Prediction of dividend yields

Ana Catarina Fernandes Moreira
Engenharia de Redes e Sistemas Informáticos
Departamento Ciência de Computadores
2022

Orientador
Nuno Moniz, Professor Auxiliar Convidado, FCUP

Coorientador
Vítor Cerqueira, Investigador, Universidade de Dalhousie

Supervisor
Alexander Thomas, Diretor Ciêntífico , Creighton AI

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ______/______/_________

MSC

Declaração de Honra

Eu, **Ana Catarina Fernandes Moreira**, inscrita no Mestrado em **Engenharia de Redes e Sistemas Informáticos** da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação/relatório de projeto reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação/relatório de projeto, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação/relatório de projeto quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Ana Moreira

Porto, 30 de Setembro de 2022

Abstract

Investors in companies are interested in forecasting future performance for various reasons. Some investors look into the changes in stock price, while others focus on the dividends paid by the stock, which can themselves be a valuable source of income.

We address the problem of predicting the performance of a company, particularly the distribution of profits through dividends by predicting the dividend yield of publicly-traded Australian companies using machine learning methodologies, and a dataset which contains daily forecasts of dividends from various banks using traditional methodologies on Australian companies. We extract their predicted dividend yield and use its past values as a feature in the model.

Our goal is to predict the dividend yield's exact value through time six months in advance, through the usage of a time window model along with a regression approach, and achieve better results than the predictions from banks.

It is demonstrated that using machine learning to predict the future value of a dividend yield gives a lower error than the predictions made by various banks using traditional methods. Furthermore, we discovered that the bank's predictions can help the performance of the model during certain periods.

Resumo

Os investidores nas empresas estão interessados em prever o desempenho futuro por diversos motivos. Alguns investidores estarão interessados nas mudanças no preço das ações, enquanto outros estarão mais focados nos dividendos pagos pelas ações, que podem ser uma valiosa fonte de rendimento.

Abordamos o problema de prever o desempenho de uma empresa, particularmente a distribuição de lucros por meio de dividendos, através de prever o *yield* do dividendo de empresas australianas cotadas na bolsa. Usando metodologias de aprendizagem máquina e um conjunto de dados que contêm previsões diárias de dividendos de vários bancos, obtidas usando metodologias tradicionais nas empresas Australianas, extraímos o *yield* do dividendo previsto e usamo-lo no modelo.

Nosso objetivo é prever o valor exato do *yield* do dividendo ao longo do tempo com seis meses de antecedência, através do uso de um *time window model* juntamente com uma abordagem de regressão, e alcançar resultados melhores do que as previsões dos bancos.

É demonstrado que o uso de aprendizagem máquina para prever o valor futuro do *yield* do dividendo apresenta um erro menor do que as previsões feitas por vários bancos usando métodos tradicionais. Além disso, descobrimos que as previsões do banco podem ajudar no desempenho do modelo em certos períodos.

Acknowledgements

I would like to thank my supervisors Professor Nuno Moniz and Professor Vitor Cerqueira for all the support throughout this process, for all suggestions that improved my thesis and accepting to join me on this challenge.

I would like to thank the partner company, Creighton AI, Iain Buchanan and Alexander Thomas for this opportunity and for providing the data which allowed this work here presented, including their valuable advice throughout the research.

A special thank you to my parents, Soledade and Rui, because without them I would never be where I am and would not be who I am today. Downey, my special comfort.

A thanks to the rest of my family, my colleagues and everybody that supported me during this adventure of doing a master's degree and the master's thesis.

Contents

List of Figures

Acronyms

LR Linear Regression

kNN Nearest Neighbors

SNR Signal Noise Ratio

MAPE Mean Absolute Percentage Error

MAE Mean Absolute Error

ARIMA Autoregressive integrated moving average

MLP Multilayer Perceptron

SVM Support Vector Machine

DAE Denoising Autoencoders

RMSE Root Mean Squared Error

AGM Annual General Meeting

BPNN Back Propagation Neural Network

NN Neural Networks

CART Classification and Regression Tree

KI Knowledge Integration

Chapter 1

Introduction

1.1 Motivation

Predicting future financial performance of a company allows investors to make a profit. Growth investors try to predict the future price in order to make profits on price increases and avoid losses. Value investors look for a steady stream of income from dividends.

Knowing the future dividend in isolation is not enough to make estimates of future performance as the price is key. A dividend of $1 on a stock worth $2 is clearly different to a one on a stock worth $100. A commonly-used metric to avoid this issue is the dividend yield, which is the dividend value divided by the price of a stock, on the day the dividend was announced. This allows comparability between stocks of different prices, and effectively puts all the stock dividends onto the same scale, simplifying predictions.

Predicting dividend yields can be difficult due to how unpredictable the stock market can be, and because financial time series data is very noisy having a low Signal Noise Ratio (SNR)[11]. Furthermore, financial markets are influenced by human behaviour, which is itself difficult to predict[8]. The markets are non-stationary, affected by global events, resulting in highly volatile data. Still, there are investors who try to predict the dividend and dividend yield values.

We focus on the dividend yield value instead of the dividend value because the dividend value alone is not enough to predict the performance of a company as the price must to also be considered. The predictive power of the dividend yield on the performance of a company has been extensively studied and evidence exists of its benefits across different global markets [9][18][3].

1.2 Objectives

Our goal is to build a model and experiment with using the bank's predictions as features in the dataset. We will run the model with the predictions from banks, and also without them. Comparing these two approaches' performance we can then conclude if using the bank's forecasts is helpful to the machine learning process, and achieve better results than the bank's models.

We face several challenges. Firstly, besides the inherent unpredictability, financial data is prone to noise arising from deficient data collection. Much of the earlier data is input by hand, which results in data prone to input-level errors. Missing values need to be filled and erroneous values need to be corrected. This necessitates an extensive data cleaning step.

Secondly, stocks are not comparable with each other. Our data shows some companies with a much bigger valuation than others. Therefore, a process of data normalisation to increase prediction power is necessary. Furthermore, the price and the dividend value distributions are skewed towards smaller values. We do not want our model to make biased predictions due to the unbalanced distribution of stock data.

Additionally, we are dealing with a large time series dataset, it includes financial data aggregated from multiple data sources, given to us by the partner company, Creighton AI. It includes stock prices, dividend information and banks' predictions. Accordingly, we must be careful in the sampling strategy to build the model efficiently.

1.3 Contributions

We focus on a case study that predicts dividends from Australia. However, our methodology is to generalise the model to other markers. Therefore, we use data common in every stock market (prices, dividends and sector of the stock) to build a model which could be used in any country.

We describe an end-to-end pipeline that allows for the normalisation of the data. The normalisation resorts to the dividend yield value – a dividend-price ratio – which includes more future performance information. E.g., if the future dividend yield is high that means a high dividend value divided by a low-price value. This not only signals a low performance of the company in terms of their stock price value, but also that the company might decrease their future dividend because if their stock price is low, they will not manage to continue paying a high dividend value. Furthermore, because the prediction tools being used are machine learning-based, and the algorithms benefit from data being of a similar scale, this ratio normalises the data and reduces the chance of biased results: higher price values generally result in higher dividend values.

We also study the detection of correlations between the features. Our approach comprises two main parts:

- forecasting the dividend yield six months prior to it being known across multiple stocks using historical data and historical predictions from the banks; and
- a set of experiments which validate the proposed method, such as comparing the errors of our predictions versus those of the banks, and further naïve baselines such as using the last value of the dividend yield.

1.4 Structure

This thesis is divided into 6 chapters. Chapter 2 provides an overview of the main financial definitions important to understand this research and related work on the usage of machine learning applied to financial and time series datasets. Chapter 3 provides a description of the dataset and the case studies being used as baselines for us to compare our results. Chapter 4 provides the problem definition, the methodology adopted to solve the problem and pre-processing steps that were taken to fulfil the criteria defined. Chapter 5 provides a summary of the models and the experimental methodology used, followed by the experimental results and a comparison between our results and the case study. Finally, some remarks, conclusions and future work will be discussed in Chapter 6.

Chapter 2

Literature Review

During this project, an investigation was carried out into how machine learning methodologies are being used for the prediction of financial time series. A search for similar work on the prediction of dividend yields and its relevance to investing strategies was also completed.

2.1 Financial Components

There are three main financial components being used and they will be described in this section: stock, dividend and dividend yield.

2.1.1 Stock

A stock represents the ownership of a fraction of a company. This usually entitles the owner of the stock to a proportion of the assets from the company and profits proportional to how much stock they own. Stocks are bought and sold predominantly on stock exchanges and are the foundation of many individual portfolios. Investors often study the performance of a company and proceed to buy the stock of the company, if the company is producing earnings to share in the rewards.

On any given day, buyers and sellers meet on the exchange and transact. Transaction prices vary throughout the day, and movements can be very volatile as new information arrives (results are announced, news events, etc.). A final close price is calculated each day from a sample of trades towards the end of the trading session. These give a good indication of the transaction price on any given day, and these are what we use in the study.

In Figure 2.1 we have an example of of the daily stock price of Apple during 2022:

Figure 2.1: Apple stock price through 2022

2.1.2 Dividend

A dividend is the distribution of a part of the profits of a company to its shareholders. Paying a dividend is a choice by the company, and in periods of poor performance, they may decide not to pay a dividend at all. Or conversely, they may continue paying the dividends investors expect, to the detriment of future investment.

Dividends are a good performance indicator of a stock, and a record of consistent dividends over a long period of time is important to many companies and shareholders, because it is widely interpreted as evidence of consistent profitability. Studies show that investors who focus only on dividend-paying stocks could receive higher returns[3].

A dividend is announced as a value per share. For example, let us assume an investor owns 1,000 shares of ABC Ltd. The company at the end of the quarter calculates its earnings and the board of directors declares a dividend. ABC Ltd. announces the dividend payment of $0.5 dollars. The investor therefore receives 1,000 (shares) x $0.5 = $500 dollars.

For a shareholder to be eligible to receive the dividend it must own the stock before the dividend expiration date. Both the dividend value and its expiration date are determined by the company's board of directors.

The dividend announcement date comes before dividend announcements signaling confidence in the future of the company. Accordingly, an increase in the regular dividend (especially if it is unexpected) often has a positive effect on the stock price. In Figure 2.2, we have an example of the dividends given by Apple between February 2021 and August 2022, where we can observe their quarterly distribution.

Announcements of dividend pay-outs are generally accompanied by a proportional increase or decrease in the stock price of the company, therefore knowing its value in advance will give an indication of what may happen to the price. A company with a long history of dividend payments

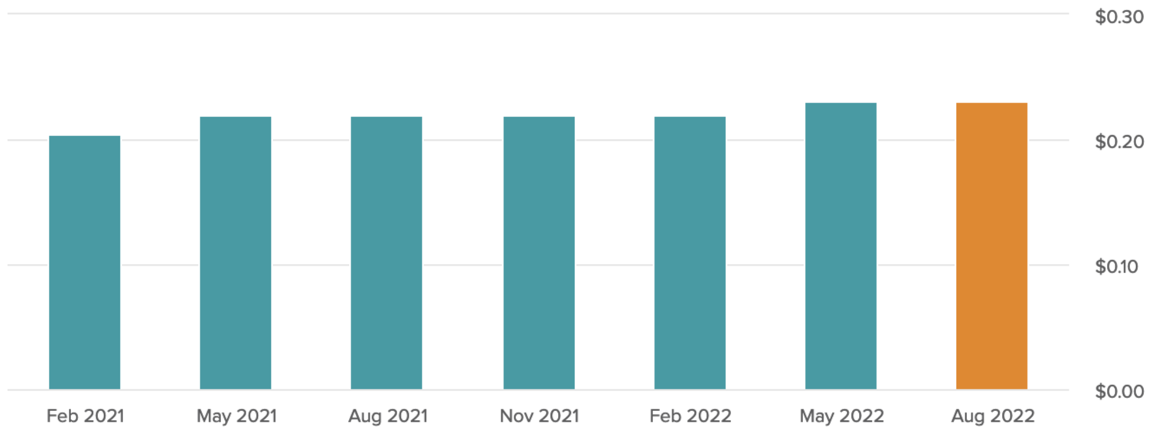

Figure 2.2: Apple dividends

that declares a reduction of the dividend amount, or its elimination, may signal to investors that the company is in trouble. E.g., AT&T Inc. cut its annual dividend in half to $1.11 on February 1, 2022, and its shares fell 4% that day.

In the end, the profit that will be achieved by an investor can be measured by a rate of return where a person who purchases a stock today for price P_t and sells it in n days for price P_{t+n} generates a rate of return on this investment of

$$r_{t+n} = \frac{P_{t+n} + \sum_{i=t}^{t+n} D_i}{P_t} \tag{2.1}$$

Where D_i is the dividend payments received during the period the stock was held.

2.1.2.1 Types of dividends

There are several dividends that a company can give.

- An interim dividend is a dividend payment made before the Annual General Meeting (AGM) of the company and the release of final financial statements, therefore it is given two to four times a year. The companies in our dataset give interim dividends twice a year;
- The final dividend occurs only once a year and is announced when the company is certain of the financial performance of the year in question. Once a final dividend is announced, it cannot be revoked under any circumstances. So, it is rigid compared to the interim dividend, which the company can cancel; and
- Special dividends is a non-recurring distribution of company assets and occurs in more rare events: exceptionally strong company earnings results or when a company wishes to make changes to its financial structure.

The dividends used in our dividend yield ratio are the final and interim dividends. Special dividends are announced sporadically so these are not used in our research.

2.1.3 Dividend Yield

Dividend yields have long been used to evaluate the expected return to investment in common stocks. Ensuring standardised feature values implicitly weights all features equally in their representation. The dividend yield is a rate between the dividend and price of a stock, these two being correlated and affected by each other. As previously stated, a high dividend yield may indicate that it is not safe and that the rate might be cut in the future because the price is low. Most importantly, this measure allows for easier comparability between companies.

There are various ways to calculate this value due to price and dividend variables differing in frequency through time: the price value changes daily while the dividend value occurs in a half-yearly or yearly period. Moving averages can be used for the price value and different point-in-time prices as well. Most commonly, the following formula represents the ratio to calculate the dividend yield:

$$DY(t) = \frac{D(t)}{P(t)} \tag{2.2}$$

Where $D(t)$ is the dividend value known on t and $P(t)$ the stock price on t as well. This is the simplest form of the dividend yield. Other variations are not tried as our focus is not to research different types of dividend yields.

As mentioned above, this ratio makes the companies comparable with each other. In Figure 2.3 we have a couple of stocks prices on the day the dividend was announced from different companies in our dataset, and in Figure 2.4 we have the respective dividend announced.

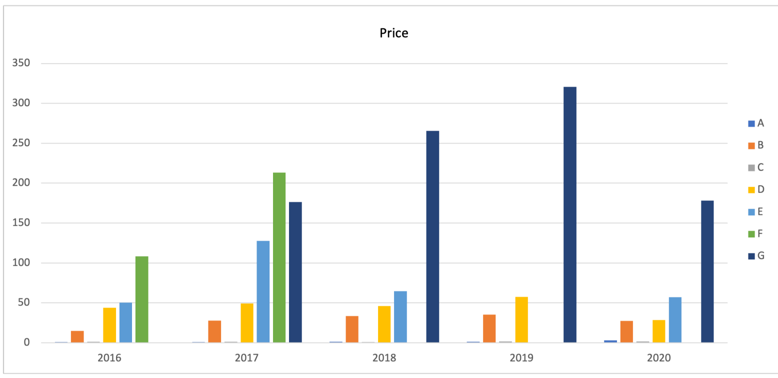

Figure 2.3: Example of prices from our dataset

The dividend and the price values have a different range of values with the prices shown going from 0 to 325 and the dividend values from 0 to 7.8. In 2017, even though F and G had

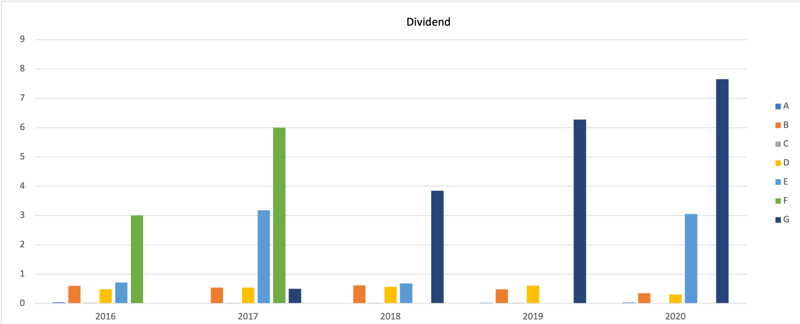

Figure 2.4: Example of dividends from our dataset

similar prices, the same does not occur in the dividend values. One could assume by the price value alone that company F and G performed well, but the dividend given by G was low.

