**Machine Learning:Face Recognition**

**A Project Work**

*Submitted in the partial fulfilment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE**

**SUBMITTED BY:**

**DARSHIT VERMA**

**17BCS4633**

**UNDER THE SUPERVISION OF:**

**Mr. Anuj Garg**



**APEX INSTITUTE OF TECHNOLOGY**

**CHANDIGARH UNIVERSITY,** **GHARUAN,MOHALI  - 140413,**

**PUNJAB**

**JULY 2019**

**ABOUT THE COMPANY**

Coding Blocks is a computer training institute located in Pitampura, New Delhi. It was founded by Mr. Manmohan Gupta, an IIT-Delhi graduate, is an ace programmer, technocrat, an entrepreneurial doyen and a mathematician. He has co-founded Software giant, **Nagarro** and **Vidyamandir Classes(VMC)**. All our mentors have been well performing students at the best of colleges, have worked at industry giants such as Sony, Cyanogen, SanDisk, Facebook, Barclays, Practo, Harman, DRDO, Cadence, American Express, Nagarro and many more. We also have GSoCers, GSoC Mentors and Google Code-in mentors in our team, in addition to people who've been published in international journals.

The institute is one of the most renowned institutes in Delhi and the entire country for its training and placement preparation programs. In 5 years of its establishment, the institute has trained more than 15000 students in the different fields of computer science and placed more than 5000 students in different companies like Amazon, Adobe, Microsoft, Goldman Sachs.

The company provides training in different fields of computer science like programming languages(C++, Java, Python), Data Structures(the most prominent field in which they work), Machine Learning, Data Science, Web Development, Android App Development, Competitive Programming, Interview Preparation, Aptitude Preparation. Data Structures, which is a requirement for every job, is one of the major working areas of Coding Ninjas.

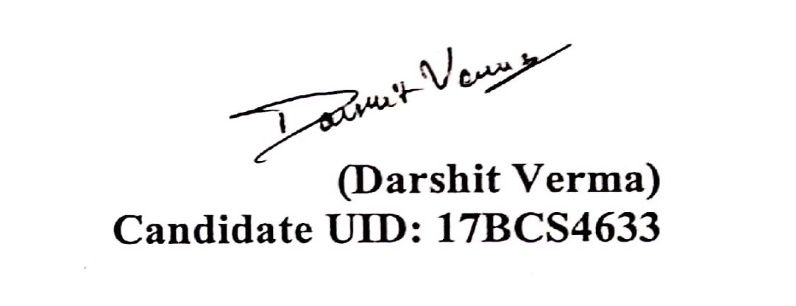
The Data Structures course of Coding Blocks is one of the best courses in this country which helps the students to grasp the basics of Computer Science well. Another area in which they work is Machine Learning and Data Science.Also, while working for Facebook, he was a core member of the team which developed a new profile on the social networking site, which we all now know as the Timeline.

Competitive Programming and advanced data structures is another major area in which the company works. This is a program which prepares the students for placements as well as major world level programming competitions like ACM – ICPC.

Apart from this, Android App Development training is conducted by Mr. Rohan Arora, who is a Google Certified Android Developer and also a member of the Google Developer’s Group (GDG). The company also trains the students in the field of web development. The different technologies taught in the web development training are HTML, CSS, JavaScript, SQL, ExpressJS, MongoDB(No SQL database), oAuth(login using social providers like Google, Facebook, etc), BootStrap, ES6, JSON, AJAX, etc

**DECLARATION**

I, **Darshit Verma**, student of **‘Bachelor of Engineering in Computer Science with a specialization in Internet of Things’,** for the **session: 2018 - 2019**, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in the Project Work entitled ‘**Face Recognition using Haarcascade File’**is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

