**Soil Fertility Prediction**

A Project Report in partial fulfillment of the degree

# BachelorofTechnology

in

# ComputerScience&Engineering

## By

**2103A51221 SHREEYA BOSE**

**UndertheGuidanceof**

# D.Ramesh

# Assistant Professor, Department of CSE.

**Submittedto**



# DEPARTMENTOFCOMPUTERSCIENCE&ENGINEERING

## SRUNIVERSITY,ANANTHASAGAR,WARANGAL





**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CERTIFICATE**

This is to certify that the Project Report entitled “Soil Fertility Prediction” is a record of Bonafide work carried out by Shreeya Bose bearing RollNo **2203A51221** during the academic year 2022-2023 in partial fulfillment of the award of the degree of ***Bachelorof Technology*** in **Computer Science Engineering**by the SR UNIVERSITY, WARANGAL.

# Supervisor HeadoftheDepartment

Mr. D. Ramesh Dr. M. Sheshikala

Asst. Professor, Assoc. Prof .& HOD (CSE)

SR University SR University

**ACKNOWLEDGEMENT**

We express our thanks to course coordinator Mr. D. Ramesh, Asst. prof. for guiding us from the beginning through the end of the course project. We express our gratitude to head of the department CS&AI, Dr. M. Sheshikala, Associate Professor for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, Dr C. V. Guru Rao, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thank to all teaching and non-teaching staff of the department for their suggestions and timely support.

# ABSTRACT

**A Soil fertility prediction is super important for farming today, because it helps farmers grow more crops and use their resources better. Were trying to make a system that can tell how fertile the soil is by using different kinds of computer programs, like K-Nearest Neighbors, Logistic Regression, Support Vector Machines, Decision Trees, Random Forest, and AdaBoost.**

**First, we fix any missing or weird data, and then we make everything the same size and shape for the model to work with. , so, the dataset is split into two parts: one for training the model and one for testing it. This is done using a method called cross-validation to check how well the model works.**

**The KNN model uses how similar soil samples are to each other to guess how fertile they are, and Logistic Regression gives a chance of being right based on the features of the soil. , SVM uses a line to sort soil samples into different groups based on their fertility, while Decision Trees split the space into sections to guess what the soil is like. , Random Forest mixes a bunch of decision trees to make predictions more accurate, and AdaBoost keeps getting better at its job by making the weights of the wrong samples bigger or smaller.**

**We are going to compare different machine learning models to find the best one for predicting soil fertility, taking into account how accurate they are, how fast they run, and how easy they are to understand.**

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Content** | **PageNo** |
| 1 | Introduction | 1 |
| 2 | Design | 3 |
| 3 | Methodology | 3 |
| 4 | DataPre-processing | 5 |
| 5 | Results | 7 |
| 6 | Conclusion | 12 |
| 7 | Future scope | 12 |
| 8 | References | 12 |

1. **INTRODUCTION:**

Soil fertility prediction stands as a crucial domain within agricultural sciences, offering invaluable insights into the soil's capacity to support plant growth and agricultural productivity. With the advent of Artificial Intelligence and Machine Learning (AIML), the agricultural landscape has witnessed a transformative shift, empowering farmers and agronomists with predictive analytics to optimize crop yield and resource management. This introduction delves into an AIML project focused on soil fertility prediction, employing a diverse range of machine learning models, including logistic regression, KNN (K-Nearest Neighbors), SVM (Support Vector Machine), and Decision Tree.

At the heart of this endeavor lies the imperative to harness the power of data-driven insights to revolutionize traditional farming practices. By amalgamating advanced computing techniques with domain expertise, this project endeavors to create a robust framework capable of accurately predicting soil fertility levels based on a multitude of factors. These factors may encompass soil composition, nutrient levels, moisture content, pH levels, and climatic conditions, among others.

One of the primary machine learning algorithms employed in this project is logistic regression. Leveraging the principles of classification, logistic regression facilitates the categorization of soil samples into distinct fertility classes, thereby enabling farmers to make informed decisions regarding crop selection and nutrient management strategies. Additionally, KNN algorithm offers a non-parametric approach to classification, wherein soil samples are assigned fertility labels based on the similarity of their features to neighboring data points.

Furthermore, SVM algorithm, renowned for its effectiveness in handling high-dimensional data, plays a pivotal role in delineating intricate patterns within the soil dataset, thereby enhancing the accuracy of fertility predictions. Meanwhile, Decision Tree algorithm offers interpretability and transparency, enabling stakeholders to gain insights into the hierarchical relationships between various soil attributes and fertility outcomes.

The significance of this AIML project transcends mere predictive modeling; it embodies a paradigm shift towards precision agriculture, wherein resources are allocated judiciously, and environmental sustainability is prioritized. By harnessing the collective prowess of machine learning models, farmers can optimize fertilizer usage, mitigate soil degradation, and foster long-term agricultural resilience.

The soil fertility prediction AIML project represents a pioneering endeavor at the intersection of agriculture and artificial intelligence. Through the integration of logistic regression, KNN, SVM, and Decision Tree algorithms, this project aims to empower stakeholders with actionable insights, thereby ushering in a new era of data-driven agriculture characterized by efficiency, sustainability, and productivity.

