Financial Factors Analysis for Acquisition Premium and Anticipation using Extreme Gradient Boosting and Deep Recurrent Neural Network

by

Mohammad Rayhan
16101117
Samiha Sultana
16301076
Annur Majid
16101038

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
December 2019

© 2019. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Mohammad Rayhan
16101117

Samiha Sultana
16301076

Annur Majid
16101038

Approval

The thesis/project titled "Financial Factors Analysis for Acquisition Premium and Anticipation using Extreme Gradient Boosting and Deep Recurrent Neural Network " submitted by

1. Mohammad Rayhan (16101117)
2. Annur Majid (16101038)
3. Samiha Sultana (16301076)

Of Fall, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on December 24, 2019.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Mahbubul Alam Majumdar, PhD
Professor
Department of Computer Science and Engineering
Brac University

Abstract

This study shows importance hierarchy of financial factors of corporations' Goodwill and tries to foresee with popular machine learning and deep learning models. Financial engineering is using mathematical model to study financial behavior. Financial engineers are hired by investment banks, commercial banks, hedge funds, insurance companies, corporate treasuries, and regulatory agencies. It is vital for each of them to asses a company's sustainability before any sort of investment. However, predicting sustainability is not deterministic. Therefore, corporate sustainability has become a mainstream business goal for stakeholders. Whether Quantitative finance impacts goodwill or has implicit insight can be a machine learning problem. Deep learning and machine learning are rapidly changing the financial services industry. Business leaders can now transform vast amounts of financial data into insightful predictions with the help of data science, creating significant savings in the bottom line. This thesis is concerned with investigating financial factors of a company's Goodwill and also fits popular machine learning and deep learning models and evaluate goodness of fit. To aid the research, a comparison between the proposed models-XGboost and Deep LSTM are conducted.

Keywords: Goodwill; Machine Learning; RNN; Xgboost; feature importance; Prediction ; deep learning

Dedication

We would like to dedicate this thesis to our loving parents and teachers.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.

Secondly, to our supervisor Md. Golam Rabiul Alam Sir for his kind support and advice in our work. He helped us whenever we needed help.

We would also like to thank our family for the mental support that they provided us throughout the journey.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

In the era of information revolution, access to financial information about corporations is publicly available. Investors can therefore have the information mostly in structured format and make use of the information on spreadsheets. Moreover, the information can be used to do financial factor analysis and predict what to anticipate from the company in the future.

With the invention of artificial intelligence, people have started desiring to be in the right track more and more. AI proved suitable to solve complex, fuzzy problem starting from playing chess to give us entertainment to saving our future by predicting using neural networks to save our tomorrow [36]. Financial analysis and forecasting is important for any running company so that they can know if they are in the right track. It is also important for the investors [35]. For example, people taking part in business wants to know why a company like Motorola had profit at some times and loss in another time period. Goodwill or acquisition premium is an intangible asset which represents the reputation of the business [30]. So, there should be a way to know which features are more relevant to the goodwill so that every company can aim that and gain profit. Feature extraction is needed to find out which features are the most important and relevant to make a forecast. Pre-processing the data should be focused to improve prediction accuracy [23]. The most important advantage is that the analysis allows the company itself to analyze its own performance and take steps to avoid the shutdown of the company [29]. A company with no clear direction and knowledge that a negative consequence is on the way can lead the life of many workers in disastrous situations due to sudden lay off from company.

29 Winning solutions reported in 2015 on Kaggle's blog, 17 solutions used XGBoost. Eight of these approaches used XGBoost solely to train the algorithm, while most others incorporated XGBoost in ensembles with neural networks. The second most popular method, deep neural networks, has been used in 11 solutions for comparison. The system's success was also experienced in the 2015 KDD Cup, where all winning teams used XGBoost in the top 10. In addition, the winning teams stated that only a small amount of ensemble methods outperformed a well-configured XGBoost. Therefore, the motivation for our work came from the frequent fluctuation of good and bad times of companies and thus we took a initiative to solve the problem in

our research using XGboost and deep neural network and compare the solutions by each model.

1.2 Problem Statement

Takeover or acquisition premium is the higher cost of acquiring a target company during a merger and acquisition. On the balance sheet of the acquirer, the acquisition premium is reported as goodwill. The importance of the brand name of a company, a strong customer base, good customer relationships, good employee relationships and any patents or proprietary technology obtained from the target company are taken into account in goodwill [20]. The main part of the intangible assets is the existing cumulative intangible assets. It plays the key role in enhancing competitiveness at its core. The academics who dedicate themselves to intangible assets divide it into various components. Stewart(1997) is of the opinion that intangible assets consist of human properties, organized assets and consumer assets [14]; Sveiby(1997) divides it into human properties, organized internal assets and structured external assets [3]; Anne Brooking(1998) breaks it into business assets, intellectual property assets, talented individual assets and asset structure. Xu Xiaojun (2004) proposed that intangible assets should include the value of the strategic elements that grow for the company [17]. Cumulative assets are divided into five components: human assets, strategic assets, technological assets, structured assets and business assets[2]. The parts are explained below :

1) **Human assets:** It reflects the potential of different levels of knowledge, skills, ability, experience and skills etc. of the staff body. It is valuable interest in capitalizing human resources. It includes the level of educational level of staff, skills (especially the level of skilled mechanics), operational ability, ability to renew knowledge and ability to innovate, etc.

2) **Strategic assets:** This represents the capacity of the business executive to consider, take, program and execute the potential of long-term business growth. This requires the strategic understanding of the business executive, ability to predict, market uppity, ability to command, ability to make decisions, organizational capacity, etc.

3) **Technological assets:** This refers to the corporate operation's intellectual property assets and application capability, production capacity, and integration potential of each technology. It includes the level of implementation of enterprise information technology, RD expenditure level, technological ability to absorb, complexity and protection of production equipment, as well as the patent of enterprise and each category of exclusive design rights etc.

4) **Structured assets:** It includes organizational structure, management system, management style, management system, development plan, corporate culture, the process that can make the business run smoothly. This involves the quality of the organizational structure of the enterprise, the design of the business community, the way of risk assessment, the management norm of the enterprise, the management process system of every type of company, the management system, etc.

5) **Business assets:** This applies to the company's market-related intangible assets.This includes the reputation or identity of the company, degree of customer satisfaction, customer loyalty, post-sale service program, each type of brand, network marketing, relationships between organization and manufacturer, customer database, ability to process market information and coordination, etc.

Actual valuation of immaterial properties often overlooked by investors or management which creates inconsistency. It often leads small company to sell their business for a lower price than actual price. Sometimes even small company tries to overvaluate their business so that the selling price can be increased. This problem can be resolved with the proper valuation of the intangible assets. Thus this paper intends to find the factors which are accounted with it.

1.3 Research Challenges

The tendency of taking over relatively smaller companies by giant company is increasing rapidly. The giants often go for the rising startups or fairly new established companies, institutes, firms and companies. They can expand their business in current horizon or they can go for the vertical business. In spite of that the giant companies merge with others to form merger or a conglomerate.

An acquisition can be two types which are portioned acquiring assets and full acquisition. Securing more than 50 percent of the targeted company stock as well as other assets is known as acquisition. The former owners, stockholders, bond holders may still own some portion of the company. Although the securing company can control the targeted company without any approval from its current stakeholders. Sometimes acquisition and takeover can use for different situations. Usually acquisition is when the board of directors of the targeted company agree with the proposal of acquisition. However, sometimes the parental company can force to buy a targeted company's stock and continues to pressure. Eventually the parental company takeover the targeted company which is also known as take over.

A company acquisition can happen for a number of reasons. The parent company sometimes are not interested in competition which is a key factor of acquisition. The other key factors are including higher market share, obtaining new technologies, expanding business, diversification of its current market in both horizontally and vertically, increasing the synergy between its products. There can be other reasons as well. Growth strategy, cost reduction and entering to a foreign marketplace are also can be strategy for mergers.

For acquiring a company, the parental company has to offer generous amount of money and other benefits so that the board of directors of the targeted company can accept it. The question comes that how much money should the targeted company may accept. There is no solid solution to this problem. The reason is assets of a company can be classified in divergent categories. Current assets, Non-current assets, physical assets. However, the assets can be both tangible and intangible. Tangible assets can be measured in terms of money. The problem arises when the

intangible assets comes in.

Intangible assets contain goodwill, patents, trademarks, copyright, brand recognition etc. However, goodwill represents the non-monetary, nonphysical value of the company. It is similar to other intangible assets. Complications comes when other intangible assets can be sold such as patents, trademarks, copyrights but not the goodwill. Although goodwill includes the other intangible assets in it. As goodwill cannot be sold, transferred or purchased separately, it has indefinite useful life which is very unlikely to found on the other assets. According to Financial Accounting Standards Board (FASB), goodwill of a company cannot be amortized from 2001. However, recently from 2014 they only allow it for the private entities.

1.4 Research Objectives

In this paper, we will use different techniques of data science such as deep neural network and ensemble random forest decision tree to examine financial factors that impact goodwill of a company and the accuracy of the learning algorithms in terms of forecasting. During the process of anticipating goodwill, this study extracts impacting features from nearly 144 features which are publicly available. 40 quarters amount of information for a company is used in the models to train and optimize the models. Eventually sample tests have been conducted to compare the accuracy.Major objectives of the study are as follows :

1) Finding Important features related to specific organization's acquisition premium.
2) Anticipate acquisition premium in future quarters using XGboost and RNN based LSTM
3) evaluation of the trained model

1.5 Organization of Research Paper

The remainder of the research paper is as follows :

Chapter 2 : This chapter discusses previous research involving acquisition premium or goodwill, machine learning methods and deep learning schemes.

Chapter 3 : This chapter explains the preparation, generation of the indicators, data set and machine learning used in the model and also explains our model.

Chapter 4 : This chapter details the accuracy and performance of the models.

Chapter 5 : This chapter concludes the findings and proposes future research

Chapter 2

Related Work

In this section, a detailed overview of previous studies of similar works will be discussed. Mostly about intangible assets, its importance as well as machine learning related works.

Intangible assets are one of the most important determinants of a company valuation. Accurate prediction of valuation of a company is very important for the investors. Prediction, without precision, may often lead to over valuation or under valuation which is noxious for both the company and the investors. The investors may acquire a company with more money than the actual price. It can be similarly disadvantageous for the company as well. The reason is the company may often sell their share for less price which actually undervalues the company. There are number of studies and analysis done by researchers. A number of valuation technique has been discovered by the researchers. In this research, neural networks and decision tree base algorithms are used. For some instances, where the neural network became complicated, may provide off target results [7].

Many models are being used to valuation of the intangible assets. Each of them has their own strength and weaknesses. Regrettably, global consensus and industry standards are lacking. Improving current solutions may lead to wider acceptance of existing models [19].

In common practice, there are eight types of models to evaluate and valuation of the intangible assets. These are: premium pricing model, the income-based model, market-based model, cost savings model, cost-based model, royalty savings model, option model and excess operating profits model [16]. Income based valuation model uses the prediction of upcoming revenues to build an estimate of the asset value [2]. This model provides high accuracy when it is feed with accurate information. Matsuura said this type of information is precise when the asset is in well defined marketplace. If the market is speculative, this model is likely to be ineffective. In addition to that, this model does not represent historical data [11].

The cost-based model measures an intangible asset by calculating the actual cost of replacing the asset. The cost-based model's assumption is that an investor would not spend more to purchase the asset than it would pay to replicate the asset [8]. The cost-based model typically does not discuss potential future advantages that can

be extracted from the resource (e.g., income from licensing). A cost-based model usually looks backwards and often requires some form of change to depreciate the asset over time [11]. In general, different companies will choose to have different costs in their design. This is why Ryan, Lai and FU said cost-based models vary widely from industry to industry and from business to business [19].

Mr. Pitkethly said Market-based valuation models by looking at the marketplace estimate the value of intangible assets [2]. Assets equivalent to those concerned are listed and, as an estimation of the value of the new assets, the licensing revenue currently generated from those comparable assets on the market is used. When it is easy to identify comparable intangible assets, market-based valuation models are relatively easy to use and can generate accurate predictions [19]. A major problem with market-based appraisal models is that it can be difficult to define a comparable intangible asset. Market-based models therefore perform well if the asset has a well-established marketplace and if the asset does not have a clearly defined marketplace, they are ineffective.

Benintendi said the excess operating profit calculation determines the value of the intangible assets by capitalizing the additional profits created by the company owning the assets above those generated by similar companies that do not benefit from the intangible assets. When using this approach, it is important to make sure that the defined excess profits are clearly due to the intangible asset in question but not due to any other variable, such as an efficient production plant or a distribution network which applies to the entire business[19]. A downside of this model is that the company evaluating the subject's profits or return on assets is likely to have some of its own intangible assets. This increases its own margins and return on assets (ROA). It can have effective manufacturing facilities which can also affect [10].

The premium pricing approach is a variation on the model of excess profit and is often used to price products in the consumer product industry where it is common for a branded product to be more costly than a non-branded counterpart. Chung, Lai and Fu said the price of this additional revenue estimated over the life of the product, net of marketing costs and other consumer support costs incurred to generate this revenue, and discounted to date, offers a brand value. Woodward said a downside of this approach is that finding a completely unbranded product these days is very difficult. In the food industry, where stores often offer branded and "personal label" goods, the store's own brand has significance [8].

The cost savings method measures the asset by measuring the cost savings present value that the company hopes to gain as a result of asset owning. This is typically the outcome of a successful operational system. Anderson said that while a company usually can calculate the costs it has saved since the new process was introduced, it can be difficult to determine if a third party would save more or less costs if it brought the same idea into its own business. [28].