****

**Date:**

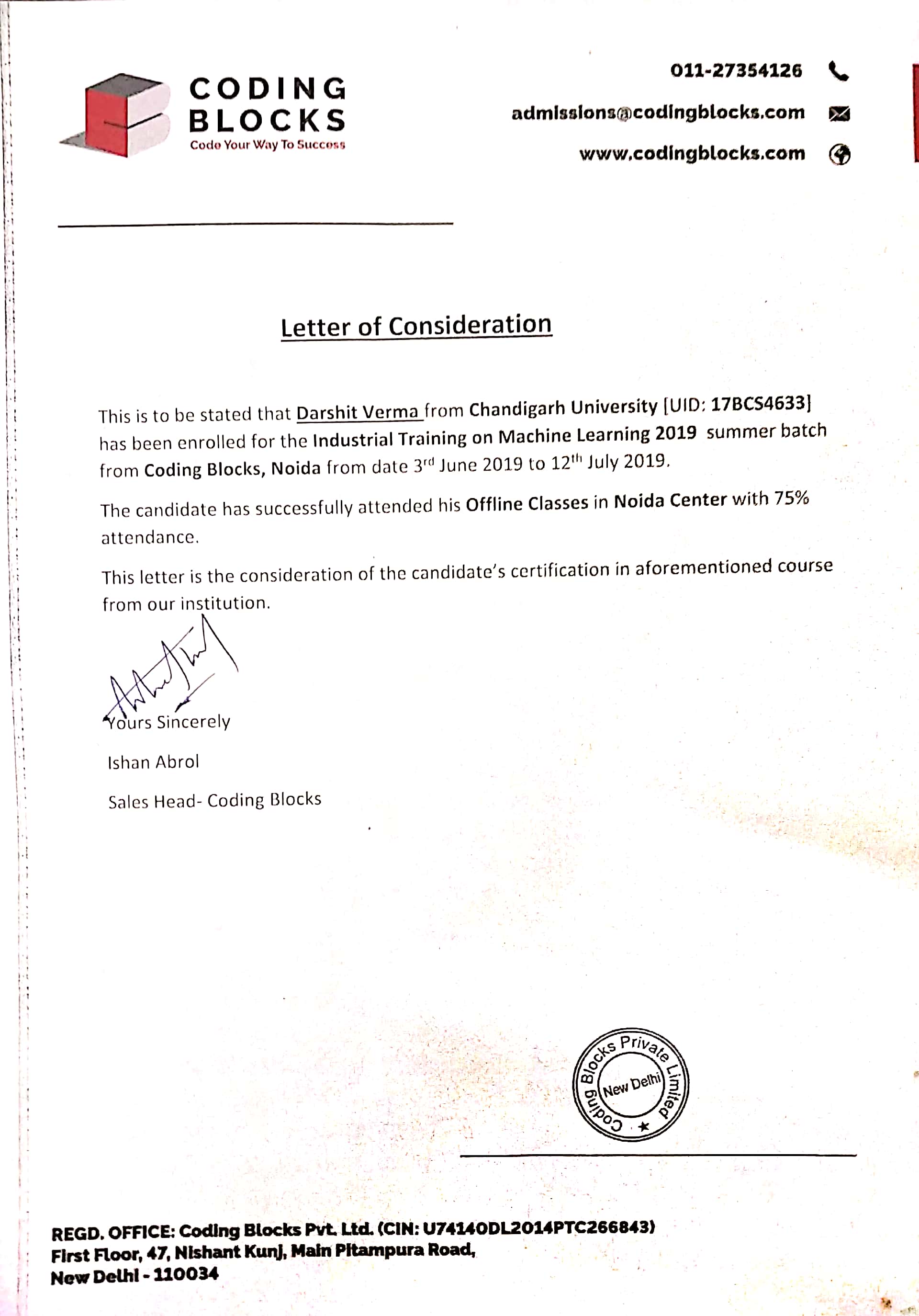
**July 18,2019.**

**Place:**

**Chandigarh University,**

**Gharuan, Mohali.**

**(Punjab)**

**CERTIFICATE**

**ACKNOWLEDGEMENT**

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I would like to thank Mr. Anuj Garg, our supervisor for the training for their efforts and teaching us Machine Learning and giving us a new skill set. I would like to thank them for wonderfully taking us through the different concepts of the subject / domain and explaining its importance in the industry and its real – time uses. I would like to thank them for their continuous guidance and their support throughout the training.

I would also like to thank the TAs for the course who were always available to solve our doubts and help us with our projects if we got stuck anywhere and also helping us to clear our concepts on various aspects of the training, either in person after the class or through whatsApp groups at any time.

I would like to thank the entire support team of Coding Blocks for seamlessly organizing the training and also bringing different opportunities to our doors and also organizing webinars / sessions with the industry leaders in different fields to keep us motivated and realize the importance of the work.

**ABSRACT**

**Machine Learning** is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

Machine Learning can be broadly classified into 2 categories: Supervised and Unsupervised Learning.

Supervised learning is a learning in which we teach or train the machine using data which is well labelled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples (data) so that supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labelled data. It can be broadly classified into 2 categories: Classification and Regression.

Unsupervised learning is the training of an artificial intelligence (AI) algorithm using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. It can be broadly classified into 2 types: Clustering and Association.

The training started with learning the syntax of Python and problem solving using Python. After that, the next part consisted of learning the python libraries necessary for Machine Learning which act as important tools while using Machine learning algorithms. The different modules covered were: Pandas, Numpy, Matplotlib. Pandas and numpy help in fetching nad manipulating the data whereas Matplotlib helps in visualising the data through bar graphs, histograms, plots, pie charts, etc.

After that, I was introduced to basics of Machine Learning and its benefits, use cases, and the library having inbuilt ML models – Sklearn. Then I was introduced to different regression and classification algorithms like Linear Regression, Gradient Descent and Logistic Regression, Decision Trees, Random Forests, KNN, Naïve Bayes, Support Vector Machines (SVMs) and Principle Component Analysis (PCA).

The next part consisted of working on Natural Language Processing (NLP), a basic introduction to Neural Networks and deep learning. After that, there was a basic introduction to Tensor Flow and Git (Version Control System) and GitHub and Unsupervised learning algorithms like k-Means Clustering.

During the training, I did several projects also like implementing Linear Regression on my own, applying Logistic Regression on titanic dataset and predicting who survived and who did not survive by applying different aspects of Machine learning on the dataset like data cleaning, feature extraction, etc. Another project I worked upon during the training was image classification on Cifar10 dataset which consists of around 60000 images of different categories.

