PREDICTING COMPANY VALUATION USING MACHINE LEARNING BASED ON COMPANY FUNDAMENTALS AND MACROECONOMIC DATA – THE INDONESIAN CASE

KEVIN WIJAYA

2540124993

PROGRAM PASCASARJANA

PROGRAM STUDI MAGISTER COMPUTER SCIENCE

UNIVERSITAS BINA NUSANTARA

TAHUN 2025

PREDICTING COMPANY VALUATION USING MACHINE LEARNING BASED ON COMPANY FUNDAMENTALS AND MACROECONOMIC DATA – THE INDONESIAN CASE

KEVIN WIJAYA

2540124993

Tesis

Sebagai Salah Satu Syarat untuk Memperoleh Gelar

Magister Computer Science

pada

PROGRAM PASCASARJANA

UNIVERSITAS BINA NUSANTARA

HALAMAN PERNYATAAN

Saya, Kevin Wijaya, 2540124993 menyatakan dengan sebenar-benarnya bahwa tesis saya berjudul "Predicting Company Valuation using Machine Learning based on Company Fundamentals and Macroeconomic Data – The Indonesian Case" adalah merupakan gagasan dan hasil penelitian saya sendiri dengan bimbingan Dosen Pembimbing.

Saya juga menyatakan dengan sebenarnya bahwa isi tesis ini tidak merupakan jiplakan dan bukan pula dari karya orang lain, kecuali kutipan dari literatur dan atau hasil wawancara tertulis yang saya acu dan telah saya sebutkan di Daftar Acuan.

Demikian pernyataan ini saya buat dengan sebenarnya dan saya bersedia menerima sanksi apabila ternyata saya ini tidak benar.

I, Kevin Wijaya, 2540124993 truly acknowledge that my thesis with title " Predicting Company Valuation using Machine Learning based on Company Fundamentals and Macroeconomic Data – The Indonesian Case" is my concept and project result with guidance from supervisor.

I, also truly acknowledge that content of this thesis is not copied and not from another people work, except my citation from literature or written interview result and already write in reference list and bibliography list.

That's my acknowledge were truly made and if in reality this acknowledge weren't true, I am willing sanction.

Jakarta, 03 Februari 2026

Yang menyatakan

Kevin Wijaya

2540124993

Judul Tesis	:	Predicting Company Valuation Using Machine Learning Based on Company Fundamentals and Macroeconomic Data – The Indonesian Case
Nama Mahasiswa	:	Kevin Wijaya
NIM	:	2540124993

Dibimbing oleh:

I Gede Putra Kusuma Negara

D5627

PERNYATAAN

STATEMENT

Dengan ini saya,

Nama	:	Kevin Wijaya
NIM	:	2540124993
Judul Tesis	:	Predicting Company Valuation using Machine Learning based on Company Fundamentals and Macroeconomic Data – The Indonesian Case

Memberikan kepada Universitas Bina Nusantara hak non-eksklusif untuk menyimpan, memperbanyak, dan menyebarluaskan tesis karya saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja, dalam bentuk format tercetak dan atau elektronik.

Menyatakan bahwa saya, akan mempertahankan hak *exclusive* saya, untuk menggunakan seluruh atau sebagian isi tesis saya, guna pengembangan karya di masa depan, misalnya bentuk artikel, buku, perangkat lunak, ataupun sistem informasi.

Hereby grant to my school, Bina Nusantara University, the non-exclusive right to archive, reproduce, and distribute my thesis, in whole or in part, whether in the form of printed and electronic formats.

I acknowledge that I retain exclusive rights of my thesis by using all or part of it in the future work or outputs, such as article, book, software, and information system.

Jakarta, 03 Februari 2026

Kevin Wijaya

2540124993

PRAKATA

Saya merasa bersyukur atas keberhasilan penyusunan proposal tesis sebagai salah satu syarat untuk menyelesaikan studi saya. Dalam proses penulisan, saya menerima banyak bantuan, bimbingan, dan dukungan dari berbagai pihak, yang dengan ini saya sampaikan rasa terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Kedua orang tua saya yang telah memberikan dukungan secara moral maupun finansial kepada saya selama perjalanan akademik ini.
2. Bapak/Ibu I Gede Putra Kusuma Negara selaku dosen pembimbing, yang telah memberikan arahan, masukan, dan ilmu yang sangat berarti dalam penyusunan tesis ini.
3. Rekan-rekan mahasiswa serta teman-teman yang telah memberikan motivasi, diskusi yang bermanfaat, serta dukungan selama penyusunan tesis ini.
4. Seluruh dosen dan staf akademik di Universitas Bina Nusantara, yang telah memberikan wawasan serta fasilitas pendukung selama penelitian ini.
5. Berbagai video YouTube yang telah membangun kompetensi dan ketertarikan saya di bidang ini.
6. Semua pihak lain yang telah membantu, baik secara langsung maupun tidak langsung, dalam penyelesaian tesis ini.

Saya menyadari bahwa proposal tesis ini masih memiliki berbagai kekurangan baik yang disadari maupun tidak. Oleh karena itu, saya menerima dengan terbuka saran dan kritik yang membangun demi penyempurnaannya. Semoga proposal tesis ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan berhasil dikembangkan menjadi tesis yang baik nantinya.

Jakarta, 03 Februari 2026

Kevin Wijaya

2540124993

Abstrak

Penilaian valuasi perusahaan sering kali terhambat oleh bias dalam metode penilaian tradisional, keterbatasan data, dan pengaruh sentimen pasar, yang membuat alokasi modal menjadi tidak efisien. Selain itu, terdapat kekurangan dataset dan penelitian mengenai penilaian berbasis pembelajaran mesin secara menyeluruh di pasar saham Indonesia. Untuk mengatasi kekurangan tersebut, penelitian ini bertujuan mengembangkan kerangka kerja berbasis *machine learning* untuk memprediksi valuasi perusahaan publik di Indonesia dengan mengintegrasikan data fundamental perusahaan dan indikator makroekonomi. Dataset dikompilasi dari 802 perusahaan yang terdaftar di Bursa Efek Indonesia pada periode 1999 Q1 – 2024 Q4, terdiri dari 24 variabel keuangan triwulanan dan 9 indikator makroekonomi tahunan. Model XGBoost, LSTM, dan TFT dilatih serta dioptimalkan *hyperparameter*-nya untuk meningkatkan kualitas hasil *output* model. *Feature importance* dinilai menggunakan mekanisme internal TFT untuk mengetahui variabel keuangan dan makroekonomi yang paling berpengaruh terhadap valuasi. Kinerja model dievaluasi menggunakan metrik *Mean Absolute Percentage Error*. Model dengan kinerja terbaik (TFT) mencapai MAPE sebesar 7.27%. Kapitalisasi pasar masa lalu dinilai sebagai variabel paling penting untuk memprediksi valuasi perusahaan. Variabel makroekonomi memainkan peran suportif penting dengan total nilai *feature importance* 20.44%. Hal ini menunjukkan bahwa menggabungkan variabel makroekonomi dengan data fundamental perusahaan penting untuk dilakukan agar bisa memprediksi valuasi perusahaan secara lebih akurat.

Kata Kunci: Forecasting valuasi, Finansial perusahaan, Variabel makroekonomi, Pembelajaran mesin, Model transformer

Abstract

Company valuation is often hindered by biases in traditional valuation methods, data limitations, and the influence of market sentiment, leading to inefficient capital allocation. There is also a lack of dataset and research about market-wide machine learning-based valuation in the Indonesian stock market. To address this gap, this study aims to develop a machine learning-based framework for predicting the valuation of publicly traded Indonesian companies by integrating company fundamentals with macroeconomic indicators. A comprehensive dataset was compiled from 802 firms listed on the Indonesia Stock Exchange covering the period 1999 Q1 – 2024 Q4, comprised of 24 quarterly financial variables and 9 annual macroeconomic indicators. XGBoost, LSTM, and TFT models were trained and hyperparameters optimized to enhance predictive performance. TFT's internal mechanisms were applied to assess feature importance, revealing the most influential financial and macroeconomic drivers of valuation. Model performance was evaluated using Mean Absolute Percentage Error metrics. The best performing model (TFT) achieved a MAPE of 7.27%. Past market capitalization is regarded as the most important feature by a significant margin in predicting company valuation. Macroeconomic variables play a significant supporting role with a 20.44% feature importance score. These findings highlight that combining macroeconomic variables with firm level fundamentals is necessary to provide a more accurate company valuation prediction.

Key Words: *Valuation forecasting, Company financials, Macroeconomic variables, Machine learning, Transformer model*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1. Background

Company valuation is the value at which investors evaluate the company to be worth. This value might be evaluated differently by different buyers. Ideally, company valuation is done by analyzing various fundamental statistics about the company, looking at valuations of companies in the same sector, considering various economic data and business climates, and comparing all of them to their historic values. However, traditional methods of valuation always contain biases due to personal opinions, institutional factors, news, etc. (Damodaran, 2024). In addition to that, most retail investors lack the foresight, patience, and knowledge necessary to do this evaluation on their own. They typically rely on surface-level statistics and news about companies to make decisions on where to invest their money (Tahir & Danarsari, 2023). Especially now that markets are more easily manipulated by using financial influencers and the prevalence of so-called "pump and dump schemes" (Li et al., 2021). Even institutional investors are prone to making hasty short-term emotionally driven decisions in their investment strategies. These irrational economic behaviors cause market inefficiencies and reduced economic growth due to improper allocation of capital (Daniel & Titman, 1999).

Previously there have been researches made to predict company value whether for startups (Dhochak & Doliya, 2020), publicly traded companies (Zhou et al., 2022), or even via machine learning (Stoyanov, 2024). However, there has not been any research for Indonesian Stock Exchange (IDX) wide machine-learning-based valuation for Indonesian companies. This is important as there are considerable differences in company valuation that vary based on company nationalities (Kuzey et al., 2014) such that a model trained on a general dataset or a dataset that consists of companies from another nation might not appropriately reflect the conditions of Indonesian companies. This can be seen through the significant difference in the average Price to Earnings (P/E) ratio between the companies listed in the Indonesian and US stock markets where US companies are on average valued more than double that of Indonesian companies for the same earnings (10.74 vs 25.55) (World PE Ratio, 2024). This research gap is partly attributed to the lack of a publicly accessible dataset that contains fundamental data for all Indonesian companies listed on the Indonesian Stock Exchange (IDX) that can be easily used to do this type of research.

This study addresses that gap by compiling a new dataset of all IDX-listed firms from 1999 Q1 to 2024 Q4, including quarterly financial fundamentals and key annual Indonesian macroeconomic indicators. The goal is to apply and compare machine learning models to forecast company valuation, identify the most influential features, and evaluate whether combining firm-level and macroeconomic data improves accuracy. Three machine learning models will be evaluated, with a state-of-the-art model in Temporal Fusion Transformers (TFT), and Extreme Gradient Boost (XGBoost) and Long Short-Term Memory (LSTM) acting as comparison to more established models. The importance of features will be derived from the TFT model's internal mechanisms.

Fundamental data are data that represents the financial state of the company (Baker et al., 2020) such as sales, assets, dividends, etc. Meanwhile macroeconomic data are data that represents the economic conditions of a state (Verma & Bansal, 2021) such as GDP growth, FDI, interest rates, etc. Both of this information can affect the valuations of a company. In doing so, the research aims to help investors decide the appropriate valuation of a company and make appropriate data-driven decisions based on that. In addition, more accurate valuations can contribute not only to improved portfolio outcomes but also to more efficient capital allocation and broader economic development.

1.2. Problem Formulation

Despite the availability of various economic and financial data in the market. The valuation of individual companies in the stock market often does not correlate with their financial and business conditions due to various market manipulations, emotional-driven decisions, or limitations in traditional market analytics. Individual investors may be unable to fully extract insights from the wealth of data available leading to suboptimal investment decisions and market inefficiencies. The key problems to be addressed in this research are:

1. How can a dataset combining historical Indonesian company fundamentals and relevant macroeconomic conditions be effectively formed to train a machine-learning model?
2. What machine learning model would be best suited to forecast a company's valuation based on the dataset?
3. What are the key features that are responsible for a company's market valuation?

1.3. Research Objectives

There are several objectives that this study intends to achieve:

1. The formation of a new dataset consisting of historical Indonesian company fundamentals and macroeconomic conditions.
2. The formation and evaluation of a new machine learning model to accurately forecast company valuation based on the dataset created.
3. Finding out key features that are responsible for a company valuation.

1.4. Research Benefits

This research hopes to produce several benefits, including:

1. Enhanced valuation accuracy: By employing a machine learning model, this research aims to aid investors in making a more objective and accurate company valuation.
2. Data-driven information available to all: This study would give retail and amateur investors access to an easy method of evaluating their prospective investments, allowing them to better allocate their capital and achieve higher returns.
3. Increasing economic development: Valuing a company according to its proper valuation lets capital be utilized more efficiently in an economy, thereby increasing economic development.

1.5. Research Scope

This study is limited in scope to the dataset used, specifically:

1. The compilation of a dataset consisting of the historical quarterly financial fundamentals of publicly traded Indonesian companies derived from income statements, balance sheets, and cash flow statements. These include 23 different variables such as total revenue, free cash flow, net debt, etc.
2. The inclusion of 9 relevant annual Indonesian macroeconomic indicators such as GDP growth, inflation, interest rates, etc. to complement the company-level data.
3. The exclusion of companies that are non-publicly traded or based outside of Indonesia and the use of macroeconomic figures with larger granularity compared to the company-level data (annual vs quarterly) may impact model accuracy and generalization.

CHAPTER II
LITERATURE REVIEW

2.1. Traditional Valuation Methods

According to (Damodaran, 2024), there are generally two approaches to any asset valuation, intrinsic and relative. Intrinsic valuation means valuing an asset based on the internal economic value of the asset. In the case of companies, these can be seen in their asset value, earnings, profit margin, etc. Relative valuation meanwhile, focuses on comparing an asset with the valuation of similar assets in the market. In the stock market, this can be achieved by comparing companies operating in the same industries.

Table 2.1. Main Traditional Valuation Methods

Balance Sheet	**Income Statement**	**Mixed**	**Discounted Cash Flow**
• Book value • Adjusted book value • Liquidation value • Substantial value	• Multiples • PER • Sales • P/EBITDA • Other multiples	• Classic • Abbreviated income • Others	• Equity cash flow • Dividends • Free cash flow • Capital cash flow • APV

As illustrated on Table 2.1, traditional valuation techniques such as balance sheets, income statements, mixed methods, and Discounted Cash Flow (DCF) have traditionally been the foundations of company valuation (Fernández, 2023). Each of these methods represents a different view on what constitutes business value and should be applied in different scenarios.

Balance sheet-based valuation methods are focused on the net worth of a company’s assets after subtracting its liabilities. The basic form of these methods would be to look at the book value, which is a measure that shows a company’s shareholders' equity on its balance sheet. A more advanced metric would be an adjusted book value that adjusts the company’s assets and liabilities in order to calculate a more accurate market value for them, especially in regards to tangible assets such as property, plant, and equipment. Liquidation value meanwhile calculates the value a company’s sale of its assets would be worth in case it closes its business operations and sells its assets at possibly reduced market value. This is usually used to evaluate companies that are experiencing difficulty in the market and risks facing bankruptcy. Substantial value takes into account the replacement value of the total value of a company's assets, especially asset heavy companies such as companies operating in the manufacturing and infrastructure sectors. It must be noted that the use of balance sheet analysis has its downsides in that it doesn't account for potential growth or intangible value.

