CHAPTER -1

INTRODUCTION

1.1 INTRODUCTION

Computational thinking is now a required course in almost all educational institutions throughout the world. Teaching programming skills is one of the most common ways to teach computational thinking. Numerous studies have demonstrated that programming, or coding as it is more often known, aids not only logical thinking and improved math achievement, but also problem solving, verbal and written skills.Real computational thinking, on the other hand, is more than just learning to code; it's a fundamental analytical skill that can be used to solve issues, design and create systems, and much more.

Most educational institutions around the world have adopted computational thinking as a mandatory course for its students. One of the most common approaches used to teach computational thinking is by teaching programming skills .Numerous studies have shown how programming or coding, as it is better known, not only helps with logical thinking and an increased performance in Math but also in problem solving, verbal and written skills .

CHAPTER -2

LITERATURE SURVEY

2.1 LITERATURE SURVEY

LITERATURE SURVEY 1:

TITLE : “Computational Thinking: What and Why?”

YEAR : 2010

AUTHOR : Jeannette M. Wing

ABSTRACT **:**

Computational thinking is used in the design and analysis of problems and their solutions, broadly interpreted. The most important and high-level thought process in computational thinking is the abstraction process. Abstraction is used in defining patterns, generalizing from instances, and parameterization. It is used to let one object stand for many. It is used to capture essential properties common to a set of objects while hiding irrelevant distinctions among them. For example, an algorithm is an abstraction of a process that takes inputs, executes a sequence of steps, and produces outputs to satisfy a desired goal. An abstract data type defines an abstract set of values and operations for manipulating those values, hiding the actual representation of the values from the user of the abstract data type. Designing efficient algorithms inherently involves designing abstract data types. Abstraction gives us the power to scale and deal with complexity

LITERATURE SURVEY 2 :

TITLE :“Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis”

YEAR : 2015

AUTHORS : Maya Israel, Jamie N. Pearson, Tanya Tapia, Quentin M. Wherfel, George Reese

ABSTRACT :

The purpose of this study was to investigate how elementary school teachers with limited computer science experience in a high-need school integrated computational thinking into their instruction. The researchers conducted a cross-case analysis across different instructional contexts (e.g., general education classrooms, library, art) that included multiple observations and interviews over four months. Major themes included: (a) a wide range of implementation models emerged depending on teaching contexts, (b) ongoing professional development and embedded coaching resulted in increasing participation in computing education, (c) teachers and administrators viewed barriers to implementing computing from a problem solving framework, and (d) struggling learners, including students with disabilities and those living in poverty, benefitted from computing education that included scaffolding, modeling, and peer collaboration.

LITERATURE SURVEY 3:

TITLE : “Experiential Education through Project Based Learning”

YEAR : 2010

AUTHORS : Manos Antonakakis, Tim April, Michael Bailey, Matthew Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, Yi Zhou

ABSTRACT :

Experiential learning is the key factor of acquiring knowledge through experiencing things. It addresses specific teaching methods, which are believed to achieve a beneficial outcome to the learning ability of students. Project Based Learning is such a modern teaching method. The core idea of Project Based Learning is to connect student's experiences with school life and to provoke serious thinking as students acquire new knowledge. While there are some negative implications related to PBL, the method can leverage the advantages of modern teaching techniques.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

CNN / Neural Networks used in the Existing system of the problem statement

3.2 DRAWBACKS OF EXISTING SYSTEM

- Less amount of accuracy score
- Small level data-set
- Applicable on small level prediction work

3.3 PROPOSED SYSTEM

We used of machine learning in the model of system like KNN(k- nearest neigbhours). KNN comes out with the best results between 80%-90%.

3.4 ADVANTAGES OF PROPOSED SYSTEM

- Increasing the accuracy score
- Large amount of feature we are taking for the training and testing.

3.5 SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS :

- OS – Windows 7, 8 and 10 (32 and 64 bit)
- RAM – 4GB
-

SOFTWARE REQUIREMENTS :

- Anaconda Navigator
- in Python Language
- Jupyter Notebook

PYTHON OVERVIEW

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- **Python is Interpreted:** Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive:** You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented:** Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language:** Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, Unix shell, and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

Python Features

Python's features include:

Easy-to-learn: Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.

Easy-to-read: Python code is more clearly defined and visible to the eyes.

Easy-to-maintain: Python's source code is fairly easy-to-maintain.

A broad standard library: Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.

Interactive Mode: Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.

Portable: Python can run on a wide variety of hardware platforms and has the same interface on all platforms.

Extendable: You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.

Databases: Python provides interfaces to all major commercial databases.

GUI Programming: Python supports GUI applications that can be created and ported to many system calls, libraries, and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.

Scalable: Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below:

- IT supports functional and structured programming methods as well as OOP.

- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- IT supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Python is available on a wide variety of platforms including Linux and Mac OS X. Let's understand how to set up our Python environment.

ANACONDA NAVIGATOR

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. It is available for Windows, mac OS and Linux.

Why use Navigator?

In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages, and use multiple environments to separate these different versions.

The command line program conda is both a package manager and an environment manager, to help data scientists ensure that each version of each package has all the dependencies it requires and works correctly.

Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages and update them, all inside Navigator.

WHAT APPLICATIONS CAN I ACCESS USING NAVIGATOR?

The following applications are available by default in Navigator:

- Jupyter Lab
- Jupyter Notebook
- QT Console
- Spyder
- VS Code
- Glue viz
- Orange 3 App
- Rodeo
- RStudio

Advanced conda users can also build your own Navigator applications

How can I run code with Navigator?

The simplest way is with Spyder. From the Navigator Home tab, click Spyder, and write and execute your code.

You can also use Jupyter Notebooks the same way. Jupyter Notebooks are an increasingly popular system that combine your code, descriptive text, output, images and interactive interfaces into a single notebook file that is edited, viewed and used in a web browser.

What's new in 1.9?

- Add support for **Offline Mode** for all environment related actions.
- Add support for custom configuration of main windows links.

Numerous bug fixes and performance enhancements.

3.6 FEASIBILITY STUDY

Feasibility study in the sense it's a practical approach of implementing the proposed model of system . Here for a machine learning projects .we generally collect the input from online websites and filter the input data and visualize them in graphical format and then the data is divided for training and testing . That training is testing data is given to the algorithms to predict the data .

1. First, we take dataset.

2. Filter dataset according to requirements and create a new dataset which has attribute according to analysis to be done

3. Perform Pre-Processing on the dataset

4. Split the data into training and testing

5. Train the model with training data then analyze testing dataset over classification algorithm

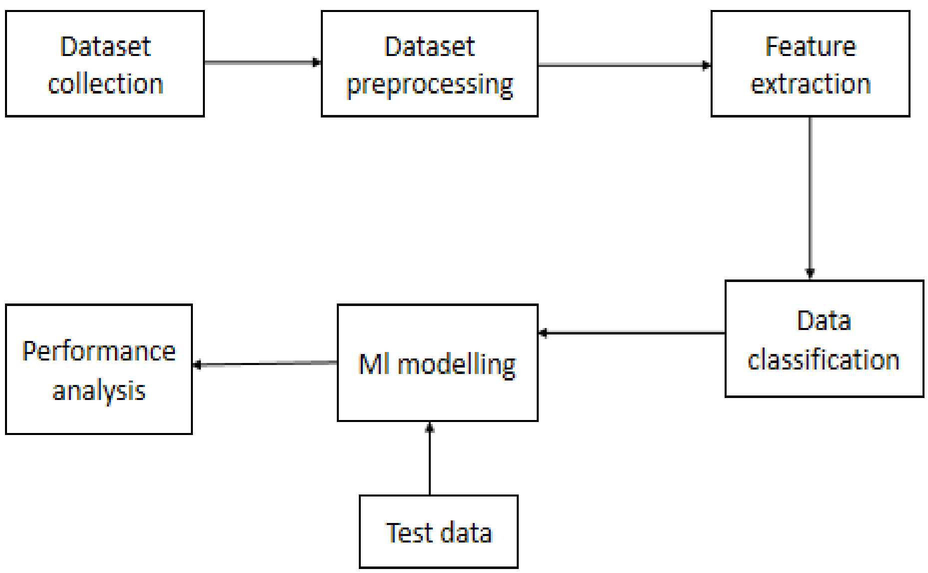
6. Finally you will get results as accuracy metrics.

7. Atlast we will creating web page using flask package with pickle python library file.

CHAPTER-4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE



4.2 MODULES

1. **DATA COLLECTION**
2. **DATA PRE-PROCESSING**
3. **FEATURE EXTRATION**
4. **EVALUATION MODE**

1. DATA COLLECTION

Data collection is a process in which information is gathered from many sources which is later used to develop the machine learning models. The data should be stored in a way that makes sense for problem. In this step the data set is converted into the understandable format which can be fed into machine learning models.

Data used in this paper is a set of cervical cancer data with 15 features . This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called *labelled data*.

2.DATA PRE-PROCESSING

Organize your selected data by formatting, cleaning and sampling from it.

Three common data pre-processing steps are:

Formatting: The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

Cleaning: Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonymized or removed from the data entirely.Sampling: There may be far more selected data available than you need to work with.

More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

3. FEATURE EXTRATION

Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Finally, our models are trained using Classifier algorithm. We use classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen classifiers were Random forest. These algorithms are very popular in text classification tasks.

4.EVALUATION MODEL

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over fitted models. There are two methods of evaluating models in data science, Hold-Out and Cross-Validation. To avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance.

