SUMMER INTERNSHIP REPORT

Area Of Online Internship	**AI/Machine/Deep Learning**
Intern Name	**Sachin Chouhan**
Name Of Institution	**INDIAN INSTITUTE OF TECHNOLOGY, INDORE**
Faculty Mentor Name	**Dr. Vimal Bhatia**
Duration	**1 MONTHS (01/10/2022 TO 01/11/2022)**
Date Of Submission	**05/11/2022**

Table Of Contents

1.Machine Learning

ML applications learn from experience (or to be accurate, data) like humans do without direct programming. When exposed to new data, these applications learn, grow, change, and develop by themselves. In other words, machine learning involves computers finding insightful information without being told where to look. Instead, they do this by leveraging algorithms that learn from data in an iterative process.

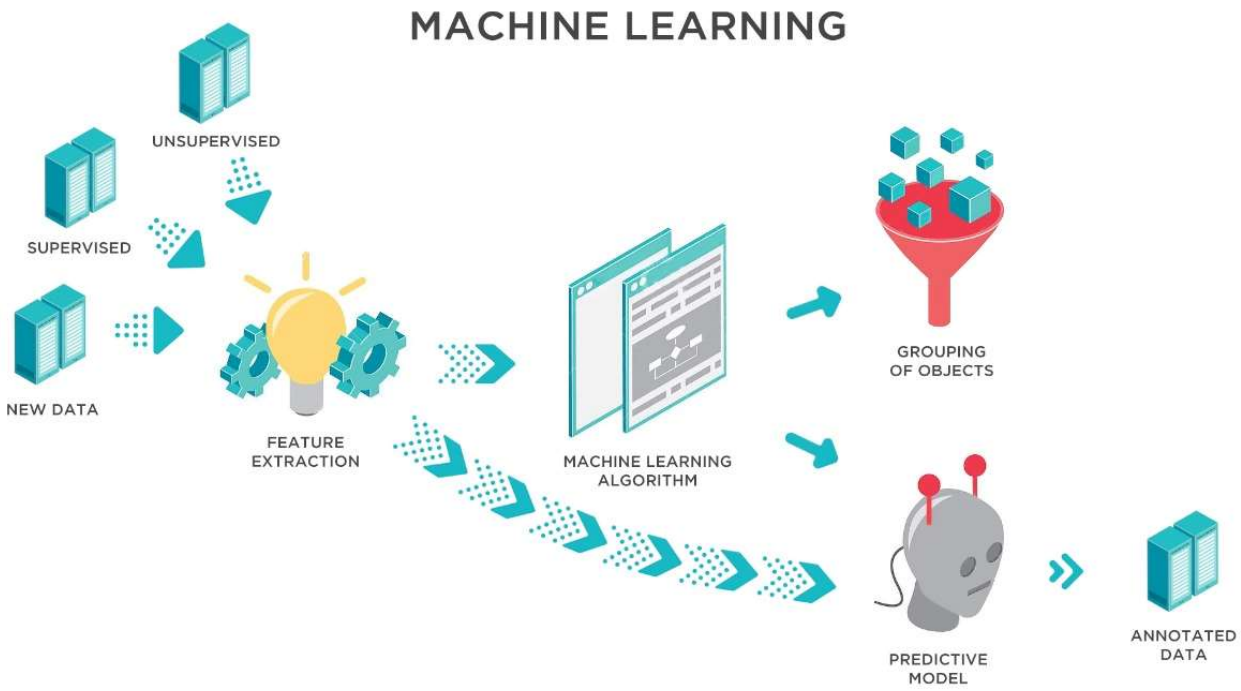

Fig. 1 Machine Learning

How does Machine Learning work?

The three major building blocks of a system are the model, the parameters, and the learner.

