

A REPORT OF TWO MONTHS INDUSTRIAL TRAINING
at
**NATIONAL INSTITUTE OF ELECTRONICS & INFORMATION
TECHNOLOGY (NIELIT)**

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE
AWARD OF THE DEGREE OF

BACHELOR OF TECHNOLOGY
(Electronics and Communication Engineering)



JULY-AUGUST 2024

SUBMITTED BY:
LAGANPREET SINGH
12202030

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

PUNJABI UNIVERSITY, PATIALA

INSTITUTE CERTIFICATE

Certificate No. : 101535

Roll No. : ST474/15



NATIONAL INSTITUTE OF ELECTRONICS & INFORMATION TECHNOLOGY, CHANDIGARH

An Autonomous Scientific Society of Ministry of Electronics & Information Technology, Govt of India

Ropar : Birla Farms, Bada Phull, Rupnagar

Chandigarh : Plot No. M925, IETE Building, Sector 30-B, Chandigarh

Certificate

This is to certify that Mr. Laganpreet Singh S/o Shri Ramandeep Singh

has successfully completed a course of Six Weeks duration from 03.06.2024 to 12.07.2024

on Artificial Intelligence and Machine Learning (Project: Milk Quality Prediction)

conducted at NIELIT Chandigarh

Course Curriculum

Python and Programming Constructs, Lists, Tuples, Dictionaries, Data Science, NumPy, Pandas, Matplotlib, Machine Learning, Predictive Analysis, Project Deployment using flask

Training Head

Date of Issue : 23.07.2024

Sanjeev

Executive Director

CANDIDATE'S DECLARATION

I “LAGANPREET SINGH” hereby declare that I have undertaken Summer-training at “NATIONAL INSTITUTE OF ELECTRONICS & INFORMATION TECHNOLOGY (NIELIT)” during a period from 03.06.2024 to 12.07.2024 in partial fulfillment of requirements for the award of degree of B.Tech (Department of Electronics and communication) at Punjabi University, Patiala. The work which is being presented in the training report submitted to Department of Electronics and Communication) at Punjabi University, Patiala is an authentic record of training work.

Signature of the Student

The summer training Viva-Voce Examination of _____

_____ has been

held on _____ and accepted.

Signature of the Examiner

ACKNOWLEDGEMENT

It is a pleasure to acknowledge many people who knowingly and unwittingly helped me, to complete my project. First of all, let us praise God for all the blessing, which carried us through all these years.

We extend our most gratitude to Ms. Anita Budhiraja and Dr. Sarwan Singh who has always stood by our side and guided, appreciated and encouraged us to get into more and more venture. Continuing the same, they enlightened us in the various stages during the development of this project and provided me with many insights and useful examples, which proved to be of immense help in the successful completion of the project.

We extend our sincere gratitude to our all teachers and guides who made unforgettable contribution. Due to their sincere efforts, we are able to work for such an organization and excel in the work entrusted upon us.

We value the friendly encouragement that we have got from all our classmates. There is still a debt we owe to our parents, friends, all the well- wishers who were directly and indirectly involved in making this project a reality.

INDEX

INSTITUTE CERTIFICATE	2
CANDIDATE'S DECLARATION	3
ACKNOWLEDGEMENT	4
INDEX	5
ABSTRACT	7
CHAPTER 1 – INTRODUCTION	10
1.1 Artificial Intelligence and Machine learning	10
1.1.1 Artificial Intelligence (AI):	10
1.1.2 Machine Learning (ML):	11
1.2 Applications of Machine Learning	13
• Image and Video Analysis:	13
• Natural Language Processing (NLP):	13
• Recommender Systems:	13
• Autonomous Vehicles:	13
• Healthcare:	14
• Financial Services:	14
• Predictive Maintenance:	14
1.3 Real World Examples of Machine Learning	14
• Speech & Image Recognition	15
• Traffic alerts using Google Map	15
• Chatbot (Online Customer Support)	15
• Google Translation	16
• Prediction	16
• Extraction	16
• g. Self-driving cars	16
• h. Ads Recommendation	16
1.4 About Python Programming Language	17
1.5 History of Python	17
1.6 Features in Python	17
• Free and Open Source	17
• Easy to code	18
• Easy to Read	18
• Object-Oriented Language	18
• GUI Programming Support	18
• High-Level Language	18
• Large Community Support	18
• Easy to Debug	18
• Python is a Portable language	19
• Python is an Integrated language	19
• Python is an Integrated language	19

• Large Standard Library	19
• Dynamically Typed Language	19
• Frontend and backend development	19
• Allocating Memory Dynamically	19
Python Latest Update	20
1.7 Role of Python in Machine Learning:	20
1.8 Jupyter Notebook	22
1.9 Google Colab:	23
1.9.1 Features of google colab	23
CHAPTER 2 – TRAINING WORK UNDERTAKEN	24
2.1 Libraries used	25
2.2 Dataset	25
ALL INFORMATION ABOUT MY DATASET	27
RESULTS	31
CONCLUSION AND FUTURE SCOPE	32
Conclusion	32
Future Scope	33
REFERENCES	34

ABSTRACT

Milk is the main dietary supply for every individual. High- quality milk shouldn't contain any adulterants. Dairy products are sold everywhere in society. Yet, the local milk vendors use a wide range of adulterants in their products, permanently altering the evaporated. Using milk that has gone bad can have serious health consequences. On October 18 of this year, the Food Safety and Standards Authority of India (FSSAI), the nation's top food safety authority, released the final result of the National Milk Safety and Quality Survey (NMSQS) and declared the milk readily available in India to be "mostly safe." According to an FSSAI survey, 68.4% of the milk in India is tainted. The quality of milk cannot be checked by any equipment or special system.

Milk that has not been pasteurized has not been treated to get rid of harmful bacteria. Infected raw milk may contain Salmonella, Campylobacter, Cryptosporidium, E. coli, Listeria, Brucella, and other dangerous pathogens. These microorganisms pose a major risk to your family's health. Manually analyzing the various milk constituents can be very challenging when determining the quality of the milk. Analyzing and discovering with the aid of machine learning can help with this endeavor. Here a machine learning-based milk quality prediction system is developed. The proposed technology has shown 99.99% classification accuracy.