Performance of each classification model is estimated base on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs.

Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

4.3 UML DIAGRAMS

The Unified Modeling Language (UML) is used to specify, visualize, modify, construct and document the artifacts of an object-oriented software intensive system under development. UML offers a standard way to visualize a system's architectural blueprints, including elements such as:

- actors
- business processes
- (logical) components
- activities
- programming language statements
- database schemas, and
- Reusable software components.

UML combines best techniques from data modeling (entity relationship diagrams), business modeling (work flows), object modeling, and component modeling. It can be used with all processes, throughout the software development life cycle, and across different implementation technologies· UML has synthesized the notations of the Booch method, the Object-modeling technique (OMT) and Object-oriented software engineering (OOSE) by fusing them into a single, common and widely usable modeling language. UML aims to be a standard modeling language which can model concurrent and distributed systems.

USECASE DIAGRAM:

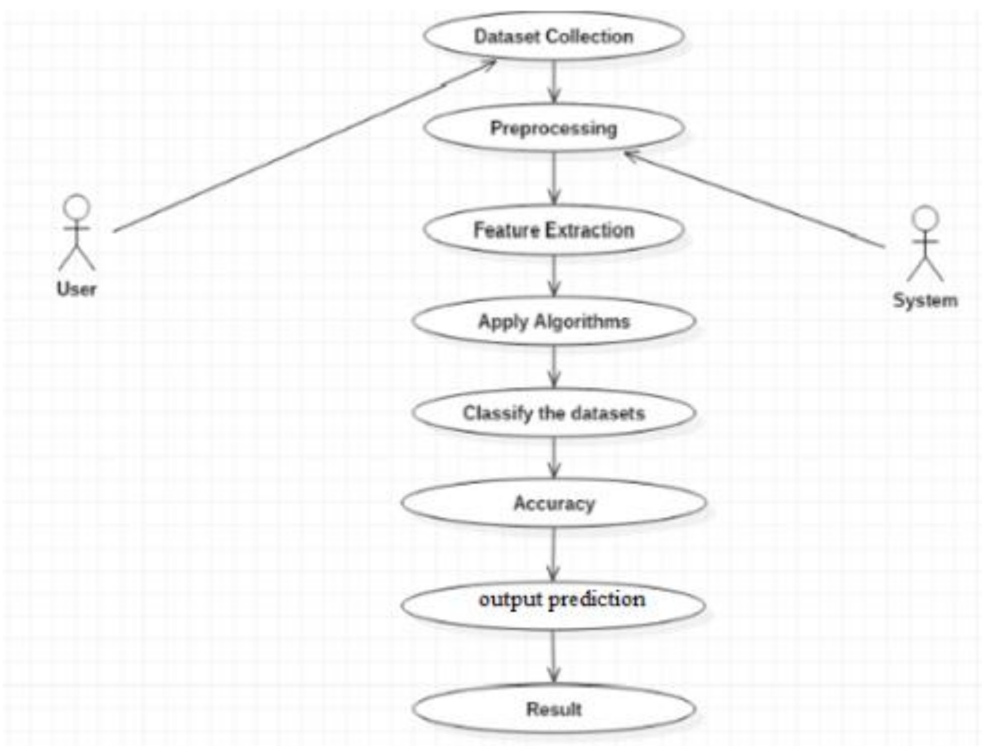
UML is a standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems.

UML was created by Object Management Group (OMG) and UML 1.0 specification draft was proposed to the OMG in January 1997.

OMG is continuously putting effort to make a truly industry standard.

UML stands for Unified Modeling Language.

UML is a pictorial language used to make software blue prints



CLASS DIAGRAM

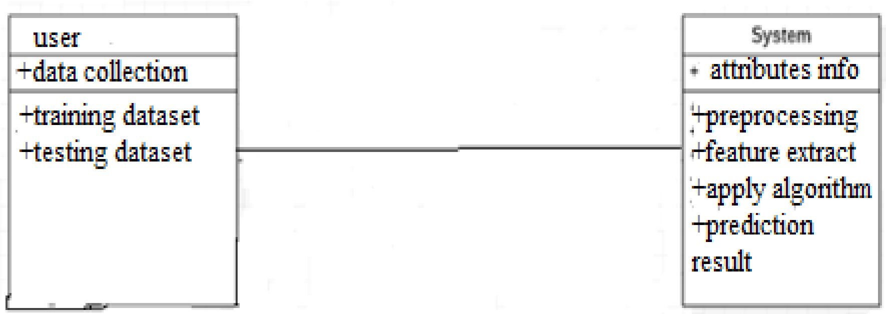
The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling.[1] The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

In the diagram, classes are represented with boxes that contain three compartments:

The top compartment contains the name of the class. It is printed in bold and centered, and the first letter is capitalized.

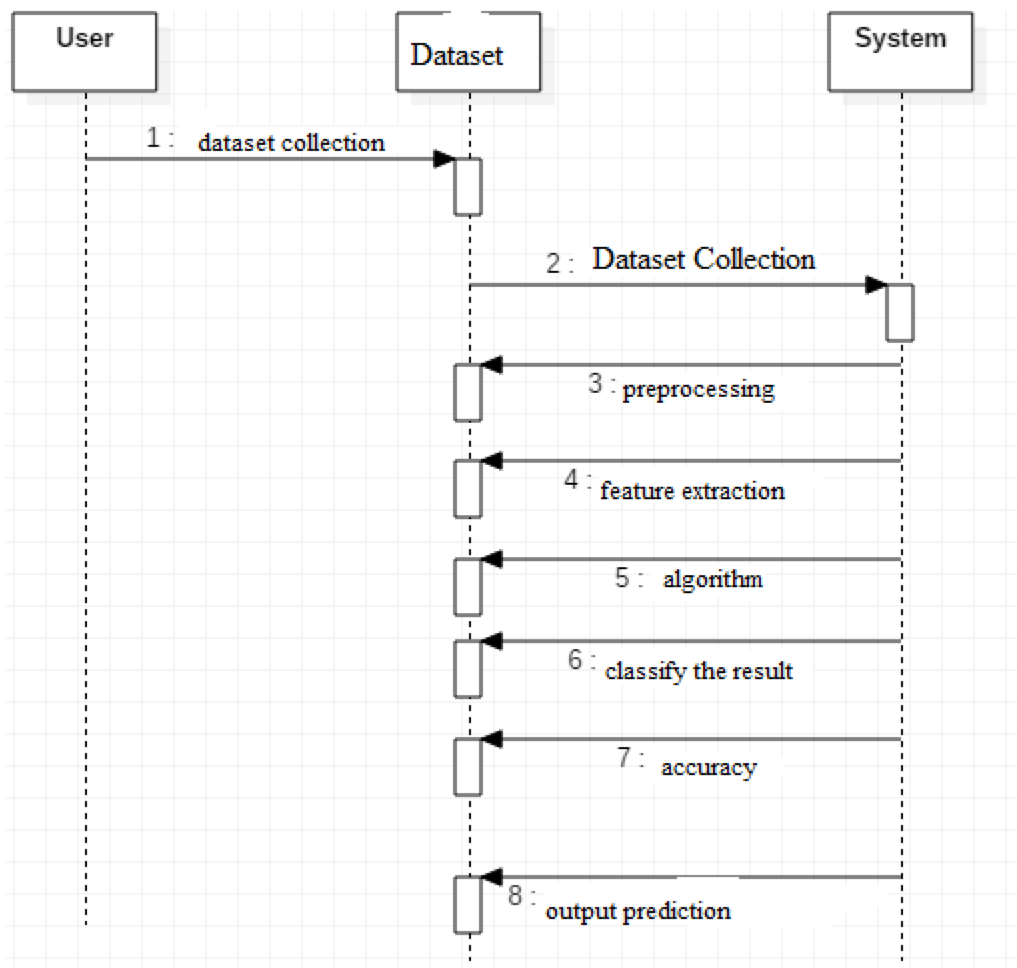
The middle compartment contains the attributes of the class. They are left-aligned and the first letter is lowercase.

The bottom compartment contains the operations the class can execute. They are also left-aligned and the first letter is lowercase.