**List of Figures**

Figure1: Linear Regression with sklearn

Figure2: Challenge: Hardwork Pays Off

Figure3: Multivariate Regression

Figure4: Boston Housing Prices

Figure5.: K-Nearest Neighbours

Figure6: Diabetes Detection Challenge

Figure7: : Gradient Descent

Figure8: Real Time Face Recognition using Haarcascade

# **CONTENTS**

# 

Title Page i

About The Company ii

[Certificate ii](#_TOC_250004)i

[Acknowledgement i](#_TOC_250003)v

[List of Figures v](#_TOC_250002)

ABSTRACT vi

[CONTENTS…](#_TOC_250000) vii

CHAPTER 1 INTRODUCTION 1

* 1. Some Methods of Machine Learning …
  2. Machine Learning Advantages …
  3. Machine Learning Applications …

CHAPTER 2 THEORY

* 1. General Theory …
  2. Working Environment …
  3. Supervised Learning
  4. Unsupervised Learning

CHAPTER 3 PROJECT AND METHODOLOGY ADOPTED

* 1. Linear Regression with sklearn (Walkatime Dataset)
  2. Multivariate Regression (Boston Housing Price)
  3. K-Nearest Neighbours(Diabetes Detection Challenge)
  4. Gradient Descent
  5. Real Time Face Recognition using Haarcascade File

CHAPTER 4 RESULTS AND DISCUSSIONS

CHAPTER 6 CONCLUSIONS AND FUTURE SCOPE OF STUDY

REFERENCES

**CHAPTER – 1: INTRODUCTION**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow thecomputerslearn automatically without human intervention or assistance and adjust actions accordingly.

**1.1. Some machine learning methods**

Machine learning algorithms are often categorized as supervised or unsupervised.

Supervised machine learning algorithmscan apply what has been learned in the past to new data using labelled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, unsupervised machine learning algorithmsare used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabelled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labelled and unlabelled data for training – typically a small amount of labelled data and a large amount of unlabelled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labelled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabelled data generally doesn’t require additional resources.

Reinforcement machine learning algorithmsis a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

**1.2 Advantages Of Machine Learning**

#### 1. **Easily identifies trends and patterns**

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviours and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

#### 2. **No human intervention needed (automation)**

With ML, you don’t need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus soft wares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

#### 3. **Continuous Improvement**

As ML Algorithms gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

#### 4. **Handling multi-dimensional and multi-variety data**

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

#### 5. **Wide Applications**

You could be an e-tailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

**1.3 Applications Of Machine Learning**

1. **Web Search Engine:** One of the reasons why search engines like google, bing etc work so well is because the system has learnt how to rank pages through a complex learning algorithm.

2. **Photo tagging Applications:** Be it facebook or any other photo tagging application, the ability to tag friends makes it even more happening. It is all possible because of a face recognition algorithm that runs behind the application.

3. **Spam Detector:** Our mail agent like Gmail or Hotmail does a lot of hard work for us in classifying the mails and moving the spam mails to spam folder. This is again achieved by a spam classifier running in the back end of mail application.

4. **Product Recommendations:** You shopped for a product online few days back and then you keep receiving emails for shopping suggestions. If not this, then you might have noticed that the shopping website or the app recommends you some items that somehow matches with your taste. Certainly, this refines the shopping experience but did you know that it’s machine learning doing the magic for you? On the basis of your behaviour with the website/app, past purchases, items liked or added to cart, brand preferences etc., the product recommendations are made.

Apart from these applications, Machine Learning is used in our day – to – day life everywhere like smartphones, social media, in fraud detection, in making predictions, solving classification problems, regression analysis, Chatbots, Face Recognition, Image Recognition and classification, Virtual Personal Assistants like Alexa, Siri, and Google are all built on top of machine learning. They are trained using different Machine Learning Algorithms.

**CHAPTER – 2: THEORY**

**2.1 General Theory**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow thecomputerslearn automatically without human intervention or assistance and adjust actions accordingly.

**Some machine learning methods**

Machine learning algorithms are often categorized as supervised or unsupervised.

Supervised machine learning algorithmscan apply what has been learned in the past to new data using labelled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, unsupervised machine learning algorithmsare used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabelled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labelled and unlabelled data for training – typically a small amount of labelled data and a large amount of unlabelled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labelled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabelled data generally doesn’t require additional resources.

Reinforcement machine learning algorithmsis a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behaviour within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

**Advantages Of Machine Learning**

#### 1. **Easily identifies trends and patterns**

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviours and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

#### 2. **No human intervention needed (automation)**

With ML, you don’t need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus soft wares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

#### 3. **Continuous Improvement**

As ML Algorithms gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

#### 4. **Handling multi-dimensional and multi-variety data**

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

#### 5. **Wide Applications**

You could be an e-tailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

**Applications Of Machine Learning**

1. **Web Search Engine:** One of the reasons why search engines like google, bing etc work so well is because the system has learnt how to rank pages through a complex learning algorithm.

2. **Photo tagging Applications:** Be it facebook or any other photo tagging application, the ability to tag friends makes it even more happening. It is all possible because of a face recognition algorithm that runs behind the application.

3. **Spam Detector:** Our mail agent like Gmail or Hotmail does a lot of hard work for us in classifying the mails and moving the spam mails to spam folder. This is again achieved by a spam classifier running in the back end of mail application.