Consulting firms such as Brand Finance (2000), Whitwell (2013), Intangible Business and AUS Consultants or Consor (Salinas, 2007) have applied the Royalty Savings Model in different ways [26].

The alternative is a preference that can be exercised at a given time, but it need not be exercised. Intangible asset owners have a variety of choices regarding their property being created and commercially used. Option models are most successful when it is possible to easily define and compare the different options. Models are more effective when constant selection values are not subject to abrupt changes in value. Choice models also work more effectively when the requirements are set and can not be exercised before they mature [19].

According to Sumi (2008), key value driver of corporate management in upcoming decades will be intangible assets such as goodwill, brand value, intellectual assets, competitiveness of products. The price of stock market not only depends on the balance sheet and income statements it is also dependent on the intangible assets. Although the data and information of these variables are not stored properly by the small companies. The integration and managing the internal figure may be poor [16].

Neural network based studies show the forecasting results are much more attractive nowadays. The accuracy is becoming higher day by day through optimizing the parameters. Newer approaches also coming into the scene to create impact. Phua, Ming and Lin (2001) defined a model where they successfully gained the accuracy of nearly 80 percent [9]. Chatterjee et al. (2000) provided a study of financial marketing with artificial neural network. Jia (2016) examined LSTM's effectiveness in predicting the stock market. He found the data on returns to be overfitting resistant and usually works better with more layers and more cells. He found that LSTM was effective at finding stock market patterns and financial data. Extensive research has been carried out on forecasting the movement of the capital market, incorporating profound learning techniques [25]. Adebiyi et al (2014) worked with Dell Incorporated's financial data. The data was analyzed by both ARIMA and Artificial Neural Network (ANN) for forecasting results. Through their study they find both worked similarly [21]. Kim specified a one-layer Neural network architecture which uses pre-trained word vectors as inputs that can be treated as vectors unique to the task or static. Word vectors are viewed as fixed inputs in that method, while they are' tuned' for a specific task in the latter. They were given better results by learning task-specific vectors through fine tuning [22].

Chapter 3

Methodology

In this chapter, dataset used and procedures that are related to this work have been discussed. Overall underlying concept of the algorithms and the related concepts are reviewed and discussed.

3.1 Dataset Description

As mentioned, this paper used financial metrics as parameters which then fed into XGBoost and LSTM RNN.Financial data comes in many shapes and forms. From a high level view, this study incorporated typical fundamental data and market related data. Two types of data used for this study are - Fundamental financial statements and market related data. several ratios are derived from these two high level publicly available data. Fundamental data is mostly accounting data, reported quarterly. A particular aspect of this data is that it is reported with a lapse.In order to maximize the lapse, quarterly recorded data are considered. Yearly collected data decreases the frequency of the inputs and daily recorded data are hardly available as well as some parameters, especially ratios will not add much information to the experiment. All sorts of trading activities of a company in the stock market is recorded as market data. They are used as parameters for the target variable- goodwill.

$$(fundamental + ratio + stock \rightarrow Goodwill) \tag{3.1}$$

Appendix A has 144 columns organised in specific order with each instances in quarterly date index format. column 39 is the target variable.From 1-86 column are fundamental data and onward are Market and ratio data.

3.2 Data Collection

Financial data used in this study are posterior to 2008 financial crisis. In this study, financial data of 7 major leading industries in the world are collected. Information are publicly available by accounting section of the company or through data vendors. For the research purpose, those company who has more than 30 percent of missing records were not taken into consideration. There are several companies who have a flat acquisition premium for a long period for example: Facebook inc. Those constant goodwill corporations were also not picked as it could mislead the result.

A model could be arbitrarily worse that will only generate the expected value of the goodwill and still would perform better since the target variable has no fluctuation. Only leading companies of specific industries were chosen that have minimum 38 quarters of records available. Every company's data is stored in previously mentioned data frame in chapter 3.1. Next, these data are indexed with date column. Thus, the statement of profit loss, balance sheet and cash flow statement, market metrics, 40 financial quarter growth ratios are prepared for each company. Nearly 40 company's data were used for the experiment. Here is a subset of the companies - assigned unique numbers as ID to store results of the research more conveniently.

3.3 Feature Selection

Each record of a company has 40 rows and 144 columns. The description of each collected data set is as follows :

```
In [5]: df.describe()
        df.info()
        df.fillna(0,inplace=True)

        <class 'pandas.core.frame.DataFrame'>
        DatetimeIndex: 40 entries, 2009-12-26 to 2019-09-28
        Columns: 144 entries, Revenue to SG&A Expenses Growth
        dtypes: float64(72), int64(72)
        memory usage: 45.3 KB
```

Figure 3.1: Data-set Description

Among which, 'goodwill' column is set as target 'Y'. A plotted graph of the target variable of company index: 08 is shown –
By looking at the goodwill plot of each company, we could not interpret any seasonality or trend otherwise the research could have concluded. Therefore, the research goal is then set to find the factors that is impacting for Y. There are several hypothetical architecture were established to map the predictor and target. Arrangement of independent and dependent variables will be structured as follows:-

Multivariate single step
Uni-variate single step
Multivariate rolling window
Uni-variate rolling window

i. Multivariate single step :
Since the problem is converted into supervised learning, the aim is to find if current quarter's financial metrics can give insight of the next quarter's acquisition premium. So far, only target variable was constructed. We are interested to make our model learn the goodwill of next quarter. Graphically –
Therefore, in multivariate single step architecture, Y is shifted one step and at the

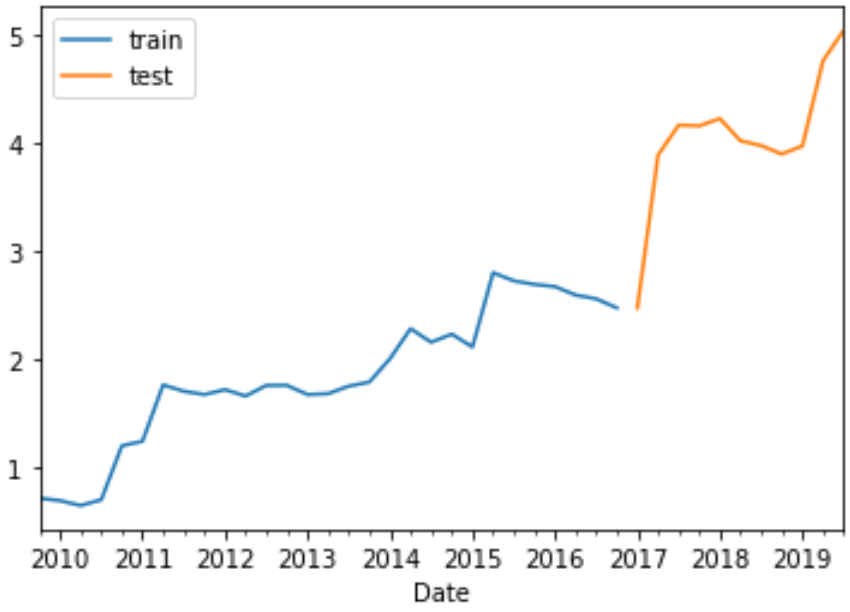

Figure 3.2: Graph of Target Variable

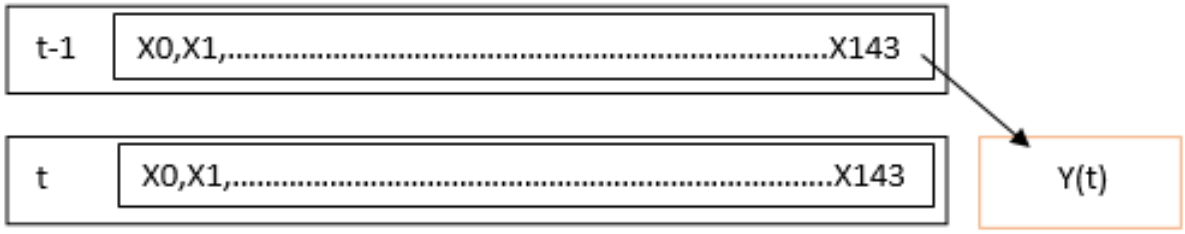

Figure 3.3: Multivariate single step

starting of the time series, target variable is dropped.

ii. Uni-variate single step
This structure is the simplest one, uses only the goodwill column as stream of time series data. Since its single step, Quarter t-1 will assign Quarter t as its target.

iii. Multivariate rolling window
Rolling window concept will be discussed in next chapter. However, to give a short overview, this technique looks back into t-n time laps and predictor is goodwill of time t. Here, n could be any size of the rolling window. If the window size is 3 then, the predictors will be last three quarters' financial data that works to predict target variable of the fourth quarter.

iv. Uni-variate rolling window
In this type of split, Goodwill of previous window size will be stored as observed

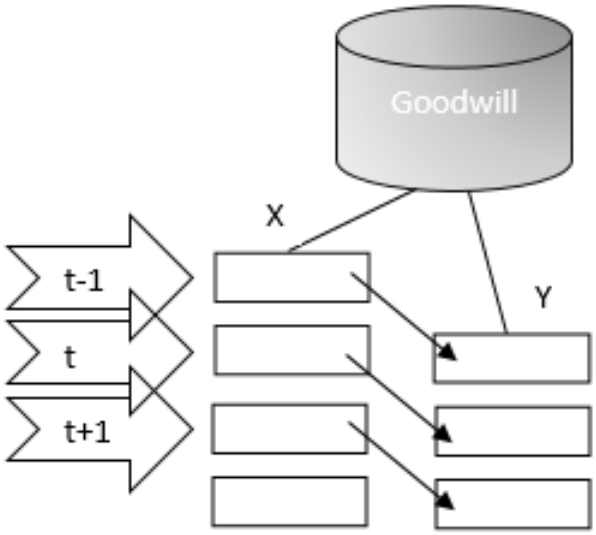

Figure 3.4: Uni-variate single step

```
In [14]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.15, random_state = 0 ,shuffle=False)
         X_train.shape, X_test.shape

Out[14]: ((33, 143), (6, 143))
```

Figure 3.5: Multivariate rolling window

and next step will be the target for the window. Synchronously, the window will shift with the target step size – in this case step forward size is one and the window size is set to three.

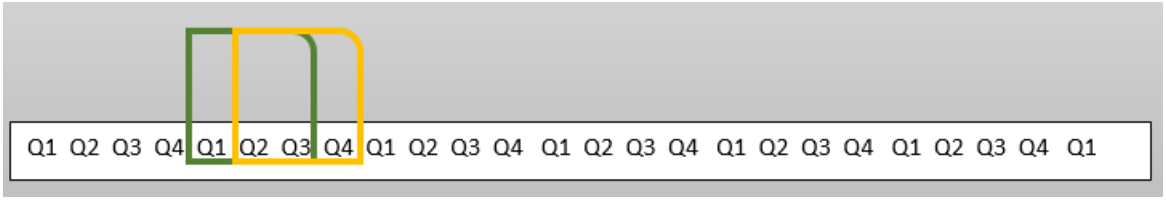

Figure 3.6: Rolling window in date-time axis

In addition, to apply K-fold cross-validation technique, the training set could be divided into training and cross-validation data. The purpose of cross-validation (CV) is to determine the generalization error of an ML algorithm, so as to prevent overfitting. CV will contribute to overfitting through hyper-parameter tuning. CV splits observations drawn from an IID process into two sets: the training set and the testing set. It is done in order to avoid leakage from one collection to the other, as this would defeat the purpose of testing unseen data. K –fold is one of the popular scheme of Cross validation. The dataset is partitioned into k subsets in the k-fold scheme. For i=1,. . . ,k ; the model will train the subsets excluding i. It uses i subset as test set during the train session.

Figure 3.7: k=3 fold cross validation scheme

However, previously we achieved over exaggerated result using cross validation and randomly splitting the collected data set. Later, we discovered that the result was completely misleading since the use case is about time series data. In finance, one reason for failure of k-fold CV is that observations can not be assumed to be drawn from an IID process. A second reason for the failure of CV is that in the process of developing a model, the test set is used multiple times, resulting in multiple testing and selection bias. Through random separation, leakage happens when there is information in the training set that also exists in the test set. Therefore, to prevent leakage from future information, data set were split in sequential manner.

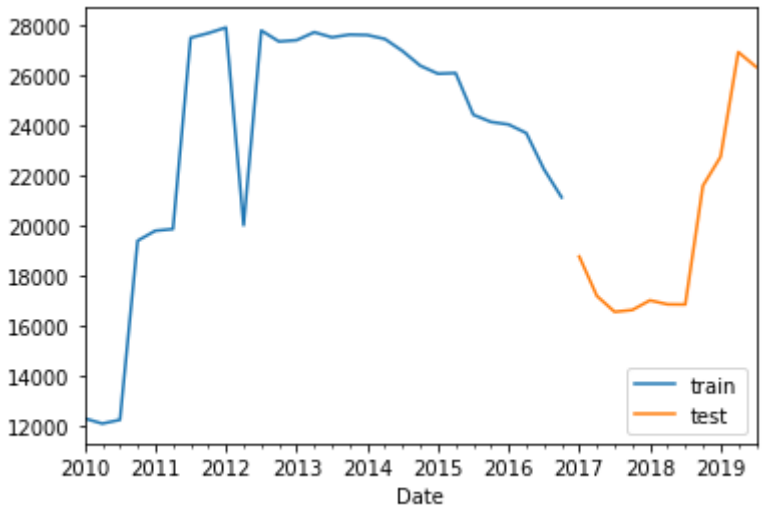

Figure 3.8: Time series train-test split without random shuffle.