Meanwhile, valuations based on income statements rely more on earnings, revenue, or the company's operating performance. Sometimes these valuation methods are also called relative valuation techniques. These valuation approaches are usually based on multiples/ratios such as Price to Earnings (PE) Ratio, Price to Sales, or Price to EBITDA, which compare a company's market value with key financial measures of its operations. These methods are widely used due to their simplicity and ease of comparability among different companies across different industries. But accounting practices, business cycles, or just temporary earnings fluctuations can significantly alter these multiple values and can cause misvaluation if not interpreted carefully.

Mixed valuation models combine aspects of both balance sheet and income statement valuations. Examples include classic and abbreviated income methods, which estimate firm value based on normalized earnings while adjusting for asset values. The main idea behind using mixed methods is to attempt to balance the stability of asset-based valuation with the forward-looking nature of earnings-based approaches. These methods are often applied in private company valuation or merger and acquisition contexts, where both asset backing and earning power are important considerations.

Discounted Cash Flow (DCF) models are the most theoretically sound among the traditional intrinsic valuation approach since the estimates for the value of the firm are based on the present value of expected future cash flows. There are several DCF variants including dividend discount models, free cash flow to equity models (FCFE), free cash flow to the firm models (FCFF), capital cash flow models, and Adjusted Present Value (APV) models. These methods incorporate growth expectations, risk, and the time value of money, making them highly flexible and conceptually robust. In general, DCF is the best method to evaluate a company as the value of a company hinges on its ability to generate future cash flows to its shareholders. However, DCF models are quite sensitive to growth rates, costs of capital, and terminal value estimates. This sensitivity can lead to a considerable degree of error in the results.

Research on Indonesian companies operating in the automotive sector (Husain et al., 2020) reveals that dividends do not play an important role in predicting company valuation and itself is not influenced by a company's profitability. However, company profitability calculated based on Return on Assets (RoA) does heavily influence Indonesian company valuation. This underscores the importance of internal financial performance over more superficial payout policies for Indonesian company valuation. These findings substantiate a view that basic financial variables are core to value estimation, as well as inspire researching data-driven techniques and machine learning models for depicting more complex, non-linear relationships between determiners of valuation.

2.2. Company Fundamental Data

A company's fundamental data is largely derived from a company's financial statements, such as income statement, balance sheet, and cash flow statement. These financial documents form the basis of traditional fundamental analysis of a company's intrinsic value. According to (Baker et al., 2020), these information can be used by investors to calculate and assess a company's financial health, operations, long-term growth and thereby derive the intrinsic value of a company.

The availability of quality data is important for the development of a good machine-learning model. A study by (A. Jain et al., 2020), listed several factors that could determine the quality of structured data such as label noise, class imbalance, data valuation, data homogeneity, data transformation, and data cleaning.

Traditional valuation methods as mentioned in Sub-chapter 2.1 by and large use fundamental data to calculate a reasonable valuation for a company. These fundamental data encapsulate a wide range of financial indicators including but not limited to revenues, assets, dividends, etc. Fundamental analyst uses these information to determine whether a publicly traded company is overvalued or undervalued relative to its market price and competitors. These data can help investors look beyond short-term price fluctuations and focus on underlying factors that matter in the long-term.

Table 2.2. Examples of Company Fundamental Data

Income Statement	Balance Sheet	Cash Flow	Others
Total Revenue	Cash	Cash from Operations	Return on Equity
Operating Profit	Total Assets	Free Cash Flow	Payout Ratio
EBITDA	Total Liabilities	Cash from Financing	Net Margin %

As listed on Table 2.2, company fundamental data includes a wide array of fundamental financial indicators derived from income statement, balance sheet, cash flow statement, and other market-derived financial information.

2.3. Macroeconomic Data

Macroeconomic variables are essential in understanding to broader national economic situation in which companies operate in. According to (Verma & Bansal, 2021), macroeconomic conditions have a significant correlation with stock performance. Figures such as GDP growth, Foreign Direct Investment (FDI), interest rates, etc. have all been discovered to have either positive or negative impacts on the performance of the stock market as a whole. This is because these macroeconomic figures reflect the condition of the economy in general, which in turn affects investor confidence, risk premiums, consumer behaviors, etc.

For instance, high GDP growth and lower unemployment typically signal a thriving economy, which leads to higher growth and stock valuations for a company due to increased expectations from investors (K. Jain, 2024). Conversely, high inflation and debt to GDP percentage may signal a turbulent economy and decreased consumer confidence (Dhanamaru, 2024). Whilst higher interest rate, increased cost of borrowing and reduces investment (X. Wang & Sun, 2023).

Assagaf et al. (2019) found that inflation, interest rate, broad money (money supply), and exchange rate all significantly affect stock returns of companies listed on the Indonesia Stock Exchange. Research in the Pakistani stock market concludes that foreign exchange rate, foreign exchange reserve, M2 (broad money), and industrial production index are the macroeconomic variables that most affect stock prices (Muhammad et al., 2009).

Table 2.3. Examples of Macroeconomic Variables

GDP Growth %	Unemployment %	Inflation %
Interest Rate %	Broad Money Growth %	Debt to GDP %

As listed on Table 2.3, a country’s macroeconomic data includes a wide array of variables which can give more context and influence the valuation derived from fundamental data. These data can give us a glimpse into the health of a national economy.

The inclusion of macroeconomic variables in this study is justified by their demonstrable influence in company valuation as influence is past research (Verma & Bansal, 2021). Incorporating them alongside company fundamental data allows a more holistic approach in evaluating a company valuation at any given time.

2.4. Data Science

Data Science is an interdisciplinary field that uses scientific methods, specifically mathematics, statistics, and computer algorithms to extract insights from available data (Dhar, 2013). These insights then can be used to improve decision making processes in organization, companies, or just in personal matters. It usually takes in large amounts of both qualitative and quantitative from various sources such as images, texts, sensors, transactions, tabular data, etc. to generate actionable insights. These large generally unstructured are popularly known as big data.

In finance, data science has increasingly been used to improve decision making processes as the industry has accumulated many structured data where insights can be easily extracted from. (J. Wang et al., 2024) Some applications where data science has been used to improve decision making include but not limited to credit risk evaluation, fraud detection, algorithmic trading, etc.

To systematize the data science pipeline, research has created a standard model called CRoss-Industry Standard Process for Data Mining (CRISP-DM).

CRISP-DM provides a framework for planning and executing data science related projects (Chapman et al., 2000). It consists of six phases:

i. Business Understanding:

This phase focuses on understanding the objectives and requirements of the project from a business perspective. For this study, the primary goal is to improve the accuracy of company valuation using machine learning models based on fundamental and macroeconomic data.

ii. Data Understanding

This phase includes data collection and exploration in order to understand its characteristics. For example, it is important to identify what variables are available, whether each dataset follows the same template, and to examine the presence of missing values.

iii. Data Preparation

Here relevant data are selected and undergo pre-processing in order to remove problematic data and transformed into a format suitable to be fed into the machine learning model. This phase may involve actions such as removal or handling of missing value, selecting relevant features, aggregating macroeconomic data to match firm-level data granularity, etc.

iv. Modeling

During modeling, selected machine learning models will be trained on the dataset. In this case models such as XGBoost, LSTM, or TFT can be used to forecast company value.

v. Evaluation

The resulting model's performance will need to be assessed using several metrics such as MAPE. This phase is important to ensure that the models produced can perform adequately in action.

vi. Deployment

Deployment involves implementing the model in a real-world situation in which it may be used to solve actual tasks. However, in academic settings this may involve documenting findings and comparing their performance to other models in different studies instead.

2.5. Machine Learning Models

There are multiple machine learning models that could be used in a time-series forecasting task. The model selection has been narrowed down to three models based on various considerations.

2.5.1. Extreme Gradient Boost (XGBoost)

Extreme Gradient Boost is the most popular derivation of the family of gradient boosting algorithms introduced by T. Chen & Guestrin (2016). It is developed as a scalable and efficient end-to-end tree boosting system. It operates by combining many weak learners, usually decision trees, by sequentially minimizing errors from previous models through gradient descend optimization.

XGBoost introduces some novel technical innovations that differentiate it from the general gradient boosting approach. These innovations include sparsity-aware algorithm for handling missing values in the data, a weighted quantile sketch for approximate tree learning, and regularization methods in order to mitigate overfitting. Another advantage of the XGBoost algorithm is its efficiency in computation due to the use of parallel processing and hardware awareness, making it suitable for large-scale datasets.

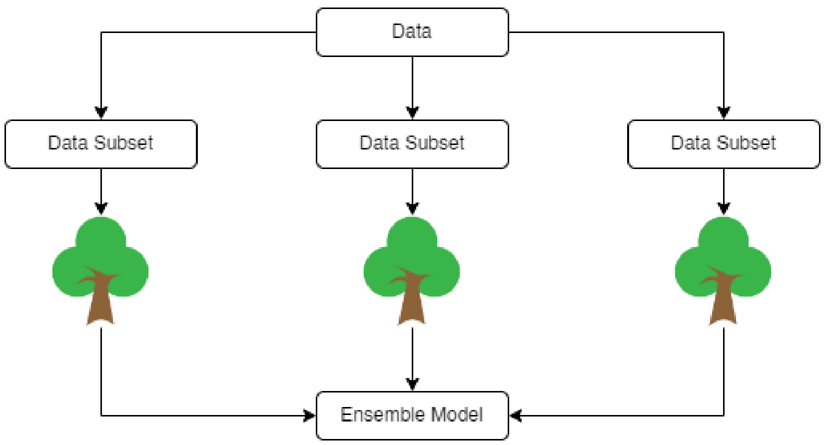

Figure 2.1 XGBoost Illustration

As illustrated on Figure 2.1, XGBoost much like other forest-based ensemble learners, operates by combining multiple weak learners into a stronger model (Breiman, 2001). It splits the dataset into various randomized data subsets in a technique called bootstrapping. And then those data subsets are used to train separate smaller tree models that will then be aggregated into one final ensemble model. Unlike neural network-based models, the temporal dependencies are not modeled inherently by XGBoost. Thus, in applications of XGBoost to time-series forecasting, the temporal information needs to be encoded manually using lagged features and engineered covariates. Despite this inherent shortcoming, XGBoost has showed good results in structured and tabular data prediction tasks, including economic and financial forecasting.

XGBoost is included in this study as a non-neural benchmark model due to its robustness, interpretability, and proven strong empirical performance in time series forecasting (Geertsema & Lu, 2023; Priel & Rokach, 2024). In view of this, the author would like to try applying XGBoost models to the forecasting task of an economic dataset to enable a comparison between deep learning–based models and traditional machine learning approaches, possibly informing us into the trade-offs that newer more complex models have.

2.5.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory is a type of Recurrent Neural Network (RNN) architecture introduced by Hochreiter & Schmidhuber (1997) to address the vanishing gradient problem present in standard RNNs. Essentially gradients tend to either diminish to explode during back-propagation of errors through time in traditional RNNs. As a result, RNNs often struggle to learn long-term dependencies present in the data. LSTM addresses this problem by introducing a memory cell structure and gating mechanisms that enable it to selectively remember and forget information so that the model will not be cluttered with noisy information minimally relevant to the training of the model.

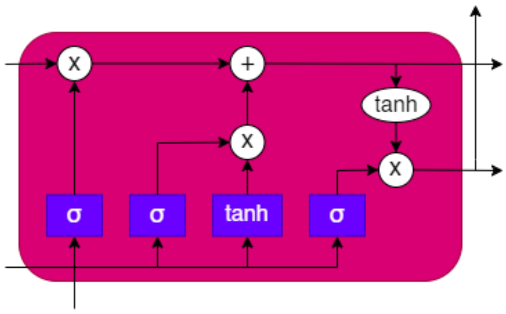

Figure 2.2 LSTM Illustration

Figure 2.2 portrays the inner working of an LSTM cell. Each LSTM unit consists of three main gates (σ): the input gate, forget gate, and output gate. All of these gates are controlled by a sigmoid activation function. These gates regulate the flow of information in and out of the cell state. The forget gate is responsible for determining which information from the previous cell state is discarded, while the input gate is responsible for controlling which new information is stored, with the output gate controlling which information is passed along to help form the next hidden state of a cell. This structure enables LSTM to selectively remember past information while forgetting noisy information minimally relevant model training.

By sequentially processing input time steps, the LSTM builds an internal representation of temporal dependencies that can be used for forecasting future values. LSTMs have been employed extensively in different contexts, including financial forecasting, macro-economic analysis, and demand analysis. Although LSTM models have demonstrated good performance in these aspects, they have been criticized for their lack of interpretability and difficulties in multi-horizon forecasting tasks without architectural extensions.

Despite that, LSTM can still be considered a strong benchmark model for time series prediction, and it has regularly been employed as a benchmark or even base model for more complex architectures, for example, attention-based and hybrid models. In this study, LSTM has been proposed as a benchmark model for comparison with more advanced models using deep learning, such as the Temporal Fusion Transformer model, to evaluate improvements that may be offered by such models.

2.5.3. Temporal Fusion Transformer (TFT)

Temporal Fusion Transformer is a newer state-of-the-art deep learning model made for time series forecasting. It was introduced by Lim et al. (2021) with a hybrid architecture combining Recurrent Neural Networks (RNNs) like LSTM with attention-based transformers. This was done in order to overcome some shortcomings of traditional RNN and transformer-based models, particularly relating to their lack of interpretability and tendency not to handle varying time series inputs effectively. The hybrid architecture can enable them to model both short-term dynamics and long-term temporal dependencies more effectively.

One of the main innovations in TFT is its ability to handle different varieties of input variables such as static covariates, time-varying known inputs, and time-varying observed inputs. These variables are dealt with by specific sub-networks. These sub-networks help the model learn the effect of different variables on the forecasting task. The variable selection network calculates weights for the different input variables at every time step. This helps the model to concentrate on important variables and minimize the contributions of unimportant ones. This is useful because aside from making the model perform tasks better, it can also be used to interpret the results because the learned weights can be analyzed to assess feature importance over time.

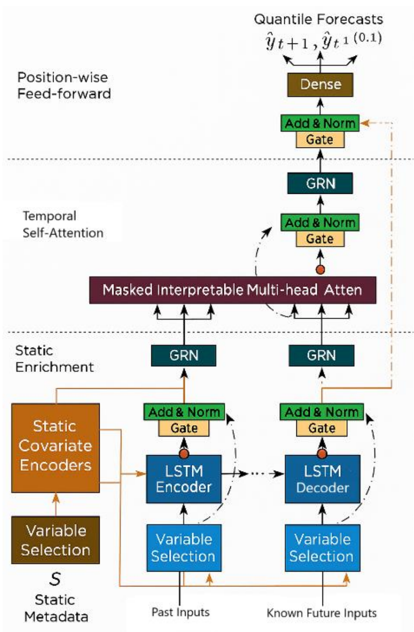

Figure 2.3 TFT Illustration

Figure 2.3 gives us a general overview of the structure of TFT. It uses an encoder–decoder architecture with LSTM layers to capture temporal dependencies in time series data. It processes time-series data through multiple components depending on the data type. The encoder layer takes as input historical observations, while the decoder generates predictions for future time steps. On top of this recurrent backbone, TFT incorporates multi-head attention layers that allow the model to selectively focus on the most relevant time steps against which to base predictions for the future. This attention mechanism enhances the model's ability to capture long-range temporal patterns that may not be efficiently learned by recurrent layers alone.

Another critical element of TFT is the Gated Residual Network (GRN), which regulates the flow of information across multiple components of the model. GRNs use both gating and residual components to prevent overfitting during the optimization process and ensure that the model can circumvent unnecessary transformations during the process. Finally, the dense layers will then output the forecasted value in the form of quantile distributions.