ABOUT THE COMPANY

NIELIT, Chandigarh was established in 1978 to provide professional services in the areas of Information, Electronics & Communication Technology (IECT). It is a pioneer institute in this region.

The permanent campus of NIELIT Chandigarh is built in approximately 12 acres of land allotted by the Govt. of Punjab at Birla Farms, Bada Phull, Ropar (Rupnagar). It has been a constant endeavor of this Centre to develop excellence in various area of Information, Electronics & Communication Technology (IECT). It is an IT and Electronics Organization with clear cut strategies and its various operations are aimed at giving its customers a total package of IT Solutions and products. The Centre ensure cost effectiveness, time to market solutions through a highly skilled work force driven by strong design principles, highest levels of qualities and ethical business practices and constantly strive to delight our customer through excellence in Service Delivery. The Centre aims to disseminate knowledge on all aspects of Information, Electronics & Communication

Technology (IECT) to the utmost user satisfaction and has built up a formidable track record for the same. The Centre has a team of highly qualified, well trained, dedicated and experienced professionals. All of them undergo extensive training for developing and enhancing professional skill throughout their career. NIELIT, Chandigarh has proven its capability of handing large project of National important.



CHAPTER 1 – INTRODUCTION – AI & ML

Artificial Intelligence and Machine Learning are two rapidly evolving fields at the forefront of technological advancements. They both revolve around the idea of creating intelligent systems capable of performing tasks that traditionally require human intelligence. While AI is a broader concept, encompassing a wide range of techniques and applications, Machine Learning is a subset of AI that focuses on creating algorithms and models that enable machines to learn and improve from experience.

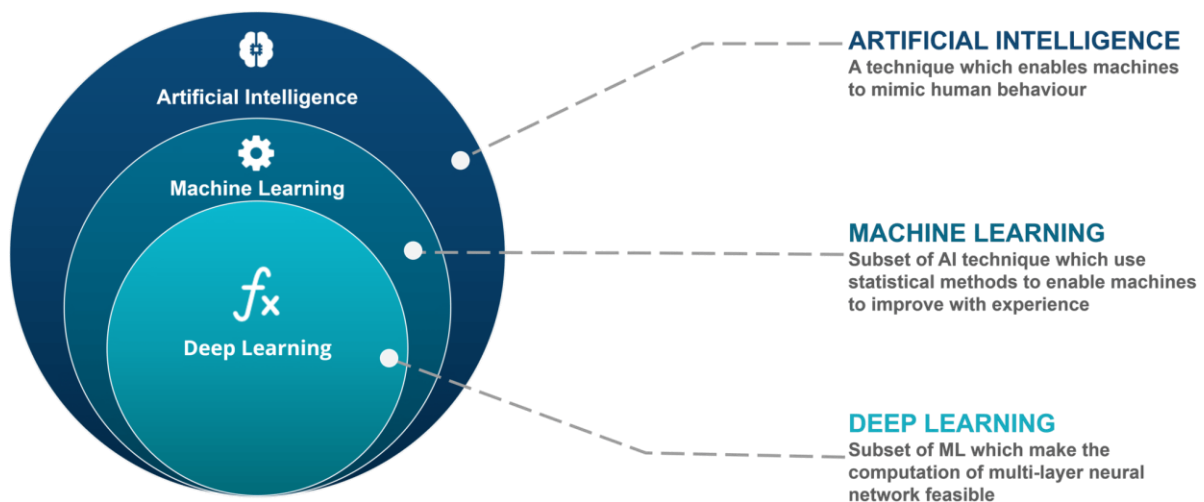


Figure 1.1 Artificial Intelligence

1.1 Artificial Intelligence (AI):

Artificial Intelligence is the theory and development of computer systems capable of performing tasks that typically require human intelligence. These tasks can include reasoning, problem solving, perception, learning, language understanding, and decision- making. The ultimate goal of AI is to create machines that can mimic, simulate, or even surpass human intelligence in various domains. AI systems can be broadly categorized into two types:

Narrow AI (Weak AI): This type of AI is designed to perform specific tasks or solve particular problems. Examples include virtual personal assistants like Siri, speech recognition systems, and recommendation algorithms used by online platforms.

General AI (Strong AI): General AI aims to possess human-like intelligence, enabling machines to understand, learn, and perform any intellectual task that a human being can do. Achieving this level of AI remains a significant challenge and is still largely theoretical.

Machine Learning (ML):

Machine Learning is a subset of AI that focuses on developing algorithms and statistical models to enable computers to learn and improve from experience automatically. Instead of being explicitly programmed, these algorithms learn patterns from data and make predictions or decisions based on that learning. There are three main types:

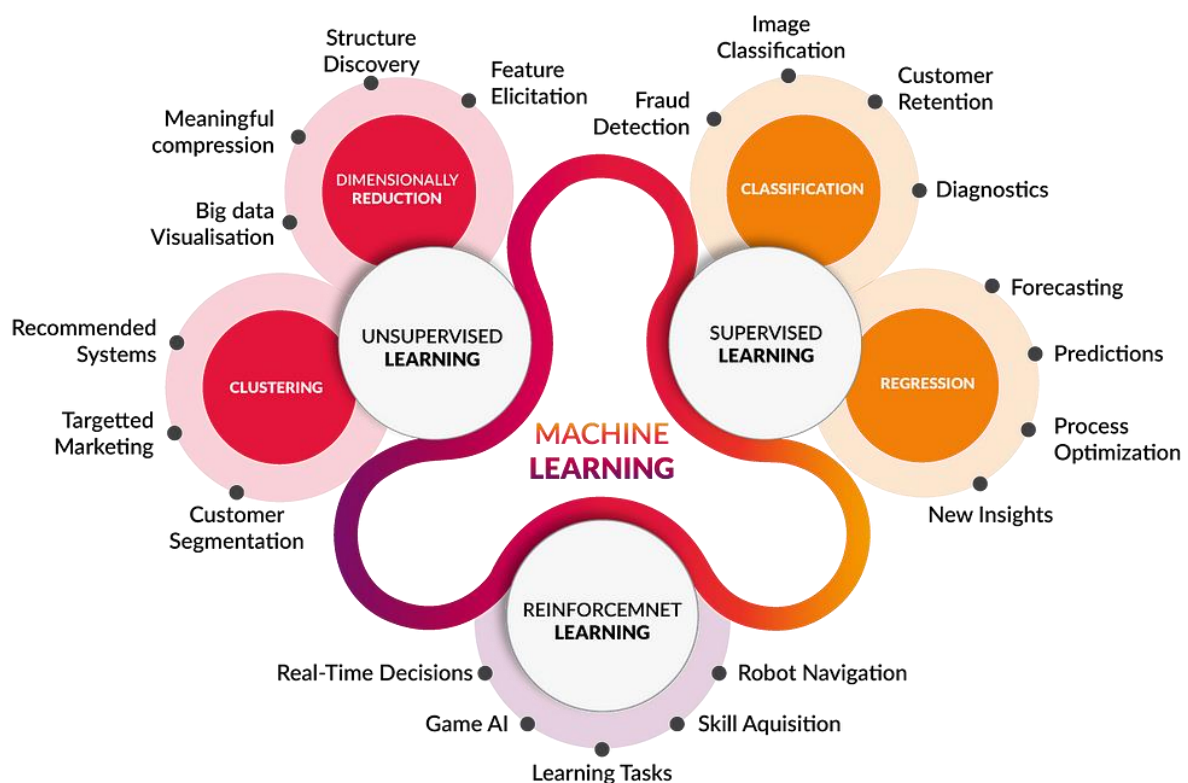


Figure 1.2 Machine Learning Types

Supervised Learning: In this approach, the algorithm is trained on labelled data, where the correct output is provided. The model learns to map inputs to outputs, making it capable of making predictions on new, unseen data.

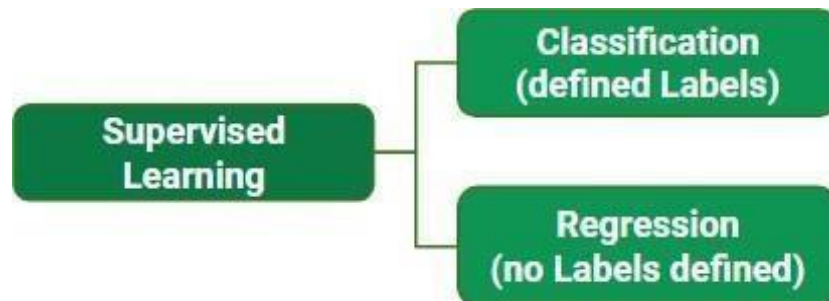


Figure 1.3 Supervised Learning

Unsupervised Learning: Unsupervised learning deals with unlabeled data, where the algorithm tries to find patterns and relationships within the data without specific guidance. Clustering and dimensionality reduction are common tasks in unsupervised learning.

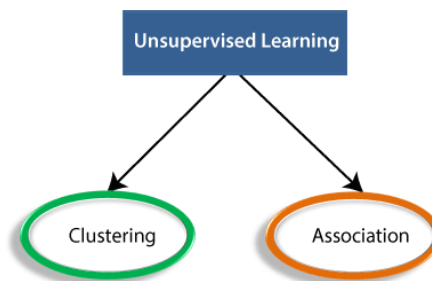


Figure 1.4 Unsupervised Learning

Reinforcement Learning: This learning paradigm involves an agent interacting with an environment and learning from feedback (rewards or penalties) received based on its actions. The agent aims to maximize its cumulative reward over time.

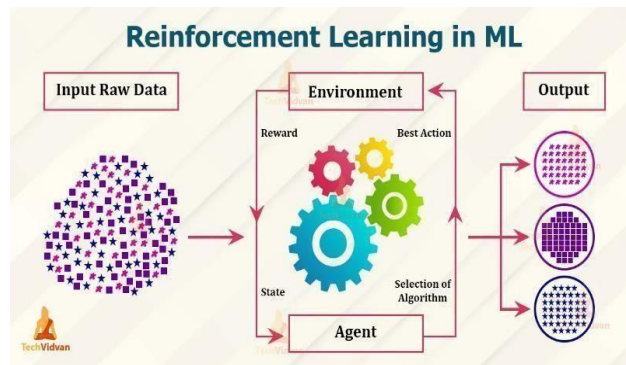


Figure 1.5 Reinforcement

Applications of Machine Learning

Machine Learning has a wide range of applications across various industries and domains. Its ability to analyse data, recognize patterns, and make predictions has led to numerous practical use cases. Here are some popular applications of Machine Learning:

Image and Video Analysis:

Machine Learning is used in image recognition and object detection tasks. Applications include facial recognition, automated tagging of images, detecting objects in photos, and medical image analysis for diagnosis.

Natural Language Processing (NLP):

NLP enables machines to understand, interpret, and generate human language. Applications include sentiment analysis, chatbots, language translation, speech recognition, and text summarization.

Recommender Systems:

Recommender systems use Machine Learning to suggest products, movies, music, or content based on a user's preferences and behavior. Examples include personalized recommendations on streaming platforms and e-commerce product recommendations.

Autonomous Vehicles:

Machine Learning plays a crucial role in developing self-driving cars. It helps in object detection, lane detection, real-time decision-making, and mapping for navigation.

Healthcare:

Machine Learning is used for medical image analysis, disease diagnosis, drug discovery, personalized treatment plans, and predicting patient outcomes.

Financial Services:

In finance, Machine Learning is applied for credit scoring, fraud detection, algorithmic trading, risk assessment, and customer service chatbots.

Predictive Maintenance:

In manufacturing and industrial sectors, Machine Learning is used for predictive maintenance, optimizing equipment maintenance schedules, and reducing downtime.

