**CHAPTER 1**

**COMPANY PROFILE**

****

|  |  |
| --- | --- |
| **Company Name** | Airobotica Services Private Limited |
| **Company Status** | Active |
| **Company CIN** | U74999KA2017PTC107285 |
| **RoC** | RoC-Bangalore |
| **Registration Number** | 107285 |
| **Class of Company** | Private |
| **Date of Incorporation** | 08 October 2017 |
| **Email** | [airoboticaa@gmail.com](mailto:airoboticaa@gmail.com) |
| **Address** | Plot No.72&73, 4th Floor, Akshay Tech Park EPIP Area Hoodi Village, Whitefield Bangalore-560066, Karnataka, India. |
| **Phone Number** | +91 9010215496 |

* 1. **COMPANY SERVICES**
* Web Design
* Web Development
* Product Management
* Marketing
* Graphic Design and Automation

**1.2 DIRECTORS**

|  |  |  |  |
| --- | --- | --- | --- |
| **DIN** | **Name** | **Designation** | **Appointment date** |
| 07965638 | SHAIK IBRAHIM KHALEEL | Director | 18 October 2017 |
| 07969321 | SHAIK MD RASOOL | Director | 18 October 2017 |

**CHAPTER 2**

**ABOUT THE COMPANY**

**2.1 VISSION**

AIROBOTICA SERVICES PRIVATE LIMITED, is a Bengaluru based firm, the date of incorporation 18th October 2017 that strives to create an automated world by bridging the world of unexplored innovative opportunities. AIRobotica is a Software Company which provide solutions on AI, ML and Data Science to top MNC’S.

AIRobotica, a company which is striving to bridge the gap between humans and Artificial Intelligence. It offers exclusive internship programs in some of the most sought- after domains like Data Science, IoT, Web Development and many more.

**2.2 MISSION**

**Data science** is a field that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Thanks to data science technologies, platforms like Netflix curates highly targeted shows by extracting customer preference from their databases. It also customizes the watchlist by leveraging customer data and viewing habits.

**Internet of Things:** Any physical object connected to the Internet and has sensors that transmit data is called an IoT device. Knowingly or unknowingly, since decades we have all been using IoT devices like smart refrigerators, smart home alarm system, etc. As per a research done by Zion, the smart home automation market could grow to a whopping $53 billion by 2022. IoT is far bigger than anyone realizes.

**Web Development:** Web developers basically create and maintain websites. According to web credibility research from Stanford University, 75% of web users make judgements about a company’s credibility based on their website design. And it takes only 0.5 seconds for users to form an opinion about your website. Web development is that crucial right now.

AIRobotica provides Internships to get trained by experts and get hands-on experience in emerging domains like Data Science, IoT and many other.

**CHAPTER 3**

**INTRODUCTION**

**3.1 PYTHON**

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.

Python is an open-source programming language that is high level and works as a general-purpose language, it is most often compared to Ruby, JavaScript, and Scheme. What sets Python apart from other programming languages is that it is simple to use, can be taught to a beginner, can be embedded into any application, and can run on all current operating systems, including Mac, Windows, and Linux. It is also one of the most powerful languages a programmer can use and is about three and five times faster to code than JavaScript and C++, respectively.

* + 1. **Characteristics of Python:**

1. **It is an open-source language:** Python is freely available to the public. Python supports the **FLOSS (Free/Libre and Open-Source Software)** model, which means you can also change it and distribute it. This allows the Python community to tweak it and improve its features continuously.
2. **It is a high-level language:** Python is a high-level language, you need not remember its system architecture, not do you need to perform memory management. This feature contributes to Python’s user-friendliness.
3. **It is interpreted:** Unlike compiled languages like C++ and Java wherein you must compile the code and then run it, Python is an interpreted language. What this means is that instead of executing the source code all at once, Python executes it line by line. This makes it easier to debug a Python code because you can do it while writing the code.
4. **It is both object-oriented and functional:** An object-oriented programming language is one that can model real-world data, while a functional language focuses on functions (code that can be reused). Python supports both object-oriented and functional programming features. Also, unlike Java, Python supports multiple inheritances.
5. **It is portable:** Python is portable and highly flexible, meaning, a Python code written for a Windows machine or a Linux machine can also run-on iOS, and vice versa – you don’t need to make any alterations in the code
6. **It is extensible and embeddable:** Python is an extensible language, as it allows you to write specific parts of your Python code in other programming languages such as C++. Similarly, you can also embed your Python code in the source code of other languages. This allows you to integrate Python’s scripting functionalities into a code written in another language.
7. **It comes with a vast collection of libraries:** Python will automatically download the extensive collection of Python libraries with it. These libraries are built-in, so you don’t have to write individual code for every single thing. Python has libraries and packages for web browsers, threading, databases, regular expressions, image manipulation, documentation-generation, unit-testing, CGI, email, and much more.

**3.1.2 Applications of Python:**

Python supports **cross-platform operating systems** which makes building applications with it all the more convenient. Some of the globally known applications such as YouTube, BitTorrent, DropBox, etc. use Python to achieve their functionality.