All of these components in TFT work together to ensure that the TFT model has adequate interpretability, which is usually difficult to find in neural network-based models for forecasting tasks. Because of these properties, TFT has become increasingly popular in recent forecasting literature, especially for multivariate and economic time series data.

2.6. Feature Importance

Feature importance refers to a field of machine learning that tries to determine the relevancy of each feature in a dataset to model prediction (Konig et al., 2021). Understanding the ranks of feature importance helps researchers identify which variables have the most impact on a model’s output. This can be used to improve model performance, better feature selection, and increased interpretability.

Theoretically this endeavor is grounded in the thinking that not all variables contribute equally to an output. For example, hours of sleep would contribute more to a person's level of tiredness compared to air quality. In high-dimensional datasets, identifying and selecting the most important variables could reduce noise and prevent overfitting.

There are many approaches to determine feature importance via machine learning models. Some methods are inextricably linked to certain machine learning models such as coefficient-based importance for linear models (Kwon, 2025) and impurity-based importance for tree-based models (Scornet, 2020). Meanwhile, there are several feature importance determining methods that can be universally used by practically every machine learning models such as the game theory-based SHapley Additive exPlanations (SHAP) values (Lundberg & Lee, 2017) and permutation importance (Breiman, 2001).

Some recent deep learning time-series forecasting models, such as the Temporal Fusion Transformer, embed feature importance mechanisms directly into their architecture. These models, through parts of their architecture like variable selection networks or attention mechanisms, learn to dynamically weight input variables and time steps with regard to their relevance for the prediction task at hand (Lim et al., 2021). This embedded interpretability means feature importance can be extracted as a part of training, providing an explanation of model behavior in a more integrated and theoretically consistent way than that provided by post training methods.

Overall, feature importance acts as an essential bridge between predictive performance and interpretability. Specifically, within financial valuation and forecasting, it allows researchers to verify if machine learning models depend on more on certain economically meaningful variables such as profitability, leverage, and macroeconomic indicators. This enhances confidence in both the model's predictions and its practical viability.

2.7. Previous Works

In recent years, machine learning has become popular as a means to more accurately evaluate companies. A research by (Bartram & Grinblatt, 2018) tries to predict company market capitalization using simple linear regression. They did this by predicting each company market capitalization per month based on fundamental data and sorting them from most undervalued to most overvalued based on the percentage difference between the predicted value and real market capitalization. They found that the 10% of stocks that are considered the most undervalued outperform the 10% that are considered most overvalued by up to 10% annually. They also found that the price gap decays to 0 after 34 months.

Another approach to stock price prediction is by combining technical data and social media sentiment analysis best exemplified by a research from (Xu & Keselj, 2019) that found using LSTM with attention mechanisms increases the prediction accuracy compared to normal LSTM. They also found that social media posts made in the intervening time where the market is closed have more impact on stock movement than posts made when the market is open.

A novel approach to stock price prediction proposed by (Patil, 2020) is using GNN as a predicting model. This research again used historical daily stock prices data as the input dataset to forecast future stock prices. They managed to achieve prediction errors of 5-10% MAPE depending on model specifications and range of time predicted.

Gu et al. (2020) proposed using various data to evaluate and predict future returns of various US stocks. They found that various fundamental data plays significant roles in determining stock prices. It then uses various machine learning algorithms to pick stock portfolios and found that in the long run machine learning, particularly neural network-based machine learning significantly outperforms the S&P 500 index returning up to 27.1% annually.

Treading the same approach as (Xu & Keselj, 2019), (Mokhtari et al., 2021) presumed that fundamental data can be inferred from sentiment analysis with data taken from social media as after a company financial statement posts would be made from it either expressing negativity or positivity. It examined AAPL stocks over the period of 2010 to 2021 accompanied with nearly 6000 tweets over that time period related to AAPL. Apparently simple linear regression performed the best on technical analysis with RMSE of 1.82, meanwhile SVM performed the best on sentiment analysis with accuracy of 0.755.

Some research also went into machine learning fundamental valuation on other asset classes such as what has been done by (Steurer et al., 2021) to the housing market. They attempted to predict property prices in the city of Graz, Austria using some fundamental data such as internal size, time of sale, available parking, balcony, age, etc. They examined several machine learning model and found that the best model is Random Forest with a MAPE of 0.195.

Two other group of researchers examined the Chinese stock market (Leippold et al., 2022) and European stock markets (Drobetz & Otto, 2021) using similar methodologies to (Gu et al., 2020). The Chinese stock market is found to rely more on liquidity-based trading signals rather than fundamental data. This is reasoned by the authors by the relative immatureness of the Chinese stock market owing to its recent establishment and lack of stock investing culture. Neural networks are still found to have the best performance among other models and stock indexes. In the European study however, they found that Support Vector Machines deliver the best performance in this scenario, delivering the highest return out of all the models tested.

Some studies of Indonesian companies have been done such as in (Kusuma & Budiartha, 2022), which explore the use of Long Short-Term Memory (LSTM) to predict future prices for 28 companies included in the LQ45 index. It concluded that LSTM managed to predict asset prices better than standard traditional methods with lower mean and standard deviation. However, both methods still produce mean values above 50% which is not accurate enough for any type of asset valuation.

Another study by (Hadrian & Kusuma, 2023), successfully modifies LSTM with an attention mechanism to increase its accuracy in predicting time series AALI stock price. It also found that adding additional input from 10 commodity prices and filtering it with feature selection manages to further improve the prediction accuracy compared to just using daily stock trading data.

Hartanto & Gunawan (2024) tries to forecast individual Indonesian companies daily stock prices using purely technical factors with TFT. They managed to forecast with 0.22% – 0.78% MAPE for those individual stock prices within a time horizon of 3 days.

Geertsema & Lu (2023) looked at Gradient Boosting Machines (GBM) as a way to better predict company valuations using fundamental data. They approach this problem in a similar way to (Bartram & Grinblatt, 2018), where they also perform portfolio-based evaluation in which they managed to achieve a 0.51%

monthly return. They do differ in that this research outputs ratios as target variables instead of directly predicting market capitalization to reduce overfitting. They also found that GBMs reduced median absolute valuation errors by around 20–30% compared to the best traditional models.

A recent study conducted by (Phuoc et al., 2024) examined Vietnamese company listed in the VN30 index to forecast their day-by day stock price using technical data (price history) fed into an LSTM model. This study yielded favorable results with most of the company being able to be forecasted with 93% or higher accuracy.

Research by (Stoyanov, 2024), shows that Large Language Model (LLM) chatbots are a viable way to get a quick company valuation assessment. However, specialized machine learning models still produce results more similar with traditional valuation methods compared to traditional valuation methods.

More often than not machine learning research in this sector focuses on time series stock projection rather than pure company valuation. A study by (L. Chen et al., 2024) tried to make a model to estimate stock prices for US companies. This research compiles and uses a dataset with 50 years of monthly US stock return combined with 178 macroeconomic variables and 46 firm-specific characteristics. They propose the use of Generative Adversarial Network (GAN) to find the conditional Stochastic Discount Factor (SDF) of assets which is the discount rate of an asset in a particular period based on macroeconomic data and the use of LSTM to predict future asset prices. They found that SDF remains stable over time and that their model successfully outperforms other benchmark models in Sharpe Ratio, explained variation, and pricing errors.

Priel & Rokach (2024) came up with a unique approach to the problems of stock pricing using fundamental data. Instead of regressing market cap or returns directly, they instead classify whether a stock is a good purchase or not and if the model outputs in the positive it will automatically generate a buy signal. This approach manages to create a 15% annual return for their portfolio and in fact over 80% of the stock chosen to be bought returns more than 15% annually. Besides that, using SHAP they found that margin of safety is the most influential features in their dataset for their model output.

Table 2.4. Previous Works Summary

No	Publication	Method / Model	Dataset	Result
1	Agnostic Fundamental Analysis (Bartram & Grinblatt, 2018)	Linear Regression	Monthly US stocks price, returns, and fundamental data March 1987 - December 2012	10% more return between Q1 (most undervalued) vs Q10 (most overvalued) stocks, 91.7% R^2
2	Stock Prediction using Deep Learning and Sentiment Analysis (Xu & Keselj, 2019)	LSTM + Attention Mechanism (technical and sentiment)	US stock data + financial tweets	Accuracy: 54.58%, MCC: 0.04780
3	Stock Market Prediction using Ensemble of Graph Theory, Machine Learning and Deep Learning Models (Patil, 2020)	Graph Neural Networks (GNN)	Top 30 US stocks data	MAPE: ~5-10% (9 days horizon)
4	Empirical Asset Pricing via Machine Learning (Gu et al., 2020)	Linear Regression, RF, Neural Networks, etc.	30,000 US stocks (1957-2016)	Neural networks: 27.1% annual return, Sharpe 2.45
5	Effectiveness of Artificial Intelligence in Stock Market Prediction Based on Machine Learning (Mokhtari et al., 2021)	SVM, Linear Regression, RF, etc. (technical and sentiment)	AAPL daily stock price data (technical) + nearly 6000 tweets (sentiment)	Technical: Linear Regression (1.56% MAPE) Sentiment: SVM (76% sentiment accuracy)
6	Metrics for Evaluating the Performance of Machine Learning Based Automated Valuation Models (Steurer et al., 2021)	Linear Regression, RF, MARS, NN, Quantile Regression	House transaction data in Graz, Austria 2015-2020 (fundamental data)	MAPE: 0.195 (RF)
7	Empirical Asset Pricing via Machine Learning: Evidence from the European Stock Market (Drobetz & Otto, 2021)	Linear Regression, RF, Neural Networks, etc.	European stocks (1990-2020)	SVM: 17.1% annual return, Sharpe 0.94
8	Machine Learning in the Chinese Stock Market (Leippold et al., 2022)	Linear Regression, RF, Neural Networks, etc.	3,900 Chinese stocks (2000-2020) + macroeconomic data	Neural networks: 27.1% annual return, Sharpe 3.45
9	The Capital Asset Pricing Model Forecast Using Artificial Intelligence (Kusuma & Budiartha, 2022)	LSTM	LQ45 daily stock prices	AI CAPM: MAPE 61.3% with SD 52.1
10	Stock Price Prediction on Indonesia Stock Market with the	LSTM + Attention Mechanism	AALI stock + commodity prices	MAPE: 17.29% (1 day horizon)

	Influence of External Factors Using Recurrent Neural Network with Attention Mechanism (Hadrian & Kusuma, 2023)			
11	Temporal Fusion Transformers for Enhanced Multivariate Time Series Forecasting of Indonesian Stock Prices (Hartanto & Gunawan, 2024)	TFT	Daily stock price for ANTM, EXCL, and ASII	MAPE: 0.22% - 0.78% (3 days horizon)
11	Relative Valuation with Machine Learning (Geertsema & Lu, 2023)	Gradient Boosting Machine (GBM)	Monthly 10% largest US stocks 1980-2019 price and quarterly fundamentals	MedianAE: 0.29–0.32; Monthly Return 0.51%
12	Applying Machine Learning Algorithms to Predict the Stock Price Trend in the Stock Market – The Case of Vietnam (Phuoc et al., 2024)	LSTM	VN30 company daily stock prices	>93% accuracy for most stocks
13	Using Artificial Intelligence to Improve the Efficiency of the Market Valuation Method (Stoyanov, 2024)	Large Language Model (LLM)	Company financial data	LLM: 13m 49s, AI: 25m 48s, Traditional: 70m 56s (LLM delivers unspecified worse results compared to AI and Traditional)
14	Deep Learning in Asset Pricing (L. Chen et al., 2024)	GAN (SDF), LSTM (price prediction)	50 years US stock + macro & firm data	Sharpe: 2.6, R^2: >90%
15	Machine Learning-based Stock Pricing using Value Investing and Quality Features (Priel & Rokach, 2024)	RF and GBM, SHAP (feature importance)	Weekly US stocks market cap, quarterly fundamental, and unspecified practitioner insights of 2161 companies, 2000-2019	Annual Return 15%; Over 80% of stocks chosen returns >80% annually

In Table 2.4 the application of various machine learning models has shown promise in predicting company valuation due to their ability to model complex relationships in data. In view of the literatures that has been done, the proposed research is justified by the lack of existing studies on machine learning-based fundamental valuation in the Indonesian market. While significant research exists for major markets, such as the U.S., Europe, and China, Indonesian studies remain limited to technical stock price prediction and traditional methods. This is crucial as the importance of integrating company fundamentals and macroeconomic data has been shown by various researches and this approach is more robust in predicting long-term returns. Furthermore, the unique characteristics of the Indonesian market, such as its retail investor dominance and differing profitability dynamics, underscore the need for tailored approaches. and the demonstrated effectiveness of the selected machine learning models.

The study justifies the use of three machine learning models, specifically Temporal Fusion Transformers (TFT) with additional comparisons to more tested methods in Extreme Gradient Boost (XGBoost) and Long Short-Term Memory (LSTM). With TFT we can evaluate state-of-the-art performance, whilst XGBoost and LSTM provides useful comparisons to well-performing mainstay machine learning and deep learning algorithms respectively. Aside from that, feature importance will be derived from the TFT model's internal mechanisms. All of these methods have also proven some merit in previous research in this area. Examining these models will hopefully address the inherent complexities of company valuation, aligning with global advancements in financial prediction while filling a critical gap in the Indonesian market.

CHAPTER III
RESEARCH METHODOLOGY

3.1. Research Framework

This research focuses on improving the valuation methods of publicly traded Indonesian companies using machine learning. Traditional valuation methods are often biased due to market manipulation, emotional decision-making, and the lack of investor knowledge (Damodaran, 2024), leading to inefficiencies in capital allocation (Daniel & Titman, 1999). While previous studies have explored machine-learning-based company valuation on other markets, there has been little research focused on the Indonesian stock market, which has unique characteristics, such as lower average P/E ratios compared to the US market (World PE Ratio, 2024).

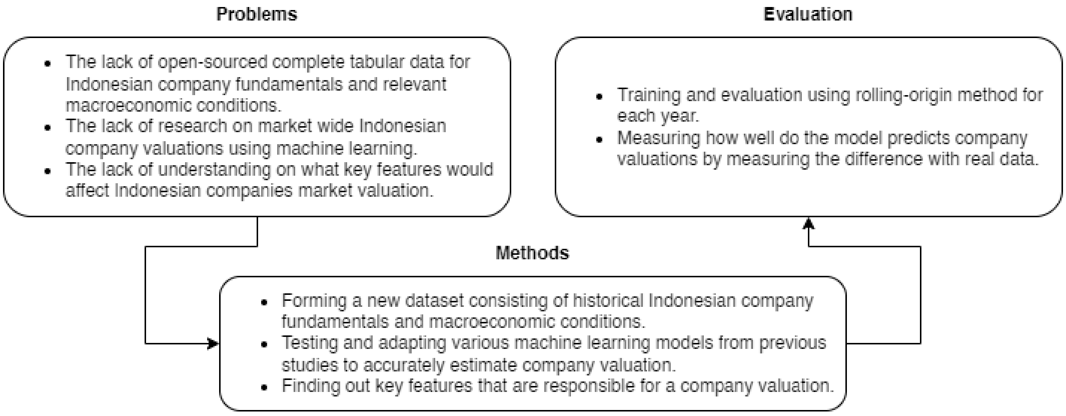

Figure 3.1. Research Framework Diagram

As mentioned in Figure 3.1, this study aims to address several key challenges in this endeavor, such as the lack of immediately usable Indonesian company fundamentals and macroeconomic data that could be immediately used in doing this research. This lack of ready-to-use data also contributes to the practically non-existent research on Indonesian company valuation based on fundamental data using machine learning. Additionally, while the key features influencing company valuation are relatively well-established in popular markets such as the US stock markets, it remains unclear what features would be important in the valuation of Indonesian companies nor its order of importance.

