciplinary research that supports informed, responsible, and adaptive decision-making in financial markets and beyond.

References:

- Aghabozorgi, S., Wah Teh, Y., (2014) “Stock market co-movement assessment using a three-phase clustering method” Expert Systems with Applications 41 (2014) 1301-1314
- Andriamasinoro, F. and Danino-Perraud, (2021). Use of artificial intelligence to assess mineral substance criticality in the French market: the example of cobalt. Mineral Economics, 34(1) ,19-37.
- Astudillo, G., Carrasco, R., Fernández-Campusano, C. and Chacón, M., (2020). Copper price prediction using support vector regression technique. applied sciences, 10(19), 6648.
- Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552–567. https://doi.org/10.1016/j.najef.2018.06.013.
- Ben Ameur, H., Boubaker, S., Ftiti, Z., Louhichi, W., Tissaoui, K., (2023). Forecasting commodity prices: empirical evidence using deep learning tools. Ann. Oper. Res. https://doi.org/10.1007/s10479-022-05076-6.
- Bujari, A., Furini, M., & Laina, N. (2017). On using cashtags to predict companies stock trends. In 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC) (pp. 25–28). IEEE. https://doi.org/10.1109/CCNC.2017.7983080.
- Cao, H., Lin, T., Li, Y., & Zhang, H. (2019). Stock Price Pattern Prediction Based on Complex Network and Machine Learning. *Complexity*, 2019, Article ID 4132485. https://doi.org/10.1155/2019/4132485.
- Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., & Oliveira, A. L. (2016). Computational intelligence and financial markets: A survey and future directions. Expert Systems with Applications, 55, 194-211.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Chung, H., & Shin, K. S. (2020). Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. *Neural Computing and Applications*, 32(12), 7897–7914. https://doi.org/10.1007/s00521-019-04236-3.

- Colaka, I., Sagiroglub, S., Yesilbudak, M. (2012) “Data mining and wind power prediction: A literature review” Renewable Energy: 46, 241-247.
- Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of Twitter streaming data for stock prediction. Procedia Computer Science, 132, 956-964.
- De Oliveira Carosia, A. E., Coelho, G. P., & Da Silva, A. E. A. (2021). Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning. *Expert Systems with Applications*, 184, 115470. https://doi.org/10.1016/j.eswa.2021.115470.
- Dessemond, C., Lajoie-Leroux, F., Soucy, G., Laroche, N., & Magnan, J.-F. (2019). Spodumene: The lithium market, resources and processes. Minerals, 9(6), 334.
- Devi, K. N., & Bhaskaran, V. M. (2015). Semantic Enhanced Social Media Sentiments for Stock Market Prediction. *International Journal of Economics and Management Engineering*, 9(2), 684–688. https://zenodo.org/records/1100501/files/10001130.pdf.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.
- Dey, S., Kumar, Y., Saha, S., & Basak, S. (2016). Forecasting to Classification: Predicting the Direction of Stock Market Price Using Xtreme Gradient Boosting. *PESIT South Campus*, 1–10.
- Fama, E. F. (1995). Random Walks in Stock Market Prices. *Financial Analysts Journal*, *51*(1), 75–80. https://doi.org/10.2469/faj.v51.n1.1861.
- Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X., & Zou, H. (2019). Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. Renewable and Sustainable Energy Reviews, 100, 186-212.
- Garrido-Hidalgo, C., Ramirez, F.J., Olivares, T., Roda-Sanchez, L. (2020) “The adoption of internet of things in a circular supply chain framework for the recovery of WEEE: the case of lithium-ion electric vehicle battery packs”. Waste Management: 103, 32-44.
- Ghoddusi, H., Creamer, G., Rafizadeh, N., (2019) “Machine learning in energy economics and finance: A review” Energy Economics 81 (2019) 709-727.
- Grant, A., Deak, D., Alto, P., Pell, R. (2020) “The CO2 Impact of the2020s’ Jade Cove Partners Battery Quality Lithium Hydroxide Supply Chain”, Jade Cove Partners, https://www.jadecove.com/research/liohco2impact.

- Grey, C. P., & Hall, D. S. (2020). Prospects for lithium-ion batteries and beyond—a 2030 vision. Nature Communications, 11(1), 6279.
- Gupta, R., Garg, N., & Singh, S. (2013). Stock market prediction accuracy analysis using kappa measure. In Proceedings of the 2013 International Conference on Communication Systems and Network Technologies (CSNT) (pp. 635–639). IEEE. https://doi.org/10.1109/CSNT.2013.136.
- Gururaj, V., Shriya, V. R., Ashwini, K. (2019). Stock Market Prediction using Linear Regression and Support Vector Machines. International Journal of Applied Engineering Research, 14(8), 1931-1934.
- Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications*, 129, 273–285. https://doi.org/10.1016/j.eswa.2019.03.029.
- Howells, K., & Ertugan, A. (2017). Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia Computer Science, 120*, 664–670. https://doi.org/10.1016/j.procs.2017.11.293.
- Hu, Z., Zhao, Y., & Khushi, M. (2021). A Survey of Forex and Stock Price Prediction using Deep Learning. *Applied System Innovation*, 4(1), 9. https://doi.org/10.3390/asi4010009.
- Hu, Z., Liu, W., Bian, J., Liu, X., & Liu, T.-Y. (2018). Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.
- Huber, J., Tammer, C., Krotil, S., Waidmann, S., Hao, X., Seidel, C., Reinhart, G. (2016) "Method for classification of battery separator defects using optical inspection", Procedia CIRP: 57, 585-590.
- Hutto, C.J. & Gilbert, E.E., 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI, USA, 1–4 June 2014. AAAI Press, pp. 216–225.
- IEA. (2022). The Role of Critical Minerals in Clean Energy Transitions. Available at: https://iea.blob.core.windows.net/assets/ffd2a83b-8c30-4e9d-980a-52b6d9a86fdc/TheRoleofCriticalMineralsinCleanEnergyTransitions.pdf [Accessed 19 Dec. 2023].