4. **Product Recommendations:** You shopped for a product online few days back and then you keep receiving emails for shopping suggestions. If not this, then you might have noticed that the shopping website or the app recommends you some items that somehow matches with your taste. Certainly, this refines the shopping experience but did you know that it’s machine learning doing the magic for you? On the basis of your behaviour with the website/app, past purchases, items liked or added to cart, brand preferences etc., the product recommendations are made.

**2.2 Working Environment**

The working environment used during the training was :

1. Anaconda

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. Package versions are managed by the package management system conda. Anaconda contains all the important python libraries, tools and frameworks that are necessary for Machine Learning. It also has the sklearn library which contains the inbuilt implementation of the majority of Machine Learning Algorithms.

2. Jupyter Notebook

Jupyter notebook is a very popular and flexible tool which lets us put our code, output of the code and any kind of visualization or plot etc. in the same document. Jupyter notebook comes pre-installed with the Anaconda distribution of Python. It helps to deal with large amount of data in a hassle-free manner and also is a good tool to visualize our data. It is web-based, hence, it does not take much space in our system and is easy to use.

3. SkLearn

Sklearn is a library built on top of other python libraries like numpy, pandas, matplotlib, scipy. It contains the inbuilt implementation of majority of Machine Learning algorithms like Linear Regression, Logistic Regression, Decision Trees, Random Forest, Support Vector Machines(SVMs), Principle Component Analysis(PCA), Naïve Bayes.

4. Matplotlib

Matplotlib is a python library that works as a data visualtion tool. Data visualisation is an important aspect of Machine learning as it helps us to detect how the features are varying and how we should train our model in order to get the best predictions for our dataset. Matplotlib helps us to plot graphs, bar graphs, histograms, pie charts, etc.

5. Numpy & Pandas

These 2 libraries are used to fetch the data and manipulate it as required. Data cleaning is an important aspect of Machine Learning. Numpy and Pandas help us to fetch the data from the dataset and clean the data, i. e, removing unwanted features from the dataset which will not contribute in making predictions or training our model. Numpy arrays and Pandas series and Data Frames are used to achieve this.

**2.3. Supervised Learning**

Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

Y = f(X)

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

Supervised learning problems can be further grouped into regression and classification problems.

* **Classification**: A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
* **Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

**2.4 Unsupervised Learning**

Unsupervised learning is where you only have input data (X) and no corresponding output variables.

The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.

These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data.Unsupervised learning problems can be further grouped into clustering and association problems.

* **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.
* **Association**:  An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Some popular examples of unsupervised learning algorithms are:

* K - means for clustering problems.

Apriori algorithm for association rule learning problems.

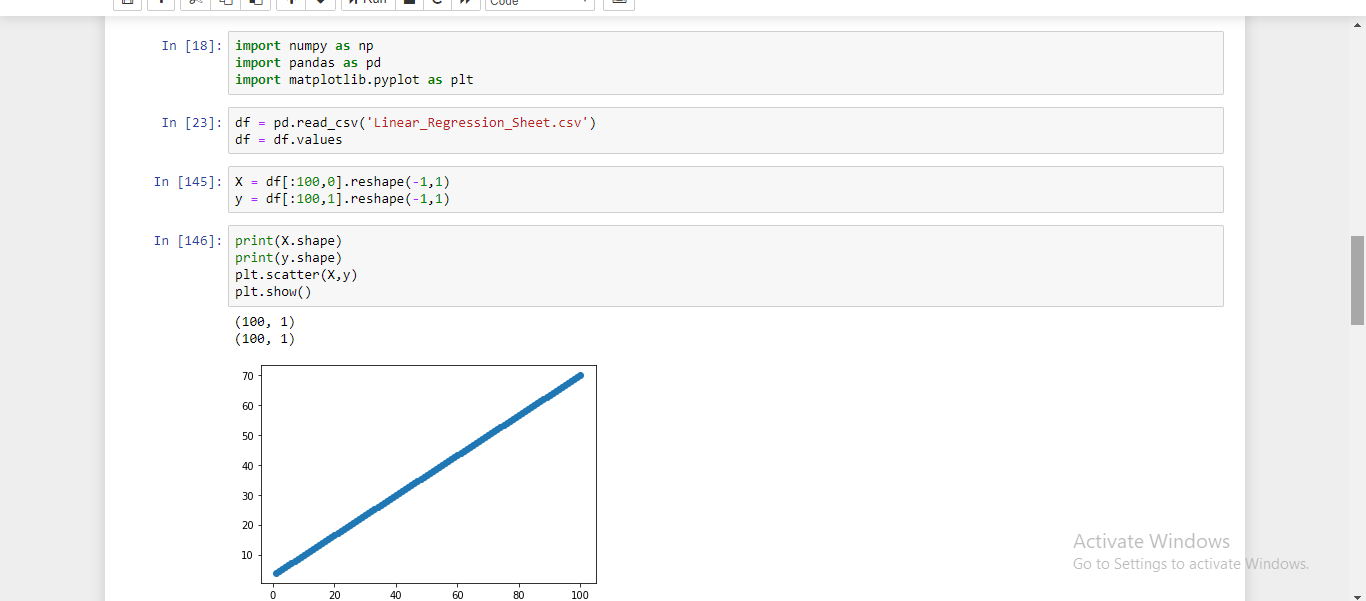
**CHAPTER 3 PROJECT AND METHODOLOGY ADOPTED**