If we divide the dividend by the price, we have the following values shown in 2.5. Notice how we can conclude that the performances in 2017 of A,B,C,D,E and F are quite similar, even though they had completely different prices and dividends given. This can lead the model to find patterns between these companies since they have a similar performance.

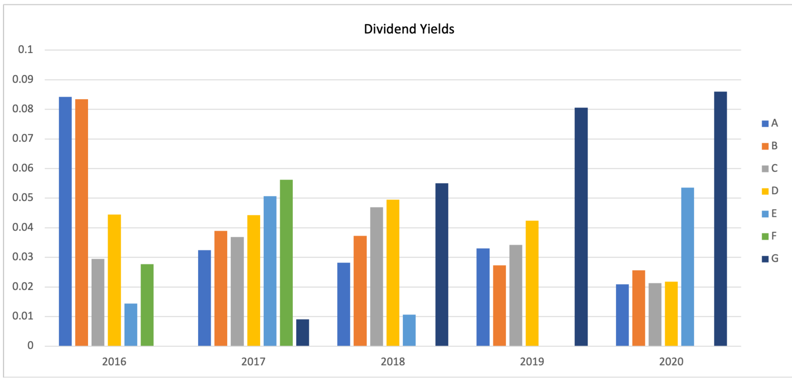

Figure 2.5: Dividend yields from previous examples

The ability of present-level dividends to forecast future stock returns is deeply researched in financial economics. The extension of the effectiveness of dividend-yield strategies in enhancing portfolio returns has been debated for many years. Some studies have suggested that little or no relationship exists between dividend yield and share price returns [10]. However, many other studies have identified a positive relationship between dividend yield and share price with [7] demonstrating that dividend yields do seem to contain information for forecasting future stock returns in the financial market from different regions. The so-called Dogs of the Dow strategy[18], a popular investment strategy that relies on dividend yield, has garnered considerable attention due to evidence suggesting that it produces abnormal returns.

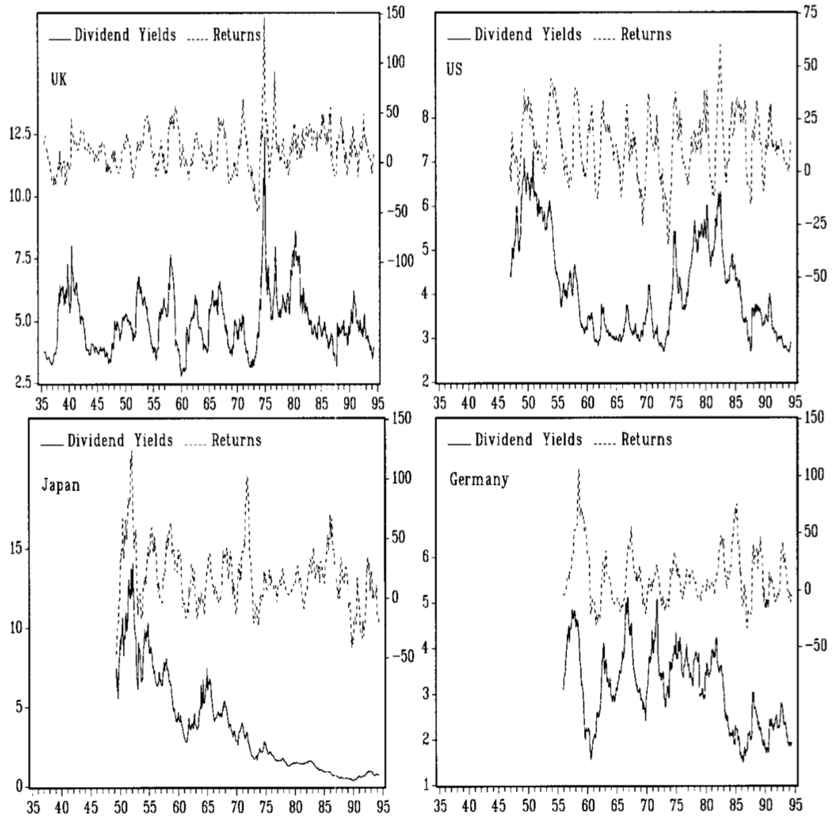

Figure 2.6: Correlation between dividend yields and stock returns found in [7] in the UK, US, Japan and Germany

2.1.4 Machine Learning

The objective of machine learning is to find patterns in a dataset. It uses algorithms that can create models from the data and make predictions. Machine learning tasks are categorised as supervised, unsupervised, or semi-supervised. In supervised learning, the goal is to learn a mapping function from a set of predictor variables to a target variable. When the predictor variables are known but the target variable is unknown, then this is a case of unsupervised learning. In semi-supervised learning, some of the target values are known.

In our research, we are dealing with supervised learning since we already know what past values of our target variable - the dividend yield - were and we can create a set of correlations between features with that target variable.

Supervised learning tasks include two categories of algorithms: classification and regression. The goal of classification is to predict a nominal target variable, i.e., categories. E.g., the problem of predicting if a stock value will increase or decrease is a classification problem, we are categorising the result. Meanwhile, regression is used to predict a numeric or continuous value, i.e., it quantifies the result, which the type of prediction we do onto the dividend yield value.

2.1.5 Time series dataset

The main features of many time series are trends and seasonal variations. In general, the trend is a systematic linear or nonlinear component that changes over time, and that does not appear to be periodic. Also, noise is an important feature for the analysis of a time series since it describes random variations or unforeseen events. Checking whether the time series has cycles is also important. A cycle is a component that reflects repeated but non-periodic events. If the data does not show trends, seasonal effects, or other time-dependent characteristics, then the time series is stationary.

2.1.5.1 Seasonality

Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur over time.

It is important to consider the effects of seasonality when analyzing stocks because it can have a big impact on an investor's profits and portfolio. A business that experiences higher sales during certain seasons may appear to make significant gains during peak seasons and significant losses during off-peak seasons. If this is not taken into consideration, an investor may choose to buy or sell securities based on the activity at hand without accounting for the seasonal change that subsequently occurs as part of the company's seasonal business cycle.

Thus multiple studies are carried out around the seasonality of financial data, including the dividend yield. In Figure 2.7 we can see a difference in the relation between average monthly returns (in percent) and dividend yield in January and for all other months for the period 1931 to 1978 found in [13].

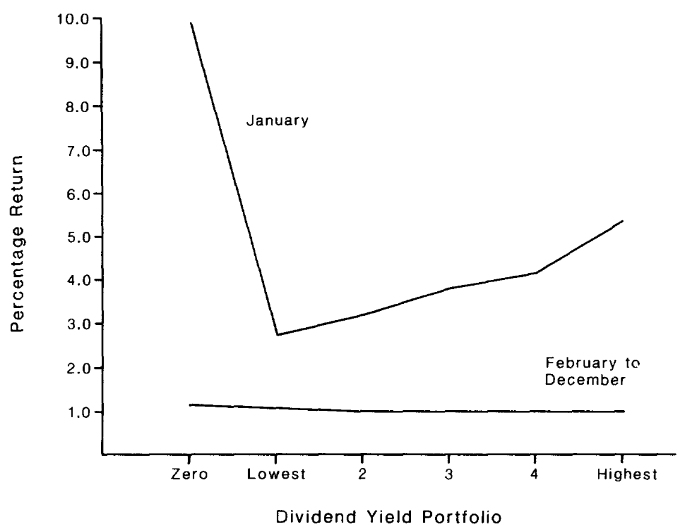

Figure 2.7: Relation between dividend yields and raw returns concentrated in the month of January found in [13]

2.1.5.2 Resorting to cosine and sine to find correlations

As time series contain cycles and the cosine and sine functions are cycling functions themselves these can be used to find relationships between features through time. Imagine the following random trend in Figure 2.8:

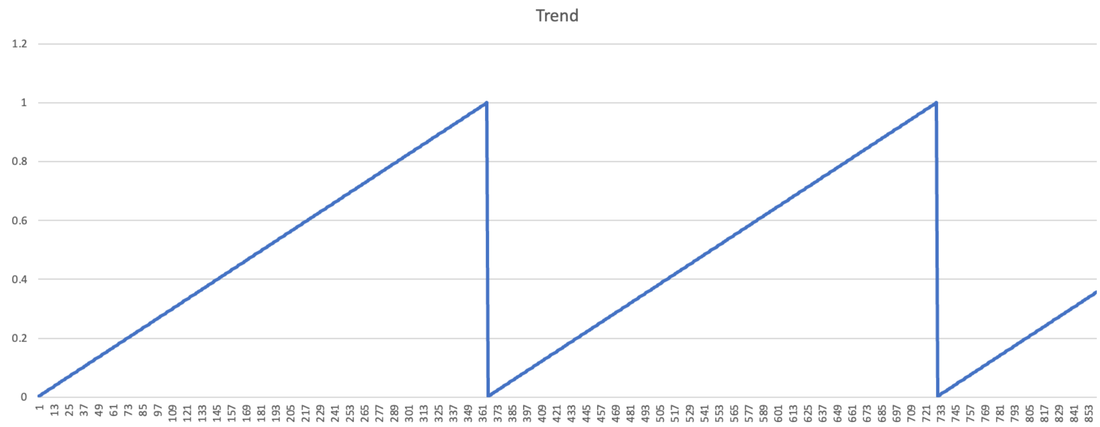

Figure 2.8: Trend example

There is a periodic repetition of the high point in the y axis being 1 and 0. If we overlap the trend graph with the cosine and sine functions we can see correlations in cosine function when y=1 and in sine function when y=0:

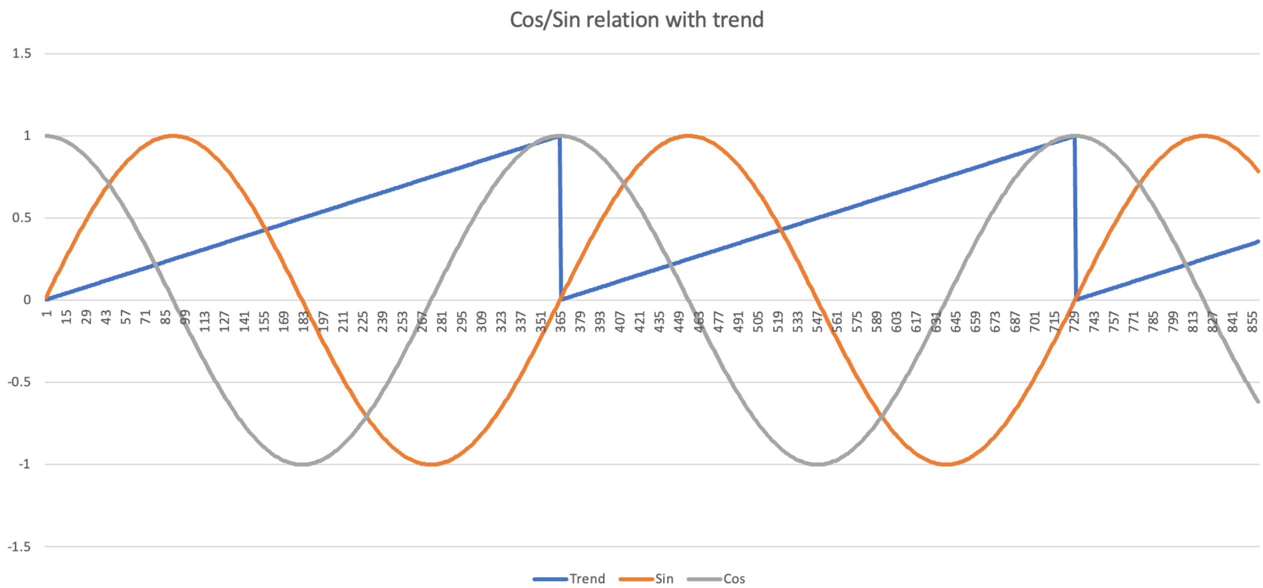

Figure 2.9: Relationship between the random trend, the cosine and sine functions

Because we have trends in our model, it is useful to use these functions to find correlations and improve further the model.

2.1.5.3 Time window model

The process of machine learning can be divided into two main parts. The first, which is training, starts from a dataset and a machine learning algorithm trying to understand what the features are and how they correlate to one another, resulting in a model. The second part is where the model is used to predict the target of unseen observations, based on the information it has learnt during the training process.

Due to the nature of it being a time series dataset, we can use data at the previous time step (training segment) to predict the data at the next time-step (test segment), i.e., use data within multiple months to predict the next month's data.

There are two main approaches to creating the training data to create the time series model:

- **Sliding window:** where a fixed size of the training data is set, changing the starting point in time in every iteration; and
- **Expanding window:** the training data grows in every iteration accumulating information through time

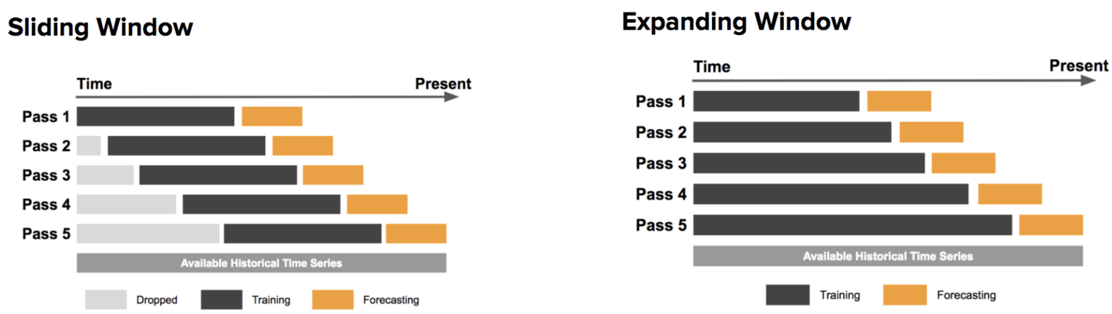

Figure 2.10: Two main window models to create the training data

Our focus is on an expanding window as the accumulation of historical data can be more benefiting to forecasting financial data than of a sliding window.

2.1.5.4 Simple Forecasting

While there are a wide range of forecasting methods, in the experimental component of the project we use two simple methods to forecast future dividends, the average and the naïve methods:

- **Average method**: The forecasts of all future values are equal to the average (or "mean") of the historical data, it can be a mean applied to all the past values or just the mean of the last x values in the past data; and

- **Naïve method**: For naïve forecasts, we simply set all forecasts to be the value of the last observations without adjusting them.

2.1.6 Evaluation metrics

The metrics used to evaluate this research of financial data are fundamental to studying the performance of our model. In financial data we are dealing with features of small scale and further in Chapter 3, where we study the distribution of the dividend and dividend yield values, we will see that these concentrate on smaller values. This means that using an error such as the Mean Absolute Error (MAE) can lead to misleading conclusions about the performance: if the values we are dealing with are small then the absolute error will be small as well.

Therefore is crucial to use a relative error measurement such as the Mean Absolute Percentage Error (MAPE). MAPE was defined by Armstrong and Collopy[2] and it is often used because of its very intuitive interpretation in terms of relative error and provides the error in terms of percentages [4].

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{2.3}$$

Where A_t is the actual value and F_t the forecasted value.

2.2 Related Work

In the financial world there is a field of quantitative analysis where the use of mathematical and statistical methods is prominent in the investment process. They decide to dig through big data and developing advanced algorithmic models for their predictions, instead of relying on intuition and old-fashioned approaches.

Such a task is not easy, [11] discusses the feasibility of creating machine learning-driven investment strategies with the main concern being the low Signal Noise Ratio (SNR) of financial data. Investors generally divide asset price movements into two components: signal and noise. Signal is the portion we can understand, model, and predict. For example, companies that reported good earnings have historically seen their stock prices go up, and the average amount by which the stock prices of the companies have risen is considered the signal. The noise component, on the other hand, consists of the unpredictable component of price movements. As financial movements are directly affected by human behaviours and these are naturally unpredictable, we end up with noisy financial datasets with periods of high volatility, such as the 2008 stock crisis, or in more recent events, the COVID-19 pandemic.

In the next section we look into public studies on predicting the financial market.

2.2.1 Traditional vs machine learning financial predictions

Studies such as the one in [16] compares the Autoregressive integrated moving average (ARIMA) model and other machine learning techniques such as Linear Regression (LR), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and a mix of Denoising Autoencoders (DAE) and SVM. ARIMA is a popular and widely used statistical method for time series forecasting. However, when predicting the S&P 500, Dow 30, and Nasdaq indices, the ARIMA accuracy of 59% was beaten by the accuracy of the other models, with the DAE-SVM combination performing the best results with an accuracy of 69%.

A similar comparison is present in this thesis since the datasets we use include predictions from banks using traditional investing methods, and there is an expectation that machine learning results will be better than those provided in the original datasets.

We extend this research by not only comparing the forecasting from banks with our forecasting, which is machine learning based, but we use those same predictions in our machine learning model. Furthermore, we observe if using predictions resulting from traditional investing approaches as features in our model could contribute to a better outcome of prediction accuracy.