- IEA. (2023). Critical Minerals Market Review 2023. Available at: https://iea.blob.core.windows.net/assets/afc35261-41b2-47d4-86d6-d5d77fc259be/CriticalMineralsMarketReview2023.pdf [Accessed 19 Dec. 2023].
- Inthachot, M., Boonjing, V., & Intakosum, S. (2016). Artificial neural network and genetic algorithm hybrid intelligence for predicting Thai stock price index trend. *Computational Intelligence and Neuroscience*, 2016, Article ID 3045254. https://doi.org/10.1155/2016/3045254.
- Irene, H., & Perry, S. (2023). Forecasting Rare Earth Stock Prices with Machine Learning. Resource Policy, 186, 104248.
- Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, 184, 115537. https://doi.org/10.1016/j.eswa.2021.115537.
- Jiao, P., Veiga, A., Walther, A., (2020) “Social media, news media and the stock market” Journal of Economic Behavior and Organization 176 (2020) 63-90.
- Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, 178, 115019. https://doi.org/10.1016/j.eswa.2021.115019.
- Kalyani, J., Bharathi, P., & Jyothi, P. (2016). Stock trend prediction using news sentiment analysis. arXiv preprint arXiv:1607.01958.
- Karaca, H., & Kozat, S. S. (2025). *Soft gradient boosting with learnable feature transforms for sequential regression*. arXiv. https://arxiv.org/abs/2509.12920
- Kianrad, A., Arani, M.N., Hasani, K., Zargar, M, Erfani, E., Chaharmahali, A.R. Investigating the impact of company announcements on stock prices: an application of machine learning on Australian lithium market. Miner Econ 37, 163–172 (2024). https://doi.org/10.1007/s13563-024-00428-z
- Kianrad, A., Hasani, K., Erfani, E., Chaharmahali, A.R. & Gandomi, A.H., 2024. Twitter sentiment analysis and its impact on stock prices of Australian lithium-related companies: A support vector machine approach. In: Proceedings of the 2024 International Conference on Soft Computing and Machine Intelligence (ISCMI 2024).
- Kumar, G., Singh, U. P., & Jain, S. (2023). Stock price forecasting based on the relationship among Asian stock markets using deep learning. *Concurrency and Computation: Practice and Experience*, 35(28), e7824. https://doi.org/10.1002/cpe.7824.

- Kwade, A., Haselrieder, W., Leithoff, R., Modlinger, A., Dietrich, F., Droeder, K. (2018) "Current status and challenges for automotive battery production technologies". Nature Energy: 3, 290–300.
- Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management*, 57(5), 102212. https://doi.org/10.1016/j.ipm.2020.102212.
- Liu, D., Gao, X., An, H., Qi, Y., Wang, Z., Jia, N., & Chen, Z. (2020). Exploring behavior changes of the lithium market in China: Toward technology-oriented future scenarios. Resources Policy, 69, 101885.
- Liu, F., Huang, X., Huang, W., & Duan, S. X. (2020). Performance Evaluation of Keyword Extraction Methods and Visualization for Student Online Comments. *Symmetry*, *12*(11), 1923. https://doi.org/10.3390/sym12111923.
- Liu, J.-X., Leu, J.-S., & Holst, S. (2023). Stock price movement prediction based on Stocktwits investor sentiment using FinBERT and ensemble SVM. *PeerJ Computer Science*, 9, e1403. https://doi.org/10.7717/peerj-cs.1403.
- Liu, L., Wu, J., Liu, L., Li, P., Li, Q., (2015) "A social-media-based approach to predicting stock comovement" Expert Systems with Applications 42 (2015) 3893-3901.
- Loria, S. (2018). Textblob Documentation. Release 0.15, 2.
- Madge, S., Bhatt, S., (2015) "Predicting Stock Price Direction using Support Vector Machines" Independent Work Report Spring 2015
- Martin, G., Rentsch, L., Höck, M., & Bertau, M. (2017). Lithium market research–global supply, future demand and price development. Energy Storage Materials, 6, 171-179.
- Miklosik, A., & Evans, N. (2021). Environmental sustainability disclosures in annual reports of mining companies listed on the Australian Stock Exchange (ASX). Heliyon, 7(7), e07505.
- Miles, T., (2016) "AI in supply chain planning" Rise of the smart machines: How concurrent planning and AI can help you win big in supply chain planning" https://www.supplychainbrain.com/ext/resources/secure_download/KellysFiles/WhitePapersAndBenchMarkReports/Kinaxis/Kinaxis_Rise_of_the_Smart_Machines.pdf.
- Misra, P., & Chaurasia, S. (2020). Data-Driven Trend Forecasting in Stock Market Using Machine Learning Techniques. *Journal of Information Technology Research (JITR)*, 13(1), 130–149. https://doi.org/10.4018/JITR.2020010109.

- Mordor Intelligence. (2023). Australia Electric Vehicle Market. Available at: https://www.mordorintelligence.com/industry-reports/australia-electric-vehicle-market [Accessed 19 Dec. 2023].
- Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., & Liu, J. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture.
- Nielsen, F. A. (2011). A new ANEW: evaluation of a word list for sentiment analysis in microblogs. Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages, 93-98.
- Nishant, R., Kennedy, M., Corbett, J., (2020) "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda". International Journal of Information Management, 53 (2020) 102104.
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data, 7*(1), 20. https://doi.org/10.1186/s40537-020-00299-5.
- Nti, I.K., Adekoya, A.F., & Weyori, B.A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53(4), 3007–3057. https://doi.org/10.1007/s10462-019-09754-z.
- Park, C.-H., & Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys, 21*(4), 786–826. https://doi.org/10.1111/j.1467-6419.2007.00519.x.
- Peachavanish, R. (2016). Stock selection and trading based on cluster analysis of trend and momentum indicators. *Lecture Notes in Engineering and Computer Science*, 1, 317–321. https://www.iaeng.org/publication/IMECS2016/IMECS2016_pp317-321.pdf.
- Phitthayanon, C. and Rungreunganun, V., (2019). Material cost prediction for jewelry production using deep learning technique. Engineering Journal, 23(6), 145-160.
- Prachyachuwong, K., & Vateekul, P. (2021). Stock trend prediction using deep learning approach on technical indicator and industrial specific information. *Information*, 12(6), 250. https://doi.org/10.3390/info12060250.
- Rahman, A. S. A., Abdul-Rahman, S., & Mutalib, S. (2017). Mining textual terms for stock market prediction analysis using financial news. In A. Mohamed, M. Berry, & B. Yap (Eds.), *Soft Computing in Data Science* (pp. 293–305). Springer. https://doi.org/10.1007/978-981-10-7242-0_25.

- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PLOS ONE, 10*(9), e0138441. https://doi.org/10.1371/journal.pone.0138441.
- Rasekhschaffe, K. C., & Jones, R. C. (2019). Machine Learning for Stock Selection. *Financial Analysts Journal, 75*(3), 70–88. https://doi.org/10.1080/0015198X.2019.1596678.
- Rezaei, H., Faaljou, H., & Mansourfar, G. (2021). Stock price prediction using deep learning and frequency decomposition. *Expert Systems with Applications, 169*, 114332. https://doi.org/10.1016/j.eswa.2020.114332.
- Saravanos, C., & Kanavos, A. (2024). Forecasting stock market volatility using social media sentiment analysis. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-024-10807-w.
- Schnell, J., Nentwich, C., Endres, F., Kollenda, A., Distel, F., Knoche, T., Reinhart, G. (2019) "Data mining in lithium-ion battery cell production". Journal of Power Sources: 413, 360-366.
- Sedighi, M., Jahangirnia, H., Gharakhani, M., & Farahani Fard, S. (2019). A novel hybrid model for stock price forecasting based on metaheuristics and support vector machine. *Data, 4*(2), 75. https://doi.org/10.3390/data4020075.
- Sharma, D. K., Hota, H. S., Brown, K., & Handa, R. (2021). Integration of genetic algorithm with artificial neural network for stock market forecasting. *International Journal of System Assurance Engineering and Management*. https://doi.org/10.1007/s13198-021-01209-5.
- Sirimevan, N., Mamalgaha, I. G. U. H., Jayasekara, C., Mayuran, Y. S., & Jayawardena, C. (2019). Stock Market Prediction Using Machine Learning Techniques. In *Proceedings of the 2019 IEEE International Conference on Advancements in Computing (ICAC)* (Vol. 1, pp. 192–197). IEEE. https://rda.sliit.lk/handle/123456789/476rda.sliit.lk.
- Smith, S., O'Hare, A. Comparing traditional news and social media with stock price movements; which comes first, the news or the price change?. J Big Data 9, 47 (2022). https://doi.org/10.1186/s40537-022-00591-6.
- Tan, T. Z., Quek, C., & Ng, G. S. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction. Computational Intelligence, 23(2), 202–223.

- Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2021). Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods. *Cognitive Computation*, 14, 372–387. https://doi.org/10.1007/s12559-021-09819-8.
- Venkatesh, C. K., & Tyagi, M. (2011). Fundamental analysis as a method of share valuation in comparison with technical analysis. *Bangladesh Research Publications Journal, 5*(3), 167–174.
- Wang, S., Bai, Y., Fu, K., Wang, L., Lu, C.-T., & Ji, T. (2023). ALERTA-Net: A Temporal Distance-Aware Recurrent Networks for Stock Movement and Volatility Prediction. *arXiv preprint arXiv:2310.18706*. https://arxiv.org/abs/2310.18706.
- W. Li and G. S. Bastos. (2020) "Stock market forecasting using deep learning and technical analysis: A systematic review," *IEEE Access*, vol. 8, pp. 185232–185242, 2020, doi: 10.1109/ACCESS.2020.3030226.
- Wessel, J., Turetskyy, A., Wojahn, O., Herrmann, C., Thiede, S. (2020) "Tracking and Tracing for Data Mining Application in the Lithium-ion Battery Production". Procedia CIRP: 93, 162-167.
- Yadav, A., Kumar, V., Singh, S., & Mishra, A. K. (2024). A Novel Approach for Forecasting Price of Stock Market Using Machine Learning Techniques. *SN Computer Science, 5*(3), Article 257. https://doi.org/10.1007/s42979-024-02916-z.
- Yu, P., & Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, 32(6), 1609–1628. https://doi.org/10.1007/s00521-019-04212-x.
- Zhang, Y., Zhang, A., Liu, L., Shen, D., (2017) "The interaction of financial news between mass media and new media: Evidence from news on Chinese stock market" Physica A 486 (2017) 535-541

Appendix A:

Ethics Application

Application ID : ETH23-8429

Application Title :

Date of Submission : Harnessing the Power of AI in Investigating Trend of Lithium-ion Batteries Related Stock Market to Support the Sustainable Clean

Energy Future

Primary Investigator : Associate Professor Eila Erfani (Chief Investigator)

Other Personnel : Mr Ahmad Kianrad (Chief Investigator)

Professor Amir H Gandomi (Chief Investigator)

Dr Karim Kaveh Hasani (Chief Investigator)

Associate Professor Amir Razmjou Chaharmahali (Chief Investigator)

Section 1: Ethics Portal

Select your application type

What type of application are you looking for?

Please do not change your application type without first consulting with the Ethics Secretariat (9514 9772).*

New application (including scope-checking for nil/negligible risk research)

Ratification of existing approval

Transfer of existing approval

Evaluation of teaching and learning activities

Amendment to existing approval

Program approval

You have selected "new application (including scope checking for nil/negligible risk research)". This option allows you to create a new form. The system will check if your application can be approved by the Faculty or whether it requires full ethics approval by the HREC. Please click "save" before continuing.

What should I know before I start?

Would you like more information on:

This system

The ethics process

Purpose of the ethics review process This question is not answered.

Section 1A: Risk evaluation

Risk A

Determining the level of risk and review

Please answer each question carefully and consecutively.

For assistance with answering these questions please refer to the National Statement on Ethical Conduct in Human Research as per the chapters listed below. If you need to contact the Research Ethics Officer you can call (02) 9514 9772

Click on the help buttons (?) for more information

You can save your application at any time by clicking on the save button on the left hand side in the toolbar. For further information and help in completing your application go to our website.

Does your research involve:

Projects involving covert observation, active concealment, or planned deception of participants

e.g. covert observation of the hand-washing behaviour of hospital employees, undisclosed role-playing by a researcher, etc. Does NOT include observation in a public place WITHOUT the use of photographs, images, video or audio footage (Chapter 2.3, p.19)

*

Yes

No

Targeted recruitment or analysis of data(?) from any of the groups listed below (or where any of these groups are likely to be significantly over-represented in the group being studied)

Women who are pregnant and the human fetus (Chapter 4.1, p. 61) Children and young people (under 18 years) (Chapter 4.2, p. 65)

People in dependent or unequal relationships (e.g. lecturer/student [except T&L], doctor/patient, employer/employee) (Chapter 4.3, p.68) People highly dependent on medical care who may be unable to give consent Chapter 4.4, p.68)

People with a cognitive impairment, an intellectual disability, or a mental illness (may include the disadvantaged/homeless) (Chapter 4.5, p. 70) People who may be involved in illegal activities (including those affected e.g. victims of domestic violence) (Chapter 4.6, p.73) Aboriginal and Torres Strait Islander Peoples (Chapter 4.7, p.77)

*

Yes

No

Targeted recruitment of people in / from countries that score <50 on the Corruption Perception Index (CPI) (check here)

This includes any cohorts from these countries, i.e. it is not restricted to marginalised groups within these countries*

Yes No Collection, use or disclosure of personal information without consent of the participant(?)

a record which may include your name, address and other details about the participant (e.g. date of birth, financial information etc.) photographs, images, video or audio footage fingerprints, blood or DNA samples

Yes

No

Collection, use or disclosure of health information(?)

personal information that is information or an opinion about the physical or mental health or a disability (at any time) of an individual; or

an individual's expressed wishes about the future provision of health services to him or her, or a health service provided, or to be provided, to an individual or other personal information collected to provide, or in providing, a health service, or

other personal information about an individual collected in connection with the donation, or intended donation, of a individual's body parts, organs, body substances, or other personal information that is genetic information about an individual arising from a health service provided to the individual in a form that is or could be predictive of the health (at any time) of the individual or of a genetic relative of the individual, or healthcare identifiers

N.B Includes information collected through physiological testing or assessment. Examples include but are not limited to EEG, EMG, BMI, blood pressure, DEXA, etc.*

Yes

No

Collection, use or disclosure of sensitive information

Racial, ethnic information, political, religious and philosophical beliefs, sexual activity or identity, and trade union membership *

Yes

No

Activity that potentially infringes the privacy or professional reputation of participants, providers or organisations e.g. observation in the workplace, collection of commercially confidential information, etc.