In various research papers, different Machine Learning Algorithms have been utilized to predict soil fertility levels. Artificial intelligence has provided a suitable framework for conducting predictions on datasets through feature extraction and data pre-processing. The machine learning algorithms commonly employed include Logistic Regression, Support Vector Machine, Linear Regression, Decision Tree, and K Nearest Neighbor.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.NO DATEOF AUTHORS NAME METHODOLOGY ACCURACY  PUBLICATION | | | | | |
| 1 | 2023 | [Aishwarya Patil1, Varshini Kulkarni, Sachin Desai](https://www.sciencedirect.com/science/article/pii/S0168169920302301" \l "b0210) | Soil nutrient analysis machine learning methods | Multiple linear regression,  neural networks | 90.0 |
| 2 | 2019 | Keerthan Kumar T G, Shubha C, Sushma S A | Using classification algorithms for crop recommendation | Random Forest | 72.74 |
| 3 | 2020 | Jagruti Raut, Dr. Sonu Mittal | Soil Fertility and Crop Recommendation using Machine Learning and Deep Learning  Techniques | *SVM, Random Forest, LSTM, K-NN, K-Means, ELM* | 93.5 |
| 4 | 2022 | N. & Choudhary | Crop recommendation using deep neural networks | Deep Neural Network, K-nearest neighbor  (KNN), Decision Tree classifier, Support Vector  Machine (SVM), Gaussian NB, Linear  discriminant analysis (LDA) | 76.57 |
| 5 | 2016 | Balakrishnan & Muthukumarasamy | A study on crop yield using classification techniques | Support Vector Machine (SVM), Naïve Bayes,  AdaSVM and AdaNaive | 67.76 |
| 6 | 2018 | Nagpal | Soil feritility using machine learning mdels | Decision Tree , Random Forest, Support Vector Machines and eXtreme Gradient Boosting | 78 |

# DESIGN:

**RequirementSpecifications**

## HardwareRequirements

## System

## RAM

## HardDisk

## Input

## Output

## SoftwareRequirements

* + - **OS**
    - **Platform**
    - **ProgramLanguage**

# 3. METHODOLOGY:

After Data pre-processing and data visualization the next step is to apply the models on the dataset. Our dataset comes under supervised learning as it contains the labeled data (target variables, feature variables). First the dataset is splitted into training set and testing set. Then the model is trained on training set and then tested on testing set.

**3.1logistic regression algorithm:**

Logistic regression is a machine learning algorithm which comes under supervised learning. It is a parametric method, where an equation is formed to solve. The equation returns continues values. These continues values should to converted to categorical values.so, we use a activation function called “sigmoid”.by using log error function we calculate the error.

* from sklearn.linear\_model import LogisticRegression
* lr=LogisticRegression()
* mm=lr.fit(x\_resem\_train,y\_resem\_train)

**3.2K-Nearest Neighbor algorithm:**

K-Nearest Neighbor algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression. This algorithm is non parametric. This is also called as lazy learning algorithm. This algorithm works by first selecting the k value which is an integer value and less than the number of rows. When a new data point is given, KNN finds the nearest neighbors to that data point based on the distance using various methods like Euclidean distance or Manhattan distance. And assigns the data point to that class.

* from sklearn.neighbors import KNeighborsClassifier
* classifier=KNeighborsClassifier(n\_neighbors=5,metric='minkowski',p=2)
* classifier.fit(x\_resem\_train,y\_resem\_train

# 3.3Desicion Tree algorithm:

# Decision tree algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression problems. This algorithm is also known as ID3 algorithm. This algorithm is non parametric method. It forms a tree from the given dataset. It has two nodes decision nodes and leaf nodes. Decision nodes are used for taking decisions and leaf nodes are the output of that decisions. The attribute selection happens by entropy and information Gini.

* from sklearn.tree import DecisionTreeClassifier
* classifier=DecisionTreeClassifier(criterion='entropy',random\_state=0)
* mm=classifier.fit(x\_resem\_train,y\_resem\_train)

# 3. 4Support vector machine algorithm:

# Support vector machine algorithm is a machine learning algorithm which comes under supervised learning. This is used for both classification and regression problems. SVM works by constructing a hyperplane or a line that separates the different classes of data points. SVM has support vectors. The distance between positive hyperplane and negative hyperplane is called margin.

* from sklearn.svm import SVC
* svm\_model=SVC(kernel='linear')
* svm\_model.fit(x\_resem\_train,y\_resem\_train)