3.4 Model Selection

3.4.1 Workflow

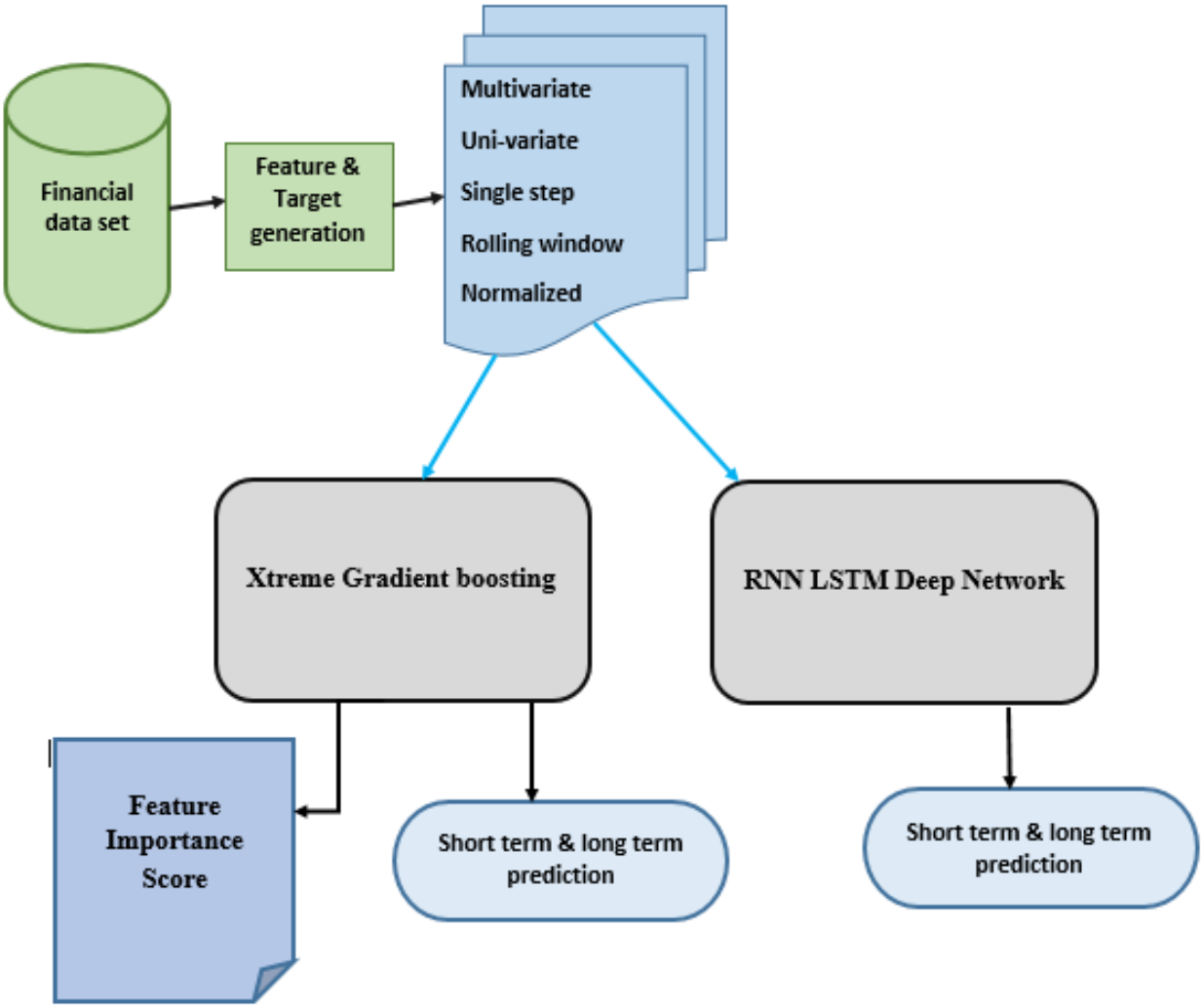

Figure 3.9: High Level Overview of the Workflow

3.4.2 Feature Importance

The purpose of this study is firstly, select those features which contribute most to our prediction variable – acquisition premium. Once we have found what features are important, we can learn more by conducting a number of experiments. Are these features all the time essential, or only in some specific environments? What induces a major change over time? Is it possible to predict these regime switches? Are those important features also applicable to other similar financial instruments? Are they appropriate for other groups of assets? In all financial instruments, what are the most important features?

In machine learning, feature subset selections can be categorized into three schemes in figure 3.10:

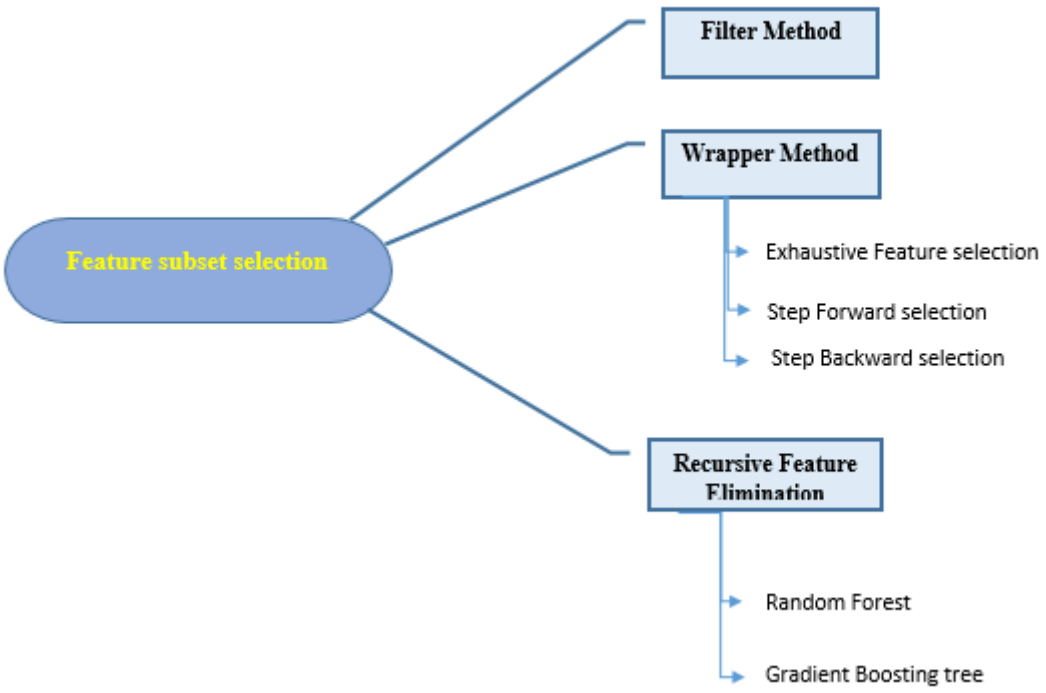

Figure 3.10: Feature subset selection scheme

Filter method searches the training set randomly, looking for the two instances nearest to the one chosen, one from the same class (near-hit) while the other from the opposite class (near-miss), and based on this iterative accumulates weights for attributes. One of the disadvantages of the filter algorithm is that it searches for all relevant features and is unable to distinguish redundant features, even if they are important in a very low degree, so each variable is given some weight. For signal type features, filter method can give meaningful insight but for our domain , filter method will not add any extra value.

The entire training set can be chosen and then evaluated against the test set in a wrapper approach or a cross-validation method can be used. This strategy has a downside when it comes to computing costs. Execution of the learning algorithm can become unfeasible for many sub-sets of features. Since this approach uses cross Validation and we restricted that approach in our domain therefore, wrapper method might not give good result.

Many predictors have their own inherent mechanisms built-in to the learning algorithm that are dedicated to the selection of features. Fortunately, The powerful algorithm we are using in this study- XGboost has this mechanism of finding feature scores.Therefore, we will refer to Recursive feature selection or Embedded method of feature subset selection in this study.

3.5 Artificial Neural Network

Artificial Neural Networks (ANN) are one of the well-established tools of machine learning study. The word "neural" suggest the neurons of human brain. This mechanism is inspired from the brain of a human. This technique imitates the operation of a human brain. The neurons of brain can change their behavior based on chemical reaction or the electrical signal which they are provided.

The way of learning through neural network is progressive. Similar to human learning system, this method contains the feedback. It can be extremely complicated. Connections of each network have different parameters. By calibrating the weight of the networks as well as changing the parameters of the networks can provide different state.

In the figure 3.11, it represents a neural network. The left layer of this neural network is known as input layer. The circles represent the input to the network. The middle layer is known as hidden layer. The reason of hidden layer is, the value of it cannot be found in the training set. Rightmost layer is the output layer. In this scenario, output layer has only unit. There can be multiple unit in the output layer.

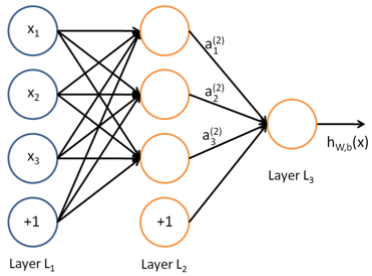

Figure 3.11: Artifical Neural Network

$$
\begin{aligned}
a_1^{(2)} &= f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)}) \\
a_2^{(2)} &= f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)}) \\
a_3^{(2)} &= f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)}) \\
h_{W,b}(x) &= a_1^{(3)} = f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)})
\end{aligned}
$$

Figure 3.12: Activation

Some circles are labeled here as **"+1"**. This units are called bias units. Bias units have only outgoing connection to other layers. However, there are no incoming connection to a bias unit. So, the value of it is not influenced by the values of any previous layers. The output without bias unit can be written as function **output = w * input (y = mx).**

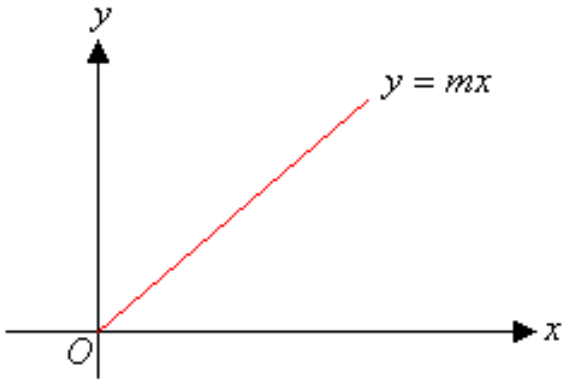

Figure 3.13: Straight Line

If we adjust the function's weight, the function's gradient adjusts to make it more steep. As a result, it can affect crucially to a number of problems. An optimal model may not pass through the origin. To setting y-intercept, the function needs to be changed accordingly. The function needs a bias unit so that it can have constant it the function. Now the function is output = w * input $+w_b(y = mx + c)$. Here, w= weight and b= bias. The w_b denotes the term c.

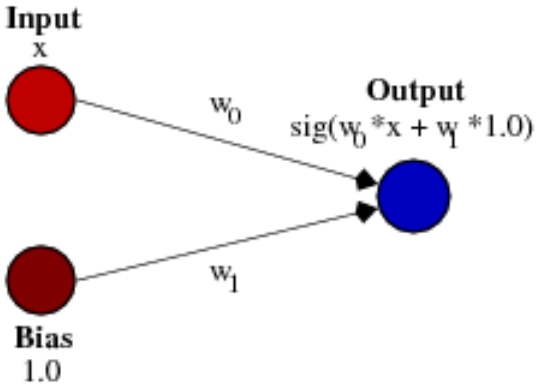

Figure 3.14: Bias

It can be seen from figure 3.14 that we have added the bias term and therefore the weight of w_b will be added to the weighted sum and fed as a constant value through the activation function. This constant term, sometimes referred to as the "intercept term" (as the linear example shows), moves the activation function to either left or right. It will also be the output when the input is zero.

Figure 3.15 shows how various weights transform the activation function (in this case Sigmoid) by scaling it up / down:

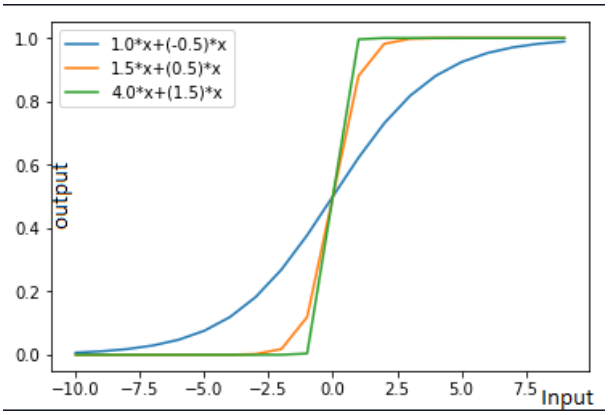

Figure 3.15: Graph of activation function trigerring without bias

But now we have the option to translate the activation function by adding the bias unit which is shown in figure 3.16 :

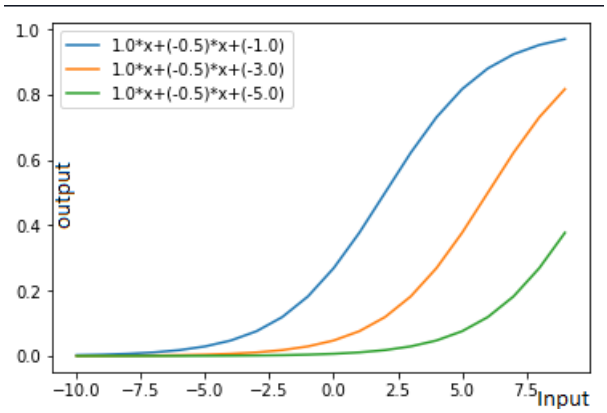

Figure 3.16: Graph of activation function trigerring with bias

Returning to the example of linear regression, if w_b=1, we will add bias • w_b= 1 • $w_b = w_b$ to the activation function. We can create a non-zero y-intercept in the line example shown in figure 3.17:

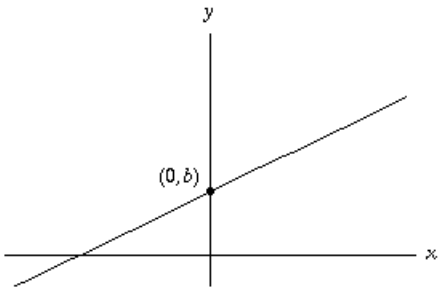

Figure 3.17: linear regression intercept

3.6 Backpropagation

Backpropagation is a common method by which a neural network is trained. This simplifies the network structure through weighted links which has less impact on the network. It uses the utility of chain and power rules. The focus of this method is to minimize the cost function by regulating the biases and weights.