Real World Examples of Machine Learning

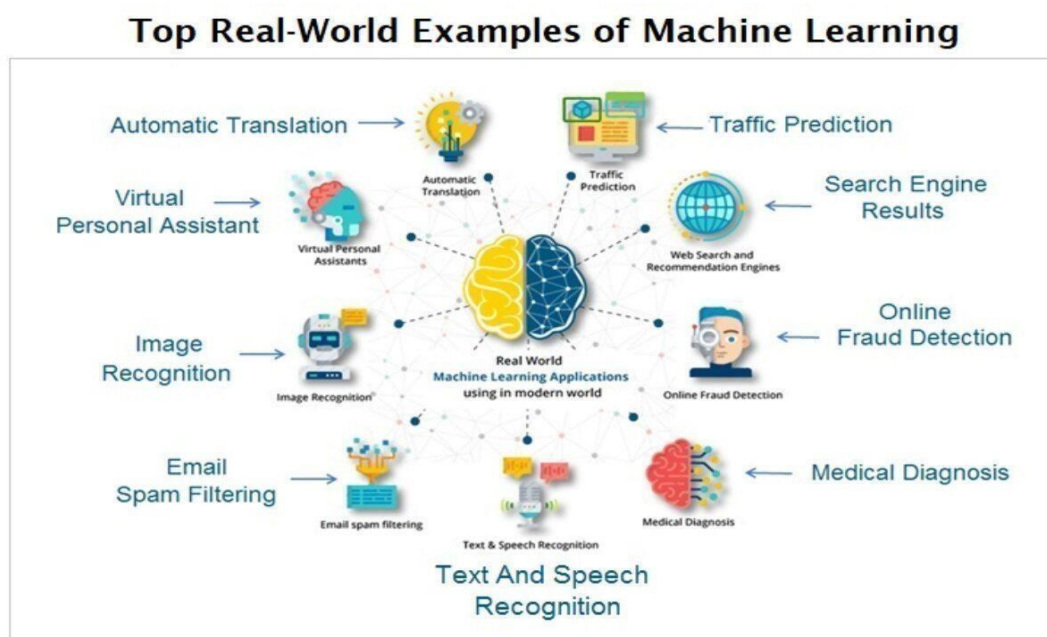


Figure 1.6 Real World Examples of Machine Learning

Speech & Image Recognition

Computer Speech Recognition or Automatic Speech Recognition helps to convert speech into text. Many applications convert the live speech into an audio file format and later convert it into a text file.

Similar to speech recognition, Image recognition is also the most widely used example of Machine Learning technology that helps identify any object in the form of a digital image.

Traffic alerts using Google Map

Google map uses different technologies, including machine learning which collects information from different users, analyse that information, update the information, and make predictions. With the help of predictions, it can also tell us the traffic before we start our journey. Machine Learning also helps identify the best and fastest route while we are in traffic using Google Maps

Chatbot (Online Customer Support)

A chatbot is the most widely used software in every industry like banking, Medical, education, health, etc. You can see chatbots in any banking application for quick online support to customers. These chatbots also work on the concepts of Machine Learning. The programmers feed some basic questions and answers based on the frequently asked queries. So, whenever a customer asks a query, the chatbot recognizes the question's keywords from a database and then provides appropriate resolution to the customer. This helps to make quick and fast customer service facilities to customers.

Google Translation

Suppose you work on an international banking project like French, German, etc., but you only know English. In that case, this will be a very panic moment for you because you can't proceed further without reviewing documents. Google Translator software helps to translate any language into the desired language. So, in this way, you can convert French, German, etc., into English, Hindi, or any other language. This makes the job of different sectors very easy as a user can work on any country's project hassle-free.

Prediction

Prediction system also uses Machine learning algorithms for making predictions. There are various sectors where predictions are used. For example, in bank loan systems, error probability can be determined using predictions with machine learning. For this, the available data are classified into different groups with the set of rules provided by analysts, and once the classification is done, the error probability is predicted.

Extraction

One of the best examples of machine learning is the extraction of information. In this process, structured data is extracted from unstructured data, and which is used in predictive analytics

tools. The data is usually found in a raw or unstructured form that is not useful, and to make it useful, the extraction process is used.

Self-driving cars

The future of the automobile industry is self-driving cars. These are driverless cars, which are based on concepts of deep learning and machine learning. Some commonly used machine learning algorithms in self-driving cars are Scale-invariant feature transform (SIFT), AdaBoost, Texton Boost, YOLO (You only look once).

Ads Recommendation

Nowadays, most people spend multiple hours on google or the internet surfing. And while working on any webpage or website, they get multiples ads on each page. But these ads are different for each user even when two users are using the same internet and on the same location. These ads recommendations are done with the help of machine learning algorithms. These ads recommendations are based on the search history of each user. For example, if one user searches for the Shirt on Amazon or any other e-commerce website, he will get start ads recommendation of shirts after some time.

About Python Programming Language

Python Language Introduction Python is a widely used general-purpose, high level programming language. It was initially designed by Guido Van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code. Python is a programming language that lets you work quickly and integrate systems more efficiently. 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.

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. Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Small Talk, and 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.



Figure 1.7 Guido Van Rossum

Features in Python

In this section we will see what are the features of Python programming language:

Free and Open Source

Python language is freely available at the official website and you can download it from the given download link below click on the Download Python keyword Download

Python. Since it is open-source, this means that source code is also available to the public. So you can download it, use it as well as share it.

Easy to code

Python is a high level programming language. Python is very easy to learn the language as compared to other languages like C, C#, Javascript, Java, etc. It is very easy to code in the Python language and anybody can learn Python basics in a few hours or days. It is also a developer-friendly language.

Easy to Read

As you will see, learning Python is quite simple. As was already established, Python's syntax is really straightforward. The code block is defined by the indentations rather than by semicolons or brackets.

Object-Oriented Language

One of the key features of Python is object oriented programming. Python supports object-oriented language and concepts of classes, object encapsulation, etc.

GUI Programming Support

Graphical User interfaces can be made using a module such as PyQt5, PyQt4, wxPython, or Tk in Python. PyQt5 is the most popular option for creating graphical apps with Python.

High-Level Language

Python is a high-level language. When we write programs in Python, we do not need to remember the system architecture, nor do we need to manage the memory.

Large Community Support

Python has gained popularity over the years. Our questions are constantly answered by the enormous Stack Overflow community. These websites have already provided answers to many questions about Python, so Python users can consult them as needed.

Easy to Debug

Excellent information for mistake tracing. You will be able to quickly identify and correct the majority of your program's issues once you understand how to interpret Python's error traces. Simply by glancing at the code, you can determine what it is designed to perform.