2.2.2 Financial Series Prediction

One of the main characteristics of market forecasting is that we are dealing with time series datasets. In general, good predictions can be made for the next day or week, since not much movement usually happens during these periods. However, when it comes to predictions for the next months the task gets harder.

An example is a study by Alexiei Dingli and Karl Sant Fournier [5] where they approached the stock's price forecasting problem for the finance and technology market through different questions:

- Will the price go up or down?
- How much will the price change?
- What will be the actual price value?

Our work relates more to the two last questions; we do not want to only discover if the dividend yield increases or decreases compared to the previous one (a classification problem), we want to know by how much and if we can get close to its actual value (a regression problem).

Whilst using Root Mean Squared Error (RMSE) they tried several regression algorithms to end up finding better results with the LR and SVM. Their best result was predicting the real price the next day in the stock market using the LR method achieving an RMSE of 0.011706 but this value increases for the weekly and monthly results, each respectively having a RMSE of 0.02047 and 0.043636 [5].

Another study done by Zhang Yudong and Wu Lenan [20] combines bacterial chemotaxis optimization (BCO) with a Back Propagation Neural Network (BPNN) to forecast the S&P500 between one day and 15 days ahead. They achieved better results against the benchmark equivalent to just using BPNN.

Further phenomena around peculiar points in time can be analysed in time series datasets. In our case a lot of changes in the predictions of our data sources happen before the dividend expiration day, possibly because the value of the stock also changes during this period.

A study by Suman Saha, Junbin Gao and Richard Gerlach [17] around this phenomenon aimed to predict the stock movement on the dividend expiry day.

[12] states that the price drop on the ex-dividend day is less than the dividend amount and positively correlated with the corresponding dividend yield in the US stock market and [1] expresses that the price change on ex-dividend day is influenced by the dividend yield and franking credit in ASX (same stocks as ours).

Having these observations, the previous study determines the existence of a non-linear relationship between input features and price movement on the ex-dividend day, making machine learning more suitable for predicting the stock movement due to their ability to capture non-linear relations [17].

We observe the relationship between dividend yield and the announcement of the dividend having the time distance in days as a feature in our model.

2.2.3 Prediction of dividend/dividend yield through Machine Learning

There is not a lot of public investigation about the prediction of dividends using machine learning specifically, or for dividend yields, although, this is where we add a substantial contribution, and the further investigation on whether the predictions done by banks can help our machine learning model.

We did find a study by Jinhwa Kim, Chaehwan Won and Jae Kwon Ba [14] where they use a statistical regression model developed by Marsh and Merton [15] as a benchmark for predicting dividends in the Korean Exchange market between the periods of 1980 to 2000. Their chosen models were Neural Networks (NN), Classification and Regression Tree (CART), and Knowledge Integration (KI).

By measuring their performance using a 'tolerance level' percentage which means the maximum deviation of the forecasted dividend from the actual dividend. They beat their benchmark model in every tolerance between 1% and 50% with the KI model demonstrating the best performance.

We believe that it is difficult to accurately predict the dividend yield six months in advance, but we expect our predictions to improve on those from the banks, or from baseline approaches such as using the last known value.

Chapter 3

Description of Dataset and Case Study

In this section, we present a brief description of the data used and some key conclusions from a first exploration before proceeding to the proposed normalisation, and more experimental tasks, along with the baselines we want to beat.

Our task is to **predict the dividend value six months in advance** and compare our results to the bank's predictions. Furthermore we extract the past predictions of the dividend yield from the banks, and introduce it to our model with the goal of answering the question of **whether bank's predictions are helpful to the machine learning process**.

3.1 Description of Dataset

There are a total of 469 stocks present in the dataset. Because the core of the dataset is dividend predictions from banks, we have per each row:

- a dividend prediction done on a particular day for a dividend that will be known on an announced date for the appropriate stock;
- the price value of the stock on that same prediction day; and
- if it has been already announced, we have the actual dividend value.

We have three predictions from banks: UBS[1], Macquarie[2] and one that is the average of several predictions from different banks created by UBS. We call these predictions Consensus.

Not every bank gave us all the data in the same format, or with the same features: Macquarie does not have the price information associated while UBS and Consensus datasets do not have the actual values of the dividends. We join the datasets together so that they would fulfil each

[1] https://www.ubs.com

[2] https://www.macquarie.com

other's data, with three columns containing the values of each prediction done by the banks on a day for a stock if the prediction exists.

Even though our dataset only contains stocks from Australia, we want to achieve results that could be applied to any region. Thus, we want to select features with information that is **common** in all stock markets. so the columns we are selecting from each dataset are prevalent to any company in any stock market, e.g., the price and sector of a stock:

Feature	Description
run_date	day of prediction
dps_expected_m	the dividend prediction done by Macquarie on the run_date
dps_expected_c	the dividend prediction done by Consensus on the run_date for the stock id
dps_expected_u	the dividend prediction done by UBS on the run_date for the stock id
dps_actual	the actual value of the dividend known on the announced_date
price	price of the stock on the run_date
announced_date	the date when the dividend value is known
run_year	year of the dividend
id	stock identifier value

These features are the base, from these, we extract the expected **dividend yield**, a unitless ratio that allows comparability power between stocks, and consequently makes it useful for the forecasting model.

With further feature engineering, we manage to build an expanding time window to see the results over the years, representing the evolution of the results for each month along with a change in the number of stocks being predicted, and the frequency of predictions increasing. These last two factors are interesting in terms of data analysis because they allow us to observe changes in the performance when more stocks are added to the dataset, and when the frequency of prediction increases.

3.1.1 Sectors

A sector is an area of the economy in which businesses share the same or related business activity, product, or service. Sectors represent a grouping of companies with similar business activities, such as the extraction of natural resources and agriculture. Dividing an economy into different sectors helps economists analyse the economic activity within those sectors and can culminate in correlations between the sector and the performance of the stocks. In Figure 3.1 we find sectors included in the dataset:

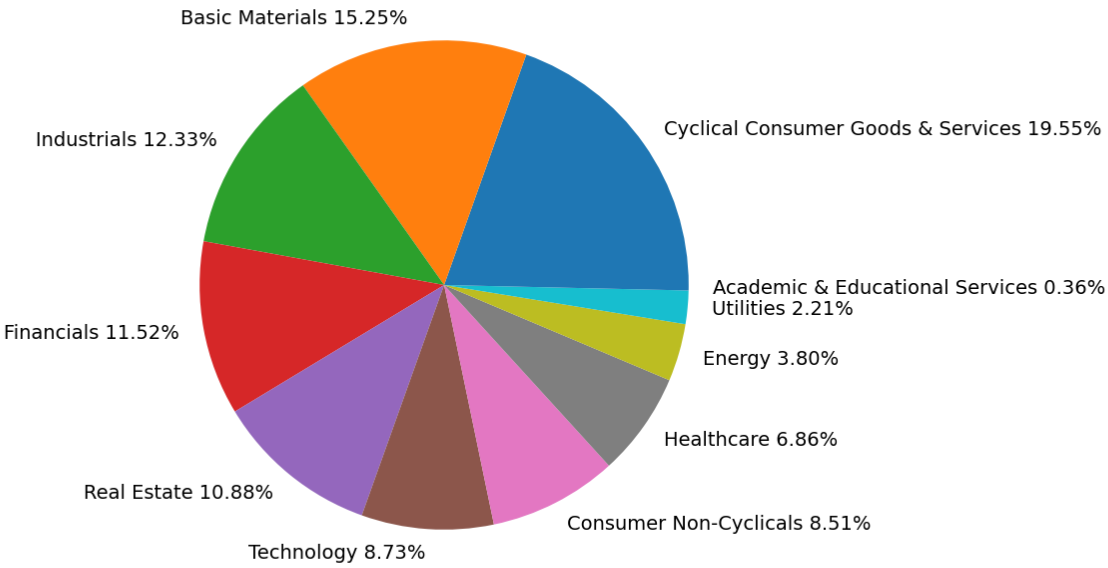

Figure 3.1: Sectors in the dataset

The sector *Academic & Educational Services* is the smaller subset of data present while *Cyclical Consumer Goods & Services* is the biggest subset present although is not a prevalent sector.

3.2 Case Study

Our baseline consists of predictions from three sources: UBS, Macquarie, and the previously described Consensus. These predictions are made on 469 publicly listed stocks from the Australian stock market. These predictions focus more on the dividend value itself than on the dividend yield value, but knowing nothing about the criteria being used, they can turn out to also be good predictions of the dividend yield.

3.2.1 Performance

Each bank has a different performance that changes overtime between 50% and 200% MAPE, which are big errors. We know nothing of the strategy being used by these banks but one assumption we have is that relative error measurements are not being used by them, else a lower MAPE could exist. Nonetheless, is this the closest we can get to a comparison point between using classical investing approaches and machine learning strategies when predicting the dividend yield, because this is the only data available to us. In this section, we discuss each bank's MAPE of their predictions done six months ahead of the dividend announced date.

The sources have a low number of predictions in the first years because they were doing monthly predictions only and then changed to doing daily predictions: Macquarie in 2018, UBS and Consensus in 2017. Such change does not have a good impact on performance. Notice how because we are using these predictions as the base for our dataset, it means that we have periods with only one date record in a month for each dividend of a stock being predicted. Later, we explain how resorting to forward-filling methodologies we were able to further complete the missing data.

We can observe a difference in the number of dividend yields predicted by Consensus and Macquarie before and after 2008. We will go deeper into this difference after the following descriptions of each performance achieved by the banks.

3.2.1.1 Macquarie

Macquarie's predictions start in the year 2003 and overtime their number of stocks which dividend yields are being predicted grows until a peak of 247 stocks, with a significant increase in 2008. In Figure 3.2 we observe the performance starts well with MAPE under 100%, but that might be justified by the low number of stocks in their dataset because when they suddenly add more stocks in 2008 with their performance error reaching a 200% peak in 2009. Furthermore, an increase in the quantity and frequency of predictions in 2018 does not help decrease the error.

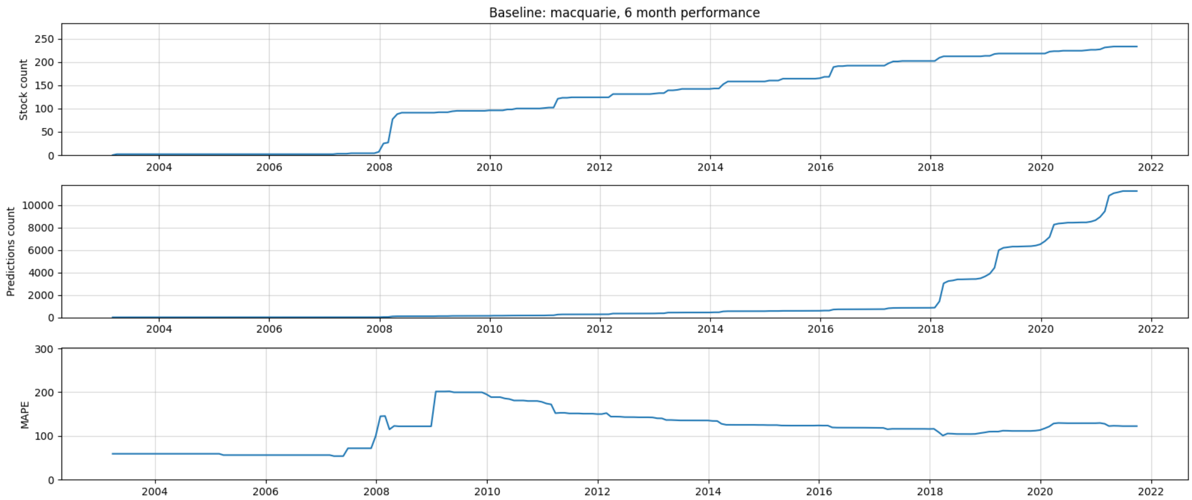

Figure 3.2: Macquarie's predictions

3.2.1.2 Consensus

Consensus's predictions start in the year 2001 and overtime the number of stocks being predicted grows until a peak of 467 stocks. The same observation done in Figure 3.2 can be done in Figure3.3: an increment in dividends being predicted in 2008 and increasing the number of predictions around 2016 does not improve the performance success in forecasting the dividend yield.

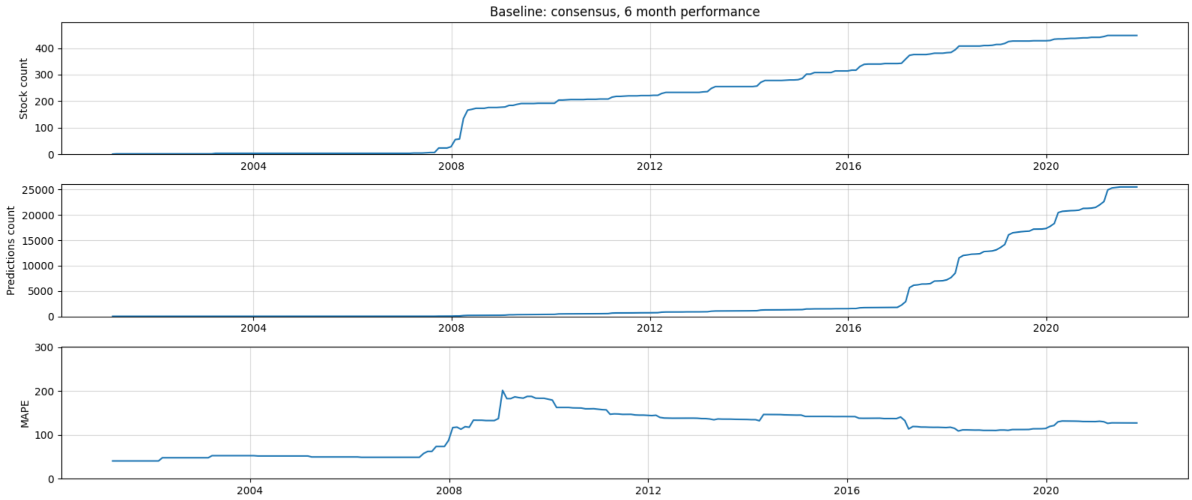

Figure 3.3: Consensus performance

3.2.1.3 UBS

In Figure 3.4, we have the UBS's predictions starting in the year 2015 and over time the number of stocks in the dataset grows until a peak of 220 stocks. Unfortunately, this source does not contain data before 2008 otherwise it would be interesting to see if its number of stocks and change in the performance would be better around that period, unlike the other sources. It does keep a steady MAPE around 100% except during 2015 when a change in the number of stocks in their dataset increases momentarily the error, but it lowers months after. It is the one with the most significant decrement in the error after the number of stocks and frequency of the predictions increases.

The overall error of the banks seems big, and adding a lot more stock to their dataset increases their error. This leads to various assumptions about the predictions from banks, like a different error being used in their valuation metrics, as previously stated. It can also be that the banks might not consider the weight the price can have on the dividend value, therefore the dividend yield that we can extract from them not being all that great, or else a lower error could be achieved from them.

But one important observation is the absence of big movements in their error over time which

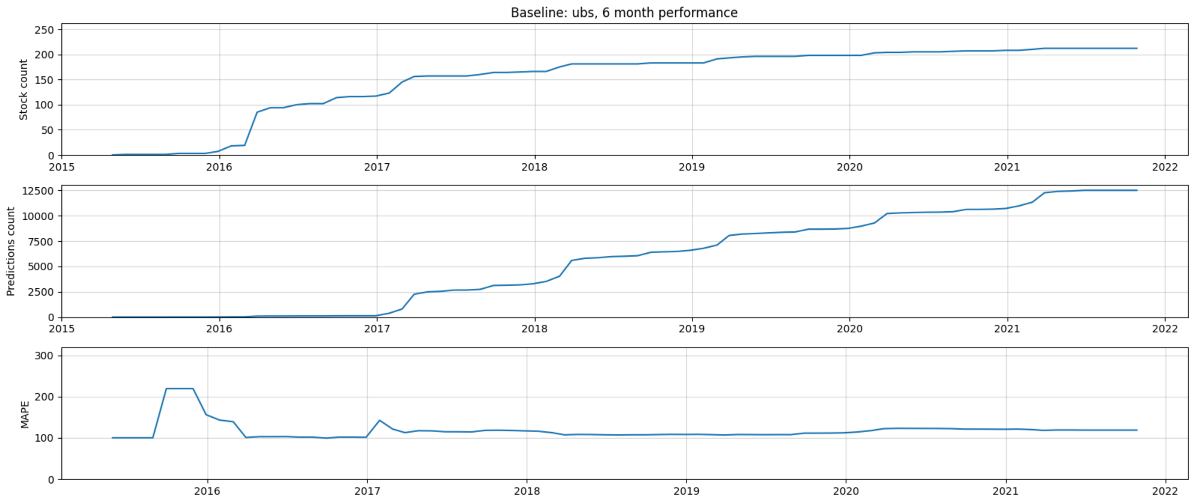

Figure 3.4: UBS performance

makes sense since, as previously mentioned, when making investment strategies investors want to reduce huge errors of their predictions, like we will see in the simple forecasting methods in the next section.

