**VISVESVARAYA TECHNOLOGICAL UNIVERSITY,**
**JNANASANGAMA, BELAGAVI – 590018**


An Internship Report
on

# DATA SCIENCE

Submitted in partial fulfillment of the award of degree
Bachelor of Engineering
in
Computer Science & Engineering

*Submitted by*

Joyline Rencita Dsouza
4SO20CS073

*Internship carried out*
*at*
CodersCave, 39, Flower Street, Saidapet, Chennai, Tamil Nadu



**Internal Guide**
Dr Saumya Y M
Associate Professor
St Joseph Engineering College

**External Guide**
Mr Manoj Kumar
Chief Executive Officer
CodersCave

Department of Computer Science and Engineering
St Joseph Engineering College
Mangaluru - 575028
**2023-24**

**VISVESVARAYA TECHNOLOGICAL UNIVERSITY,**
**JNANASANGAMA, BELAGAVI – 590018**

An Internship Report
on

# DATA SCIENCE

**Submitted in partial fulfillment of the award of degree**

**Bachelor of Engineering**
**in**
**Computer Science & Engineering**

***Submitted by***

**Joyline Rencita Dsouza**
**4SO20CS073**

**Department of Computer Science and Engineering**
**St Joseph Engineering College**
**Mangaluru - 575028**
**2023-24**

# St Joseph Engineering College
# Mangaluru - 575028
# Department of Computer Science and Engineering

# CERTIFICATE

Certified that the Internship Work title **Data Science** was carried out by **Ms JOYLINE RENCITA DSOUZA** bearing **USN 4SO20CS073**, a bonafide student of final year B.E. in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi, during the year 2023-24. Further, it is certified that all corrections/suggestions indicated during Internal Evaluation have been incorporated in this report.

| | | |
|---|---|---|
| _____ | _____ | _____ |
| **Dr Saumya Y M** | **Dr Sridevi Saralaya** | **Dr Rio D'Souza** |
| **Internal Guide** | **Head of the Department** | **Principal** |

## External Viva Voce Examination

| Name of the Examinar | Signature with Date |
|---|---|
| 1._____ | _____ |
| 2._____ | _____ |

CODERS CAVE

# CERTIFICATE

## OF COMPLETION

This certificate is presented to

*Joyline Renata Dsouza*

for successful completion of 1 Month Virtual Internship in **Data Science** with an outstanding performance at CodersCave.

MANOJ KUMAR
FOUNDER & CEO

CODERS CAVE

## Letter of Recommendation

### To whom it may concern !

This is to certify that Joyline Rencita Dsouza had successfully completed the CodersCave Virtual Internship Program with a Excellent Performance at CodersCave for a duration of 1 month.

Throughout their internship, Joyline Rencita Dsouza demonstrated an outstanding work ethic, eagerness to learn, and a genuine passion for their chosen field. They consistently exhibited a high level of professionalism and adaptability, adapting quickly to new challenges and responsibilities.

She would be a valuable assest to any employer and I recommend him strongly for any endeavours.

Regards,
Team CodersCave

+91 95970 28220
contact@coderscave.in

# DECLARATION

I, **Joyline Rencita Dsouza**, bearing **USN 4SO20CS073**, student of final year B.E. in Computer Science and Engineering, St Joseph Engineering College, Mangalore, hereby declare that the Internship Work titled **"Data Science"** has been duly executed by me from **August - September 2023**, at **CodersCave, 39, Flower Street, Saidapet, Chennai, Tamil Nadu**. Further, the "Tasks Performed" section of this report represents the work done solely by me and does not contain any statements falsely claiming work done by others, as my own.

**Date: 16th February 2024**

**Place: Mangaluru** **Joyline Rencita Dsouza**

# ACKNOWLEDGEMENT

The joy and satisfaction that accompany the successful completion of any task would be incomplete without thanking those who made it possible. I consider myself proud to be a part of St Joseph Engineering College, the institution which moulded me in all my endeavours. I express my sincere gratitude to the management for providing state of the art facilities and support for the smooth completion of the Internship.

I would like to offer my earnest gratitude to my external guide, **Mr Manoj Kumar**, Internship Coordinator and Chief Executive Officer, CodersCave platform, for providing me with valuable support throughout the period of my internship.