Python is a Portable language

Python language is also a portable language. For example, if we have Python code for Windows and if we want to run this code on other platforms such as Linux, Unix, and Mac then we do not need to change it, we can run this code on any platform.

Python is an Integrated language

Python is also an Integrated language because we can easily integrate Python with other languages like C, C++, etc.

Python is an Integrated language

Python is also an Integrated language because we can easily integrate Python with other languages like C, C++, etc.

Large Standard Library

Python has a large standard library that provides a rich set of modules and functions so you do not have to write your own code for every single thing. There are many libraries present in Python such as regular expressions, unit-testing, web browsers, etc.

Dynamically Typed Language

Python is a dynamically-typed language. That means the type (for example- int, double, long, etc.) for a variable is decided at run time not in advance because of this feature we don't need to specify the type of variable.

Frontend and backend development

With a new project py script, you can run and write Python codes in HTML with the help of some simple tags <py-script>, <py-env>, etc. This will help you do frontend development work in Python like javascript. Backend is the strong forte of Python it's extensively used for this work cause of its frameworks like Django and Flask.

Allocating Memory Dynamically

In Python, the variable data type does not need to be specified. The memory is automatically allocated to a variable at runtime when it is given a value. Developers do not need to write `int y = 18` if the integer value 15 is set to y. You may just type `y=18`.

Python Latest Update

The latest Python version is 3.12.6 and in this Python versions improve an efficiency, developer experience, and performance.

Improved Performance: Python 3.12 introduces several optimizations, such as support for the BOLT binary optimizer, which provides an overall 5% performance improvement.

Enhanced Error Messages: This version includes better error messages, with more detailed suggestions for fixing common typos and mistakes, making it more user-friendly, especially for beginners.

New Debugging Features: A new debugging and profiling API was introduced, which will be especially useful for developers when profiling their applications.

Role of Python in Machine Learning:

Python plays a significant role in the field of machine learning and is widely regarded as one of the most popular and versatile programming languages for this domain. Its role in machine learning can be attributed to several key factors:

Rich Ecosystem of Libraries: Python offers an extensive collection of libraries and frameworks specifically designed for machine learning and data science. Some of the most prominent ones include:

NumPy: A fundamental library for numerical computations, which forms the backbone of many other machine learning libraries.

Sandas: A powerful library for data manipulation and analysis, often used for data preprocessing.

Scikit-learn: A comprehensive library for various machine learning algorithms, including classification, regression, clustering, and more.

TensorFlow and PyTorch: Leading deep learning frameworks used for building and training neural networks.

Keras: An easy-to-use high-level neural networks API built on top of TensorFlow and Theano (now integrated with TensorFlow as tf.keras).

Ease of Use: Python's simple and clean syntax makes it easy to read and write, reducing the learning curve for newcomers to machine learning. Its readability also facilitates better collaboration among team members.

Community and Documentation: Python has a vast and active community of data scientists, machine learning researchers, and developers. This community actively contributes to the development of machine learning tools, libraries, and frameworks. Additionally, there is extensive documentation available for most popular Python libraries, making it easier to get started and find solutions to common problems.

Versatility and Integration: Python is a versatile language that can be used for various tasks beyond machine learning. Its ability to integrate well with other languages and technologies makes it a preferred choice for incorporating machine learning models into larger applications.

Data Processing and Visualization: Python's ecosystem includes libraries like pandas and matplotlib, which enable efficient data preprocessing, cleaning, and visualization. These steps

are essential for preparing the data before feeding it into machine learning models.

Support for Deep Learning: With the rise of deep learning, Python has become the go-to language for building, training, and deploying deep neural networks. TensorFlow and PyTorch, both primarily Python-based, are among the most popular frameworks for deep learning.

Open-Source and Free: Python is an open-source language, which means it is freely available for anyone to use and modify. This accessibility has contributed to its widespread adoption in academia and industry.

Python is Portable and Extensible: This is an important reason why Python is so popular in Machine Learning. A lot of cross-language operations can be performed easily on Python because of its portable and extensible nature. There are many data scientists who prefer using Graphics Processing Units (GPUs) for training their ML models on their own machines and the portable nature of Python is well suited for this. Also, many different platforms support Python such as Windows, Macintosh, Linux, Solaris, etc. In addition to this, Python can also be integrated with Java, .NET components or C/C++ libraries because of its extensible nature.

Jupyter Notebook

Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text tool among data and narrative text tool among data .It is a popular tists, researchers, and developers for popular scientists, researchers, and developers for computing and data analysis interactive computing and data analysis. interactive computing and data analysis.

Jupyter Notebook's key features:

Cell-based Structure: Notebooks are organized into cells, which can be of two types

Code Cells: Used for writing and executing code (Python, R, Julia, etc.).

Markdown Cells: Used for writing formatted text, including headings, lists, images, and more, using the Markdown syntax.

Code Execution: Code cells allow you to write and execute code in a step-by-step manner. To run a code cell, you can either click the "Run" button in the toolbar or use the keyboard shortcut "Shift + Enter." The output of the code execution is displayed directly below the cell.

Kernel: A notebook runs a kernel, which is essentially the computational engine that executes the code within the notebook. For example, if you write Python code, the notebook uses a Python kernel to execute it. You can switch kernels to work with different programming languages.

Saving and Exporting: Notebooks are automatically saved, but you can also manually save them by clicking the "Save" button or using the keyboard shortcut "Ctrl + S" (or "Cmd + S" on Mac). You can export notebooks to various formats, such as HTML, PDF, or plain Python scripts.

Rich Media Support: Notebooks allow you to embed images, videos, audio, and other media directly in the cells. This is particularly useful when creating interactive data visualizations.

Google Colab:

Google Colab (short for Collaboratory) is a cloud-based platform provided by Google that offers a Jupyter Notebooklike environment with access to free GPU and TPU (Tensor Processing Unit) resources. Colab is especially popular among data scientists, machine learning researchers, and students due to its ease of use, free access to computational resources, and the ability to collaborate and share notebooks easily.

Features of google colab

Jupyter Notebook Environment: Colab provides a Jupyter Notebook environment similar to what you would find in Jupyter Notebook. It supports both code cells (for writing and executing code) and markdown cells (for formatted text).

