**ABSTRACT**

The project involved gaining a good understanding of a Blockchain and Artificial Intelligence. Task is to predict the potential vote reward of a new post on Steem Blockchain. Below are key tasks

* Prepare a dataset by collecting information of various posts
* How to train Neural Network using the above dataset
* Predict potential vote reward for new post

**LIST OF FIGURES**

Fig 1.1 Blockchain Formation

Fig 1.2 Total rewards paid

Fig 1.3 Features of Steem Blockchain

Fig 2.1 Comparison between Artificial Intelligence, Machine and Deep Learning

Fig 2.2 Various Types of Learning

Fig 2.3 Neural Network

Fig 2.4 Decision Tree

Fig 2.5 Support Vector Machine

Fig 4.1 Sample Dataset

Fig 4.2 loss vs epochs graph

Fig 4.3 Performance of 0.5hr NN

Fig 4.4 Performance of 1hr NN

Fig 4.5 Performance of 2hr NN

Fig 4.6 Performance of 3hr NN

Fig 4.7 Performance of 4hr NN

Fig 4.8 Performance of 6hr NN

Fig 4.9 Performance Final NN

**LIST OF TABLES**

Table 4.1: Contains list of features collected for each post from Steem Blockchain

**CHAPTER 1**

**INTRODUCTION**

* 1. **Overview**

Steem is a social blockchain that grows communities and makes immediate revenue streams possible for users by rewarding them for sharing content. It’s currently the only blockchain that can power real applications via social apps like Steemit. There are content creators and content curators. Creators get reward for writing their content on blockchain while Curators get rewards for promoting Creators content.

The idea behind the project is to gather information of posts on Steem blockchain and to design a Robot to predict potential upvote reward of the post using Steem Blockchain data and Artificial Intelligence. The potential winning posts will be upvoted before they land on trending section.

* 1. **Problem Statement**

The goal is to predict the Payout of a post on Steem Blockchain at different intervals.

* + 1. Collect data of posts from Steem blockchain
    2. Preprocess the collected
    3. Train a neural network using back-propagation
    4. Predict the pay-out of new post before they land on Trending Section

Prediction should be done after:

* 15 min of post creation
* 30 min of post creation
* 1 hour of post creation
* 2 hour of post creation
* 4 hour of post creation
* 6 hour of post creation
  1. **Objective**

The Objective of the project is to find profitable post and vote them to maximize curation reward.

* 1. **What is Steem Blockchain**

First, we need to know what **blockchain** is,

A **blockchain**, originally **block chain**, is a growing list of [records](https://en.wikipedia.org/wiki/Record_(computer_science)), called *blocks*, which are linked using [cryptography](https://en.wikipedia.org/wiki/Cryptography). Each block contains a [cryptographic hash](https://en.wikipedia.org/wiki/Cryptographic_hash_function) of the previous block, a [timestamp](https://en.wikipedia.org/wiki/Trusted_timestamping), and transaction data (generally represented as a [Merkle tree](https://en.wikipedia.org/wiki/Merkle_tree) root hash).

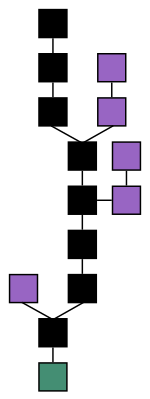
[](https://en.wikipedia.org/wiki/File:Blockchain.svg)

Fig.1.1 Blockchain formation. The main chain (black) consists of the longest series of blocks from the genesis block (green) to the current block. Orphan blocks (purple) exist outside of the main chain.

By design, a blockchain is resistant to modification of the data. It is "an open, [distributed ledger](https://en.wikipedia.org/wiki/Distributed_ledger) that can record transactions between two parties efficiently and in a verifiable and permanent way". For use as a distributed [ledger](https://en.wikipedia.org/wiki/Ledger), a blockchain is typically managed by a [peer-to-peer](https://en.wikipedia.org/wiki/Peer-to-peer) network collectively adhering to a [protocol](https://en.wikipedia.org/wiki/Protocol_(communication)) for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, which requires consensus of the network majority. Although blockchain records are not unalterable, blockchains may be considered [secure by design](https://en.wikipedia.org/wiki/Secure_by_design) and exemplify a distributed computing system with high [Byzantine fault tolerance](https://en.wikipedia.org/wiki/Byzantine_fault_tolerance). [Decentralized](https://en.wikipedia.org/wiki/Decentralized) consensus has therefore been claimed with a blockchain.

Blockchain was invented by Satoshi Nakamoto in 2008 to serve as the public transaction ledger of the cryptocurrency bitcoin. The invention of the blockchain for bitcoin made it the first digital currency to solve the double-spending problem without the need of a trusted authority or central server.

**Now we’ll know about Steem Blockchain**

**STEEM** is a social blockchain that grows communities and makes immediate revenue streams possible for users by rewarding them for sharing content. It’s currently the only blockchain that can power real applications via social apps like Steemit.

Shareholders of social media platforms made billions of dollars from user-generated content. The content creators? They made nothing.

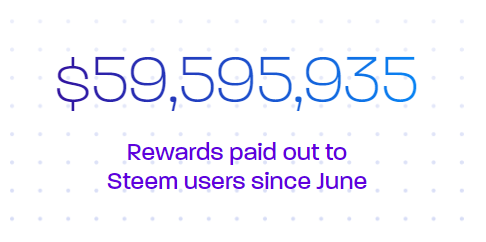
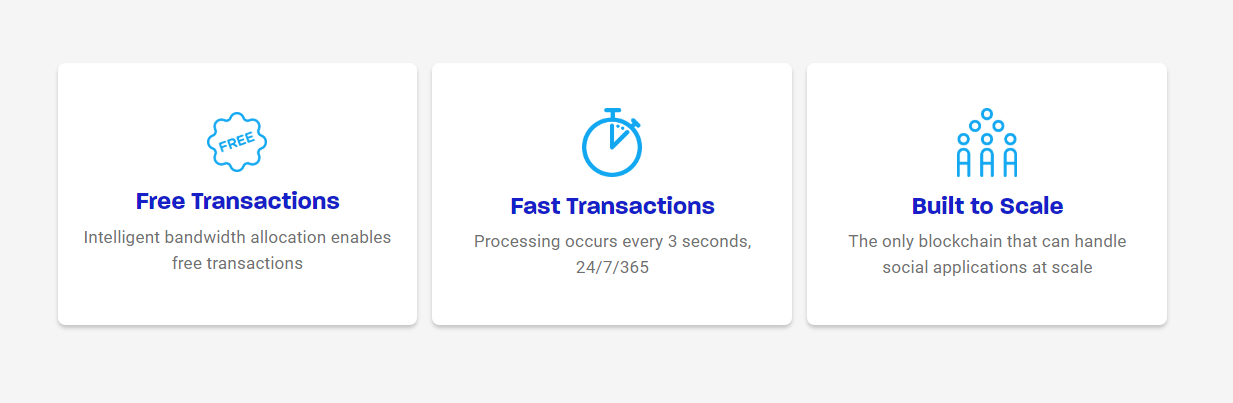


Fig 1.2 Total rewards paid

Steem flips the model and returns the value to the people who contribute the most. Users become platform stakeholders, maintaining control over their data, and earning cryptocurrency rewards for each contribution they make.