- Model is the system which makes predictions
- The parameters are the factors which are considered by the model to make predictions

- The learner makes the adjustments in the parameters and the model to align the predictions with the actual results

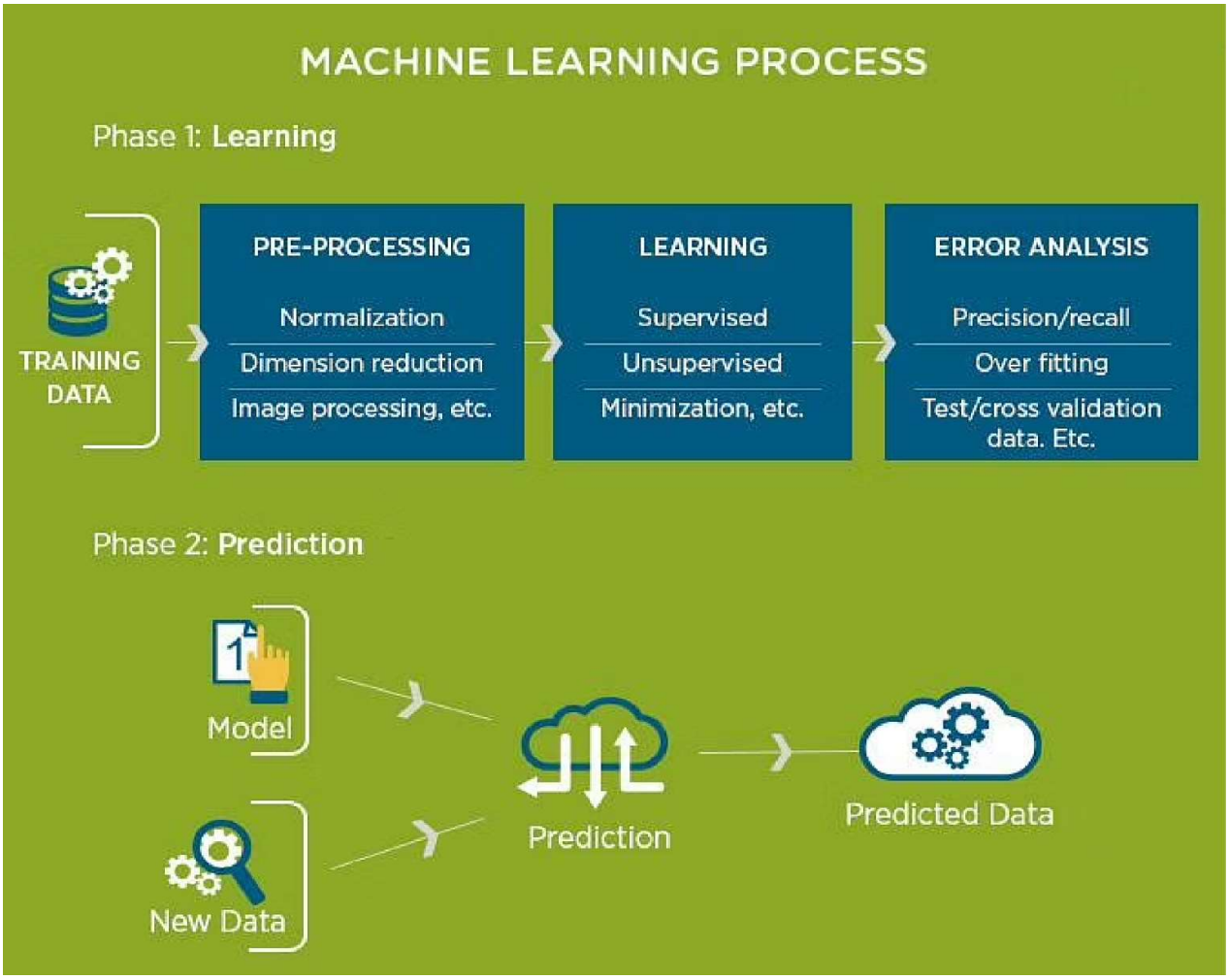

Fig. 2 Machine Learning Process

Let us build on the Stock Predictor example from above to understand how machine learning works. A machine learning model here has to predict if a Stock will be in profit or loss. The parameters selected are the previous closing and news related to the stock. The first step is:

1. Learning from the training set

This involves taking a sample data set of Stock for which the Previous Day data and stock news is specified. Now, we have to define the description of each classification, that is in profit or loss or long-term gain, in terms of the value of parameters for each type. The model can use the description to decide if a new Stock should be trade or

not.

You can represent the values of the parameters, ‘Previous High’ and ‘Stock Fundamentals’ as ‘x’ and ‘y’ respectively. Then (x,y) defines the parameters of each Stock in the training data. This set of data is called a training set. These values, when plotted on a graph, present a hypothesis in the form of a line, a rectangle, or a polynomial that fits best to the desired results.