.Free GPU and TPU Support: One of the major advantages of Colab is the provision of free access to GPUs and TPUs. This is particularly useful for training deep learning models that require substantial computational power.

Cloud-based: Colab runs entirely in the cloud, which means you don't need to install anything on your local machine. You can access your notebooks from any device with an internet connection and a web browser.

Collaboration: Colab allows multiple users to collaborate on the same notebook

simultaneously. You can share the notebook with others and work on it together in realtime, making it a valuable tool for team projects and remote collaborations.

Pre-installed Libraries: Colab comes with many popular Python libraries pre-installed, such as TensorFlow, PyTorch, NumPy, pandas, matplotlib, and more. This saves time on setup and configuration, enabling you to focus on your analysis or machine learning tasks.

File Storage: Colab provides limited file storage in Google Drive. You can save and load files directly from your Google Drive, making it easy to keep your data and notebooks organized.

Cloud-based: Colab runs entirely in the cloud, which means you don't need to install anything on your local machine. You can access your notebooks from any device with an internet connection and a web browser.

Collaboration: Colab allows multiple users to collaborate on the same notebook simultaneously. You can share the notebook with others and work on it together in real-time, making it a valuable tool for team projects and remote collaborations.

Pre-installed Libraries: Colab comes with many popular Python libraries pre-installed, such as TensorFlow, PyTorch, NumPy, pandas, matplotlib, and more. This saves time on setup and configuration, enabling you to focus on your analysis or machine learning tasks.

File Storage: Colab provides limited file storage in Google Drive. You can save and load files directly from your Google Drive, making it easy to keep your data and notebooks organized.

Code Snippets and Examples: Colab provides a range of code snippets and examples for different use cases, which can help users get started quickly and learn new techniques.

CHAPTER 2 – TRAINING WORK UNDERTAKEN

Project introduction – **MILK QUALITY PREDICTION** (Milk production optimization)

Milk is considered as a most useful supplement in everyone's daily life. Hence better health it is advisable to consume good-quality milk. A perishable product is a milk. Tons of milk can deteriorate with just one gram of milk that is of low quality or structure, resulting in significant financial losses. In a very short period of time, millions of bacteria can grow in spoilt milk. In this way, circumstances that jeopardize human health may arise if people ingest milk or dairy products. In the USA each year around 48 million medical cases come due to food contamination. Because improper maintenance of dairy products can pose a transmission risk for a variety of pathogens and cause outbreaks of brucellosis, listeriosis, tuberculosis, etc., developing nations like India confront more simultaneous concerns. According to an FSSA Milk is considered as a most useful supplement in everyone's daily life. Hence better health it is advisable to consume good-quality milk. Developing nations like India confront more simultaneous concerns. According to an FSSA the dairy industry plays a crucial role in the global food supply chain, with milk being one of the most widely consumed agricultural products worldwide. Ensuring the quality and safety of milk is paramount to maintain consumer trust and satisfaction. Traditionally, milk quality assessment has relied on manual testing methods, which are labor-intensive, time-consuming, and prone to errors. Moreover, these methods often provide delayed results, hindering timely interventions to address potential quality issues. assessment by providing faster, more reliable, and cost-effective solutions. In this context, this paper proposes a novel approach for milk quality prediction using a machine learning algorithm.

Accurate milk quality prediction using machine learning brings several benefits:

Enhanced Product Quality: Consistently ensures high-quality milk and dairy products by accurately identifying variations and contaminants early, protecting brand reputation and consumer trust.

Cost Reduction: Reduces waste and resource usage by optimizing quality control processes, lowering the costs of manual inspection and reducing losses from spoiled or rejected products.

Improved Animal Health Monitoring: Detects anomalies related to animal health (such as

infections) through early signs in milk composition, enabling timely interventions and reducing veterinary expenses.

Operational Efficiency: Automates quality checks, allowing quicker decision-making, faster processing times, and efficient supply chain management from farm to consumer.

Regulatory Compliance: Helps maintain compliance with quality standards by providing reliable data for audit and reporting, thus reducing the risk of regulatory penalties.

Predictive Maintenance: Identifies issues in the milking process or equipment through abnormal milk data, preventing equipment failures and costly repairs. **Better Resource Management:** Enables more sustainable practices by monitoring and optimizing feed, veterinary care, and environmental conditions based on predicted quality outcomes.

With the rapid advancements in sensor technology and machine learning algorithms, there is a growing interest in developing automated systems for milk quality prediction. These systems leverage the power of hardware-enabled sensors to capture real-time data from milk samples and employ machine learning techniques to analyze this data and predict key quality parameters accurately.

Benefits

Improved Accuracy: Machine learning models have demonstrated improved accuracy in predicting milk quality compared to traditional methods, leveraging large datasets and sophisticated algorithms.

Real-Time Monitoring: These models enable real-time monitoring and prediction, allowing for timely interventions to ensure high milk quality

Cost-Effectiveness: Automation of milk quality prediction reduces the need for manual testing, leading to cost savings in the dairy industry.

Scalability: Machine learning systems can easily scale to handle large volumes of data from multiple sources, making them suitable for large dairy farms and cooperatives.

Data-Driven Decisions: The adoption of machine learning fosters a data-driven approach, helping stakeholders make informed decisions based on predictive analytics.

Enhanced Food Safety: By accurately predicting potential quality issues, machine learning contributes to enhanced food safety and consumer confidence in dairy product

Libraries used

Pandas - Pandas is a popular Python library for data manipulation and analysis.

It provides data structures for handling tabular and timeseries data, including the Data Frame and Series objects. Pandas offers a wide range of functions for data cleaning, filtering, grouping,

merging, and reshaping. It also supports data visualization and integration with other Python libraries, such as

NumPy and Matplotlib. Pandas' intuitive and powerful API has made it a go-to library for data analysts and scientists.



NumPy - NumPy is the fundamental package for scientific computing in Python which provides a multidimensional array object. Other mathematical operations can be performed using this but simply speaking we just need it

to convert our images into some form of an array so that we can store the model that has been trained.