Unlike most blockchains that are too slow and expensive to be used for apps, Steem is fast, free, and scalable.



**Fig 1.3 Features of Steem Blockchain**

In the field of crypto-currencies, the unique properties of STEEM make it both “smart” and “social” compared to others, such as bitcoin and ether. This stems from two new token features. The first is a pool of tokens dedicated to incentivizing content creation and curation (called the “rewards pool”). The second is a voting system that leverages the wisdom of the crowd to assess the value of content and distribute tokens to it. These two unique properties when combined are referred to as Proof-of-Brain, which is an entendre based on Proof-of-Work, meant to emphasize the human work required to distribute tokens to community participants.

The rate that new tokens are generated was set to 9.5% per year starting in December 2016, and decreases at a rate of 0.01% every 250,000 blocks, or about 0.5% per year. The inflation will continue decreasing at this rate until it reaches 0.95%, after a period of approximately 20.5 years.

Of the supply of new tokens created by the Steem blockchain every year, 75% of those tokens compose the “rewards pool” which are distributed to content creators and content curators. 15% are distributed to vested token holders, and 10% are distributed to Witnesses, the block producers cooperating inside Steem’s DPoS consensus protocol.

* **STEEM.**STEEM is the foundational cryptocurrency of the Steemit network. STEEM can be exchanged for bitcoin or other cryptocurrencies via several prominent exchanges. In order to cast votes, however, you need to turn your STEEM into Steem Power (SP). This process is known as “powering up.” You can also convert STEEM into Steem Dollars.
* **Steem Power (SP).** When you turn STEEM into SP, you are investing it in the network as equity. One unit of STEEM, vested as SP, equates to one vote. Those with the most SP to contribute have the greatest influence over which content is elevated to the top. Upvotes and downvotes, or “flags,” from users holding large amounts of SP are worth more than users with less SP. Additionally, around 90% of new STEEM that is created goes to users who hold a lot of SP, which creates an incentive for participants to invest their earnings back into the network. Rate limits are in place on converting SP back into STEEM, known as “powering down.”
* **Steem Dollars.** Steem Dollars, valued at 1:1 with the US dollar, are a unit used to represent a short-term debt. Holding Steem Dollars is described in Steem’s whitepaper as a way of lending the community the value of one US dollar, which is designed to foster growth.

**How Steem is different from Bitcoin?**

Both Steem and bitcoin were designed with a particular function in mind. Bitcoin was created to serve as a digital, peer-to-peer currency. Steem is, at its core, a social media platform. Many social media platforms use voting systems, such as upvotes, shares, or “likes,” to prioritize content. Steem incorporates these concepts into a unique, blockchain-based platform that pays users in STEEM for their contributions to the network. While STEEM and bitcoin are both cryptocurrencies, STEEM’s usage is specific to social networks.

**Bitcoin** is based on **Proof of Work**. In Proof of Work (PoW) mechanism, computers in the network (miners) work to solve a cryptographic puzzle repetitively consisting of the mathematical function (hash). Take Bitcoin as an example, the first computer or network of computers (i.e., a mining pool) that found a hash with specific properties are rewarded with few Bitcoins roughly every 10 minutes.

**Steem** is based on **Delegated Proof of Stake.** Proof of Stake (PoS) doesn’t require computers to perform repetitive computations thus being more environmentally friendly. PoS replace miners with validators where they will lock up some of their coins as stake (deposit). Delegated Proof-of-Stake (DPoS) is a more efficient PoS mechanism. DPoS uses a reputation system and real-time voting to achieve consensus. Community members vote for super representatives to secure their network and super representatives will be rewarded by validating transactions for the next block.

**CHAPTER 2**

**TOOLS AND TECHNOLOGY USED**

This project is developed using data collected from Steem Blockchain. Artificial intelligence and its Regression models were used to predict the potential payout of new post created on Steem Blockchain. The main Programming Language used is Python because of its good community support in the field of Machine Learning.

* 1. **Artificial Intelligence**

Artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.

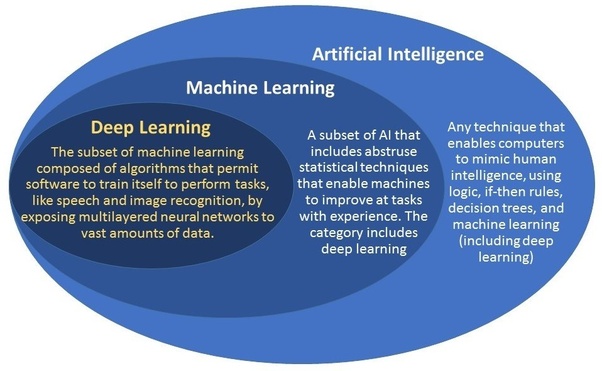


Fig 2.1 Comparison between Artificial Intelligence, Machine and Deep Learning

**Machine Learning**

Machine learning (ML) is the study of algorithms and mathematical models that computer systems use to progressively improve their performance on a specific task. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task.

**Deep Learning**

Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised.

Deep learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases superior to human experts.

**Types of learning algorithms**

The types of machine learning algorithms differ in their approach, the type of data they input and output, and the type of task or problem that they are intended to solve.