2. Measure error

Once the model is trained on a defined training set, it needs to be checked for discrepancies and errors. We use a fresh set of data to accomplish this task. The outcome of this test would be one of these four:

True Positive: When the model predicts the condition when it is present
True Negative: When the model does not predict a condition when it is absent
False Positive: When the model predicts a condition when it is absent
False Negative: When the model does not predict a condition when it is present
machine learning process

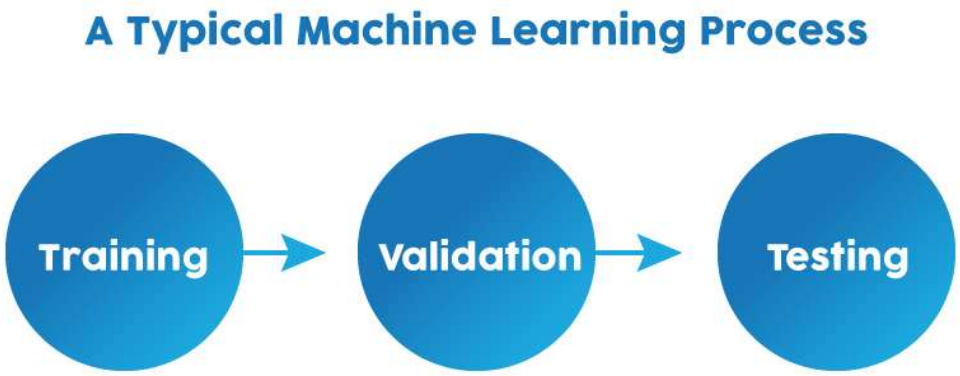

Fig. 3 Machine Learning Process

The sum of FP and FN is the total error in the model.

3. Manage Noise

For the sake of simplicity, we have considered only two parameters to approach a machine learning problem here that is the Previous day high and positive news about the stock. But in reality, you will have to consider hundreds of parameters and a

broad set of learning data to solve a machine learning problem.

- The hypothesis then created will have a lot more errors because of the noise. Noise is the unwanted anomalies that disguise the underlying relationship in the data set and weakens the learning process. Various reasons for this noise to occur are:
- Large training data set
- Errors in input data
- Data labelling errors
- Unobservable attributes that might affect the classification but are not considered in the training set due to lack of data

You can accept a certain degree of training error due to noise to keep the hypothesis as simple as possible.

4. Testing and Generalization

While it is possible for an algorithm or hypothesis to fit well to a training set, it might fail when applied to another set of data outside of the training set. Therefore, It is essential to figure out if the algorithm is fit for new data. Testing it with a set of new data is the way to judge this. Also, generalisation refers to how well the model predicts outcomes for a new set of data.

When we fit a hypothesis algorithm for maximum possible simplicity, it might have less error for the training data, but might have more significant error while processing new data. We call this is underfitting. On the other hand, if the hypothesis is too complicated to accommodate the best fit to the training result, it might not generalise well. This is the case of over-fitting. In either case, the results are fed back to train the model further.

Types of Machine Learning

There are three main types:

- **Supervised learning:** In this type, the model is trained on a labelled dataset, meaning that each example in the training data has a known label. The model can then make predictions on new, unlabelled data.

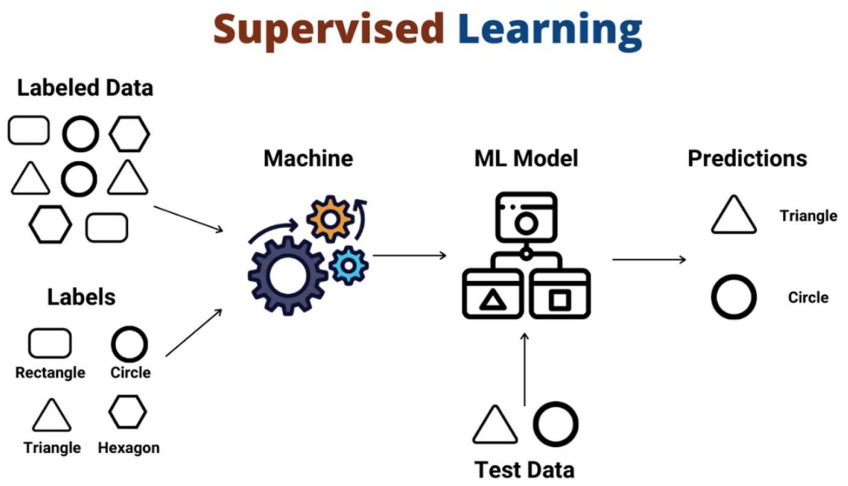

Fig. 4 Supervised Learning

- **Unsupervised learning:** In this type, the model is trained on an unlabeled dataset, meaning that the examples in the training data do not have known labels. The model can then find patterns or groups in the data.