Sklearn - scikit-learn (often abbreviated as sklearn) is a popular and widely-used machine learning library for Python. It provides a rich set of tools for various machine learning tasks, including classification, regression, clustering, dimensionality reduction, model selection, and preprocessing of data. scikit-learn is built on top of other foundational Python libraries like



NumPy and SciPy, making it a fundamental part of the Python data science ecosystem.

Dataset

We provided some general information about milk quality prediction datasets and it typically This dataset is manually collected from observations. It helps us to build machine learning models to predict the quality of milk.

This dataset consists of 7 independent variables ie. pH, Temperature, Taste, Odor, Fat, Turbidity, and Color.

Generally, the Grade or Quality of the milk depends on these parameters. These parameters play a vital role in the predictive analysis of the milk.

Link of dataset used – <https://www.kaggle.com/datasets/cpluzshrijayan/milkquality>

Usage: The target variable is nothing but the Grade of the milk. It can be

Target

Low (Bad)

Medium (Moderate)

High (Good)

If Taste, Odor, Fat, and are satisfied with optimal conditions then they will assign 1 otherwise 0.

Temperature and ph are given their actual values in the dataset.

We have to perform data preprocessing, and data augmentation techniques to build statistical.

ALL INFORMATION ABOUT DATASET

```
mq.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1059 entries, 0 to 1058
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype  
---  -
0   pH              1059 non-null   float64
1   Temperature     1059 non-null   int64  
2   Taste          1059 non-null   int64  
3   Odor            1059 non-null   int64  
4   Fat             1059 non-null   int64  
5   Turbidity       1059 non-null   int64  
6   Colour          1059 non-null   int64  
7   Grade           1059 non-null   object  
dtypes: float64(1), int64(6), object(1)
memory usage: 66.3+ KB
```

Training process

First step is to upload required libraries in jupyter notebook that will be used for milk quality prediction.

```
import pandas as pd
import numpy as np
```

Then we will describe our data and check the minimum and maximum values of different columns.

```
mq.describe()
```

	pH	Temperature	Taste	Odor	Fat	Turbidity	Colour
count	1059.000000	1059.000000	1059.000000	1059.000000	1059.000000	1059.000000	1059.000000
mean	6.630123	44.226629	0.546742	0.432483	0.671388	0.491029	251.840415
std	1.399679	10.098364	0.498046	0.495655	0.469930	0.500156	4.307424
min	3.000000	34.000000	0.000000	0.000000	0.000000	0.000000	240.000000
25%	6.500000	38.000000	0.000000	0.000000	0.000000	0.000000	250.000000
50%	6.700000	41.000000	1.000000	0.000000	1.000000	0.000000	255.000000
75%	6.800000	45.000000	1.000000	1.000000	1.000000	1.000000	255.000000
max	9.500000	90.000000	1.000000	1.000000	1.000000	1.000000	255.000000

Then we will check the datatypes of the columns in the dataset

```
mq.dtypes
```

pH	float64
Temprature	int64
Taste	int64
Odor	int64
Fat	int64
Turbidity	int64
Colour	int64
Grade	object
dtype:	object

Now we will check how many entries are there in Grade column

```
mq['Grade'].value_counts()
```

Grade	
low	429
medium	374
high	256
Name:	count, dtype: int64

Now we will replace the strings into integer so that our whole dataset will be converted into numeric form.

```
# change the grades into [0,1,2]
mq['Grade'] = mq['Grade'].replace(['low','medium','high'],[0,1,2])
```

Then we will divide all columns into input and output variables 'x' and 'y' respectively

```
x=mq.iloc[:, :-1]  
y=mq.iloc[:, -1]
```

```
x.head()
```

	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour
0	6.6	35	1	0	1	0	254
1	6.6	36	0	1	0	1	253
2	8.5	70	1	1	1	1	246
3	9.5	34	1	1	0	1	255
4	6.6	37	0	0	0	0	255

```
y.head()
```

```
0    2  
1    2  
2    0  
3    0  
4    1  
Name: Grade, dtype: int64
```

X is our input

Y is our output

Now we split our dataset for training and testing.

```
from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
```

Now we use the DecisionTreeClassifier algorithm

```
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()
dt.fit(x_train,y_train)
```

▼ DecisionTreeClassifier
DecisionTreeClassifier()

To check the accuracy, we will perform given function below

```
from sklearn.metrics import accuracy_score
y_pred = dt.predict(x_test)
accuracy_score(y_test,y_pred)
```

0.9905660377358491

Prediction

```
# predict the test
y_pred = dt.predict([[6.6,35,1,0,1,0,254]])
y_pred
```

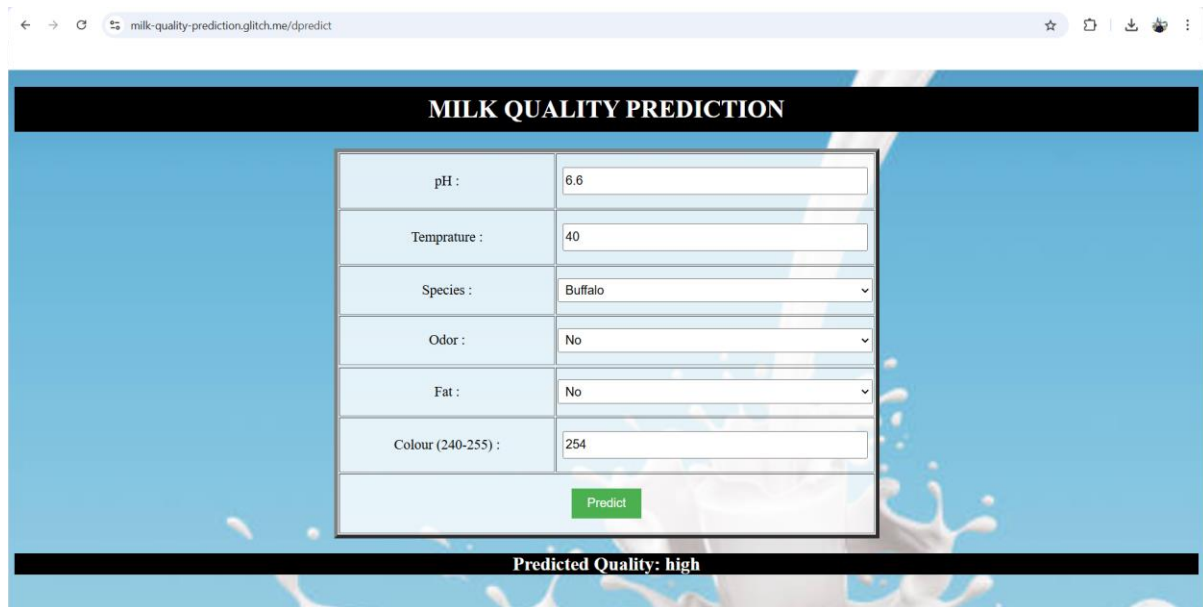
```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but DecisionTreeClassifier was fitted with feature names
  warnings.warn(
array([2])
```

