A Novel Machine Learning Based Method for Stock Price Crash Prediction

Faizan Saifullah

00000206529

Supervisor

Dr. Hasan Sajid

ROBOTICS AND INTELLIGENT MACHINE ENGINEERING

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY (NUST)

ISLAMABAD, OCTOBER 2020

A Novel Machine Learning Based Method for Stock Price Crash Prediction

FAIZAN SAIFULLAH

00000206529

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Robotics and Intelligent Machine Engineering

Thesis Supervisor:

Dr. Hasan Sajid

Thesis Supervisor's Signature: ____

ROBOTICS AND INTELLIGENT MACHINE ENGINEERING

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

OCTOBER, 2020

Declaration

I certify that this research work titled "*A Novel Machine Learning Based Method for Stock Price Crash Prediction*" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

FAIZAN SAIFULLAH

00000206529

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student

FAIZAN SAIFULLAH

00000206529

Signature of Supervisor

Acknowledgements

I am thankful to my Creator Allah Subhana-Watala to have guided me throughout this work at every step and for every new thought which You setup in my mind to improve it. Indeed, I could have done nothing without Your priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life.

I would also like to express special thanks to my supervisor Dr. Hasan Sajid for his help throughout my thesis and also for Machine Learning courses which he has taught me. I can safely say that I haven't learned any other engineering subject in such depth than the ones which he has taught.

I would again like to pay special thanks to Dr. Hasan Sajid again for his tremendous support and cooperation. Each time I got stuck in something, he came up with the solution. Without his help I wouldn’t have been able to complete my thesis. I appreciate his patience and guidance throughout the whole thesis.

I would also like to thank Dr. Jawad Khan, Dr. Karam Dad and Dr. Omer Gillani for being on my thesis guidance and evaluation committee and express my special thanks to Miss Hafsa Inam for her tremendous support and help. I am also thankful to Miss Hasti Sadeghi and Lecturer Hamza Asif for their support and cooperation.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my exceptional parents and adored siblings whose tremendous support and cooperation led me to this wonderful accomplishment.

Abstract

Corporate Governance is a very well-known and accepted technique for the assessment of company performance in the stock market and to predict and ensure that the company's stock value will not fall. This practice throughout the globe helps ensure the safety of investor's money and also keeps company's stakeholders and shareholders on board with the truth about the real strength of the company and its worth. The board of directors in the corporate governance keeps transparency between the managers and the owners of the company. However, it has been seen that the composition and characteristics of this board affect the overall performance of the company. Therefore, in order to assess the performance of this board of governors/directors in light of the characteristics and composition of board, I have implemented a new technique of machine learning that can assess if the company's stock value will crash in the stock market or not, depending upon the characteristics and composition of the board. This thesis uses data from Bloomberg Platform, Osiris and Corporate Library covering 500 banks and financial institutions to validate our algorithm. Moreover, in this thesis, I have compared the empirical results of this algorithm with the baseline known algorithms of SVM and logistic regression. Results show that the proposed algorithm is more accurate than the baseline methods. The thesis concludes with the effects and role of corporate governance features in stock price crash prediction.

Key Words: *Machine Learning, Deep Learning, Corporate Governance Performance, Stock Price Crash Prediction, Financial Risk, Firm Performance*

Table of Contents

List of Figures

List of Tables

INTRODUCTION

The research work in this dissertation has been presented in two parts. First part is related to the baseline machine learning investigations on the stock price crash data. The objective of this part is to study the effect of corporate governance features on stock price crash and to make a model that can predict stock price crash with the help of corporate governance features using baseline machine learning methods. The second part includes the usage of neural networks for predicting stock price crash and the effects of changing the neural network model parameters on the accuracy of prediction. With the help of results, it has been concluded in the thesis that the neural networks have a higher accuracy in predicting stock price crash using corporate governance parameters as the features.

Machine Learning in Corporate Governance

Machine learning which is also known nowadays as deep learning is being used in almost all categories of science and technology and all walks of life. . And since, deep learning is being used in every field of science now, from car tracking using safe city cameras to coronavirus prediction from x-rays, its good time to test the deep learning method for finance sector too. Till 2019, finance sector had been using methods like regression and machine learning baseline methods along with rules for assessing outcomes like stock price prediction and governance system performance. Hence, less work has been done in this field. And in order to predict valuable outcomes like stock price crash prediction and stock price crash risk prediction, machine learning can prove to be very useful. Moreover, predicting these outcomes on the basis of corporate governance system will make it a complete system that can help prevent companies from crashing.

There are several ways to assess the corporate governance system performance and then linking it to stock price crash prediction. These ways involve variables from company's financial and non-financial reports that can highlight the performance of corporate governance in a company. Out of the several identifiers, the two most commonly used variables are

NCSKEW (negative conditional skewness) and DUVOL (Down to Up Volatility). With the help of these variables, we can predict stock price crash and simultaneously at the same time also assess the performance of the governance system of the company.

The system of corporate governance is implemented wherever there is a need to have a third party within a corporate that can link the upper non-technical group of people i.e. shareholders and investors with the technical group of people i.e. managers and CEOs. The requirement of such system for managers and CEOs is to properly convey their concerns and updates to the shareholders. And as for shareholders, they keep an eye on the managers and help in preventing managers from hiding the true performance of the company.

Over the years, corporate governance system has remained related with the events involving shareholder investment loses, unethical management and company crashes. And, the structure and function of corporate governance board plays an important role in it. According to some authors, corporate governance is the transparency and facilitation piece in the puzzle. Its main objective is to align the interests of investors and managers and help prevent companies from crashing by addressing the needs of both parties.

The stock price crash risk is an important factor while making investment decisions and a key part of risk management. It is an important parameter for owners by which they can safeguard the success and growth of company and to the investors for making their investments. With the help of such a tool, there will be a clear image in the eyes of shareholders about the structure and composition of board of governors and the effect of that structure on the stock price crash of their company. Moreover, the investors can then choose the right structure of corporate governance system for their company that will then make sure that managers don’t present a false picture of the performance of the company just for the sake of their own performance based incentives. And the final result of all this will be a successful and growing company with very few losses to suffer.

Research Work

For this purpose, in this thesis, I have used the methods of machine learning that were primarily used by other authors and then I have implemented a neural network modified according to the needs of the data. And with both methods, I have gathered the results relating to the accuracy of a neural network model that can predict the stock price crash on the basis of some of the features of corporate governance and I have compared it with the results of the baseline methods that I implemented.

The corporate governance features used in this study for assessment of stock price crash are divided into four categories i.e. ‘board structure and effectiveness’, ‘ownership structure’, ‘accounting opacity and auditing’ and ‘managerial incentives’. With the help of these four categories, we can fully incorporate the effects of corporate governance system on the stock price crash and it can present the effect of changing these parameters of Board of Governors’ composition on the stock price crash.

In the proposed model, the assessment of the Board of Governors is made every week on the basis of these categories. For that specific week, the model incorporates the board structure of Board of Governors, the ownership structure of institution and insiders, the accounting opacity and auditing features and the managerial incentives to evaluate the effect of each category on the stock price crash and therefore helps in the composition of the corporate governance system. It can also predict whether the stock price of a company will crash or not depending upon the system characteristic of corporate governance that are fed in it as inputs. And this assessment is linked to stock price crash with the help of variables i.e. NCSKEW and DUVOL. But for simplicity of our model, we have only used NCSKEW for now.

2 LITERATURE REVIEW

In the past, a lot of work has been done in the field of predicting stock price crash using machine learning. Baseline methods like random forest, logistic regression and SVM have been used to predict stock price crash and stock price crash risk. In the recent past, some authors have also worked on using neural networks and very promising results have been achieved.

Neural networks have been used since long ago for finance sector problems. The use of neural networks in finance covers a broad spectrum. From bankruptcy to fraudulent activities, neural networks have proven to be very useful. Many authors have used neural networks in finance sector. Hwang & Lin (2000) and Lin et al. (2003, 2004) have examined the use of neural networks for detecting fraudulent financial reporting and management fraud. In bankruptcy prediction, a number of authors including but not limited to Alam et al. (2005), Kim and Kang (2010) and Hu and Tseng (2010) have done work. Nowadays, in the recent past, the attention has shifted to stock price crash risk prediction using machine learning methods.