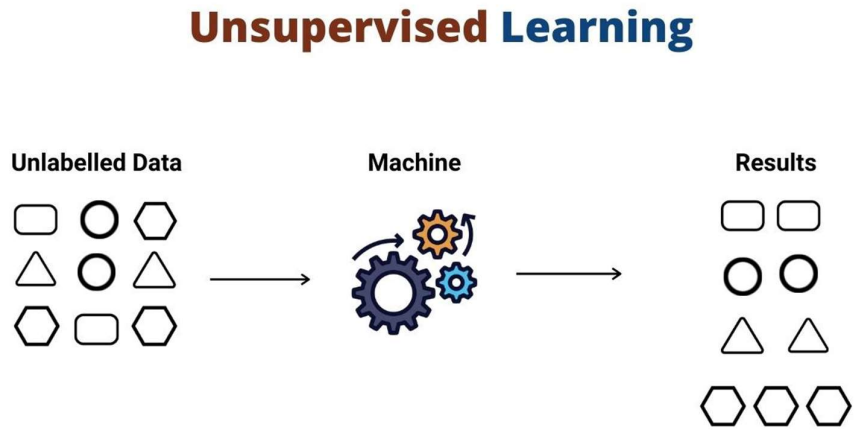

Fig. 5 Unsupervised Learning

- **Reinforcement learning:** In this type, the model is trained by interacting with an environment where it receives rewards or punishments for its actions. The model can then be used to make decisions in new situations to maximize its rewards.

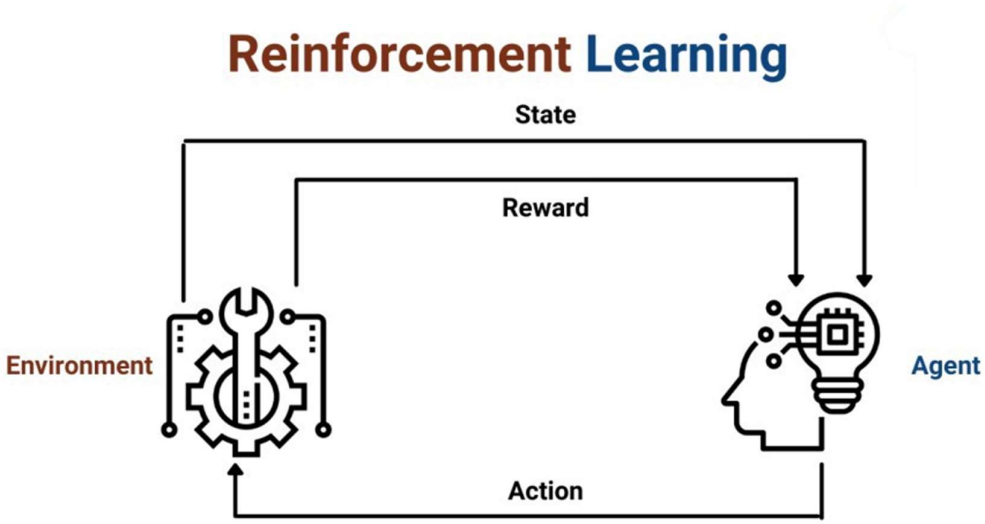

Fig. 6 Reinforcement Learning

Advantages and Disadvantages

Everything comes with a few advantages and disadvantages. In this section, let's talk about a few of the basic advantages and disadvantages of ML.

Advantages:

- It can be used for pattern detection.
- It can be used to make predictions about future data.
- It can be used to generate new features from data automatically.
- It can be used to cluster data automatically.
- It can be used to detect outliers in data automatically.

Disadvantages:

Some disadvantages include the potential for biased data, overfitting data, and lack of explainability.