I owe my profound gratitude to my internal guide **Dr Saumya Y M**, Associate Professor, Department of Computer Science and Engineering, St Joseph Engineering College for her valuable guidance and support during the entire period of my internship.

I am grateful to **Dr Sridevi Saralaya**, Head of the Department, Computer Science and Engineering, for her support and encouragement.

I am indebted to my respected Principal, **Dr Rio D'Souza** for his valuable guidance and encouragement throughout the Internship program.

I am extremely thankful to **Rev. Fr. Wilfred Prakash D'Souza**, Director and **Rev. Fr. Kenneth Crasta**, Assistant Director for providing all the facilities and timely support for the completion of the Internship.

I wish to express my sincere gratitude to all the Faculty and Technical staff of the Department and friends and family for their valuable help and support during the period of my internship .

# Executive Summary

I carried out my Internship as Data Science Intern at CodersCave, located in Chennai, Tamil Nadu, India, from August 16, 2023, to September 16, 2023. CodersCave is a service-based company distinguished for its expertise in Software Development and Training. CodersCave specializes in providing software development and training services, standing out as a leader committed to creating practical learning opportunities for individuals in the constantly changing tech field.

The primary goal of my internship was to contribute actively to real-world projects at CodersCave by applying data science concepts. Utilizing tools such as Python and machine learning, my tasks revolved around hands-on learning experiences focused on data analysis and interpretation. This approach aligned seamlessly with CodersCave's mission, fostering practical application of data science concepts.

The outcomes of my internship were significant, encompassing both technical and non-technical realms. On the technical side, engaging in data analysis, machine learning techniques, and Python programming provided a robust foundation in real-world data science applications. This experience not only refined my skills but also encouraged the exploration of new technologies and tools within the field. As per the non-technical side, the internship played a pivotal role in the development of effective communication and time management skills, offering valuable insights into collaborative dynamics within a tech-focused organization. I was able to complete the tasks and projects assigned to me within the given time. This internship also provided me an opportunity to apply acquired knowledge to real work experiences.

**Joyline Rencita Dsouza**

# Table of Contents

# List of Figures

# Chapter 1

# Company Profile

## 1.1 Brief History

CodersCave, located in Chennai, Tamil Nadu, India, stands out as a prominent leader in Software Development and Training. Specializing in project-based learning, our commitment is to provide future technology training and project development skills. Dedicated to sculpting a resilient tech future for developers, CodersCave focuses on practical learning experiences, offering hands-on opportunities for individuals to navigate the ever-evolving tech landscape. With a commitment to excel, CodersCave provides a platform that bridges the gap between theoretical knowledge and practical application in the dynamic field of technology.

## 1.2 Services Offered By The Company

- Software Training and Development
- IEEE Projects
- Ph. D Guidance & Assistance
- Internship Program
- Manpower Consultant & Placement

## 1.3 Contact Details

Address : 39, Flower Street, Saidapet, Chennai, Tamil Nadu 600015

Phone : +91 959 702 8220

Email : contact@coderscave.in

Official Website : https://www.coderscave.in/

# Chapter 2

# About the Department

## 2.1 Introduction

The internship serves as a valuable platform for students, providing active involvement in real-time projects and facilitating a smooth transition into practical, hands-on experiences. The program introduces participants to the dynamic field of data science. The department's primary focus is on offering interns real-world exposure and the application of data science concepts. Guided by passionate professionals with extensive industry experience and certifications, the internship is designed to immerse participants in the practical aspects of data science, encompassing a diverse range of tools and technologies.

## 2.2 Roles and Responsibilities

During the internship, my responsibilities encompassed daily assignments facilitated by an external mentor. These tasks were meticulously crafted to consolidate and apply theoretical knowledge acquired in academic settings to practical scenarios. Within the realm of data science, my project implementation spanned a broad spectrum of essential functions, including comprehensive data preprocessing, utilization of machine learning algorithms for predictive analytics, rigorous statistical analysis, proficient data visualization techniques, and thorough model evaluation protocols. This immersive

hands-on experience with an array of tools and technologies prevalent in the field significantly augmented both my theoretical understanding and practical proficiency in data science.