Having little data before 2008, it makes sense to start our time series model in that same year. The following table shows each MAPE from each bank starting in 2008:

MAPE results from the banks	
Bank	MAPE
Consensus	124.21%
Macquarie	135.70%
UBS	118.89%

With UBS with the lowest MAPE of 118.89% it is our goal to achieve a lower error.

3.2.2 Simple forecasting methods

Besides the predictions from banks, we can extract dividend yield's predictions using simple forecasting methods, and see if the results are better than the baseline, or even useful features for the experimental modeling process. There are naïve methods and an average method were used on the actual dividend yield value. The results calculated using MAPE can be seen in Figures 3.5, 3.6, 3.7 and 3.8.

- Use the last dividend yield value
- Use the second last dividend yield value
- Use the third last dividend yield value
- Average of the three last yield values

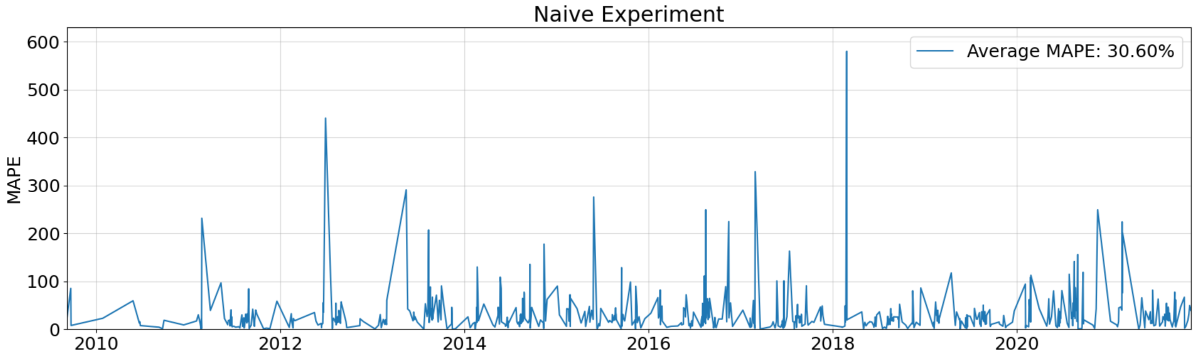

Figure 3.5: MAPE of naive predictions using the average of the last dividend yield

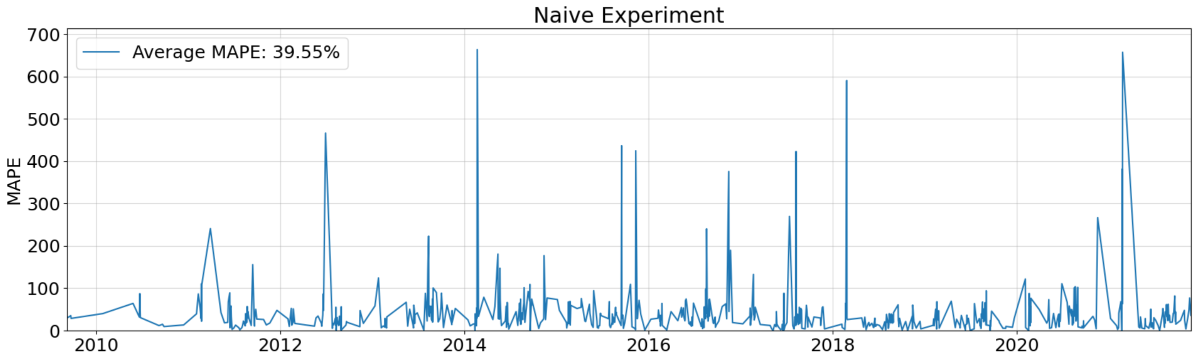

Figure 3.6: MAPE of naive predictions using the average of the second last dividend yield

Looking at the graphs we might have an initial assumption that using previous dividend yield values might not be a good idea because there are periods with huge disparities in the error. This is likely due to periods of high volatility in the market. E.g., the huge error present in 2021 might translate to big changes in the dividend yield value perhaps because of the impact the COVID-19 pandemic had on the stock market.

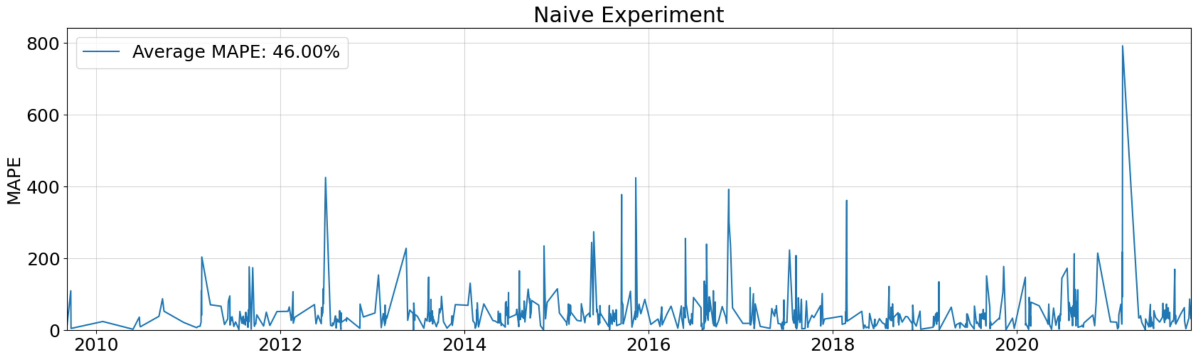

Figure 3.7: MAPE of naive predictions using the average of the third last dividend yield

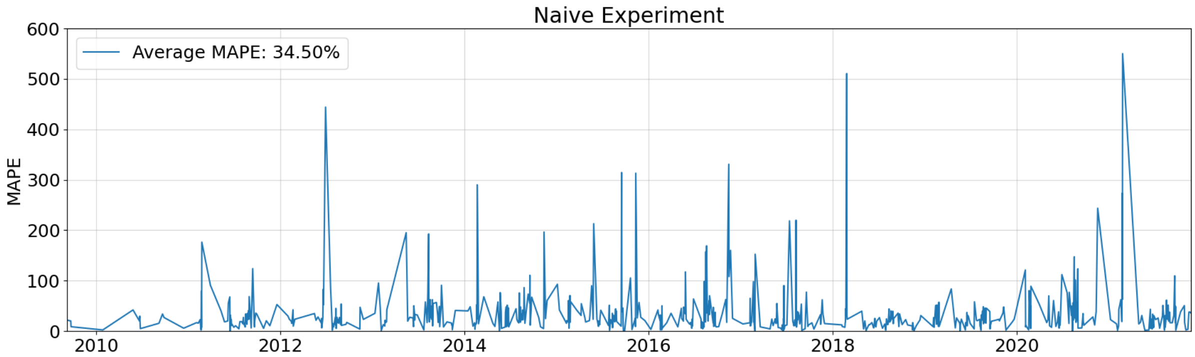

Figure 3.8: MAPE of naive predictions using the average of the three previous dividend yields

MAPE results from the naïve experiment	
Value used	MAPE
Last actual value	30.60%
Second last actual value	39.55%
Third last actual value	46.00%
Average of three last actual values	34.50%

This experiment shows a much **lower average error** than the one achieved by the banks, but a much **higher disparity in the predictions' errors**. High errors in their predictions is all investors want to avoid as gross mistakes can ruin investment strategies by banks. Even though the average errors of the naïve predictions are lower, banks might not choose this approach of selecting the latest values due to the risk associated when sizable changes occur between dividend yields.

We can also observe that the error increases as the previous value used becomes older, this confirms the fact that when making predictions in financial data these become poorer as the time horizon increases.

Chapter 4

Experimental Analysis

4.1 Methodology

In this section, we describe the methodology adopted to solve the target problem. We perform some exploratory data analysis on the data supplied by the partner institution, and we proceed to several data transformations so to accomplish a new, and richer dataset for further analysis. Such can be divided into three parts:

- **Data acquisition**: The first step is to understand the terminologies of the stock market, and read financial literature with the goal of understanding the dividend and dividend yield process. This is important to understand how possible is to predict the exact future dividend yield value, and set a goal for the project. To further refine the goal, and have a point of comparison of what we want to improve, we study the predictions from banks, and their performance. The following step is to put together the information from our sources of data. It consists of the same daily predictions used for the baseline that includes dividends predicted by banks using traditional investing methodologies, and respective forecast dividend yields that was extracted. This data also includes daily pricing of a stock, and sector information which we join in a single time series dataset. Some difficulties are encountered, predominantly because the information each source provides is different (some banks do not provide the stock price in their dataset or might not provide the actual dividend yield values), so we have to standardise the information between them;
- **Data cleaning and transformation**: With all the information available, and some domain knowledge obtained, the gathered data is aggregated into one dataset, to further analyse. Then, we clean the data, correcting or removing wrong records, and analysing them from the perspective of time series. The missing data is fetched from other pricing and dividend datasets available to us, which also are used to cross-check the data given by banks, and correct it; and
- **Modelling and optimisation**: The last step of the methodology is to obtain a regression

model that can predict the dividend yield value in our dataset, with data six months in advance to the dividend yield value being known. In the end, we end up with a general model where we do not want to favour high or low-value companies, so we can confidently include all the existing companies in our dataset.

4.2 Data Analysis

4.2.1 Distribution of the data

With the goal of normalizing the data, we study the distribution of the dividend and the dividend yield value. By normalisation we mean that we have a dataset with only dividend values whose distribution has a very long tail. We aim to reduce that long tail of more uncommon bigger dividend values. This task is difficult due to the very highly noisy data we are dealing with, so we do not expect to find the perfect theoretical distribution that matches the dividend yield distribution.

In essence, we want to answer the question "What is the probability that the values of dividend and the dividend yield present in our dataset could have been drawn from that probability distribution?". If that probability is high then it means that our feature has that same distribution as of the reference one, and we can afterwards make many data transformations such as a logarithmic transformation of the dividend yield values.

We start by looking into the distribution of the dividend value. It is crucial to see if there is an improvement in the skewness of the dividend amount distribution when compared to the dividend yield ratio, so we can further justify the prediction power of the dividend yield.

So first lets look at the distribution of the dividend value in Figure 4.1:

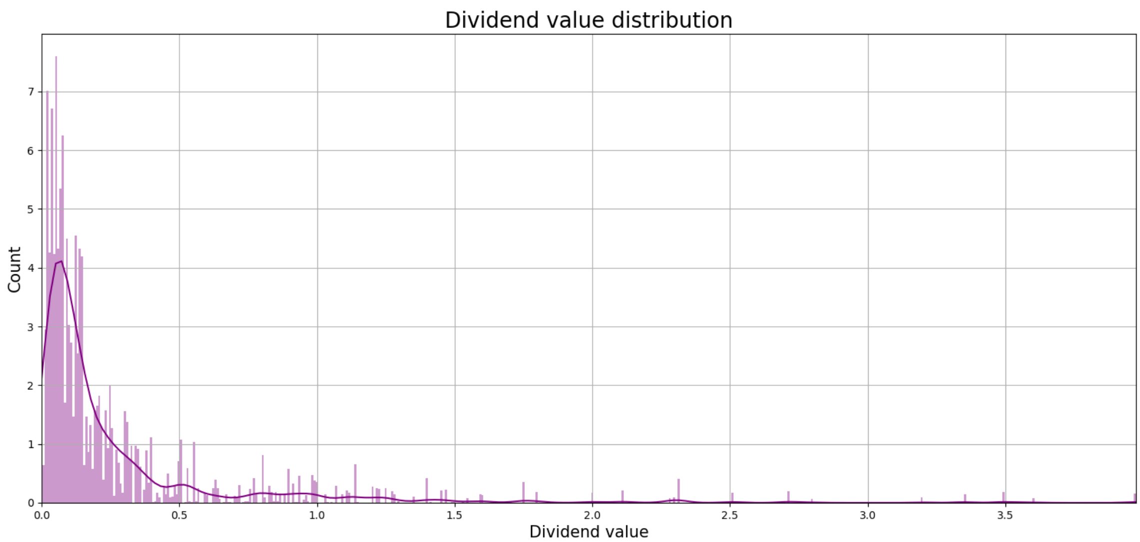

Figure 4.1: Dividend value distribution

With the line being drawn representing the kernel density which estimates the probability density function of a random variable. As we mentioned before, the distribution of stock data such as the price and the dividend value has a higher density on smaller values, and then a very long tail of bigger values due to the existence of a couple of companies with a high valuation. Therefore, we expect to see a distribution like the one plotted in Figure 4.2.

Comparing the distribution of our observed data with a selection of theoretical distributions, we can see a high similarity with the *lognorm* distribution and therefore it makes sense to do a logarithmic transformation.

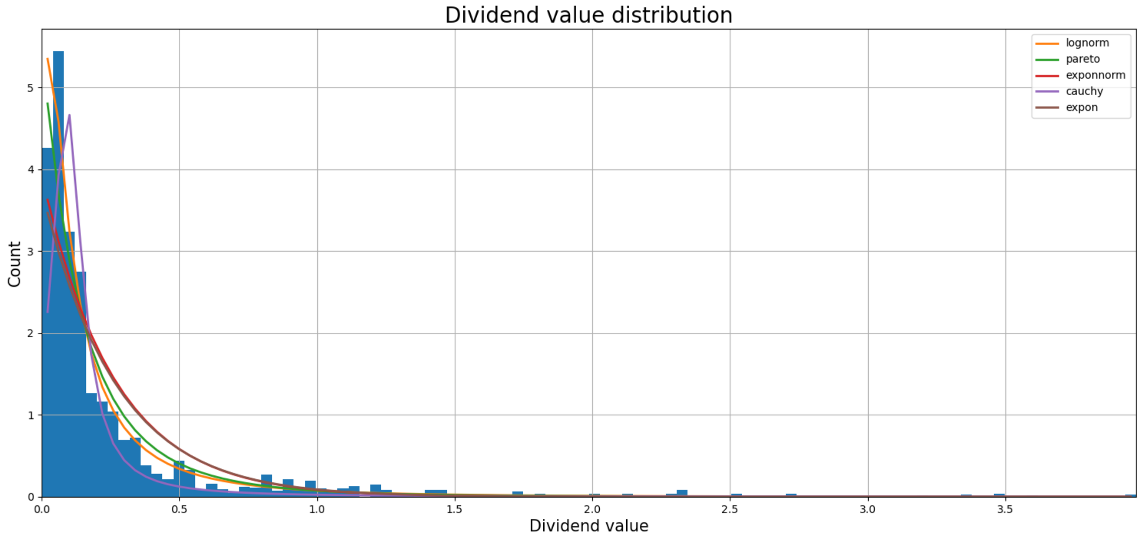

Figure 4.2: Comparison of distributions to the dividend distribution

Logarithmic transformation is a data transformation method in which it replaces each variable x with a $log(x)$. Our original continuous data does not follow the bell curve, and therefore we choose to logarithmic transform this data to make it as **normal** as possible, so that the statistical analysis results from this data become more valid. In other words, the logarithmic transformation reduces or removes the skewness of our original data. This makes sense given that our data is non-negative, and only includes companies that give dividends, so we do not have zero dividends.

So, after doing the logarithmic transformation, we have the distribution plotted in Figure 4.3. The tail has significantly reduced therefore we managed to reduce some of the skewness in the dividend value, but this is still not in a normalised form. It is now that we investigate the dividend yield values and see if we can further achieve more normalised data.