1. **Web Development:** Python can be used to make web-applications at a rapid rate It is because of the frameworks Python uses to create these applications. There is *common-backend logic*that goes into making these frameworks and a number of libraries that can help integrate protocols such as HTTPS, FTP, SSL etc. and even help in the processing of JSON, XML, E-Mail and so much more.
2. **Game Development:** Python is also used in the development of interactive games. There are libraries such as PySoy which is a 3D game engine supporting Python 3, PyGame which provides functionality and a library for game development. Games such as Civilization-IV, Disney’s Toontown Online, Vega Strike etc. have been built using Python.
3. **Artificial Intelligence and Machine Learning:** Machine Learning and Artificial Intelligence are the talks of the town as they yield the most promising careers for the future. We make the computer learn based on past experiences through the data stored or better yet, create algorithms which makes the computer learn by itself. Python support for these domains with the **libraries** that exist already such as Pandas, Scikit-Learn, NumPy and so many more.
4. **Desktop GUI:** We use Python to program **desktop applications**. It provides the Tkinter library that can be used to develop user interfaces. There are some other useful toolkits such as the wxWidgets, Kivy, PYQT that can be used to create applications on several platforms.
5. **Enterprise-level/Business Applications:** Business Applications are different than our normal applications covering domains such as e-commerce, ERP and many more. They require applications which are scalable, extensible and easily readable and Python provides us with all these features. Platforms such as Tryton is available to develop such business applications.

**3.2 MACHINE LEARNING**

The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that “it gives computers the ability to learn without being explicitly programmed”. And in 1997, Tom Mitchell gave a “well-posed” mathematical and relational definition that “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Machine learning is a branch of artificial intelligence (AI) focused on building applications that learn from data and improve their accuracy over time without being programmed to do so.

In data science, an algorithm is a sequence of statistical processing steps. In machine learning, algorithms are 'trained' to find patterns and features in massive amounts of data in order to make decisions and predictions based on new data. The better the algorithm, the more accurate the decisions and predictions will become as it processes more data.

This encompasses basically all types of data science algorithms, supervised, unsupervised, segmentation, classification, or regression". Few important areas where machine learning can be applied are Handwriting Recognition, Language Translation, Speech Recognition, Image Classification, Autonomous Driving.

* + 1. **Classification of Machine learning:**

Machine learning implementations are classified into three major categories, depending on the nature of the learning “signal” or “response” available to a learning system which are as follows:

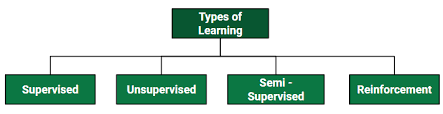
**

Fig. 3.1 Classifications of Machine Learning

1. **Supervised learning:** When an algorithm learns from example data and associated target responses that can consist of numeric values or string labels, such as classes or tags, in order to later predict the correct response when posed with new examples comes under the category of Supervised learning.

* **Image Classification:** You train with images/labels. Then in the future you give a new image expecting that the computer will recognize the new object.
* **Market Prediction/Regression:** You train the computer with historical market data and ask the computer to predict the new price in the future

1. **Unsupervised learning:** Whereas when an algorithm learns from plain examples without any associated response, leaving to the algorithm to determine the data patterns on its own. Unsupervised learning can be a goal in itself .
   * **Clustering:** You ask the computer to separate similar data into clusters, this is essential in research and science. Here, a set of inputs is to be divided into groups. Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task. As you can see in the example below, the given dataset points have been divided into groups identifiable by the colors red, green and blue.



Fig. 3.2 Clustering example

* + **High Dimension Visualization:** Use the computer to help us visualize high dimension data.
  + **Generative Models:** After a model captures the probability distribution of your input data, it will be able to generate more data. This can be very useful to make your classifier more robust.

1. **Reinforcement learning:** When you present the algorithm with examples that lack labels, as in unsupervised learning. A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). The program is provided feedback in terms of rewards and punishments as it navigates its problem space.
   * **Classification:** Inputs are divided into two or more classes, and the learner must produce a model that assigns unseen inputs to one or more (multi-label classification) of these classes. This is typically tackled in a supervised way. Spam filtering is an example of classification, where the inputs are email (or other) messages and the classes are “spam” and “not spam”.
   * **Regression:** It is also a supervised learning problem, but the outputs are continuous rather than discrete. For example, predicting the stock prices using historical data.



Fig. 3.3 Classification and Regression

1. **Semi-supervised learning:** where an incomplete training signal is given, a training set with some (often many) of the target outputs missing. Problems where you have a large amount of input data and only some of the data is labeled, are called semi-supervised learning problems. These problems sit in between both supervised and unsupervised learning. For example, a photo archive where only some of the images are labeled, (e.g. dog, cat, person) and the majority are unlabeled.
   * 1. **How machine learning works:**

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

* Model is the system that makes predictions.
* The parameters are the factors that 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.

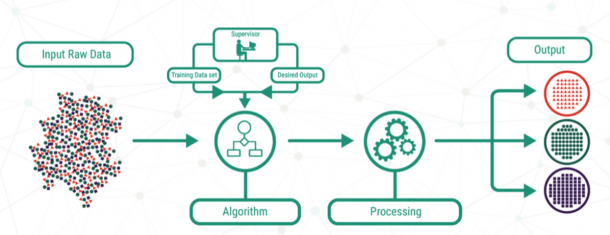


Fig. 3.4 Machine Learning Process

Let us build on the beer and wine example from above to understand how machine learning works. A machine learning model here has to predict if a drink is a beer or wine. The parameters selected are the colour of the drink and the alcohol percentage. The steps are:

* **Learning from the training set:**

This involves taking a sample data set of several drinks for which the colour and alcohol percentage are specified. Now, we have to define the description of each classification, that is wine and beer, in terms of the value of parameters for each type. The model can use the description to decide if a new drink is wine or beer.

You can represent the values of the parameters, ‘colour’ and ‘alcohol percentages’ as ‘x’ and ‘y’ respectively. Then (x, y) defines the parameters of each drink 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.

* **Measuring of 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.

* **Testing and Generalization:**

While it is possible for an algorithm or hypothesis to fit well into 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, generalization 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 generalize well. This is the case of over-fitting. In either case, the results are fed back to train the model further.