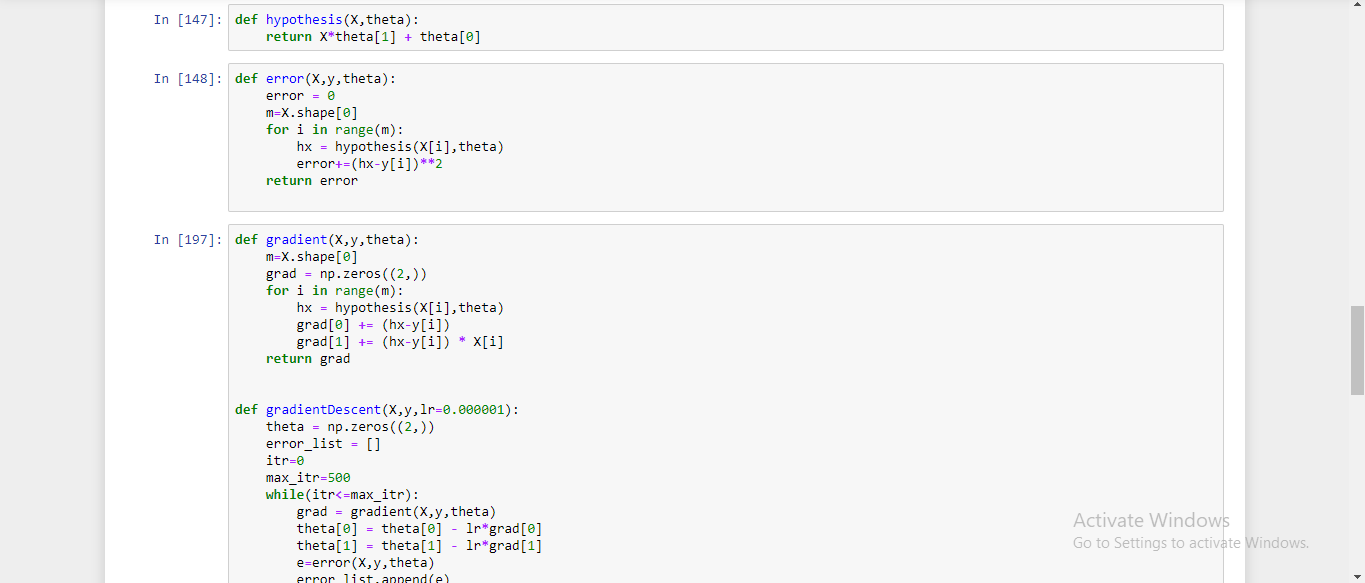
**Linear Regression**

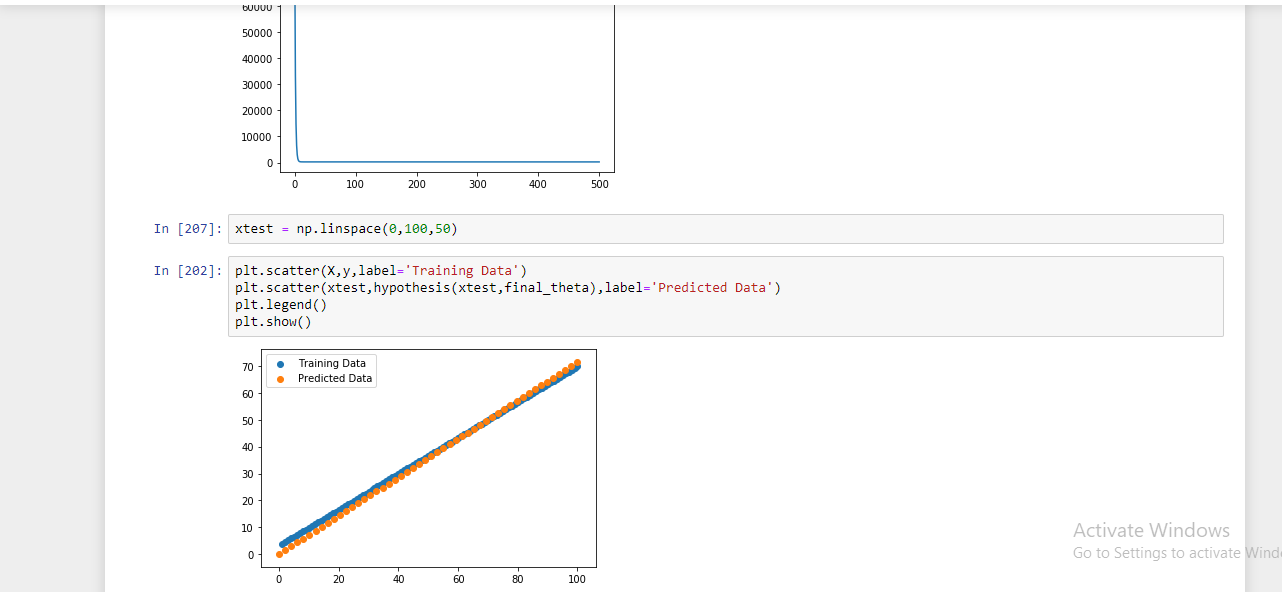
Linear regression is used for finding linear relationship between target and one or more predictors. There are two types of linear regression- Simple and Multiple.