Commercially confidential information = Any information which is not in the public domain or publicly available, and where disclosure may undermine the economic interest or competitive position of the owner of the information (TGA adopted definition from European Medicines Agency (EMA)).

N.B. if canvassing opinion via consensus methods i.e. Delphi (?), answer “No” here

*

Yes

No

Establishment of a register or databank of identifiable data for possible use in future research projects (Chapter 3.2, p.27) (?)

*

Yes No

Collection, transfer(?) and/or banking of human biospecimens. e.g. tissue, blood, urine, sputum etc.(?)

*

Yes

No

Any significant alteration to routine care or service provided to participants e.g. deviation from standard care or usual practice *

Yes

No

Prospective assignment of human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes(?)

(Chapter 3.14-3.17) *

Yes

No

Potential for participants to experience harm (i.e. anything more than discomfort)(?)

e.g. physical, psychological, devaluation of personal worth, social, economic and/or legal (Chapter 2.1, p.12) *

Yes No

Risks B

For assistance with answering these questions please refer to the National Statement on Ethical Conduct in Human Research as per the chapters listed below.

Collection, use or disclosure of personal information (?) WITH consent of the participant e.g. surveys, interviews, focus groups, publication, etc.

N.B. if administering anonymous surveys and you have the participant's contact information, answer "Yes" here N.B. if canvasing opinion via consensus methods i.e. Delphi (?), answer “No” here

*

Yes

No

Involves direct communication or contact with participants (e.g. UTS staff/students, patients, consumers, professionals, industry partners or members of the public, etc.)

e.g. interviews or focus groups, etc (face-to-face, via phone, email, online etc.)

N.B. if canvasing opinion via consensus methods i.e. Delphi (?), answer "No" here

*

Yes

No

Involves participants who have a pre-existing relationship with the researcher(s) and/or those involved with recruitment and/or consent e.g. relative, friend, co-worker, lecturer/student [T&L].

N.B. if canvasing opinion via consensus methods i.e. Delphi (?), answer "No" here

*

Yes

No

People unable to give free informed consent due to difficulties in understanding the Information Sheet or Consent Form. e.g. language difficulties

*

Yes

No

People in other countries that score >50 on the Corruption Perception Index (CPI) (check here) Chapter 4.8, p.80

N.B. if canvasing opinion via consensus methods i.e. Delphi (?), answer “No” here

*

Yes

No

Low Risk

Section 2: Project information

Project title

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Application ID (automatically generated):

ETH23-8429

Application Title:*

Harnessing the Power of AI in Investigating Trend of Lithium-ion Batteries Related Stock Market to Support the Sustainable Clean Energy Future

Please note that the HREC is now granting a standard approval period for the research proposals.

The approval period for your project will be specified in your approval letter.

Please also note that research should not commence until ethics approval has been granted. The Committee cannot grant retrospective approval for data that has already been collected.

Ethics category code (automatically selected):*

Human

Is this a resubmission of a previous application?*

Yes

No

Is this a pilot study? *

Yes

No

Has a pilot study been conducted as part of this project? *

Yes

No

Please save and continue to the next page

Consultation

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Have you undertaken any consultation in preparing this application?*

Yes

No

Please describe*

I have discussed this with my supervision team, other PhD candidates and industry experts.

Please save and continue to the next page

Section 3: Personnel

Investigators

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Are there external investigators or personnel listed on this protocol?*

Yes

No

Is this application for a student project?*

Yes No

Student applicants:

1. Please note that once your application is submitted it will go directly to your supervisor and not to the Committee.

2. We strongly recommend notifying your supervisor that you have submitted your application in case of any technical issues, to avoid potential delays in the review process.

3. Once your supervisor endorses your application it will go to your Local Research Office for endorsement before coming to the Ethics Secretariat for review.

4. Your electronic application must be endorsed by your supervisor by the Local Research Office (LRO) submission deadline.

5. Please also ensure that the Primary AOU at the end of this page is updated to your supervisor's AOU. This will show in the table under 'Internal personnel listed below', once you add them. If you need any assistance with this please contact Research.Ethics@uts.edu.au or call 9514 9772. Please note that this is particularly important if you have a dual role as a staff/student as your application could go to the wrong faculty for review through the automated process.

Positions in the personnel table

Position type: In the personnel table use the following positions from the drop-down list: Chief Investigator/Supervisor 1Chief Investigator (students must not be listed as Chief Investigator)

Co Investigator 3Assoc. Investigator

Co Supervisor Co-Supervisor

Research Student 5Research Student

Project Administrator 7Project Administrator

Note: Further options are available in the drop down list.

Instructions on how to add a person to the personnel table:

1. Click on "Add"

2. Start typing the details (first name, last name or Staff ID) in the search bar.

3. Click on "Add selected"

4. The extra information panel will open, select their position from the drop-down list. If they are the primary contact (e.g. Chief Investigator/Supervisor), tick "Yes" under

'Primary contact' and then select "OK"

Student research: Students must add their supervisors to their application and must mark their primary supervisor as a Chief Investigator and as a primary contact. Students must be listed as "5Research student" under the column 'Position' to ensure the application is properly submitted to their supervisor.

Ratifications/Transfers: If this list differs from that of the original application, you must provide evidence that any additional investigators have been added via amendment to the lead/external HREC [attach relevant amendments and evidence of approval].

Internal personnel listed on this ethics protocol:*

1 Primary No

ID

Surname Kianrad

Given Name Ahmad

Full Name Mr Ahmad Kianrad

Position Chief Investigator

Type Domestic

AOU FEIT.School of Computer Science

Managing Unit Faculty of Engineering & Information Technology

Email Address Ahmad.Kianrad@student.uts.edu.au

Work Number

2 Primary

ID [redacted]

Surname Gandomi

Given Name Amir H

Full Name Prof Amir H Gandomi

Position Chief Investigator

Type Internal

AOU FEIT.The Data Science Institute

Managing Unit Faculty of Engineering & Information Technology

Email Address Amirhossein.Gandomi@uts.edu.au

Work Number 5081

3 Primary Yes

ID [redacted]

Surname Erfani

Given Name Eila

Full Name Associate Professor Eila Erfani

Position Chief Investigator

Type Honorary

AOU FEIT.School of Computer Science

Managing Unit Faculty of Engineering & Information Technology

Email Address Eila.Erfani@uts.edu.au

Work Number 1098

If any details are incorrect or missing please contact the Ethics Secretariat on (02) 9514 9772 or by email.