2.Python

Python is famous for its readability and relatively lower complexity as compared to other programming languages. ML applications involve complex concepts like calculus and linear algebra which take a lot of effort and time to implement. Python helps in reducing this burden with quick implementation for the ML engineer to validate an idea. You can check out the Python Tutorial to get a basic understanding of the language. Another benefit of using Python is the pre-built libraries. There are different packages for a different type of applications, as mentioned below:

- Numpy, OpenCV, and Scikit are used when working with images
- NLTK along with Numpy and Scikit again when working with text
- Librosa for audio applications
- Matplotlib, Seaborn, and Scikit for data representation
- TensorFlow and Pytorch for Deep Learning applications
- Scipy for Scientific Computing
- Django for integrating web applications
- Pandas for high-level data structures and analysis

Here is a summary:

Fig. 7 Python Features

3.Regression

Regression is a technique for examining the relationship between independent variables or features and a dependent variable or outcome. It is a method for predictive modelling in machine learning in which an algorithm is used to predict continuous outcomes.

Regression problems are one of the most common applications for machine learning models, particularly in supervised machine learning. Algorithms are trained to understand the relationship between independent variables and an outcome or dependent variable. The model can then be used to predict the outcome of new and unknown input data, or to fill a gap in missing data.

Regression analysis is an essential component of any forecasting or predictive model, and it is a common method in machine learning powered predictive analytics. Along with classification, regression is a popular application for supervised machine learning models. This training model approach required labelled input and output training data. Machine learning regression models require accurate labelled training data to understand the relationship between features and outcome variables.

Because regression is a key component of predictive modelling, it can be found in a wide range of machine learning applications. Regression analysis, whether used to power financial forecasting or predict healthcare trends, can provide organisations with critical insight for decision-making. It's already being used in a variety of industries to forecast house prices, stock or share prices, and map salary changes.

What is the definition of machine learning regression?

Regression is a technique for determining the relationship between two or more independent variables or features and a dependent variable or outcome. After estimating the relationship between the independent and dependent variables, outcomes can be predicted. Regression is a statistical field of study that plays an important role in machine learning forecast models. It is used in predictive modelling to predict continuous outcomes, so it is useful for forecasting and predicting data outcomes. In general, machine learning regression entails drawing a line of best fit through the data points. To find the best fit line, the distance between each point and the line is minimised.

Along with classification, regression is a key application of supervised machine learning. Classification is the categorisation of objects based on learned features, whereas regression

is the prediction of continuous outcomes. Both are predictive modelling issues. Because classification and regression models rely on labelled input and output training data, supervised machine learning is essential in both cases. The features and output of the training data must be labelled so that the model can understand the relationship.

Regression analysis is used to understand the relationship between different independent variables and the dependent variable or outcome. Models are trained to forecast or predict trends and outcomes will be trained using regression techniques. These models will learn the relationship between input and output data from labelled training data. It can then predict future trends or predict outcomes from unseen input data or be used to understand gaps in historical data.

As with all supervised machine learning, special care ought to be taken to make sure the labelled coaching information is representative of the general population. If the training data isn't representative, the prophetical model are overfit to data that doesn’t represent new and unseen data. this may lead to inaccurate predictions once the model is deployed. because of multivariate analysis involves the relationships of options and outcomes, care should be taken to incorporate the correct choice of features too.

What are regression models used for?

Machine learning regression models are mostly used in predictive analytics to foresee trends and predict outcomes. Regression models will be trained to understand the link between several independent variables and a result. As a result, the model can comprehend the numerous aspects that may result in a desired conclusion. The generated models can be applied in a variety of ways and circumstances. Outcomes may be predicted from previously unseen data, market swings can be forecast and accounted for, and campaigns can be tested by adjusting several independent variables.

In practise, a model will be trained on labelled data to comprehend the link between data attributes and the dependent variable. The model can predict the outcome of new and previously unseen data by estimating this relationship. This could be used to forecast missing historical data as well as estimate future outcomes. In a sales setting, an organisation could utilise regression machine learning to forecast next month's revenue based on a variety of criteria. In a medical setting, an organisation could forecast health trends in the general population over time.