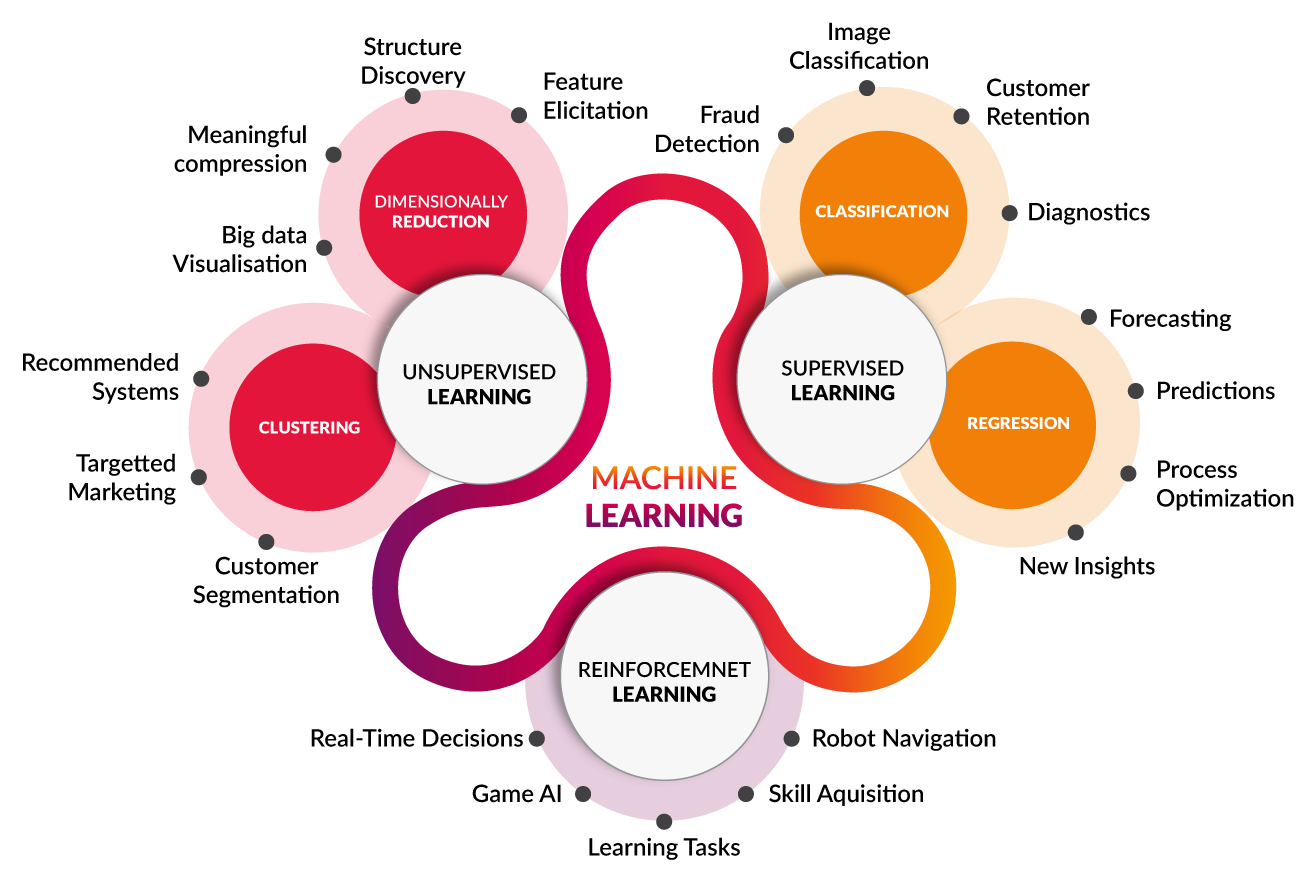


Fig 2.2 Various Types of Learning

* **Supervised and semi-supervised learning**

Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs. The data is known as training data, and consists of a set of training examples. Each training example has one or more inputs and a desired output, also known as a supervisory signal.

* **Unsupervised learning**

Unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms therefore learn from test data that has not been labelled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data.

* **Reinforcement learning**

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Due to its generality, the field is studied in many other disciplines, such as game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, statistics and genetic algorithms.

* 1. **Machine Learning Models**

1. **Artificial neural networks**

Artificial neural networks (ANNs), or connectionist systems, are computing systems vaguely inspired by the biological neural networks that constitute animal brains.[47] The neural network itself is not an algorithm, but rather a framework for many different machine learning algorithms to work together and process complex data inputs.[48] Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules.

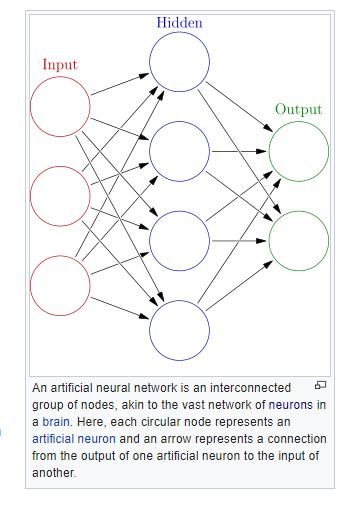
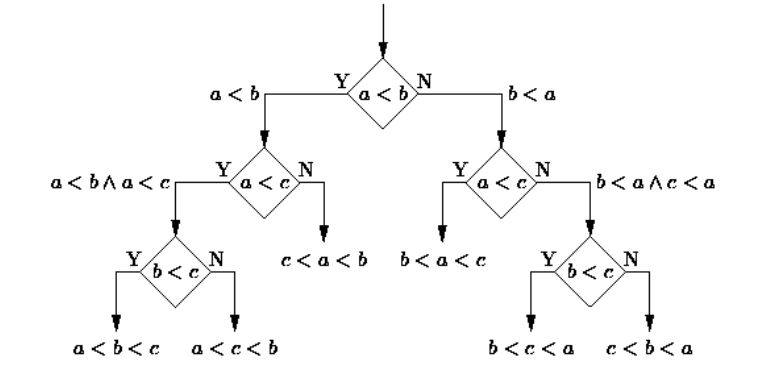


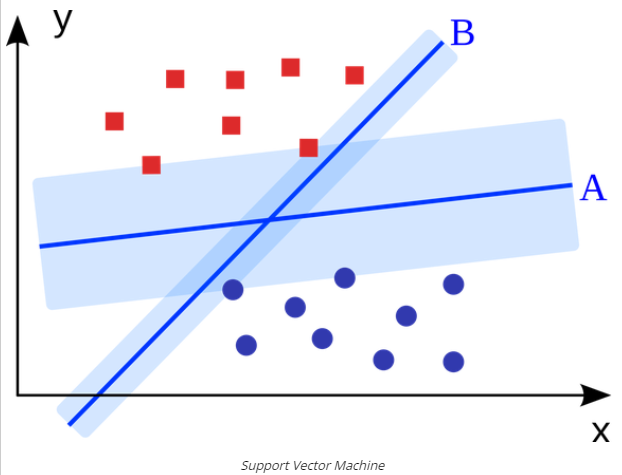
Fig 2.3 Neural Network

1. **Decision Trees:** A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance-event outcomes, resource costs, and utility. Take a look at the image to get a sense of how it looks like.



**Fig 2.4 Decision Tree**

1. **Support Vector Machines:** SVM is binary classification algorithm. Given a set of points of 2 types in N dimensional place, SVM generates a (N — 1) dimensional hyperplane to separate those points into 2 groups. Say you have some points of 2 types in a paper which are linearly separable. SVM will find a straight line which separates those points into 2 types and situated as far as possible from all those points.



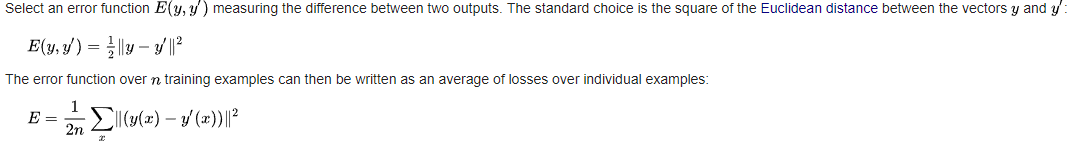
**Fig 2.5 Support Vector Machine**

* 1. **Backpropagation**

Backpropagation is a method used in artificial neural networks to calculate a gradient that is needed in the calculation of the weights to be used in the network.[1] Backpropagation is shorthand for "the backward propagation of errors," since an error is computed at the output and distributed backwards throughout the network’s layers.[2] It is commonly used to train deep neural networks,[3] a term referring to neural networks with more than one hidden layer.