The ResearchMaster database has a very large number of external personnel so please conduct a search for them before adding them in the text box below. Please contact the Ethics Secretariat on 9514 9772 if you cannot find an external investigator through the system.

External personnel listed on this ethics protocol:

*

1 Primary

ID

Surname Hasani

Given Name Karim

Full Name Dr Karim Kaveh Hasani

Position Chief Investigator

Type External

AOU

Managing Unit

Email Address

Work Number

2 Primary

ID

Surname Razmjou Chaharmahali

Given Name Amir

Full Name Dr Amir Razmjou Chaharmahali

Position Chief Investigator

Type External

AOU

Managing Unit

Email Address @uts.edu.au

Work Number

Please provide additional (or preferred) contact details of any of the people listed on the project if necessary (4000 character limit)

Ahmad.Kianrad@student.uts.edu.au

Eila.Erfani@uts.edu.au

Amir H Gandomi <amirhossein.gandomi@uts.edu.au> Amir RAZMJOU CHAHARMAHALI <amir.razmjou@ecu.edu.au>

kaveh.hasani@flinders.edu.au</amir.razmjou@ecu.edu.au></amirhossein.gandomi@uts.edu.au>

Please provide details of any formal qualifications (REF NS 1.1(e)) of each person listed on the project

(4000 character limit)*

Associate Professor Eila Erfani completed her PhD in Information Systems in 2016. She is a Senior Lecturer.

Mr. Ahmad Kianrad is a PhD student in Information Systems. He holds Masters' degree in Information Technology Engineering.

Dr Amir Razmjoo Chaharmahali completed his PhD in 2012. He is a Lecturer.

Dr Amir Hossein Gandomi completed his PhD in 2015. He is a Professor Of Data Science.

Please outline the experience of each person listed on this project relevant to this application (4000 character limit)*

Associate Professor Eila Erfani:

Eila pursues research and teaching on digital transformation, data and business analytics, and digital ethics. She is fascinated by the socio-technical impacts of digital technologies and data-based services on individuals, society and the environment. She has diverse experience in academic leadership and education, most recently as Senior Lecturer and Deputy Head of School (Teaching and Learning) at the School of Information Systems and Modelling, University of Technology Sydney (UTS). She contributed to the design and delivery of undergraduate, and postgraduate programs online short courses, micro-credentials, and Graduate Certificates. She is the winner of the teaching excellence award.

Associate Professor Amir Razmjou:

Associate Professor Razmjou is Board Director of the Membrane Society of Australasia (MSA) and an editorial board member of the Journal of Desalination. His current research areas are: Direct Lithium Extraction Technologies, Ion-selective membranes, Material discovery using artificial intelligence and machine learning, Sensors, Controlled resale of ionic drugs and PFAS removal technologies. Dr Amir Hossein Gandomi:

Professor Amir H. Gandomi is among the world's most-cited researchers for his work in the fields of global optimisation and big data analytics, in particular, using machine learning and evolutionary computations. He has served as associate editor, editor and guest editor in several prestigious journals, such as AE of IEEE TBD, IEEE Networks, and IEEE IoTJ. He regularly delivers keynote addresses at major conferences.

Primary AOU*

FEIT.School of Computer Science

Managing Unit

Faculty of Engineering & Information Technology

Please save and continue to the next page

Student details

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Degree being undertaken (500 character limit)*

PhD

Have you been successful in your doctoral/masters assessment? *

Yes

No

Please make sure you attach a copy of your Stage 1 confirmation (or the stage 1 panel's report) in the attachments section.

Students, please read carefully: Once you have completed this application and followed the submission instructions, your application will go to your supervisor for review.

Once your supervisor has reviewed and endorsed your application it will come to the Ethics Secretariat for a pre-review. This pre-review process helps ensure that your application is complete, has all necessary attachments, and that the quality of responses to the questions meets the Committee's expectations. Your application should therefore be submitted as early as possible. If you do not submit your application in time, it may be delayed and held off until the next closing date.

Section 4: Funding and Disclosure of interests

Funding details

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Have you received funding in relation to this research?*

Yes

No

Do you intend to apply for funding in the future?*

Yes

No

Please save and continue to the next page

Disclosure of Interests

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

Do any members of the research team (including persons not listed in this application), have any financial or non-financial interests related to this research?*

Yes

No

Please save and continue to the next page

Section 5: Methodology

Description

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

The purpose of this section is to place your research in context for the HREC and demonstrate your ability to conduct the research. The HREC may only approve research which is methodologically sound. Remember to use simple language that can be understood by people from a variety of backgrounds. Avoid jargon and acronyms.

What are the hypotheses/goals/aims/objectives of your research? Please include a brief description using plain English explaining your research aims (approximately 100 words)

(4000 character limit)*

Providing Insights that support stakeholders of lithium-ion batteries related companies in their informed decision making.

Investigating the impacts of various factors on the lithium-related stock market in Australia.

Conducting an empirical and evidence-based study on forecasting lithium-related stock market.

Scattered data from various sources such as historical stock price data, traditional news data and data from social media will be investigated. This study will help stakeholders, policymakers, and investors of lithium-related stock market in Australia make informed decisions with societal benefits.

Note: Clinical Trials, Recruitment of Participants and Data Collection are dealt with later in the application so you do not need to describe them in detail below

Please provide a brief description of the research design including research questions and proposed methods for conducting the research (approximately 250 words) (4000 character limit)*

Research Questions:

RQ1: What is the trend of lithium-ion batteries related stock market in Australia?

RQ2: How and to what extent do traditional news data impact the trend of lithium-ion batteries related stock market in Australia? RQ3: How and to what extent do social media data impact the trend of lithium-ion batteries related stock market in Australia?

Research methodology:

As stock market is naturally volatile, it is very difficult to predict stock trends by using traditional statistical and regression methods and time series approaches, so new methodologies and more complicated algorithms need to be used. One of the popular machine learning techniques used by researchers to predict the stock trends is Support Vector Machine (SVM). After feature extraction and pre-processing, we obtain a dataset containing the several indicators as features and price as label. Then, we randomly split the dataset into training and test sets. Using the training set, we train our empirical models and forecast the test set. In this study, we used two empirical models of OLS regression and XGBoost decision tree. OLS stands for ordinary linear regression model and is one of the traditional statistical methods and XGBoost means extreme gradient boosting and is one of best-performing machine learning algorithms. It forms many decision trees to forecast but then integrates many weak trees to form a stronger tree.