Generally, supervised machine learning models are used for classification or regression

problems. Classification is the process of training a model to categorise an object based on its features. This could include facial recognition software or detecting spam emails within a firewall. To understand the specific features that classify a labelled object, a model will be trained on labelled input and output data. A regression problem, on the other hand, occurs when a model is used to predict continuous outcomes or values. This could be a model that predicts salary increases, home prices, or retail sales. To understand the strength of relationships between data features and output, the model is trained on labelled input and output data.

Regression is used to uncover patterns and relationships within a dataset, which may subsequently be applied to new and previously overlooked data. As a result, regression is a critical component of machine learning in finance, and it is frequently used to estimate portfolio performance or stock costs and trends. Models can be trained to grasp the relationship between a variety of different features and a desired output. In most circumstances, machine learning regression offers organisations with insight into certain outcomes. However, because this approach has the potential to affect an organization's decision-making process, the explainability of machine learning is a crucial aspect.

Machine learning regression models are commonly used for the following purposes:

- Forecasting continuous outcomes such as housing values, stock prices, or sales.
- Predicting the performance of future retail sales or marketing efforts to ensure that resources are used effectively.
- Predicting customer or user trends, such as on streaming services or retail websites.
- Analysing datasets to determine the links between variables and output.
- Predicting interest rates or stock values based on a multitude of factors.
- Making time series visualisations.

What kinds of regression are there?

Regression can be carried out using a variety of different techniques in machine learning. Machine learning regression is accomplished using a variety of well-known techniques. The various methods could use various numbers of independent variables or handle various kinds of data. A different relationship between the independent and dependent variables may also be assumed by different kinds of machine learning regression models. For instance, because linear regression approaches presuppose a linear relationship, they are ineffective for datasets having nonlinear interactions.

The following categories of regression analysis can be used to classify some of the most popular regression approaches in machine learning:

- Simple Linear Regression
- Multiple linear regression
- Logistic regression

What is it Simple linear regression?

To reduce inaccuracy between the line and the data points, simple linear regression plots a straight line inside the data points. It is among the simplest and most fundamental varieties of machine learning regression. In this instance, it is assumed that the connection between the independent and dependent variables is linear. Because just one independent variable and the dependent variable are examined, this strategy is straightforward. Because the straight line of greatest fit is used in simple linear regression, outliers could arise often.

What is multiple linear regression?

When more than one independent variable is involved, the multiple linear regression method is used. A multiple linear regression technique is polynomial regression. When there are several independent variables, this kind of multiple linear regression is employed. When more independent variables are present, it provides a superior fit when compared to simple linear regression. The outcome would be a curving line fitted to the data points when plotted in two dimensions.

What is logistic regression?

When the dependent variable may only take one of two possible values, such as true or false or success or failure, logistic regression is utilised. It is possible to forecast the likelihood that a dependent variable will occur using logistic regression models. The output values must typically be binary. The relationship between the dependent variable and independent variables can be visualised using a sigmoid curve.

4.Recurrent Neural Network(RNN)

Neural Networks is one of the most popular machine learning algorithms and also outperforms other algorithms in both accuracy and speed. Therefore it becomes critical to have an in-depth understanding of what a Neural Network is, how it is made up and what its reach and limitations are.

What Is a Neural Network?
A Neural Network consists of different layers connected to each other, working on the structure and function of a human brain. It learns from huge volumes of data and uses complex algorithms to train a neural net.

Here is an example of how neural networks can identify a dog's breed based on their features.

- The image pixels of two different breeds of dogs are fed to the input layer of the neural network.
- The image pixels are then processed in the hidden layers for feature extraction.
- The output layer produces the result to identify if it's a German Shepherd or a Labrador.
- Such networks do not require memorizing the past output.

Several neural networks can help solve different business problems. Let's look at a few of them.

- **Feed-Forward Neural Network:** Used for general Regression and Classification problems.
- **Convolutional Neural Network:** Used for object detection and image classification.
- **Deep Belief Network:** Used in healthcare sectors for cancer detection.
- **RNN:** Used for speech recognition, voice recognition, time series prediction, and natural language processing.

What Is a Recurrent Neural Network (RNN)?

A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, natural language processing

(nlp), speech recognition, and image captioning; they are incorporated into popular applications such as Siri, voice search, and Google Translate. Like feedforward and convolutional neural networks (CNNs), recurrent neural networks utilize training data to learn. They are distinguished by their "memory" as they take information from prior inputs to influence the current input and output. While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of recurrent neural networks depend on the prior elements within the sequence. While future events would also be helpful in determining the output of a given sequence, unidirectional recurrent neural networks cannot account for these events in their predictions.