# **Simple Linear Regression**

Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable. It looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other. For example, using temperature in degree Celsius it is possible to accurately predict Fahrenheit. Statistical relationship is not accurate in determining relationship between two variables. For example, relationship between height and weight.

The core idea is to obtain a line that best fits the data. The best fit line is the one for which total prediction error (all data points) are as small as possible. Error is the distance between the point to the regression line. 





**Challenge: Hardwork Pays Off**

In this challenge, you were given walkatime data of students and how they performed in the evaluation exam. The task is to predict the score you will get given the amount of time you spend on coding daily.

**Input** You are given one feature corresponding to time noted by walkatime.

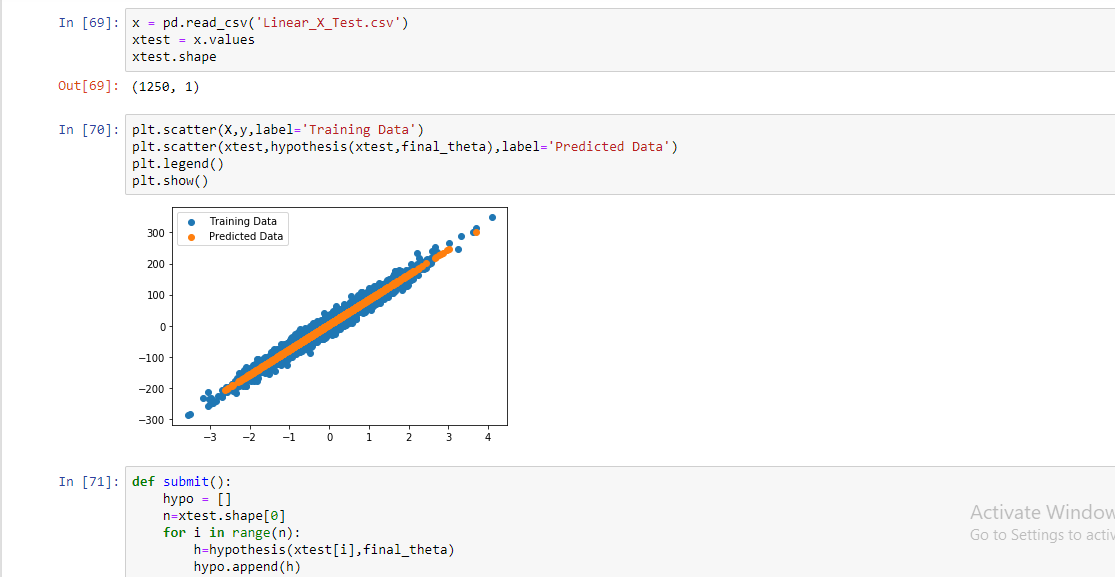
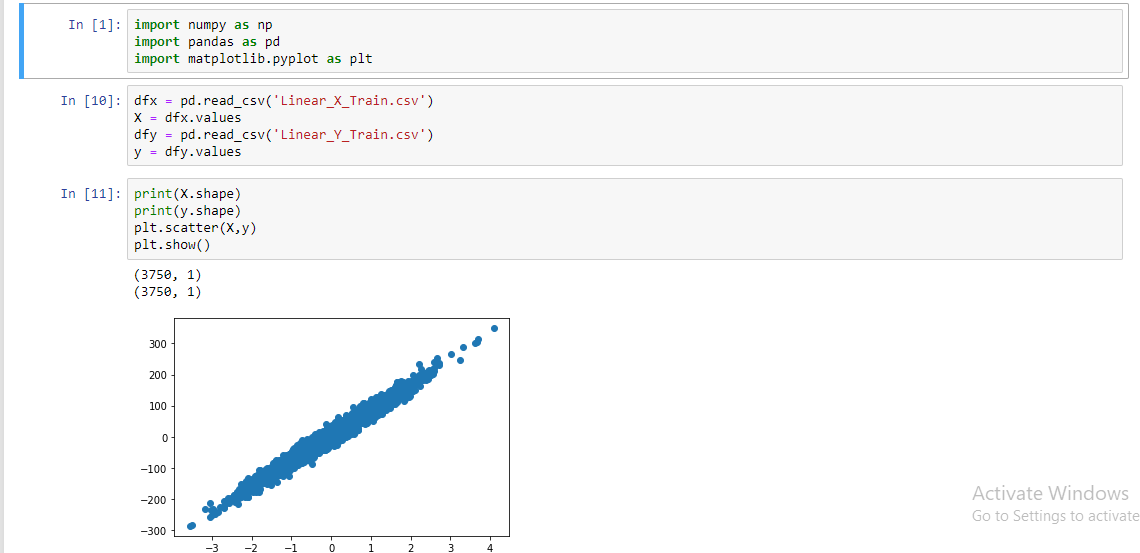
**Output** A scalar denoting the level of perfomance student achived by devoting the given time.

Figure1: sklearn Linear Regression

**Multivariate Regression**

Multivariate regression is a technique that estimates a single regression model with more than one outcome variable. When there is more than one predictor variable in a multivariate regression model, the model is a multivariate multiple regression.