- Data Pre-processing
- Machine Learning Algorithms
- Statistical Analysis
- Data Visualisation
- Model Evaluation
- Domain Knowledge
- Feature Engineering

# Chapter 3

# Tasks Performed

## 3.1 Weekly Report

### 3.1.1 Week 1:

- Brief introduction about the company was addressed and self introductions were proposed.
- Basic technical questions were asked and made to interact with each other.
- Explored Python programming features and fundamentals. Also covered Python basics, including data types, variables, and loops.
- Received support during the internship for setting up the environment in Visual Studio and downloading necessary packages.
- Was introduced to and instructed on the extensive use of Jupyter notebooks for coding exercises, data analysis, and project implementation.
- Discussed module installation with PIP, focusing on libraries like NumPy, Matplotlib, Scikit-learn, Seaborn, and other essential packages.
- Engaged in hands-on Python exercises for practical understanding.

### 3.1.2 Week 2:

- Discussed the basic concepts of Machine Learning.

- Gained theoretical knowledge of Machine Learning categories, including Supervised Learning, Unsupervised Learning, Reinforcement Learning, and Semi-supervised Learning.
- Introduced to various Statistical Concepts, crucial for understanding Machine Learning.
- We actively participated in the session which was organized to classify the real world examples of ML categories.
- Learned concepts like Correlation, Covariance, and Statistical Distributions including Binomial and Normal distribution, along with the relevant mathematical formulas.
- Applied Statistical methods to test hypotheses, resembling tasks encountered in Machine Learning.
- Explored Probability theories, emphasizing their significance in decision-making within Machine Learning.
- Learned and applied Exploratory Data Analysis theory for rescaling, standardizing, and normalizing data.

### 3.1.3 Week 3:

- First, We were assigned individual projects, and I was assigned to predict if a patient is diabetic or not using machine learning.
- Utilized datasets from open sources, such as Kaggle, for project implementation.
- Learned different Machine Learning Algorithms like Linear and Logistic regression, Naive Bayes, K-Means Clustering, K-Nearest Neighbour Classification algorithm, Random Forest, and Support Vector Classifier.

- Pre-processing techniques necessary for Machine Learning and their importance was taught to us by our mentor.
- Received an introduction to testing and training data with Python codes for simple manipulations, such as displaying the data and generating graphs.

### 3.1.4 Week 4:

- Our project focused on making a Machine Learning model to tell if someone has diabetes or not.
- Handled data by removing duplicates, outliers, and addressing null values.
- Explored and analyzed data specifically related to Diabetes prediction through Exploratory Data Analysis (EDA).
- Constructed predictive models employing various algorithms including Logistic regression, Naive Bayes, K-Nearest Neighbour Classification algorithm, Random Forest, Support Vector Classifier, Decision Tree Classifier and Artificial Neural Network.
- Evaluated model effectiveness by analyzing key metrics such as precision, recall, f1-score, and support, to determine if a person has diabetes or not.
- Additionally, the project was uploaded to our GitHub repository, and the repository link was submitted to our mentor before the deadline.

# Chapter 4

# Project Implementation

## Title : Diabetes Prediction using Machine Learning

## 4.1 Project Description

During my internship, I embarked on a project centered around diabetes analysis with the primary objective of predicting an individual's diabetic status before traditional medical analysis, utilizing Exploratory Data Analysis (EDA) and Machine Learning (ML) concepts. The project aimed to streamline predictions through ML, allowing for early identification of diabetes risk factors[4]. We employed various models and techniques, resulting in diverse levels of accuracy in predicting outcomes. This project explored the details of analyzing diabetes and demonstrated how Machine Learning can improve early healthcare actions.

## 4.2 Features:

- The dataset is made to ignore the exceptional cases which will degrade the model's efficiency. It is taken care of in the pre-processing of the dataset.
- Performing EDA on the cleaned dataset also helps in looking for relationships between different variables.
- Identifying relationships between variables.

- The efficient Exploratory Data Analysis that is carried out will make sure that only the necessary and important parameters in the dataset are considered for the fair evaluation of the models.
- Since different Machine Learning models are being created, the most efficient one is being used to predict the outcome of the given values.