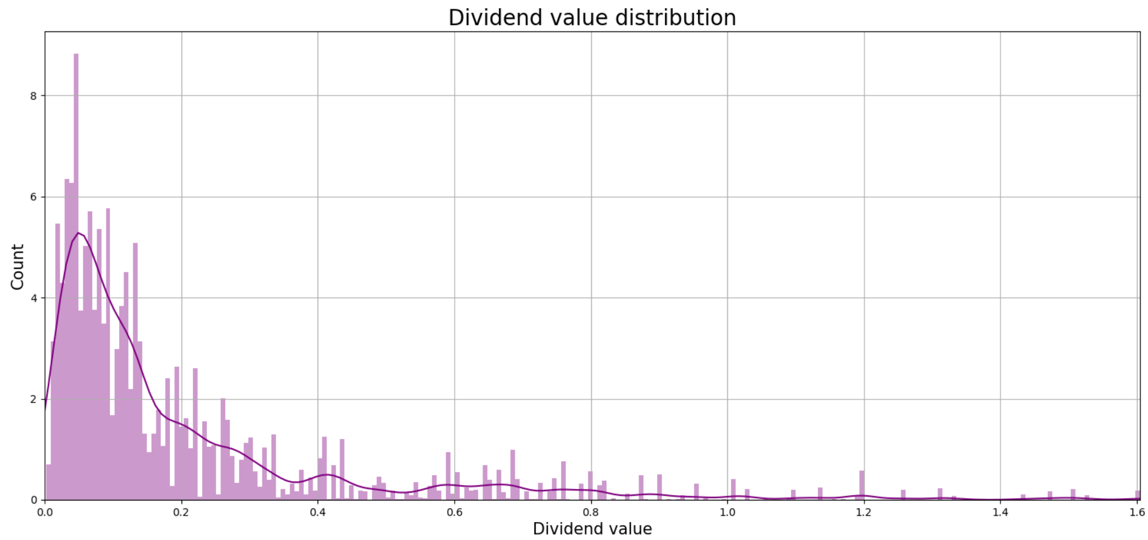

Figure 4.3: Distribution of the dividend value after logarithmic transformation

Lets observe the dividend yield distribution:

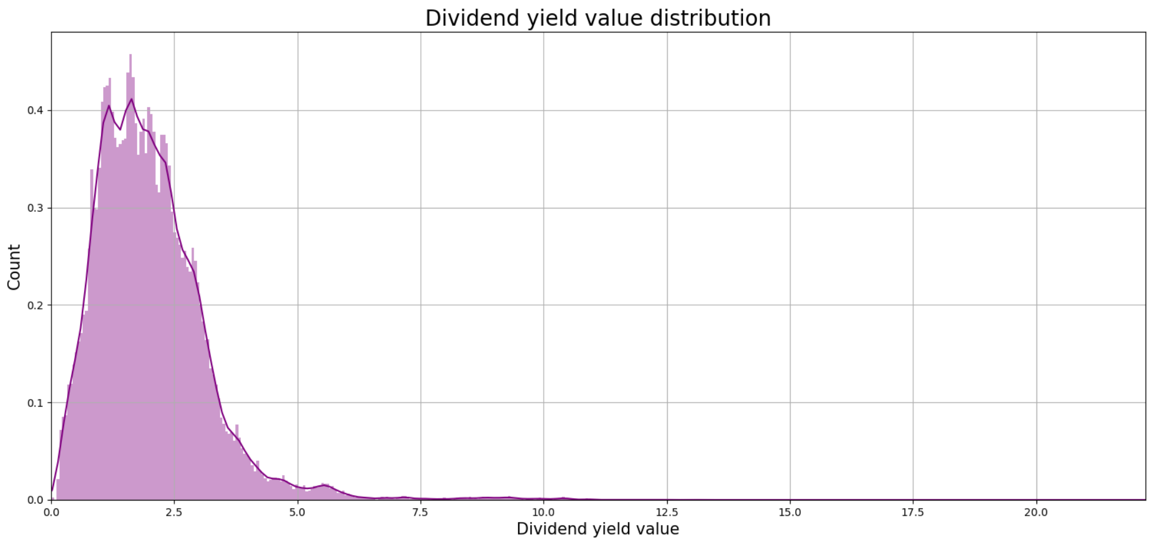

Figure 4.4: Distribution of the dividend yield value

We can see a more bell shape kernel density, but we still have a very long tail. I Figure 4.5 we compare it with other distributions, and again *lognorm* seems to be the most similar, we also have the *exponorm* being quite similar.

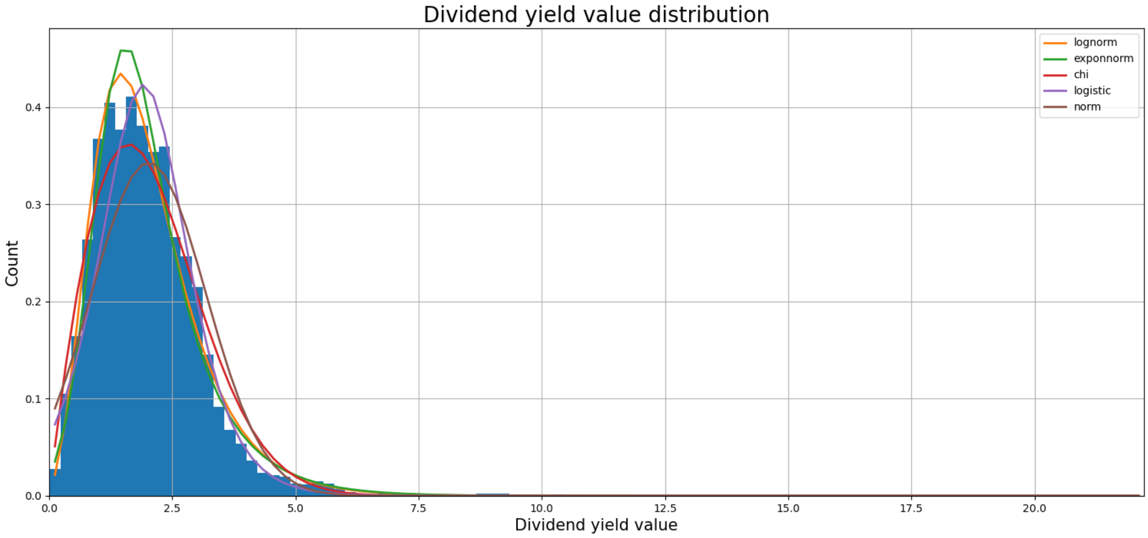

Figure 4.5: Comparison of distributions to the dividend yield distribution

Lets see how the values transform after doing a logarithmic transformation:

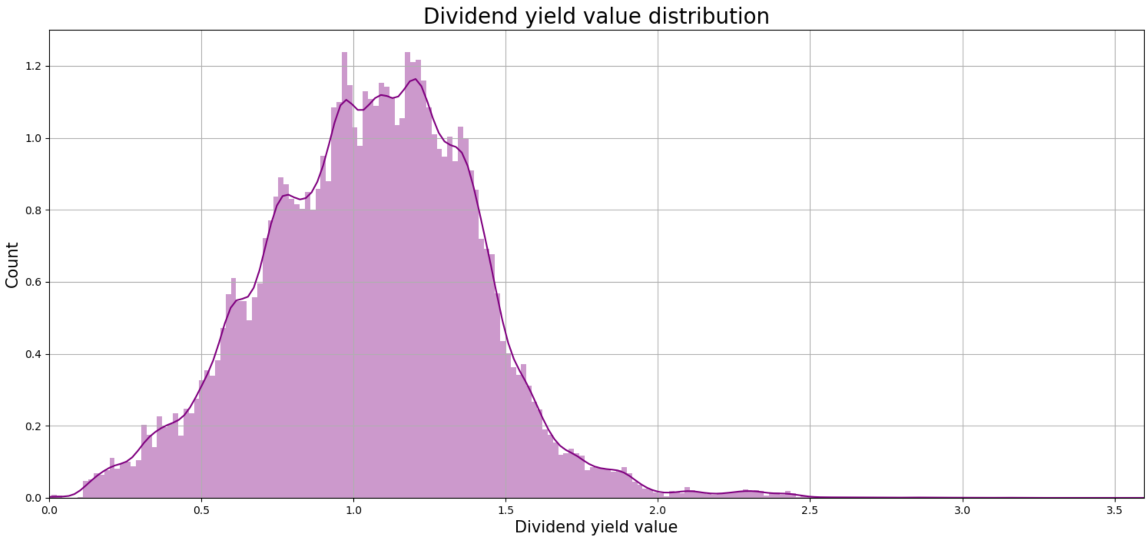

Figure 4.6: Distribution of the dividend yield value after logarithmic transformation

We have achieved a more normal distribution (although we still have some skewness), and therefore we use the logarithmic transformation of the past values of the dividend yield as a feature, and experiment with its future values as a target. We convert it back to the original scaled value by doing the reverse of the logarithm, the exponential transformation, when calculating errors in the evaluation process.

4.3 Pre-processing

4.3.1 Using the predictions done by banks as features

We have three different predictions each from the three sources we have data from. However, not all times those sources make predictions for the same dividends, so we have missing predictions from different sources throughout our dataset. We have always at least one prediction from the three therefore we decided to fill the missing predictions from others with zero, and afterwards averaging their predictions done on dividends into one single feature, the *dps_expected*:

$$\text{expected dividend} = \frac{\text{UBS's dividend} + \text{Macquarie's dividend} + \text{Consensus's dividend}}{3} \tag{4.1}$$

We replace the missing predictions by zero to avoid them having weight in the previous average done in 4.1. We now proceed to **extract their predicted dividend yield** by dividing their averaged expected dividend by the price on the day d of the prediction:

$$\text{expected dividend yield} = \frac{\text{expected dividend}}{\text{price}} \tag{4.2}$$

4.3.2 Forward-filling

Time series data rarely comes perfectly clean so, as expected, we have days without data. We decided to fill those gaps with a fill forward, where the value from the previous day is used to fill the missing value.

Our dataset consists of predictions from banks, so the existing days correspond to the days where the banks did predictions. As seen in Chapter 2 those predictions change from monthly predictions to daily predictions between 2016 and 2019 depending on each bank's strategy, meaning that, before 2016 we have a single day for each dividend being predicted. After we apply forward-filling, we expect to complete the remainder of the pricing, and dividend data in that month. Before forward-filling, we have 162155 rows and afterwards, we end with 497056 rows meaning 206.531% increase of the data.

Having forward-filling applied we now know what days were missing in the dataset, and we use those days to fetch pricing and dividend data from other datasets, so we do not end up with a lot of repetitive values. Consider the example where we have a single date record per month for each dividend being forecast in that period before 2016, when forward-filling that date records we will have every day of the month with the same price of that date, because the forward-filling picked up the last known price. So, to avoid this, we resorted to other datasets available to us with pricing and dividend data to better fill the missing values.

In Figure 4.7 we see huge drops in the number of daily predictions through time. The increase of the daily predictions comes from the increment of the number of stock added to the predictions, as we saw in Chapter 3.

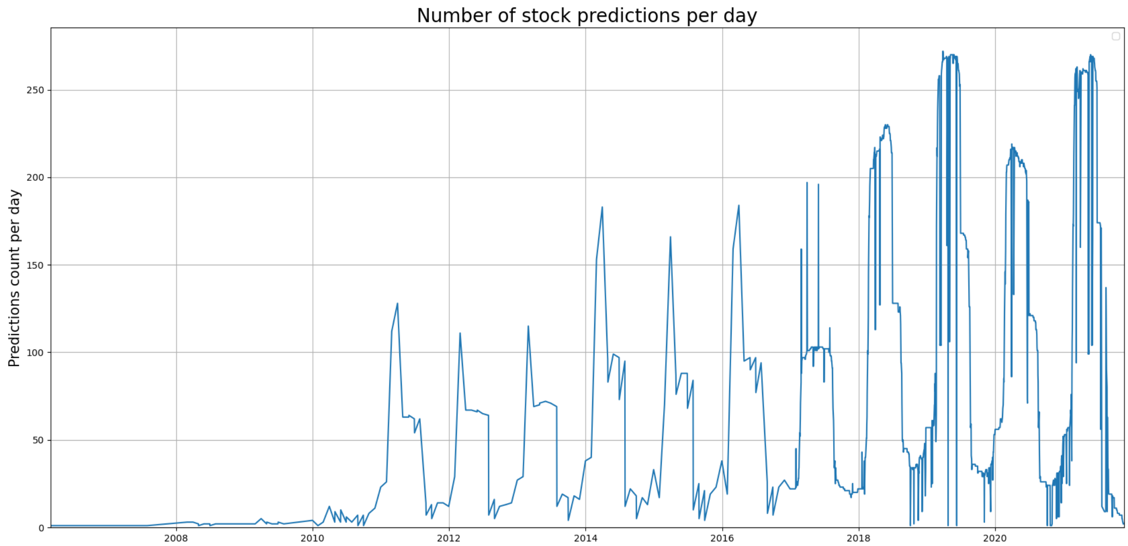

Figure 4.7: Predictions per day

We suspect these huge drops to be due to the Australian stock market being closed on weekends, and so trading data only exists on weekdays. To confirm this, we can see in Figure 4.8 the count of predictions per weekday in our data, with Consensus being the only source with trading data on Sunday (it will not be pricing or dividend data since there is none on the weekends but predictions done on Sundays).

To fill these gaps we decided to fill the days with no predictions with the last previous value, and now we have a more complete dataset in Figure 4.9.

In Figure 4.9 there is still a noticeable trend in the dataset, but since we are more interested in the distance between the prediction's run date and the dividend's announced date we moved on to remove that trend in the new time distance feature.

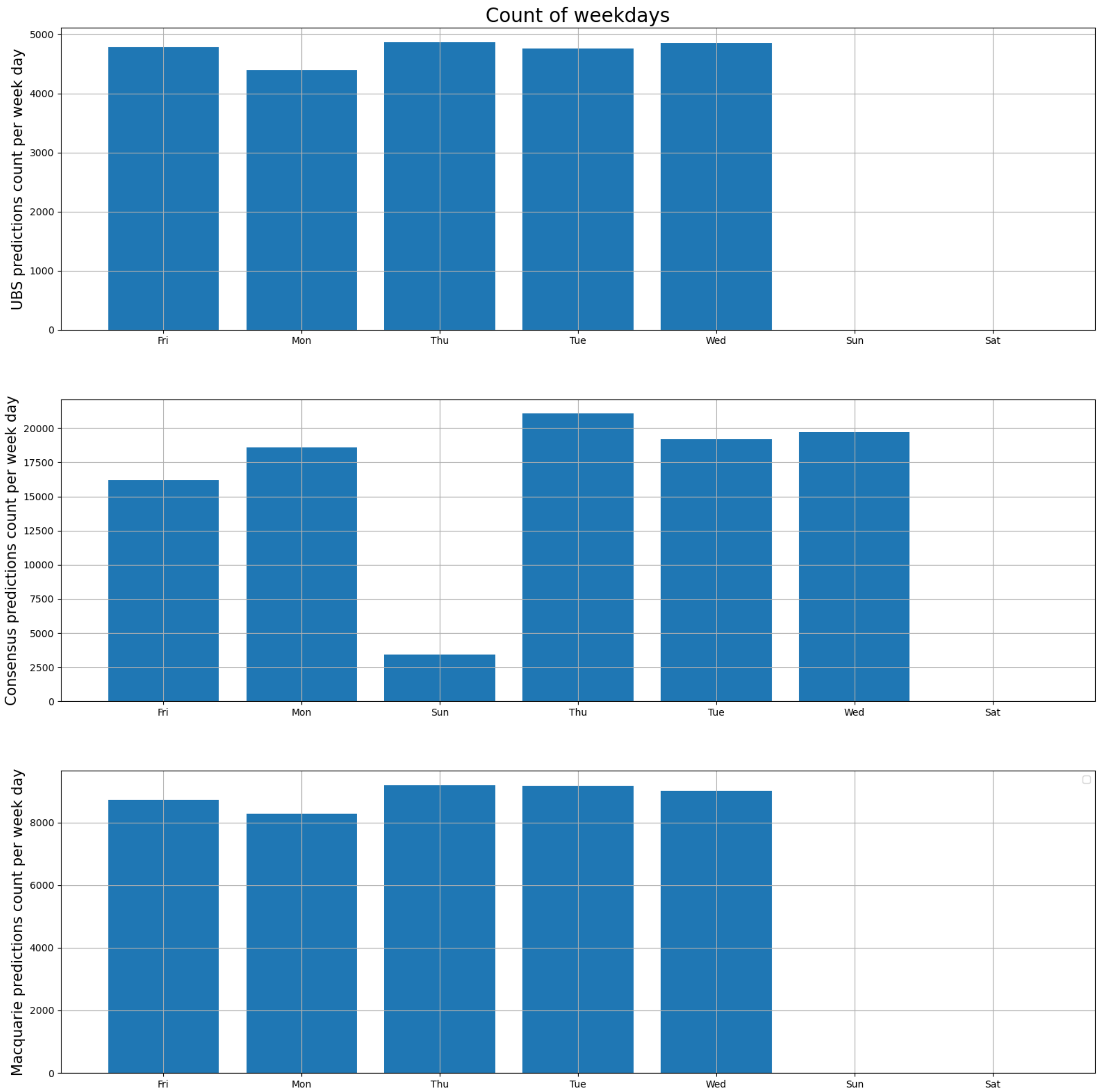

Figure 4.8: Predictions per weekday

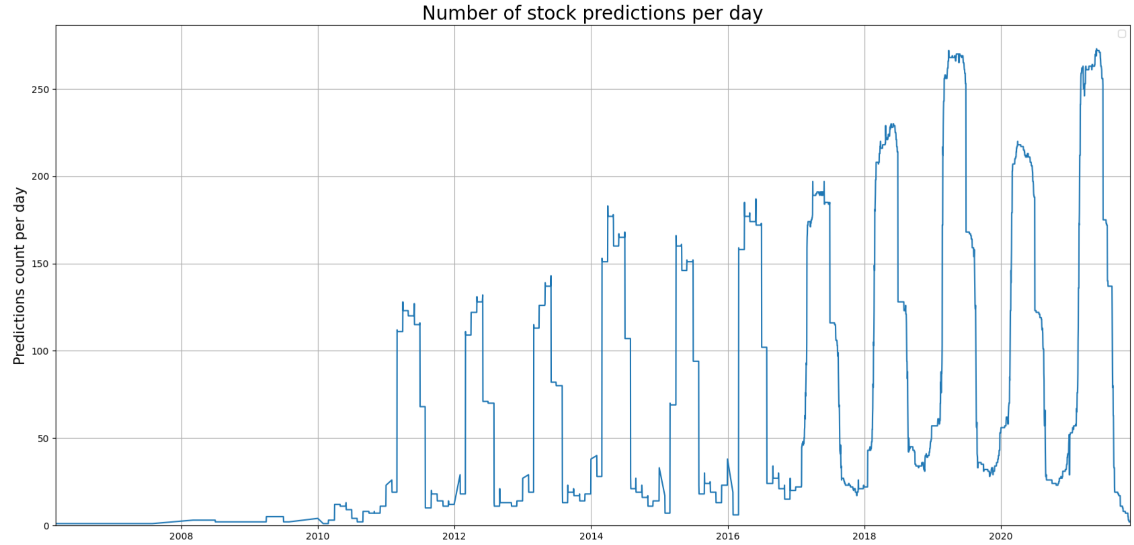

Figure 4.9: Predictions per day after forward-filling

4.3.3 Time distance

Time distance is a new feature we add which is the measurement of how many days are in between the run date of the row data, and the date we know the dividend yield value. This is a useful feature to filter data, and consequently later helpful to build our train and test for the predictive model. For example, it helps to fetch all the data that exists six months in advance to the dividend yield value since that is the target distance we are aiming for.