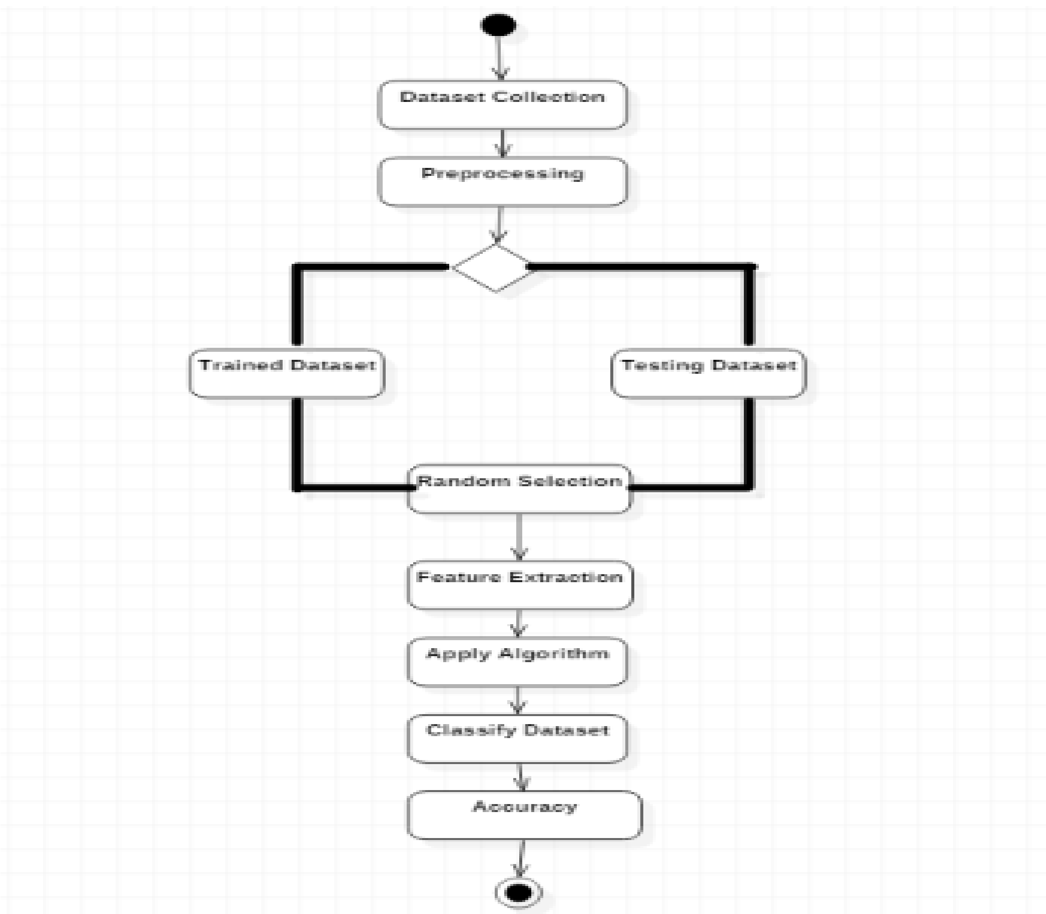
SEQUENCE DIAGRAM:

Sequence Diagrams Represent the objects participating the interaction horizontally and time vertically. A Use Case is a kind of behavioral classifier that represents a declaration of an offered behavior. Each use case specifies some behavior, possibly including variants that the subject can perform in collaboration with one or more actors. Use cases define the offered behavior of the subject without reference to its internal structure. These behaviors, involving interactions between the actor and the subject, may result in changes to the state of the subject and communications with its environment. A use case can include possible variations of its basic behavior, including exceptional behavior and error handling.



ACTIVITY DIAGRAMS-:

Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



CHAPTER-5

SYSTEM IMPLEMENTATION

5.1 TECHNOLOGIES

TensorFlow

the most famous deep learning library in the world is Google's TensorFlow. Google product uses machine learning in all of its products to improve the search engine, translation, image captioning or recommendations.

To give a concrete example, Google users can experience a faster and more refined the search with AI. If the user types a keyword a the search bar, Google provides a recommendation about what could be the next word.

Google wants to use machine learning to take advantage of their massive datasets to give users the best experience. Three different groups use machine learning:

- Researchers
- Data scientists
- Programmers.

They can all use the same toolset to collaborate with each other and improve their efficiency.

Google does not just have any data; they have the world's most massive computer, so TensorFlow was built to scale. TensorFlow is a library developed by the Google Brain Team to accelerate machine learning and deep neural network research.

It was built to run on multiple CPUs or GPUs and even mobile operating systems, and it has several wrappers in several languages like Python, C++ or Java.

In this tutorial, you will learn

TensorFlow Architecture

Tensor flow architecture works in three parts:

- Pre processing the data
- Build the model
- Train and estimate the model

It is called Tensor flow because it takes input as a multi-dimensional array, also known as **tensors**. You can construct a sort of **flowchart** of operations (called a Graph) that you want to perform on that input. The input goes in at one end, and then it flows through this system of multiple operations and comes out the other end as output.

This is why it is called TensorFlow because the tensor goes in it flows through a list of operations, and then it comes out the other side.

Where can Tensor flow run?

TensorFlow can hardware, and software requirements can be classified into

Development Phase: This is when you train the mode. Training is usually done on your Desktop or laptop.

Run Phase or Inference Phase: Once training is done Tensorflow can be run on many different platforms. You can run it on

- Desktop running Windows, macOS or Linux
- Cloud as a web service
- Mobile devices like iOS and Android

You can train it on multiple machines then you can run it on a different machine, once you have the trained model.

The model can be trained and used on GPUs as well as CPUs. GPUs were initially designed for video games. In late 2010, Stanford researchers found that GPU was also very good at matrix operations and algebra so that it makes them very fast for doing these kinds of calculations. Deep learning relies on a lot of matrix multiplication. TensorFlow is very fast at computing the matrix multiplication because it is written in C++. Although it is implemented in C++, TensorFlow can be accessed and controlled by other languages mainly, Python.

Finally, a significant feature of Tensor Flow is the Tensor Board. The Tensor Board enables to monitor graphically and visually what TensorFlow is doing.

List of Prominent Algorithms supported by TensorFlow

- Linear regression: tf. estimator .Linear Regressor
- Classification :tf. Estimator .Linear Classifier
- Deep learning classification: tf. estimator. DNN Classifier
- Booster tree regression: tf.estimator.BoostedTreesRegressor
- Boosted tree classification: tf.estimator.BoostedTreesClassifier

5.2 ALGORITHM

Domain Specification

MACHINE LEARNING

Machine Learning is a system that can learn from example through self-improvement and without being explicitly coded by programmer. The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results.

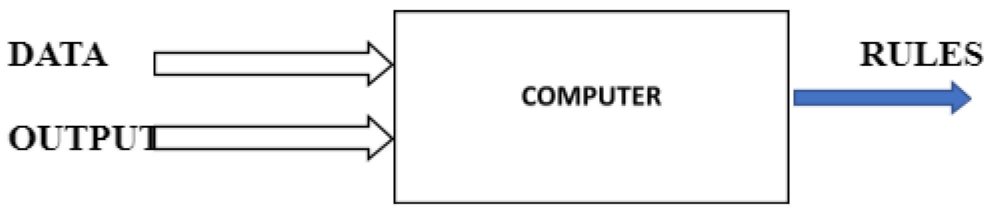
Machine learning combines data with statistical tools to predict an output. This output is then used by corporate to makes actionable insights. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input, use an algorithm to formulate answers.

A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation.

Machine learning is also used for a variety of task like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.

Machine Learning vs. Traditional Programming

Traditional programming differs significantly from machine learning. In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.



Machine Learning

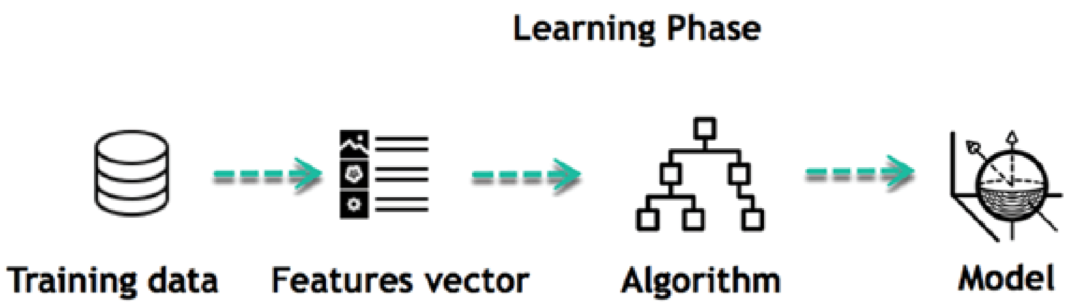
Machine Learning

How does Machine learning work?

Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if its feed a previously unseen example, the machine has difficulties to predict.

The core objective of machine learning is the **learning** and **inference**. First of all, the machine learns through the discovery of patterns. This discovery is made thanks to the **data**. One crucial part of the data scientist is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a **feature vector.** You can think of a feature vector as a subset of data that is used to tackle a problem.

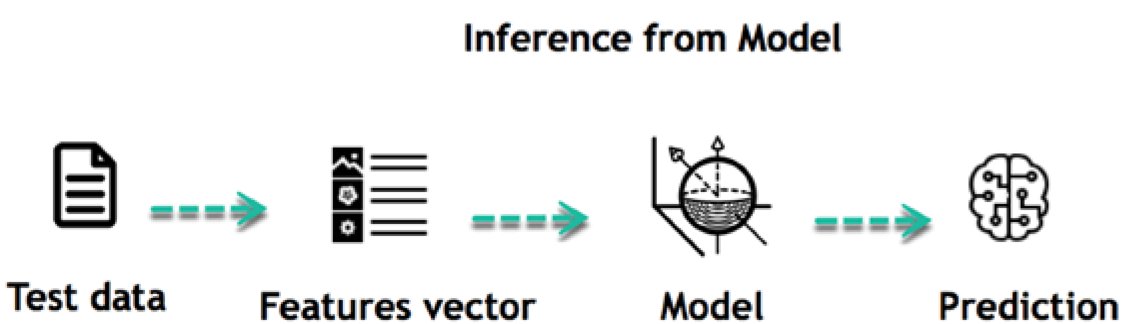
The machine uses some fancy algorithms to simplify the reality and transform this discovery into a **model**. Therefore, the learning stage is used to describe the data and summarize it into a model.



For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood to go to a fancy restaurant. It turns out the machine finds a positive relationship between wage and going to a high-end restaurant: This is the model

Inferring

When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need to update the rules or train again the model. You can use the model previously trained to make inference on new data.

