**Company Turnover Prediction**

**1.Introduction**

In [computer science](https://en.wikipedia.org/wiki/Computer_science), **artificial intelligence** (**AI**), sometimes called **machine intelligence**, is [intelligence](https://en.wikipedia.org/wiki/Intelligence) demonstrated by [machines](https://en.wikipedia.org/wiki/Machine), in contrast to the **natural intelligence** displayed by humans.

The term "artificial intelligence" is used to describe machines that mimic "cognitive" functions that humans associate with other [human minds](https://en.wikipedia.org/wiki/Human_mind), such as "learning" and "problem solving".

Modern machine capabilities are classified as AI include successfully [understanding human speech](https://en.wikipedia.org/wiki/Natural_language_understanding), competing at the highest level in [strategic game](https://en.wikipedia.org/wiki/Strategic_game) systems,[autonomously operating cars](https://en.wikipedia.org/wiki/Autonomous_car), intelligent routing in [content delivery networks](https://en.wikipedia.org/wiki/Content_delivery_network), and [military simulations](https://en.wikipedia.org/wiki/Military_simulations).

**1.1 Philosophy of AI**

AI has close connections with  [philosophy](https://en.wikipedia.org/wiki/Philosophy)  because both share several concepts and these include intelligence,action, [consciousness](https://en.wikipedia.org/wiki/Consciousness), [epistemology](https://en.wikipedia.org/wiki/Epistemology), and even [free will](https://en.wikipedia.org/wiki/Free_will).Furthermore, technology is concerned with the creation of artificial animals or artificial people (or, at least, artificial creatures) so the discipline is of considerable interest to philosophers.

These factors contributed to the emergence of the **philosophy of artificial intelligence.** Some scholars argue that the AI community's dismissal of philosophy is detrimental.

**1.2 Goals of AI**

Modern artificial intelligence techniques are pervasive and are too numerous to list here. Frequently, when a technique reaches mainstream use, it is no longer considered artificial intelligence; this phenomenon is described as the [AI effect](https://en.wikipedia.org/wiki/AI_effect).

HealthCare - Artificial intelligence is breaking into the healthcare industry by assisting doctors. Microsoft has developed AI to help doctors find the right treatments for cancer.

Military-Military drones capable of autonomous action are widely considered a useful asset.

**1.3 What Contributes to AI?**

Artificial Intelligence is no doubt the next generation technology everyone is looking forward to. China who wants to be a world leader in Artificial Intelligence has added AI in the school curriculum of high-school students. Now, you can imagine the importance of AI in the coming future.

Next generation computing architecture - Machine Learning and Artificial Intelligence require speedy processors. The traditional microprocessors and CPUs are not meant for Machine Learning. They require a new breed of processors. This made the rise of GPU. The Graphics Processing Units which used to be the part of high-end gaming PC’s and workstations have seen a tremendous growth after the evolution of AI.

Availability of cloud - Data Scientists require large data sets and historic data for ML models to predict with more accuracy. In order to predict the weather or solve problems like detecting cancers, data scientists need a large amount of data to analyze. The efficiency of ML models depends merely on the size of data. The more the data, the better is the accuracy.

Advanced DeepLearning - Artificial Neural Network is replacing the conventional Machine Learning system. The advancement in Machine Learning Models with the new technologies like Capsule Neural Network and Transfer Learning have changes the way ML models are trained and implemented. These techniques are so advanced that they are able to produce the most accurate results with a limited data.

**2.Python Programming**

Python is an [interpreted](https://en.wikipedia.org/wiki/Interpreted_language), [high-level](https://en.wikipedia.org/wiki/High-level_programming_language), [general-purpose](https://en.wikipedia.org/wiki/General-purpose_programming_language) [programming language](https://en.wikipedia.org/wiki/Programming_language). Created by [Guido van Rossum](https://en.wikipedia.org/wiki/Guido_van_Rossum) and first released in 1991.

Python's design philosophy emphasizes [code readability](https://en.wikipedia.org/wiki/Code_readability) with its notable use of [significant whitespace](https://en.wikipedia.org/wiki/Off-side_rule). Its language constructs and [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming) approach aims to help programmers write clear, logical code for small and large-scale projects.