So having the trend of number of predictions as we get closer to dividend yield value:

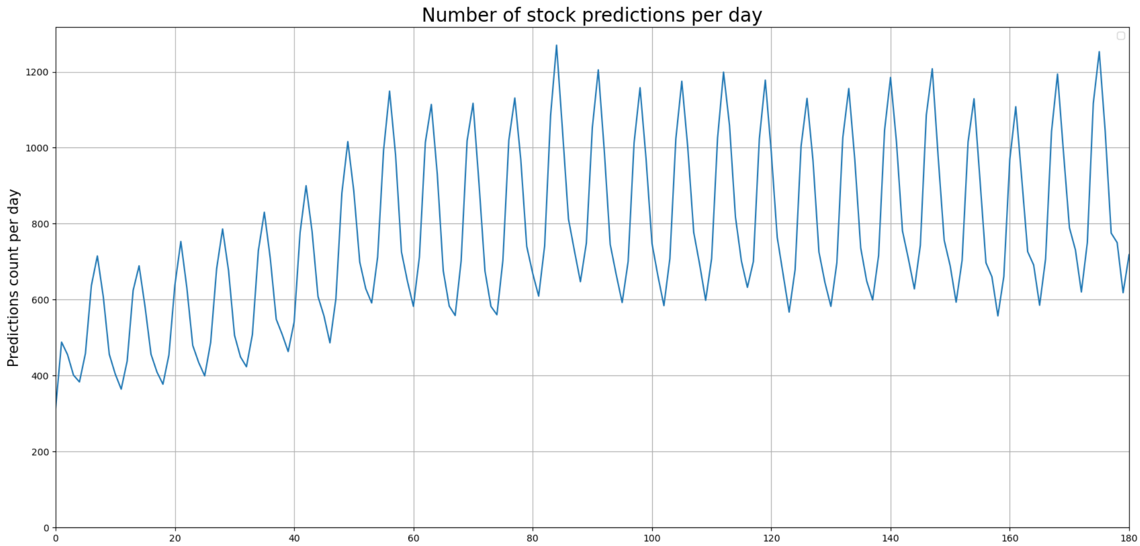

Figure 4.10: Number of predictions per time distance

We can remove this trend, therefore we do the exact same fill forward as previously mentioned and the removal of the trend is noticeable in Figure 4.11.

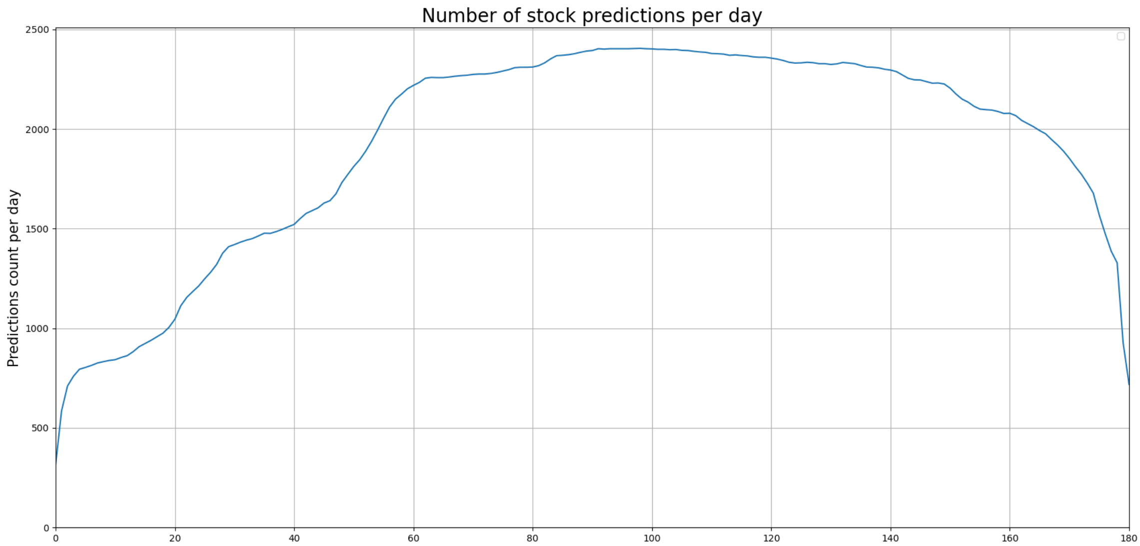

Figure 4.11: Number of predictions per time distance trend removal with forward-filling

4.3.4 Seasonality features

Due to the findings such as the one in [13] where there was a relation between dividend yields and stock returns depending on the month, we decided it would be an experiment to then add the number of the month as a feature in the model. Also as presented in Chapter 2, the cosine and sine functions are interesting functions that can help find correlations between points in time, and other variables. Thus, we add the trigonometric transformation of the month number generating two new features:

- *month_cos*, the cosine transformation of the month number
- *month_sin*, the sine transformation of the month number

4.4 Volatility features

Volatility is the rate at which a value increases or decreases over a particular period. Stock price volatility is studied and used a lot in financial strategies, e.g., usually higher stock price volatility often means higher risk, and helps an investor estimate the fluctuations that may happen in the future.

In periods of crisis, financial assets become highly volatile, we can try to signal the model about periods of high volatility by using it as a feature.

Historical volatility is based on historical prices, and represents the degree of variability in the returns of an asset. So, when there is a rise in historical volatility, the price will also move more than normal. At this time, there is an expectation that something will or has changed. If the historical volatility is dropping, on the other hand, it means any uncertainty has been eliminated, so things return to the way they were. Thus, trying different ways of using historical periods of volatility in our historical dividend, pricing and dividend yield values.

We calculated the volatility of certain values on averages of their values in 30 days, and used them in some experiments of our models obtaining poor results. After doing a Pearson correlation we found that there was no correlation between the volatility and the main financial features of the dataset: the price, dividend and, most importantly, the dividend yield of the company. In the end, we end up with the additional features:

Name	Description
div_yield_actual_volatility	Volatility of the real values of the dividend yield
div_yield_expected_volatility	Volatility of the forecast the dividend yields from the banks
log_div_yield_actual_volatility	Volatility of the logarithmic scale of the real values of the dividend yield
log_div_yield_expected_volatility	Volatility of the logarithmic scale of the forecast the dividend yields from the banks
price_volatility	Volatility of the price
prev_value_1_volatility	Volatility of the last known value of the dividend yield
prev_value_2_volatility	Volatility of the second last known value of the dividend yield
prev_value_3_volatility	Volatility of the third last known value of the dividend yield
average_volatility	Volatility of the average of the three last known values of the dividend yield

We can check the lack of correlation in Figure 4.12. Because of this, we do not use the volatility as a feature.

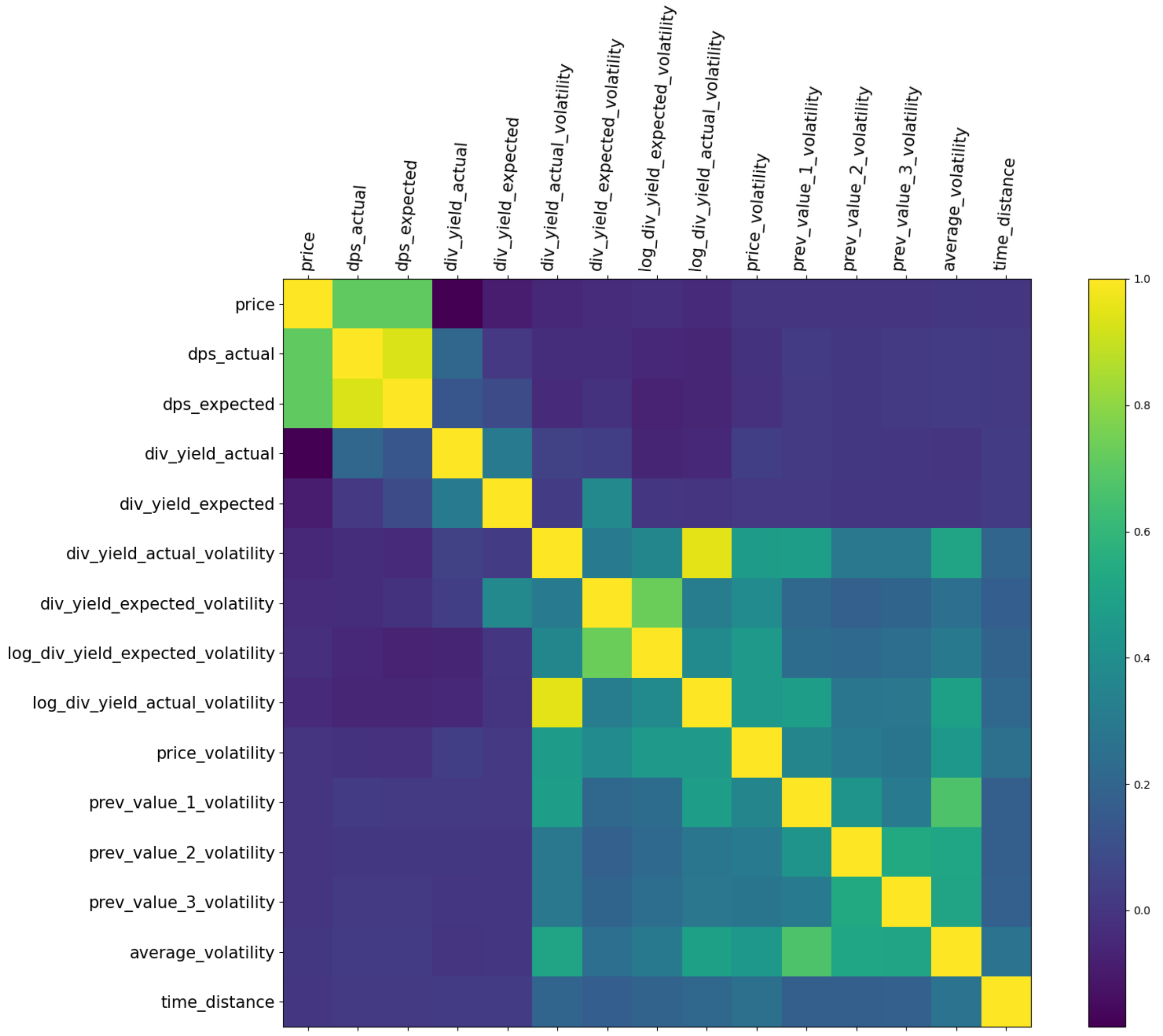

Figure 4.12: Pearson correlation of volatility features

4.5 Data cleaning

4.5.1 Cleaning the dividend yield data

We only care about the periodic dividends, so no special dividends since those are impossible to predict, therefore they were dropped. There were also duplicated and incorrect entries found.

4.5.2 Features correlation

Correlation between variables that we decided to use in the model, is tested using Pearson correlation, the initial guess that price and the actual dividend value are correlated can be verified in the test show in Figure 4.13:

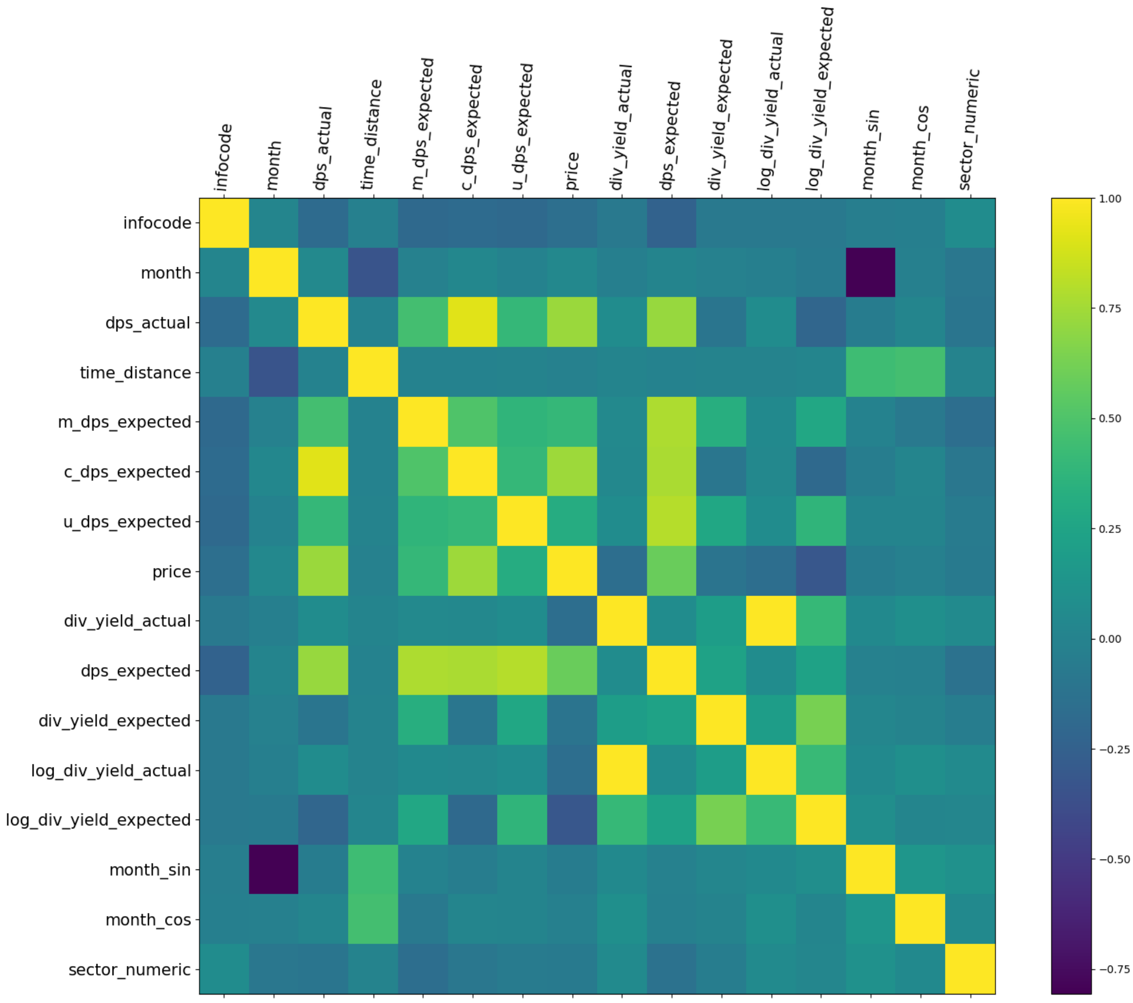

Figure 4.13: Pearson correlation

We can also check that the expected dividend value also correlates with the price, meaning that in the investing methodology used by the banks to predict the dividend they have the price of the stock in the equation.

We do not remove the features with low correlation initially because we wanted to test their effect in the model. However, this observation will later be helpful for us to understand our model results, and what features to drop later.

4.6 Time window aggregating model

To test the accuracy of the model, we predict on data whose values we already know. As previously observed in Chapter 2, there is little data before 2008 so we start predicting dividend yield values known after the start of that year. Thus, from the beginning of 2008 up to the end of 2021, we went over each month and predicted six months in advance the value of the dividend yield during that month, calculating the Mean Absolute Percentage Error (MAPE) of those predictions done within the month. The final error is the average MAPE of these dividend yields' predictions.