In early research, authors tried to find stock price crash predictors. Different authors associated different features of corporate governance with crash of stock price. Chen et al. (2001) found that firms with high past returns, past return skewness and high differences in investors' opinions have a high probability of crash risk. Hutton et al. (2009) conclude that transparency in reported earnings is important for the stability of capital markets. They also found that firms with high market-to-book (MTB) exhibit more crash risk. According to Kim et al. (2015) and Callen & Fang (2015) there is a positive relation between firms' profitability (ROA) and its crash risk. Kim et al. (2014) stated that big firms are more likely to crash, whereas firms with high leverage are less likely to crash. In all above researches, all authors used logistic regression or other conventional approaches. Qunfeng Liao (2016) is the first one to use neural networks.

According to Qunfeng Liao (2016), neural networks predict stock price crash risk with better accuracy than logistic regression model and random forest model. Qunfeng built the model with eight financial ratios used by Hutton et al. (2009), Kim et al. (2011a, 2011b) and Kim et al. (2014).

He used the data between 1990 and 2013 in his research. Following Kim et al. (2011a, 2011b), he measured predictors at the end of year‘t-1’ and stock price crash in year ‘t’. The results in his paper show an overall prediction accuracy of 73.48% which is higher than logistic model accuracy of 72.62% and random forest prediction of 71.81%.

Managers have incentives to overstate financial performance by withholding bad news as long as possible because of their compensation contracts and career concerns (Ball, 2009; Graham et al., 2005; Kothari et al., 2009; Lafond & Watts, 2008). According to previous literature, a prominent factor of stock price crash risk is the managerial tendency of withholding bad news from. Particularly, when firm performance falls below investors’ expectations, managers tend to hide the bad news to protect their wealth, reputation, and jobs (Amihud & Lev, 1981; Holmstrom, 1979; Benmelegh et al., 2010; Gormley & Matsa, 2011). Corporate governance mechanisms can help prevent managerial opportunistic behaviors, and reduce stock price crash risk (Shleifer & Vishny, 1997; Healy et al. 1999; An & Zhang, 2013). Usually, a board of director is widely believed to play an important role in corporate governance, particularly in the monitoring of the top management (Fama & Jesen, 1983). Furthermore, independent directors should ensure that financial decisions are made in the best interests of all shareholders (Donaldson & Preston, 1995).

An extensive body of literature suggests that corporate governance mechanisms can help to prevent sub-optimal managerial behavior (Shleifer and Vishny, 1997; Healy et al., 1999). Good corporate governance practices discipline investments (Masulis et al., 2007), prevent earnings management (Xie et al., 2003), improve information disclosure process (Armstrong et al., 2012; Karamanou and Vafeas, 2005), and align interests of managers and shareholders (Benmelegh et al., 2010 among others). Ironically, the structure of executives’ compensation - which is supposed to align interests of managers and shareholders - may also trigger agency problems. Accordingly, Healy (1985), Beneish (1999), Ke (2005), Burns and Kedia (2006), Johnson et al. (2009), Kedia and Philippon (2010) argue that stock-based compensation leads to accounting fraud, misreporting, and earnings mismanagement, followed by the stock price overvaluation and collapse.

Benmelegh et al. (2010) demonstrate that stock-based CEO compensation can cause stock price crashes. They identify conditions under which stock-based compensation leads to suboptimal investment, misreporting, and a subsequent sharp decline in equity prices. Benmelegh et al. (2010) argue that CEOs of medium – to high-growth firms initially have to invest intensively in order to make a better use of growth opportunities. When growth rates slow down, CEOs can camouflage growth decline by making suboptimal investment decisions, resulting in subsequent stock price collapse. Kim et al. (2011b) provide empirical evidence supporting results of Benmelegh et al. (2010).

An and Zhang (2013) explore the relationship between institutional investors' ownership and stock price crash risk, and conclude that strong monitoring by dedicated institutional investors attenuates managers' bad-news hoarding, and so prevents rapid stock price drop. Andreou et al. (2013) consider several corporate governance characteristics and their effects on firm-specific future stock price crashes. They find that future stock price crashes are positively related to institutional ownership, percentage of directors who hold company's shares, and opacity of financial reports. Conversely, the percentage of independent directors on the audit committee and auditor's industry experience are negatively related to stock price crashes.

A lot of work has been done on the study of effect of corporate governance structure on the stock price crash risk. A very recent work that has been done in relation to this is by Kyeongmin Jeon. According to Kyeongmin Jeon (2019), irrespective of the size of the board, if there are independent and expert directors' present in the Board of Governors, corporate governance may have an impact on the reduction of stock price crash risk. He concludes that an effective Board of Governors will be that which will have independent directors in Board of Governors, who are highly qualified and their sole purpose is to keep the company's stock price up.

A more close work to Machine Learning, though not on corporate governance assessment, but instead bankruptcy prediction with the help of machine learning algorithms using financial data of companies was also recently done by Maciej et al. [2] in which they use Extreme Gradient Boosting model to predict bankruptcy of a company based on financial data.

In recent past, several authors have worked in the field of assessing corporate governance using machine learning. In a recent work in 2018 by Elvis et al. [1], work has been done on the assessment of corporate governance framework by determining a structure function. This structure function helps in modeling the functioning of the system and identifying the effect of components on the company failure or success.

There are four general categories in which the corporate governance system is mostly divided i.e. ‘board structure and effectiveness’, ‘ownership structure’, ‘accounting opacity and auditing’ and ‘managerial incentives’. Panayiotis et al. (2016) investigated whether ownership structure, accounting opacity, board structure & processes and managerial incentives attributes relate to future stock price crash risk. Panayiotis applied Principal Component Analysis on the 21 attributes that comprise these four corporate governance dimensions reveals that they can explain between 13.1% and 23.0% of a one standard deviation in crash risk. Transient institutional ownership, CEO stock option incentives and the proportion of directors that hold equity increase crash risk, whilst insiders’ ownership, accounting conservatism, board size and the presence of a corporate governance policy mitigate crash risk. Overall these relationships are more pronounced in environments that accentuate agency risk.

The data that I have used is also divided into these four categories. Carrying on the work, I have applied baseline implementations including SVM and Logistic Regression and then proposed Neural Networks methodology on the data comprised of the four divisions of corporate governance. The methodology shows promising results with Neural Networks concluding that my proposed methodology is better in performance and more extensive than the previous ones.

3 NEURAL NETWORKS MODELS AND MACHINE LEARNING TECHNIQUES METHODOLOGY

In this thesis, I have presented a neural network model that takes as input several variables which are corporate governance features and uses the variable NCSKEW to predict stock price crash value. I have then compared the results of this neural network with the baseline methods of machine learning i.e. Logistic Regression and SVM.

3.1.1 Computing Machine

The computing machine used for this research is a core i-7 with NVIDIA-GTX 1050 to process the data as speedily as possible. The algorithms have been run on python programming language using PyCharm user interactive environment.

3.1.2 Neural Networks

The methodology used in this thesis is the neural networks that are comprised of layers of neurons that interact and find the relation between the input data (features) and the required output (labels). These neurons possess the similarity with the neurons of brains in structure and function. Just like a brain neuron, which has a nucleus that processes the information and the dendrites that pass on the information, the neurons in the neural networks also have the same structure with neurons connected between the layers and activation functions to process the information. The information that is sent forward with all the variables of the neural networks is then calculated at the end using a cost function to classify the output into a label. If there is great difference between the predicted value and the actual label, the data is sent backwards using the method of backpropagation. The purpose of backpropagation is to re-adjust the weights that are used to predict the labels. Finally, when the model has optimized the weights, the parameters are saved and those are used for any new data that comes to predict the labels. The features or input data in our case is the corporate governance parameters and the labels in our case are whether the stock crashed or not.

The following model shows the neural network structure of a single layer network also called as perceptron. Such a neural network gives a single output and has no hidden layers in between:

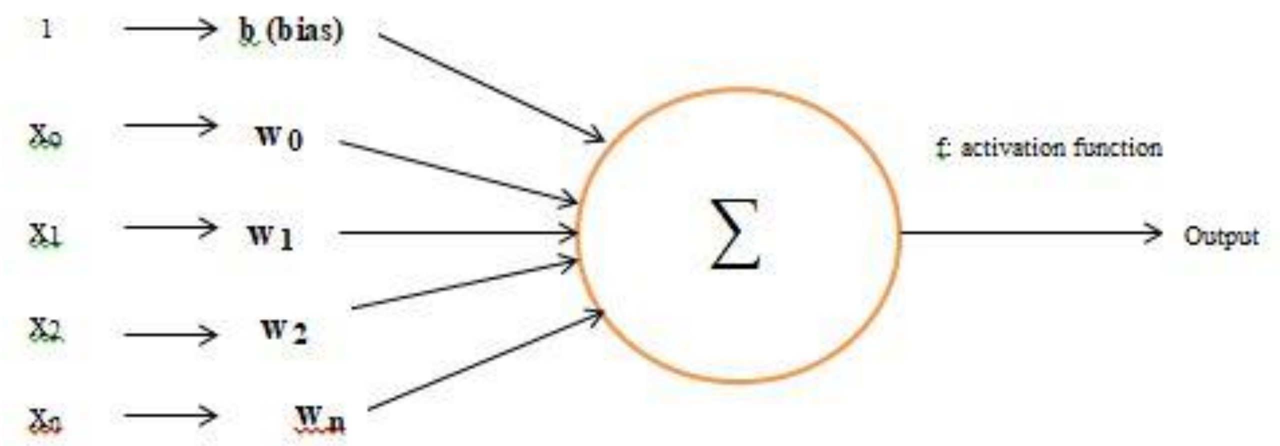

Figure 1 Perceptron Widrow and M. A. Lehr

In the above figure, in a single layer neural network, there are various inputs that are labelled as x_0, x_1, x_2 uptil $x_{n.}$ These inputs are multiplied with weights of the network i.e. w_0, w_1, w_2 uptil $w_{n.}$ Finally, the result obtained from these multiplications is inserted into activation functions and the final output is obtained.