# 4.DATASETPREPROCESSING:

# DATASET DESCRIPTION

# Attributes:

# N

# P

# K

# pH

# EC

# OC

# S

# Zn

# Fe

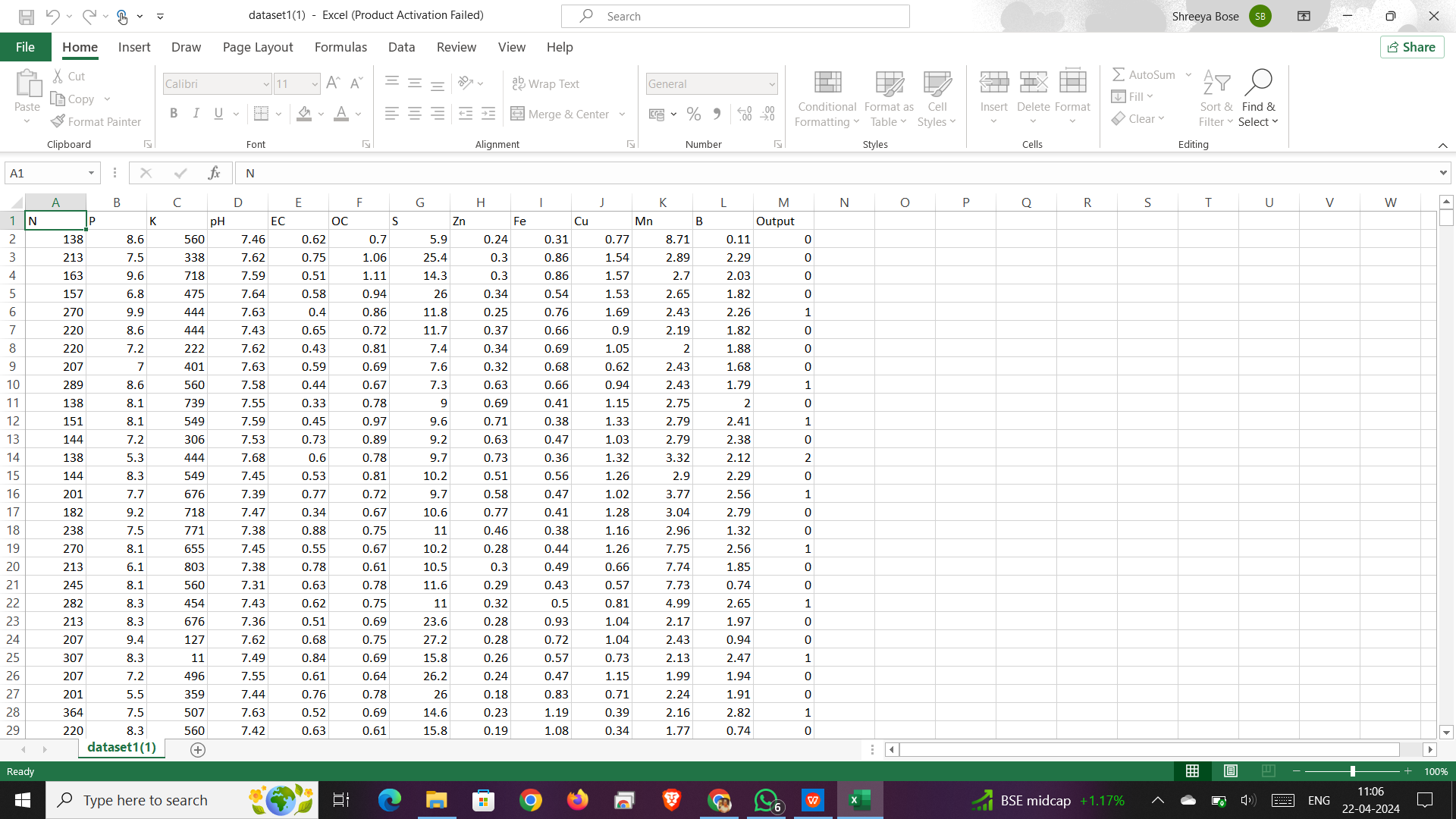