**3.2.3 Applications of Machine Learning:**

* Automatic Translation
* Email Spam Filtering
* Image Recognition
* Medical Diagnosis
* Online Fraud Detection
* Search Engine Results
* Text and Speech Recognition
* Traffic Prediction
* Virtual Personal Assistant

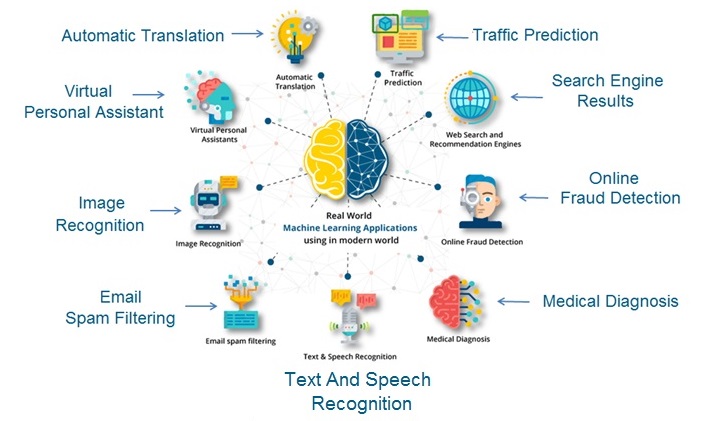
****

Fig. 3.5 Applications of Machine Learning

**3.3 ARTIFICIAL INTELLIGENCE**

“The science and engineering of making intelligent machines, especially intelligent computer programs.” - John McCarthy

Now if we talk about AI, it is completely a different thing from Machine learning and deep learning, actually deep learning and machine learning both are the subsets of AI. There is no fixed definition for AI, you will find a different definition everywhere, but here is a definition that will give you idea of what exactly AI is. “*AI is a ability of computer program to function like a human brain*”

AI means to actually replicate a human brain, the way a human brain thinks, works and functions. Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.

The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal. A subset of artificial intelligence is machine learning, which refers to the concept that computer programs can automatically learn from and adapt to new data without being assisted by humans. Deep learning techniques enable this automatic learning through the absorption of huge amounts of unstructured data such as text, images, or video.

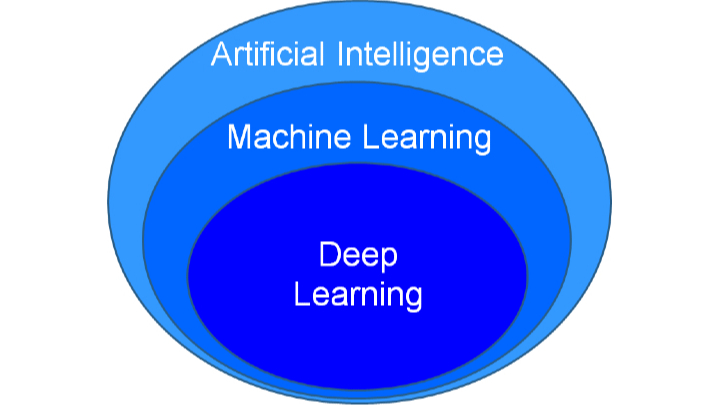


Fig. 3.6 Relationship between AI, ML and DL

The Intelligence is Intangible. It is composed of:

* Reasoning
* Learning
* Problem Solving
* Perception
* Linguistic Intelligence

The objectives of AI research are reasoning, knowledge representation, planning, learning, natural language processing, realization, and ability to move and manipulate objects. There are long-term goals in the general intelligence sector.

**3.3.1 Applications of AI:**

* **Gaming:** AI plays important role for machine to think of large number of possible positions based on deep knowledge in strategic games. for example, chess, river crossing, N-queens’ problems and etc.
* **Natural Language Processing:** Interact with the computer that understands natural language spoken by humans.
* **Expert Systems:** Machine or software provide explanation and advice to the users.
* **Vision Systems:** Systems understand, explain, and describe visual input on the computer.
* **Speech Recognition:** There are some AI based speech recognition systems have ability to hear and express as sentences and understand their meanings while a person talks to it. For example, Siri and Google assistant.
* **Handwriting Recognition:** The handwriting recognition software reads the text written on paper and recognize the shapes of the letters and convert it into editable text.
* **Intelligent Robots:** Robots are able to perform the instructions given by a human.
* **Healthcare:** AI has taken a critical step in helping people with looking after patients as well. The automated bots and healthcare applications ensure proper medication and treatment of patients in the facilities.
* **Banks:** AI for banking, which is of far higher value for banks, is in fraud detection. It can be hard for humans to understand patterns, but machines are good at it. This is where fraud prevention AI comes into play.

**3.4 DATA VISUALIZATION**

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets. The term is often used interchangeably with others, including information graphics, information visualization and statistical graphics.

Data visualization provides a quick and effective way to communicate information in a universal manner using visual information. The practice can also help businesses identify which factors affect customer behaviour; pinpoint areas that need to be improved or need more attention; make data more memorable for stakeholders; understand when and where to place specific products; and predict sales volumes.

The benefits of data visualization include the following:

* The ability to absorb information quickly, improve insights and make faster decisions.
* An increased understanding of the next steps that must be taken to improve the organization.
* An improved ability to maintain the audience's interest with information they can understand.
* An easy distribution of information that increases the opportunity to share insights with everyone involved.
* Eliminate the need for data scientists since data is more accessible and understandable.
* An increased ability to act on findings quickly and, therefore, achieve success with greater speed and less mistakes.

**3.5 DATA CLEANSING**

Data cleaning is one of the important parts of machine learning. It plays a significant part in building a model. However, proper data cleaning can make or break your project. Professional data scientists usually spend a very large portion of their time on this step.