Backpropagation requires the derivative of the loss function with respect to the network output to be known, which typically (but not necessarily) means that the desired target value is known. For this reason, it is considered to be a supervised learning method, although it is used in some unsupervised networks such as autoencoders. Backpropagation is also a generalization of the delta rule to multi-layered feedforward networks, made possible by using the chain rule to iteratively compute gradients for each layer. It is closely related to the Gauss–Newton algorithm and is part of continuing research in neural backpropagation.

**Loss Function**



**Optimization Algorithm**

The optimization algorithm repeats a two-phase cycle, propagation and weight update. When an input vector is presented to the network, it is propagated forward through the network, layer by layer, until it reaches the output layer. The output of the network is then compared to the desired output, using a loss function. The resulting error value is calculated for each of the neurons in the output layer. The error values are then propagated from the output back through the network, until each neuron has an associated error value that reflects its contribution to the original output.

* 1. **Python Libraries Used:**

1. **Steem**: Python library to interact with Steem blockchain.
2. **csv**: Python library for storing data in .csv file
3. **pandas**: For data manipulation and interaction.
4. **NumPy**: Fast and efficient operations on array
5. **matplotlib**: Used for data visualization
6. **keras**: Creating, training and testing neural network models
7. **sklearn**: For pre-processing the dataset
8. **ast**: Convert string into dictionary data structure
9. **math**: General mathematical functions
10. **datetime**: Handles different formats of date and time
    1. **Platform and IDE**

Forthis project Windows platform and PyCharm IDE by JetBrains was used.

**CHAPTER 3**

**HISTORY: IS ARITIFICIAL INTELLIGENCE NEW?**

**Early Days**

During the Second World War, noted British computer scientist Alan Turing created the Enigma Machine, which could decipher German code. The Enigma Machine laid the foundations for Machine Learning. According to Turing, who came up with the idea in 1950, a machine that could converse with humans without the humans knowing that it is a machine would win the “imitation game” and could be said to be “intelligent”.

In 1956, American computer scientist John McCarthy organised the Dartmouth Conference, at which the term ‘Artificial Intelligence’ was first adopted. Research centres popped up across the United States to explore the potential of AI. Researchers Allen Newell and Herbert Simon were instrumental in promoting AI as a field of computer science that could transform the world.

**Getting Serious About AI Research**

In 1951, a machine known as Ferranti Mark 1 successfully used an algorithm to master checkers. Subsequently, Newell and Simon developed General Problem Solver algorithm to solve mathematical problems. Also, in the 50s John McCarthy, often known as the father of AI, developed the LISP programming language which became important in machine learning.

In the 1960s, researchers emphasized developing algorithms to solve mathematical problems and geometrical theorems. In the late 1960s, computer scientists worked on Machine Vision Learning and developing machine learning in robots. WABOT-1, the first ‘intelligent’ humanoid robot, was built in Japan in 1972

**AI Winters**

However, despite this well-funded global effort over several decades, computer scientists found it incredibly difficult to create intelligence in machines. To be successful, AI applications (such as vision learning) required the processing of enormous amount of data. Computers were not well-developed enough to process such a large magnitude of data. Governments and corporations were losing faith in AI.

**New Millennium, New Opportunities**

Exponential gains in computer processing power and storage ability allowed companies to store vast, and crunch, vast quantities of data for the first time. In the past 15 years, Amazon, Google, Baidu, and others leveraged machine learning to their huge commercial advantage. Other than processing user data to understand consumer behaviour, these companies have continued to work on computer vision, natural language processing, and a whole host of other AI applications. Machine learning is now embedded in many of the online services we use. As a result, today, the technology sector drives the American stock market.

**CHAPTER 4**

**WORK DONE: STEEM PAY-OUT BOT**

The idea behind the project is to gather information of posts on Steem blockchain and to design a Robot to predict potential upvote reward of the post using Steem Blockchain and Artificial Intelligence. The profitable posts will be upvoted before they land on trending section.

The project is divided into two parts:

1. Dataset generation
2. Prediction of Pay-out of new posts.
   1. **Dataset Generation:**

Dataset comprises of data of old posts on Steem Blockchain in Trending Section. Each post can be uniquely identified by its **author** and **permlink.** So, a post is picked from trending section, its author and permlink is stored in data variable.

This is passed as parameter to **Steem API** to get information of the post stored in blockchain

**post = s.steemd.get\_content(post['author'], post['permlink'])**

Information about is obtained using:

**author\_data = s.steemd.get\_account(post['author'])**

|  |  |  |  |
| --- | --- | --- | --- |
| Tag Score | Replies | Followers | No of Posts |
| Author Reputation | Promotion | SBD author has | Liquid Steem |
| Steem Power | Reblogged | Vote Speed | Comment Speed |
| Pay-out Speed |  |  |  |

Table 4.1: Contains list of features collected for each post from Steem Blockchain

The Dataset contains data of more than 30,000 posts.

As Steem API can get only 100 posts in one instruction, the above instruction can be placed in a loop to get more data.