# Cu

# Mn

# B

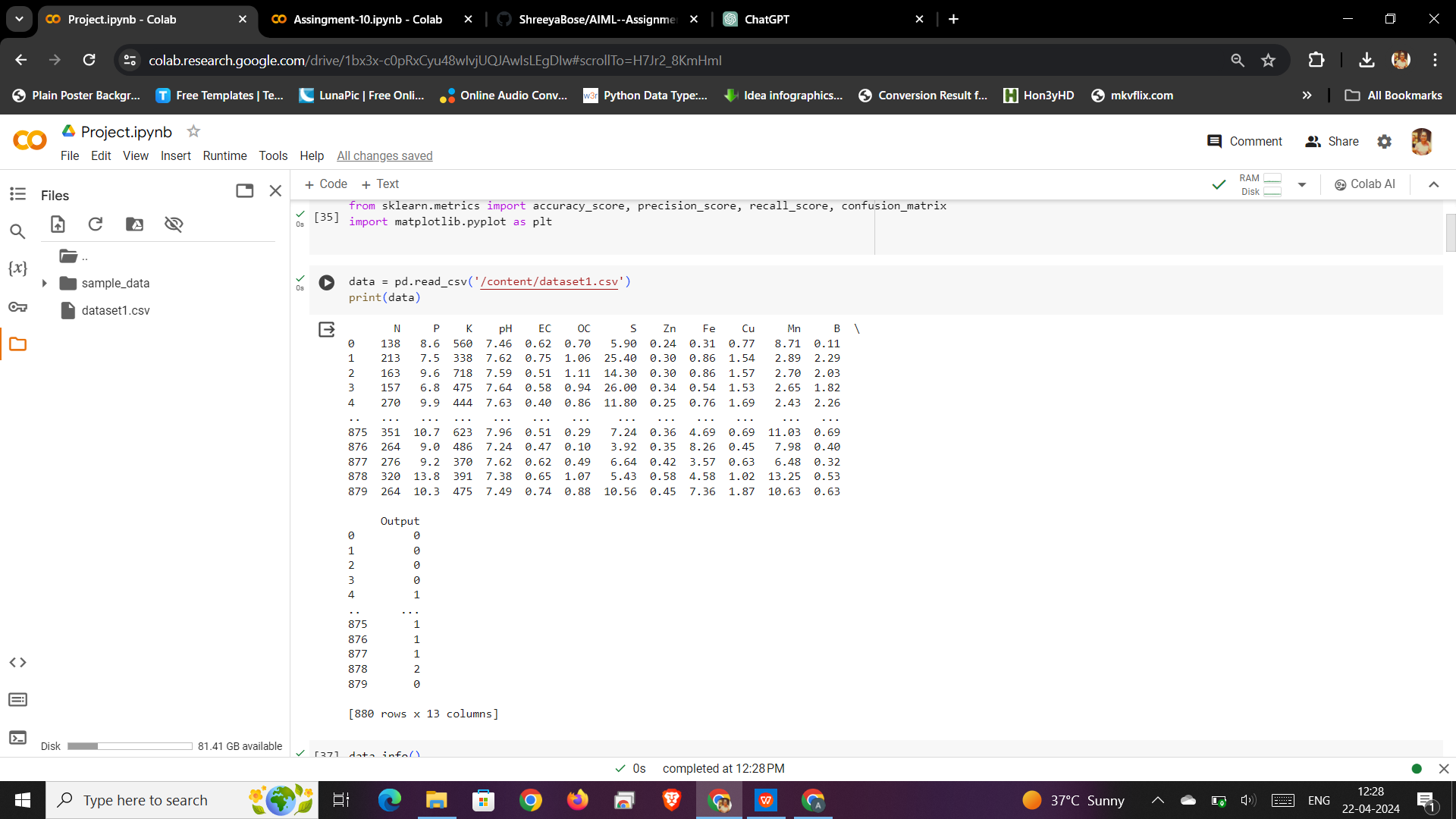
# Output

**Dataset**

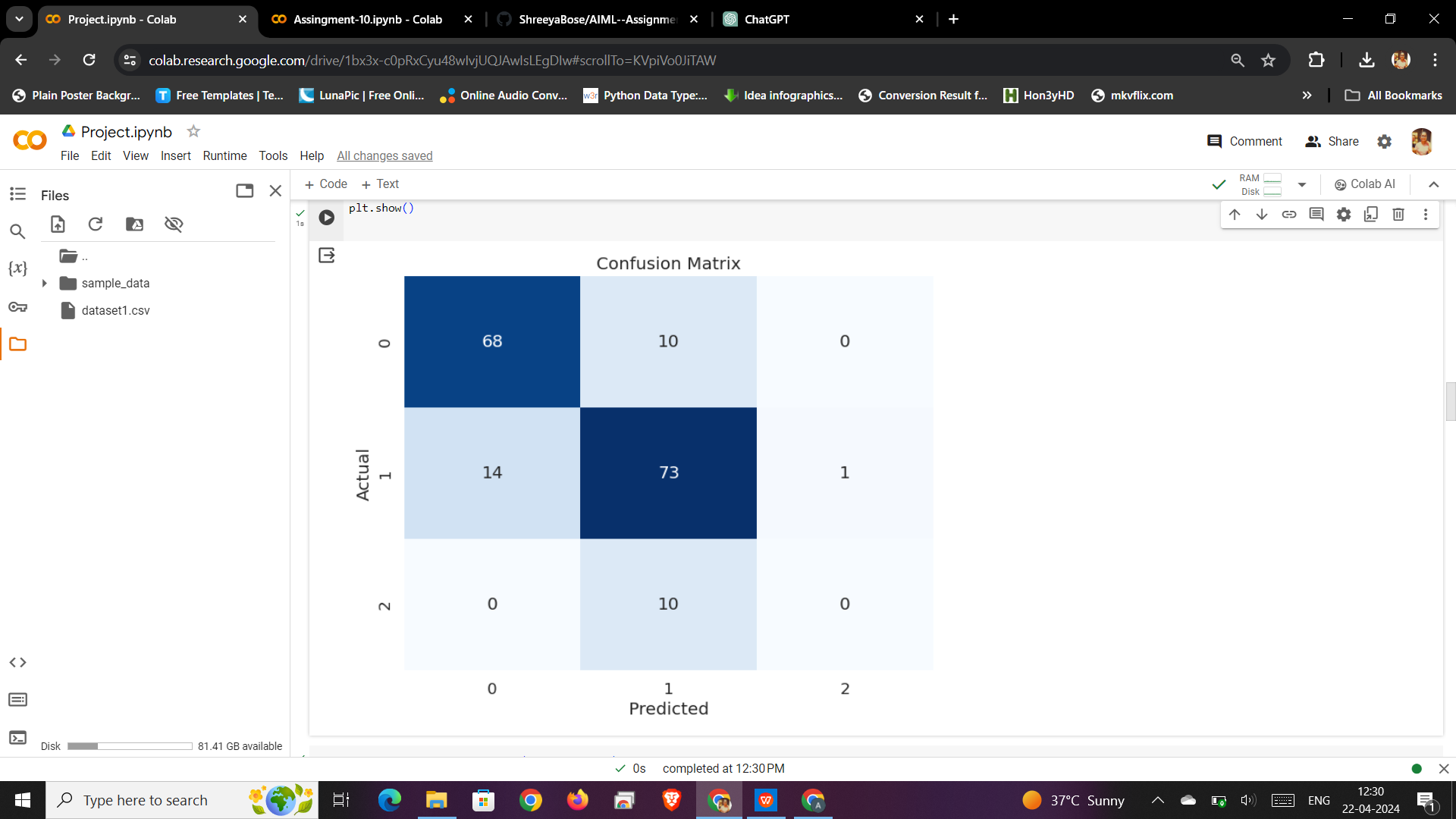


# RESULTS:

**Dataset:**