RNN works on the principle of saving the output of a particular layer and feeding this back to the input in order to predict the output of the layer.

How Does Recurrent Neural Networks Work?

In Recurrent Neural networks, the information cycles through a loop to the middle hidden layer.

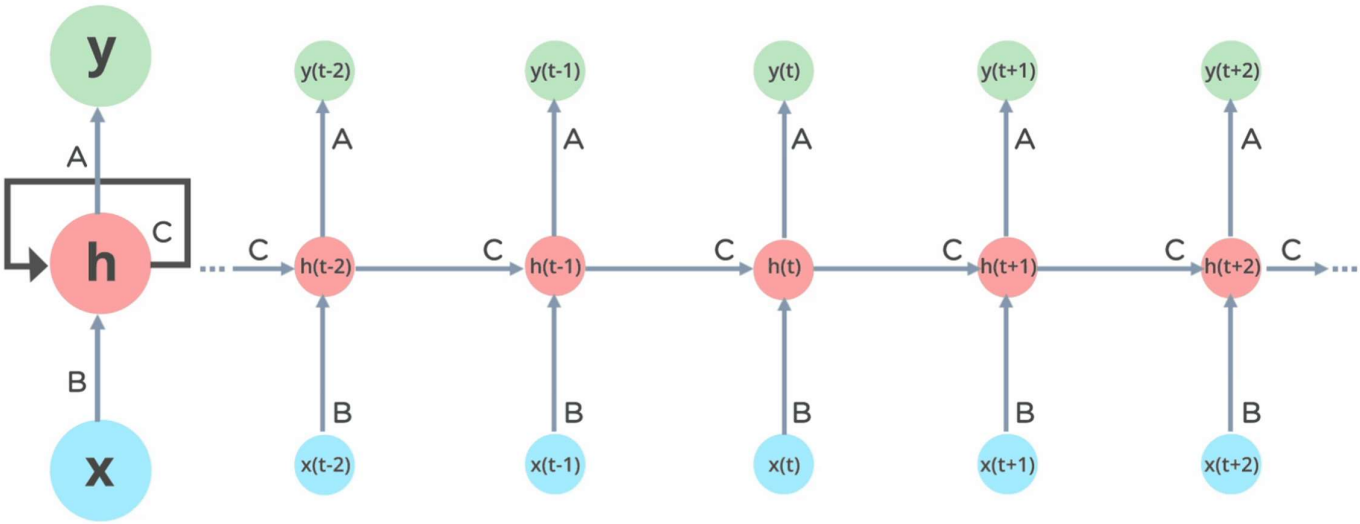

Fig. 8 Recurrent Neural Network Working

The input layer 'x' takes in the input to the neural network and processes it and passes it onto the middle layer.

The middle layer ‘h’ can consist of multiple hidden layers, each with its own activation functions and weights and biases. If you have a neural network where the various parameters of different hidden layers are not affected by the previous layer, ie: the neural network does not have memory, then you can use a recurrent neural network.

The Recurrent Neural Network will standardize the different activation functions and weights and biases so that each hidden layer has the same parameters. Then, instead of creating multiple hidden layers, it will create one and loop over it as many times as required.

5.Stock Prediction

Predict the stock market price of next few days using previous stock market data (equity or indices) using machine learning or Deep learning.
1. Use News headlines as Data for prediction.
2. Use previous Equity data of Day open, close, low, high for prediction.
3. Any other stock Relative data.

What is the Stock Market?
A stock market is a public market where you can buy and sell shares for publicly listed companies. The stocks, also known as equities, represent ownership in the company. The stock exchange is the mediator that allows the buying and selling of shares.

The Importance of the Stock Exchange

- Stock exchanges assist businesses in raising financing.
- It contributes to personal wealth creation.
- The stock market is a leading indication of the state of the economy.
- It is a popular way for investors to invest in firms with strong growth potential.

Stock Price Prediction using machine learning helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to gain significant profits. Predicting how the stock market will perform is a hard task to do. There are other factors involved in the prediction, such as physical and psychological factors, rational and irrational behaviour, and so on. All these factors combine to make share prices dynamic and volatile. This makes it very difficult to predict stock prices with high accuracy.