The study aims to address several key challenges on this endeavor, such as the formation of a comprehensive dataset combining historical financial data of Indonesian companies with macroeconomic indicators, identifying the best machine-learning model for forecasting company valuation, and determining key variables that influences companies market value.

This study seeks to address these issues by developing a machine learning model for estimating future company valuation to provide Indonesian stock market investors with objective and data-driven insights, helping them make better-informed decisions. Additionally, this research will produce a publicly available dataset company fundamentals + macroeconomic dataset that could be used for further research. A further objective is to determine which features from the dataset provides the most insight into company valuation in Indonesia. It is hoped that this research can enhance capital allocation and investment decisions for investors in the Indonesian stock market leading to better economic growth and investment returns.

The model will be developed by testing several machine learning architectures, including Extreme Gradient Boost (XGBoost), Long Short-Term Memory (LSTM), Temporal Fusion Transformer (TFT). Each model will be optimized by tuning key parameters to improve prediction accuracy. They will output a company's expected future valuation based on the data given. These results will then be evaluated and compared with each other to identify the best approach to estimating company valuations.

It is hypothesized that while the results may not be very accurate in the short term, over longer prediction horizons they are expected to perform better than models relying purely on technical time-series patterns. This is because company fundamentals change more gradually and reflect underlying economic value, whereas short-term market movements are often driven by irrationality, volatility, and sentiment-driven noise that make short-term forecasting inherently difficult. Therefore, the addition of fundamental and macroeconomic data should result in more informed and accurate prediction as they give explanatory power to the underlying asset and economic conditions.

3.2. Research Stages

Figure 3.2 represents the general workflow in completing this study. There are roughly eight stages that must be done in order to finish this research from topic selection to writing the research paper itself. All research starts with selecting a topic that interests you. In this case, after reviewing numerous academic papers, a gap was identified in the existing literature: the lack of research about market-wide machine learning-based company valuation models using fundamental data tailored to the Indonesian stock market. This combined with one of the author's interests and background in finance made us choose 'forecasting company valuation based on company fundamentals and macroeconomic data using machine learning' as the topic to address this research gap and provide a data-driven, localized solution.

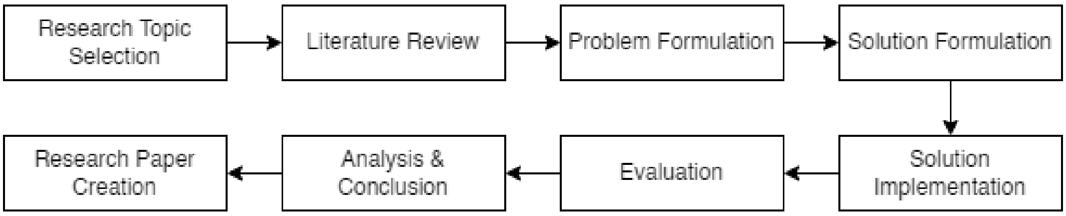

Figure 3.2. Research Procedure Flowchart

After selecting a topic, a comprehensive literature review was conducted on various topics and past research into company valuation. This gave the author a good foundational understanding of company valuation methods starting from traditional valuation methods, related works using machine learning methods in global markets, and other works in the field. In doing so, it highlighted several problems that needed to be addressed in order to realize the objectives of this project.

The problems found are threefold: the lack of a dataset combining firm-level financial data and macroeconomic indicators for Indonesian companies, the choice of machine learning models to be used, and the lack of understanding on what key features are would affect Indonesian companies market valuation. The solutions to these problems can be found in forming a new dataset, selecting several of the best performing machine learning models from the previous research, and doing feature importance analysis.

With the proposed solutions identified, the next step involves determining how to implement them effectively within the scope of the research. To build the dataset, access was gained to Refinitiv (LSEG, 2025), a service provided by the author's university that has historical financial data per company. These combined with macroeconomic data from various sources on the internet formed the foundation for the intended dataset. With the dataset established, the model selection process narrowed down three candidates: XGBoost, LSTM, and TFT based on performance in previous research and their diverse methodologies. These models are then trained and fine-tuned with the post-processed dataset to predict company valuation. Then, data features are preselected based on traditional valuation research and data availability, then their importance ranked using TFT's internal mechanism.

The resulting models will then be evaluated based on their difference to the real valuation using metrics such as MAPE. If needed, experimental parameters will be adjusted and re-tested to improve model performance. We will then end it with a conclusion to summarize the research results. All of this work will end up being written into a thesis that will be presented to the university and a research paper that will be published into a journal.

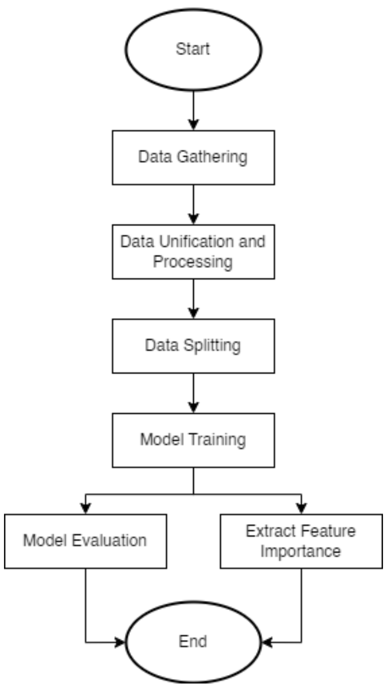

Figure 3.3. Research Methods Flowchart

Figure 3.3 represents the technical steps undertaken in completing the experimental phase of the study. In general, the process begins with gathering historical fundamental financial data from each publicly traded company in Indonesia and Indonesia's historical macroeconomic data. Then all the data will be processed and unified into a machine-learnable file. Following this, the selected models are trained and evaluated based on valuation error metrics.

3.3. Data Gathering & Dataset Creation

There has been no publicly available fundamental financial data for more than 3 years of publicly traded Indonesian company and certainly no data that laid them into one file. In view of this the author decided to gather these data from a financial database service called Refinitiv and manually combine them. Meanwhile the Indonesian historical macroeconomic dataset is publicly accessible from both the World Bank (World Bank Group, 2025) and International Monetary Fund (IMF) (IMF, 2025) website. The macroeconomic data used in this study will be primarily based on the World Bank dataset with supplementation on central government debt and inflation metrics from the IMF dataset. Finally, both the fundamental data and the macroeconomic data will be merged into one file.

Table 3.1 gives us a snippet of how the raw company financial data looked like. Each company data is downloaded one by one from the service, each one on separate files. Each of these files are subdivided into multiple worksheets, each containing critical parts of a company's financial statements such as financial summary, income statement, balance sheet, cash flow, and valuation. Each company data extracted originates from their quarterly financial statements originating as far as the data is available. The values in Table 3.1 are listed in million(s) rupiah. Each financial statement includes hundreds of variables, many of which are either empty or redundant, being captured more effectively by other, more comprehensive variables.

Table 3.1 Example of Raw Company (GOTO) Fundamental Data

Field Name	31-12-2024	30-09-2024	30-06-2024	31-03-2024	31-12-2023
Selected Income Statement Items					
Revenue from Business Activities - Total	4,231,884	3,925,166	3,658,186	4,079,226	4,274,773
Gross Profit - Industrials/Property - Total	2,194,092	1,998,359	1,667,476	1,876,968	2,323,908
Operating Profit before Non-Recurring Income/Expense	-189,568	-323,838	-785,499	-941,969	-1,479,240
Earnings before Interest, Taxes, Depreciation & Amortization (EBITDA)	-51,875	-204,002	-638,716	-602,106	-845,652
Income before Discontinued Operations & Extraordinary Items	-925,615	-1,693,249	-1,908,885	-937,106	-80,919,924

Table 3.2. Example of Raw Indonesian Macroeconomic Data

Indicator Name	2000	2001	2002	2003	2004
Poverty gap at $3.65 a day (2017 PPP) (%)	33.6	31.5	22.4	21.9	22.3
Poverty headcount ratio at $2.15 a day (2017 PPP) (% of population)	43.8	40.3	23.5	22.8	24.2
Income share held by highest 10%	25.6	26.3	25.9	24.6	25.3
Income share held by second 20%	13	12.8	13	13.2	12.9
Out-of-pocket expenditure per capita, PPP	42.24901474	45.23192314	50.71107314	60.22711442	60.33403495

(current international $)					

Table 3.2 gives us a snippet of how the raw Indonesian macroeconomic data looked like. Most of the macroeconomic data needed in this study comes from the World Bank data repository (World Bank Group, 2025) (IMF, 2025). However, both central government debt and inflation data are entirely derived and supplemented by the IMF data repository respectively as there are problems and incompleteness on the World Bank data. The World Bank database included macroeconomic data of Indonesia since 1960, although most of the cells are empty, especially concerning older data before the 1990s. The World Bank database contains over 1500 variables regarding the nation of Indonesia. However, only a few of them correlated to the macroeconomic conditions of the country, the rest such as diabetes prevalence or homicide rates are not considered in this study.

Both fundamental and macroeconomic data that have been gathered will then be combined into one file and processed. Each company will also be classified based on sectors based on data from IDX. Empty data where emptiness can be presumed as 0 like total debt or payout ratio, are filled with 0 value. Empty data that can be calculated via other data such as ratio and/or TTM-based metrics are calculated whenever possible. While the rest of the data that has empty values are dropped from the dataset. Furthermore, data from companies that are suspended in the stock market at the time of the data retrieval will be removed.

Since each company presents financial information with slight variations in reporting standards or formats, standardization will be carried out based on FCC codes on the dataset and the variable naming based on generally accepted financial accounting standards. This step ensures that variables remain comparable across different companies and sectors, and any discrepancies in units or naming conventions will be normalized or reclassified.

The dataset may also contain invalid data in the form of outliers, anomalies, or duplicate records. Outliers and anomalies are assumed to be just general market behavior as companies that report erroneous financial statements are not allowed to continue trading in the IDX to begin with. Duplicate records, if present, will be detected through cross-referencing and eliminated to avoid redundancy and potential data leakage.

Table 3.3. Fundamental Data Used in Dataset

Income Statement	Balance Sheet	Cash Flow	Others
Total Revenue	Cash	Cash from Operations	Industry Group
Operating Profit	Total Assets	Cash from Investing	Payout Ratio
EBIT	Total Liabilities	Cash from Financing	Net Margin %
EBITDA	Total Equity (BV)	Capital Expenditure	Debt/Equity %
Net Income	Total Capital	Free Cash Flow	Earnings Retention %
	Net Debt		Return on Equity
	Debt - Total		Return on Assets

Table 3.4. Macroeconomic Variables Used in Dataset

GDP Growth %	Unemployment %	Inflation %
Debt to GDP %	Interest Rate %	USD to Rupiah Exchange Rate
Industrial Growth %	Broad Money Growth %	Total Reserves in Months of Imports

The input variables in this dataset are fundamentally separated into two: fundamental data and macroeconomic variables. Table 3.3 lists 24 fundamental variables which are selected to be in the dataset. These variables are fundamental financial indicators derived from income statement, balance sheet, cash flow statement, and other market-derived financial information. Meanwhile Table 3.4 lists the 9 macroeconomic variables chosen for inclusion in this dataset. These variables are derived from the realities of the Indonesian economic conditions. Both of these sets of variables are chosen because they either represent or have demonstrated relevance in measuring the financial conditions of a company and/or the health of a national economy. Further information on the selection can be found in Sub-chapter 2.2 and 2.3.

3.4. Data Preparation

After the dataset is finalized, additional data preparation steps are going to be done to ensure compatibility and optimal training process for the machine learning models. These steps are done roughly the same way across models, with slight differences on XGBoost due to the different model architecture making it unsuitable for the same data inputs/outputs as the neural network models.

Because the dataset is of a multivariate time series nature, each company's data points need to be ordered chronologically so that the model can know the temporal order of the data points. In order to do this, a continuous time index is created each training cycle as representation of the temporal standing of each company data point. This time index is particularly needed for neural network-based models that can learn temporal dependencies.

All continuous variables are normalized using standard normalization (Z-score normalization). This means that all variables are centered at 0 with a standard deviation (σ) of 1. This particular normalization technique makes gradient descent stabler in neural networks (Lim et al., 2021). Normalization has been shown to immensely benefit machine learning models, in particular neural networks as it has been shown to improve numerical stability and accelerate convergence during neural network training (Ioffe & Szegedy, 2015). The feature normalization process will be done per company ticker on the neural network models. This is because each ticker/company will become their own time series in a neural network model. This is because normalization prevents variables with larger numerical ranges, such as total assets or revenue, from disproportionately influencing the learning process.

Categorical variables, such as economic sector and company tickers, are encoded as integer identifiers using a label encoder prior to model training for the deep learning models. Meanwhile in XGBoost, categorical variables are just ordinally encoded for convenience as tree-based models do not treat them ordinally.

On the other hand, target variables are normalized using a base 10 log function for all the models. Meanwhile, although the XGBoost model does not need any normalization on the target variable because tree-based models are invariant to variable transformations, log transformation is still done to ensure consistency wherever possible.

In order to evaluate model performance properly, there will be a rolling-window validation in which each cycle will contain 8 years of data starting from 2000 Q1 – 2007 Q4 and then moving the time window 1 year forward every cycle. In each cycle, the 8 years will be divided sequentially into 6 years for training, 1 year for validation, and 1 year for testing. This time-based splitting strategy reflects real-world forecasting conditions and prevents the use of future information during model training.

As these are forecasting tasks, there needed to be lagged input variables coming from past observations to feed the model in order for it to forecast future target variables. For the neural network model, in each company, a fixed number of past observations spanning 4 quarters (encoder length) structured as a sequence is used as model input to predict company valuation over the subsequent 4 quarters (prediction length). This sequence-based representation allows RNN-based models to explicitly learn temporal dependencies across multiple time steps.

Meanwhile XGBoost, while not inherently sequence-based, is also configured to perform multi-output forecasting in this study. Multiple lagged features are manually created by adding several historical observations from the past quarters as separate input variables, which will allow the model to grasp the temporal relationship in a tabular format. This is done for all of the continuous variable in the input. As for the output, XGBoost will be trained under an experimental multi-output regression setting, where each target will represent a particular forecast horizon, in this case one to four quarters ahead. This enables XGBoost to generate multiple future predictions simultaneously while still maintaining a tree-based learning structure.

Most of the preprocessing for the neural networks, including sequence construction, lag handling, and alignment between encoder and prediction windows, are done automatically within a data preparatory function in pytorch forecasting called TimeSeriesDataset. However, for XGBoost the author needed to manually pre-process the dataset using various different functions, mostly from the sklearn library. These preprocessing includes, but not limited to, the creation of lagged features, target vectors for each forecast horizon, and feature scaling.

As a result of these preparation steps, the dataset was transformed from a raw published dataset into a structured, normalized, and temporally consistent input suitable for both the tree-based and neural network-based models. Table 3.5 summarizes the data preparation steps explained in this Sub-chapter.