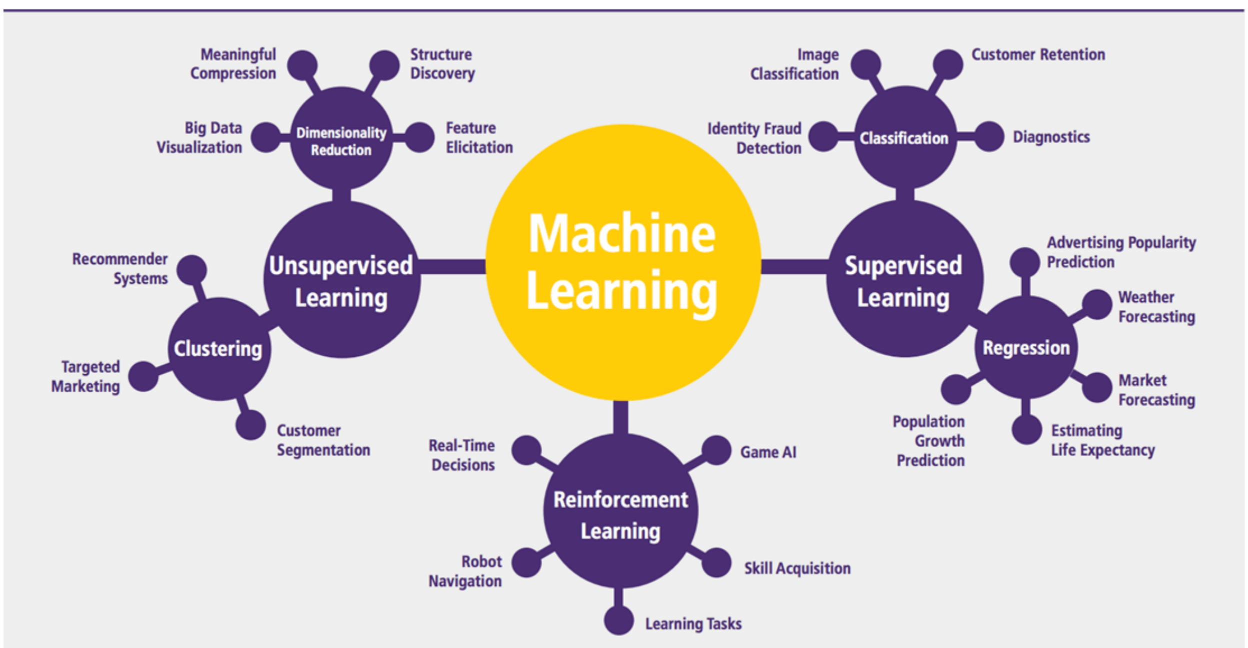


The life of Machine Learning programs is straightforward and can be summarized in the following points:

- Define a question
- Collect data
- Visualize data
- Train algorithm
- Test the Algorithm
- Collect feedback
- Refine the algorithm
- Loop 4-7 until the results are satisfying
- Use the model to make a prediction

Once the algorithm gets good at drawing the right conclusions, it applies that knowledge to new sets of data.

Machine learning Algorithms and where they are used?



Machine learning can be grouped into two broad learning tasks: Supervised and Unsupervised. There are many other algorithms

Supervised learning

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans.

You can use supervised learning when the output data is known. The algorithm will predict new data.

There are two categories of supervised learning:

Algorithm Name	**Description**	**Type**
Linear regression	Finds a way to correlate each feature to the output to help predict future values.	Regression
Logistic regression	Extension of linear regression that's used for classification tasks. The output variable 3is binary (e.g., only black or white) rather than continuous (e.g., an infinite list of potential colors)	Classification
Decision tree	Highly interpretable classification or regression model that splits data-feature values into branches at decision nodes (e.g., if a feature is a color, each possible color becomes a new branch) until a final decision output is made	Regression Classification
Naive Bayes	The Bayesian method is a classification method that makes use of the Bayesian theorem. The theorem updates the prior knowledge of an event	Regression Classification

	with the independent probability of each feature that can affect the event.	
Support vector machine	Support Vector Machine, or SVM, is typically used for the classification task. SVM algorithm finds a hyperplane that optimally divided the classes. It is best used with a non-linear solver.	Regression (not very common) Classification
Random forest	The algorithm is built upon a decision tree to improve the accuracy drastically. Random forest generates many times simple decision trees and uses the 'majority vote' method to decide on which label to return. For the classification task, the final prediction will be the one with the most vote; while for the regression task, the average prediction of all the trees is the final prediction.	Regression Classification
AdaBoost	Classification or regression technique that uses a multitude of models to come up with a decision but weighs them based on their accuracy in predicting the outcome	Regression Classification
Gradient-boosting trees	Gradient-boosting trees is a state-of-the-art classification/regression technique. It is focusing on the error committed by the previous trees and tries to correct it.	Regression Classification

- Classification task
- Regression task

Classification

Imagine you want to predict the gender of a customer for a commercial. You will start gathering data on the height, weight, job, salary, purchasing basket, etc. from your customer database. You know the gender of each of your customer, it can only be male or female. The objective of the classifier will be to assign a probability of being a male or a female (i.e., the label) based on the information (i.e., features you have collected). When the model learned how to recognize male or female, you can use new data to make a prediction. For instance, you just got new information from an unknown customer, and you want to know if it is a male or female. If the classifier predicts male = 70%, it means the algorithm is sure at 70% that this customer is a male, and 30% it is a female.

The label can be of two or more classes. The above example has only two classes, but if a classifier needs to predict object, it has dozens of classes (e.g., glass, table, shoes, etc. each object represents a class)

Regression

When the output is a continuous value, the task is a regression. For instance, a financial analyst may need to forecast the value of a stock based on a range of feature like equity, previous stock performances, macroeconomics index. The system will be trained to estimate the price of the stocks with the lowest possible error.

Unsupervised learning

In unsupervised learning, an algorithm explores input data without being given an explicit output variable (e.g., explores customer demographic data to identify patterns)

You can use it when you do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you

Algorithm	Description	Type

K-means clustering	Puts data into some groups (k) that each contains data with similar characteristics (as determined by the model, not in advance by humans)	Clustering
Gaussian mixture model	A generalization of k-means clustering that provides more flexibility in the size and shape of groups (clusters	Clustering
Hierarchical clustering	Splits clusters along a hierarchical tree to form a classification system. Can be used for Cluster loyalty-card customer	Clustering
Recommender system	Help to define the relevant data for making a recommendation.	Clustering
PCA/T-SNE	Mostly used to decrease the dimensionality of the data. The algorithms reduce the number of features to 3 or 4 vectors with the highest variances.	Dimension Reduction

Application of Machine learning

Augmentation:

Machine learning, which assists humans with their day-to-day tasks, personally or commercially without having complete control of the output. Such machine learning is used in different ways such as Virtual Assistant, Data analysis, software solutions. The primary user is to reduce errors due to human bias.

Automation:

Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots performing the essential process steps in manufacturing plants.

Finance Industry

Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns inside the data but also to prevent fraud.

Government organization

The government makes use of ML to manage public safety and utilities. Take the example of China with the massive face recognition. The government uses Artificial intelligence to prevent jaywalker.

Healthcare industry

Healthcare was one of the first industry to use machine learning with image detection.

Marketing

Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers develop advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, marketing department relies on AI to optimize the customer relationship and marketing campaign.

Example of application of Machine Learning in Supply Chain

Machine learning gives terrific results for visual pattern recognition, opening up many potential applications in physical inspection and maintenance across the entire supply chain network.

Unsupervised learning can quickly search for comparable patterns in the diverse dataset. In turn, the machine can perform quality inspection throughout the logistics hub, shipment with damage and wear.

For instance, IBM's Watson platform can determine shipping container damage. Watson combines visual and systems-based data to track, report and make recommendations in real-time.

In past year stock manager relies extensively on the primary method to evaluate and forecast the inventory. When combining big data and machine learning, better forecasting techniques have been implemented (an improvement of 20 to 30 % over traditional forecasting tools). In term of sales, it means an increase of 2 to 3 % due to the potential reduction in inventory costs.

Example of Machine Learning Google Car

For example, everybody knows the Google car. The car is full of lasers on the roof which are telling it where it is regarding the surrounding area. It has radar in the front, which is informing the car of the speed and motion of all the cars around it. It uses all of that data to figure out not only how to drive the car but also to figure out and predict what potential drivers around the car are going to do. What's impressive is that the car is processing almost a gigabyte a second of data.

Deep Learning

Deep learning is a computer software that mimics the network of neurons in a brain. It is a subset of machine learning and is called deep learning because it makes use of deep neural networks. The machine uses different layers to learn from the data. The depth of the model is represented by the number of layers in the model. Deep learning is the new state of the art in term of AI. In deep learning, the learning phase is done through a neural network.

Reinforcement Learning

Reinforcement learning is a subfield of machine learning in which systems are trained by receiving virtual "rewards" or "punishments," essentially learning by trial and error. Google's DeepMind has used reinforcement learning to beat a human champion in the Go games. Reinforcement learning is also used in video games to improve the gaming experience by providing smarter bot.

One of the most famous algorithms are:

- Q-learning
- Deep Q network
- State-Action-Reward-State-Action (SARSA)
- Deep Deterministic Policy Gradient (DDPG)

Applications/ Examples of deep learning applications

AI in Finance: The financial technology sector has already started using AI to save time, reduce costs, and add value. Deep learning is changing the lending industry by using more robust credit scoring. Credit decision-makers can use AI for robust credit lending applications to achieve faster, more accurate risk assessment, using machine intelligence to factor in the character and capacity of applicants.