Because of the belief that, “**Better data beats fancier algorithms**”. If we have a well-cleaned dataset, we can get desired results even with a very simple algorithm, which can prove very beneficial at times.

Obviously, different types of data will require different types of cleaning. However, this systematic approach can always serve as a good starting point.

****

Fig. 3.7 Data Cleansing steps

**3.5.1 Removal of unwanted observations:**

This includes deleting duplicate/ redundant or irrelevant values from your dataset. Duplicate observations most frequently arise during data collection and irrelevant observations are those that don’t actually fit the specific problem that you’re trying to solve.

* Redundant observations alter the efficiency by a great extent as the data repeats and may add towards the correct side or towards the incorrect side, thereby producing unfaithful results.
* Irrelevant observations are any type of data that is of no use to us and can be removed directly.

**3.5.2 Fixing Structural errors:**

The errors that arise during measurement, transfer of data or other similar situations are called structural errors. Structural errors include typos in the name of features, same attribute with different name, mislabeled classes, i.e., separate classes that should really be the same or inconsistent capitalization.

For example, the model will treat America and america as different classes or values, though they represent the same value or red, yellow and red-yellow as different classes or attributes, though one class can be included in other two classes. So, these are some structural errors that make our model inefficient and gives poor quality results.

**3.5.3 Managing Unwanted outliers:**

Outliers can cause problems with certain types of models. For example, linear regression models are less robust to outliers than decision tree models. Generally, we should not remove outliers until we have a legitimate reason to remove them. Sometimes, removing them improves performance, sometimes not. So, one must have a good reason to remove the outlier, such as suspicious measurements that are unlikely to be the part of real data.

**3.5.4 Handling missing data:**

Missing data is a deceptively tricky issue in machine learning. We cannot just ignore or remove the missing observation. They must be handled carefully as they can be an indication of something important. The two most common ways to deal with missing data are:

1. Dropping observations with missing values. Dropping missing values is sub-optimal because when you drop observations, you drop information.
   * The fact that the value was missing may be informative in itself.
   * Plus, in the real world, you often need to make predictions on new data even if some of the features are missing!
2. Imputing the missing values from past observations. Imputing missing values is sub-optimal because the value was originally missing but you filled it in, which always leads to a loss in information, no matter how sophisticated your imputation method is.
   * Again, “missingness” is almost always informative in itself, and you should tell your algorithm if a value was missing.
   * Even if you build a model to impute your values, you’re not adding any real information. You’re just reinforcing the patterns already provided by other features.

Both of these approaches are sub-optimal because dropping an observation means dropping information, thereby reducing data and imputing values also is sub-optimal as we fil the values that were not present in the actual dataset, which leads to a loss of information.

**3.6 MACHINE LEARNING MODEL**

A machine learning training model is a process in which a machine learning (ML) algorithm is fed with sufficient training data to learn from.

A training model is a dataset that is used to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that have an influence on the output. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output. The result from this correlation is used to modify the model. This iterative process is called “model fitting”. The accuracy of the training dataset or the validation dataset is critical for the precision of the model.

There are 7 primary steps involved in creating a machine learning model:

**3.6.1 Defining the problem:**

Defining the problem statement is the first step towards identifying what an ML model should achieve. This step also enables recognizing the appropriate inputs and their respective outputs; Questions like “what is the main objective?”, “what is the input data?” and “what is the model trying to predict?” must be answered at this stage.

**3.6.2 Data Collection:**

After defining the problem statement, it is necessary to investigate and gather data that can be used to feed the machine. This is an important stage in the process of creating an ML model because the quantity and quality of the data used will decide how effective the model is going to be. Data can be gathered from pre-existing databases or can be built from the scratch.

**3.6.3 Preparing the Data:**

The data preparation stage is when data is profiled, formatted and structured as needed to make it ready for training the model. This is the stage where the appropriate characteristics and attributes of data are selected. This stage is likely to have a direct impact on the execution time and results. This is also at the stage where data is categorized into two groups – one for training the ML model and the other for evaluating the model. Pre-processing of data by normalizing, eliminating duplicates and making error corrections is also carried out at this stage.

**3.6.4 Assigning Appropriate Model/ Protocols:**

Picking and assigning a model or protocol has to be done according to the objective that the ML model aims to achieve. There are several models to pick from, like linear regression, k-means and Bayesian. The choice of models largely depends on the type of data that is being used. For instance, image processing convolutional neural networks would be the ideal pick and k-means would work best for segmentation.

**3.6.5 Training the Machine Model or “The Model Training”:**

This is the stage where the ML algorithm is trained by feeding datasets. This is the stage where the learning takes place. Consistent training can significantly improve the prediction rate of the ML model. The weights of the model must be initialized randomly. This way the algorithm will learn to adjust the weights accordingly.

**3.6.6 Evaluating and Defining Measure of Success:**

The machine model will have to be tested against the “validation dataset”. This helps assess the accuracy of the model. Identifying the measures of success based on what the model is intended to achieve is critical for justifying correlation.

**3.6.7 Parameter Tuning:**

Selecting the correct parameter that will be modified to influence the ML model is key to attaining accurate correlation. The set of parameters that are selected based on their influence on the model architecture are called hyperparameters. The process of identifying the hyperparameters by tuning the model is called parameter tuning. The parameters for correlation should be clearly defined in a manner in which the point of diminishing returns for validation is as close to 100% accuracy as possible.

**3.7 DATASET**

“It is a collection of data records for computer processing”

A dataset is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the dataset in question. It lists values for each of the variables, such as height and weight of an object. Each value is known as a datum. The dataset may comprise data for one or more members, corresponding to the number of rows.