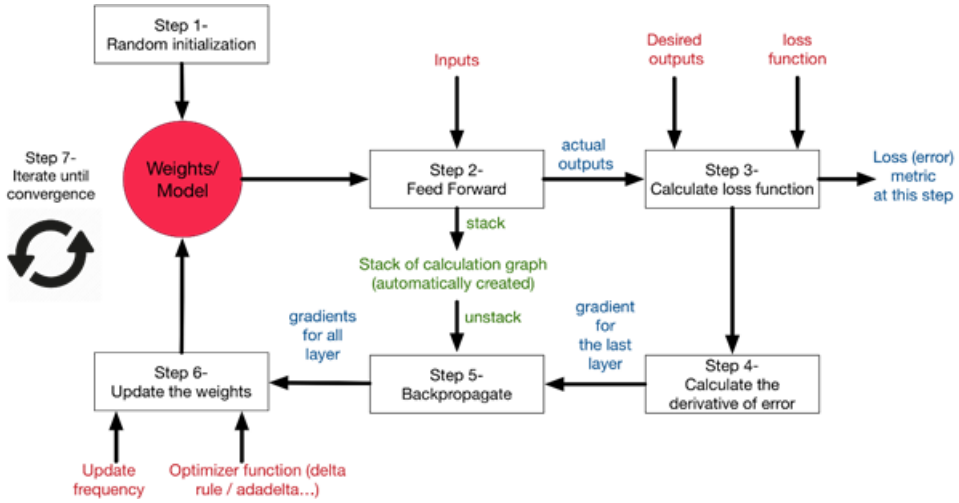

Figure 3.18: Overview of neural networks learning process [33]

Model Initialization Neural networks can begin from anywhere, as in genetic algorithms and theory of evolution. Thus, a model's random initialization is a common practice. The reasoning behind this is that it can reach to the pseudo-ideal model from wherever it begins, if it is sufficiently perseverant and through an iterative learning process.

Forward propagation After initializing the model at random, the next natural step to take is to test its efficiency. Taking the input of the network and pass through it to the network layer and measure the model's actual output directly. This phase is called forward propagation as the calculation stream goes in the direction forward from the input to neural network then to the output [33].

Loss function If a problem is possible to generalize that is what called loss function. It is essentially a performance metric on how well the NN performs to achieve its

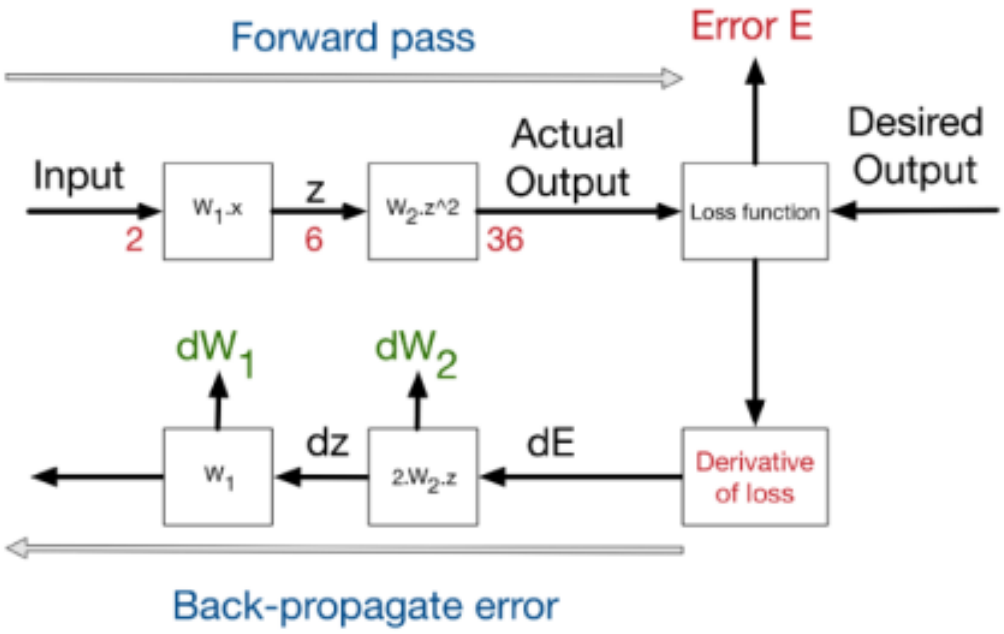

Figure 3.19: Diagram of forward and backward paths

goal of generating results as close to the target values as possible[33]. The straightforward approach of the function equation is

$$function = desired\_output\text{-}actual\_output \tag{3.2}$$

However, the function provides positive values when prediction less than desired output. If the opposite is true i.e. if the prediction greater than desired output, the function provides negative value. To reflect the absolute error, it can be often written as

$$function = Absolute(desired\_output\text{-}actual\_output) \tag{3.3}$$

Under some circumstances, the total error of the function can be same for summation of a number of small errors or summation of few big errors. As it is a prediction work, it is preferable to have the distribution of a number of small errors than a few errors of larger values. Here lamba/2m is regularization term. h-theta is output term.

$$J(\Theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}\left[y_k^{(i)}\log((h_\Theta(x^{(i)}))_k) + (1-y_k^{(i)})\log(1-(h_\Theta(x^{(i)}))_k)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{j,i}^{(l)})^2$$

(3.4)

```
do

   for Each training example named example

      prediction = neural-net-output (network, example) //
      forward pass

      actual = teacher-output(example)

      compute error (prediction - actual) at the output units

      compute {displaystyle Delta w_{h}} for all weights from
      hidden layer to output layer // backward pass

      compute {displaystyle Delta w_{i}} for all weights from
      input layer to hidden layer   // backward pass continued

      update network weights // input layer not modified by error
     estimate

until every example classified correctly or any stopping
criterion satisfied

return the network
```

Figure 3.20: Pseudo Algorithm

Gradient Descent
This is a method that helps to know the w(weight) and b(bias) parameters in a way that minimizes cost of the function. Generally, the cost function of logistic regression is convex in nature. It has only one global minima. This the reason for taking this function. The squared error can have more than one minima.

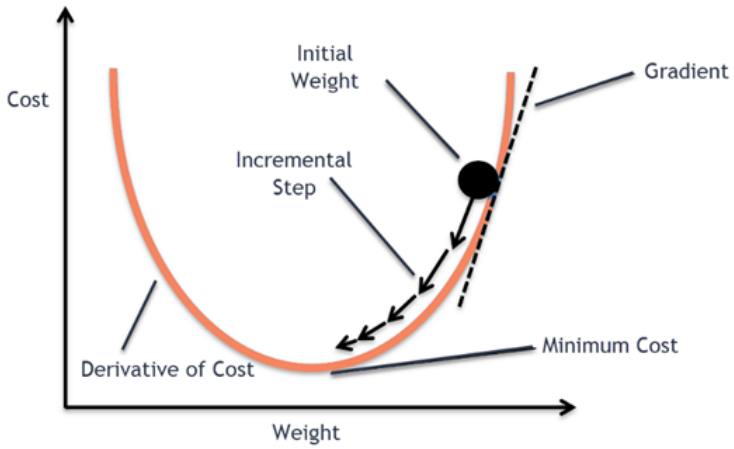

Figure 3.21: Gradient Descent

3.7 Deep Neural Network

Deep learning is part of an artificial neural network-based component of machine learning methods. Deep-learning networks are differentiated by their level of detail from the more general single-layer neural networks. Older neural network models

such as the first representations were shallow consisting of one input and one output layer, and at most one hidden layer in between. Over three stages of learning (including input and output) count as "deep" learning. Deep neural network can overshadow the performance of shallow feed neural network. The reason is, in a shallow network, it has lower number of hidden layers. It produces less accurate results than the deep neural network.

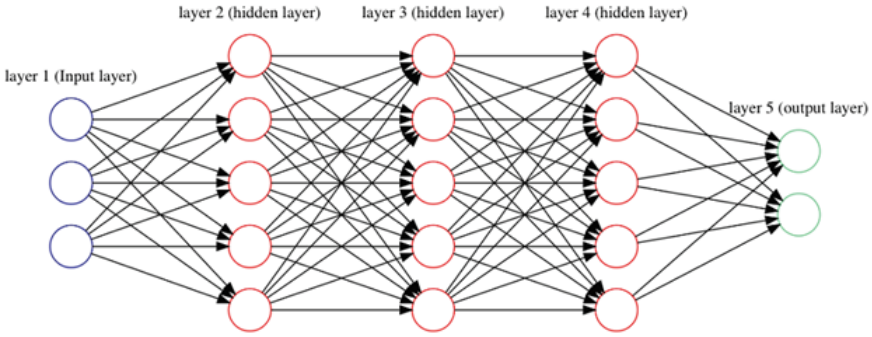

Figure 3.22: Deep layered(3) architecture

3.8 Recurrent Neural Networks

Recurrent Neural Network remembers the past and their actions are guided and informed by what they have learned from the past. Basic feed networks still remember things, but they remember what they discovered during learning process. In a fixed size vector, a regular neural network is used as an input that limits its use in situations involving an input with no predetermined length.

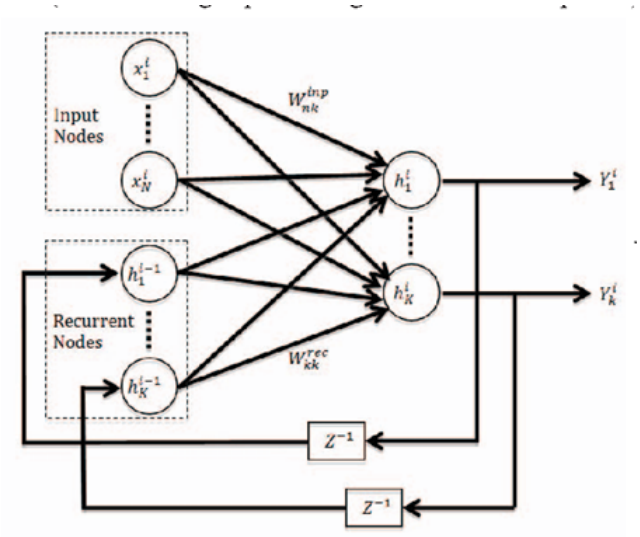

Figure 3.23: Recurrent Neural Network

In particular, the introduction of time is the difference between vanilla feed forward

networks. The output of the hidden layer is reused in RNN. It fed back the output to itself.

Loss function — The loss function L of all-time steps is described in the recurrent neural network based on the loss at each time step. This function is for minimizing the cost of the model.

$$\mathcal{L}(\hat{y}, y) = \sum_{t=1}^{T_y} \mathcal{L}(\hat{y}^{<t>}, y^{<t>})$$

(3.5)

Backpropagation through time- At each point in time, backpropagation is completed. At timestep T, the weight matrix W derivative of the loss L is expressed as

$$\frac{\partial \mathcal{L}^{(T)}}{\partial W} = \sum_{t=1}^{T} \frac{\partial \mathcal{L}^{(T)}}{\partial W}\bigg|_{(t)}$$

(3.6)

The problem with basic recurrent neural networks

Regular recurrent neural networks are not used very often [27]. The key rationale is the issue of the vanishing gradient. Ideally, it is preferable to have long memories for recurrent neural networks so that the network can link information relationships in time at significant distances. With more time steps, chances of backpropagation going out of the context. The gradient, with more steps, going towards zero. As a consequence, the significance of this value is not impactful. The weight adjustment is also less accurate. So, it does not learn from previous data with more time. Thus, regular recurrent neural network are not commonly used. With long short term memory cell, this problem can overcome.

3.9 LSTM

Long Short Term Memory (LSTM) networks are a version of Recurrent Neural Network (RNN). This method is capable of learning long term experience from its extended memory. This networks are well fitted, where the gap between the information is long, in the problem. From their previous context, LSTM can connect with recent data. It can take a single data point and process it to sequences of data through its feedback connection. Long Short Term Memory basically contains three types of gates which are input gate layer, forget gate layer, output gate layer.