Moreover, for more detailed analysis and estimation, we assigned companies into three categories, based on the capital market. Companies with capital market less than two billion dollars (According to ASX) are labelled as small cap, companies with their capital market between two and five billion dollars are labelled as medium cap, and companies with capital market above five billion dollars are categorised as large cap. We ran our linear regression and XGBoost models for all three categories separately and compared the results. The following Table shows the details of the news categories.

What do you hope the outcome(s) of this research will be? (4000 character limit)*

This study adds to the growing body of research that aims to leverage machine learning to improve investment strategies and support informed decision-making in financial markets. This study provides insights into how these techniques can be applied to other markets and assets. Ultimately, continued research in this area will help investors, stakeholders, and company managers make more informed decisions and achieve better investment outcomes.

Who do you think will benefit from this research? (4000 character limit)*

Investors, stakeholders, and company managers and policymakers of lithium-related stock market in Australia.

Please provide a brief description of the significance of your research (approximately 100 words)

(4000 character limit)*

This study will help stakeholders, policymakers, and investors of lithium-related stock market in Australia make informed decisions with societal benefits.

Please save and continue to the next page

Literature review & references

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Please give a brief literature review. The aim is to explain how your research fits

into the context of other research in the area (REF NS 1.1(c)) (4000 character limit with spaces) Please note that you cannot paste links into the online form

*

Climate change and deterioration of our natural environment are critical problems. Studies have shown that the use of energy sources with less carbon emission plays an important role in environmental sustainability and protection. On the other hand, electrical devices (e.g., mobiles and laptops) that use lithium batteries as their energy source are widely used in everyday life. The demand for producing lithium batteries that are environmentally friendly has increased.

Therefore, providing insights that support lithium market stakeholders in making informed decision is vital. The Industrial Revolution led to a surge in greenhouse gas (GHG) emissions, which have become a global concern since the latter half of the 20th century. This has contributed to the urgent need for GHG reduction as the world moves towards a cleaner and more sustainable energy future. Identifying trends in the market for renewable energy materials and technologies is crucial to achieving this goal.

Lithium-ion batteries are a critical technology for a cleaner energy future as they provide a sustainable means of storing electrical energy from renewable sources like wind and solar power. They also have a lower environmental impact compared to traditional fossil fuels, with longer lifespans, less maintenance, and fewer greenhouse gas emissions during their lifecycle (Zubi et al. 2018). Additionally, recycling, and reusing lithium-ion batteries can further reduce their environmental impact.

Identifying trends in the market for lithium-ion batteries is significant for investors to make informed decisions on where to allocate their resources. By analyzing price changes, investors can identify which lithium-ion battery technologies are likely to experience growth in the coming years, allowing them to allocate resources more effectively. It can also help policymakers and industry leaders make more informed decisions about where to invest in research and development.

Leveraging data analytics and machine learning algorithms is a promising approach to identifying patterns in lithium-ion battery prices, facilitating informed decision-making and accelerating the transition towards sustainable energy. Applying machine learning techniques in predicting stock prices in financial markets is advantageous compared to traditional methods because it can handle vast amounts of complex data and adapt to changing market conditions in real-time (Gururaj et al. 2019), Machine learning algorithms can identify non-linear patterns and correlations in large datasets, providing accurate and reliable predictions that improve over time (Yun et al. 2021).

This study will harness the power of AI to provide evidence and insight for the trend of lithium related stock market in Australia. In addition, it will investigate the impact of various factors that might potentially impact the trend. This study has practical contribution and will support different stakeholders of lithium related stock market in making data-driven decisions.

Please list the references only used in the literature review and cited in your application

NOTE: Do not include references you have not used in this application (4000 character limit)

*

Gururaj, V., Shriya V.R., Ashwini K., 2019. “Stock Market Prediction using Linear Regression and Support Vector Machines,” International Journal of Applied Engineering Research ISSN 0973-4562, Volume 14, pp. 1931-1934.

Yun, K. K, Yoon, S. W., Won, D., 2021. “Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process,” Expert Systems with Applications, Volume 186, p.115716.

Zubi, G., Dufo-López, R., Carvalho, M., Pasaoglu, G., 2018. “The lithium-ion battery: State of the art and future perspectives,” Renewable and Sustainable Energy Reviews, Volume 89, pp. 292-308.

Please save and continue to the next page

Methods and methodologies

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

In order to consider your research, the HREC will need to know what it will involve for your participants (REF NS 3.1)

What kinds of methods and methodologies will you use in your research? (More than one box may be checked)*

Quantitative

Qualitative

Does your research involve collection and/or use of secondary data? (e.g. existing / routinely collected data etc.)*

Yes

No

Please save and continue to the next page

Qualitative

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

What qualitative methodology and methods will be using in this research?

Section 1: Qualitative methodology*

Auto-ethnography

Historical research

Action research

Narrative enquiry

Biographical research

Case study

Phenomenology

Indigenous research paradigm

Discourse analysis Grounded theory

Other *(Please describe below)

Please describe other methodologies*

Interview with experts

Section 2: Qualitative methods*

Participants observation

Covert observation

On-line research

Psychological testing/assessment

Verbal protocol

Journaling

Artifact analysis

Document/Policy analysis

Access to records

Audio/video recording

Life story or oral history

Focus groups

Structured interviews

Semi-structured interviews

Unstructured interviews

Other * (Please describe below)

Please describe how interviews will be conducted, including how many participants will be involved (from each participant group if there is more than one group/cohort), the amount of time required of participants for this, whether it will be recorded, and any other information applicable (4000 character limit)*

We will conduct interviews until saturation point of data

Please save and continue to the next page

Section 6: Research participants/subjects part 1

Participant involvement

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

What time commitment will the research involve for your participants?

NOTE: This information must be included in any information to participants

(4000 character limit)*

We will interviews for about 45 minutes.

In what location will the research/data collection take place?

NOTE: This information must be included in any information to participants (4000 character limit)*

Online

What travel, if any, does the research involve for your participants?

NOTE: This information must be included in any information to participants (4000 character limit)*

NA

Please include any additional information relating to participants that you think relevant

NOTE: This information must be included in any information to participants (4000 character limit)*

NA

Describe and justify any benefit, payment or compensation the participants will receive. For research being conducted with Aboriginal and Torres Strait Islander People, the described benefits from research should have been discussed with and agreed to by the Aboriginal or Torres Strait Islander research stakeholders. (REF NS 2.1) and 4.7.8 & 4.7.9) (4000 character limit)*

NA

Please save and continue to the next page

Recruitment of participants

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

In line with the National Statement, the definition of participants includes not only those humans who are the primary focus of the research but also those who will be affected by the research.

The HREC regards the principle of respect for persons as of paramount importance. (REF NS 1.1 (d), 1.6-1.9, 1.10, 2.1).