## Steps For Implementation of Machine Learning Models:

1. **Importing the ML model libraries and requirements.**
2. **Predicting the Test set Results.**
3. **Calculation of Confusion Matrix and Accuracy Score.**

## 4.3 Project Snapshots:

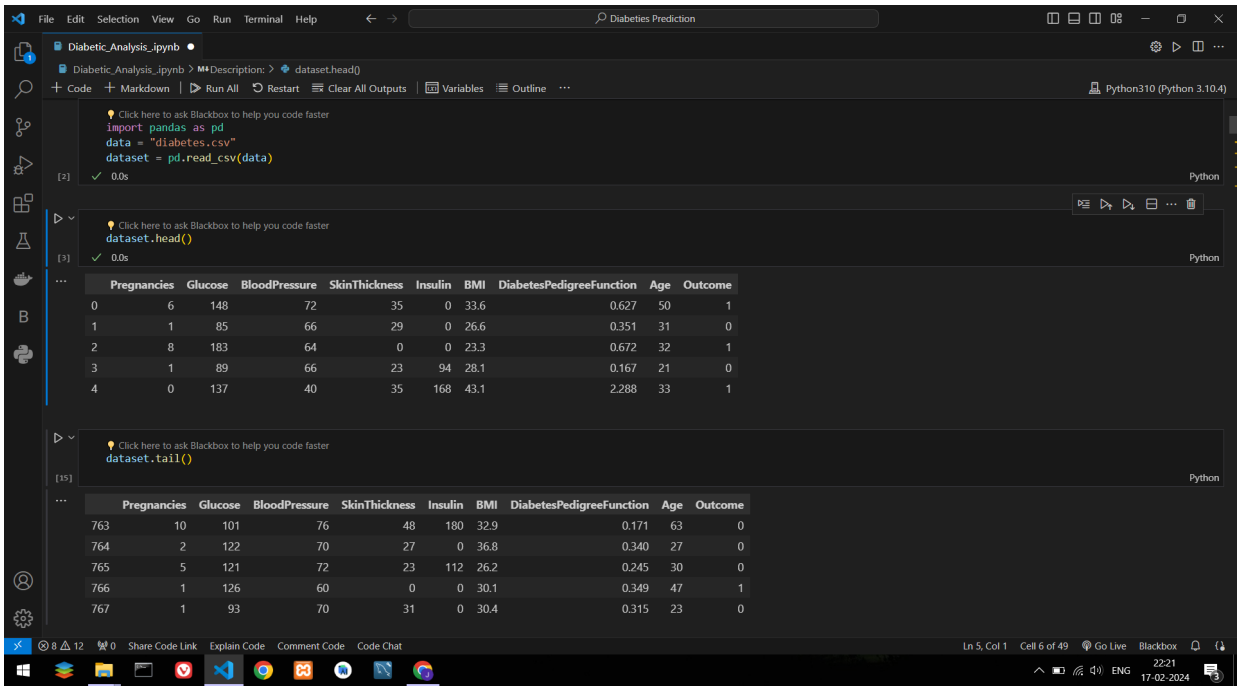


**Figure 4.1: The Initial Dataset Table**

Figure 4.1 showcases key attributes of the initial dataset, including glucose levels, blood pressure, age, skin thickness, insulin levels, BMI, and other pertinent factors [4]. These attributes serve as foundational components for subsequent analysis and insights generation in our study.