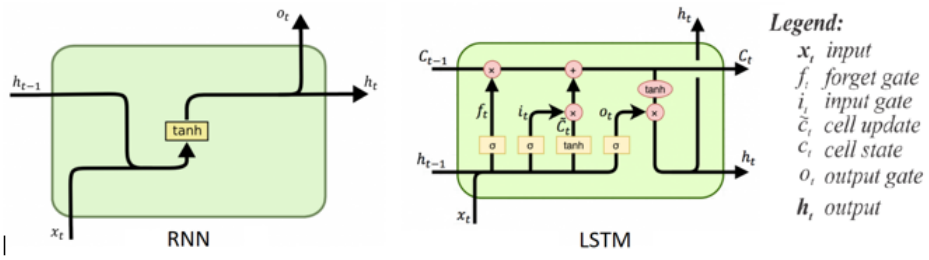

Figure: RNN and LSTM

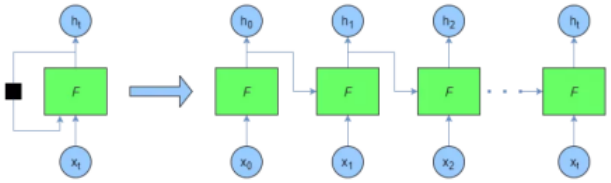

Figure 3.24: Unrolled Recurrent Neural Network

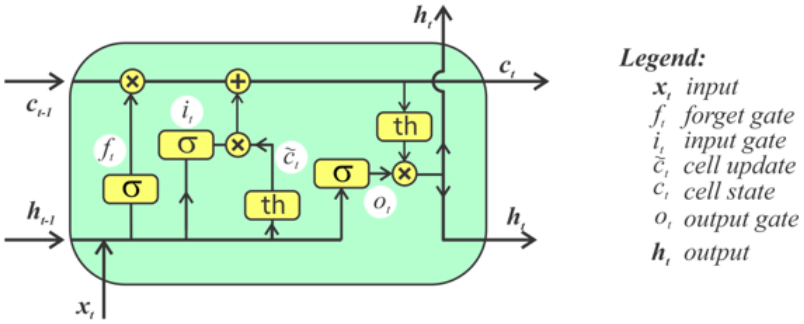

Figure 3.25: General Representation of LSTM

In each neural network cell, the computation has 2 steps.

$$z = WTx + B \tag{3.7}$$

$$propagated\_output, a = activation\_function(z) \tag{3.8}$$

Here, W is the coefficient matrix. It is shared across all the layers. The intercept matrix is represented with b. T represents the layer. X denotes the input, or, the previous layer output. There is activation function. it can be sigmoid, tanh, ReLU, or leaky ReLU.

A forget gate layer, to decide to discard information, implements a sigmoid function layer.

$$f_t = \sigma(Wf.[h_{t-1}, X_t] + bf) \tag{3.9}$$

Then, an input gate layer decides which values are needed to be updated. This is also a sigmoid function.

$$it = \sigma(Wi.h_{t-1}, X_t] + b_i) \tag{3.10}$$

After that, new candidate values, Ct are created as a vector by a tanh layer. These values are added to the state.

$$C_t = tanh(W_c.[h_{t-1}, X_t] + b_c) \tag{3.11}$$

Then the final Ct calculated discarding the ft and propagated to a sigmoid function and a tanh function to get our final output.

$$C_t = f_t * C_{t-1} + it * C_t \tag{3.12}$$

$$ot = \sigma(W_o.[h_{t-1}, X_t] + b_o) \tag{3.13}$$

$$h_t = o_t * tanh(C_t) \tag{3.14}$$

3.10 Xtreme Gradient Boosting

Boosting is a powerful concept in machine learning.In this study we have used an advanced modification of tree boosting-Xgboost, was developed as a research project at university of Washington by Tianqi Chen and Carlos Guestrin in 2016. Xgboost can be used in classification, regression both. ANN are known to work better for images, videos and unstructured data whereas XGboost performs well for small to medium structured/tabular data. Before diving into the tool XGboost which has been used in this study, let's understand the pre-requisite terms to understand XG-boost and existing models before XGboost was there.

Since XGboost is scalable tree boosting technique, we start with how a regression tree works.A simple regression tree partitions the input space and assign a constant target in each specific region. For example, in financial domain we want to construct a regression tree which has two input space – interest rate and earning per share and the target is to predict the market price of the company's bond. Constructed tree will partition several regions in the $twodimensionalspace \rightarrow$ Interest rate and EPS growth. After partitioning, separated regions will be assigned to constant market price of the company's share. Mathematically, if the inputs belong to a region $\rightarrow$ Rj then output Cj.

$$f(x) = \sum_{j=1}^{J} Cj1[X \in Rj] \tag{3.15}$$

The following graph in fig.3.26. shows how simple regression tree will have separate confident regions:

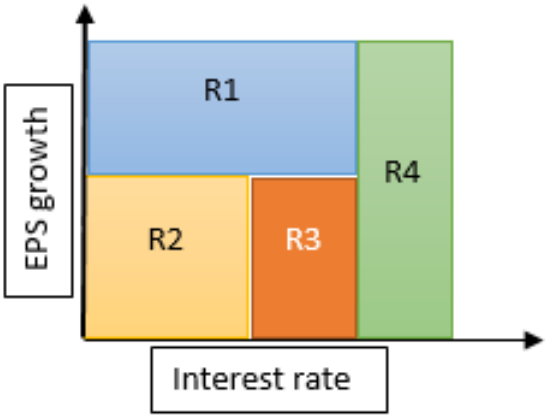

Figure 3.26: Regression Tree

The regions will be assigned with the regression tree so that the empirical loss is minimized. Here, the optimization goal is as follows:

$$[Rj, Cj]_{j=1}^{J} = \theta \tag{3.16}$$

$$min_\theta \frac{1}{N} \sum_{i=1}^{N} (y_i - f(x_i)^2 \tag{3.17}$$

Regression trees typically use mean squared loss. Since the parameters in regression tree equation is, Rj and Cj; the mean square error is depending on these parameters. Equation 3.17 calculates the residuals or the difference from predicted and real value,squares the residuals, divides by the total number of data point N - which is the mean squared error and the goal of the regression tree is to split feature nodes such that equation 3.17 is optimized.

A typical regression tree is very naive approach and gives less accuracy.Individual regression tree filters out information and makes decision. A single tree might not consider all information related to the outcome. Therefore,single model does not guarantee to give high accuracy. Researchers came with the idea of using bunch of trees and average the results to minimize error from a single tree[1]. In modern days we have extensive computational power to implement thousands of trees for prediction. This idea is a fusion from single regression tree model. Intuitively, we can assign one financial officer to track the trend of Goodwill and predict before hand.However, one analyst can definitely go wrong. On contrary, large group of financial analysts will be collectively better than an individual analyst. This idea of 'wisdom of crowd' is known as ensemble in computer science. If we create an

ensemble of regression trees and average their collective decision then we can create a more powerful regression model than a single tree. The idea of ensemble is not restricted to regression tree, we can choose any supervised prediction model that we usually implement individually such as - SVM, Elastic Net,Decision Tree, Random Forest, KNN, Logistic Regression and even neural network. The supervised learning models we use in ensemble are called base model or base learner. If we pick a specific type of learner and provide them the same training set then it will be a homogeneous ensemble. In homogeneous ensemble, features and hyper parameter are uncommon in between base models. We also have option to use various learning models in heterogeneous ensemble model (Reid 2007) [13]. Mathematically, ensemble methods can be represented as in Fig.3.27 and Eq. (3.18) (Chen and Guestrin 2016).[24]

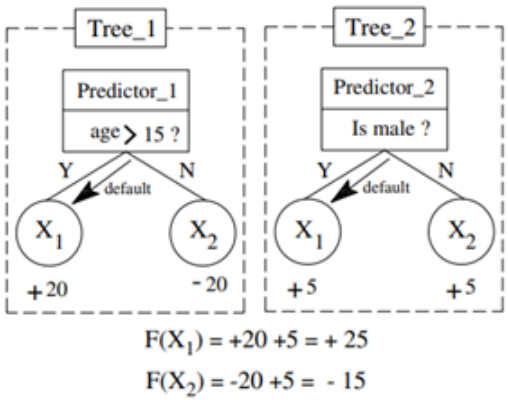

Figure 3.27: Ensemble model using two decision trees as base learner

In previous section we gave the intuition of averaging the regression result produced by multiple base learners. For classification and regression task ensemble methods compromise voting, stacking, bagging, boosting in general.Soon we will dig deeper into boosting method. For understanding ensemble learning, voting and averaging are two of the simplest methods. Averaging is popularly used for regression and voting is used for classification. The ensemble learning methods can effectively deal with the problems of high-dimension data, complex data structures, and small sample size which is suitable in the financial data set used in this paper [5]. Fig 3.27 accumulates two decision trees with same training data set having same features and label. We provide the train data D with features x and target y ,the data has m number of features and number of instance is n, then D=$\{(x_i; y_i)\}$ ($x_i \in R^m, y_i \in R$). Number of tree is a real value K, in figure 3.27 K=2. Equation 3.18 adds each prediction of kth decision tree. Later, if we divide the equation 3.18 with K then the result gives the average prediction by K base learners.

$$\hat{y} = \sum_{k=1}^{K} (f_k)(X_i) \tag{3.18}$$

Since we have the idea of ensemble and regression tree,now we can move into boosting as we described in previous section that boosting can be applied in ensemble.We now give an intuition on boosting and later move to gradient boosting and lastly we move to our main interest which is extreme gradient boosting. Previously we have known that in ensemble,number of models are created at once and then we feed the train data to each of the model.In boosting,the process of building base learner is iterative. To elaborate the idea, let's say we implement one regression tree with the train data,we now have one single model which learns some target labels better than others. The target labels which are not quite well predicted by the tree model is given more weights and then we construct another regression tree in next iteration.The second tree focuses on the weighted targets and try to predict them better than the first model. It is expected that the second model now has some. It starts by fitting an initial model (e.g. a tree or linear regression) to the data. Then a second model is built that focuses on accurately predicting the cases where the first model performs poorly. The combination of these two models is expected to be better than either model alone. The process of boosting can repeat many times. Each successive model attempts to correct for the shortcomings of the combined boosted ensemble of all previous models. Boosting is not a new algorithm, rather it is a concept that can be applied to specific machine learning models. Boosting can improve generalization by decreasing bias error (Schapire et al. 1998).[4]

Gradient boosting is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The name gradient boosting arises because target outcomes for each case are set based on the gradient of the error with respect to the prediction. Each new model takes a step in the direction that minimizes the prediction error, in the space of possible predictions for each training case.

XGBoost is a traditional gradient boosting tree algorithm with a regularization parameter. Once the regularization term is removed, the XGBoost is converted back to the traditional gradient tree boosting. It is a varient of Gradient Boosting Machine. A differentiable convex cost function can be replaced by Taylor series as a second-order approximation for fasting optimization (Friedman et al. 2000)[6].

Another key advantage of XGBoost is handling the missing values for which defaults direction is identified. Accordingly, no effort is needed for cleaning the collected data (Fan et al. 2008) [15]. The unique advantage of the XGBoost is its scalability, so it can process noisy data. XGBoost applies parallel computing to effectively reduce computational complexity and learn faster (Chen and Guestrin 2016)[24]. The unique advantage of XGBoost is its scalability to fit high-dimension data without overfitting based on the following Eq.4.18

$$L() = (x+a)^n \sum_{k=0}^{n} l(\overline{y_i} * y_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3.19}$$

where

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda ||w^2|| \tag{3.20}$$

Where L represents a differentiable convex cost function that determines the difference between the predicted output y^i and the actual output yi. Ω is to avoid overfitting and smooth the learnt weights (Wi). The second term penalizes the complexity of the regression tree functions.

3.11 Why other regression models are not chosen for acquisition premium prediction ?

Many time series researches suggested ARIMA performed well in financial data. We also tried implementing ARIMA in this research. However, our feature space does not hold some vital properties that ARIMA model assumes. Let's imagine we have X_1,....,X_n information. The most fundamental assumption is that X_i are independent. Independence is a nice property, as we can obtain a lot of useful results from using it. The problem is that this property does not hold at times (or frequently, depending on the view). There is only one way for two random variables to be independent, but they can be dependent in different ways. Stationarity is therefore one way of modeling the function of dependency. It turns out that many nice results that hold for independent random variables (law of large numbers, central limit theorem to name a few) hold for stationary random variables (we can simply say sequences) and it turns out that a lot of data can be called stationary, so the principle of stationarity is very useful in modeling non-independent data.

Stationarity is uniquely defined, i.e. data is either stationary or not, so there is only way for data to be stationary, but there are plenty of ways for it to be non-stationary. Again it turns out that after certain transformation a lot of data is stationary. ARIMA model is a non-stationary model. This implies that after differentiation, the data becomes stationary.

In the sense of regression, stationarity is significant because if the data is stationary, the same results that apply to independent data holds.

Most importantly, our feature space is multi-dimensional space. ARIMA mostly works with uni-variate scheme. We could only use the uni-variate single step feature set for ARIMA model. Most of the corporations'acquisition premium of several didn't turn out stationary after differentiating.
The things that are missing from ARIMA algorithms are :

- Not specially intuitive
- No way to build in our underlying understanding about how it works
    * Random walk element
    * Cyclical element
    * External regressors
- Some systems cycle more slowly or stochastically than can be easily described with an ARIMA Model

XGboost uses parallel computation power and has been ruling recent data science competitions including Netflix prize and bunch of Kaggle competitions [12]. It is high time to make efficient use of the parallel computing machines for domains such as finance which produces huge number of vector spaces of features. Since the data set used is tabular, which is suitable for tree based model, unlike most other researches, XGboost should produce good result in our use case. On the other hand deep learning operates differently than XGboost but as mentioned, many studies compared XGboost with Deep neurals since they outperformed in various domain. XGboost is the technique of the ensemble that consecutively builds on weak learners to produce one final strong learner.

Chapter 4

Experimental Results

4.1 Performance Evaluation

Performance analysis metrics help to assess how the established model works with unknown data. It also helps to set confidence bound in future data predictions. In this domain, for anticipating task we want our models to predict each quarter's acquisition premium with high accuracy and precision. For regression analysis, analysts are concerned with how far or close the predicted y is from the actual y. In the graph, if X-axis has the real y value and Y-axis has predicted y, for best case $y=\hat{y}$. To measure how good or bad the model predicts, there are several metrics used in statistics.