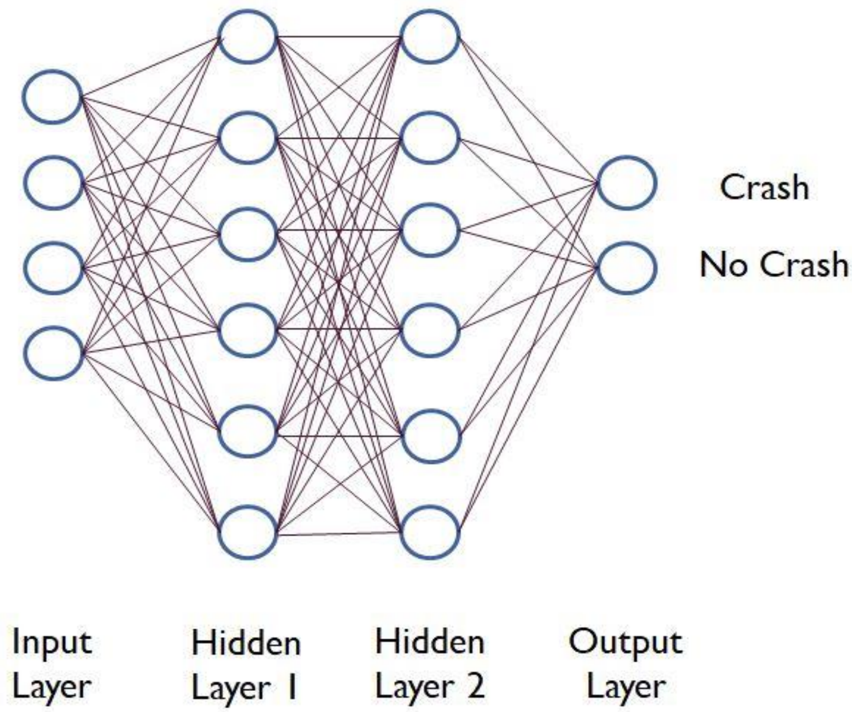

Figure 2 Neural Network J. Tang, C. Deng and G. Huang

This final output is then compared with the original output and when there is deviation, the neural network returns the output backwards. This backward output is then adjusted with derivatives and the new set of weights in the next cycle.

3.1.3 Activation Functions

The Activation function is important for an ANN to learn and make sense of something really complicated. Their main purpose is to convert an input signal of a node in an ANN to an output signal. This output signal is used as input to the next layer in the neural network when there are multiple layers. Activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The motive is to introduce non-linearity into the output of a neuron. If we do not apply activation function then the output signal would be simply linear function (one-degree polynomial). Now, a linear function is easy to solve but they are limited in their complexity, have less power. Without activation function, our model cannot learn and model complicated data.

3.1.4 Types of Activation Functions

There are various types of activation functions that are used widely but the most prominent and frequently used in machine learning are:

1. Threshold Activation Function
2. Sigmoid Activation Function
3. Hyperbolic Tangent Function (tanh)
4. Rectified Linear Units (ReLu)

1. Threshold Activation Function

A threshold function is the activation function that gives the value of either 0 or 1. This activation function is rigid in performance but is used in simpler tasks.

$$f(x)=\begin{cases}0 \text{ if } 0>x\\ 1 \text{ if } x\geq 0\end{cases}$$

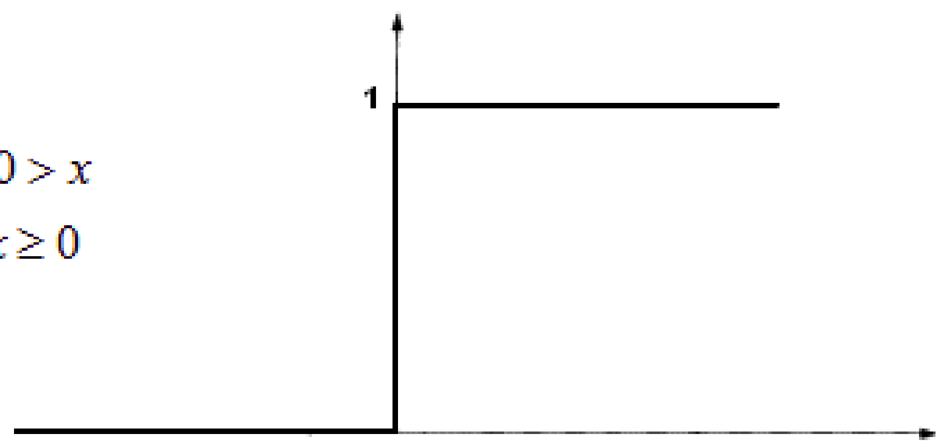

Figure 3 Guang-Bin Huang, Qin-Yu Zhu, K. Z. Mao

2. Sigmoid Activation Function

The second one that is most widely used is the sigmoid activation function. It converts the input into output and gives a value between 0 and 1. This activation function is used in the tasks where probability is required.

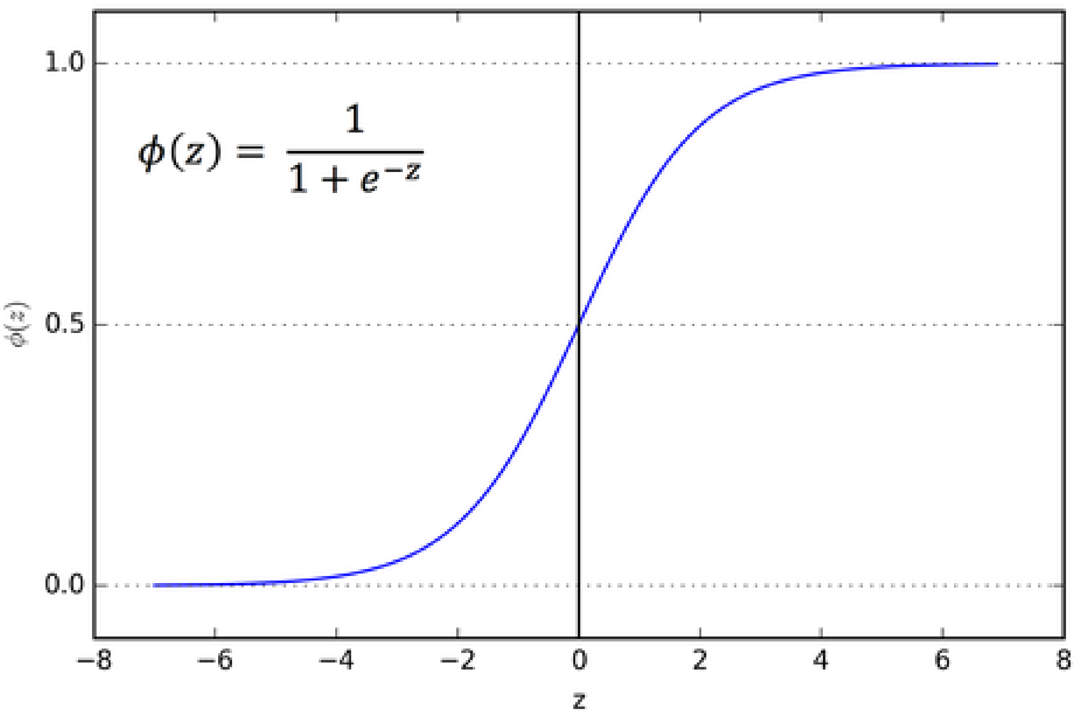

Figure 4 Sigmoid Activation Function, R. Murugadoss and M. Ramakrishnan

3. Hyperbolic Tangent Function (tanh)

The third activation function is the hyperbolic tangent function which is very similar to sigmoid function but is used in the tasks where the range of output lies between -1 and 1.