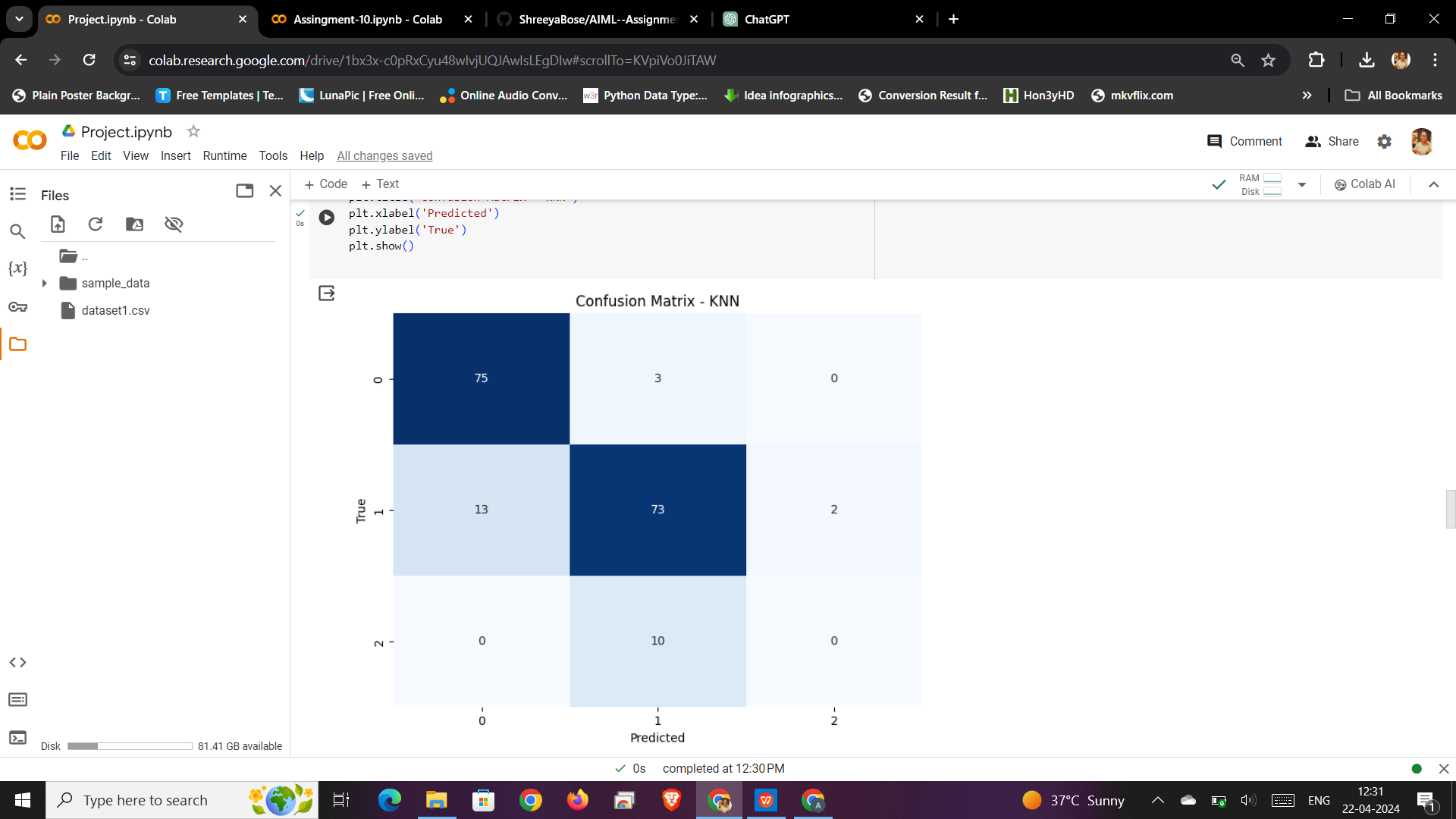


**Logistic Regression:**



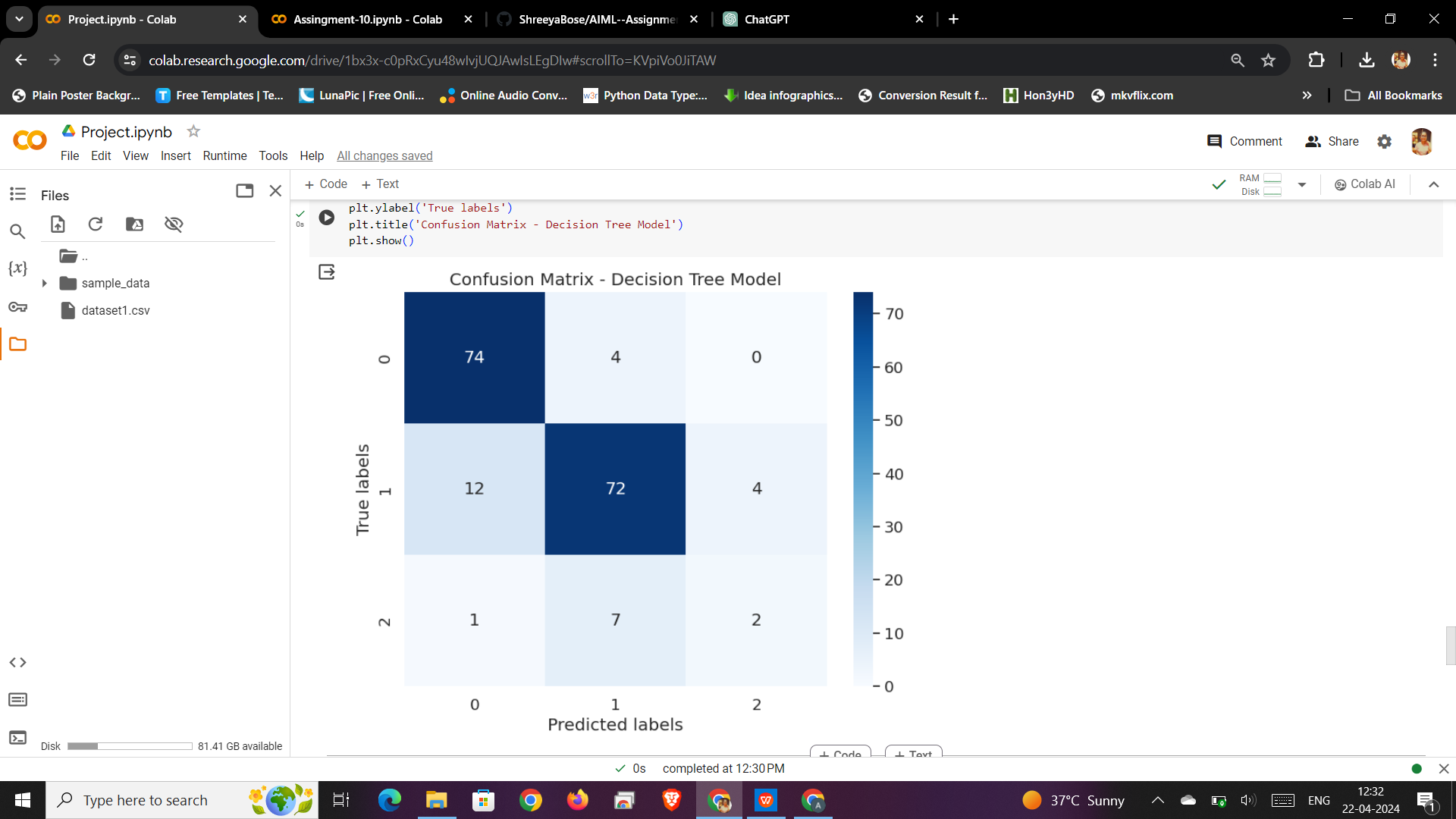
According to the graph of logistic regression it shows an accuracy of 0.8011363636363636. This indicates that the logistic regression model has a relatively high accuracy in predicting soil fertility levels.