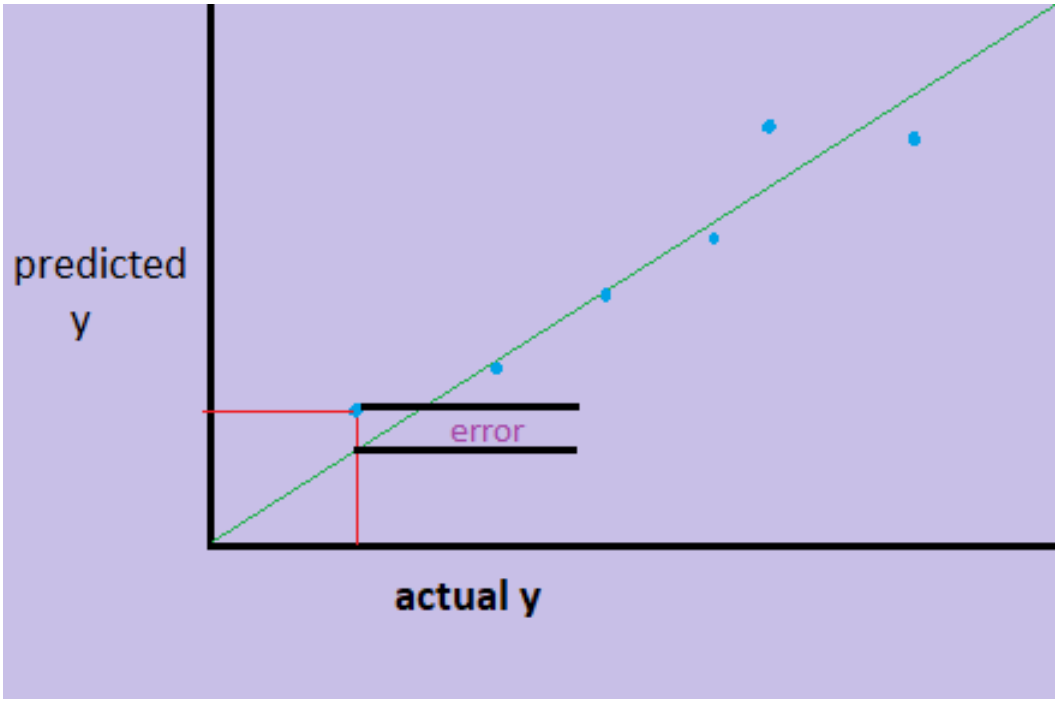

Figure 4.1: Performance Analysis

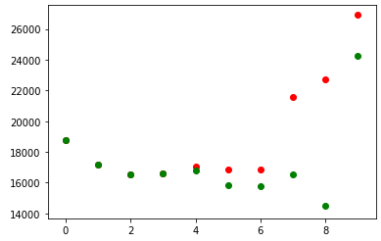

Figure 4.2: Xboost Prediction

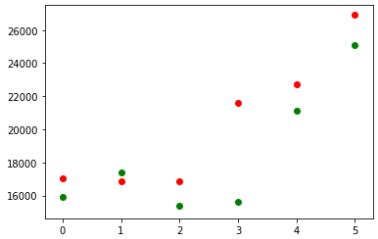

Figure 4.3: Multivariate LSTM prediction

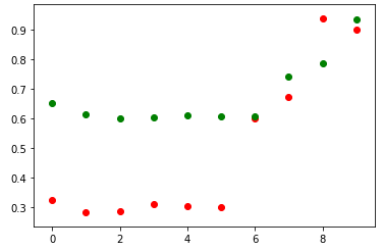

Figure 4.4: Uni-variate LSTM prediction

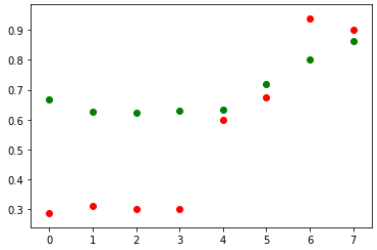

Figure 4.5: Univariate Rolling window LSTM prediction

Figure 4.2-4.5 shows an anticipation of target variable using proposed models of a single company named coca cola in consumer goods industry (index : 30 of Appendix-B) . **Green** dots represent the predicted goodwill and **red** dots are real goodwill. We are now interested to evaluate which models better anticipate the target variable.

The predictions made on a single company of retail industry called HomeDepot (index : 13 of Appendix B) is shown in Figure 4.6 to 4.10.

Assessment techniques for the predictive regression model may include absolute error, mean squared error (MSE), root-mean squared error (RMSE), determination coefficient (R2) and explained variance scores. To penalize the error squared error is used in many cases. However, analyst can have the flexibility while penalizing, any powered error can be set to determine the error of the model depending on the domain. In our case, Root Mean squared error seems feasible since the amount in

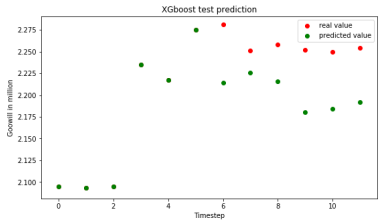

Figure 4.6: XGBoost Prediction

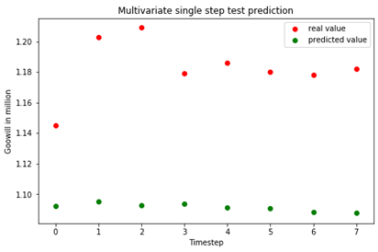

Figure 4.7: Multivariate Single Step LSTM prediction

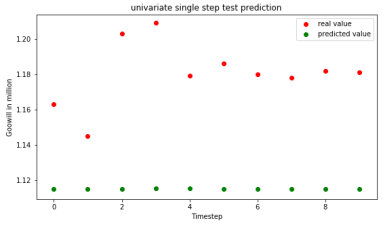

Figure 4.8: Univariate Single Step LSTM prediction

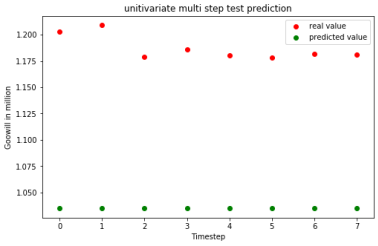

Figure 4.9: Univariate Multi Step LSTM prediction

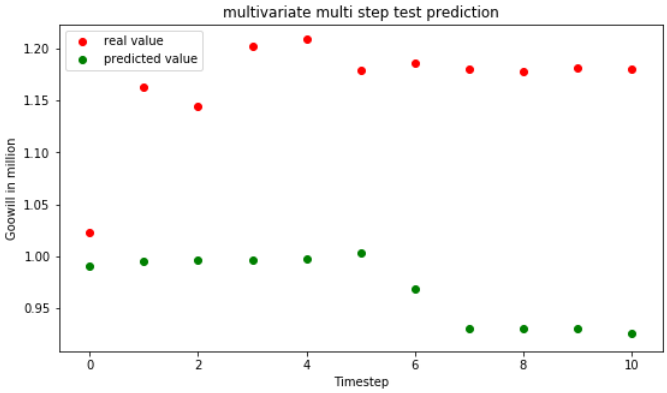

Figure 4.10: Mulivariate Multi Step LSTM prediction

million can be interpreted.

4.2 RMSE

Root Mean Square Error is a method for thoroughly measuring errors in the standard residual deviation. This quadratic method of scoring penalizes most of the bigger errors. This means that when major errors are particularly undesirable, RMSE is more useful. RMSE's formula stands-

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \tag{4.1}$$

Here,
n = the number of observations
y_j = true value
$\hat{y}_j$ = the predicted value

4.3 R Squared

A metric of coefficient of determination. R square can be interpreted as the proportion of the variance for a dependent variable that is explained in a regression model by an independent variable or variables. Higher R squared values generally mean identical to the expected values and real values. A low value of R squared, on the other hand, simply means the values are not that similar.

$$R^2 = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2} \tag{4.2}$$

Here,
$\hat{y}$ = y predicted
y= y actual
$\bar{y}$ = y mean

4.4 Explained Variance

Explained variance is used to calculate the difference between a prediction and actual data. In other words, it is the part of the total variance of the model that is explained by factors that actually exist and is not due to variance of error. Higher percentages of the stated variance indicate a greater correlation power and it also means the predictions are better.

$$explained\_variance(y, \hat{y}) = 1 - \frac{Var\{y - \hat{y}\}}{Var\{y\}} \tag{4.3}$$

Here,
Var$\{y - \hat{y}\}$=[$\sum$((y-$\hat{y}$) $- E(y - \hat{y}))^2$]/n

So that, if $E(y - \hat{y})$ or mean(error)=0, then explained variance = R squared.

4.5 Results

First, we start with the prediction results, averaged by same industry metrics are then compared between the trained models. We will start with explained variance score and see which model can score close to one. Table 4.1 consists with the explained variance score of five pipelines averaged by the same industry. Highest scores are marked in different shade.

From variance explain score we can interpret if our models can understand the distribution of the target variable. XGboost describes the distribution of each industry's goodwill better than other architectures. Uni-variate single step model is worse in almost all industries. From these scores we can interpret XGboost can fit better than others and hence, residuals will be lower. To find better picture we will look at other metrics of evaluation.

Table 4.1: Explained Variance of 7 chosen industries' prediction result

	Recurrent Long Short Term memory				XGBoost
	Univariate		Multivariate		Multivariate
	Single Step	Rolling Window	Single Step	Rolling Window	Single Step
IT	0.536217	0.505658	0.441628	0.2525619	0.9995535
Business	0.18894	0.69923	0.003418	0.016892	0.939297
Retail	0.00837	0.495226	0.016238	0.026287	0.899734
Apparel	0.07137	0.009597	0.441868	0.079276	0.986989
Health	0.57155	0.0195983	0.036568	0.244522	0.987994
Consumer Goods	0.00268	0.31994	0.7314420	0.78713	0.898625

Table 4.2: R square of 7 chosen industries' prediction result

	Recurrent Long Short Term memory				XgBoost
	Univariate		Multivariate		Multivariate
	Single Step	Rolling Window	Single Step	Rolling Window	Single Step
IT	0.065894	0.01402	0.915346	0.5412251	0.99954
Business	0.008309	0.608924	0.071949	0.42974	0.974560
Retail	0.0023570	0.103648	0.699315	0.004438	0.99471
Apparel	0.07630	0.018614	0.359113	0.319068	0.9733760
Health	0.920765	0.3078943	0.86799	0.5787364	0.98742
Consumer Goods	0.001800	0.16883	0.477951	0.79713	0.889178

Variance explained score is a good evaluation metric if mean error is supposed to be zero. If mean score is zero then R squared and variance score would be equal. It turns out that mean error is not zero therefore, R squared value is different from explained variance score. Since these two metrics are correlated therefore, the comparison gives similar result. From table 4.2 XGboost has the highest R squared score in all industries. Although we expected our rolling window models to describe most of the variance since these models are given much information from past. Among LSTM models, multivariate rolling window comes up with better R squared in consumer goods industry. On the other hand, Uni-variate rolling window model has better R squared score in business and retail industry. In terms of R square measure, XGboost ourperforms all type of RNN structures used in the study. XGboost predicts closer values of goodwill.

Usually, the coefficient of determination (R Squared) changes from 0 to 1. An R squared value 1.0 is the best possible result, it means there is no difference between the predicted and real value. A value of 0.0 means that the predicted value and mean is same. In our case, no model came up with zero R squared value that means the models are better than just predicting mean value.

Table 4.3 represents root mean squared error categorized by industry. Lower RMSE values ensure low residuals and more accuracy. Minimum RMSE of each industry is marked with darker shade. RMSE values are always positive as the difference is

Table 4.3: RMSE of 7 chosen industries' prediction result

	Recurrent Long Short Term memory				XgBoost
	Univariate		Multivariate		Multivariate
	Single Step	Rolling Window	Single Step	Rolling Window	Single Step
IT	19331.9401	21266.8811	1007.15748	11647.8394	379.47111
Business	1718.5254	239.7792	63.781048	379.9622	17.133932
Retail	723.4557	1060.3210	154.79879	601.4108	47.423828
Apparel	2495.5714	612.4849	285.330	919.66049	179.56390
Health	5032.435	3863.8799	3527.5891	4548.399	2884.99693
Consumer Goods	0.0033725	0.00502344	0.0026258	0.0103978	0.001000408

squared. An array with many elements will have a larger error and so we average them by number of elements and finally square root to reverse the damage of forcefully making the value positive. We can interpret the RMSE in millions. Apparently, the best result showed up in consumer goods industry. Each architecture has low RMSE in consumer brand. For this particular industry, each pipeline can be used in order to anticipate the intangible asset.

In summary, XGboost is more reliable than other pipelines in predicting goodwill value of organizations using their quantitative financial history. All models can be deployed together in consumer goods industry.

4.6 F Score Results

While interpreting Feature importance score, it requires domain knowledge in financial field. F score will basically help CFO to make educated guess about the firm. CFO may ignore several high scored features for example: Ratios related to intangible assets. Intuitively Goodwill may have strong relation with intangible asset ratios. Therefore, those ratios may not give proper insight of the financial feature spaces. F score will help analysts identify features that have strong explanatory power for acquisition premium. The result may include from trivial outputs to much unexpected output. Thus, helps analysts to connect the dots while interpreting goodwill value. Investors can run the F score analysis and track the importance features for future gains. More importantly, Acquirers may fetch the F score result and forecast particular factor to make educated guess on future acquisition premium. CFO can focus on controlling the goodwill factors to increase a company's goodwill value. In summary, F score reveals the so called 'black box' with trivial and surprising results[18]. Moreover, F score will help to make Capital budgeting decision. Company may take decision in accepting future deals and investments to fixed assets with the goal of optimizing the F score metrics. However, Capital budgeting is not the scope of this research. This study figured out the F score of various leading industries' goodwill and made decision accordingly. For better interpretation, the following section included one F score result from each industry that explains most of the F score output in that particular industry. In each F score result, top ten important features were ranked in ascending order.