**3.7.1 CSV File:**

A CSV File (comma-separated values) is a text file that stores data in the form of columns, separated by commas, and rows are distinguished by line breaks. A CSV file is a special type of file that you can create or edit in Excel. Instead of storing information in columns, CSV files store data separated by commas. When text and numbers are saved in a CSV file, it is easy to move them from one program to another. For example, you can export Google contacts to a CSV file and then import them into Outlook.

It is a very simple way to represent the information. Normally it is used to import or export databases of some applications.

* 1. **TERMINOLOGIES**
* **Model:** A model is a specific representation learned from data by applying some machine learning algorithm. A model is also called hypothesis.
* **Feature:** A feature is an individual measurable property of our data. A set of numeric features can be conveniently described by a feature vector. Feature vectors are fed as input to the model. For example, in order to predict a fruit, there may be features like color, smell, taste, etc.

Note: Choosing informative, discriminating and independent features is a crucial step for effective algorithms. We generally employ a feature extractor to extract the relevant features from the raw data.

* **Target (Label):** A target variable or label is the value to be predicted by our model. For the fruit example discussed in the features section, the label with each set of input would be the name of the fruit like apple, orange, banana, etc.
* **Training:** The idea is to give a set of inputs (features) and its expected outputs (labels), so after training, we will have a model (hypothesis) that will then map new data to one of the categories trained on.
* **Prediction:** Once our model is ready, it can be fed a set of inputs to which it will provide a predicted output (label).
* **Preprocessing**: Preprocessing data is a first step in the workflow to prepare raw data in a format that the network can accept.
* **Overfitting:** A statistical model is said to be over fitted when we feed it a lot more data than necessary, which indicates that the model is too complex for the problem that it is solving.
* **Training Dataset**: Training dataset is to determine or learn the optimal combinations of variables that will generate a good predictive model.
* **Validation Dataset**: A validation dataset is a dataset of examples used to tune the hyperparameters (i.e. the architecture) of a classifier.
* **Testing Dataset**: A test dataset is a dataset that is independent of the training dataset, but that follows the same probability distribution as the training dataset. If a model fit to the training dataset also fits the test dataset well, minimal overfitting has taken place.
* **Loss and Accuracy:** Loss function is used to find error or deviation in the learning process. Some of the losses are mean square error, mean absolute error, hinge error etc.

Accuracy is defined as the proportion of correct predictions in all predictions made.

**CHAPTER 4**

**TASK PERFORMED**

I found my internship to be rewarding, fulfilling, challenging, and enjoyable. My internship had a duration of 4 Weeks. I had the opportunity to develop a console-based application and some tasks performed as shown in this chapter. It was also very exciting and interesting to interact with so many people who collaborate to make the tasks complete.

**4.1 WEEKLY PROGRESS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. NO** | **Week** | **Work Assigned** | **Work Completed** |
| 1. | Week 1 (25th July 2020 – 1st Aug 2020) | • Joining the Internship program understanding the company and its technology.  • Understand about basics of Python and Machine Learning Importance in Real World.  • Given assignment | • Understand about company and its technology.  • Made a fast Recap about Python.  • Understand the use of Machine Learning in Real World.  • Completion of given assignment. |
| 2. | Week 2 (3rd Aug 2020 – 10th Aug 2020) | • Introduction to Machine Learning concepts learning and understanding the libraries like sklearn, pandas, numpy, matplotlib., etc.  • Understanding the CSV file format and how to retrieve the data using python libraries.  • Given Article -“DataScience on HealthCare”. | • Understanding the concept of Machine Learning and Explored libraries like sklearn, pandas, numpy, matplotlib., etc.  • Learned CSV file for data retrieve using python libraries.  • Completion of Article -“DataScience on HealthCare”. |
| 3. | Week 3 (11th Aug 2020 – 18th Aug 2020) | • Understanding the ML Algorithms like Logistic Regression, KNN, Naïve Bayes, Random Forest, Decision Tree., etc.  • Understanding How to select the data for training and testing the ML Model.  • Understanding How the data is pre-processed and prepared before model training and testing.  • Given assignment. | • Learned the ML Algorithms like Logistic Regression, KNN, Naïve Bayes, Random Forest, Decision Tree., etc.  • Learned How to select the data for training and testing the ML Model.  • Learned How the data is pre-processed and prepared before model training and testing.  • Completion of given assignment. |
| 4. | Week 4 (19th Aug 2020 – 28th Aug 2020) | • Discuss the ideas for hands-on Project.  • To start the project Implementation | • Discussed the ideas for hands-on project.  • Implemented the project Successfully. |

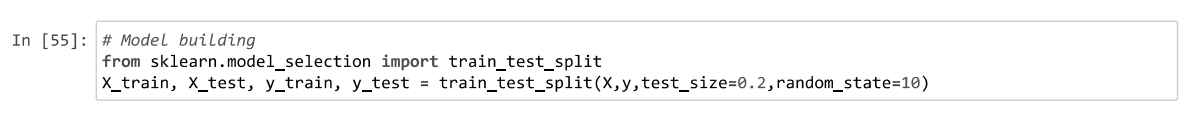
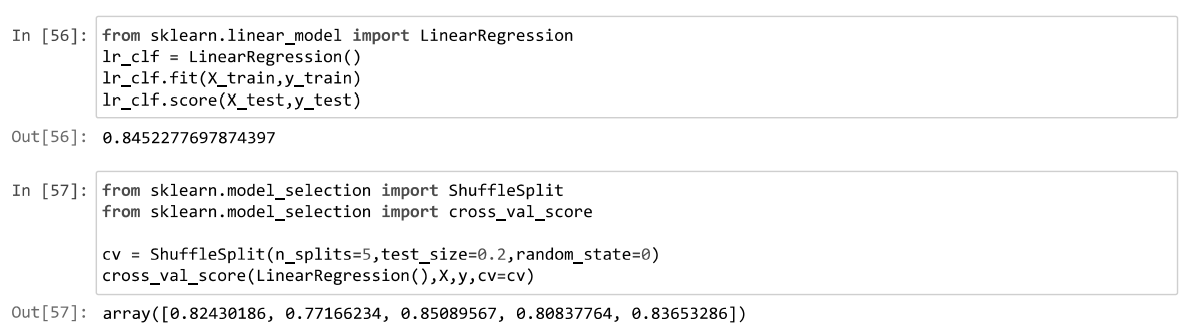
**4.2 PROJECT DESCRIPTION**