**Csv library** is used to store dataset in the .csv file.

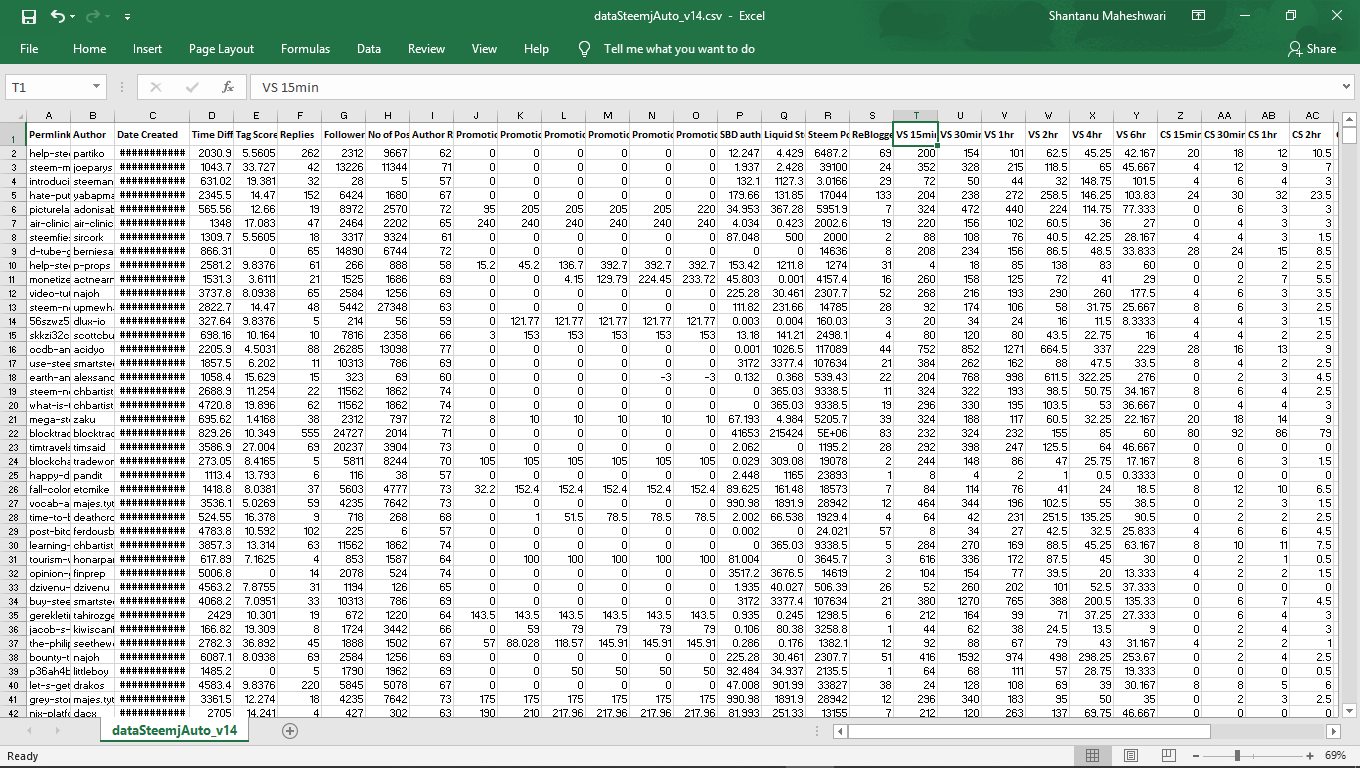


Fig 4.1 Sample Dataset

* 1. **Training and Prediction using Artificial Neural Network**

This stage is divided into 3 parts:

* + 1. Data Preprocessing
    2. Training of ANN
    3. Prediction of Payout using trained ANN

**Data Preprocessing:**

**Sklearn** python library is used. Missing data is adjusted and incorrect data is removed. Remaining data is scaled using Minmax Scalar (). Data is split randomly into 2 parts; Training set comprises of 80% and rest 20% for Test set.

**Pandas** library is used to load dataset from .csv file. Its also used for operations on dataset.

X\_train, X\_test, y\_train, y\_test = ttsplit (X, y, test\_size = 0.2, random\_state = 0) y\_train = y\_train.reshape(len(y\_train),1) y\_test = y\_test.reshape(len(y\_test),1)

**Training of Neural Network:**

**Keras** library is used for creating NN models, and training them.

model = Sequential ()

model.add(Dense(units=7, input\_dim=13, activation='relu', kernel\_initializer='normal'))

model.add(Dense(units=1, activation='linear', kernel\_initializer='normal'))

model.summary()

model.compile(loss='mse', optimizer='adamax', metrics=['mse','mae'])

regressor = model.fit(X\_train\_scale, y\_train, epochs=1500, verbose=2, batch\_size=10)

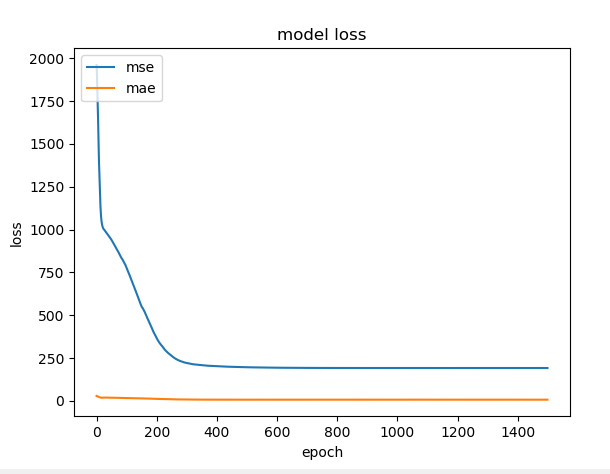


Fig 4.2 loss vs epochs graph

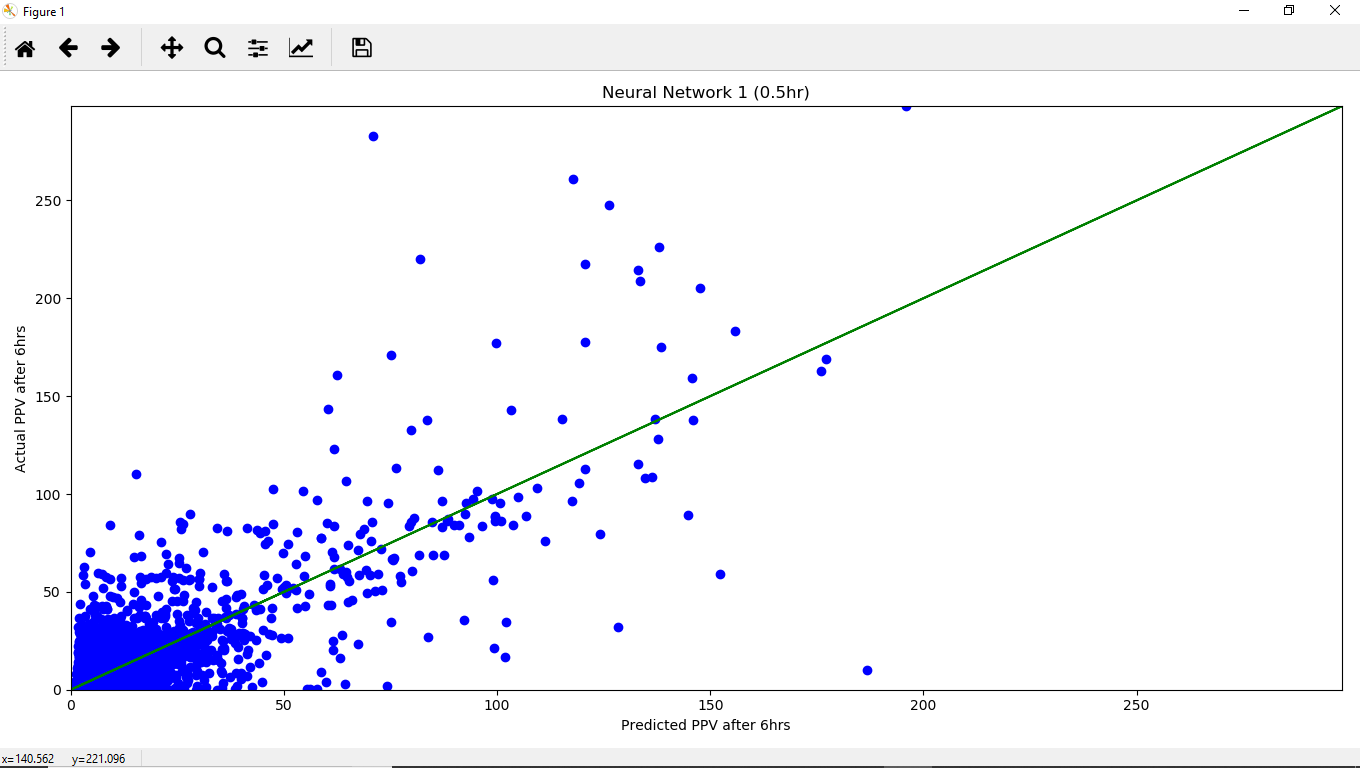
**Prediction of pay-out against Test set:**