NOW WE ARE DONE WITH BASIC ALGORITHM

RESULT

Here, shows real life model where we give some inputs and submit it and obtain result like 'Low', 'Medium', 'High'

Link for model – <https://milk-quality-prediction.glitch.me>



The screenshot shows a web browser window with the URL `milk-quality-prediction.glitch.me/dpredict`. The page has a blue background with a milk splash graphic. At the top, there is a black header with the text "MILK QUALITY PREDICTION" in white. Below this, there is a form with a light blue border. The form contains several input fields and a "Predict" button. The inputs are: pH (6.6), Temperature (40), Species (Buffalo), Odor (No), Fat (No), and Colour (240-255) (254). The "Predict" button is green. Below the form, there is a black footer with the text "Predicted Quality: high" in white.

Input	Value
pH :	6.6
Temperature :	40
Species :	Buffalo
Odor :	No
Fat :	No
Colour (240-255) :	254

Predicted Quality: high

Conclusion and Future Scope

Conclusion:

In conclusion, leveraging machine learning for milk quality prediction offers transformative potential across the dairy industry. By analyzing complex patterns within milk data, machine learning models can detect quality issues, such as contamination, adulteration, or bacterial presence, with a high degree of accuracy. These predictive models enable early intervention and reduce dependency on manual testing, which can be costly and time-consuming. As a result, dairy producers can enhance both the safety and consistency of their products, building greater consumer confidence and safeguarding public health.

One of the most significant advantages of using machine learning for milk quality prediction is its role in operational efficiency. Machine learning models automate and streamline quality assessment processes, reducing the need for labor-intensive testing while accelerating decision-making. This efficiency allows for a quicker response to potential quality issues, minimizing the chance of contaminated milk reaching consumers. Moreover, by identifying patterns and trends in real-time, producers can adjust operations dynamically, optimizing factors like storage conditions and processing methods to maintain peak milk quality.

Machine learning also plays a critical role in animal health monitoring, a crucial aspect of sustainable dairy production. By analyzing deviations in milk composition, machine learning algorithms can detect early signs of health issues in livestock, such as infections or nutritional deficiencies. This capability enables dairy farmers to address health problems promptly, ensuring the well-being of their herds and reducing veterinary costs. Improved animal welfare directly correlates with better milk quality, as healthier animals produce higher-quality milk, contributing to a more productive and ethical industry.

Additionally, integrating machine learning in milk quality prediction contributes to regulatory compliance. Dairy producers must adhere to stringent quality and safety standards, and machine learning provides reliable data to demonstrate compliance. These models can offer consistent documentation for audits, reducing the risk of penalties and bolstering the industry's overall reputation. As machine learning continues to advance, it has the potential to reshape regulatory frameworks, encouraging data-driven practices that enhance transparency and accountability within the dairy sector.

In summary, machine learning for milk quality prediction is not only a tool for ensuring product quality but also a comprehensive solution for enhancing the efficiency, sustainability, and

reliability of dairy production. By harnessing the power of predictive analytics, dairy producers can protect consumer health, reduce operational costs, and promote more ethical farming practices. As the dairy industry increasingly embraces this technology, machine learning will likely become an integral component in setting higher standards for milk quality and safety worldwide.

Future Scope:

Real-Time Monitoring and IoT Integration: ML-powered sensors and IoT devices can continuously monitor milk quality in real-time, from collection points to processing plants. This would enable immediate corrective actions, reducing waste and improving supply chain efficiency.

Precision Dairy Farming: With predictive analytics, ML can optimize feed quality, animal health, and environmental conditions tailored to individual animals, enhancing milk quality and yield per animal.

Automated Quality Assessment and Sorting: Automated systems at dairy plants could instantly assess milk quality and sort batches accordingly, streamlining processing and ensuring consistent quality in consumer products.

Predictive Maintenance for Dairy Equipment: By analyzing machine data and milk quality output, ML can predict equipment failures and contamination risks, reducing downtime and improving operational reliability.

Customized Product Development: Based on quality predictions, dairy processors could develop custom milk products tailored to consumer demands (e.g., low-fat or protein-rich milk) by preselecting milk based on its composition.

Disease Outbreak Prediction and Control: By analyzing trends in milk composition data, ML could help detect early signs of animal diseases, facilitating preventive action and minimizing the spread of infections in herds.

Enhanced Regulatory Compliance: ML models could assist in automatically ensuring that milk meets safety and regulatory standards, making compliance reporting easier and more efficient.

for dairy farms and processing plants.

Sustainable Dairy Management: ML can optimize resource use (e.g., water, feed) to lower environmental impacts while maintaining or improving milk quality, aligning with goals for sustainable agriculture.

Blockchain Integration for Traceability: ML in conjunction with blockchain can enhance traceability, allowing consumers to verify the quality and origin of milk, thus building greater transparency and trust in dairy products.

Advanced Quality Prediction Models: As ML algorithms become more sophisticated, we can expect even more precise predictions of milk quality, factoring in complex variables like genetics, seasonal changes, and nutritional factors.

REFERENCES

- | | |
|---------------------------|---|
| 1. W3SCHOOLS LINK: | https://www.w3schools.com/ |
| 2. JAVAPPOINT LINK: | https://www.javatpoint.com/ |
| 3. GEEKS FOR GEEKS LINK : | https://www.geeksforgeeks.org/ |
| 4. KAGGLE LINK: | https://www.kaggle.com/ |