Table 3.5. Data Preparation Configuration

Component	Configuration	Applied To
Data Structure	Multivariate time series, grouped by company ticker	All models
Time Index	Continuous integer index per company, ordered chronologically	All models
Train–Validation–Test Split	Rolling-window validation (8 years total per cycle)	All models
Training Window	6 years (24 quarters)	All models
Validation Window	1 year (4 quarters)	All models
Testing Window	1 year (4 quarters)	All models
Encoder Length	4 quarters (historical input window)	Neural networks
Prediction Length	4 quarters (forecast horizon)	Neural networks
Continuous Feature Scaling	Standard normalization (Z-score normalization)	All models
Categorical Encoding	Label encoding	Neural networks
Categorical Encoding (XGBoost)	Ordinal encoding	XGBoost
Target Transformation	Log transformation per company	All models
Lagged Features	1-year lag (4 quarters)	XGBoost

3.5. Modelling

Once the final data has been created and pre-processed, then multiple machine learning models will be trained based on the fundamental and macroeconomic data collected to output the predicted valuation. In the training process hyper parameter tuning is performed for each model to examine the best parameters to use in order to achieve maximum performance.

This study aims to look into several models and examine the best parameters for each one. This study intends to use a new state-of-the art forecasting model called Temporal Fusion Transformer (TFT) to maximize forecasting accuracy. Aside from TFT, this study will also look at a more traditional machine learning and deep learning method in XGBoost and LSTM respectively.

XGBoost is an ensemble machine learning model based on gradient-boosted decision trees (T. Chen & Guestrin, 2016). It improves prediction performance by sequentially building trees that correct the errors of previous trees using gradient descent optimization. Due to its strong performance in previous researches on structured tabular data and its robustness to feature scaling, XGBoost will serve as a competitive non-neural network comparison in this study.

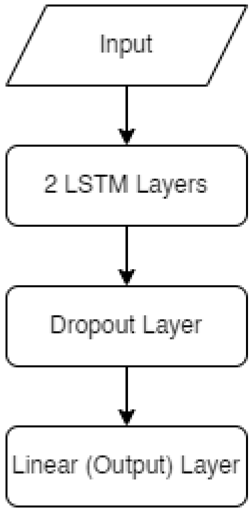

Figure 3.4. LSTM Architecture

LSTM is a type of recurrent neural network (RNN) designed to model sequential data by capturing both short-term and long-term temporal dependencies (Hochreiter & Schmidhuber, 1997). It addresses the vanishing gradient problem commonly found in traditional RNNs through the use of gating mechanisms that regulate the flow of information within the network. As can be seen on Figure 3.4, we will be using an LSTM model with a core of 2 layers of LSTM. Then the data will pass through a dropout layer and get outputted via a linear layer with an output size according to the time step needed, which in this case is 4 quarters. This structure should enable the model to learn temporal patterns while mitigating overfitting.

TFT is a deep learning model specifically developed for interpretable multi-horizon time-series forecasting (Lim et al., 2021). It combines RNN-LSTM components for local temporal feature extraction with attention-based Transformer layers for long-term dependency modeling. TFT also incorporates various mechanisms such as variable selection networks that dynamically weight input features at each time step. This allows the model to focus on the most relevant variables for forecasting. Aside from that, gated residual networks (GRN) regulate the flow of information between model components in order to improve training stability. Meanwhile the multi-head attention mechanism further enables TFT to identify which historical time steps contribute the most to future valuation predictions.

Table 3.6. Summary of Models Configuration

Component	Configuration
Encoder Length	4 (quarters)
Prediction/Decoder Length	4 (quarters)
Loss Function (LSTM and TFT)	SMAPE
Loss Function (XGBoost)	MSE
Optimizer (LSTM and TFT)	Adam
RNN Layers (LSTM and TFT)	2 LSTM
Early Stopping (LSTM and TFT)	40 Epochs
Learning Rate Scheduler	15 Epochs
Other Hyperparameters	Hyperparameter Tuned

All of these models will take in the full processed data, including lagged features and output the forecasted market capitalization for each of the time steps. TFT and LSTM as deep learning models are going to be using an early stopping function which stops after 40 epoch of the same or higher validation loss to prevent overfitting and not waste compute. Table 3.6 summarizes the general models configuration for this study.

As mentioned in the previous Sub-chapter, although all models in this study are trained on the same underlying fundamental and macroeconomic variables, the structure and dimensionality of model inputs differ substantially due to architectural differences between sequence-based neural networks and tree-based models.

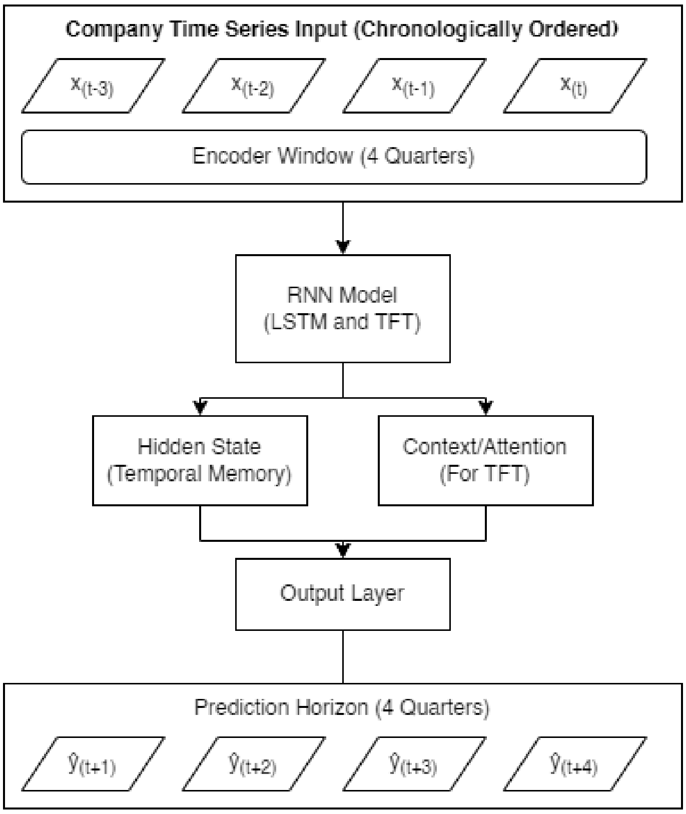

Figure 3.5. RNN-based Multi Horizon Forecasting Flowchart

For the RNN-based models (LSTM and TFT) inputs are presented as sequences. This can be seen in Figure 3.5 that describes the flow of data in the RNN-based models. The input data are modeled from each company separately, forming an independent multivariate time series, ordered chronologically. Each training instance feeds into the model with a fixed-length historical window of size four quarters on the encoder length. In every step in time, the model accepts several continuous and categorical variables, which form a three-dimensional input tensor with shape (batch size, time steps, number of features). Thus, this type of representation will allow RNN-based models to capture temporal dependencies explicitly since the model analyzes information step by step through time.

On the other hand, the XGBoost algorithm does not handle data in the form of sequences and hence is unable to capture relationships over time for various time steps. In order to bypass this limitation, the temporal aspects are addressed by changing the input data format so that it has explicitly engineered lagged features corresponding to the 4 previous quarters. We also made extra columns corresponding to the forecasted features in the future 4-timesteps forward. Consequently, the number of input variables for XGBoost increases linearly with the number of lags and features, whereas RNN-based models maintain a constant feature dimension per time step. This means that the XGBoost input variable per datapoint is 4x more numerous for the continuous variables corresponding to the 4 lagged timesteps.

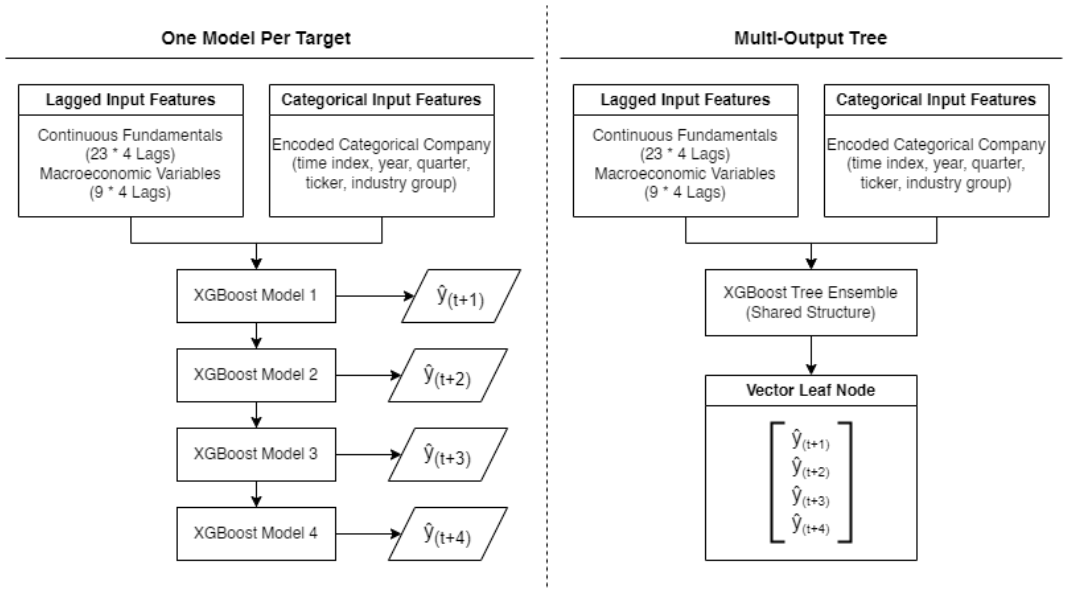

Figure 3.6. Multi-Output XGBoost Model Approach Comparison

In order to forecast multiple timesteps the XGBoost in our model will utilize an experimental multi output XGBoost model. There are really two ways to obtain a multi output XGBoost model. One is by training a different model for each output and the other by making a single tree with vector-valued leaves. This allows the tree structure to be shared across all outputs. These approaches are illustrated in detail on Figure 3.6 The first approach has been more proven but lacks any temporal dependency. Meanwhile the second approach can capture more temporal associations between timesteps but is still very experimental and less stable. We will examine both approaches to see which is better.

3.6. Model Interpretation

Temporal Fusion Transformer (TFT) incorporates interpretability directly in its architecture through variable selection networks and temporal attention mechanisms (Lim et al., 2021). Unlike traditional feature importance method which usually are obtained post-training, these components are learned jointly with the forecasting task during training. Because the weights are learned during training, they reflect the true predictive contribution of each variable rather than relying on post-training analysis.

3.6.1. Variable Selection Networks (Feature Importance)

TFT uses variable selection networks at both the encoding and decoding stages to assign dynamic weights to each input variable at every timestep. Each input feature is passed through a Gated Residual Network (GRN) and then assigned a weight using a softmax operation. These weights represent the relative contribution (percentage) that each feature influences the model's predictions at a given timestep (Medina Hernández et al., 2025).

This dynamic weight allows us to know not only which variables are generally important, but also how their importance shifts over time depending on the underlying patterns in the company fundamentals or macroeconomic conditions. For example, macroeconomic indicators may receive higher weights during periods of market instability, while firm-level financial metrics may be more important during more stable periods. This mechanism enables the model to adaptively prioritize different sources of information rather than relying on a fixed feature ranking.

In order to gain a global understanding of the feature importance, the weights will be aggregated across all timesteps in a given validation year and across the entire rolling validation process. The resulting averaged weights represent the overall contribution of each input variable to the forecasting task and are used as the primary measure of feature importance in this study.

3.6.2. Temporal Attention Mechanism (Timestep/Temporal Importance)

Aside from identifying important variables, TFT also provides a temporal understanding of which past time steps/lagged features the model focuses on when forecasting future values through its attention mechanisms. Temporal attention operates by assigning higher weights to timesteps that the model deems to contain the most relevant information for predicting company valuation.

This mechanism allows the model to distinguish between short-term and long-term dependencies. By examining these temporal weights, it is possible to determine whether the model is relying more on certain data like more recent timesteps or take in the general historical trends. This is particularly relevant in financial time series tasks, where valuation may be influenced both by recent market movements and long-term company performance.

The attention weights will be extracted from the trained model in each set in the rolling validation (every year) and be averaged out to produce a representative temporal importance data. This global data will provide insights as to how in general does the model think is the most influential timestep in determining company valuation.

3.6.3. Visualization

Together, these will give us a clear explanation of which factors most influence a company's valuation and how the model distributes its focus across both variables and time. The feature importance will be represented in a sideways bar chart that shows the aggregated contribution of each input variable across the validation set. Meanwhile, the temporal attention will be represented via a line chart showing the weighted attention across timesteps based on the validation set. These visualizations help contextualize the model's decision-making process.

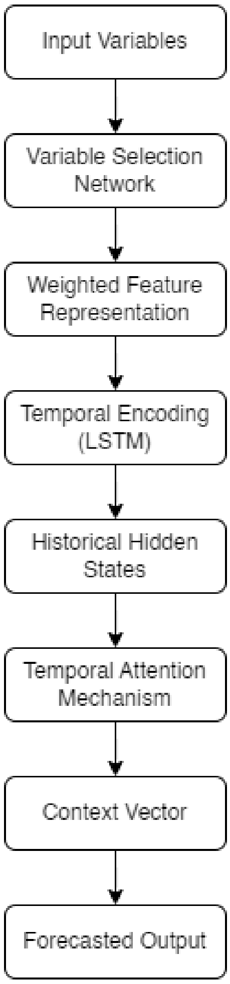

Figure 3.7. Conceptual Representation of Feature Importance in the TFT

Figure 3.7 represents the conceptual flow of feature importance extraction in TFT. Feature relevance is first determined through the variable selection network, which assigns dynamic importance weights to each input variable. Subsequently, a temporal attention mechanism identifies the most relevant historical timesteps for forecasting future company valuation.

3.7. Model Evaluation

The model error evaluation is done in a rolling-origin fashion. Each model will be trained on data based on six years of data, validated on data within the immediate year after the training, then tested on data within one year after the validation. The data window will then be shifted one year forward after each attempt until there is no data left to be tested. This process will be repeated with data from 2000 Q1 until 2024 Q4. This is illustrated in detail in Figure 3.8.

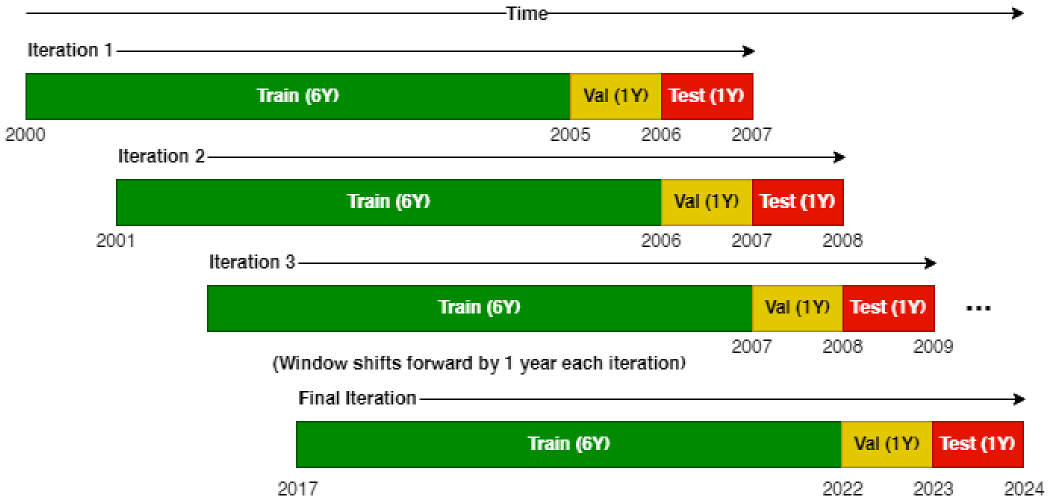

Figure 3.8. Rolling Origin Evaluation Representation

Hyperparameter tuning will be performed on all the models based on their performance when tested on the testing set. Due to computational limitations, the hyperparameter tuning will only be done on a single testing window instead of a full rolling-origin evaluation. Specifically, the testing window that will be used is the one whose validation data is in 2016, with a training window from 2010-2015. This is because in this period the stock market movement is relatively stable without any major market downturns, while still being relatively modern subset of the data. Each model will have separate hyperparameters that will need to be tested and adjusted differently to reach the optimal configuration.