### Boston Housing Prices

### The data was drawn from the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. The attributes are deﬁned as follows (taken from the UCI Machine Learning

### Repository1): CRIM: per capita crime rate by town 2. ZN: proportion of residential land zoned for lots over 25,000 sq.ft. 3. INDUS: proportion of non-retail business acres per town 4. CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise) 5. NOX: nitric oxides concentration (parts per 10 million) 1https://archive.ics.uci.edu/ml/datasets/Housing 123 20.2. Load the Dataset 124 6. RM: average number of rooms per dwelling 7. AGE: proportion of owner-occupied units built prior to 1940 8. DIS: weighted distances to ﬁve Boston employment centers 9. RAD: index of accessibility to radial highways 10. TAX: full-value property-tax rate per $10,000 11. PTRATIO: pupil-teacher ratio by town 12. B: 1000(Bk−0.63)2 where Bk is the proportion of blacks by town 13. LSTAT: % lower status of the population 14. MEDV: Median value of owner-occupied homes in $1000s We can see that the input attributes have a mixture of units.

### 

**K-Nearest Neighbours**

K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection.

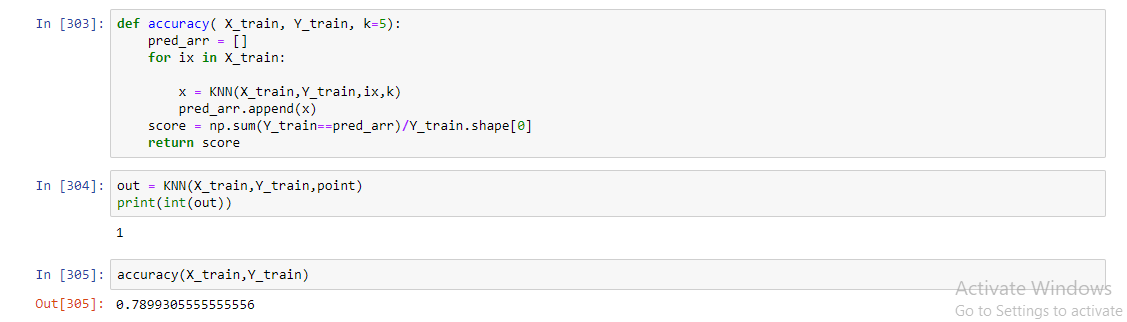
It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data).

We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.

**Diabetes Detection Challenge on KNN Implementation**

In this problem we are given a Diabetes Data set consisting of following features -['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome']and our task is to predict whether a person is suffering from diabetes or not (Binary Classification)

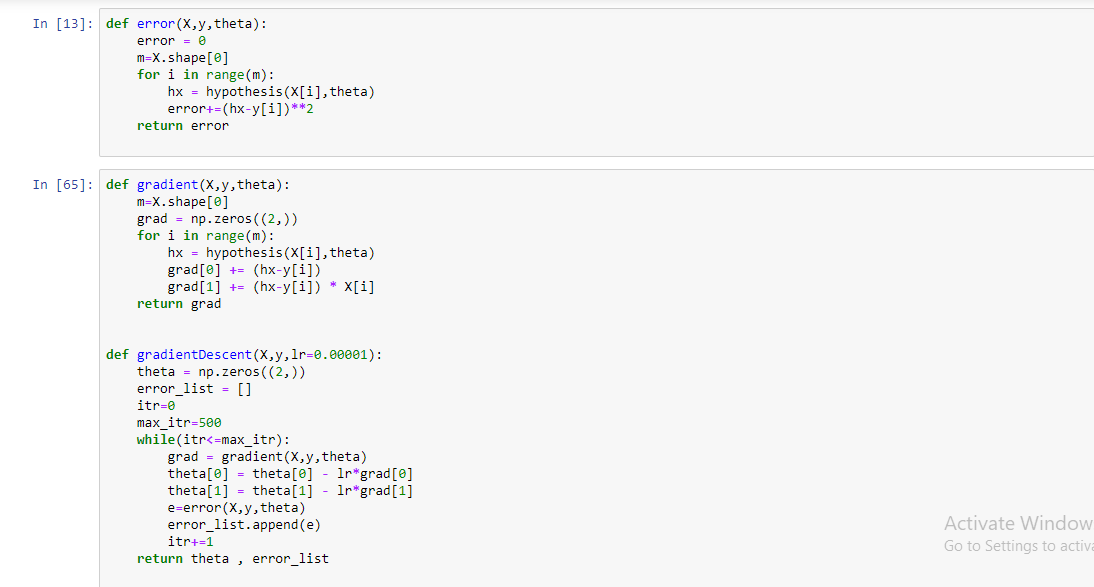
1) Plot a bar graph showing number of classes and no of examples in each class.

2) Classification Task, classify a person as 0 or 1 (Diabetic or Not) using K-Nearest Neighbors classifier.

**Gradient Descent**

**Task –** Implement Gradient descent on your own, i. e, its cost function, learning rate, and other parameters.

**Approach –** Linear Regression considers linear relationship between the input and the output variables whereas Gradient Descent traces a curve and based on its gradient tries to train the model. Gradient Descent brings an improvement into the cost function of linear regression and brings about more accuracy in its prediction. To find the gradient / i. e, optimal point, we trace the curve in its vicinity and try about to find a line that fits the best as a gradient line, i. e, a line for which the slope is increasing in its vicinity.