Feature importance from IT industry -
Among ten Computer and technology service corporations - feature importance varied from firm to firm. However, few factors were commonly visible in factor analysis result. The following ten factors are common in other IT companies' top ranked F score. We marked 'Interest Debt per share' as the most important factor in IT sector's goodwill.

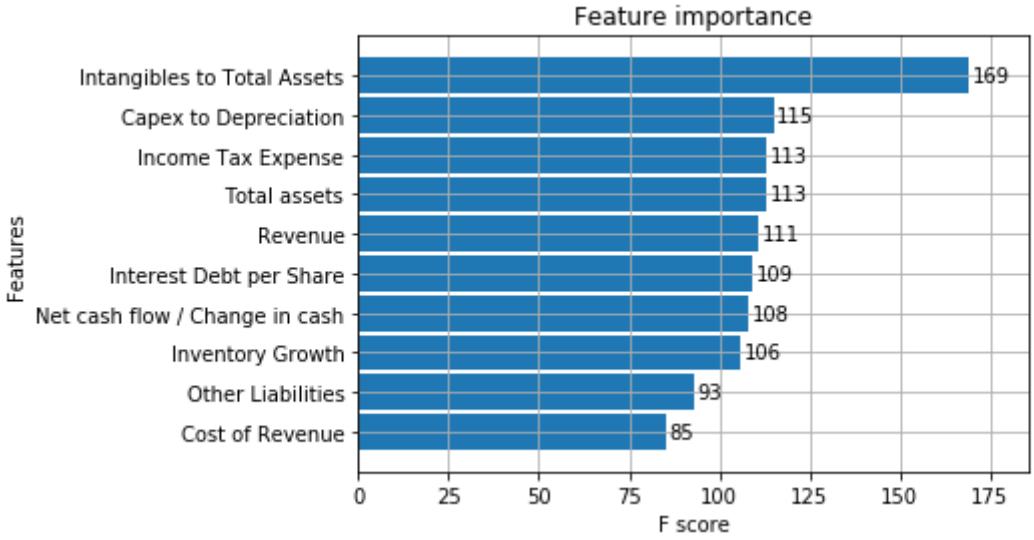

Figure 4.11: HP Company Feature Importance

Feature importance from food industry-
The following F score output is a subset which is present in almost all other companies in the same sector. In this industry – 'EV to free cash flow' is marked as the highest ranked financial factor to the acquisition premium.

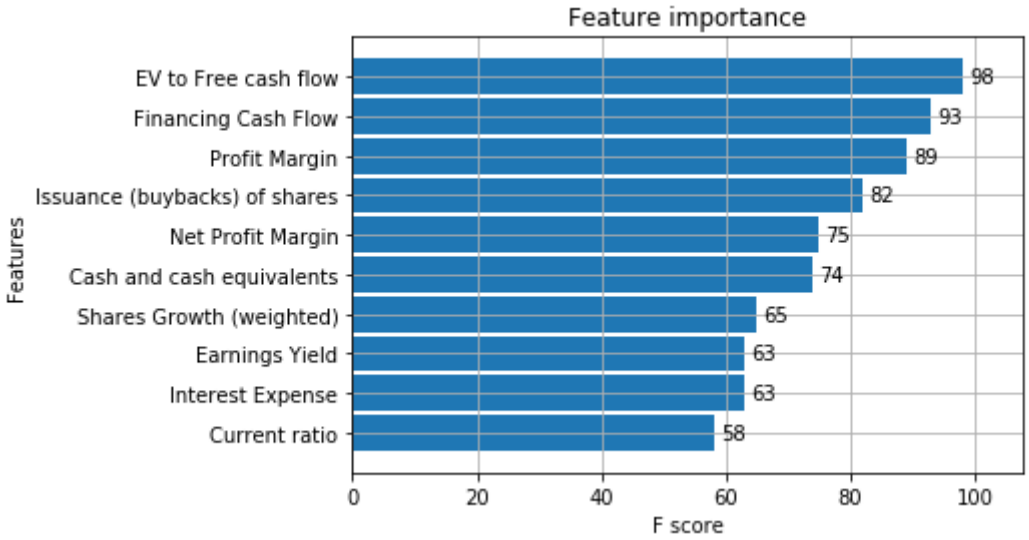

Figure 4.12: B & G Food Company Feature Importance

Feature importance from business and service industry-
'PB ratio', 'EBIT Growth', 'POCF ratio' –are top ranked in most of the corporations

in business and service industry.

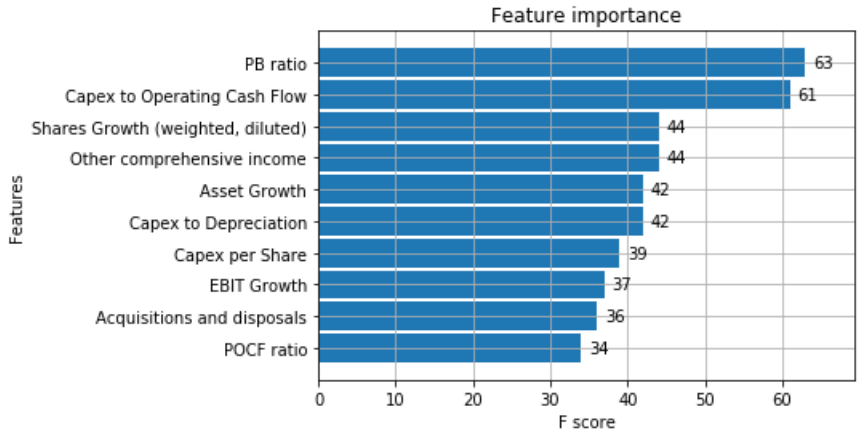

Figure 4.13: Mastercard Company Feature Importance

Feature importance from retail industry-

For retail industry, inventory growth and liquidity ratio such as current ratio are really key determinant for the brand value or intangible asset of the retail company.Xgboost was able to find this insight in this particular industry's goodwill. For example, The Home Depot provides a convenient way for customers to buy equipment for home improvement. Thus, they increase their sales and upgrade inventory faster. As a result their brand value increased and became a household name among craftsmen, builders and DIY-ers. Moreover, Current ratio is simply = current assets/current liabilities.Current asset is a measure of ability to overcome short term debt.Typically an financial analyst would not care about the current ratio while tracking down goodwill value whereas our parallel implementation of gradient boosting trees identifies current asset as importance feature for goodwill value.

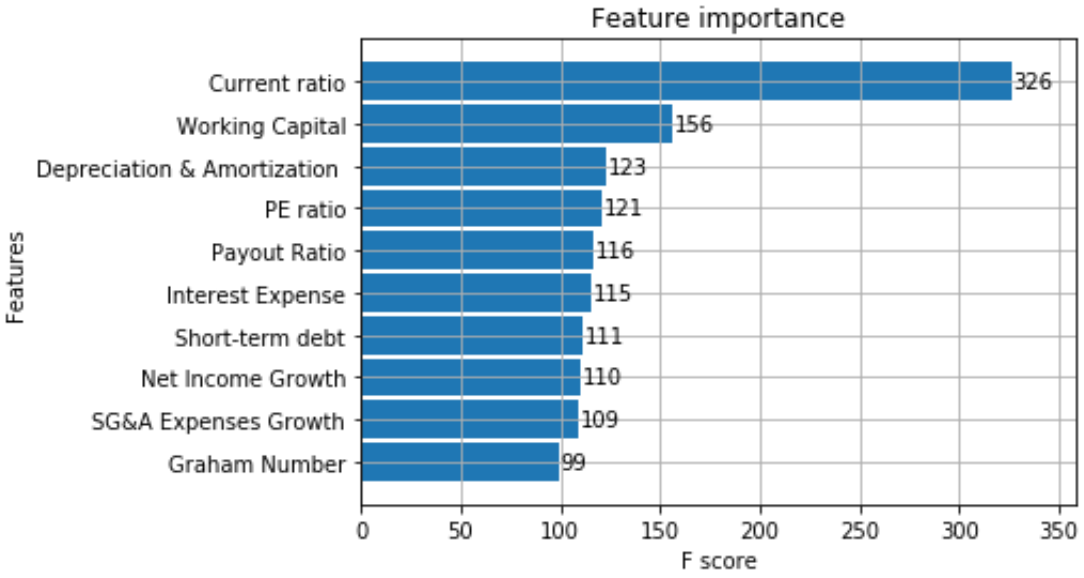

Figure 4.14: HomeDepot Company Feature Importance

Feature importance from apparel industry-

Each industry operates differently in their own target market so is apparel industry. 'SGA Expense' ranked high in this industry. SGA includes the cost of manufacturing and distributing goods and services and the cost of running the business [34]. SGA plays a key role in the productivity of a company and in determining its break-even point, where revenue generated and expenses incurred are the same. When trying to boost profitability, it is also one of the easiest places to look at. Cutting operating expenses, such as wages for non-sales staff, may typically be achieved easily and without affecting production or sales processes. SGA is also one of the first areas that executives search for in mergers or acquisitions to eliminate redundancies. There are a variety of redundant posts and staff after a merger. This field is an easy goal for a management team that aims to boost profits quickly. SGA is not assigned to a particular product and as a result not included in the cost of sold goods (COGS). Industries like garments and textile, selling, general, and administrative expenses (SGA) are vital for profit generation and a target to merger, therefore, Xgboost really figured out the meaningful insight of the given target.

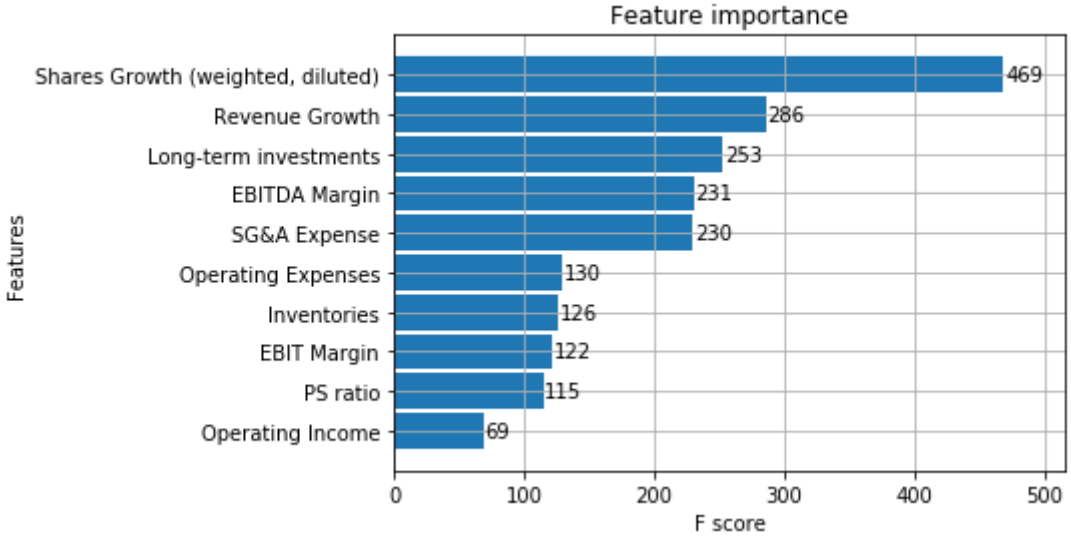

Figure 4.15: Ralph Company Feature Importance

Feature importance from Health and medicine industry-
Among the leading healthcare and medicine brands , united health group is still ranked top in 2019[31]. More cash flow in this sector means more trading of the services. Therefore, ' 'Marketcap' is a key factor for the target factor- goodwill.

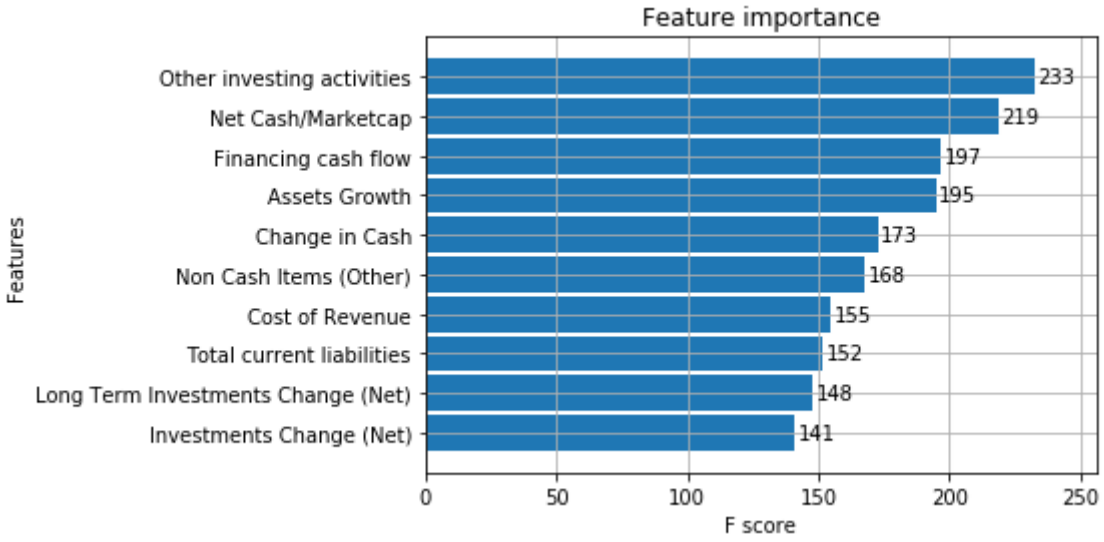

Figure 4.16: United Health Group Incorporated Feature Importance

Feature importance from consumer goods industry-
Leading consumer goods have been dominating the market for a long time. It's really surprising how in practice the goodwill factor are different from IFRS metrics. Interpreting the key factors derived by Xgboost makes sense to get insight about a particular industry. Companies like Coca-Cola, Pepsico, Johnson Johnson – more current assets ensures more cash and more short term investments. At a time when financial information was not as readily available, Graham used this approach, and net-networks were more recognized as a measure of corporate valuation. In defining a viable business as a network, the review focused only on the current assets and liabilities of the company, without taking into account any tangible assets or long-term liabilities. Advances in the processing of financial data now allow investors to quickly access all the financial statements, ratios and other metrics of a business[32]. In consumer brands, cash and equivalents is a good tangible metric to the brand value of the company.