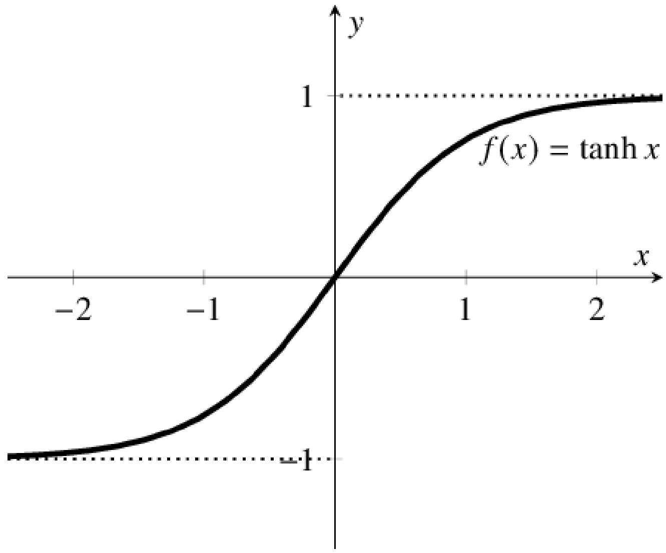

Figure 5 Hyperbolic Tangent Function, A. Wibisono, M. R. Alhamidi

4. Rectified Linear Units (ReLu)

The fourth and most widely used activation function nowadays is the rectified linear unit that has the highest performance and is used uniquely when the output in negative range is not desired. This activation function provides output between 0 and ∞.

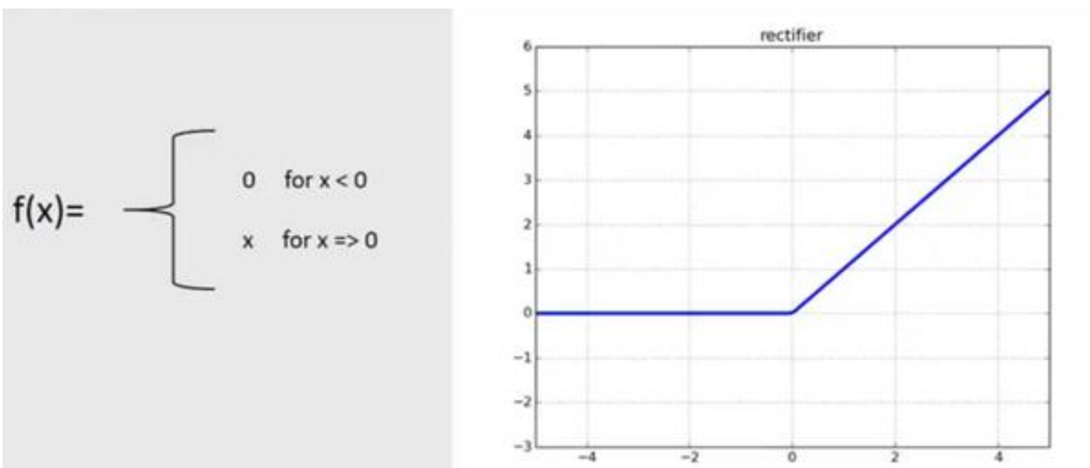

Figure 6 Rectified Linear Units, Y. Ying, J. Su

3.1.5 Backpropagation

Looking at an analogy may be useful in understanding the mechanisms of a neural network. Learning in a neural network is closely related to how we learn in our regular lives and activities — we perform an action and are either accepted or corrected by a trainer or coach to understand how to get better at a certain task. Similarly, neural networks require a trainer in order to describe what should have been produced as a response to the input. Based on the difference between the actual value and the predicted value, an error value also called **Cost Function** is computed and sent back through the system.

For each layer of the network, the cost function is analyzed and used to adjust the threshold and weights for the next input. Our aim is to minimize the cost function. The lower the cost function, the closer the actual value to the predicted value. In this way, the error keeps becoming marginally lesser in each run as the network learns how to analyze values. We feed the resulting data back through the entire neural network. The weighted synapses connecting input variables to the neuron are the only thing we have control over. As long as there exists a disparity between the actual value and the predicted value, we need to adjust those wights. Once we tweak them a little and run the neural network again, A new Cost function will be produced, hopefully, smaller than the last. We need to repeat this process until we scrub the cost function down to as small as possible.

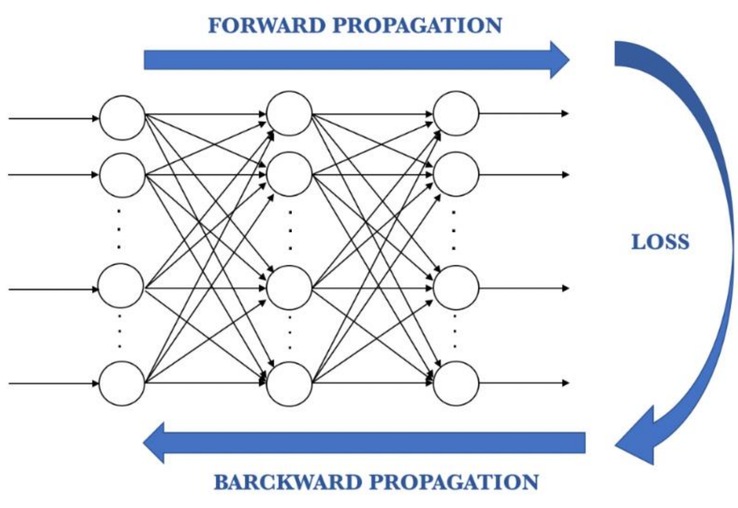

Figure 7 Back Propagation

The procedure described above is known as **Back-propagation** and is applied continuously through a network until the error value is kept at a minimum.

3.1.6 Neural network on Corporate Governance

In order to implement neural networks in corporate governance, we defined the 15 variables as inputs and put them in the neural network. On the basis of NCSKEW values, we added the column of stock crash in our data that will tell us crash if 0 and non-crash if 1.

Data and Collaboration

This research work has been done in collaboration with Miss Hasti Sadeghi who collected the data from Bloomberg Platform, Osiris and Corporate Library. The data is in the form of weekly data which has been collected from 500 banks & Financial Institutions in USA, covering 9 years from 2008 to 2018.

Features

There are fifteen features in the collected data that directly or indirectly depict the performance of corporate governance system and help in predicting stock price crash. These are:

Table 1 Corporate Governance Features

Feature	Abbreviation
%Institutional shares outstanding	%INST_OWN
%Insider shares outstanding	%INS_OWN
% of Independent Directors on Audit Committee	%AUD_CMT_IND

% of Independent Directors on Board	%BRD_IND
Average Age of the Board's Members	BRD_AVG_AGE
% of Women on Board	%WOMEN_ON_BRD
The Number of Members Sitting on the Board	BRD_SIZE
CEO Stock Awards	CEO_INC_STK
CEO Option Awards	CEO_INC_OPT
% Board Compensation Paid in Stocks	%BRD_COM_STK
Return on Equity	ROE
Leverage	LEV
Natural Logarithm of Market Value of Equity	SIZE
Market Value to Book Value of Equity	MV_TO_BV
De-trended average weekly stock trading volume	DTURN

The label used against them is NCSKEW i.e. Negative Conditional Skewness. These features are divided into four categories i.e. 'Board Structure and Effectiveness', 'Ownership Structure', 'Accounting Opacity and Auditing' and 'Managerial Incentives' and are associated with stock crash of companies.

Data Normalization

Before input into model, the data was formatted and normalized to be used with the functions of our neural network. The data has been converted into fraction values. So, for example, if there is a value under ‘%WOMEN_ON_BRD’ which is -0.025, it means that on that week the number of women on board was reduced. We did this conversion also to have an accurate statistical mean.

3.1 Proposed Method

Our machine learning model consists of 4 hidden layers, one output layer and one input layer. The data is fed into the neural network after normalization and goes through several epochs to give us good prediction accuracy.