In the case of XGBoost, hyperparameter tuning is focused on improving accuracy and preventing overfitting (Mehdary et al., 2024; Optuna, 2025). The learning rate controls how much each additional tree contributes to the final prediction. The maximum tree depth defines how complex each tree can be, which affects the model's capacity to learn intricate patterns. Additionally, the number of boosting rounds determines how many trees are built, with more trees potentially increasing accuracy at the risk of inducing overfitting problems.

For LSTM, hyperparameter tuning focuses on controlling the network's capacity to learn temporal dependencies and avoid overfitting (Saeed et al., 2025).

The hidden size determines the dimensionality of the LSTM cell state, where larger values allow the model to capture more complex sequential patterns but increase computational cost. The dropout rate controls the regularization strength by randomly dropping units during training to reduce overfitting. The learning rate governs how quickly the model updates its weights during optimization, with lower learning rates offering more stable convergence. Additionally, the gradient clipping value prevents exploding gradients by constraining the maximum allowable gradient norm.

For TFT, there are several key parameters that can be tuned to increase performance (Lim et al., 2021). The number of attention heads and hidden units determines the model's capacity to capture complex temporal relationships and feature interactions. The dropout rate helps prevent overfitting by randomly deactivating neurons during training, improving generalization, with potentially worse model performance as tradeoffs. The learning rate and gradient clip influence convergence behavior, where smaller learning rates generally yield more stable but slower optimization.

The hyperparameter tuning itself will be done via Bayesian optimization to smartly choose parameter values that will be tested in the next iteration using a probabilistic model so that less training is needed (Snoek et al., 2012). In this study, Bayesian optimization is implemented via Optuna (Optuna, 2025), which automates the hyperparameter value searching process by construction a model that estimates how different parameter configurations affect validation performance. Instead of having to try every single possible combination like grid search, Bayesian optimization adaptively chooses values to test by prioritizing configurations that are more likely to improve the objective function, based on previous trials. This makes tuning process significantly more computer efficient, whilst still maintaining good hyperparameter search behavior and granular values.

This is achieved by using an acquisition function, in this case Optuna used log expected improvement (logEI) for single-objective optimization to balance exploring new hyperparameter values and keeping promising hyperparameter values the same. As more trials are conducted, the Bayesian model will become progressively more accurate in searching for good hyperparameter value for a given dataset and task.

In this case we are running 100 hyperparameter tuning trial runs for all the models on 2014 validation data only due to compute constraints. If a trial run is judged by the algorithm to not have promising results it will be pruned early without training to its full epoch cycle.

3.8. Performance Measure

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{yi} \right| \tag{3.1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3.2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.4}$$

Model performance in this study is primarily evaluated using Mean Absolute Percentage Error (MAPE) as shown in Equation (3.1). MAPE measures the average percentage of forecast errors compared to the actual values, providing an intuitive understanding of how far predictions deviate from the real world on average. This metric is particularly suitable for regression and forecasting tasks, as it expresses accuracy in percentage terms and allows for easy comparison across models and datasets with different scales (Hyndman & Koehler, 2006). All graphs, yearly testing/validation results, and final results are shown using this metric unless otherwise stated.

Aside from that there are three other metrics commonly associated with regression problems that we are going to use. These metrics are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The formulas for these metrics are depicted in equation (3.2) –

(3.4). Unlike MAPE these are only going to be shown for the final averaged out results for each model.

Mean Absolute Error (MAE) measures the absolute difference between predicted and actual values. This metric provides a straightforward interpretation of prediction error in the same unit as the target variable. However, due to its un-normalized metric, it means that it's not particularly sensitive to small changes in value, such as those in smaller companies. It also doesn't really mesh that well with valuation prediction due to the fact that company valuation differs a lot between smaller and bigger companies, thereby rendering it not consistent and less informative in a dataset full of other companies.

Root Mean Squared Error (RMSE) penalizes larger prediction errors more heavily due to the squared error term. As a result, RMSE is particularly effective in highlighting cases where the model generates significantly different values from actual values. This metric aids in evaluating the model's overall reliability, particularly in financial forecasting tasks where large errors can lead to serious practical repercussions.

The coefficient of determination (R^2) measures the amount of variance in the target variable that is able to be explained by the model (James et al., 2013). A value of 1 would indicate that the model is able to explain all of the variability in the company valuation, whilst a lower value indicates progressively less explanatory power by the model. A value lower than 0 indicates that it is worse than predicting the average of the dataset. Unlike error-based metrics, R2 provides a more direct insight into the goodness of the model than just the magnitude of prediction errors.

By using a combination of MAPE, MAE, RMSE, and R^2, this study hoped to provide a more comprehensive view of the quality of the model through both relative and absolute prediction accuracy as well as explanatory ability.

CHAPTER IV
RESULTS AND DISCUSSIONS

4.1. Data Gathering Result & Dataset Creation

Table 4.1. gives an illustration on the dataset format and its data. Each row represents a unique data entry for each company on each quarter. Each entry contains general data about the company, the period of the financial statement, and combinations of several firm-level quarterly fundamental financial data with annual macroeconomic indicators.

Table 4.1. Sample of Dataset Created

Ticker	Economic Sector	Date	Market Cap (Millions)	Total Assets (Millions)	…	Inflation (%)
ASII	Consumer Non-Cyclicals	31-12-2024	198,369,410	472,925,000	…	1.57
BBRI	Financials	30-09-2023	612,177,119	1,851,964,853	…	3.7
GJTL	Consumer Cyclicals	31-03-2022	2,178,000	18,916,159	…	4.2

This dataset is composed of historical financial data taken from 928 publicly traded companies in the Indonesian stock market and the corresponding macroeconomic data for that year. Companies whose data are incomplete, only release annual financial statements, unable to be retrieved, or suspended from trading activities are dropped. This led to a final dataset consisting of 802 companies. This represents approximately 86.4% of the initially collected firms. The dataset spans from a time period between 1999 Q1 – 2024 Q4.

Because firm-level financial data are reported quarterly while macroeconomic data from our sources are published on an annual basis, they are aligned by assigning the corresponding annual macroeconomic values to all quarters within the same year. This ensures that all the data points have an appropriately associated macroeconomic variables to them.

Missing values in companies' financial data are usually caused by shoddy reporting during certain time periods for certain companies, the incomplete documentation on our sources, and/or the fact that certain companies have different reporting standards/templates. This is especially prominent on older data points and newer companies. These incomplete data points will be tried to be filled using methods outlined in Sub-chapter 3.4, while those unable to be filled will be discarded. This process resulted in a clean dataset suitable for further use in various machine learning or statistical analysis task.

Table 4.2. Dataset Summary

Item	Value
Initial companies	928
Final companies	802
Time period	1999 Q1 – 2024 Q4
Frequency	Quarterly
Fundamental variables	24
Macroeconomic variables	9
Total input variables	33
Target variable	Market Capitalization
Total observations	37372

There are 33 input variables in this dataset consisting of 24 fundamental variables and 9 macroeconomic variables. This is the result of filtering down the hundreds of fundamental and macroeconomic variables in the original source data used to make this dataset based on importance in previous research. Meanwhile, the output variable here is market cap as a by stand for company value. Table 4.2 details specific details regarding the dataset compiled in this study.

4.2. Data Preparation Result

After the dataset was finalized in Sub-chapter 4.1, additional data preparation steps were done to ensure compatibility and optimal training process for the machine learning models. These steps were done roughly the same way across models, with slight differences on XGBoost due to the different model architecture making it unsuitable for the same data inputs/outputs as the neural network models.

As a brief summary, the dataset was structured as a multivariate time series grouped by company and ordered chronologically using a continuous time index. Continuous variables were standardized using Z-score normalization, while categorical variables were encoded as integer identifiers, with label encoder to be used in neural network models. For neural networks, the target variable was log-transformed on a per-company basis to stabilize scale differences, whereas XGBoost did not require target transformation due to its tree-based architecture.

Model evaluation was conducted using a rolling-window validation strategy to reflect realistic forecasting conditions and prevent information leakage. Each evaluation cycle consisted of eight years of data, split sequentially into six years for training, one year for validation, and one year for testing. Neural network models used four quarters of historical data (encoder length) in a sequence to forecast market capitalization over the subsequent four quarters (prediction length). Due to its architectural constraints of not being able to sequentially learn temporal data, XGBoost needed to be fed explicitly engineered lagged features corresponding to the 4 previous quarters to predict the explicitly engineered next 4 quarters in the same data point. For more information, readers can look at Sub-chapter 3.4.

Table 4.3. Sample of Dataset after Pre-processing

Ticker	**Economic Sector**	**Time Index**	**Market Cap**	**Total Assets**	**…**	**Inflation**
0	16	27	16.42	0.064	…	0.80
1	27	27	12.23	-0.195	…	0.80
2	23	27	15.10	-0.174	…	0.80

As a result of these preparation steps, the dataset was transformed from a raw published dataset into a structured, normalized, and temporally consistent input suitable for both the tree-based and neural network-based models. A sample of the dataset after these preprocessing, specifically for the neural networks can be seen on Table 4.3. As can be seen from the sample, time index has been added to the dataset, the categorical variables have been encoded into integers using label encoder, the target variable (Market Cap) transformed using log transformation, and the rest of the numerical variables have been transformed using standard scaler.

4.3. Training and Validation Result

4.3.1. eXtreme Gradient Boost (XGBoost)

After hyperparameter tuning it was found that the values found in Table 4.4 represent the best combination of hyperparameter for the XGBoost model in this study. There are two model types evaluated here, the older one model per target model and the newer multi-output tree model. These values were found using a Bayesian optimization algorithm to efficiently find optimal hyperparameter after testing with the validation data. The validation data taken into account here is from the 2016 period with a training window from 2010-2015 as it is a relatively stable, but modern period for the Indonesian stock market.

Table 4.4. Optimal Hyperparameter for XGBoost Models

Model Types	Max Depth	Learning Rate	Boosting Rounds	Subsample	Min Child Weight	Gamma
One Model Per Target	1	0.027	4915	0.556	8	5.907
Multi-Output Tree	1	0.085	1724	0.233	7	5.326

Table 4.5 and 4.6 lists the XGBoost validation performance metrics aggregated across all the window cycle tested from the two different approaches. It seems that both models are not able to learn temporal patterns well in this dataset. This is reflected in the pretty high MAPE in general and the results of the hyperparameter tuning where we find that the best max depth possible for both approaches is 1. This indicates that when the model tries to learn overly specific patterns it tends to overfit, so the model decides to barely learn any patterns at all and overgeneralizes its prediction. Both approaches perform better on different years. However in general, the more traditional one model per target approach outputs slightly more accurate forecast with a MAPE of 32.64 ± 6.44% compared to the multi-output tree approach who has both a lower MAPE and higher standard

deviation of 32.99 ± 7.03%. Because of this all later mention of XGBoost will refer to the better performing one model per target approach unless otherwise specified.

Table 4.5. XGBoost One Model Per Target Validation Results Across All Years

Year	Average MAPE	Mean of Average MAPE ± σ
2006	31.35%	32.64 ± 6.44%
2007	35.63%	
2008	34.02%	
2009	31.64%	
2010	34.31%	
2011	30.38%	
2012	32.07%	
2013	31.00%	
2014	23.99%	
2015	36.20%	
2016	24.58%	
2017	26.63%	
2018	26.18%	
2019	25.35%	
2020	42.83%	
2021	30.94%	
2022	42.47%	
2023	47.96%	

Table 4.6. XGBoost Multi-Output Tree Validation Results Across All Years

Year	Average MAPE	Mean of Average MAPE ± σ
2006	25.21%	32.99 ± 7.03%
2007	35.41%	
2008	34.79%	
2009	32.52%	
2010	35.53%	
2011	30.56%	
2012	31.84%	
2013	30.81%	
2014	24.30%	
2015	36.31%	

2016	24.76%	
2017	27.79%	
2018	26.93%	
2019	26.36%	
2020	42.72%	
2021	32.79%	
2022	44.51%	
2023	50.67%	

4.3.2. Long-Short Term Memory (LSTM)

After hyperparameter tuning it was found that the values found in Table 4.7 represent the best combination of hyperparameter for the LSTM model in this study. These values were found using a Bayesian optimization algorithm to efficiently find optimal hyperparameter after testing with the validation data. The validation data taken into account here is from the 2016 period with a training window from 2010-2015 as it is a relatively stable, but modern period for the Indonesian stock market.

Table 4.7. Optimal Hyperparameter for LSTM Model

Hidden Size	Gradient Clip	Learning Rate	Dropout Rate
81	0.26	0.003	0.29

Using the optimal hyperparameter configuration, the LSTM model is retrained and evaluated across all rolling validation cycles as described in Sub-chapter 3.7. Model convergence behavior is first examined to ensure stable training dynamics. Figure 4.1 illustrates the training and validation loss curves for the best-performing validation year after 2010, which in this case is the year of 2014. This is done because the data before 2010 are pretty sparse and often have missing entries compared to the data after 2010. As can be seen, the training loss decreases steadily throughout the training process, this indicates that the model has successfully learned patterns from the dataset. The validation loss follows a similar downward trend to the training line, albeit starting from a higher start and having a more erratic pattern. Both loss data decreases sharply until around epoch 100 and then starts to level out, although still decreasing steadily each epoch. They began to converge around epoch 100-150, with both still decreasing slowly but surely. In this year, we have a somewhat rare case where the validation loss overtook the training loss slightly after they converge, possibly due to normal variations in the data.

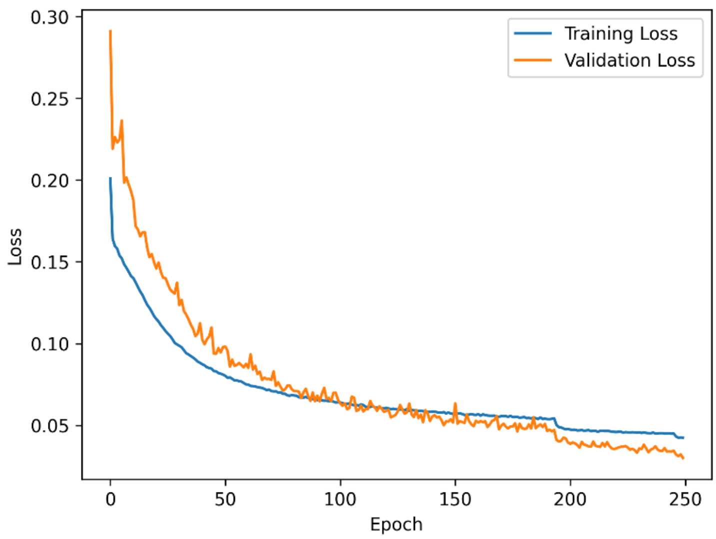

Figure 4.1. LSTM Best Training and Validation Convergence Plot (2014 Validation)

The training process kept going until epoch 250 without triggering the 40-epoch patience mechanism monitoring the validation loss. This indicated that if the training was ran longer with more epoch, the loss could still decrease below the endpoint that we see here. However, this is not done due to the heavy requirement of the model, where each time window the model needs anywhere between 18 – 63 minutes to be trained on the author’s RTX 2060 mobile laptop. This all adds up to around 22 hours of training time for the whole rolling window validation. Overall, the convergence behavior demonstrates that the selected hyperparameter configuration enables effective learning for the LSTM model.