Python is a [multi-paradigm programming language](https://en.wikipedia.org/wiki/Multi-paradigm_programming_language). [Object-oriented programming](https://en.wikipedia.org/wiki/Object-oriented_programming) and [structured programming](https://en.wikipedia.org/wiki/Structured_programming) are fully supported, and many of its features support [functional programming](https://en.wikipedia.org/wiki/Functional_programming) and [aspect-oriented programming](https://en.wikipedia.org/wiki/Aspect-oriented_programming) .

**2.1 Objectives of research**

The accounting concept that calculates how quickly a business conducts its operations .Most often,the turnover of a company is used to understand how quickly a company collects cash from accounts receivable or how fast the company sells its inventory and to predict whether the company’s turnover reaches the target or not.

Purpose of predicting a company turnover,as it is precisely impossible to predict annual revenues precisely,particularly for a new business but it is critically important for companies to create high quality revenue budgets.So, to maximize the odds of being in the right ballpark relative to actual results ,we stick to a few key fundamentals.

**2.2 Problem Statement**

Given a group of rows and columns, which gives us the information about the company on some key fundamentals.Each column defines a key and gives the values .

As the values may be sometimes be positive,negative or even zero, and these values may be responsible for the turnover of a company and to ensure that these values may not affect the turnover of a company so drastically.

The task was thus to build a model which could make prediction to classify the turnover of a company for a particular data is true or false with all its key fundamentals.

**3. Review of Literature**

The topic is to build a model having different layers like input, hidden ,output.

Solution is based on the following 3 parts:

1. Pre-processing
2. Models
3. Prediction

**3.1 Pre-processing**

Importing Libraries

We are going to use most essential libraries for data pre-processing:

[**numpy**](http://www.numpy.org/) : numpy is a library which supports mutli-dimensional arrays in Python.

[**pandas**](https://pandas.pydata.org/pandas-docs/stable/): pandas provides expressive data structures for relational data.

Trained our models on 3 datasets with different pre-processing:

* original dataset : the original dataset with the unwanted information is also viewed for understanding.
* original dataset without the unwanted data: We drop the columns of unwanted data by using the “drop” command on the dataset and by setting the axis value.
* The original dataset is sliced into the dependent and independent variables along with the data,and makes us clear what the values of the variables affect the turnover of a company.

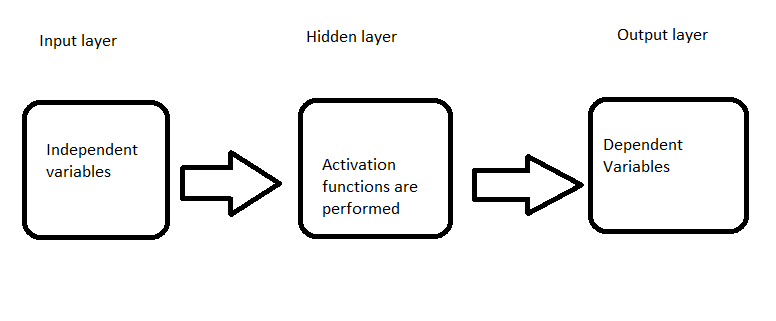
**3.2 Models**

In this , we create a model by importing “sequential” from keras.models.

And we create the layers that are input layer,hidden layers and output layer.

And for the creation of layers we import “dense” from keras.layers.As we know we will have one input layer and many number of hidden layers and one output layer.And according the dataset taken we take the input dimension as 11 as we have 11 independent variables and we have given four hidden layers to get the prediction accurately.

And or the compilation we use an “adam” optimizer.And to calculate the error we use “mean square error”.In the input layer all the data required on which functions are performed to get the exact turnover prediction of a company.In the hidden layers we perform the activation functions .And in the output layer we get the output .



**Input layer**:

The purpose of the input layer is to receive as input the values of the explanatory attributes for each observation. Usually, the number of input nodes in an input layer is equal to the number of explanatory variables. ‘Input layer’ presents the patterns to the network, which communicates to one or more ‘hidden layers’.

The nodes of the input layer are passive, meaning they do not change the data. They receive a single value on their input and duplicate the value to their many outputs. From the input layer, it duplicates each value and sent to all the hidden nodes.