How will you initially select and contact your participants? More than one box may be checked, if appropriate*

Advertisement/flyer

E-mail

Telephone

Internet

Organisation

Personal contact

Letter

Other contact method to be used

Outline how you will obtain participants' contact details*

Through organizational website and my professional network and my supervision team

Please describe your recruitment plan/strategy*

I will approach participants (experts) via email and after achieving the consent based on their preference I will conduct online interviews.

How many participants do you intend to recruit? (If you are intending to recruit different groups of participants, please answer all relevant questions for each group, e.g. control group, test group, etc) (4000 character limit)*

12-25

Explain how and why you have chosen this number. If the research is quantitative, explain the power calculations; if the research is qualitative, explain why the proposed number is likely to result in adequate data based on evidence/literature. For guidance, check our Fact Sheets.*

Based on Erfani et al 2015, we will reach to saturation point of data by conducting interview with 12-25 experts

Describe your inclusion and exclusion criteria for participants*

Inclusion criteria: Expertise in Machine Learning and analysis of structured and semi-structured data

Please save and continue to the next page

Consent

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Will you be obtaining written consent?*

Yes

No

Please provide sample documents in attachments list at the end of the application form

Please use the following HREC templates when creating an information sheet and consent form: HREC templates

Do you believe there will be any special issues relating to consent in your research? (REF NS 1.13, 2.2, 2.3, Chapter 4)*

Yes

No

Are the participants able to consent fully? (REF NS Chapter 2, 4.4, 4.5)*

Yes

No

Please save and continue to the next page

Limited disclosure

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

Does this research involve limited disclosure to participants? (REF NS 2.3)*

Yes

No

Please save and continue to the next page

Ethical considerations specific to participants

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

Indicate if your research will involve the following populations (as per the National Statement) other than as incidental participants

(i.e. they are not included in the design of the project but may be participants) (REF NS Chapter 4)

*

Women who are pregnant and the human foetus

Children and young people

People in dependent or unequal relationships

People highly dependent upon medical care who may be unable to give consent

People with a cognitive impairment, an intellectual disability or a mental illness

People who may be involved in illegal activities

People who are incarcerated

Aboriginal and Torres Strait Islander Peoples

People in other countries

None of the above

If your research is being conducted in Australia, does it involve Culturally and Linguistically Diverse (CALD) People (other than incidentally)?*

Yes

No

Does your research involve Defence or the Department of Veteran Affairs in any way?*

Yes

No

Please save and continue to the next page

Section 7: Research participants/subjects part 2

Risk/harm

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

Risk or harm could be described as damage or hurt to the wellbeing, interests or welfare of an individual, institution or group. Harm could range from physical hurt or damage such as illness or injury, to psychological or emotional hurt or damage, such as embarrassment or distress. Please note that as a researcher, you are not necessarily immune from risk yourself and should give careful consideration to this question (REF NS 2.1).

NOTE:

It is really important that you carefully consider all potential risks that could occur, even if they seem negligible.

Please do not provide one-word answers to any of the questions below.

Describe, as best as you can, any possible risks to research participants, subjects and related groups

NOTE: This information must be included in any information to participants (4000 character limit)*

Interviews will be conducted online in a secure environment

How would you categorise the magnitude of potential risk? (e.g. inconvenience, discomfort, harmful, painful)

Explain why you believe this is so (4000 character limit)*

The questions posed to participants revolve around their experience with machine learning algorithms. The estimated risk primarily pertains to inconvenience due to the time participants spend providing responses to these questions.

How would you categorise the likelihood of risk? (i.e. slight, possible, likely, probable, unavoidable)

Explain why you believe this is so (4000 characters)*

For experts with machine learning expertise, the risk is relatively low since the questions do not require them to reflect on issues that they would not have otherwise thought about. However, there is still a potential inconvenience related to the time commitment required to provide thoughtful and accurate responses. Participants may need to allocate additional time, which could interfere with their regular schedules and other commitments.

What strategies will you use to minimise and/or manage the risks? (4000 character limit)*

To mitigate the inconvenience related to time commitment, researchers can employ several strategies. Firstly, pre-screening participants based on their areas of expertise allows for tailored questionnaires, focusing on relevant topics and avoiding unrelated questions. This customization minimizes the time required for responses. Secondly, clear communication regarding the estimated time commitment before participation helps participants plan and prepare accordingly. Lastly, providing flexible response options, like allowing participants to answer the question later, accommodates their schedules and reduces the pressure of rushing through the task.

Discuss likely or possible risk to researchers (including yourself), and your strategies for minimising such risks (4000 character limit)*

There are no foreseen risks to researchers in this project. All data collection will be conducted remotely, eliminating the need for researchers to travel to alternate locations such as participant homes or places of employment. Consequently, the risk to researchers is considered negligible.

Please save and continue to the next page

Pre-existing relationships

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

Are there likely to be any pre-existing relationships with research participants? (e.g. employer/employee, colleague, friend, relation, student/teacher, etc) (REF NS 4.3)*

Yes

No

Will you be recruiting UTS staff and/or students as research participants?*

Yes

No

Please save and continue to the next page

External organisations

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

Indicate if your research will involve any of the following:*

Institution

Organisation

Community Group

None of the above

Please save and continue to the next page

Section 8: Data

Data collection & use

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

Does your research involve access to student records at this University?*

Yes

No

Provide an analysis plan outlining how the aims/objectives will be met, the statistical methods to be used, and who will be carrying out the analysis. *

I will extract the stock market data of lithium-related companies in Australia and will utilize machine learning to investigate the impacts of historical stock data, traditional data and social media data on trend of stock market in Australia.

Describe any foreseeable future use of this data; such as sharing with other researchers, secondary use for related research, publishing for unrelated research and non-research purposes and any other possible uses. Please note this information must be included in the participant information sheet.*

NA

Do you have a research data management plan?*

Yes

No

Please save and continue to the next page

Section 9: Additional information

Other ethical issues

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button. For further information and help in completing your application go to our website

If there are any additional ethical issues which you do not believe have been covered by this form, please explain them for the HREC: (4000 character limit)*

NA

Please save and continue to the next page

Section 10: Attachments

Attachments

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.

For further information and help in completing your application go to our website

I have attached the following supporting documents

Doctoral or Masters assessment (i.e. panel's report)*

Yes

N/A

Sample research advertisement/announcement

Yes

N/A

Participant Information Sheet(s)*

Yes

No

Informed consent form(s)*

Yes

N/A

Explanations of any technical terms used *

Yes

N/A

Research data management plan (RDMP)*

Yes

No

Standard Operating Procedures

N.B. May include a distress or disclosure protocol [see UTS HREC Disclosure Guidelines under University policies and guidelines], Faculty of Health - Low Risk protocol; procedures for participant screening, physiological, or biological sampling and/or laboratory or safety procedures where relevant.