**K Nearest Neighbor:**



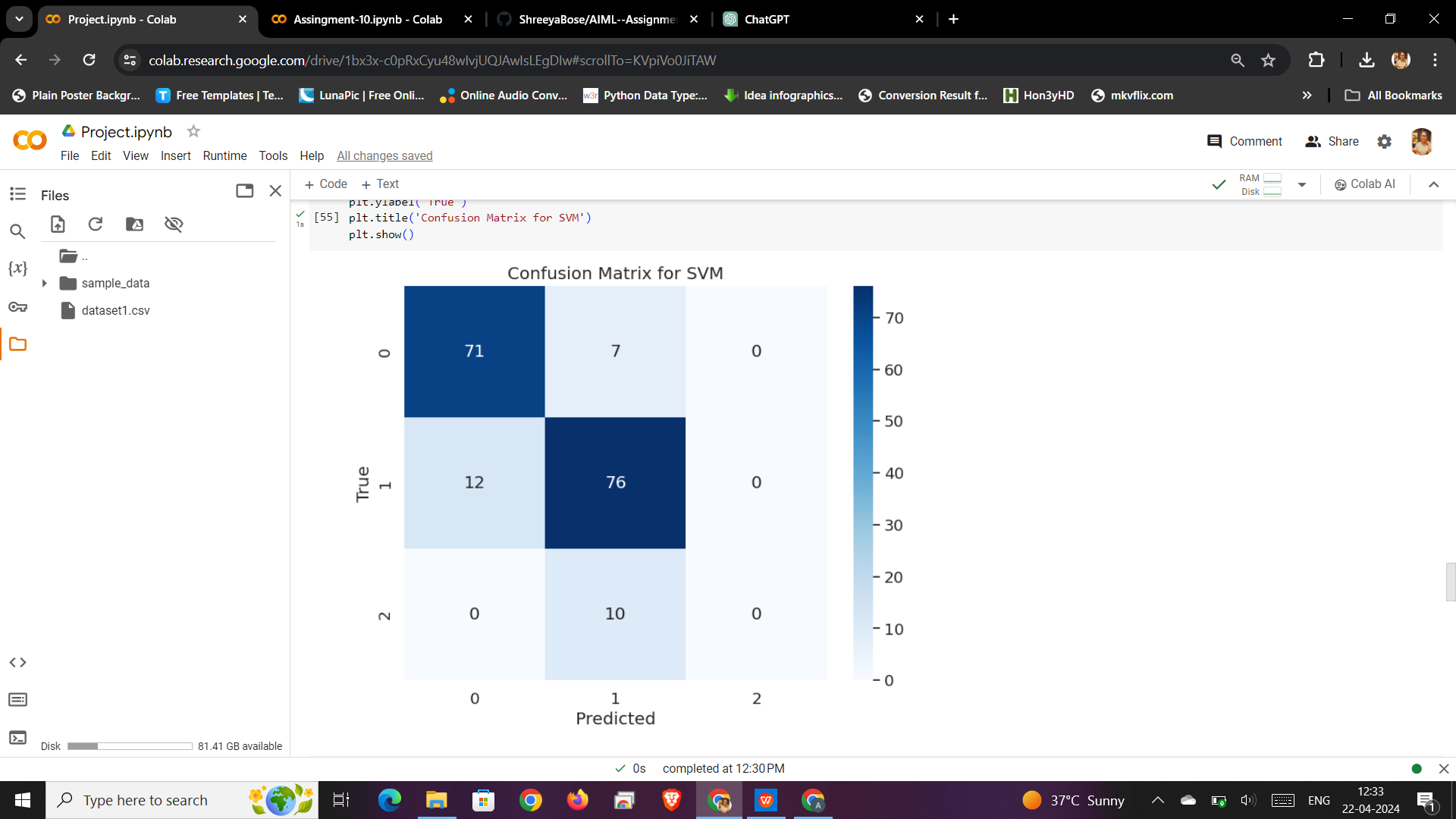
According to the KNN algorithm is a machine learning algorithm used for both classification and regression. The accuracy of the KNN model is reported to be 0.8409090909090909. This suggests that the KNN algorithm is effective in predicting soil fertility levels based on the dataset of soil attributes.

**Decision tree:**



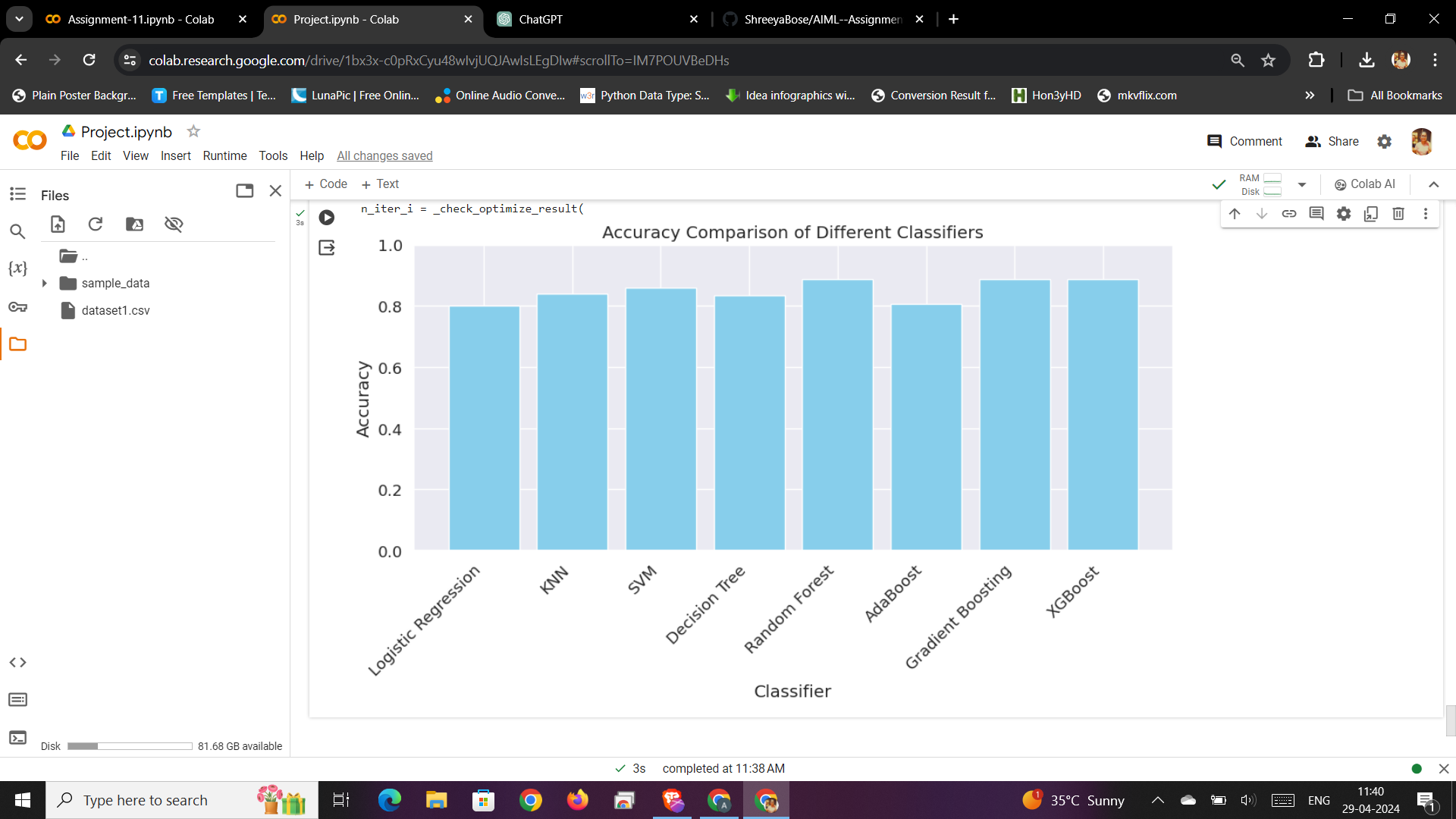
The decision tree model has an accuracy of 0.8409090909090909, which is slightly lower than Random Forest and XG Boost. Therefore, from the decision tree graph, we can conclude that while it is effective for predicting soil fertility but there are other models with higher accuracy that should also be considered.