**Hidden layer**:

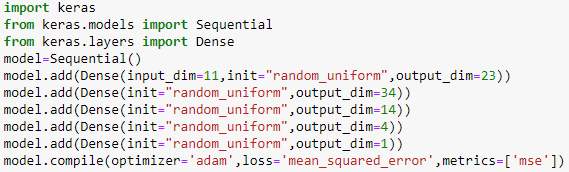
The Hidden layers apply given transformations to the input values inside the network. In this, incoming arcs that go from other hidden nodes or from input nodes connected to each node. It connects with outgoing arcs to output nodes or to other hidden nodes. In hidden layer, the actual processing is done via a system of weighted ‘connections’. There may be one or more hidden layers. The values entering a hidden node multiplied by weights, a set of predetermined numbers stored in the program. The weighted inputs are then added to produce a single number.

**Output layer**:

The hidden layers then link to an ‘output layer‘. Output layer receives connections from hidden layers or from input layer. It returns an output value that corresponds to the prediction of the response variable. In classification problems, there is usually only one output node. The active nodes of the output layer combine and change the data to produce the output values.

The ability of the neural network to provide useful data manipulation lies in the proper selection of the weights. This is different from conventional information processing.

The creation of Models and addition of layers can be shown in the following code that is used .

****

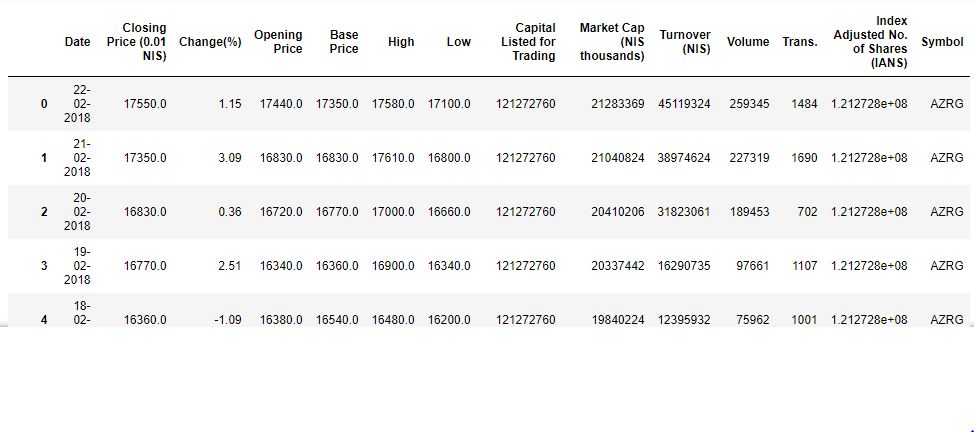
**3.3 Data Collection**

The dataset used for predicting the turnover of a company is the set of a company called “top-35-israeli-stocks-5-years-data”. And it defines the turnover of that company day by day.And the dataset consists of 1000 rows and 14

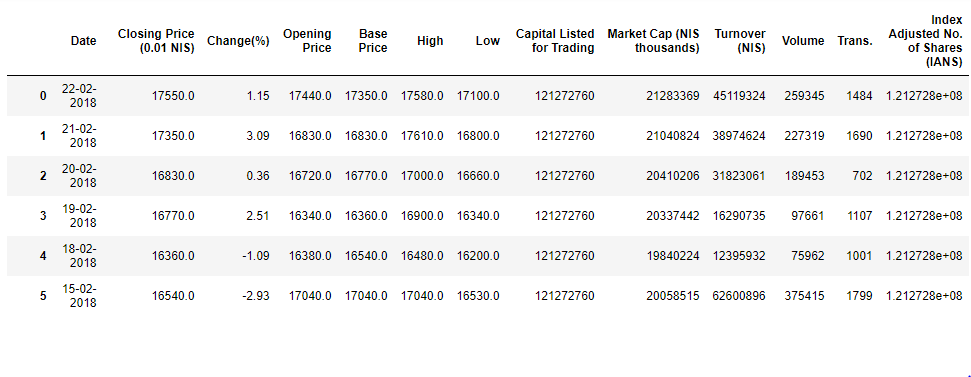
columns.

Data: The dataset we are using for predicting the turnover of a company is taken from Kaggle competition which can be found at Kaggle. Dataset has a large number of comments from Wikipedia talk page edits. They have been labeled by human raters for toxic behavior.