Table 4.8 lists the LSTM validation performance metrics aggregated across all the window cycle tested. We can see that MAPE performance remains relatively stable around 4% - 8% across the years with the notable exceptions of the year 2022 and 2023 where MAPE spikes to around 25% - 30%. These two years are also the only time where the model training stops early due to the 40 epoch patience mechanism because they are unable to lower the validation loss further during training. We are not sure on what exactly causes this spike. An explanation would be the increasing amount of IPOs of startups in the Indonesian stock exchange causes the market to be saturated with lower market cap companies. These companies usually have really different financial and valuation characteristics due to the fact that they are not yet profitable or are trading at massive P/E ratios due to the expectation of future growth. The increased number of companies datapoints also can make the hyperparameters to be unsuitable for the model.

Table 4.8. LSTM Validation Results Across All Years

Year	Average MAPE	Mean of Average MAPE ± σ
2006	3.86%	8.57 ± 7.01%
2007	4.85%	
2008	4.40%	
2009	4.74%	
2010	7.40%	
2011	5.27%	
2012	5.20%	
2013	5.32%	
2014	5.16%	
2015	6.94%	
2016	9.41%	
2017	6.36%	
2018	7.17%	
2019	7.06%	
2020	7.88%	
2021	7.55%	
2022	25.64%	
2023	30.13%	

Another explanation after looking at the data would be the fact that it seems while the model is able to predict market downturns like the 2008 global financial crisis and 2020 COVID recession quite well, after around 2 years of these economic downturns we can see higher relative MAPE, specifically in 2010 and 2023-2024. There is even a spike 2 years after the 2014 general election, which transfers power to President Joko Widodo from President Susilo Bambang Yudhoyono, an event that comes with decreased economic growth. These could indicate that major economic changes in the training dataset could "poison" the model into outputting worse forecasts.

Aggregating and averaging out the data, we can get a mean error of 8.57% for LSTM with a standard deviation of 7.01%. The high standard deviation is due to the spike of errors in the 2022-2023 data.

4.3.3. Temporal Fusion Transformer (TFT)

After hyperparameter tuning it was found that the values found in Table 4.9 represent the best combination of hyperparameter for the TFT model in this study. These values were found using a Bayesian optimization algorithm to efficiently find optimal hyperparameter after testing with the validation data. The validation data taken into account here is from the 2016 period with a training window from 2010-2015 as it is a relatively stable, but modern period for the Indonesian stock market.

Table 4.9. Optimal Hyperparameter for TFT Model

Hidden Size	**H. Continuous Size**	**Attention**	**Gradient Clip**	**Learning Rate**	**Dropout Rate**
87	40	7	0.087	0.002	0.177

Using the optimal hyperparameter configuration, the TFT model is retrained and evaluated across all rolling validation cycles as described in Sub-chapter 3.7. Model convergence behavior is first examined to ensure stable training dynamics. Model convergence behavior is first examined to ensure stable training dynamics. Figure 4.2 illustrates the training and validation loss curves for the best-performing validation year after 2010, which in this case is the year 2010 itself. This is done because the data before 2010 are pretty sparse and often have missing entries compared to the data after 2010. As can be seen, both the training loss and the validation loss steadily go down throughout the training process. This indicates that the model has successfully learned patterns from the dataset. The validation loss here, weirdly has a lower number almost consistently across the entire training process. This might be due to the sparser validation data where the amount of validation companies are smaller than the training companies just due to the amount of companies who have inconsistent data reporting. This then leaves out only usually bigger, more established companies who tend to have a stabler stock price in the validation data and therefore results in a lower error loss number.

It can be seen that both losses decreased rapidly at the start of the training process up until about epoch 100 before starting to slow down. However, they never stopped decreasing even till the end at epoch 250. This indicated that if the training process were to continue it is likely that the model would have output better forecast still. It was decided, however, not to rerun these with even higher epochs due to the time constraints and the time and compute power needed to do the rolling window validation on TFT. Each training window requires 140-400 minutes of training time depending on the time window. This amounted to around 75 hours of training time on the author's RTX 2060 mobile laptop. Overall, the convergence behavior demonstrates that the selected hyperparameter configuration enables effective learning for the TFT model.

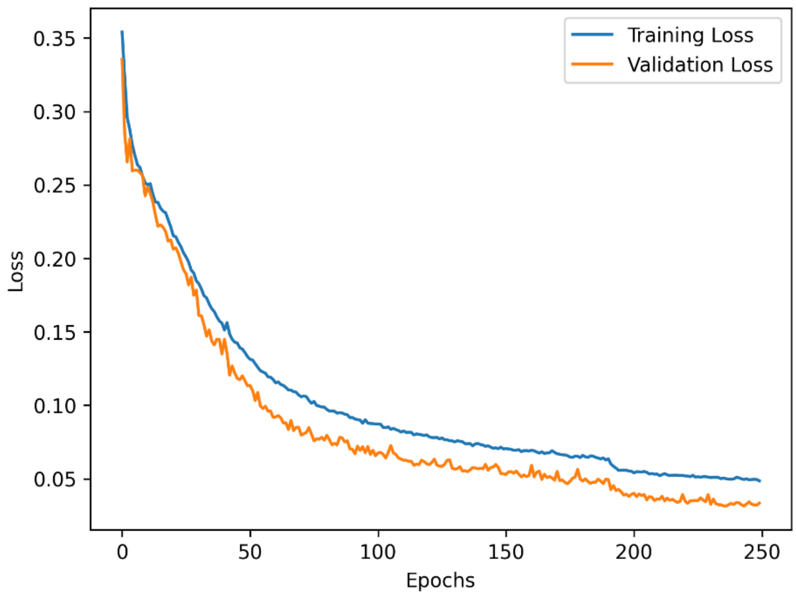

Figure 4.2. TFT Best Training and Validation Convergence Plot (2010 Validation)

Table 4.10 lists the TFT validation performance metrics aggregated across all the window cycle tested. The model seems to be able to forecast well across all the training-validation window with error rates ranging from 2.37% - 6.52%. There do not seem to be any major spikes in the validation error rates unlike the LSTM model, evidenced by the low 1.15% standard deviation for a 4.25% MAPE. Overall, TFT shows the lowest and most consistent error results in the validation set compared to the other models.

Table 4.10. TFT Validation Results Across All Years

Year	Average MAPE	Mean of Average MAPE ± σ
2006	2.37%	4.25 ± 1.15%
2007	3.79%	
2008	4.06%	
2009	2.57%	
2010	2.67%	
2011	4.38%	
2012	3.51%	
2013	3.43%	
2014	3.09%	
2015	4.85%	
2016	3.88%	
2017	5.33%	
2018	5.61%	
2019	5.24%	
2020	4.68%	
2021	4.99%	
2022	6.52%	
2023	5.52%	

4.3.4. Summary

In summary, both LSTM and TFT provide good forecasting ability for company valuations, while XGBoost predictive capability remains much to be desired. TFT slightly overperforms LSTM over most years and crucially is much more stable in forecasting during times of major economic changes. After looking at the training and validation loss graphs, both neural-network models also seem to still be able to be trained further to better fit the data by increasing the epochs on most training windows. This is, however, not done due to the immense computational requirements to redo the experiments with more epochs and the time limitation of this project. Table 4.11 shows a summarized comparison of each model's error rates.

Table 4.11. Validation Error Comparison

No.	Models	Mean of Average MAPE ± σ
1	XGBoost	32.64 ± 6.44%
2	LSTM	8.57 ± 7.01%
3	TFT	4.25 ± 1.15%

4.4. Testing Result

The error testing results achieved by all the models can be seen in Table 4.12. As can be seen, the TFT model achieved superior per year (4 quarters) MAPE performance of 7.27%, outperforming XGBoost (33.23%), and LSTM (11.91%). This result suggests that the TFT architecture successfully captured both short-term and long-term dependencies, making it suitable for forecasting tasks with multiple time-varying and static features. While the LSTM model's result is not quite as good as the TFT model it is still decently accurate, especially for such a long period of time. Meanwhile the XGBoost model struggled with higher error rate.

Table 4.12. Average MAPE Over Time Testing Result

Year	XGBoost	LSTM	TFT
2007	7.68%	6.41%	3.06%
2008	64.94%	4.26%	2.34%
2009	32.83%	5.73%	3.92%
2010	34.99%	12.21%	3.10%
2011	23.13%	4.89%	3.58%
2012	27.63%	4.84%	4.75%
2013	33.83%	4.04%	4.41%
2014	24.47%	4.60%	3.81%
2015	38.89%	5.79%	3.67%
2016	23.50%	4.46%	5.24%

2017	26.80%	6.54%	4.79%
2018	24.85%	5.69%	5.89%
2019	24.53%	6.00%	5.92%
2020	32.55%	6.57%	5.03%
2021	30.67%	6.01%	5.31%
2022	46.17%	28.60%	24.57%
2023	47.08%	44.81%	32.26%
2024	53.56%	52.91%	9.18%
Mean ± Std Dev	**33.23 ± 12.84%**	**11.91 ± 14.23%**	**7.27 ± 7.72%**

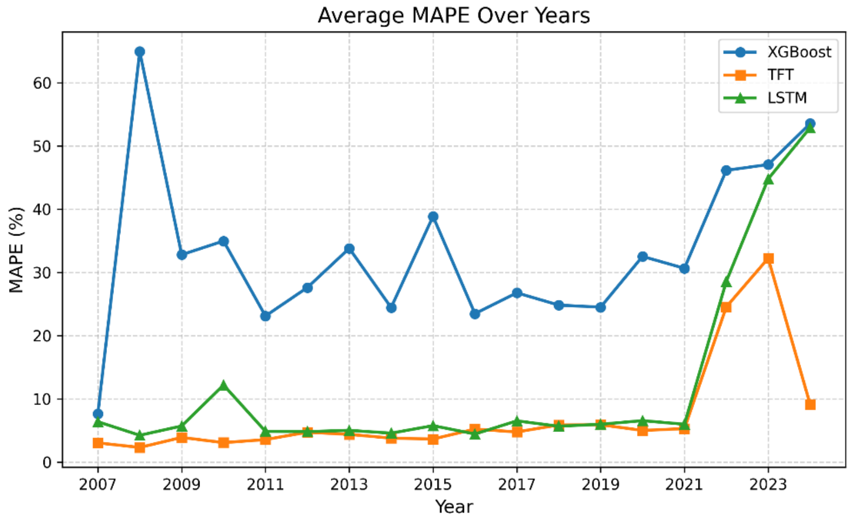

Figure 4.3. Average MAPE Over Time Testing Result

As can be seen from the graph in Figure 4.3, the MAPE values of both models fluctuate between year to year due to stock market volatility. This seems more apparent on the XGBoost model, as can be seen on the high peaks near 2008 and 2020 onward corresponding to the stock market downturns in the 2008 financial crisis and COVID-19 recession respectively. In general, the TFT model forecast seems a lot more resistant to shocks in the stock market and has a lower average MAPE. This indicates the superior adaptability of the TFT model in handling multivariate time series forecasting tasks, especially when the data exhibits significant shifts in temporal dynamics that simpler models cannot easily capture.

Table 4.13. Metrics of TFT (Best Model)

MAPE	MAE	RMSE	R^2
7.27%	3,423,747	51,446,146	0.936

Table 4.13 shows the comprehensive result metrics of the best performing model, which in this case, is the TFT model. Both MAE and RMSE are not easily interpretable with numbers this large and on such varying scales as market capitalization. However, the R^2 metrics show that the model on average can explain about 93.6% of the variation in its outputted prediction. Due to these results, it was felt that the TFT model exhibits satisfactory abilities to forecast company valuation reliably in medium-term timesteps. This consistent performance suggests its potential usefulness as a decision-support tool for investors and policymakers alike. Overall, the results reaffirm that integrating deep learning architectures like TFT into financial forecasting can enhance predictive power and robustness beyond what is achievable with traditional valuation methods or conventional machine learning methods such as XGBoost and LSTM.

4.5. Analysis of Feature Importance

After the models have been trained, the TFT model's internal mechanism will be used to interpret the impact of individual features on the output. Figure 4.4 visualizes the global feature importance across the dataset. Company financial variables are highlighted in blue, macroeconomic variables in red, and other additional variables in yellow. This study found that the top contributing variables for company valuation are features such as Market Cap, Return on Assets (RoA), and Net Debt. The importance of RoA and net debt aligns with traditional valuation theory, as companies with higher RoA generally are considered more efficient with their assets and capitals in general and so command higher market capitalizations, while companies with lower debt are more stable. Other important financial indicators like total capital and EBITDA seem to rate quite low, most likely because the model already derived this intrinsic information from other variables, most likely past market cap.

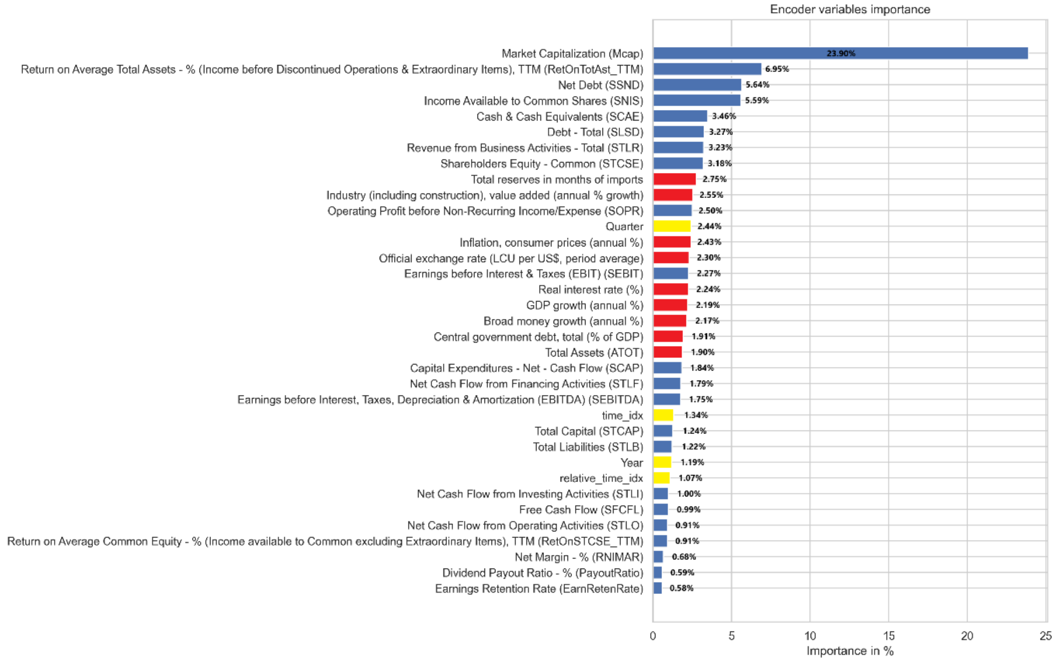

Figure 4.4. TFT Feature Importance Result

Macroeconomic variables seem to play a lesser supporting role in company valuation prediction with the most important macroeconomic variables appearing only in rank 9 (total reserves) and 10 (industry growth). Most other macroeconomic variables are clumped in the middle in terms of importance, usually each having ~2% importance in determining company valuation. This, however, does add up to a significant 20.44% role in determining company valuation. These highlights the significant role of macroeconomic conditions in shaping valuations. These results indicate that, while as expected, financial performance metrics play more role in predicting company valuation, the use of macroeconomic indicators are not without merit as shown in this analysis.