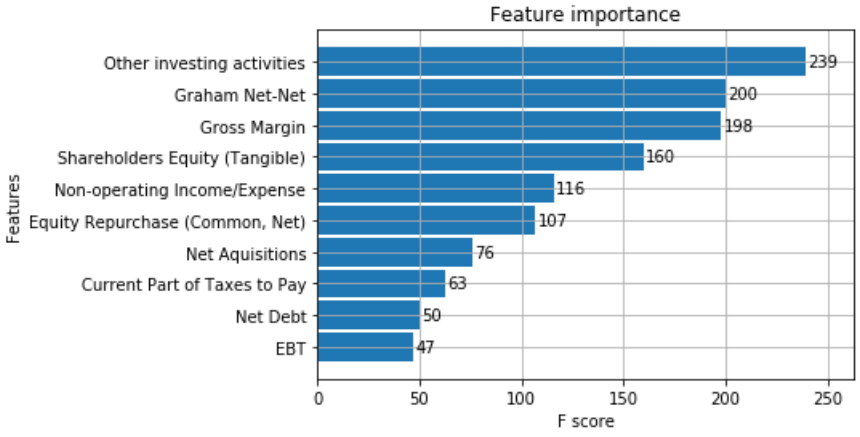

Figure 4.17: Coca Cola Feature Importance

Chapter 5

Conclusion

The objective of this paper was to identify the influential features of individual sectors to evaluate the acquisition price premium. The financial statement of a company combining with the machine learning algorithms assist to find out the decisive features. There is no specific common global procedure of evaluating a company's price premium which actually increase the difficulty to learn the actual real valuation. This research can guide a long-term investor to decide where the money should be invested. Short term investor often goes for the stock data. Whenever the price of a company increases, they sell it and increase their profit. However, the investors who actually want to acquire a company may not get proper assistance from the stock data. With this in mind, various techniques of machine learning techniques have been used.

The data used in this research was second hand data. Second hand data often varies from the actual data. The approach of valuation of different organization is non identical. This creates gap between companies in different sector as a company of different sector may often use different methods to evaluate itself. Thus, universal features for all sectors was providing conflicting result. The data of this research is based on USA. This can also limit the other external factors.

Finance is not some random values on paper. It is actually combined with both arts and science. The real essence of data may still to be discovered. In future with help of more financial statements, real world obligations can help the research to provide better anticipation. Moreover, the outcome of this research is not yet to have an applied usage. By adding more global companies, this research can provide a real-world usage. With the help of other potential data such as PESTEL analysis, porter's generic strategy and five forces, decision pattern and strategies can potentially improve the accuracy. All these remains in the area of future research.

In future, it is hoped to overcome the limitations by having a contract with a company to have a first hand data and daily data for 5 years at minimum. Financial data with time interval of weekly or fortnightly can give birth to many other research areas and can raise the accuracy of the predictions.

Bibliography

[1] L. K. Hansen and P. Salamon, "Neural network ensembles", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, pp. 993–1001, 1990.

[2] R. Pitkethly, "The valuation of patents: Review of patent valuation methods with consideration of option-based methods", *Said Business School. University of Oxford*, 1997.

[3] K. E. Sveiby, *The new organizational wealth: Managing & measuring knowledge-based assets*. Berrett-Koehler Publishers, 1997.

[4] R. E. Schapire, Y. Freund, P. Bartlett, W. S. Lee, *et al.*, "Boosting the margin: A new explanation for the effectiveness of voting methods", *The annals of statistics*, vol. 26, no. 5, pp. 1651–1686, 1998.

[5] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization", *Machine learning*, vol. 40, no. 2, pp. 139–157, 2000.

[6] J. Friedman, T. Hastie, R. Tibshirani, *et al.*, "Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors)", *The annals of statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[7] N. Gradojevic, J. Yang, *et al.*, *The application of artificial neural networks to exchange rate forecasting: The role of market microstructure variables*. Bank of Canada, 2000.

[8] K. G. Rivera and D. Kline, "Discovering new value in intellectual property", *Harvard business review*, vol. 55, 2000.

[9] P. K. H. Phua, D. Ming, and W. Lin, "Neural network with genetically evolved algorithms for stocks prediction", *Asia-Pacific Journal of Operational Research*, vol. 18, no. 1, p. 103, 2001.

[10] S. Benintendi, *Intellectual property valuation: One important step in a successful asset management system*, 2003.

[11] J. H. Matsuura, "An overview of intellectual property and intangible asset valuation models.", *Research Management Review*, vol. 14, no. 1, pp. 33–42, 2004.

[12] J. Bennett, S. Lanning, *et al.*, "The netflix prize", in *Proceedings of KDD cup and workshop*, New York, NY, USA., vol. 2007, 2007, p. 35.

[13] S. Reid, "A review of heterogeneous ensemble methods", *Department of Computer Science, University of Colorado at Boulder*, 2007.

[14] T. A. Stewart, *The wealth of knowledge: Intellectual capital and the twenty-first century organization*. Crown Business, 2007.

[15] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification", *Journal of machine learning research*, vol. 9, no. Aug, pp. 1871–1874, 2008.

[16] T. Sumi, "Intangible asset value evaluation and mot", in *PICMET'08-2008 Portland International Conference on Management of Engineering & Technology*, IEEE, 2008, pp. 17–23.

[17] L.-x. Jian, J.-t. He, and Z.-t. Wang, "The research on intangible assets composition and characteristics of enterprises, universities and governments", in *2009 First International Conference on Information Science and Engineering*, IEEE, 2009, pp. 4597–4600.

[18] M. Giuliani and D. Brännström, "Defining goodwill: A practice perspective", *Journal of Financial Reporting and Accounting*, vol. 9, no. 2, pp. 161–175, 2011.

[19] R. P. Chung, K. K. Lai, and Y. Fu, "A new model on intangible assets valuation", in *2013 Sixth International Conference on Business Intelligence and Financial Engineering*, IEEE, 2013, pp. 181–185.

[20] S. P. Ferris, N. Jayaraman, and S. Sabherwal, "Ceo overconfidence and international merger and acquisition activity", *Journal of Financial and Quantitative Analysis*, vol. 48, no. 1, pp. 137–164, 2013.

[21] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of arima and artificial neural networks models for stock price prediction", *Journal of Applied Mathematics*, vol. 2014, 2014.

[22] Y. Kim, "Convolutional neural networks for sentence classification", *arXiv preprint arXiv:1408.5882*, 2014.

[23] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques", *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, 2015.

[24] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, ACM, 2016, pp. 785–794.

[25] H. Jia, "Investigation into the effectiveness of long short term memory networks for stock price prediction", *arXiv preprint arXiv:1603.07893*, 2016.

[26] G. Rubio, C. M. Manuel, and F. Pérez-Hernández, "Valuing brands under royalty relief methodology according to international accounting and valuation standards", *European journal of management and business economics*, vol. 25, no. 2, pp. 76–87, 2016.

[27] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems.*" O'Reilly Media, Inc.", 2017.

[28] A. Andersen, "Co., 1992. the valuation of intangible assets-special report no. p254", *The Economist Intelligence Unit, London*,

[29] *Financial statement analysis*, https : / / www . readyratios . com / reference / analysis/financial_statement_analysis.html, (Accessed on 12/09/2019).

[30] *Goodwill definition*, https://www.investopedia.com/terms/g/goodwill.asp, (Accessed on 12/09/2019).

[31] *Healthcare_25_free.pdf*, https://brandfinance.com/images/upload/healthcare_25_free.pdf, (Accessed on 12/10/2019).

[32] *Net-net definition*, https://www.investopedia.com/terms/n/net-net.asp, (Accessed on 12/10/2019).

[33] *Neural networks and back-propagation explained in a simple way*, https://medium.com/datathings/neural-networks-and-backpropagation-explained-in-a-simple-way-f540a3611f5e, (Accessed on 12/10/2019).

[34] *Selling, general & administrative expense (sg&a) definition*, https://www.investopedia.com/terms/s/sga.asp, (Accessed on 12/10/2019).

[35] *The importance of forecasting for businesses of all sizes — centage*, https://www.centage.com/the-importance-of-financial-forecasting/, (Accessed on 12/09/2019).

[36] *The past decade and future of ai's impact on society — openmind*, https://www.bbvaopenmind.com/en/articles/the-past-decade-and-future-of-ais-impact-on-society/, (Accessed on 12/09/2019).

Appendix A

Table 5.1: Features

0	Revenue
1	Revenue Growth
2	Cost of Revenue
3	Gross Profit
4	R&D Expenses
5	SG&A Expense
6	Operating Expenses
7	Operating Income
8	Interest Expense
9	Earnings before Tax
10	Income Tax Expense
11	Net Income - Non-Controlling int
12	Net Income - Discontinued ops
13	Net Income
14	Preferred Dividends
15	Net Income Com
16	EPS
17	EPS Diluted
18	Shares (basic)
19	Shares (weighted)
20	Shares (weighted, diluted)
21	Dividend per Share
22	Gross Margin
23	EBITDA Margin
24	EBIT Margin
25	Profit Margin
26	Free Cash Flow margin
27	EBITDA
28	EBIT
29	Consolidated Income
30	Earnings Before Tax Margin
31	Net Profit Margin
32	Cash and cash equivalents
33	Short-term investments
34	Cash and short-term investments
35	Receivables
36	Inventories
37	Total current assets
38	Property, Plant & Equipment Net
39	Goodwill and Intangible Assets
40	Long-term investments
41	Tax assets
42	Total non-current assets
43	Total assets
44	Payables
45	Short-term debt
46	Total current liabilities
47	Long-term debt
48	Total debt
49	Deferred revenue
50	Tax Liabilities
51	Deposit Liabilities
52	Total non-current liabilities
53	Total liabilities
54	Other comprehensive income
55	Retained earnings (deficit)
56	Total shareholders equity
57	Investments
58	Net Debt
59	Other Assets
60	Other Liabilities
61	Depreciation & Amortization
62	Stock-based compensation
63	Operating Cash Flow
64	Capital Expenditure
65	Acquisitions and disposals
66	Investment purchases and sales
67	Investing Cash flow
68	Issuance (repayment) of debt
69	Issuance (buybacks) of shares
70	Dividend payments
71	Financing Cash Flow
72	Effect of forex changes on cash
73	Net cash flow / Change in cash
74	Free Cash Flow
75	Net Cash/Marketcap
76	Revenue per Share
77	Net Income per Share
78	Operating Cash Flow per Share
79	Free Cash Flow per Share
80	Cash per Share
81	Book Value per Share
82	Tangible Book Value per Share
83	Shareholders Equity per Share
84	Interest Debt per Share
85	Market Cap
86	Enterprise Value
87	PE ratio
88	PS ratio
89	POCF ratio
90	PFCF ratio
91	PB ratio
92	PTB ratio
93	EV to Sales
94	EV to EBITDA
95	EV to Operating cash flow

96 EV to Free cash flow
97 Earnings Yield
98 Free Cash Flow Yield
99 Debt to Equity
100 Debt to Assets
101 Net Debt to EBITDA
102 Current ratio
103 Interest Coverage
104 Income Quality
105 Dividend Yield
106 Payout Ratio
107 SG&A to Revenue
108 R&D to Revenue
109 Intangibles to Total Assets
110 Capex to Operating Cash Flow
111 Capex to Revenue
112 Capex to Depreciation
113 Stock-based compensation to Revenue
114 Graham Number
115 Graham Net-Net
116 Working Capital
117 Tangible Asset Value
118 Net Current Asset Value
119 Invested Capital
120 Average Receivables
121 Average Payables
122 Average Inventory
123 Capex per Share
124 Free Cash Flow per Share
125 P/TB per Share
126 Gross Profit Growth
127 EBIT Growth
128 Operating Income Growth
129 Net Income Growth
130 EPS Growth
131 EPS Diluted Growth
132 Shares Growth (weighted)
133 Shares Growth (weighted, diluted)
134 Dividends per Share Growth
135 Operating Cash Flow growth
136 Free Cash Flow growth
137 Receivables growth
138 Inventory Growth
139 Asset Growth
140 Book Value per Share Growth
141 Debt Growth
142 R&D Expense Growth
143 SG&A Expenses Growth

Appendix B

Table 5.2: dictionary of the selected company

	industy	company
0	IT	HP
1	IT	amd
2	IT	microsoft
3	IT	amazon
4	IT	ibm
5	IT	motorola
6	IT	tessco
7	business	Visa inc
8	business	mastercard
9	business	american express
10	business	paypal
11	business	western union
12	business	capital one
13	retail	homedepot
14	retail	walmart
15	retail	costco
16	retail	dollar tree
17	apparel	Raplh lauren
18	apparel	nike
19	food	tyson food
20	food	b&g food
21	food	tree house
22	food	lifeway
23	food	J & J snacks
24	health	united health group
25	health	mckesson corp
26	health	abbott laboratories
27	health	thermo fisher scientefic
28	health	medtronic
29	health	cvs health
30	consumer goods	coca cola
31	consumer goods	johnson and johnson
32	consumer goods	pepsico inc
33	consumer goods	procter and gamble

34	consumer goods	Philip Morris International
35	consumer goods	Altria Group