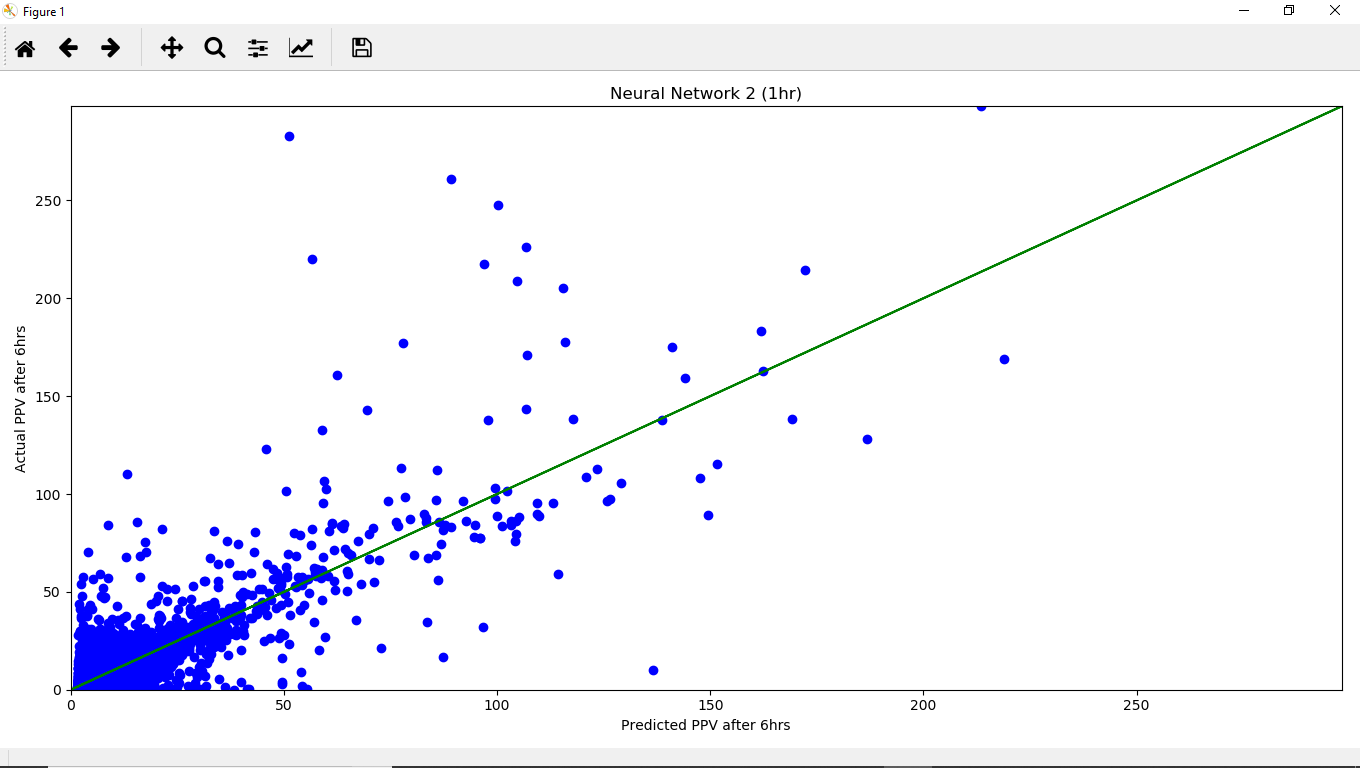
y\_pred = model.predict(X\_test\_scale)

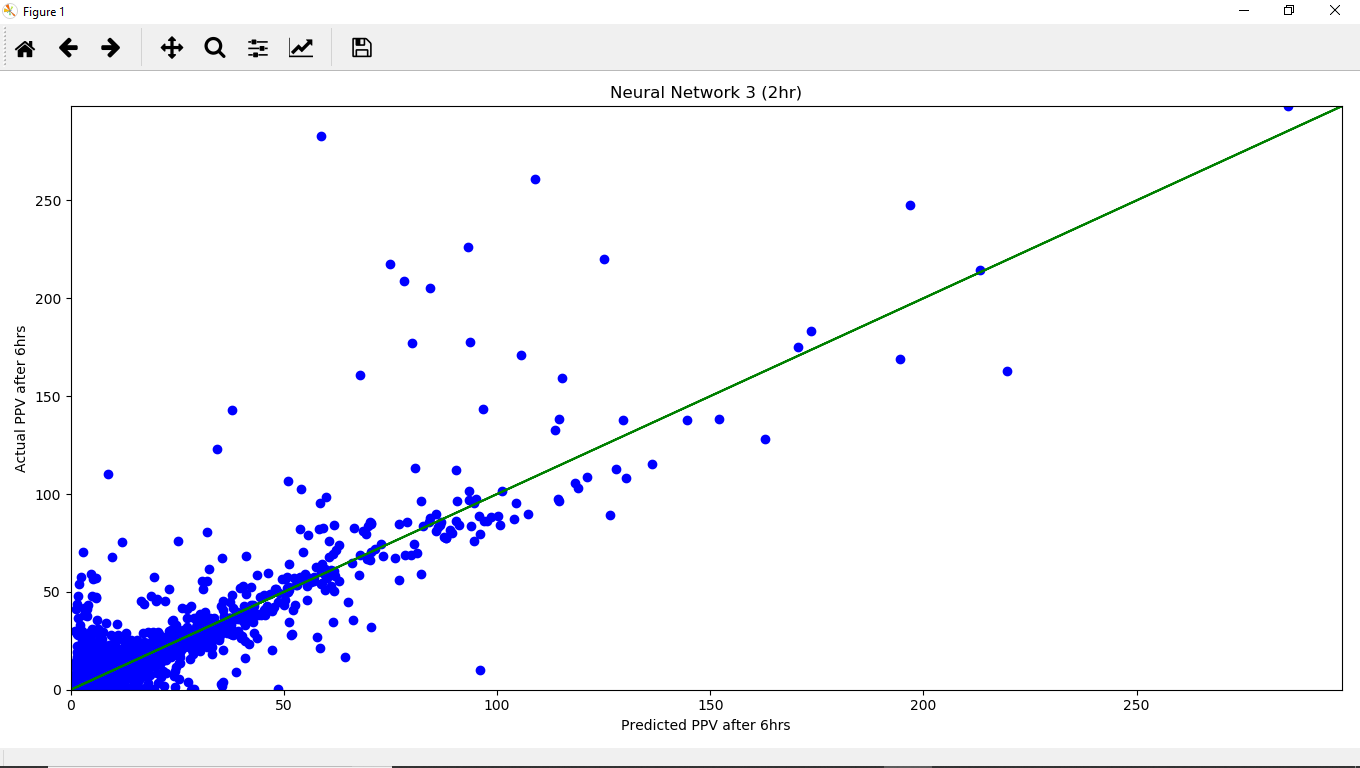
errorNN = abs(y\_test-y\_pred)