**Title: House Price Prediction**

* House Price Prediction is a web-based application that predicts the price of house.
* The Flask Framework is used to handle the HTTP requests from application.
* HTML, CSS and JavaScript is used for the Front End.
* This Machine Learning Project which is trained the model using the previously available data which have different locations in Bangalore.
* This model takes the Area (Square Feet), BHK, Bathrooms and Locations, as a parameter to predict the approximate price.

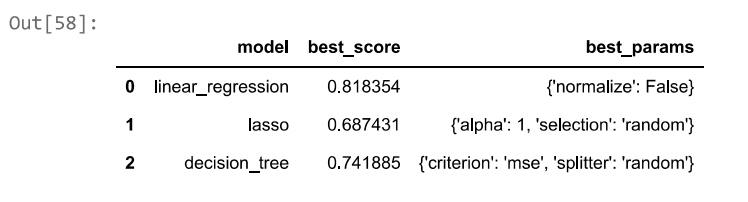
**4.2.1 CODE:**

**Model Training:**

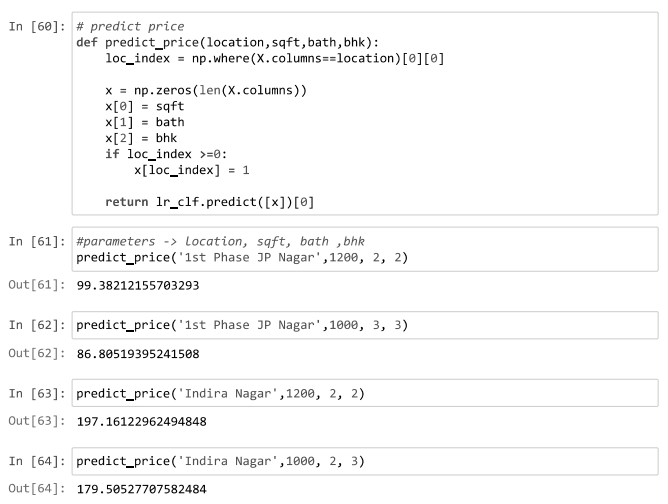


**Selecting The Best Model:**

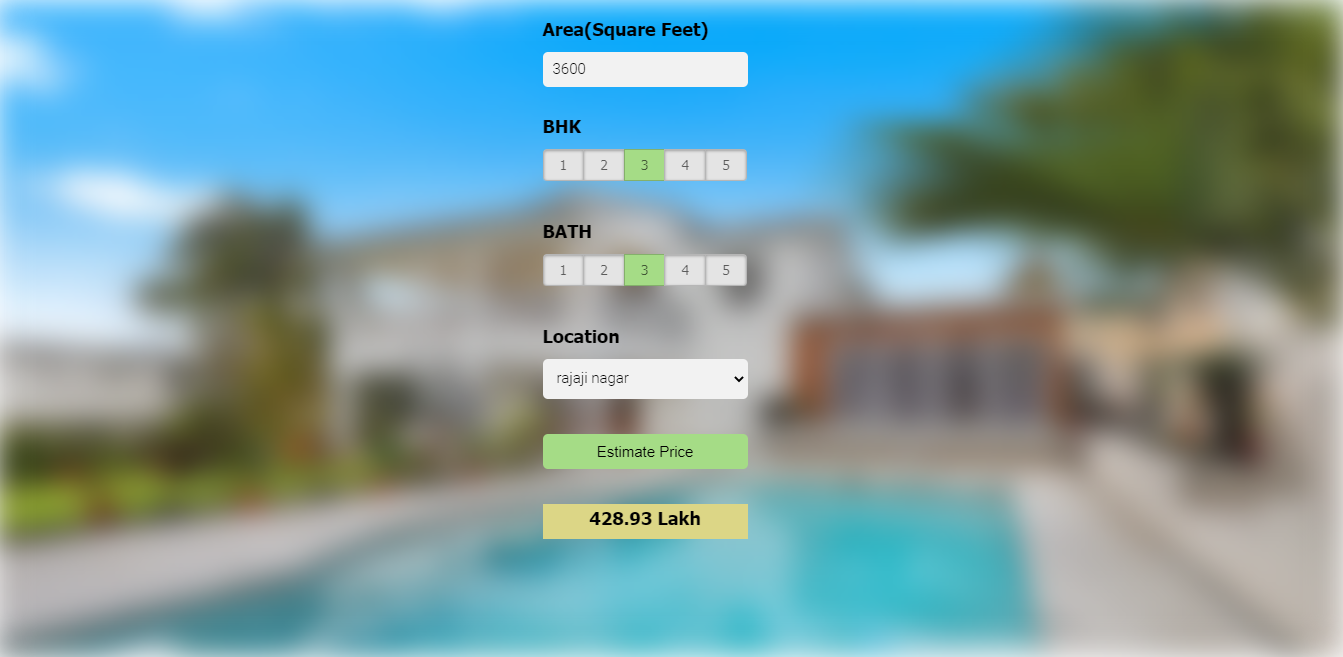




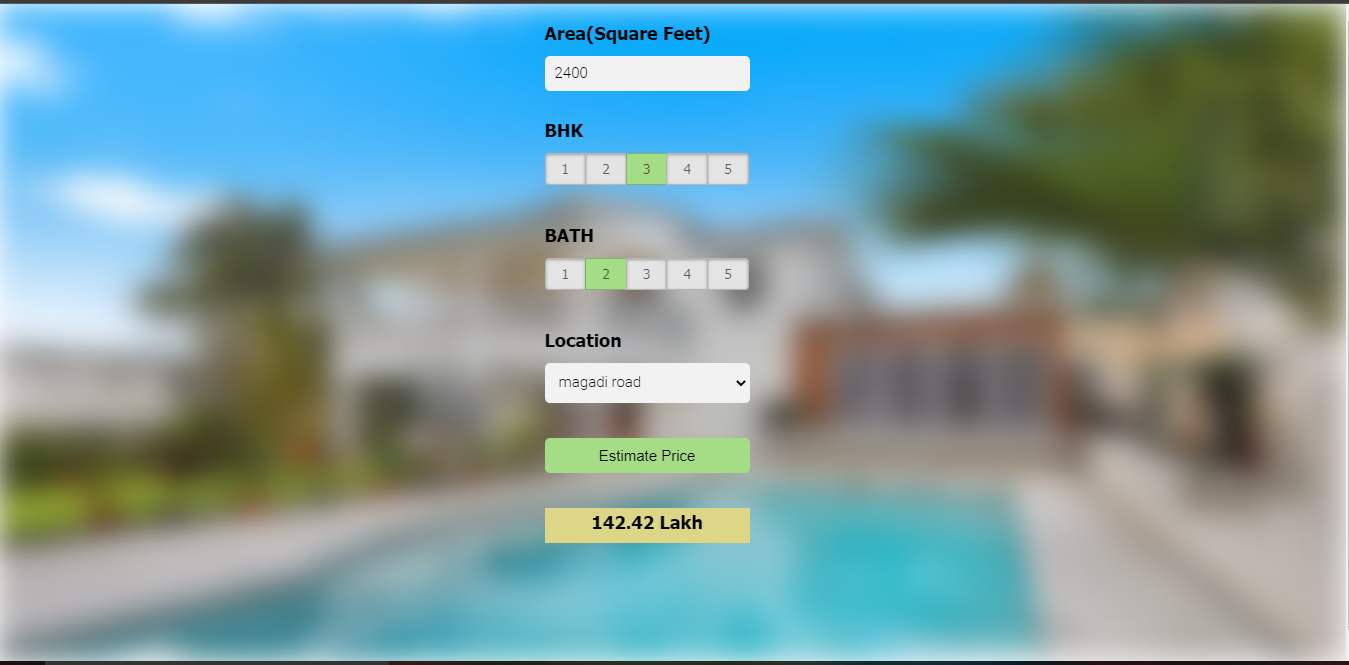
**Model Testing:**



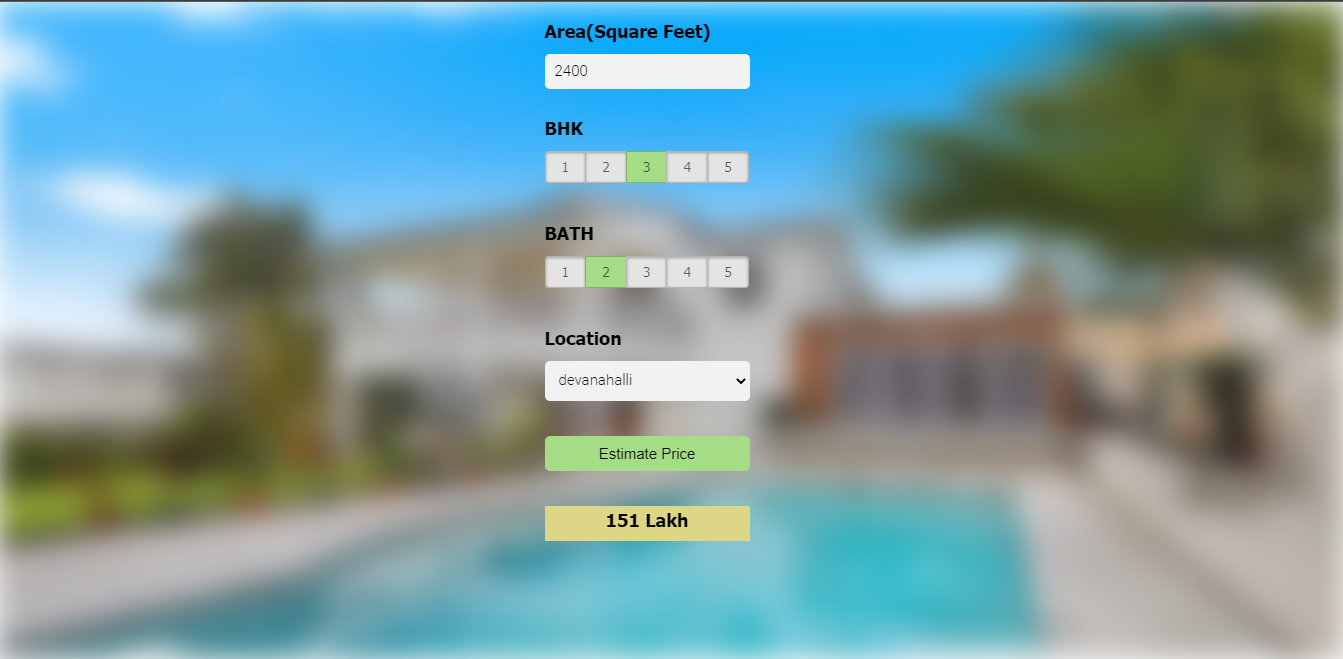
**Web App Snapshots**

**House Price Prediction at Rajaji Nagar:**

*fig 3.2.1 (A) House Price Prediction at Rajaji Nagar*

**House Price Prediction at Magadi Road:**

*fig 3.2.1 (B) House Price Prediction at Magadi Road*

**House Price Prediction at Devanahalli:**

*fig 3.2.1 (C) House Price Prediction at Devanahalli*

**CHAPTER 5**

**REFLECTION**

**5.1 Work Experience**

* The learning environment at AIROBOTICA is good.
* Mentor was very informative and helpful.
* Gained knowledge on Project development life cycle.
* Gained knowledge on how to use python, various libraries of python, how to code python for Machine Learning.
* Learned how scripts can be written in python.
* Learned how to effectively utilize the tools and resources to complete a task.
* Improved communication skills.
* Different case studies and problem statements helped in improving the problem solving and critical thinking skills.
* Improved presentation skills.
* Gained management skills.

**5.2 Problems/Challenges**

* The internship was much different from college academics and it takes little time to adjust to the environment and schedule.
* Initially, it is difficult to understand the problem statements.
* We learned how to use python, and how to use Machine learning, configure it.

**5.3 Technical Outcomes**

* Developing a technical artifact requiring new technical skills.
* Using profession-specific terminology appropriately.
* Effectively utilizing the tools and resources to complete a task.
* Creating training materials.
* Maintaining and troubleshooting technology.
* Analyzing or visualizing data to create information.
* Writing requirements documentation.
* Selecting appropriate technologies.
* Acquiring and evaluating information.