**Real Time Face Recognition using Haarcascade**

Face detection is one of the fundamental applications used in face recognition technology. Facebook, Amazon, Google and other tech companies have different implementations of it. Before they can recognize a face, their software must be able to detect it first. Amazon has developed a system of real time face detection and recognition using cameras. Facebook uses it mostly on photos that their users upload in order to suggest tagging friends.

## Requirements

* **Any operating system that will support OpenCV and** **Python**(**Windows, Linux, MacOS**)
* **Python**
* **OpenCV-Python**
* **Haar Cascades Data File**
* **Photo images for testing**

**CHAPTER-4: Result and Discussion**

As a result of the different Machine Learning concepts learnt during the training and the projects undertaken, I have learnt the following things –

1. How to handle classification Problems

2. Regression Analysis

3. Different Machine Learning Algorithms and their implementation

4. Applications of Machine Learning in the industrial sector

5. The linear regression project helped to learn how to handle the diabetes dataset and how Machine Learning can prove to be helpful in the health care sector. If we can check the symptoms of the patient beforehand and with the help of ML algorithms, we can check if he/she is likely to suffer from diabetes or any other disease, then we can help him take precautions so that he does not suffer from that disease.

6. The Gradient Descent project ( House Price Prediction) helped to make predictions about the market and learn about its dynamic nature and vulnerability. This could help the people to find out the areas where to buy their house, when to buy their house, when not to buy their house. For example, if a person is looking for a house, then we could help him find out the best house as per his/her budget and also the best suited area for him/her (e.g. area with less crime rate and good locality,etc.)

7. The Multivariate Regression project helped to understand the most important aspects of Machine Learning, i. e, Data Cleaning and handling null values. The regression project helps us to take the data from the past events and check whether they followed certain principles or not.

**CHAPTER-5: Conclusion and Future Scope**

A subset of artificial intelligence (AI), machine learning (ML) is the area of computational science that focuses on analysing and interpreting patterns and structures in data to enable learning, reasoning, and decision making outside of human interaction. Simply put, machine learning allows the user to feed a computer algorithm an immense amount of data and have the computer analyse and make data-driven recommendations and decisions based on only the input data. If any corrections are identified, the algorithm can incorporate that information to improve its future decision making.

In today’s world, Machine Learning finds uses in every field we can think of, either it be predicting the outcomes, smartphones, virtual personal assistants like Alexa and Siri. Every smart system in today’s world is trained using Machine Learning Algorithms and it is one of the most promising sub-fields in the field of computer science. Machine Learning has immensely helped in creating enhanced user experience like customized ads for every user through Google Adwords. Google Adwords uses Machine Learning to look into what kind of websites/services, a particular user uses and based on that, it recommends similar ads to the person. Similarly, the people you may know feature of Facebook also using Machine Learning to suggest us the people we should add.

Machine Learning is currently one of the most developing and opportunistic field. In terms of jobs, Machine learning promises to bring a lot of new jobs in the future. The global machine learning as a service (MLaaS) market is rising expeditiously mainly due to the Internet revolution. The process of connecting the world virtually has generated vast amount of data which is boosting the adoption of machine learning solutions. This is because deployment of machine learning improves the speed and accuracy of functions performed by the system.

Apart from this, adoption of advanced analytics technologies from several industry verticals such as healthcare and life sciences, BFSI, retail, telecom, and manufacturing is contributing towards the growth of machine learning as a service market. Machine learning as a service solutions are also adopted across industry verticals to enhance the decision making capability of machines.

Different job opportunities based websites like Indeed have termed Machine Learning as the biggest job producing field in the near future and currently, the average salary of a Machine Learning engineer lies at around 10-12 lakhs in India and the highest salary lies around 18-20 lakhs and the no of job opportunities in this field is growing at a very fast pace which establishes the fact that Machine Learning is the next big thing of the IT Industry

In terms of higher education also, Machine Learning is growing. In the premier institutes of India, Machine Learning is now offered as a specialization in the Masters courses which leads to the development of better Machine Learning engineers and better IT industry.

**REFERENCES**

1. <https://scikit-learn.org/stable/documentation.html> - Official sklearn documentation. Sklearn is a library in python that contains the inbuilt implementation of majority of Machine Learning Algorithms. It helps to directly use the Algorithms and carry out our work effortlessly.

2. <https://pandas.pydata.org/pandas-docs/stable/> - Official Pandas documentation – Pandas is a library that we use for data fetching and data manipulation. Pandas Series and Data frames help to fetch the data from the dataset and perform manipulations on it.

3. <https://www.numpy.org/devdocs/> - Numpy official documentation – Numpy is a library that is used to fetch and manipulate the dataset just like Pandas.

4. <https://matplotlib.org/users/index.html> – Matplotlib official documentation – It is a library that we use for data visualisation. Using it, we can visualise the data with the help of graphs, histograms, pie charts, etc.

5. <https://www.kaggle.com/> - kaggle is the home of data science. Kaggle is the world’s biggest website that hosts competitions on data science and machine learning. It also contains several problems that help us to learn the foundation of machine learning.