We first start with a training dataset that contains data until 2008, so our starting point is predicting the dividends of July 2008 meaning that our train data contains data up until January 2008. We then predict the August 2008 dividends, implicating that the training dataset has information until February 2008, and so on. Notice how the diagram in Figure 4.14 each next iteration we will accumulate one month of data to be used to predict a further six months.

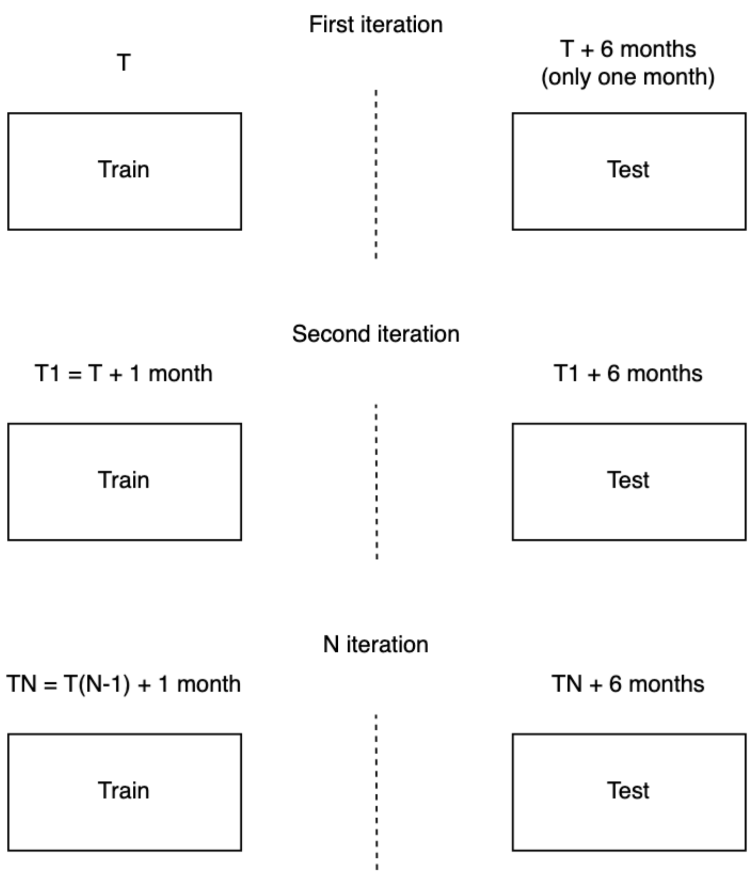

Figure 4.14: Window aggregating model diagram

Here is the example for the year 2008:

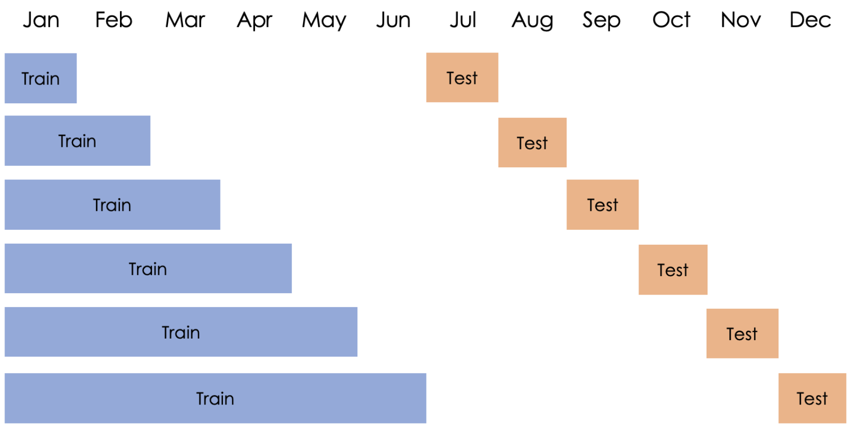

Figure 4.15: Window aggregating model for 2008

Chapter 5

Results

5.1 Learning Algorithms

Several machine learning algorithms and their respective regression versions are used to solve the problem, namely: LR, kNN, Random Forest and Bagging. We chose these algorithms because, due to it being a big dataset, it could take too long to get an initial idea of the performance. These algorithms test different approaches. The main goal of testing different approaches was to analyse which type of algorithm is more suitable for the problem in question. The software Python was used to implement the scientific experiments.

5.2 Model performance

In the following sections we will go through the performance of the model. Like we saw in the naïve experiment, the low average error of all predictions is not enough to say it is good if there are periods where we have huge disparities in the error. So, we look not only into the average error throughout the whole aggregated window, but also at the average error of all predictions in the slices of the window.

5.2.1 Summary of the results

The first attempt is to train the model to predict the dividend yield six months in advance, and see the performance without and with the bank's predictions and with, and without logarithmic transformation.

When using a logarithmic transformation on the past dividend yield, the future values (target variable) must also be transformed. When measuring the error we convert them it back to their non-logarithmic form through exponentiation, giving us back the original dividend yield.

Mean Absolute Percentage Error (MAPE) results				
Algorithm	With bank's predictions		Without bank's predictions	
	w/ log.	w/out log.	w/ log.	w/out log.
LR	540.46%	88.69%	514.10%	92.37%
kNN (k=1)	83.36%	84.71%	85.74%	88.06%
Bagging	0.63%	41.55%	0.52%	44.17%
Random Forest	0.56%	41.55%	0.51%	44.17%

5.2.1.1 Nearest Neighbors (kNN)

kNN presents an error varying between 83.36% when using logarithmic transformation with the bank's predictions, and 88.06% without the bank's predictions with no logarithmic transformation. Observing the error throughout the aggregating window, the same conclusion can be taken from the naïve experiment: although the average MAPE is lower, we have periods with disparities in the error, which is bad. Using the bank's predictions improves the average MAPE however in 2008 it adds a huge error as we can see in Figure 5.1:

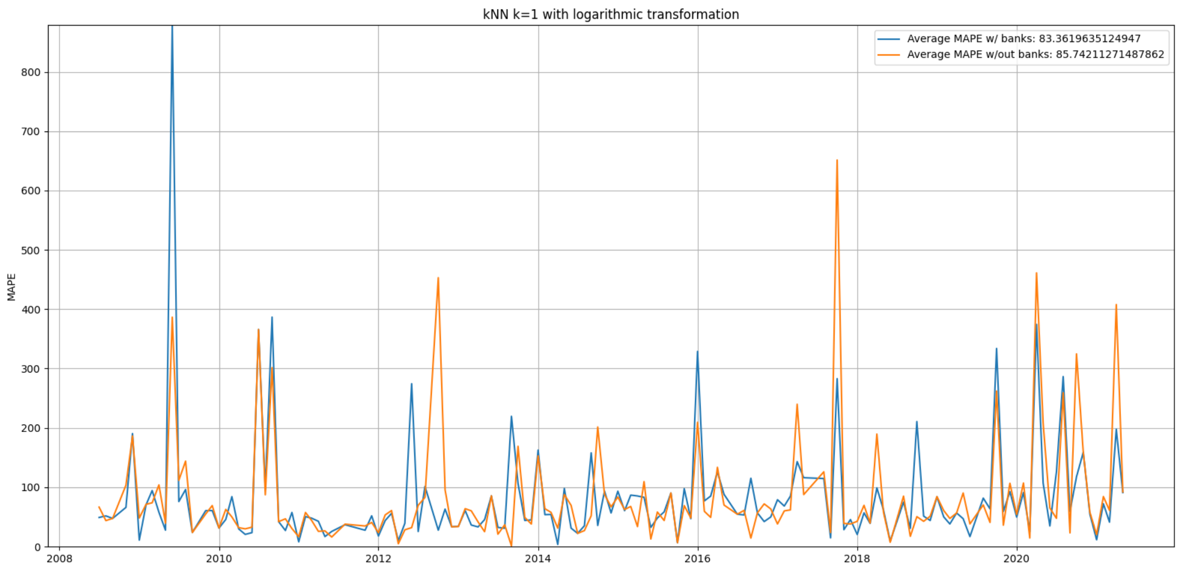

Figure 5.1: Result for kNN k=1 with logarithmic transformation

This is likely due to the changes in the performance of the banks during that time like we saw in Macquarie and Consensus:

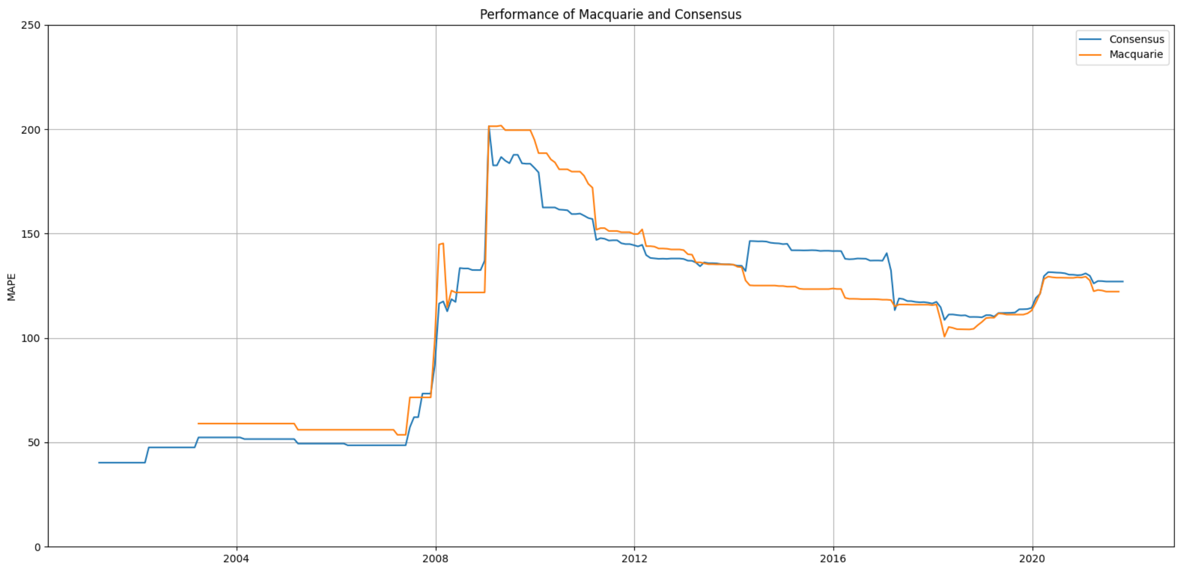

Figure 5.2: Performance of Macquarie and Consensus

5.2.1.2 Linear Regression

LR performs the worst with logarithmic transformation, probably due to either not being the correct algorithm or the huge error is brought about by a few outliers that are predicted extremely poorly, like we saw in the naïve experiment.

Whether including the bank's predictions does not affect the performance significantly throughout the aggregating window, except during 2010 as we can observe in Figure 5.3:

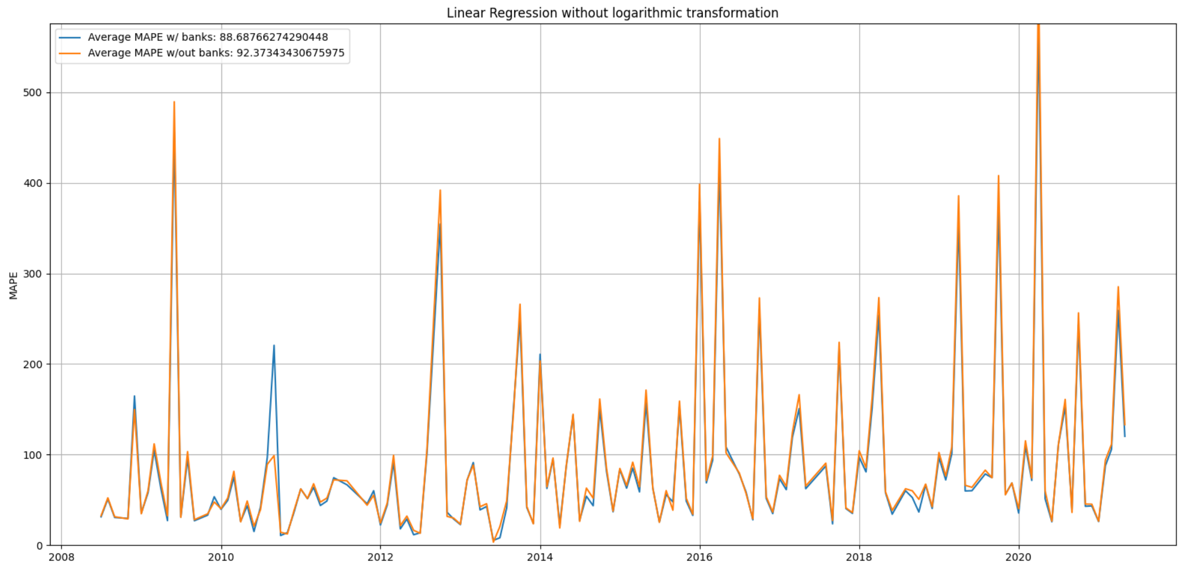

Figure 5.3: Result for Linear Regression (LR) without logarithmic transformation

5.2.1.3 Random Forest and Bagging

As expected, Random Forest and Bagging have the same performances due to their execution being based on the same logic of creating data samples. Using Random Forest and Bagging substantially lower the error when using logarithmic transformation in the preprocessing, with an error of 0.51% and 0.52% respectively as more data is aggregated, observed in Figure 5.4:

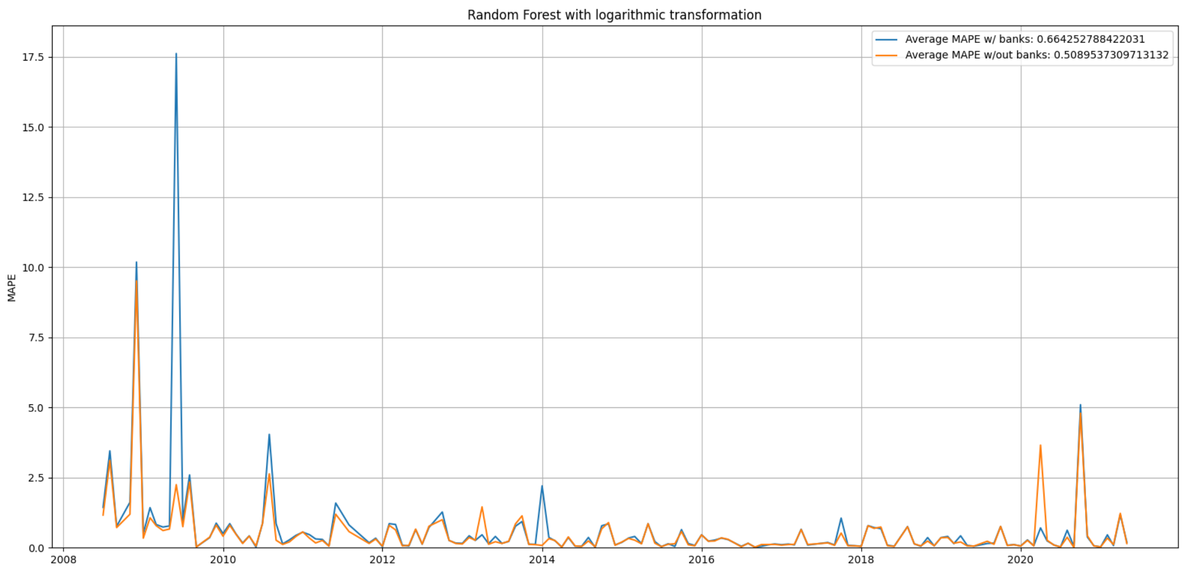

Figure 5.4: Result for Random Forest with logarithmic transformation

Without the bank's predictions, not only the average error is low, but also the error throughout the aggregating window is low when compared to the forecast dividend yield of the banks, and the naïve experiment. The highest error of 9.50% occurs on dividend yields from 2009.

In Figure 5.4 we also see the biggest error of 17.55% with the bank's predictions occurring also in 2009. Again, it might be correlated with the banks' errors present in Figure 5.2.