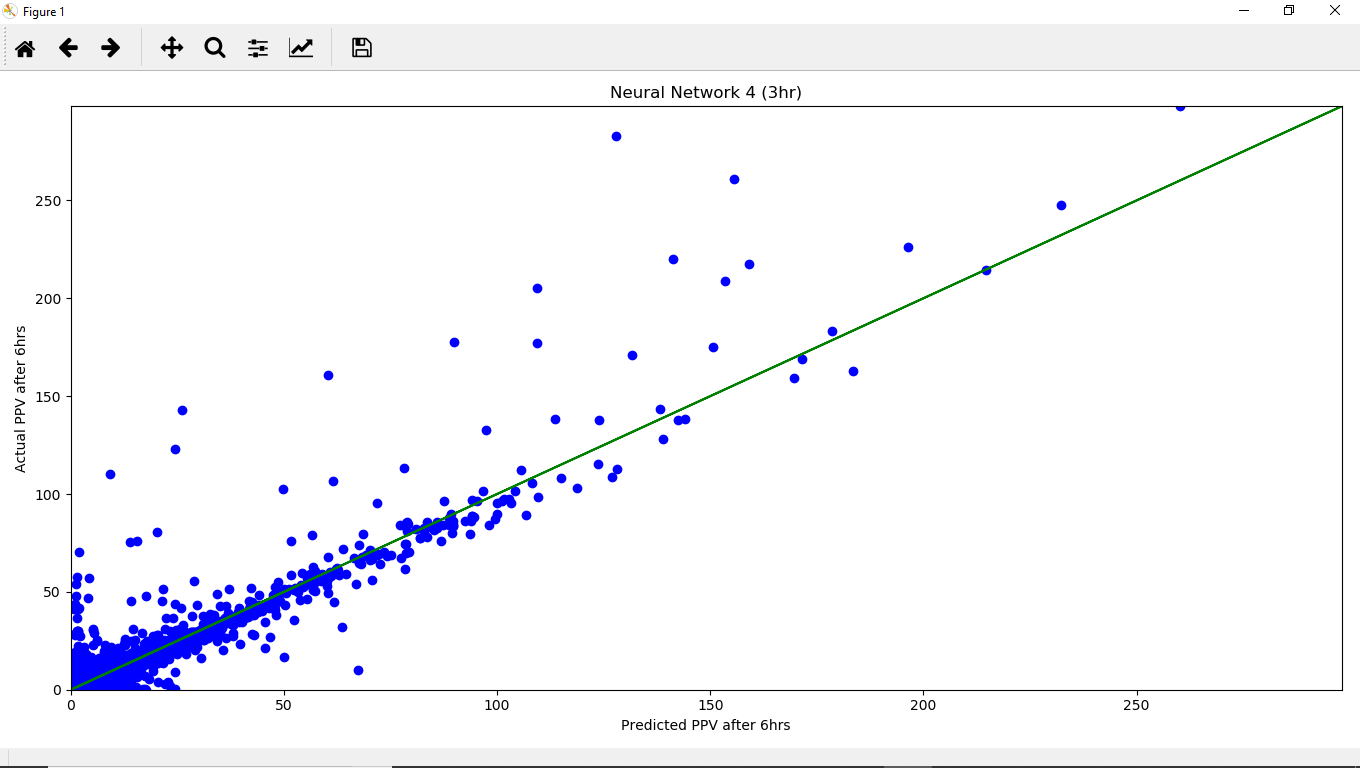
print(sum(errorNN)/len(y\_test))