**5.4 Non-Technical Outcomes**

* Demonstration and understanding of professional customs and practices.
* Organizing and maintaining information.
* Applying knowledge to the task.
* Negotiating and arriving at a decision.
* Working with diverse populations.
* Identifying, understanding, and working with professional standards.
* Working in cross-cultural and/or multicultural settings.
* Improving problem-solving and critical thinking skills.
* Monitoring and correcting performance.
* Exercising leadership and behaving professionally, ethically, and effectively.
* Dressing appropriately and addressing colleagues and superiors appropriately.
* Adapting effectively to changing conditions.
* Participating as a member of a team.
* Developing appropriate workplace attitudes.
* Developing individual responsibility.

**5.5 Reason for choosing Machine Learning**

**5.5.1 Career Opportunities:**

* Machine learning is the shining star of the moment.
* Every industry looking to apply AI in their domain, studying machine learning opens world of opportunities to develop cutting edge machine learning applications in various verticals – such as cyber security, image recognition, medicine, or face recognition.
* Several machine learning companies on the verge of hiring skilled ML engineers, it is becoming the brain behind business intelligence.

**5.5.2 Jobs on the rise:**

* The major hiring is happening in all top tech companies in search of those special kind of people (machine learning engineers) who can build a hammer (machine learning algorithms).
* The job market for machine learning engineers is not just hot but it’s sizzling. Machine Learning Jobs on Indeed.com - 2,500+(India) & 12,000+(US)

**5.6 Benefits of Doing an Internship**

**5.6.1 Gain Valuable Work Experience:**

An internship provides the opportunity to gain hands-on work experience that is not possible to get in the classroom and also companies train interns and help in gaining the experience required to get a job.

**5.6.2 Have an Edge in the Job Market:**

Employers are usually more concerned with the work experience than the qualifications and internships are often the only way to get the work experience that is required to secure a job, so pursuing an internship has a vital role in the resume. Many employers prefer or require applicants to pursue internships or relevant work experience. Hence by having an internship experience the resume gets weightage.