Without logarithmic transformation, the bank's predictions affect the model positively. They were helpful to this algorithm in their average error measurement going from 44.54% to 40.71% when adding the bank's predictions:

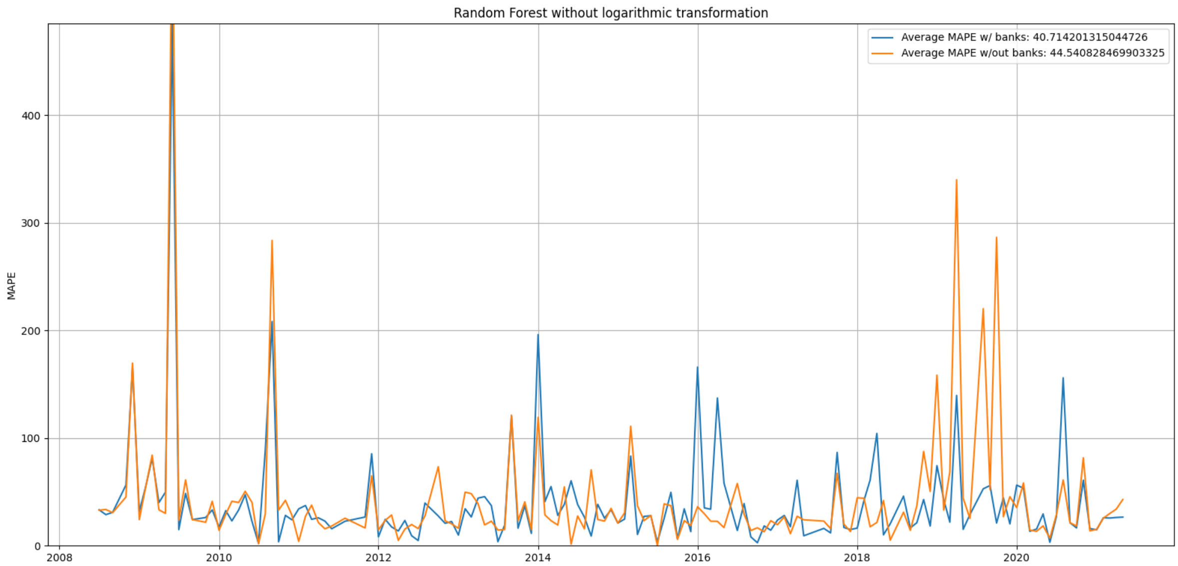

Figure 5.5: Result for Random Forest without logarithmic transformation

We can see the difference between the lines when the forecast from bank is used, and when it is not. With the period between 2018 and 2020 having the most noticeable differences.

The **feature importance during fitting** confirms this. In Figure 5.6 we have the feature importance for predicting the dividend yields using Random Forest without logarithmic transformation, and with the banks' predictions. These predictions are for November 2021, which is at the end of our window aggregated model, meaning that at this point, almost all historical data is being used. We see that the original dividend value (*dps_actual*) and the price value have a big importance in predicting the actual dividend yield value. The dividend yield from the banks (*div_yield_expected*) also has significant importance on the actual dividend yield value, so, as we mentioned before, **the banks' predictions are useful for the machine learning process**.

We also see the sector having small importance. Unfortunately, we fail to make seasonality relevant for the model since it has low importance compared to other features. This could have been improved with more time spent looking into the trends of the dataset, and trying a different type of seasonality, or even variations on the Fourier series.

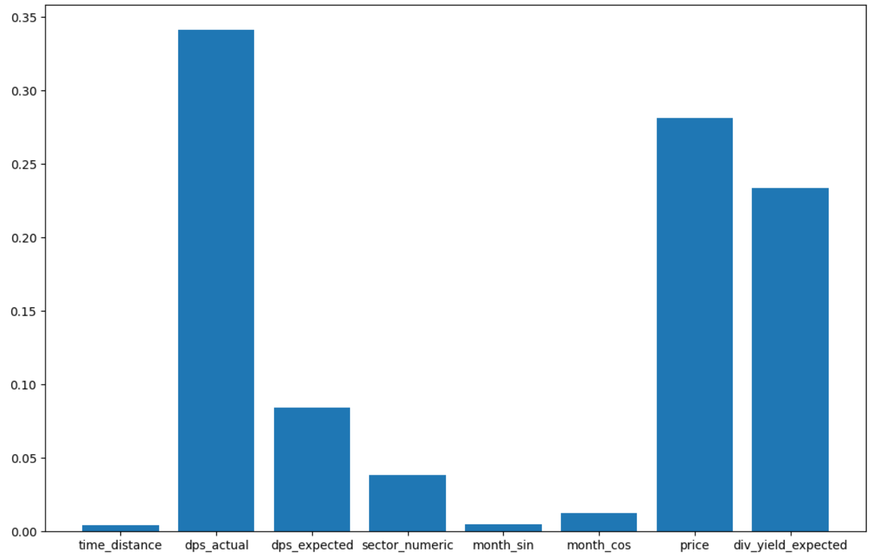

Figure 5.6: Feature importance for Random Forest during the fitting process without logarithmic transformation and with bank's predictions for November 2021

With the logarithmic transformation, the feature importance during fitting is different, the model **focuses only on the past actual dividend yields of the model**, as we see in Figure 5.7. This may be correlated to the naïve experiment observations, the model figures that the best strategy is to look at the past values of the target variable, and can correct the huge disparities that occurred in the error of the naïve experiment.

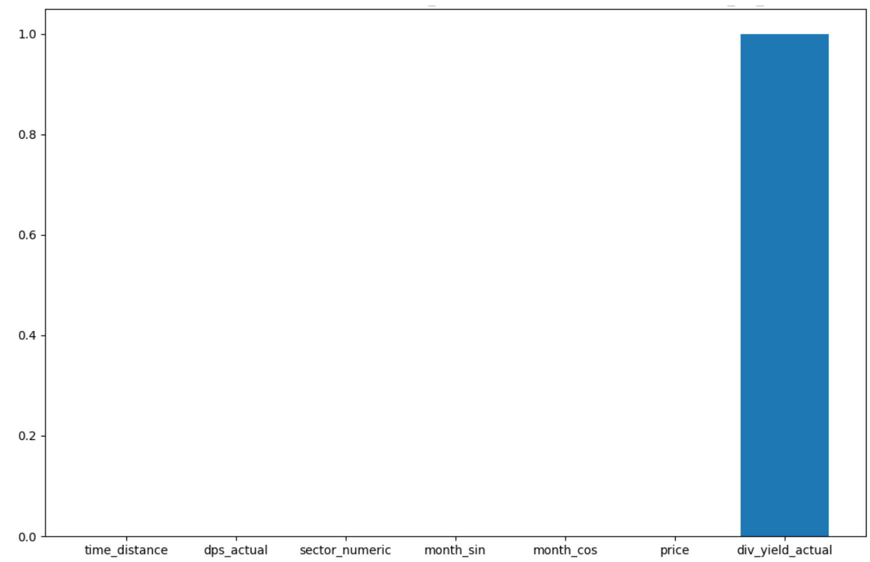

Figure 5.7: Feature importance for Random Forest during the fitting process with logarithmic transformation and without bank's predictions for November 2021

5.2.2 Discussion

Depending on the usage of normalization of the data, it can either be helpful or not. It can be the wrong pre-processing to use in certain algorithms, consequently leading to huge errors, as we saw with LR. Or, it can be spectacular at achieving better results, than not using normalization at all, like we observed in Random Forest and Bagging.

The banks can be useful or error inducers, depending on the algorithms. Adding their predictions increased the error in 2008 when using kNN, meaning that this algorithm is sensitive to the changes in the banks. However, with Random Forest and Bagging, it helped them in the volatility present in their predictions between 2018 and 2020, when having the dividend yield in its original scale.

5.2.2.1 Baseline vs our results

Banks are good at not making huge mistakes like the naïve experiment, linear regression, and kNN did. However, Random Forest and Bagging were able to not only have absence of disparities in the errors throughout the whole aggregated window, but also the final average error is very low too.

Also in the Random Forest and Bagging algorithms, we can see a significant decrease in the error through the years. This means that the **aggregation of data highly benefits these machine learning approaches**. The same does not occur in the banks, we see in Figures 3.2, 3.3 and 3.4 that their performance increased when adding more stock's data.

As previously mentioned, the banks either can add errors, or are helpful to the model, meaning that they can increase, or decrease the disparities in the MAPE of the predictions.

Chapter 6

Conclusions

In this thesis, we predict the dividend yield value, where its unitless characteristic allows comparability between different stocks averaging the weight the price could have in statistical predictions. This consequently results in less biased outcomes, especially when used as input in machine learning tools. This value not only signals a low performance of the company in terms of their stock price value, but also signals future changes in the dividend value, with a high dividend yield value implicating a future dividend cut.

Predicting dividend yields can be difficult, not only due to how unpredictable the stock market can be, but also because stock market data is very noisy. It is hugely affected by global events and human behaviour, which results in highly volatile data. Using the previous latest known value as the prediction, although it has an average error between 30.60% when using the latest value, and 46.00% with the third latest value which is lower than the smallest error from the banks of 118.89%. However, there are periods in the naïve experiment with huge errors surpassing the 700% MAPE measurement, specially in periods of higher volatility in the market, and investors do not want those high errors. It is fundamental to avoid gross mistakes, which ruin investment strategies.

Using these same predictions done by banks that resort to traditional investing methodologies on Australian companies, we extract their predicted dividend yield, and use it on our model as input. Furthermore, we use data common in every stock market (prices, dividends, and sector of stock) to have a model that could be applied in any stock market.

Additionally, further scaling of data is performed due to the existing high skewness in the distribution of the dividend yield, in an extensive process of the normalisation of this value. During the project implementation, we faced several obstacles that had to be overcome. Working with a large amount of data is very challenging, especially with the limited available time. Missing data is either fill forwarded, or fetched from other datasets available to us. There is a lot of noise present in financial data, and that is observed all over our experiments, and model results. In addition to these challenges, we still dealt with changes in the target time definition as we could not guess how far we could make good predictions so for a while the six-month period was questionable.

We demonstrate that using machine learning to predict the future value of a dividend yield is better than using traditional investing methodologies, and further naïve methodologies. We obtain an improvement not only in the average error, where the lowest was 0.51% when using Random Forest, but throughout all the dividend yields' predictions throughout the aggregated window model, with the highest error of 9.50% occurring in 2009.

It is also found that the bank's predictions can be helpful features in the machine learning model. In the model where we use Random Forest or Bagging without normalisation, and with the bank's predictions, the average MAPE is of 40.71%. This is lower than the 44.54% MAPE without the bank's predictions. We also observed in Figure 5.5 that the help was significant in the periods between 2018 and 2020.

In short, the project objectives were successfully met, the various main challenges presented were overcome, and some refinements in our model proved that machine learning techniques can make better quantitative guesses of dividend yields, than the ones done by traditional investing strategies. And also, that those investing strategies' data can be useful for the model.

6.1 Future Work

6.1.1 Deep Learning

Due to the lack of time, we did not try deep learning approaches. Furthermore, because this investigation might be a base for commercial usage it was important to understand the behaviour of our model, which deep learning can make difficult due to its black box nature. Additionally, Deep Learning favours from high Signal Noise Ratio (SNR) ratio[19] while, as we mention throughout the project, financial data has a low SNR.

However, while researching the state of the art we found a study[6] that focused on proving whether Deep Learning modelling was useful for time series forecasting. As we have concluded already, they also proved that machine learning approaches achieve better forecasting performance than traditional statistical models such as the ARIMA. They also found that overall that the regression algorithm used - Gradient Boosting Regression Tree - showed better performance than the deep neural network models except for one example on financial data from the *NASDAQ100* stock prices using the same error measurement as ours:

Model	MAPE
Autoregressive integrated moving average (ARIMA)	0.0184
DARNN	**0.0043**
GBRT	0.0257

This can show hope of achieving good results with deep learning when doing forecasting on a financial time series dataset if the SNR is high enough. Meaning a next step could experiment with this type of modelling, try to increase the dataset SNR so it becomes useful, and we can make conclusions on whether deep learning is a good methodology, or not.

Bibliography

[1] Andrew B. Ainsworth et al. 'Taxes, order imbalance and abnormal returns around the ex-dividend day'. English. In: *International Review of Finance* 18.3 (Sept. 2018), pp. 379–409. ISSN: 1369-412X. DOI: 10.1111/irfi.12155 (cit. on p. 16).

[2] J.Scott Armstrong and Fred Collopy. 'Error measures for generalizing about forecasting methods: Empirical comparisons'. In: *International Journal of Forecasting* 8.1 (1992), pp. 69–80. ISSN: 0169-2070. DOI: https://doi.org/10.1016/0169-2070(92)90008-W (cit. on p. 14).

[3] C. Mitchell Conover, Gerald R. Jensen and Marc W. Simpson. 'What Difference Do Dividends Make?' In: *Financial Analysts Journal* 72.6 (2016), pp. 28–40. DOI: 10.2469/faj.v72.n6.1. eprint: https://doi.org/10.2469/faj.v72.n6.1 (cit. on pp. 1, 6).

[4] Arnaud de Myttenaere et al. 'Mean Absolute Percentage Error for regression models'. In: *Neurocomputing* 192 (2016). Advances in artificial neural networks, machine learning and computational intelligence, pp. 38–48. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2015.12.114 (cit. on p. 14).

[5] Alexiei Dingli and Karl Fournier. 'Financial Time Series Forecasting - A Machine Learning Approach'. In: *Machine Learning and Applications: An International Journal* 4 (Sept. 2017), pp. 11–27. DOI: 10.5121/mlaij.2017.4302 (cit. on p. 15).

[6] Shereen Elsayed et al. *Do We Really Need Deep Learning Models for Time Series Forecasting?* 2021. DOI: 10.48550/ARXIV.2101.02118 (cit. on p. 52).

[7] Eugene F. Fama and Kenneth R. French. 'Dividend yields and expected stock returns'. In: *Journal of Financial Economics* 22.1 (1988), pp. 3–25. ISSN: 0304-405X. DOI: https://doi.org/10.1016/0304-405X(88)90020-7 (cit. on pp. 9–10).

[8] Vladimir Filimonov and Didier Sornette. 'Quantifying reflexivity in financial markets: Toward a prediction of flash crashes'. In: *Physical Review E* 85.5 (May 2012). DOI: 10.1103/physreve.85.056108 (cit. on p. 1).

[9] Wai Mun Fong and Zhehan Ong. 'Dividends, Profitability, and Lifecycle Investing'. In: *The Journal of Investing* 25.2 (2016), pp. 55–63. ISSN: 1068-0896. DOI: 10.3905/joi.2016.25.2.055. eprint: https://joi.pm-research.com/content/25/2/55.full.pdf (cit. on p. 1).

[10] WILLIAM N. GOETZMANN and PHILIPPE JORION. 'Testing the Predictive Power of Dividend Yields'. In: *The Journal of Finance* 48.2 (1993), pp. 663–679. DOI: https://doi.org/10.1111/j.1540-6261.1993.tb04732.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1993.tb04732.x (cit. on p. 9).

[11] Ronen Israel, Bryan Kelly and Tobias Moskowitz. 'Can Machines 'Learn' Finance?' In: *SSRN Electronic Journal* (Jan. 2020). DOI: 10.2139/ssrn.3624052 (cit. on pp. 1, 14).

[12] Avner Kalay. 'The Ex-Dividend Day Behavior of Stock Prices: A Re-Examination of the Clientele Effect'. In: *The Journal of Finance* 37.4 (1982), pp. 1059–1070. ISSN: 00221082, 15406261 (cit. on p. 16).

[13] Donald B. Keim. 'Dividend yields and stock returns: Implications of abnormal January returns'. In: *Journal of Financial Economics* 14.3 (1985), pp. 473–489. ISSN: 0304-405X. DOI: https://doi.org/10.1016/0304-405X(85)90009-1 (cit. on pp. 11, 36).

[14] Jinhwa Kim, Chaehwan Won and Jae Kwon Bae. 'A knowledge integration model for the prediction of corporate dividends'. In: *Expert Systems with Applications* 37.2 (2010), pp. 1344–1350. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2009.06.035 (cit. on p. 16).

[15] Terry A. Marsh and Robert C. Merton. 'Dividend Behavior for the Aggregate Stock Market'. In: *The Journal of Business* 60.1 (1987), pp. 1–40. ISSN: 00219398, 15375374 (cit. on p. 16).

[16] Xin-Yao Qian and Shan Gao. 'Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods'. In: (June 2017) (cit. on p. 15).

[17] Suman Saha, Junbin Gao and Richard Gerlach. 'Stock Movement Prediction on Ex-Dividend Day Using Event Specific Features and Machine Learning Techniques'. In: *2021 International Joint Conference on Neural Networks (IJCNN)*. 2021, pp. 1–10. DOI: 10.1109/IJCNN52387.2021.9533799 (cit. on p. 16).

[18] Sue Visscher and Greg Filbeck. 'Dividend-Yield Strategies in the Canadian Stock Market'. In: *Financial Analysts Journal* 59.1 (2003), pp. 99–106. ISSN: 0015198X (cit. on pp. 1, 9).

[19] Tongtong Yuan et al. *Signal-to-Noise Ratio: A Robust Distance Metric for Deep Metric Learning*. 2019. DOI: 10.48550/ARXIV.1904.02616 (cit. on p. 52).

[20] Yudong Zhang and Lenan Wu. 'Stock market prediction of SP 500 via combination of improved BCO approach and BP neural network'. In: *Expert Systems with Applications* 36.5 (2009), pp. 8849–8854. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2008.11.028 (cit. on p. 16).