*

Yes

No

Please explain why any of the above items have not been attached (either softcopy/hardcopy) and when they will be provided (4000 character limit)*

Those items are not applicable

Documents attached to this application:

How to attach documents

1. Click on 'Add'

Ensure the fields are as follows:

Document type- soft copy

Name: Include the document name and version number Description: This field is optional

2. You can then either select the file you want to upload OR drag and drop it where it says 'Drop file here'

3. Click on 'OK'

Note: Please use the following HREC templates when creating an information sheet, consent form, verbal script, etc.: HREC templates. All submitted documents should be titled, and have have version control included in the footer.*

1 Document type Soft copy

Name CA1 Form_AhmadKianrad

Reference (Document Title) CA1 Form_AhmadKianrad.pdf

Description CA1 Form Ahmad Kianrad

2 Document type Soft copy

Name DataManagementPlan_Stash_AhmadKianrad

Reference (Document Title) DataManagementPlan_Stash_AhmadKianrad.pdf

Description DataManagementPlan AhmadKianrad

3 Document type Soft copy

Name HREC-Template Participant information sheet-consent form - 07 Oct 2022(3)_AhmadKianrad

Reference (Document Title) HREC-Template Participant information sheet-consent form - 07 Oct 2022(3)_AhmadKianrad.docx

Description

Reminder to student applicants:

1. Please note that once your application is submitted it will go directly to your supervisor and not to the Committee.

2. We strongly recommend notifying your supervisor that you have submitted your application in case of any technical issues, to avoid potential delays in the review process.

3. Once your supervisor endorses your application it will go to your Local Research Office for endorsement before coming to the Ethics Secretariat for review.

4. Your electronic application must be endorsed by your supervisor by the Local Research Office (LRO) submission deadline.

5. Please also ensure that the Primary AOU listed at the end of the Investigators page is updated to your supervisor's AOU. This will show in the table under 'Internal personnel listed below', once you add them. If you need any assistance with this please contact Research.Ethics@uts.edu.au or call 9514 9772. Please note that this is particularly important if you have a dual role as a staff/student as your application could go to the wrong faculty for review through the automated process.

Declaration

Declaration

I have answered all questions in the risk assessment truly and completely to the best of my knowledge

I will notify the UTS Human Research Ethics Committee of any variation to this research that may alter the level of risk associated with it

This research will be undertaken in compliance with the UTS Research Policy or any replacement or amendment thereof

This research will be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research

Please click on the "Submit" button in the Actions menu.

Confirmation

Confirmation by Local Research Office Low Risk

Application type*

Research (student project)

Internal personnel listed on this ethics protocol*

1 Primary No

ID

Surname Kianrad

Given Name Ahmad

Full Name Mr Ahmad Kianrad

Position Chief Investigator

Type Domestic

AOU FEIT.School of Computer Science

Managing Unit Faculty of Engineering & Information Technology

Email Address Ahmad.Kianrad@student.uts.edu.au

Work Number

2 Primary

ID

Surname Gandomi

Given Name Amir H

Full Name Prof Amir H Gandomi

Position Chief Investigator

Type Internal

AOU FEIT.The Data Science Institute

Managing Unit Faculty of Engineering & Information Technology

Email Address Amirhossein.Gandomi@uts.edu.au

Work Number 5081

3 Primary Yes

ID

Surname Erfani

Given Name Eila

Full Name Dr Eila Erfani

Position Chief Investigator

Type Honorary

AOU FEIT.School of Computer Science

Managing Unit Faculty of Engineering & Information Technology

Email Address Eila.Erfani@uts.edu.au

Work Number 1098

External personnel listed on this ethics protocol*

1 Primary

ID

Surname Hasani

Given Name Karim

Full Name Dr Karim Kaveh Hasani

Position Chief Investigator

Type External

AOU

Managing Unit

Email Address

Work Number

2 Primary

ID [redacted]

Surname Razmjou Chaharmahali

Given Name Amir

Full Name Dr Amir Razmjou Chaharmahali

Position Chief Investigator

Type External

AOU

Managing Unit

Email Address [redacted]@uts.edu.au

Work Number

Confirmed by:*

James Tan

Faculty reference number if applicable (e.g. faculty TRIM number, otherwise write "N/A")

n/a

Please select "Local Research Office review" from the dropdown list.*

FEIT research manager

Date of review*

01/08/2023

The Local Research Office has certified that:

This application has been deemed low risk by the Faculty/School/Centre in accordance with the National Statement on Ethical Conduct in Human Research (Chapter 2.1).

All information in this application and supporting documentation is correct and as complete as possible.

This research will be undertaken in compliance with the UTS Research Ethics and Integrity Policy or any replacement or amendment

This research will be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research.

Yes

No

Confirmation by ADR

Application type

Human

Internal personnel listed on this ethics protocol

1 Primary No

ID

Surname Kianrad

Given Name Ahmad

Full Name Mr Ahmad Kianrad

Position Chief Investigator

Type Domestic

AOU FEIT.School of Computer Science

Managing Unit Faculty of Engineering & Information Technology

Email Address Ahmad.Kianrad@student.uts.edu.au

Work Number

2 Primary

ID

Surname Gandomi

Given Name Amir H

Full Name Prof Amir H Gandomi

Position Chief Investigator

Type Internal

AOU FEIT.The Data Science Institute

Managing Unit Faculty of Engineering & Information Technology

Email Address Amirhossein.Gandomi@uts.edu.au

Work Number 5081

3 Primary Yes

ID

Surname Erfani

Given Name Eila

Full Name Associate Professor Eila Erfani

Position Chief Investigator

Type Honorary

AOU FEIT.School of Computer Science

Managing Unit Faculty of Engineering & Information Technology

Email Address Eila.Erfani@uts.edu.au

Work Number 1098

External personnel listed on this ethics protocol

1 Primary

ID

Surname Hasani

Given Name Karim

Full Name Dr Karim Kaveh Hasani

Position Chief Investigator

Type External

AOU

Managing Unit

Email Address

Work Number

2 Primary

ID

Surname Razmjou Chaharmahali

Given Name Amir

Full Name Dr Amir Razmjou Chaharmahali

Position Chief Investigator

Type External

AOU

Managing Unit

Email Address @uts.edu.au

Work Number

Faculty reference number if applicable (e.g. faculty TRIM number) n/a

Declaration:

I am aware that this research is being conducted within this Faculty/School/Centre.

I am satisfied that the researchers have met all Faculty/School/Centre requirements in relation to this research

I am satisfied that this research has been deemed low risk by the Faculty/School/Centre in accordance with the National Statement on Ethical Conduct in Human Research

(Chapter 2.1).

This research will be undertaken in compliance with the UTS Research Ethics and Integrity Policy or any replacement or amendment thereof.

This research will be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research.

Yes

No

Date Approved/Endorsement*

20/08/2023

Comments

This question is not answered.