Underwrite is a Fintech company providing an AI solution for credit makers company. underwrite.ai uses AI to detect which applicant is more likely to pay back a loan. Their approach radically outperforms traditional methods.

AI in HR: Under Armour, a sportswear company revolutionizes hiring and modernizes the candidate experience with the help of AI. In fact, Under Armour Reduces hiring time for its retail stores by 35%. Under Armour faced a growing popularity interest back in 2012. They had, on average, 30000 resumes a month. Reading all of those applications and begin to start the screening and interview process was taking too long. The lengthy process to get people hired and on-boarded impacted Under Armour's ability to have their retail stores fully staffed, ramped and ready to operate.

At that time, Under Armour had all of the 'must have' HR technology in place such as transactional solutions for sourcing, applying, tracking and onboarding but those tools weren't useful enough. Under armour choose **HireVue**, an AI provider for HR solution, for both on-demand and live interviews. The results were bluffing; they managed to decrease by 35% the time to fill. In return, the hired higher quality staffs.

AI in Marketing: AI is a valuable tool for customer service management and personalization challenges. Improved speech recognition in call-center management and call routing as a result of the application of AI techniques allows a more seamless experience for customers.

For example, deep-learning analysis of audio allows systems to assess a customer's emotional tone. If the customer is responding poorly to the AI chatbot, the system can be rerouted the conversation to real, human operators that take over the issue.

Apart from the three examples above, AI is widely used in other sectors/industries.

Artificial Intelligence

Difference between Machine Learning and Deep Learning

	Machine Learning	**Deep Learning**
Data Dependencies	Excellent performances on a small/medium dataset	Excellent performance on a big dataset
Hardware dependencies	Work on a low-end machine.	Requires powerful machine, preferably with GPU: DL

		performs a significant amount of matrix multiplication
Feature engineering	Need to understand the features that represent the data	No need to understand the best feature that represents the data
Execution time	From few minutes to hours	Up to weeks. Neural Network needs to compute a significant number of weights
Interpretability	Some algorithms are easy to interpret (logistic, decision tree), some are almost impossible (SVM, XGBoost)	Difficult to impossible

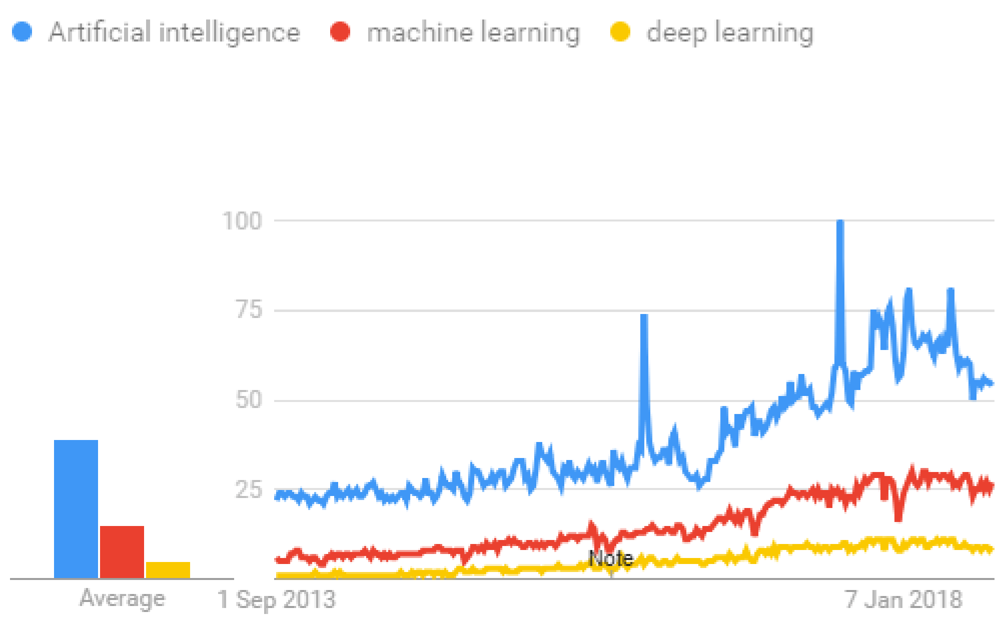
When to use ML or DL?

In the table below, we summarize the difference between machine learning and deep learning.

	Machine learning	**Deep learning**
Training dataset	Small	Large
Choose features	Yes	No
Number of algorithms	Many	Few

Training time	Short	Long

With machine learning, you need fewer data to train the algorithm than deep learning. Deep learning requires an extensive and diverse set of data to identify the underlying structure. Besides, machine learning provides a faster-trained model. Most advanced deep learning architecture can take days to a week to train. The advantage of deep learning over machine learning is it is highly accurate. You do not need to understand what features are the best representation of the data; the neural network learned how to select critical features. In machine learning, you need to choose for yourself what features to include in the model.



5.3 PROTOTYPING

a. Source code

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

```
import os

for dirname, _, filenames in os.walk('/kaggle/input'):

    for filename in filenames:

        print(os.path.join(dirname, filename))


!pip install plotly_express


import matplotlib.pyplot as plt

import seaborn as sns

import plotly_express as px

%matplotlib inline


import warnings

warnings.filterwarnings('ignore')


df=pd.read_csv('mismanaged_plasticwaste.csv')


df.head()


df.info()
```

* Feature Engineering

Renaming the columns

```
df=df.rename(columns={'Total_MismanagedPlasticWaste_2010 (millionT)': 'Total_mT_2010',
                'Total_MismanagedPlasticWaste_2019 (millionT)': 'Total_mT_2019',
                'Mismanaged_PlasticWaste_PerCapita_2010 (kg per year) ':'PerCapita_kg_2010',
                'Mismanaged_PlasticWaste_PerCapita_2019 (kg per year) ':'PerCapita_kg_2019'})
```

Getting country codes for mapping

```
!pip install pycountry
```

```
import pycountry
```

```
def findCountry (country_code):
    try:
        return pycountry.countries.get(name=country_code).alpha_3
    except:
        return 'NaN'
```

```
df['country_code'] = df.apply(lambda row: findCountry(row.Country) , axis = 1)


df_no=df[df['country_code']== 'NaN']

df_no['Country'].unique()


pd.set_option('display.max_rows', 200)


np.array(df[['Country','country_code']])


df[['Country','country_code']]=[['Albania', 'ALB'],
       ['Algeria', 'DZA'],
       ['Angola', 'AGO'],
       ['Anguilla', 'AIA'],
       ['Antigua and Barbuda', 'ATG'],
       ['Argentina', 'ARG'],
       ['Aruba', 'ABW'],
       ['Australia', 'AUS'],
       ['Bahamas', 'BHS'],
       ['Bahrain', 'BHR'],
       ['Bangladesh', 'BGD'],
       ['Barbados', 'BRB'],
```

['Belgium', 'BEL'],

['Belize', 'BLZ'],

['Benin', 'BEN'],

['Bermuda', 'BMU'],

['Bosnia and Herzegovina', 'BIH'],

['Brazil', 'BRA'],

['British Virgin Islands', 'VGB'],

['Brunei', 'BRN'],

['Bulgaria', 'BGR'],

['Burkina Faso', 'BFA'],

['Cambodia', 'KHM'],

['Cameroon', 'CMR'],

['Canada', 'CAN'],

['Cape Verde', 'CPV'],

['Cayman Islands', 'CYM'],

['Channel Islands', 'CHI'],

['Chile', 'CHL'],

['China', 'CHN'],

['Christmas Island', 'CXR'],

['Cocos Islands', 'CCK'],

['Colombia', 'COL'],

```
['Comoros', 'COM'],

['Congo', 'COG'],

['Cook Islands', 'COK'],

['Costa Rica', 'CRI'],

["Cote d'Ivoire", 'CIV'],

['Croatia', 'HRV'],

['Cuba', 'CUB'],

['Curacao', 'CUW'],

['Cyprus', 'CYP'],

['Democratic Republic of Congo', 'COD'],

['Denmark', 'DNK'],

['Djibouti', 'DJI'],

['Dominica', 'DMA'],

['Dominican Republic', 'DOM'],

['Ecuador', 'ECU'],

['Egypt', 'EGY'],

['El Salvador', 'SLV'],

['Equatorial Guinea', 'GNQ'],

['Eritrea', 'ERI'],

['Estonia', 'EST'],

['Faeroe Islands', 'FRO'],
```