Understanding Long Short Term Memory Network
Here, you will use a Long Short Term Memory Network (LSTM) for building your model to predict the stock prices of Google.

LTSMs are a type of Recurrent Neural Network for learning long-term dependencies. It is commonly used for processing and predicting time-series data.

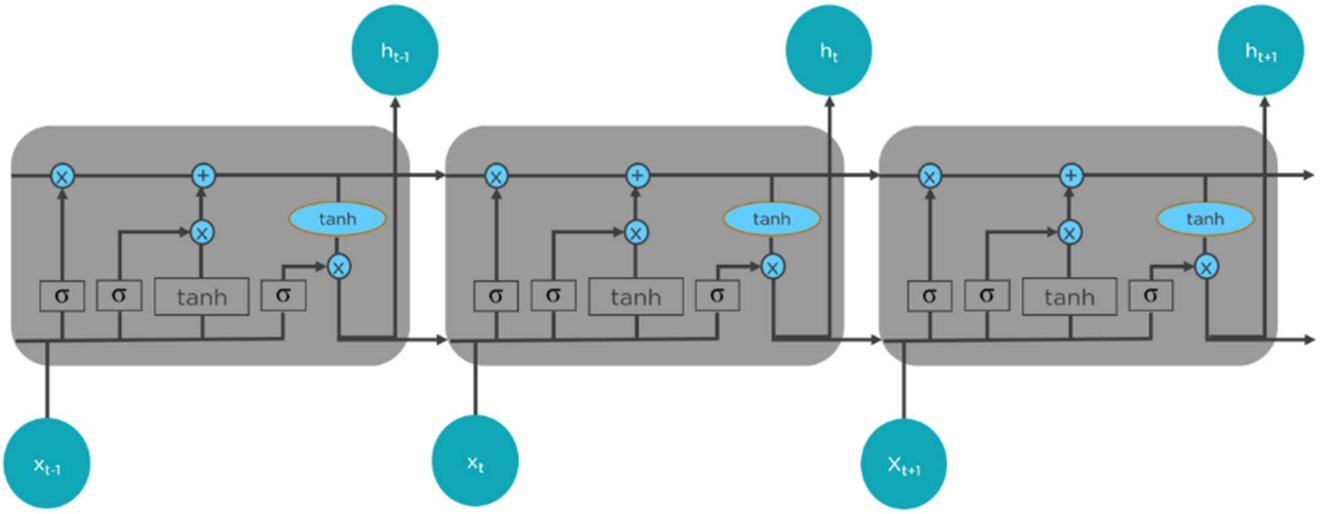

Fig. 9 LSTM

From the image on the top, you can see LSTMs have a chain-like structure. General RNNs have a single neural network layer. LSTMs, on the other hand, have four interacting layers communicating extraordinarily.

LSTMs work in a three-step process.

- The first step in LSTM is to decide which information to be omitted from the cell in that particular time step. It is decided with the help of a sigmoid function. It looks at the previous state (ht-1) and the current input xt and computes the function.
- There are two functions in the second layer. The first is the sigmoid function, and the second is the tanh function. The sigmoid function decides which values to let through (0 or 1). The tanh function gives the weightage to the values passed, deciding their level of importance from -1 to 1.
- The third step is to decide what will be the final output. First, you need to run a sigmoid layer which determines what parts of the cell state make it to the output. Then, you must put the cell state through the tanh function to push the values between -1 and 1 and multiply it by the output of the sigmoid gate.
-

With this basic understanding of LSTM, you can dive into the hands-on demonstration part of this tutorial regarding stock price prediction using machine learning.

6.Working Model-

With the help of the Regression and LSTM (RNN) We train our model and predict the future open and close price of the stock for any given data set.

The code link is as follow –

https://drive.google.com/drive/folders/1Rd5pR2nBh2qYOY_UwbkLyIH5JCWfTkxq?usp=sharing

7.REFERENCES

[1.] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock Market prediction using machine learning

[2.] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International.

[3.] M. Billah, S. Waheed and A. Hanifa, "Stock market Prediction using an improved training algorithm of neural network," 2016.

[4.] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International.

[5.] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai,2010, pp. 59-63.