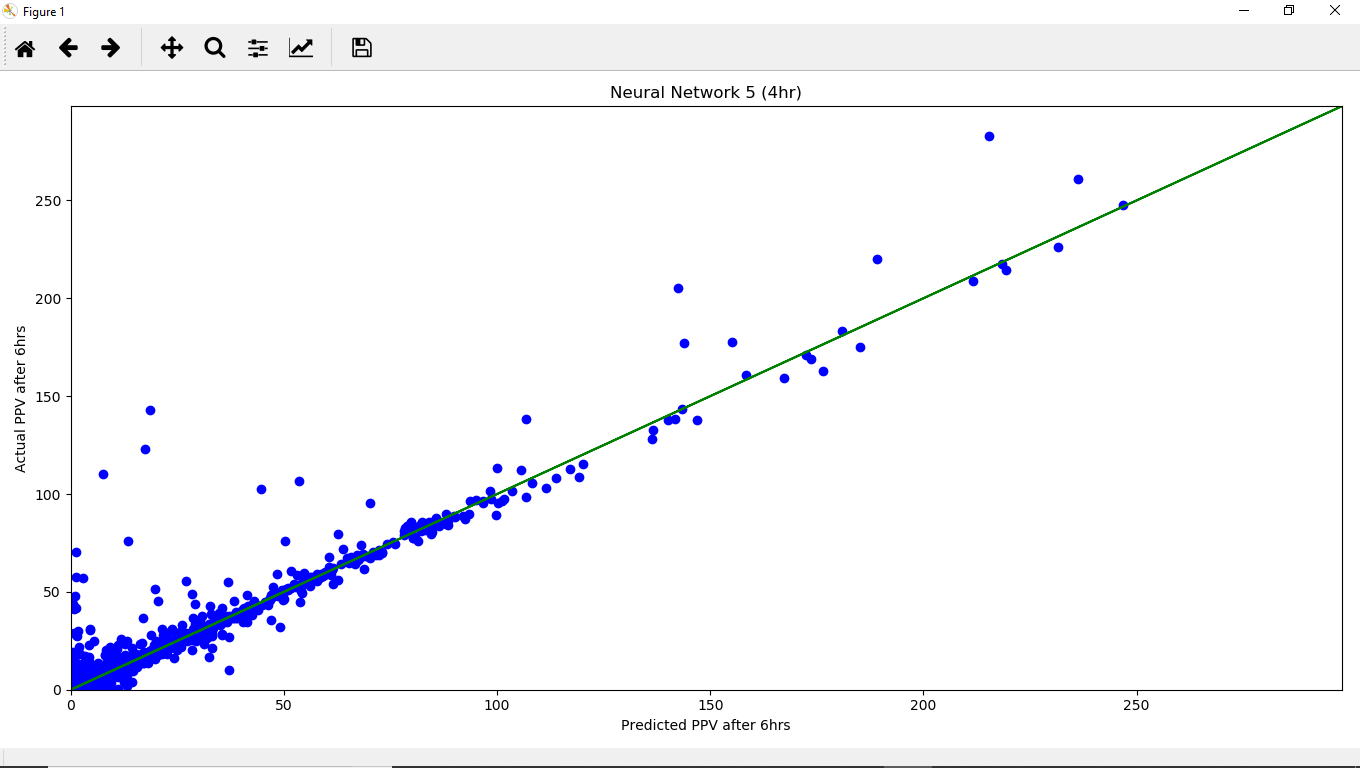
* 1. **Performance of Neural Network**

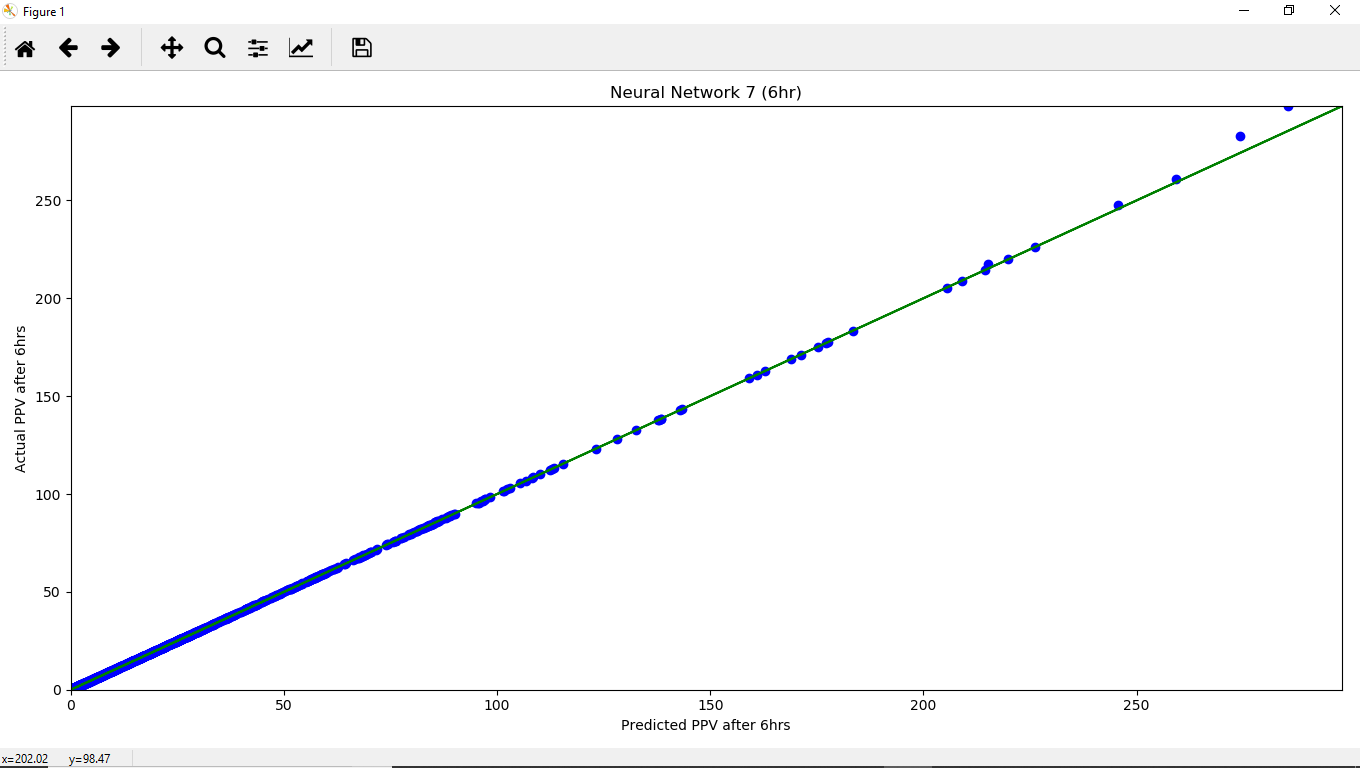
**Fig 4.3 Performance of 0.5hr NN**

**Fig 4.4 Performance of 1hr NN**

**Fig 4.5 Performance of 2hr NN**

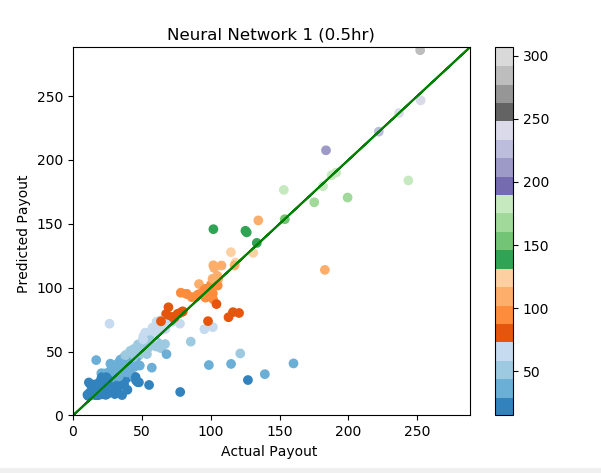
**Fig 4.6 Performance of 3hr NN**

**Fig 4.7 Performance of 4hr NN**

****

**Fig 4.8 Performance of 6hr NN**

**FINAL OUTPUT**



**Fig 4.9 Performance of Final NN**

**CHAPTER 5**

**CONLUSION AND FUTURE SCOPE**

The above ANN trained performed well in predicting payout of new posts on Steem Blockchain.

The Whole Project was divided into 3 Phases:

* + 1. Prediction using Statistical Data
    2. Prediction using Post Content (Natural Language Processing)
    3. Prediction using Media embedded in the post.

The ANN gave error of about $5 when tested against Test Set. Its work is to predict payout after 6hours of post creation using only first half an hour data.

This project can be used in future to continuously search for good post on blockchain and automatically vote them.

New content creators can also pay you to vote for their posts.

Blockchain is the future of many businesses but not all.

The key value proposition of blockchain is to provide transparency and efficiency. Businesses need to do their own homework to access whether or not they should adopt blockchain, when to adopt and what division to adopt. For example, a local restaurant who is looking for technology to improve table turnover is better off to use some existing non blockchain applications. There is no value added benefit here from a blockchain solution. However, if you are doing business with other parties all across the globe and want to know the credential of a new trading partner before signing a contract, you should probably look for some blockchain solutions. There is no way a central body can build a global registry / KYC platform / reputation book for all businesses across the globe. But decentralized blockchain solutions can. It heads back the control, data, and decision making to value creaters.

Many of the businesses can improve efficiency by adopting blockchain, especially in finance, supply chain, healthcare and gaming. But there is no one size fits all answer to blockchain adoption. We have seen mainstream players (e.g. IBM in hyperledger & JPM's Quorum) tapped blockchain/DLT and the positive outcome from these projects.

**REFERENCES**

1. Steem White Paper (Research Paper)
2. Steem Blue Paper (Research Paper)
3. Stack overflow: https://stackoverflow.com
4. <https://medium.com/tron-foundation/an-easy-to-understand-guide-to-pow-pos-dpos-consensus-mechanism-and-super-representative-eb1f5504a8e>
5. <https://steemit.com/steemit/@steemitguide/steemit-financial-instruments-explained-steem-steem-power-and-steem-dollar>
6. https://www.wikipedia.org/