['Falkland Islands', 'FLK'],

['Fiji', 'FJI'],

['Finland', 'FIN'],

['France', 'FRA'],

['French Guiana', 'GUF'],

['French Polynesia', 'PYF'],

['Gabon', 'GAB'],

['Gambia', 'GMB'],

['Georgia', 'GEO'],

['Germany', 'DEU'],

['Ghana', 'GHA'],

['Gibraltar', 'GIB'],

['Greece', 'GRC'],

['Greenland', 'GRL'],

['Grenada', 'GRD'],

['Guadeloupe', 'GLP'],

['Guam', 'GUM'],

['Guatemala', 'GTM'],

['Guernsey', 'GGY'],

['Guinea', 'GIN'],

['Guinea-Bissau', 'GNB'],

['Guyana', 'GUY'],

['Haiti', 'HTI'],

['Honduras', 'HND'],

['Hong Kong', 'HKG'],

['Iceland', 'ISL'],

['India', 'IND'],

['Indonesia', 'IDN'],

['Iran', 'IRN'],

['Iraq', 'IRQ'],

['Ireland', 'IRL'],

['Israel', 'ISR'],

['Italy', 'ITA'],

['Jamaica', 'JAM'],

['Japan', 'JPN'],

['Jordan', 'JOR'],

['Kazakhstan', 'KAZ'],

['Kenya', 'KEN'],

['Kiribati', 'KIR'],

['Kuwait', 'KWT'],

['Latvia', 'LVA'],

['Lebanon', 'LBN'],

```
['Lesotho', 'LSO'],
['Liberia', 'LBR'],
['Libya', 'LBY'],
['Lithuania', 'LTU'],
['Macau', 'MAC'],
['Madagascar', 'MDG'],
['Malaysia', 'MYS'],
['Maldives', 'MDV'],
['Malta', 'MLT'],
['Marshall Islands', 'MHL'],
['Martinique', 'MTQ'],
['Mauritania', 'MRT'],
['Mauritius', 'MUS'],
['Mexico', 'MEX'],
['Micronesia', 'FSM'],
['Monaco', 'MCO'],
['Montenegro', 'MNE'],
['Montserrat', 'MSR'],
['Morocco', 'MAR'],
['Mozambique', 'MOZ'],
['Myanmar', 'MMR'],
```

['Namibia', 'NAM'],

['Nauru', 'NRU'],

['Netherlands', 'NLD'],

['Netherlands Antilles', 'ANT'],

['New Caledonia', 'NCL'],

['New Zealand', 'NZL'],

['Nicaragua', 'NIC'],

['Nigeria', 'NGA'],

['Niue', 'NIU'],

['Norfolk Island', 'NFK'],

['North Korea', 'NaN'],

['Northern Mariana Islands', 'MNP'],

['Norway', 'NOR'],

['Oman', 'OMN'],

['Pakistan', 'PAK'],

['Palau', 'PLW'],

['Palestine', 'PSE'],

['Panama', 'PAN'],

['Papua New Guinea', 'PNG'],

['Peru', 'PER'],

['Philippines', 'PHL'],

```
['Poland', 'POL'],

['Portugal', 'PRT'],

['Puerto Rico', 'PRI'],

['Qatar', 'QAT'],

['Reunion', 'REU'],

['Romania', 'ROU'],

['Russia', 'RUS'],

['Saint Helena', 'SHN'],

['Saint Kitts and Nevis', 'KNA'],

['Saint Lucia', 'LCA'],

['Saint Martin', 'MAF'],

['Saint Pierre and Miquelon', 'SPM'],

['Saint Vincent and the Grenadines', 'VCT'],

['Samoa', 'WSM'],

['Sao Tome and Principe', 'STP'],

['Saudi Arabia', 'SAU'],

['Senegal', 'SEN'],

['Seychelles', 'SYC'],

['Sierra Leone', 'SLE'],

['Singapore', 'SGP'],

['Sint Maarten', 'SXM'],
```

```
['Slovakia', 'SVK'],

['Slovenia', 'SVN'],

['Solomon Islands', 'SLB'],

['Somalia', 'SOM'],

['South Africa', 'ZAF'],

['South Korea', 'KOR'],

['Spain', 'ESP'],

['Sri Lanka', 'LKA'],

['Sudan', 'SDN'],

['Suriname', 'SUR'],

['Sweden', 'SWE'],

['Syria', 'SYR'],

['Taiwan', 'TWN'],

['Tanzania', 'TZA'],

['Thailand', 'THA'],

['Timor', 'TLS'],

['Togo', 'TGO'],

['Tokelau', 'TKL'],

['Tonga', 'TON'],

['Trinidad and Tobago', 'TTO'],

['Tunisia', 'TUN'],
```

```
    ['Turkey', 'TUR'],
    ['Turks and Caicos Islands', 'TCA'],
    ['Tuvalu', 'TUV'],
    ['Ukraine', 'UKR'],
    ['United Arab Emirates', 'ARE'],
    ['United Kingdom', 'GBR'],
    ['United States', 'USA'],
    ['Uruguay', 'URY'],
    ['Vanuatu', 'VUT'],
    ['Venezuela', 'VEN'],
    ['Vietnam', 'VNM'],
    ['Western Sahara', 'ESH'],
    ['Yemen', 'YEM'],
    ['Zimbabwe', 'ZWE']]

df.head()

df['var_Total_mt']=df['Total_mT_2019']-df['Total_mT_2010']
df['var_PerCapita']=df['PerCapita_kg_2019']-df['PerCapita_kg_2010']

df.describe()
```

* Visualization & Global Mapping

Comparing total mismanaged plastic waste in 2010 and in 2019

```
x1=df['Total_mT_2010'].sum()

x2=df['Total_mT_2019'].sum()

x=['2010','2019']

height = [x1,x2]

plt.bar(x,height)
```

Total amount of mismanaged plastic waste has doubled between 2010 and 2019 !

Global Map of mismanaged plastic waste in 2010

```
fig = px.choropleth(df,locations='country_code',color='Total_mT_2010',scope='world',color_continuous_scale=px.colors.sequential.GnBu,

                   range_color=(0,13000000),title='Total_mT_2010',height=800

   )

fig.show()
```

Global Map of mismanaged plastic waste in 2019

```
fig = px.choropleth(df,locations='country_code',color='Total_mT_2019',scope='world',color_continuous_scale=px.colors.sequential.GnBu,
                    range_color=(0,13000000),title='Total_mT_2010',height=800
    )
fig.show()
```

Top 10 countries which have mismanaged plastic waste in 2010

```
colors = ['blue', 'blue', 'blue', 'blue','blue','blue','red','blue','red','blue']
df.sort_values(by='Total_mT_2010', ascending=False).head(10).plot.bar(x='Country',y='Total_mT_2010',figsize=(8,5),color=colors)
```

8 countries are in Asia and 2 countries are in Africa out of top 10 countries in 2010.

Top 10 countries which have mismanaged plastic waste in 2019

```
colors = ['blue', 'blue', 'blue', 'yellow','red','red','blue','red','red','blue']
```

```
df.sort_values(by='Total_mT_2019',
ascending=False).head(10).plot.bar(x='Country',y='Total_mT_2019',figsize=(8,5),color=colors)
```

5 countries are in Asia, 4 countries are in Africa and 1 country is in South America out of top 10 countries in 2019.

Top 10 countries which have increased mismanaged plastic waste between 2010 and 2019

```
colors = ['blue', 'blue', 'yellow', 'blue','red','red','blue','red','blue','red']
df.sort_values(by='var_Total_mt',
ascending=False).head(10).plot.bar(x='Country',y='var_Total_mt',figsize=(8,5),color=colors)
```

India has increased mismanaged plastic waste overwhelmingly between 2010 and 2019.

Global Map of mismanaged plastic waste per capita in 2010

```
fig =
px.choropleth(df,locations='country_code',color='PerCapita_kg_2010',scope='world',color_conti
nuous_scale=px.colors.sequential.GnBu,
                range_color=(0,70),title='PerCapita_kg_2010',height=800
   )
fig.show()
```

Global Map of mismanaged plastic waste per capita in 2019

```
fig = px.choropleth(df,locations='country_code',color='PerCapita_kg_2019',scope='world',color_continuous_scale=px.colors.sequential.GnBu,
                    range_color=(0,70),title='PerCapita_kg_2019',height=800
    )
fig.show()
```

Top 10 countries which have mismanaged plastic waste per capita in 2010

```
df.sort_values(by='PerCapita_kg_2010', ascending=False).head(10).plot.bar(x='Country',y='PerCapita_kg_2010',figsize=(8,5))
```

Top 10 countries which have mismanaged plastic waste per capita in 2019

```
df.sort_values(by='PerCapita_kg_2019', ascending=False).head(10).plot.bar(x='Country',y='PerCapita_kg_2019',figsize=(8,5))
```

Top 10 countries which have most mismanaged plastic waste per capita increase between 2010 and 2019

```
df.sort_values(by='var_PerCapita',
ascending=False).head(10).plot.bar(x='Country',y='var_PerCapita',figsize=(8,5))
```

The Comoros is a volcanic archipelago off Africa’s east coast, in the warm Indian Ocean waters of the Mozambique Channel.

https://www.google.co.jp/maps/place/Comoros/@-26.5185058,-8.1862356,3z/data=!4m5!3m4!1s0x1898e3036408a48d:0xf70a7fbee4dfd4db!8m2!3d-11.6455!4d43.3333

Trinidad and Tobago is a dual-island Caribbean nation near Venezuela, with distinctive Creole traditions and cuisines.

https://www.google.co.jp/maps/place/Trinidad+and+Tobago/@-0.77942,-113.2231317,3.4z/data=!4m5!3m4!1s0x8c3607976350b6c5:0xff082855c639f127!8m2!3d10.691803!4d-61.222503

* EDA

I think that we have to consider some factors which have influence to plastic waste. For example, 'Comoros' and 'Trinidad and Tobago' are No1 and No2 in Top 10 countries which have most mismanaged plastic waste per capita and their increase. It may come from Plastic currents. But on the other hand, we also consider where people make plastic waste which means 'Plastic input from municipal solid waste and wastewater' and 'Plastic input into oceans'

So I look into the articles which show 'Plastic currents','Plastic input from municipal solid waste and wastewater' and 'Plastic input into oceans'.

1.Plastic currents

Discarded plastic moving around the ocean – on the surface, in the water column and on the sea floor – sometimes comes to rest. The geographical distribution of marine plastic debris is strongly influenced by the entry points and the different transport pathways, which are in turn determined by the density of plastic debris coupled with prevailing currents, wind and waves (Rech et al., 2014).