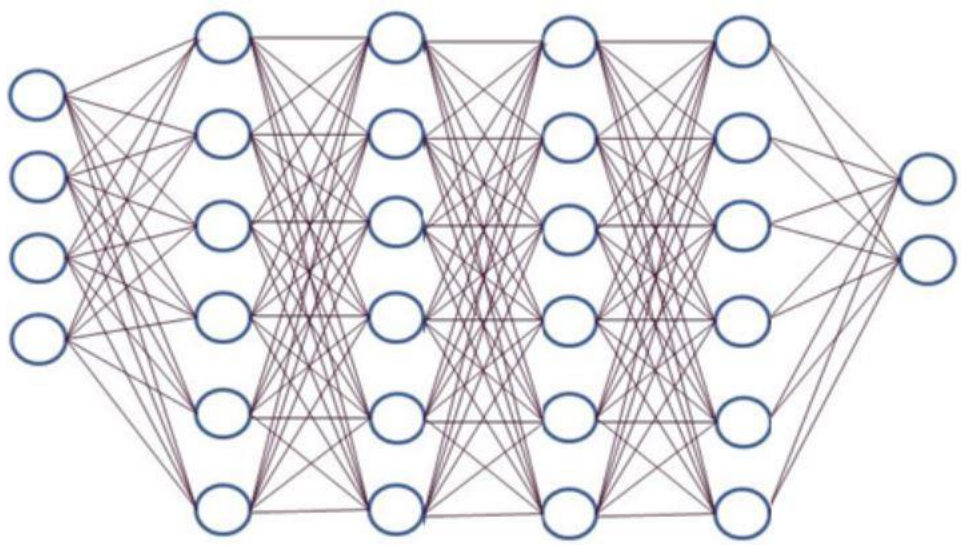

Figure 8 Neural Network Implementation

There are several variables that affect the accuracy and efficiency of the neural network. Among them include the number of hidden layers, hidden layer size, learning rate to reduce cost, batch size being fed into neural network, optimizer parameters, activation functions and output function.

After several thousand iterations, we came to conclusion that there is very little advantage to having more than four hidden layers. However, it adds to computational power significantly if we increase the number of layers. Moreover, the size of the hidden layer at which we got best accuracies was between 100 and 128 and it was best at 128 for two hidden layers and 120 for 4 hidden layers. We kept the batch size to 100, since in the earlier tests; we tested our model on few hundred examples. For optimization of algorithm, we tested gradient descent, stochastic gradient and Adam. The best results were obtained from Adam optimizer however the difference was small. And, in between the layers, the best results were obtained through ReLU activations. Lastly, we applied LogSoftmax at the end of our neural network and computed loss with Negative Log Likelihood (NLLLoss) applied to it.

This is the algorithm applied using pytorch:

```
class DeepNeuralNetwork(nn.Module):
    def __init__(self, inputSize, hidden1Size, hidden2Size, hidden3Size, hidden4Size, numClasses):
        super(DeepNeuralNetwork, self).__init__()
        self.fc1 = nn.Linear(inputSize, hidden1Size)
        self.relu1 = nn.ReLU()
        self.fc2 = nn.Linear(hidden1Size, hidden2Size)
        self.relu2 = nn.ReLU()
        self.fc3 = nn.Linear (hidden2Size, hidden3Size )
        self.relu3 = nn.ReLU ()
        self.fc4 = nn.Linear (hidden3Size, hidden4Size )
        self.relu4 = nn.ReLU ()
        self.fc5 = nn.Linear(hidden4Size, numClasses)
        self.logsm1 = nn.LogSoftmax(dim=1)

    def forward(self, x):
        out = self.fc1(x)
        out = self.relu1(out)
        out = self.fc2(out)
        out = self.relu2(out)
        out = self.fc3(out)
        out = self.relu3(out)
        out = self.fc4(out)
        out = self.relu4(out)
        out = self.fc5(out)
        out = self.logsm1(out)
        return out


dnn = DeepNeuralNetwork(inputSize, hidden1Size, hidden2Size, hidden3Size, hidden4Size, numClasses)
```

Figure 9 Neural Network Code

```
lossFN = nn.NLLLoss()
optimizer = torch.optim.Adam(dnn.parameters(), lr=learningRate)
```

Figure 10 Loss Function

```
for epoch in range(0, numEpoch):
    for i, data in enumerate(trainLoader, 0):
        labels = Variable(data[:, -1])
        data = Variable(data[:, 0:15].float())
        optimizer.zero_grad()
        outputs = dnn(data)
        loss = lossFN(outputs, labels.long())
        loss.backward()
        optimizer.step()
```

Figure 11 Back Propagation

4 RESULTS AND DISCUSSIONS

The tests were conducted using two sensitivity levels i.e. Normal Sensitivity and High Sensitivity. The Normal Sensitivity tests highlight Crash Prediction and High Sensitivity Tests highlight Crash Risk Prediction. In the normal sensitivity tests, data was formatted for labels such that positive to less positive value of NCSKEW was taken as non-crash. In high sensitivity tests, data was formatted such that positive to less positive value of NCSKEW was considered as a crash.

In this chapter, I have performed model performance analysis to predict stock price crash with the help of machine learning methods. Firstly, I have compared the prediction accuracy scores of a neural network on normal sensitivity data with the high sensitivity data. In this section, I have concluded that the prediction accuracies of high sensitivity data are very low and hence it is not useful to predict stock price crash risk. However, the prediction accuracies of stock price crash are in a reasonable range. In the second section, I have done ablation study on the analysis of effects of changing model parameters. In the third section, I have collected model performance results on the basis of categories of corporate governance. In the fourth section, I have discussed another variable that is considered for stock price crash other than NCSKEW. Finally, in the last section, I have done comparison of neural network performance with the baseline methods.

1. Normal Vs High Sensitivity

There is a difference in prediction accuracy when it comes to the sensitivity of the data. If our data is prepared to predict stock price crash risk, the model accuracy will be different than on the data which is used to predict stock price crash. I have studied the performance of neural network on both normal sensitivity data and high sensitivity data on the basis of learning rate, batch size and number of epochs. Following are the results:

a) General Comparison:-

The first two tests were conducted to study the comparison of the accuracies of the sensitivity and they have been studied with reference to learning rate.

First test used NCSKEW as label and all 15 features; 100 training examples and 97 test examples. Files are named as "hettest.csv" and "hettrain.csv".

NumEpoch = 1000 (Epochs)

Batch Size = 10

- **For learning rate: 0.001;**

	Loss	**Accuracy**
1	0.0167	87%

- **For learning rate: 0.01;**

	Loss	**Accuracy**
1	0	86%

Second test used NCSKEW as label and all 15 features; 200 training examples and 100 test examples. Files are named as "hettest22.csv" and "hettrain22.csv".

NumEpoch = 1000 (Epochs)

Batch Size = 10

For Learning Rate = 0.01;

	Loss	Accuracy
1	0.7187	46%

- **For Learning Rate = 0.001;**

	Loss	Accuracy
1	0	51%

In this comparison, I gathered results that conclude that firstly, the model using high sensitivity data has very less accuracies in comparison to the model using normal sensitivity data. Secondly, as we can see that the use of learning rate 0.001 results in better results than 0.01 learning rate.

b) High Sensitivity:-

After understanding that the accuracy of high sensitivity data model is far less than the normal sensitivity data model, I have performed the further analysis of accuracy on high sensitivity data using number of epochs and batch size. If number of epochs is reduced to 100; with learning Rate remaining at 0.001, we get:

	Loss	Accuracy
1	0.0094	51%

This accuracy is same. However, the loss value has increased since the model isn't getting enough time to train properly.

Similarly, then I changed batch Size from 10 to 50;

	Loss	**Accuracy**
1	0.0926	50%

The results show that the accuracy has decreased with an increase in batch size. To cater for this effect, the literature suggests that by increasing learning rate, the accuracy can increase. However, by further increasing learning rate, the accuracy again decreases.

We observed using high sensitivity data i.e. for stock price crash risk prediction, our accuracy of the model is very less. Moreover, there is a range to select number of epochs and batch size. Deviating further from this range our accuracy can further decrease and our loss can increase.

In the **third test,** I tried to increase the accuracy of model using high sensitivity data to bring it closer to the accuracies of normal sensitivity data. For this purpose, I removed the last three features i.e. SIZE, MV_TO_BV & DTURN. These three features are indirectly linked to managerial incentives which may also have effect on the stock price crash prediction low accuracy. For third test, we get:

	Loss	**Accuracy**
1	0.6907	58%

It can be seen from the results, the accuracy of high sensitivity model has increased by 7% when we decrease the number of features. Hence it can be concluded; that the model can make more sense of the data with less number of features but the effect is not enough to bring the accuracy values closer to normal sensitivity model.

In the **fourth test,** we removed two more features i.e. Insider Shares Outstanding and Institutional Shares Outstanding related to the category “Ownership Structure”.

In our results, for hidden layer sizes 128 for each layer;

	Loss	**Accuracy**
1	0.6650	56%

Since, these two variables play vital role in predicting the stock price crash risk prediction, our accuracy decreased. It is very important that we select the variables with utmost observation so that our model can make the most sense out of the data.