And the dataset can be viewed as follows

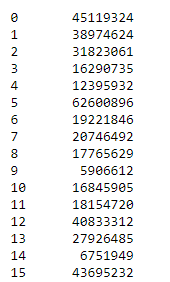


And the dataset after removing the unwanted columns look like

****

And the above all parameters defined as the independent variables and these are belonged to the variable “x”.

And the dependent variable “y” which gives the value of turnover of a company.And can be seen as

****

**3.4 Data Modelling**

In this data modelling we define the algorithm that is used for prediction and “ANN” is the algorithm used in this company turnover prediction.

**ANN ALGORITHM:**

It intended to simulate the behavior of biological systems composed of “neurons”. ANNs are computational models inspired by an animal’s central nervous systems. It is capable of machine learning as well as pattern recognition. These presented as systems of interconnected “neurons” which can compute values from inputs.

A neural network is an oriented graph. It consists of nodes which in the biological analogy represent neurons, connected by arcs. It corresponds to dendrites and synapses. Each arc associated with a weight while at each node. Apply the values received as input by the node and define Activation function along the incoming arcs, adjusted by the weights of the arcs.

A neural network is a machine learning algorithm based on the model of a human neuron. The human brain consists of millions of neurons. It sends and process signals in the form of electrical and chemical signals. These neurons are connected with a special structure known as synapses. Synapses allow neurons to pass signals. From large numbers of simulated neurons neural networks forms.

An Artificial Neural Network is an information processing technique. It works like the way the human brain processes information. ANN includes a large number of connected processing units that work together to process information. They also generate meaningful results from it.

We can apply Neural network not only for classification. It can also apply for regression of continuous target attributes.

Neural networks find great application in data mining used in sectors. For example economics, forensics, etc and for pattern recognition. It can be also used for data classification in a large amount of data after careful training.

**Structure of a Neural Network:**

The structure of a neural network also referred to as its ‘architecture’ or ‘topology’. It consists of the number of layers, Elementary units. It also consists of Interconchangend Weight adjustment mechanism. The choice of the structure determines the results which are going to obtain. It is the most critical part of the implementation of a neural network.

The simplest structure is the one in which units distributes in two layers: An input layer and an output layer. Each unit in the input layer has a single input and a single output which is equal to the input. The output unit has all the units of the input layer connected to its input, with a combination function and a transfer function. There may be more than 1 output unit. In this case, the resulting model is a linear or logistic regression.This is depending on whether transfer function is linear or logistic. The weights of the network are regression coefficients.

By adding 1 or more hidden layers between the input and output layers and units in this layer the predictive power of neural network increases. But a number of hidden layers should be as small as possible. This ensures that the neural network does not store all information from learning set but can generalize it to avoid overfitting.

Overfitting can occur. It occurs when weights make the system learn details of learning set instead of discovering structures. This happens when size of learning set is too small in relation to the complexity of the model.

A hidden layer is present or not, the output layer of the network can sometimes have many units, when there are many classes to predict.

**Advantages and Disadvantages of a Neural Network:**

Neural networks perform well with linear and nonlinear data but a common criticism of neural networks, particularly in robotics, is that they require a large diversity of training for real-world operation. This is so because any learning machine needs sufficient representative examples in order to capture the underlying structure that allows it to generalize to new cases.

Neural networks works even if one or few units fail to respond to network but to implement large and effective software neural networks, much processing and storage resources need to be committed. While the brain has hardware tailored to the task of processing signals through a graph of neurons, simulating even a most simplified form on Von Neumann technology may compel a neural network designer to fill millions of database rows for its connections – which can consume vast amounts of computer memory and hard disk space.

Neural network learns from the analyzed data and does not require reprogramming but they are referred to as black box” models,and provide very little insight into what these models really do. The user just needs to feed it inputs and watch it train and await the output.