In Figure 4.5, it can be seen that while all the time steps play a significant role, the most important lag data is the data from 4 quarters ago. This indicates that the model assigns higher attention weights to information from the same quarter in the previous year, suggesting the presence of seasonal or cyclical patterns in companies' financial data. This behavior persists even to the point comparatively neglecting the lag data closer to the time of prediction. The second most important timestep does seem to be the newest timestep (-1 quarter) which is the closest data. The result highlights the TFT's capability to capture both long-term dependencies and temporal relevance, emphasizing the influence of recurring patterns in shaping future values.

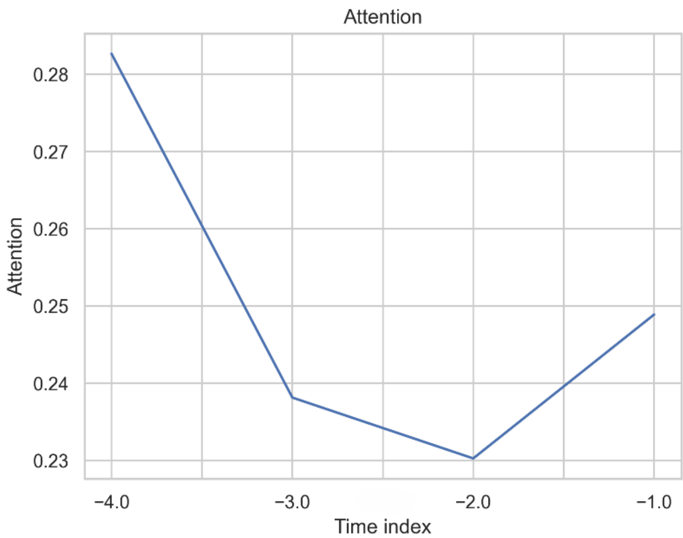

Figure 4.5. TFT Temporal Attention

CHAPTER V

CONCLUSION AND FUTURE WORKS

5.1. Conclusion

This study was conducted with three main objectives, which are dataset construction, valuation forecasting using machine learning models, and the identification of key features influencing company valuation. The conclusions corresponding to each objective are presented as follows:

The first objective was to construct a new dataset containing fundamental and macroeconomic data for publicly traded Indonesian companies in the Indonesian Stock Exchange (IDX). To achieve this, financial statement data from all Indonesian listed companies were scoured and integrated with relevant macroeconomic indicators covering the period 1999 Q1 – 2024 Q4. After data cleaning and filtering to process or remove incomplete data points, a final dataset consisting of 802 companies covering the period containing 24 fundamental variables and 9 macroeconomic indicators was successfully compiled. This dataset provides a comprehensive base for research relating to Indonesian companies financials in a machine learning or statistical manner.

The second objective was to develop and evaluate various machine learning models to forecast company valuation using the constructed dataset. Three models, namely Extreme Gradient Boost (XGBoost), Long Short-Term Memory (LSTM), and Temporal Fusion Transformer (TFT) were trained and evaluated to determine the best approach for this task. Market capitalization is forecasted every quarter up to a year in the future. The results show that TFT achieved the lowest Mean Absolute Percentage Error (MAPE) of 7.27%, outperforming the other models in prediction accuracy. The LSTM model still manage to perform quite well earning a respectable 11.91% MAPE. However, the XGBoost model fails to learn a significant amount of pattern and only managed a disappointing 33.23% MAPE.

The third objective of the study was to identify the key features that influence company valuation. This was addressed using the built-in interpretability mechanisms of the TFT model. Feature importance analysis identified past Market Capitalization, Return on Assets, Net Debt as the most influential variables, highlighting the role of both internal financial metrics in shaping valuations.

Macroeconomic variables play a lesser, but still significant, supporting role in shaping company valuation forecasts accounting for 20.44% of the feature importance score. These findings suggest that combining macroeconomic variables with firm-level data can improve valuation accuracy and provide meaningful insights for investors.

5.2. Future Works

While the study offers a valuable dataset and demonstrates a practical methodology, several limitations remain and open opportunities for further research in this area. The first limitation relates to the mismatch in data granularity between quarterly company data and annual macroeconomic indicators. While company fundamentals are available on a quarterly basis, most macroeconomic variables are only reported annually. This may limit the model's ability to capture shorter-term dynamic changes in macroeconomic conditions. Future research could address this issue by incorporating higher-frequency macroeconomic data or just higher frequency data in general.

This study also has the limitation of not including qualitative and market-based information that may affect company valuation. Factors like investor sentiment, news events, and developments particular to the industry are not reflected in financial statements or macroeconomic indicators. Future research could incorporate additional data sources, like sentiment analysis from news articles or social media, to enhance the input context and possibly improve forecasting performance.

Another limitation is that this research examines market capitalization as a proxy for company valuation. Although market capitalization is a useful variable to know, it does not provide a direct indication of investment performance or returns. This provides another avenue for future works to try deriving returns from the data, using that as target variable, and evaluating the model performance by its stock picking ability based on portfolio returns every quarter. This would enable a more straightforward evaluation of the practical investment implications of the proposed models.

Finally, future works could explore the use of different model architectures to improve prediction accuracy and robustness. Investigating various encoder and prediction lengths and analyzing more cross-company results could also yield further insights and enhancements beyond the present approach.

BIBLIOGRAPHY

Assagaf, A., Murwaningsari, E., Gunawan, J., & Mayangsari, S. (2019). The effect of macro economic variables on stock return of companies that listed in stock exchange: Empirical evidence from Indonesia. *International Journal of Business and Management*, *14*(8), 108. https://doi.org/10.5539/ijbm.v14n8p108

Baker, H. K., Filbeck, G., & Kiymaz, H. (2020). *Equity markets, valuation, and analysis*. John Wiley & Sons.

Bartram, S. M., & Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, *128*(1), 125–147. https://doi.org/10.1016/j.jfineco.2016.11.008

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., & others. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS Inc*, *9*(13), 1–73.

Chen, L., Pelger, M., & Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, *70*(2), 714–750. https://doi.org/10.1287/mnsc.2023.4695

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Damodaran, A. (2024). *The little book of valuation: How to value a company, pick a stock, and profit*. John Wiley & Sons.

Daniel, K., & Titman, S. (1999). Market efficiency in an irrational world. *Financial Analysts Journal*, *55*(6), 28–40. https://doi.org/10.2469/faj.v55.n6.2312

Dhanamaru, M. (2024). A study on impact of persistent inflation on stock valuations in current context. *International Journal of Research Publication and Reviews*. https://api.semanticscholar.org/CorpusID:273805213

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64–73. https://doi.org/10.1145/2500499

Dhochak, M., & Doliya, P. (2020). Valuation of a startup: Moving towards strategic approaches. *Journal of Multi-Criteria Decision Analysis*, *27*(1–2), 39–49. https://doi.org/10.1002/mcda.1703

Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: evidence from the European stock market. *Journal of Asset Management*, *22*(7), 507–538. https://doi.org/10.1057/s41260-021-00237-x

Fernández, P. (2023). Company valuation methods. The most common errors in valuations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.274973

Geertsema, P., & Lu, H. (2023). Relative valuation with machine learning. *Journal of Accounting Research*, *61*(1), 329–376. https://doi.org/10.1111/1475-679X.12464

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Hadrian, & Kusuma, G. P. (2023). Stock price prediction on Indonesia stock market with the influence of external factors using recurrent neural network with attention mechanism. *Communications in Mathematical Biology and Neuroscience*, *2023*.

Hartanto, S., & Gunawan, A. A. S. (2024). Temporal Fusion Transformers for Enhanced Multivariate Time Series Forecasting of Indonesian Stock Prices. *International Journal of Advanced Computer Science and Applications*, *15*(7). https://doi.org/10.14569/IJACSA.2024.0150713

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Husain, T., Sarwani, Sunardi, N., & Lisdawati. (2020). Firm's value prediction based on profitability ratios and dividend policy. *Finance & Economics Review*, *2*(2), 13–26. https://doi.org/10.38157/finance-economics-review.v2i2.102

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

IMF. (2025, January 17). *IMF DATAMAPPER*. Www.Imf.Org.

Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*.

Jain, A., Patel, H., Nagalapatti, L., Gupta, N., Mehta, S., Guttula, S., Mujumdar, S., Afzal, S., Sharma Mittal, R., & Munigala, V. (2020). Overview and importance of data quality for machine learning tasks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3561–3562. https://doi.org/10.1145/3394486.3406477

Jain, K. (2024). Economic indicators and stock market performance: An analysis of GDP, unemployment rates, and inflation. *IOSR Journal of Economics and Finance*. https://api.semanticscholar.org/CorpusID:273640326

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Konig, G., Molnar, C., Bischl, B., & Grosse-Wentrup, M. (2021). Relative feature importance. *2020 25th International Conference on Pattern Recognition (ICPR)*, 9318–9325. https://doi.org/10.1109/ICPR48806.2021.9413090

Kusuma, N. P. N., & Budiartha, I. K. (2022). The capital asset pricing model forecast using artificial intelligence. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, *5*(1), 808–819.

Kuzey, C., Uyar, A., & Delen, D. (2014). The impact of multinationality on firm value: A comparative analysis of machine learning techniques. *Decision Support Systems*, *59*, 127–142. https://doi.org/10.1016/j.dss.2013.11.001

Kwon, T. Y. (2025). Feature importance in linear models with ensemble machine learning: A study of the Fama and French five-factor model. *Finance Research Letters*, *71*, 106406. https://doi.org/10.1016/j.frl.2024.106406

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, *145*(2), 64–82. https://doi.org/10.1016/j.jfineco.2021.08.017

Li, T., Shin, D., & Wang, B. (2021). Cryptocurrency pump-and-dump schemes. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3267041

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal* *of Forecasting*, *37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

LSEG. (2025). *LSEG Data & Analytics*. Www.Lseg.Com/En/Data-Analytics.

Lundberg, S., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*.

Medina Hernández, J., Rodríguez, J. P., McMahon, C. R., Sequeira, A. M. M., & Eguíluz, V. M. (2025). Improving prediction region accuracy in marine animal movement with temporal fusion transformer. *Scientific Reports*, *16*(1), 297. https://doi.org/10.1038/s41598-025-29520-2

Mehdary, A., Chehri, A., Jakimi, A., & Saadane, R. (2024). Hyperparameter optimization with genetic algorithms and XGBoost: A step forward in smart grid fraud detection. *Sensors*, *24*(4), 1230. https://doi.org/10.3390/s24041230

Mokhtari, S., Yen, K. K., & Liu, J. (2021). *Effectiveness of artificial intelligence in stock market prediction based on machine learning*. https://doi.org/10.5120/ijca2021921347

Muhammad, S. D., Hussain, A., Ali, A., & Jalil, M. A. (2009). Impact of macroeconomics variables on stock prices: Empirical evidence in case of KSE. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1683357

Optuna. (2025). *Optuna*. https://optuna.org.

Patil, P. (2020). *Stock market prediction using ensemble of graph theory, machine learning and deep learning models* [San Jose State University]. https://doi.org/10.31979/etd.38nc-j52r

Phuoc, T., Anh, P. T. K., Tam, P. H., & Nguyen, C. V. (2024). Applying machine learning algorithms to predict the stock price trend in the stock market – The case of Vietnam. *Humanities and Social Sciences Communications*, *11*(1), 393. https://doi.org/10.1057/s41599-024-02807-x

Priel, R., & Rokach, L. (2024). Machine learning-based stock picking using value investing and quality features. *Neural Computing and Applications*, *36*(20), 11963–11986. https://doi.org/10.1007/s00521-024-09700-3

Saeed, A., Asif, R., Rehman, A., Hassan, S., Bharany, S., & Hamam, H. (2025). AI-Based Energy Management and Prediction System for Smart Cities.

Jordan Journal of Electrical Engineering, *11*(2), 1. https://doi.org/10.5455/jjee.204-1728140025

Scornet, E. (2020). *Trees, forests, and impurity-based variable importance*.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). *Practical bayesian optimization of machine learning algorithms*.

Steurer, M., Hill, R. J., & Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research*, *38*(2), 99–129. https://doi.org/10.1080/09599916.2020.1858937

Stoyanov, S. (2024). Using artificial intelligence to improve the efficiency of the market valuation method. *Finance, Accounting and Business Analysis (FABA)*, *6*(2), 217–227.

Tahir, A. N., & Danarsari, D. N. (2023). App-based investment platform and investment decision making: A study of retail investor behavior in Indonesia. *Eduvest - Journal of Universal Studies*, *3*(7), 1258–1272. https://doi.org/10.59188/eduvest.v3i7.859

Verma, R. K., & Bansal, R. (2021). Impact of macroeconomic variables on the performance of stock exchange: a systematic review. *International Journal of Emerging Markets*, *16*(7), 1291–1329. https://doi.org/10.1108/IJOEM-11-2019-0993

Wang, J., Xu, J., Cheng, Q., & Kumar, R. (2024). Research on finance credit risk quantification model based on machine learning algorithm. *Academic Journal of Science and Technology*. https://api.semanticscholar.org/CorpusID:269451211

Wang, X., & Sun, X. (2023). Effects of national easing monetary policy on Chinese enterprises' investment. *Revista de Cercetare Si Interventie Sociala*. https://api.semanticscholar.org/CorpusID:266274746

World Bank Group. (2025, September 18). *Indonesia*. Https://Data.Worldbank.Org.

World PE Ratio. (2024, December 19). *Global P/E ratios by country*. Https://Worldperatio.Com.

Xu, Y., & Keselj, V. (2019). Stock prediction using deep learning and sentiment analysis. *2019 IEEE International Conference on Big Data (Big Data)*, 5573–5580. https://doi.org/10.1109/BigData47090.2019.9006342

Zhou, G., Liu, L., & Luo, S. (2022). Sustainable development, ESG performance and company market value: Mediating effect of financial performance. *Business Strategy and the Environment*, *31*(7), 3371–3387. https://doi.org/10.1002/bse.3089

RIWAYAT HIDUP

Penulis dilahirkan di Kota Semarang, pada 14 Juli 2003, anak ke pertama dari 2 orang bersaudara pasangan Bapak Herman Wijaya dan Ibu Hartini Ronggo Warsito. Penulis telah menyelesaikan program sarjana dari Universitas Bina Nusantara pada tahun 2024 dengan IPK 3,90. Sebelumnya penulis bekerja sebagai back-end developer pada perusahaan PT Inti Utama Solusido (Pharos Group), di kota Jakarta, pada tahun 2024.

Pada tahun 2024, penulis melanjutkan Pendidikan pascasarjana sesuai dengan program master track BINUS dalam Program Studi Computer Science, BINUS Graduate Program, Universitas Bina Nusantara. Selamat mengikuti perkuliahan, penulis telah mengikuti beberapa kegiatan akademik. Pada tahun 2023, penulis mengikuti Konferensi Internasional ICCSCI yang dilaksanakan secara online dan mempresentasikan sebuah artikel berjudul “Tackling Clickbait with Machine Learning: A Comparative Study of Binary Classification Models for YouTube Title”. Penulis juga telah berhasil men-submit artikel ilmiah yang dihasilkan dari penelitian Skripsi pada Journal of Applied Data Science (Scopus Q3) dan Tesis pada Journal of Indonesian Economy and Business (Scopus Q3). Penulis dapat dihubungi melalui wijayakevin2003@gmail.com.