**5.6.3 Transition into a Job:**

Employers see interns as prospective employees, so by performing well one can finish internships and continue working with the company full time. Internships are the number one way for employers to find new staff and employees to find a new job with experience.

**5.6.4 Machine Learning Opportunities:**

Internships are a great way to meet people in a specific field of our interest. An internship allows meeting people who might help in getting a job later on and give the contacts of the industry to which we are interested and break into it. Plus, references from people in the industry will really add weight to the application.

**CHAPTER 6**

**CONCLUSION**

Machine Learning is evolving into a self-supporting discipline and producing professionals with distinct and complementary skills relative to professionals in computer information. However, regardless of its potential eventual disciplinary status, the evidence points to the robust growth of Machine Learning that will indelibly shape the undergraduate students of the future. In fact, fueled by growing student interest and industry demand, machine learning will likely become a staple of the undergraduate experience. ML can be used to predict and draws insights; it has a wide range of applications and is a demanding skill forever in the IT Industry. It is essential that both processes are sustained as new pathways emerge at institutions.

During the internship, the Fundamentals of ML to Understanding advanced concepts, top libraries, and using them to bring solutions to the problem was learned effectively.

Machine learning has the great ability to revolutionize the diabetes risk prediction with the help of advanced computational methods and availability of large amount of epidemiological and genetic diabetes risk dataset. Detection of diabetes in its early stages is the key for treatment. This work has described a machine learning approach to predicting diabetes levels. The technique may also help researchers to develop an accurate and effective tool that will reach at the table of clinicians to help them make better decision about the disease status.

**REFERENCES**

1. www.wikipedia.com
2. https://www.ieee.org/
3. https://machinelearningmastery.com
4. https://www.kaggle.com/
5. Deepti Sisodia et al., **“Prediction of Diabetes using Classification Algorithms”**, *International Conference on Computational Intelligence and Data Science (ICCIDS),* pp. 1578–1585, 2018.
6. S. Dewangan.et.al., **“Diabetes Prediction Using Machine Learning Techniques”,** *Int. Journal of Engineering Research and Application,* pp.-09-13, 2018.