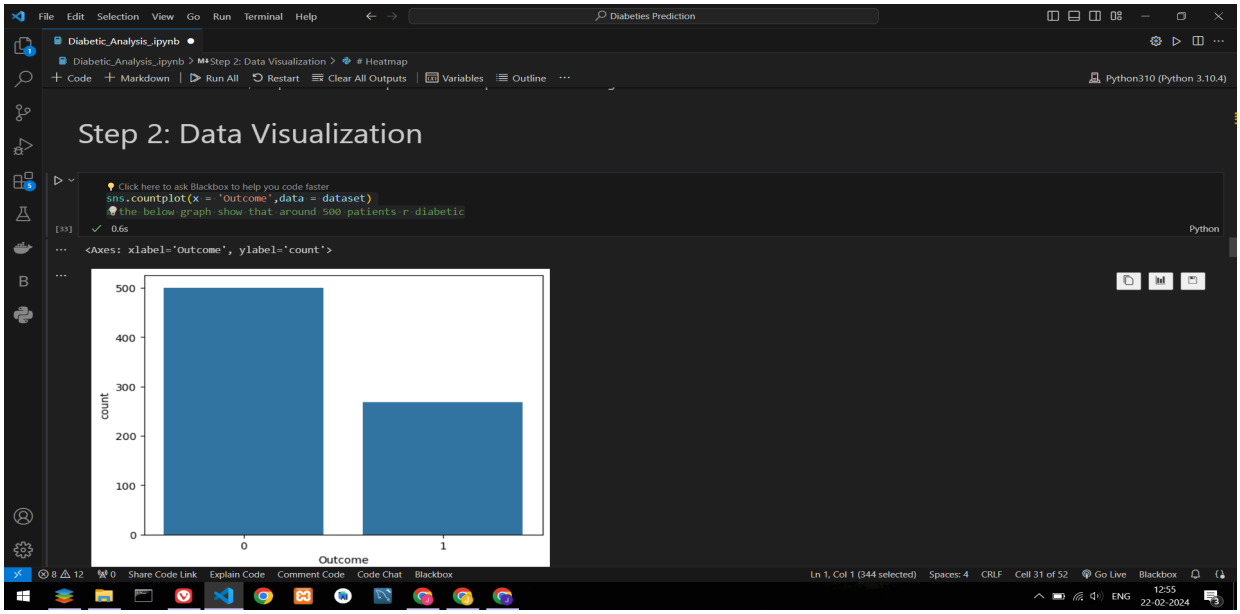


**Figure 4.2: Data Visualisation**

Figure 4.2 shows the count plot of diabetic and non-diabetic patients

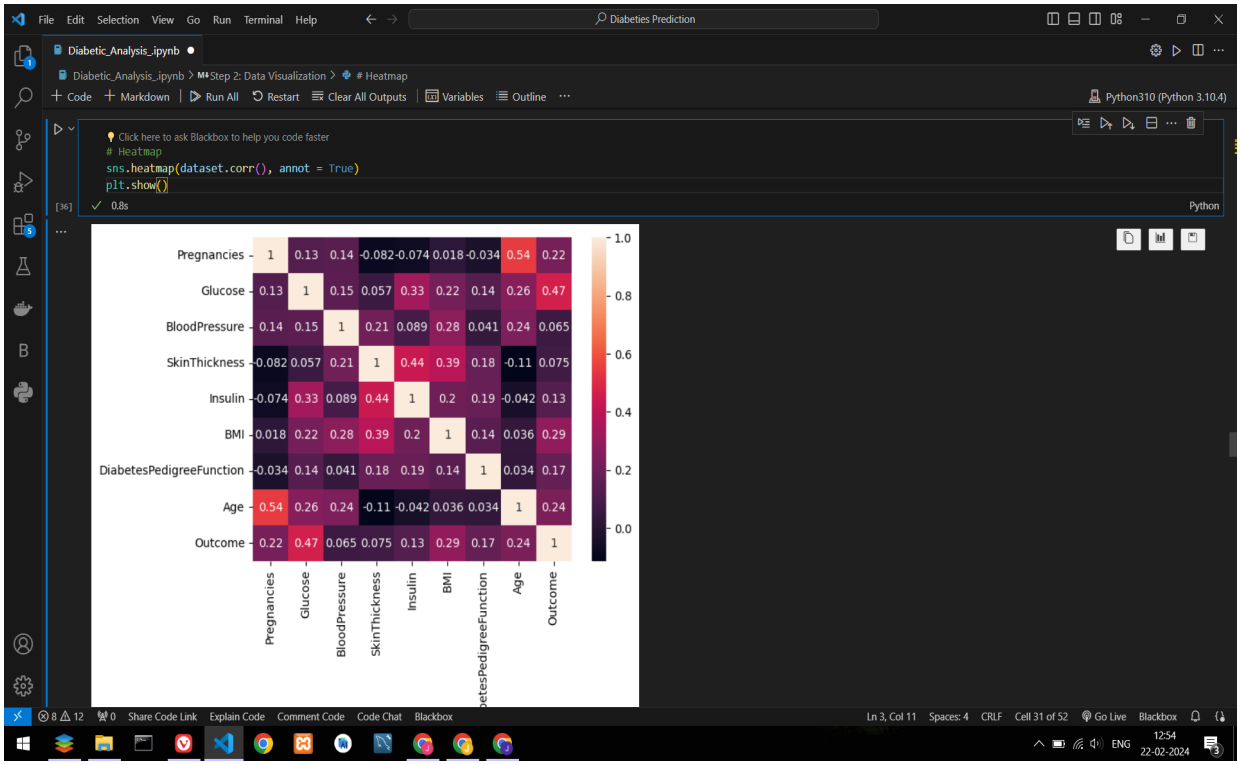


**Figure 4.3: Heatmap**

Figure 4.3 presents heatmap which