Finally, I changed the hidden layer size and found the value with the highest accuracy on this data. The highest accuracy that we get is for hidden layer sizes of 120 each;

	Loss	**Accuracy**
1	0.6589	59%

So, there is an increase in our accuracy but further changing the sizes doesn't change accuracy any further.

After performing analyses on high sensitivity data, it can be concluded that the accuracy of high sensitivity data cannot be further improved. A lower learning rate with higher epoch's number and lower batch size are the best parameters for better accuracy results.

c) Normal Sensitivity:-

After understanding that with lower learning rate but a higher epoch's number, we can get better accuracy results. Finally, I have studied the effect of changing number of epochs and batch size and to understand the effect on normal sensitivity data also.

If number of epochs is reduced to 100; with learning Rate remaining at 0.001, we get:

	Loss	**Accuracy**
1	0.0954	87%

This clearly shows the same behavior of number of epochs that was in high sensitivity data model by increase of loss value and same accuracy.

Similarly, I have changed the batch size from 10 to 20:

- **For learning rate: 0.001;**

	Loss	**Accuracy**
1	0.0026	86%

- **For learning rate: 0.001;**

	Loss	**Accuracy**
1	0	85%

It can be seen that the accuracy of the model decreases by increasing batch size. For this purpose, we need to increase learning rate but that then again affects our prediction accuracy.

2. Analysis on Effects of Changing Model:-

a) Adding data

In this test, we used all of our data which was:

3113 training examples

1000 test examples

For a batch size of 100 and number of epochs 1000, we got following results:

	Loss	**Accuracy**
1	0.0000	82%

Our model's main objective was to reduce the loss value to as close as zero. So, for a higher loss value, our accuracy was better:

	Loss	**Accuracy**
1	0.2576	86%

Hence, it can be seen from the results, that the accuracy has been improved with a better loss value. The more data the network sees, the better it can learn from the data.

b) Changing Number of layers

In this test, we increased number of layers. We chose number of layers as 2, 4 and 6. Our model accuracy, with epoch 1000 and batch size 100 becomes:

Number of Layers	**Loss**	**Accuracy**
2	0	82%
4	0	84%
6	0	84%

As it can be seen, that the model accuracy improved when the number of layers becomes 4 but there is no further effect or benefit of increasing number of layers to 6.

c) Changing Layer Size

For the analysis of effect of layer size, we chose 50, 100 and 150. The model performance on changing layer size is:

Layer Size	Accuracy (%)	Loss
50	80	0
100	81	0
150	80	0

Then, we used layer size of 128 that is mostly used in the literature, we get better accuracy. Lastly, we tweaked around with the number and got the best accuracy at 120 for four hidden layers network.

Layer Size	Accuracy (%)	Loss
128	82	0
120	84	0

d) Optimizer Performance

Optimizer is one of the most integral parts of the neural networks. A good optimizer can provide better training in finding the global minima of our loss function. For our dataset, using ‘Gradient Descent’ and ‘Stochastic Gradient Descent’ result in less accuracy and more loss value instead of Adam optimizer. However, the effect is very less.

e) Learning Rate

The literature states that the learning rate is the most crucial element of designing a neural network. With the wrong learning rate, we can get stuck in local minima. But with using learning rate of 0.001, it brings better results than using 0.01. I have compensated for the decrease in learning rate by increasing number of epochs which gives the network more time to train. As shown in section 1, the accuracy on high sensitivity data is 51% with learning rate of 0.001 in comparison to 46% with learning rate of 0.01.Similarly, the accuracy on normal sensitivity data is 87% with learning rate of 0.001 in comparison to 86% with learning rate of 0.01.

4.1.1 Best Output:-

- Accuracy(Normal Sensitivity – Less data): 87%
- Accuracy(Normal Sensitivity – Full data): 82%
- Accuracy(Normal Sensitivity – Full data): 84% (4 hidden layers)
- Accuracy (High Sensitivity – Less data): 59%

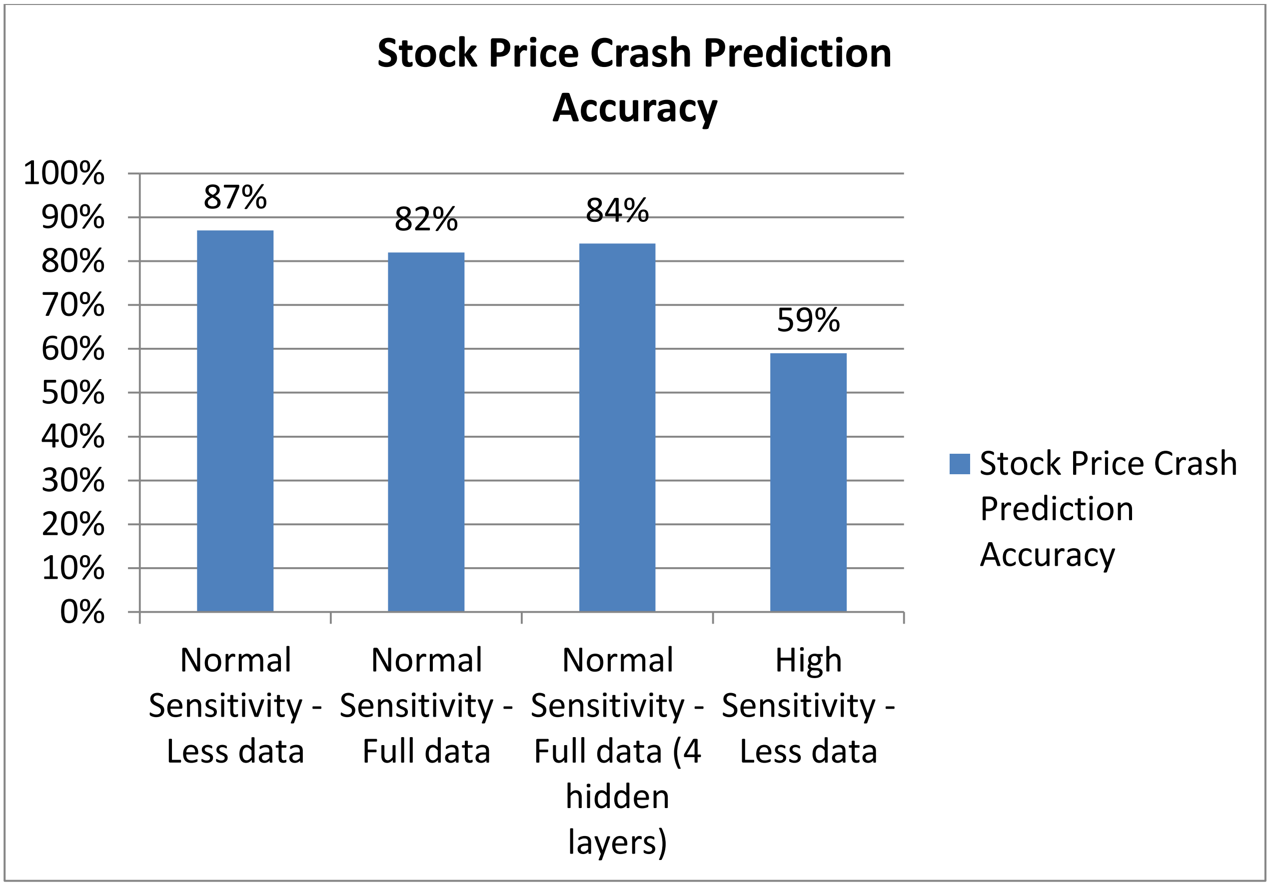

Figure 12 Accuracy Results from Neural Networks

3. Category-wise Results

With normal sensitivity, I then performed analysis on corporate governance performance on the basis of categories. There are four categories in which our data is divided, I have performed accuracy tests on these four categories to compare and understand that which category can provide higher accuracy and has more meaningful relation with the stock price crash.

Categories of Corporate Governance which our data covers are:

1. Ownership Structure
2. Accounting Opacity and Auditing
3. Board Structure and Effectiveness
4. Managerial Incentives

1. Ownership Structure:-

In the ownership structure category, there are two variables that I am using as features i.e. % Institutional Shares Outstanding and % Insider Shares Outstanding. When I use only these two features to predict stock price crash, I get accuracy of 82% in 1000 epochs.

	Loss	**Accuracy**
1	0.1591	82%

In addition to the previous research done on the relationship of ownership structure with the stock price crash, the results also suggest that ownership structure has direct relationship with stock price crash. The prediction accuracy of 82% clearly states that if the model is given the ownership structure of a corporate governance system, it can fairly predict stock price crash.