**Partitioning into testing and training**

Here X is the independent data and y is the dependent data

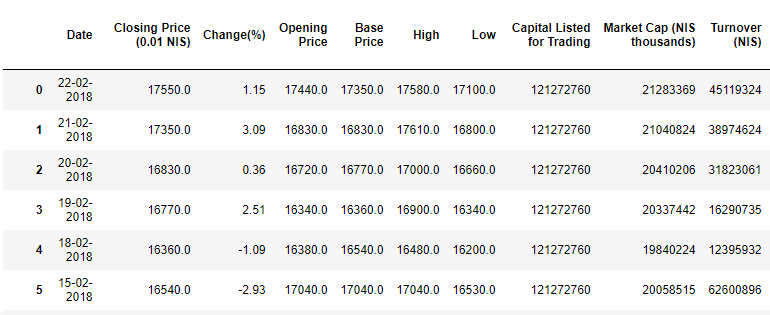
We partition independent and dependent data. In X the

independent variables are present which are having independent data

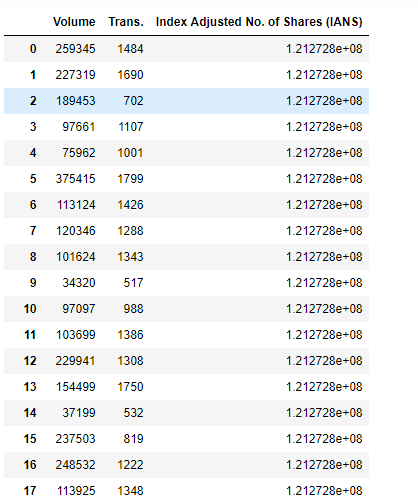
and the y variable contains the dependent variable that contains dependent data as from the above we know that the turnover will be the dependent variable .

As from the above basic dataset that is defined in the model section it is clear that the dependent variable is in the middle of the dataset so we need to partition the dataset so that the dependent and independent variables can be known clearly. And from the below it will be clear

x1=dataset.iloc[0:9999,1:9]



x2=dataset.iloc[0:9999,10:]



And by concatenating we get the value of “x” that the set contains all the

independent variables and looks like the above “x”,shown in the model section.

And the concatenation can be defined by

x=pd.concat([x1,x2], axis=1)

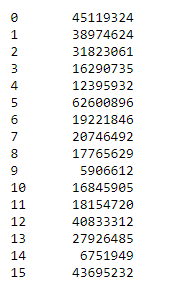
And “y” can be defined as

y=dataset.iloc[0:9999,9]

The “x” looks like



And the “y” looks like



Test and Train sets are partitioned as follows. In Cross-Validation the Train-Test-Split method is used.

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.3, random\_state = 0)

**4. Data Preprocessing**

import numpy as np

import pandas as pd

dataset=pd.read\_csv(r'C:\Users\satish\Downloads\top-35-israeli-stocks-5-years-data.zip') #it gives the path of the dataset on which the operations need to be performed.

x1=dataset.iloc[0:9999,1:9]

x2=dataset.iloc[0:9999,10:]

x=pd.concat([x1,x2], axis=1)

y=dataset.iloc[0:9999,9]

**5.Prediction and testing the samples:**

y\_pred=model.predict(x\_test)

y\_pred=(y\_pred>0.5)

For testing the sampling and to know the prediction for a particular sample it can be known by passing the values as below

y\_pred\_single=model.predict(np.array([[16830,0.36,16720,16770,17000,16660,1.27E+08,20410206,189453,702,1.21E+08]]))

y\_pred\_single

y\_pred\_single=(y\_pred\_single>0.5)

y\_pred\_single

**6. Findings and suggestions**

We get some help by using the following sources:

<https://www.kaggle.com/haimfeld87/top-35-israeli-stocks-5-years-data>

<https://en.wikipedia.org/wiki/Artificial_intelligence>

<https://en.wikipedia.org/wiki/Neural_network>

<https://www.tutorialspoint.com/artificial_neural_network/>

**7.Conclusion:**

Using the developed model we can predict the company's turnover based on the few parameters like opening price ,closing price, base price, capital for trading, market capital and the volume.

Our model is developed using the Artificial Neural Network, which facilitate in creating the neural model and training it using a sample data set. Here the training and testing of model plays a crucial role in prediction.

Since, many of the companies need to know the average turnover inorder to avoid huge loss, our model helps to predict the turnover on few factors. This model takes a few factors into consideration and then predicts the turnover as the predicted output.

We can also put our model into next phase by helping many of the small scale and large scale industries, factories and also the startup companies.