`![32241425111_d37d8e16d1_b.jpg](attachment:6d57261a-7601-4690-a5f1-434fe09a1ab4.jpg)`

reference: https://www.grida.no/resources/6913

2.Plastic input from municipal solid waste and wastewater

Debris released by human activity on land can be washed by surface runoff or blown by wind into rivers and other watercourses and ultimately be transported into the ocean. Debris can also be directly dumped or discharged from boats or sewage plants into rivers (Rech et al., 2014). Plastic is very efficiently transported downstream due to its near-neutral buoyancy and may reach the ocean after only a few days (Kabat et al., 2012). Rivers transport plastic debris and, because the average journey is much shorter than the time needed for plastic to degrade, the majority ultimately reaches the ocean. Debris can also become stranded on riverbanks or entangled in vegetation; it may then be remobilized by wind or surface runoff to continue its journey downstream (Williams and Simmons, 1997). During high discharge events caused by heavy rainfall or human-controlled water releases, plastic and other debris can be exported far offshore from the river mouth. Dispersal of debris is also more efficient along coasts that experience high wave energy and/or large tides or other dynamic current regimes (Galgani et al., 2000; Carson et al., 2013; Lechner et al., 2014; Rech et al., 2014).

`![32322457236_6157bd4a1e_b.jpg](attachment:f4e3eeb0-3d88-4298-beba-97b21a114962.jpg)`

reference:https://www.grida.no/resources/6925

3.Plastic input into the oceans

Despite knowledge of the role played by rivers, there are no global estimates of the amount of man-made debris reaching the ocean at river mouths. Therefore, of the estimated 4.8 to 12.7 million tonnes of litter which enter the marine environment in 2010 from land-based sources within a 50 km-wide coastal zone (Jambeck et al., 2015), the proportion delivered by rivers is unknown. Debris originating more than 50 km inland from the coast would also need to be added to the figures above. The quantity and composition of anthropogenic debris delivered by a particular river also depends on the intensity and character of the socio-economic activities and population density in the river basin. The implementation of environmental protection and waste treatment measures may help to reduce the leakage of debris. The distribution and extent of impervious surfaces (built-up areas) in watersheds has been used as a proxy for the input of plastic debris through watercourses, as it is directly related to both urbanization and runoff volume (Lebreton et al., 2012).

`![31519203414_12f1147779_b.jpg](attachment:9dbf8c75-1896-4009-989c-8f9e68031898.jpg)` reference: https://www.grida.no/resources/6906

* Clustering

```
df1=df.drop(['Country','country_code'],axis=1)
df1.head()

from sklearn.cluster import KMeans

kmeans_model = KMeans(n_clusters=5, random_state=10).fit(df1)

labels = kmeans_model.labels_
labels

df1['cluster']=labels
df1['Country']=df['Country']

sns.pairplot(df1,hue='cluster')

df1.groupby('cluster').mean()
```

```
df1[df1['cluster']==1]

df1[df1['cluster']==2]

df1[df1['cluster']==3]

df1[df1['cluster']==4]

df1['country_code']=df['country_code']

fig = px.choropleth(df1,locations='country_code',color='cluster',scope='world',
                    range_color=(0,4),title='Map by Clusters',height=800
   )
fig.show()
```

By clustering, I concluded that India and China play the important role in Mismanaged Plastic Waste, and south east asian countries, some african countries and Brazil are also key to reduce Mismanaged Plastic Waste. The important thing is that Mismanaged Plastic Waste is coming from not only 'Plastic input from municipal solid waste and wastewater' and 'Plastic input into oceans' but also from 'Plastic currents'. It means that plastic waste which is made in a country move to other countries. So each countries have to cooperate to cope with this problem.

* The future of plastic waste

Just for your inforamtion is that I also find the article which tells us the future of plastic waste.

Plastic waste produced and mismanaged

The rapid rise in the use of oil and gas during the last half century has been accompanied by the development of a range of petroleum products, some of which, like petrochemicals, have other important applications beyond energy production. The global production of petroleum-derived plastic has also increased dramatically, from 1.5 million tonnes in 1950 to more than 300 million tonnes in 2014 (Plastics Europe, 2015; Velis, 2014). Some people have described this dramatic increase in the use of plastics as the “Age of Plastics” (Stevens, 2002) or “Our Plastic Age” (Thompson et al., 2009). If the current trend where production increases by approximately 5 per cent a year continues, another 33 billion tonnes of plastic will have accumulated around the planet by 2050 (Rochman et al., 2013).

CHAPTER- 6

TESTING

6.1 Testing

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software Testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks at implementation of the software. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs.

Software Testing can also be stated as the process of validating and verifying that a software program/application/product:

- Meets the business and technical requirements that guided its design and Development.
- Works as expected and can be implemented with the same characteristics.

TESTING METHODS

Functional Testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- Functions: Identified functions must be exercised.
- Output: Identified classes of software outputs must be exercised.
- Systems/Procedures: system should work properly

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

Test Case for Excel Sheet Verification:

Here in machine learning we are dealing with dataset which is in excel sheet format so if any test case we need means we need to check excel file. Later on classification will work on the respective columns of dataset .

Test Case 1 :

SL #	TEST CASE NAME	DESCRIPTION	STEP NO	ACTION TO BE TAKEN (DESIGN STEPS)	EXPECTED (DESIGN STEP)	Test Execution Result (PASS/FAIL)
1	Excel Sheet verification	**Objective:** There should be an excel sheet. Any number of rows can be added to the sheet.	Step 1	Excel sheet should be available	Excel sheet is available	Pass
			Step 2	Excel sheet is created based on the template	The excel sheet should always be based on the template	Pass
			Step 3	Changed the name of excel sheet	Should not make any modification on the name of excel sheet	Fail
			Step 4	Added 10000 or above records	Can add any number of records	Pass

RESULTS

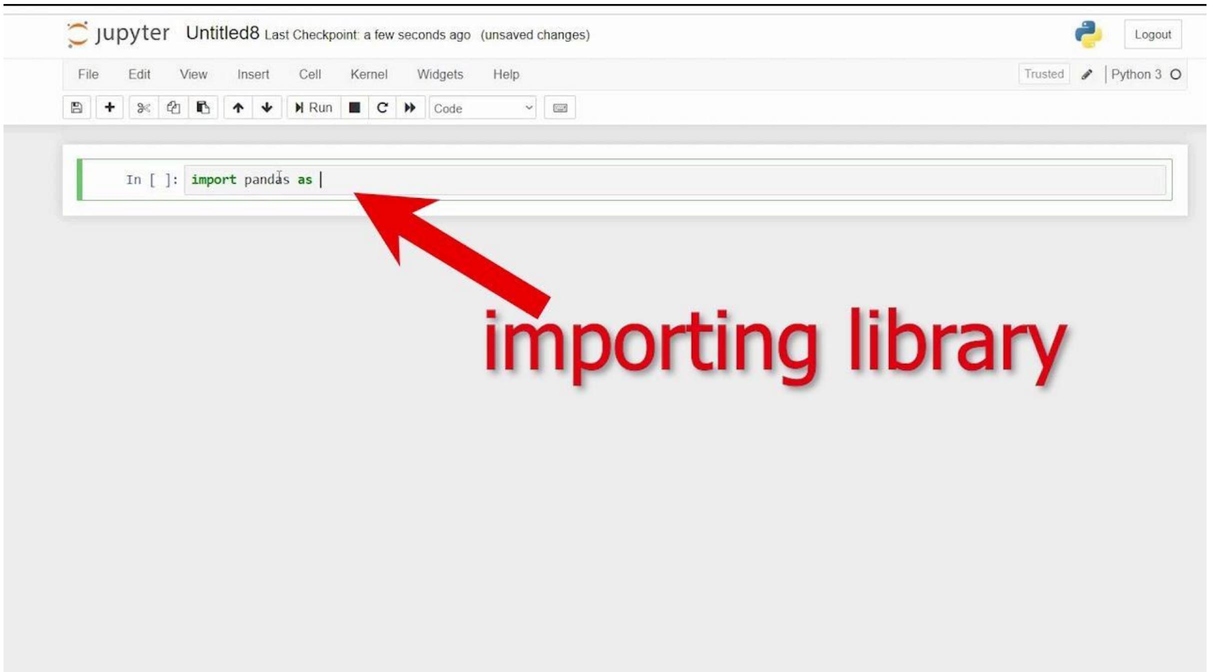
While reflecting on the outcomes of this project, the strengths were clearly obvious in that students were able to apply their coding skills to investigate and retrieve data of current issues the world is facing. Students also learnt to be more critical in coming up with targeted questions that were relevant for extraction from the data that is available. Though these were positive outcomes, it would be worth implementing some changes in the methodology for the next round to investigate and review how the results unfold. As a recommendation, instead of having every group exploring their own subject matter of interest, it could also be possible for the whole class to come up with one broad topic that everyone agrees on. Following that, students can be divided into groups to come up with their own question sets focusing on one sub category of the main topic. For the subject of waste management for example, possible aspects that student groups could focus on include – types of waste most produced, popular methods of waste disposal, cost involved in waste management, landfills around the world, ocean dumping etc. In this way the project could turn out to be more comprehensive in nature with all strands of the topic covered for further discussion in the class as a whole.