2. Accounting Opacity and Auditing:-

In the Accounting Opacity and Auditing Category of corporate governance system, we have one variable in our data i.e. % of Independent Directors on Audit Committee. The research states that more the number of independent directors on board, the less chance there is that the stock will crash.

	Loss	**Accuracy**
1	0.3048	86%

The percentage of independent directors does not change in our data set. And, for same value of this variable, there are few entries in which the stock price crashed. Therefore, the accuracy is 86% because for these entries the model predicted that the stock price will not crash.

3. Board Structure and Effectiveness

In the board structure and effectiveness category, there are four variables that we have used as features i.e. % of Independent Directors on Board, Average Age of the Board's Members, % of Women on Board and the Number of Members sitting on the Board.

The model can predict stock price crash on the basis of board structure and effectiveness with an accuracy of:

	Loss	**Accuracy**
1	0.1604	77%

In comparison to the other results, accuracy in this category is lower. The reason behind this is that there is no certainty that depending upon the presence of women, the stock price crash will occur more or less. Neither there is any concrete relation between the number of members on board and the probability of stock price crash because as documented in previous researches, sometimes by increasing number of board members, the stock price is more likely to crash and sometimes it is less likely to crash. Similarly, the remaining two variables of % of independent directors on board and average age of the board's members, also present different behaviors at different points of stock price crash. At some stock price crashes, the number of independent directors is increasing and at some crashes it decreases. And there are some crashes where more aged members of the board are present and at some there are less aged board members.

4. Managerial Incentives:-

In this category, out of six variables, four features are directly linked to managerial incentives i.e. CEO Stock Awards, CEO Option Awards, % Board Compensation Paid in Stocks and Return on Equity.

The model prediction accuracy about stock price crash on the basis of these managerial incentives data is:

	Loss	Accuracy
1	0.3442	82%

DUVOL:-

The second parameter that is a predictor of stock price crash is DUVOL. There exists a partial relationship between DUVOL and ROE. In many of the cases, when DUVOL is increasing, meaning the weeks with high volatility in stock price, return on equity of board members increases but the stock crash probability also increases.

The main relationship between DUVOL and stock price crash is that the company stock is unstable, the more the chances of crash. However, it is not necessarily true always.

4. Comparison with Baseline Methods

a) SVM Results

The same data was inserted in SVM algorithm. Using python language in PyCharm interactive environment, I ran the SVM algorithm on the same data set comprising 15 features and the accuracy is lower than the neural networks.

The data was split into testing and training dataset using sklearn_model_selection option into 80/20. The data on which the model performed training and set its parameters was 80% of the total data. The data for testing phase on which predictions were done and accuracy was measured was 20%. The accuracy measured through SVM algorithm is 84%. This accuracy in comparison to neural networks is slightly less and will have drastic effect on results if the shareholders' investment is dependent on it.

The code for svm was:

```
data = pd.read_csv("full_data1.csv")

print(data.head())

y = data.CRASH
x = data.drop('CRASH', axis=1)

#print(x.head())
#print(y.tail())


x_train, x_test, y_train, y_test = sklearn.model_selection.train_test_split(x, y, test_size=0.2)


clfn = svm.SVC()
clfn.fit(x_train, y_train)

y_pred = clfn.predict(x_test)

accu = metrics.accuracy_score(y_test, y_pred)

print(accu)
```

b) Logistic Regression

The second baseline method with which the model was compared is logistic regression method. The same technique was adopted for this algorithm also in which the algorithm was asked to predict between crash and non-crash states. The data was inserted with all fifteen features and the accuracy was 75%.

c) Comparison of Prediction Accuracies

After taking results from both baseline methods and neural networks, the results show that the performance of neural networks is better than the baseline methods. The study shows promise that with the help of neural networks, better predictions regarding stock price crash can be made. Following are the results for comparison:

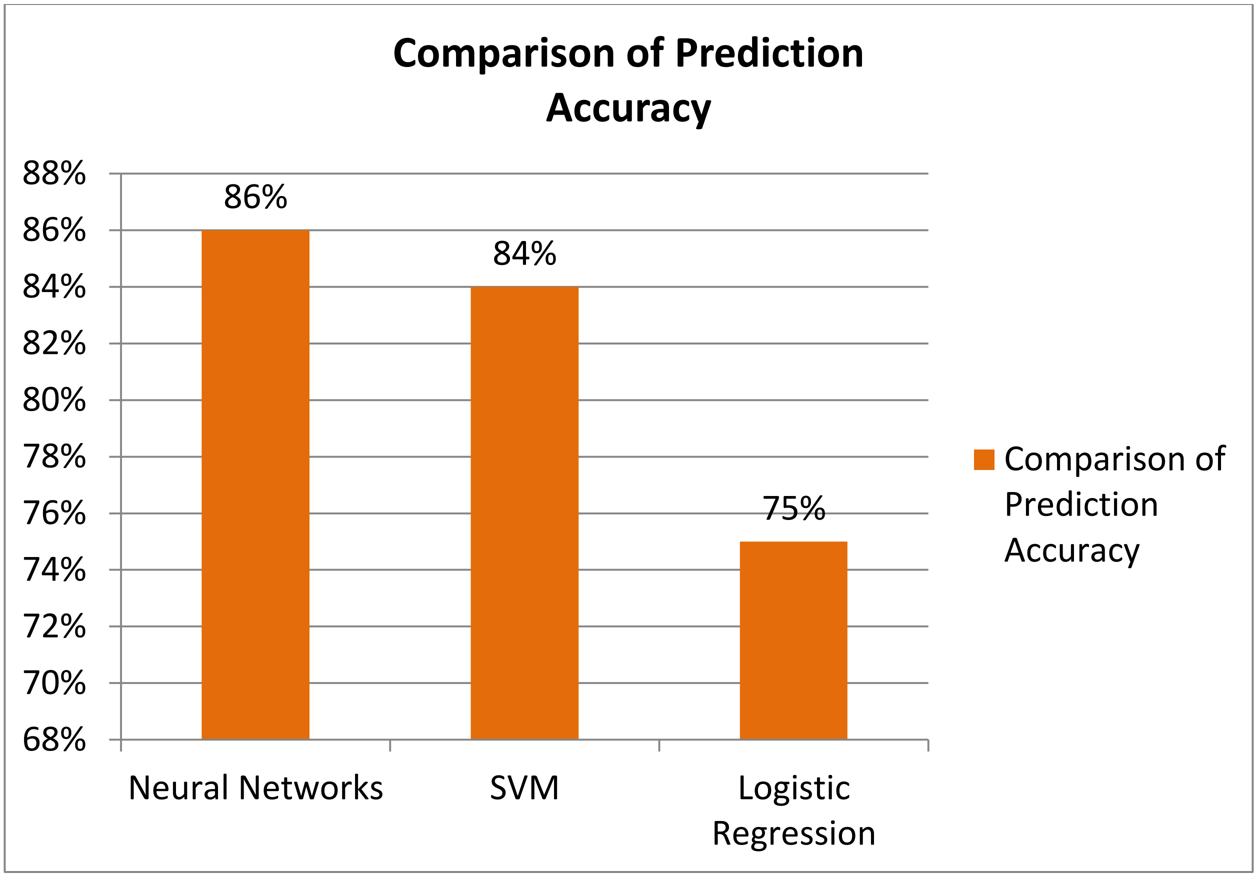

Figure 13 Comparison of Accuracy with Baseline Methods

5 CONCLUSION

Our machine learning model suggests that machine learning is a reliable technique for predicting stock price crash before it actually happens and hence be able to mitigate the effects or even prevent it from happening. The model works especially really well using the Adam optimizer for the fine tuning of gradients. The test that was conducted using normal sensitivity is also accepted globally for stock price crash prediction. However, the high sensitivity test, though a little less accurate, is an additional measure to raise yellow flag over data being received about corporate governance from a company over a week. With this technique, accuracy can further be increased by using selective features from the four categories discussed above. The best part is that no rules are required prior to training a model or even for testing as used by previous systems.

This is a really reliable and useful technique that the stakeholders and shareholders can use to secure their capital investments. It is a deployable model and it can wield useful results beneficial for companies on large scale. However, a lot of work needs to be done in this field to further improve results and to analyze the effects of different features on stock price crash prediction. For future work, assessment needs to be done on the relation of DUVOL to stock price crash prediction using this machine learning technique so that the shareholders and investors can better control their companies. And finally, the most useful and critical technique i.e. the use of LSTMs for real-time data monitoring needs to be implemented so that the accuracies can be further improved and adequate time availability for shareholders for response to issue can be increased.