**Support Vector machine:**



From the graph, we can conclude that the SVM model has an accuracy of 0.8352272727272727. This indicates that the SVM algorithm is effective in enhancing the accuracy of soil fertility predictions. But the accuracy of the SVM model should be considered in conjunction with other machine learning models to determine the most suitable approach for predicting soil fertility levels.

**Comparision of Accuracy between all the models:**



From the comparison of accuracy between all the models, we conclude that the Random Forest and XG Boost models have the highest accuracy at 0.8863636363636364, followed by the K-Nearest Neighbor model at 0.8409090909090909. These models have shown to be effective in predicting soil fertility levels based on the dataset of soil attributes. Additionally, the findings underscore the effectiveness of machine learning techniques in soil fertility prediction and highlight the importance of selecting appropriate models tailored to the specific characteristics of the dataset. This information can empower farmers and agricultural practitioners with valuable tools for optimizing soil management practices and enhancing crop yield potentials in precision agriculture contexts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **MACHINE LEARNING MODEL** | **ACCURACY** | **PRECISION** | **RECALL** |
| 1 | Logistic regression | **0.8011363636363636** | **0.7599897479913215** | **0.8011363636363636** |
| **2** | K-Nearest Neighbor | **0.8409090909090909** | **0.802130381510667** | **0.8409090909090909** |
| **3** | Support vector Machine | **0.8352272727272727** | **0.7877094889823223** | **0.8352272727272727** |
| **4** | Decision Tree | **0.8409090909090909** | **0.8322174041265715** | **0.8409090909090909** |
| **5** | Random Forest | **0.8863636363636364** | **0.8357566346696782** | **0.8863636363636364** |
| **6** | AdaBoost | **0.8068181818181818** | **0.7962791605292644** | **0.8068181818181818** |
| **7** | Gradient Boost | **0.8806818181818182** | **0.8871473354231975** | **0.8806818181818182** |
| **8** | XG Boost | **0.8863636363636364** | **0.8922413793103449** | **0.8863636363636364** |

# From the results above, the Random Forest and XG Boost models gives the highest accuracy, precision, and recall values among the models tested. Both models achieve an accuracy of 0.886, precision above 0.835, and recall of 0.886. Therefore, based on these metrics, the Random Forest and XG Boost models are the most effective for predicting soil fertility in this dataset.

# 6. CONCLUSION:

# Based on the provided dataset and the performance metrics of different machine learning models, the Random Forest and XG Boost models are found to be the most effective for predicting soil fertility levels. These models achieved the highest accuracy, precision, and recall values among the models tested, indicating their ability to accurately classify soil fertility levels based on the given parameters.

In conclusion, the insights gained from this study can empower farmers and agricultural practitioners with valuable tools for optimizing soil management practices and enhancing crop yield potentials in precision agriculture contexts. By using these models, farmers can make data-driven decisions about soil management, crop selection, and fertilizer application, leading to improved crop yields and reduced environmental impact.

1. **FUTURE SCOPE :**

The soil fertility prediction AIML project opens avenues for further research and application in precision agriculture. Future enhancements could involve integrating advanced sensing technologies such as IoT devices and drones for real-time data collection, enabling dynamic updates to fertility models. Additionally, incorporating spatial analysis techniques could improve the accuracy of predictions by considering spatial variability within fields. Moreover, exploring the integration of machine learning models with domain-specific knowledge, such as agronomic principles and crop-specific requirements, could enhance the predictive capabilities and practical relevance of the system, paving the way for more effective and sustainable soil management practices.

# 8 . REFERENCES:

1. https://www.sciencedirect.com/science/article/pii/S0168192315007546
2. https://www.sciencedirect.com/science/article/pii/S0168169913001257

[3] https://doi.org/10.1109/Agro-Geoinformatics.2014.6910609

[4] https://doi.org/10.1109/JCSSE.2016.7748856

[5] https://doi.org/10.1109/ICCTIDE.2016.7725357