CHAPTER-7

RESULTS

7.1 Screenshots

i. Importing packages



ii. Data Collection

```
[2]: df = pd.read_csv('file.csv')

[3]: df
```

[3]:

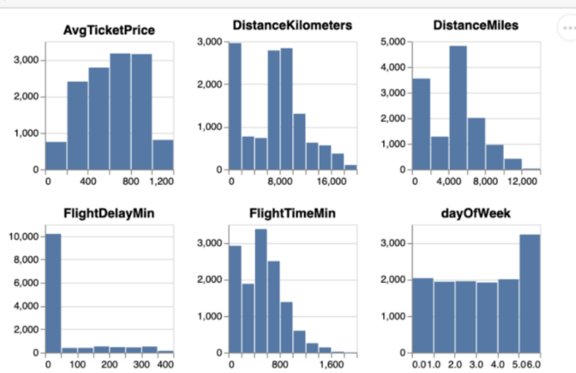
	Name	Score
0	a	90
1	b	80
2	c	95
3	d	20

iii. Data preprocessing

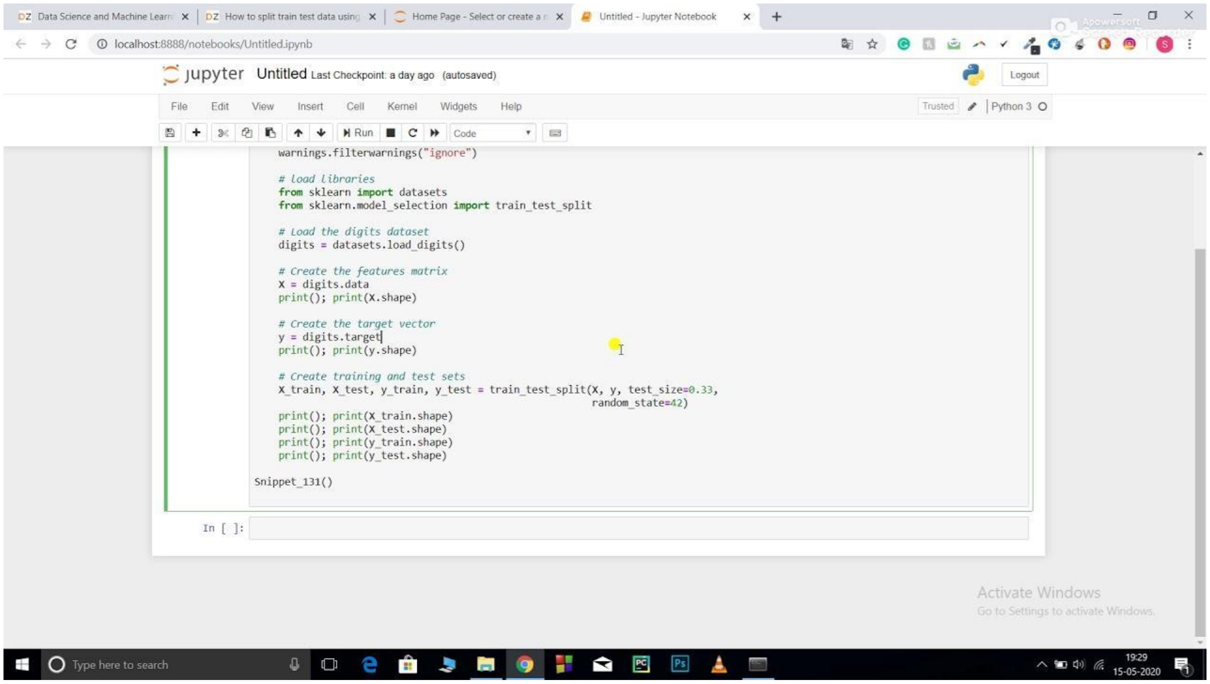
```
In [155]: #drop the records with age missing in inp0 and copy in inp1 dataframe.
          inp0['age'].dropna(inplace=True)

In [156]: inp0['age'].isnull().sum()

Out[156]: 0

In [157]: inp0.isnull().sum()

Out[157]: age          20
          salary        0
          balance       0
          marital       0
          targeted      0
          default       0
```

iv. Feature Extraction

```
In [9]: data = df_melt

        chart = alt.Chart(data).mark_bar().encode(
            alt.X('value:Q', bin=True, title=''),
            alt.Y('count()', title=''),
            tooltip=[
                alt.Tooltip('value:Q', bin=True, title='x'),
                alt.Tooltip('count()', title='y')
            ]
        ).properties(
            width=130,
            height=130
        )

        alt.ConcatChart(
            concat=[
              chart.transform_filter(alt.datum.attribute == value).properties(title=value)
              for value in sorted(data.attribute.unique())
            ],
            columns=3
        ).resolve_axis(
            x='independent',
            y='independent'
        ).resolve_scale(
            x='independent',
            y='independent'
        )

Out[9]:
```



v. Training and Testing



vi. Evaluation model

```
from sklearn.ensemble import RandomForestRegressor
ran_for = RandomForestRegressor()
ran_for.fit(X_train, Y_train)
Y_pred_ran_for = ran_for.predict(X_test)
r2=r2_score(Y_test, Y_pred_ran_for)
print("R2 score:", r2_score(Y_test, Y_pred_ran_for))
```

```
R2 score: 0.7226846570984489
```

CHAPTER-8

CONCLUSION

8.1 CONCLUSION

This research has elucidated how a project in a Computational Thinking course such as the ICT 2013 has the potential to create and stimulate a range of opportunities for

students to increase their awareness of current global issues. The possibility of taking this project to other areas of interest and to another level is unlimited and up to the creativity of the faculty and their students to explore. A project of this nature thus creates additional value for students by developing multiple skills such as coding, data analytics, collaborative learning and higher order thinking all of which are essential for them to be better prepared for the world of work.

CHAPTER-9

FUTURE ENHANCEMENTS

9.1 FUTURE ENHANCEMENTS

Projected increase in future plastic use will result in a concomitant increase in post-consumer plastic waste. For instance, by 2025 the global urban population is estimated to generate > 6 Mt of solid waste daily. Even using the present fraction of ~10% plastics in the solid waste stream, this amounts to over 200 Mt of waste plastics: this was the entire global plastic resin production in 2002 . The discouragingly slow growth in recycling rates and the likely increase in single-use products, both exacerbate this situation. Packaging products are almost always discarded with their functional characteristics virtually intact, permitting both facile re-use and recycling, however only about 9.4% of plastics is presently recycled in the US mainly due to collection costs, lack of requisite infrastructure and poor demand by processors for recycled plastic granulate.

CHAPTER-10

REFERENCES

10.1 REFERENCES

[1] T. C. Hsu, S. C. Chang, and Y. T. Hung, “How to learn and how to teach computational thinking: Suggestions based on a review of the literature,” Comput. Educ., 2018. Authorized licensed use limited to: University of New South Wales. Downloaded on July 26,2020 at 18:44:26 UTC from IEEE Xplore. Restrictions apply.

[2] J. M. Wing, “Computational Thinking: What and Why?,” thelink - Magaizne Varnegie Mellon Univ. Sch. Comput. Sci., 2010.

[3] J. M. Wing, “Computational thinking,” Communications of the ACM. 2006.

[4] M. Israel, J. N. Pearson, T. Tapia, Q. M. Wherfel, and G. Reese, “Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis,” Comput. Educ., 2015.

[5] J. Larmer and J. H. Mergendoller, “Seven essentials for project-based learning,” Educ. Leadersh., 2010.

[6] S. Ekberg, S. Danby, S. Houen, C. Davidson, and K. J. Thorpe, “Soliciting and pursuing suggestions: Practices for contemporaneously managing student- centred and curriculum-focused activities,” Linguist. Educ., 2017.

[7] J. Barnes, Applying Cross-curricular approaches creatively. 2018.

[8] D. Hoornweg and P. Bhada, “What a Waste. A Global Review of Solid Waste Management,” Urban Dev. Ser. Knowl. Pap., 2012.

[9] D. Asamoah, D. Doran, and S. Schiller, Teaching the Foundations of Data Science: An Interdisciplinary Approach. 2015.

[10] S. Gowrishankar, A. Veena, S. Gowrishankar, and A. Veena, “Introduction to Data Science,” in Introduction to Python Programming, 2019.

[11] R. Alber, "Deeper Learning: A Collaborative Classroom Is Key," Blog: eduTopia, 2012. [Online]. Available: https://www.edutopia.org/blog/deeper- learning-collaboration-key-rebecca-alber.

[12] S. Bell, "Project-Based Learning for the 21st Century: Skills for the Future," Clear. House A J. Educ. Strateg. Issues Ideas, 2010.

[13] M. Eodice, A. E. Geller, and N. Lerner, The meaningful writing project: Learning, teaching, and writing in higher education. 2017.

[14] G. Rodríguez, N. Pérez, G. Núñez, J. E. Baños, and M. Carrió, "Developing creative and research skills through an open and interprofessional inquiry-based learning course," BMC Med. Educ., 2019.

[15] K. D. Dahm and R. P. Nidhal Carla Bouaynaya; Ramachandran, "Board 34: Use of Big Data Analytics in a First-year Engineering Project," 2019.

[16] I. I. The World Bank, "CSV file-Country level dataset," Data Catalog (https://datacatalog.worldbank.org). [Online]. Available: https://datacatalog.worldbank.org/dataset/what- waste-global-database.

[17] R. Subaih, N. Maktoom, and B. Saeed, "Python Code for ICT 2013 Project," Microsoft Azure. Azure Notebook, Dubai, 2019.