6 APPENDIX A

Name	Definition
Stock Price	The stock price is the highest amount someone is willing to pay for the stock, or the lowest amount that it can be bought for.
Corporate Governance	Corporate governance is the combination of rules, processes or laws by which businesses are operated, regulated or controlled. The term encompasses the internal and external factors that affect the interests of a company's stakeholders, including shareholders, customers, suppliers, government regulators and management.
NCSKEW	It is measured as the inverse of the third central moment of firm-specific weekly return scaled by the variance of firm-specific weekly return raised to 3/2.
DUVOL	Down to Up Volatility; An up (down) week is defined as a week when the firm-specific weekly return is above (below) the annual mean.

7 REFERENCES

[1] Ruffles, P.C.; 2001, “Expanding the Horizons of Gas Turbine in Global Markets”; *ISABE 2001-1010*

[2] Dring, R. P., Joslyn, H. D., Hardin, L. W., and Wagner, J. H., 1982, "Turbine Rotor-Stator Interaction," ASME Journal of Engineering for Power, vol.104, pp.729-742

[3] Arndt, N., 1993, “ Blade Row Interaction in a Multistage Low Pressure Turbine”, ASME Journal of turbomachinery, Vol. 115, pp. 370-376

[4] Denton J.D., 1993, “ Loss Mechanisms in Turbomachines:, ASME Journal of Turbomachinery, Vol.115

[5] Denton J.D., 1993, “ Loss Mechanisms in Turbomachines:, ASME Journal of Turbomachinery, Vol.115

[6] R.E. Walraevens, A.E. Gnllus, “Stator-Rotor-Stator Interaction inAn Axial Flow Turbine And its Influence on Loss Mechanisms”, AGARD CP 571, 1995, UK, pp 39(1-13).

[7] Dawes, W.N., 1994, “ A Numerical Study of the Interaction of Transonic Compressor Rotor Over Tip Leakage Vortex with the Following Stator Blade Row”, ASME Paper No. 94-GT-156

[8] Sharma, O,P., Renaud, E., Butler, T.L., Milasps, K., Dring, R.P., and Joslyn, H.D., 1988, “ Rotor- Stator Interaction in Multi- Stage- Axial Flow Turbines”, AIAA Paper No. 88-3013

[9] Busby, J.A., Davis, R.L., Dorney, D.J., Dunn, M.G., Haldeman , C.W., Abhari, R.S., Venable, B.L., and Delany, R.A., 1999, “ Influence of Vane-Blade Spacing on Transonic Turbine Stage Aerodynamics: Part II- Time- Resolved Data and Analysis:, ASME Journal of Turbomachinery , Vol. 121, pp. 673-682

[10] Collar (1947), "The Expanding Domain of Aeroelasticity," *Journal of the Royal Aeronautical Society* 51-1.

[11] Bell Loyed, “Three dimensional Unsteady flow analysis in vibrating Turbine Cascades”, Durham University, 1999.

[12] Jeff Green, PHD Thesis, “Controlling Forced Response of a High Pressure Turbine Blade”, Royal Institute of Technology, Stockholm, 2006.

[13] Ball, R. (2009). Market and political/regulatory perspectives on the recent accounting scandals. *Journal of Accounting Research*, *47*(2), 277-323

[14] Graham, J.R., Harvey, C.R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of accounting and economics*, *40*(1), 3-73.

[15] Kothari, S.P., Shu, S., & Wysocki, P.D. (2009). Do managers withhold bad news? *Journal of Accounting Research*, *47*(1), 241-276.

[16] LaFond, R., & Watts, R.L. (2008). The information role of conservatism. *The Accounting Review*, *83*(2), 447-478

[17] Hölmstrom, B. (1979). Moral hazard and observability. *The Bell journal of economics*, *10*(1), 74-91.

[18] Benmelegh, E., Kandel, E., & Veronesi, P. (2010). Stock-based compensation and CEO (dis) incentives. Quarterly,*Journal of Economics*, 125, 1769-1820.

[19] Amihud, Y., & Lev, B. (1981). Risk reduction as a managerial motive for conglomerate mergers. *The bell journal of economics*, *12*(2), 605-617.

[20] Amihud, Y., & Lev, B. (1981). Risk reduction as a managerial motive for conglomerate mergers. *The bell journal of economics*, *12*(2), 605-617.

[21] An, H., & Zhang, T. (2013). Stock price synchronicity, crash risk, and institutional investors. *Journal of Corporate Finance*, *21*, 1-15.

[22] Healy, P.M., Hutton, A.P., & Palepu, K.G. (1999). Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary accounting research*, *16*(3), 485-520.

[23] Shleifer, A., & Vishny, R.W. (1997). A survey of corporate governance. *The Journal of Finance*, *52*(2), 737-783

[24] Fama, E.F., & Jensen, M.C. (1983). Separation of ownership and control. *Journal of law and economics*, *30*(2) 301-325.

[25] Fama, E.F., & Jensen, M.C. (1983). Separation of ownership and control. *Journal of law and economics*, *30*(2) 301-325.

[26] Masulis, R., Mobbs, S. “Are all inside directors the same? Evidence from the external directorship market.” *The Journal of Finance* 66, no. 3 (2011): 823-872.

[27] Beneish, M. “Incentives and penalties related to earnings overstatements that violate GAAP.” *Accounting Review* 74 (1999): 425-457.

[28] Healy, P. “The effect of bonus schemes on accounting decisions.” *Journal of Accounting and*

Economics 7 (1985): 85-107.

[29] Burns, N., Kedia, S. “The impact of performance-based compensation on misreporting.” *Journal of Financial Economics* 79 (2006): 35-67.

[30] Kedia, S., Philippon, T. “The economics of fraudulent Accounting.” *Review of Financial Studies* 23 (2010): 939-961

[31] Kim, J., Li, Y., Zhang, L. “CFOs versus CEOs: Equity incentives and crashes.” *JOurnal of Financial Economics* 101 (2011b): 713-730.

[32] Kim, J., Li, Y., Zhang, L. “Corporate tax avoidance and stock price crash risk: firm-level analysis.” *Journal of Financial Economics* 100 (2011a): 639-662.

[33] An, H., Zhang, T. “Stock price syncronicity, crash risk, and institutional investors.” *Journal of Corporate Finance* 21 (2013): 1-15.

[34] Andreou, P., Antoniou, C., Horton, J., Louca, C. “Corporate governance and firm-specific stock price crashes.” *Working Paper* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2029719), 2013: 1-49.

[35] Panayiotis C. Andreou, Corporate Governance and Firm-specific Stock Price Crashes, European Financial Management, Vol. 22, No. 5, 2016, 916–956 doi: 10.1111/eufm.12084

[36] Nagesh Singh Chauhan, Introduction to Artificial Neural Networks, https://www.kdnuggets.com/2019/10/introduction-artificial-neural-networks.html

[37] J. Tang, C. Deng and G. Huang, "Extreme Learning Machine for Multilayer Perceptron," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 4, pp. 809-821, April 2016, doi: 10.1109/TNNLS.2015.2424995.

[38] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: perceptron, Madaline, and backpropagation," in *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415-1442, Sept. 1990, doi: 10.1109/5.58323.

[39] Guang-Bin Huang, Qin-Yu Zhu, K. Z. Mao, Chee-Kheong Siew, P. Saratchandran and N. Sundararajan, "Can threshold networks be trained directly?," in *IEEE Transactions on Circuits*

and Systems II: Express Briefs, vol. 53, no. 3, pp. 187-191, March 2006, doi: 10.1109/TCSII.2005.857540.

[40] R. Murugadoss and M. Ramakrishnan, "Universal approximation using probabilistic neural networks with sigmoid activation functions," *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, Unnao, 2014, pp. 1-4, doi: 10.1109/ICAETR.2014.7012920.

[41] A. Wibisono, M. R. Alhamidi, A. Nurhadiyatna and W. Jatmiko, "Hyperbolic tangent activation function on FIMT-DD algorithm analysis for airline big data," *2017 International Workshop on Big Data and Information Security (IWBIS)*, Jakarta, 2017, pp. 31-36, doi: 10.1109/IWBIS.2017.8275099.

[42] Y. Ying, J. Su, P. Shan, L. Miao, X. Wang and S. Peng, "Rectified Exponential Units for Convolutional Neural Networks," in *IEEE Access*, vol. 7, pp. 101633-101640, 2019, doi: 10.1109/ACCESS.2019.2928442.

[43] I. Arora and A. Saha, "Comparison of back propagation training algorithms for software defect prediction," *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, Noida, 2016, pp. 51-58, doi: 10.1109/IC3I.2016.7917934.