displays feature correlations which describes interfeature relationships.

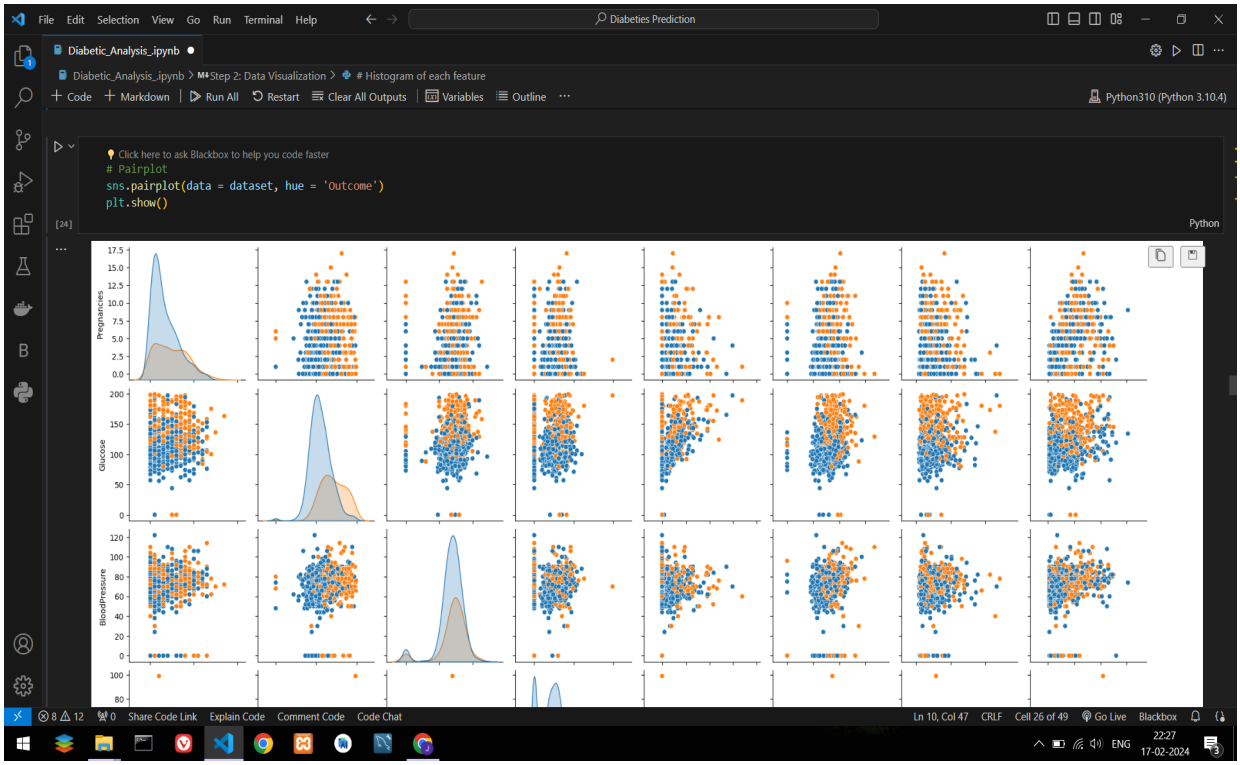


**Figure 4.4: PairPlot**

Figure 4.4 presents pair plot which visualises the relationship between each pair of variables in a dataset.

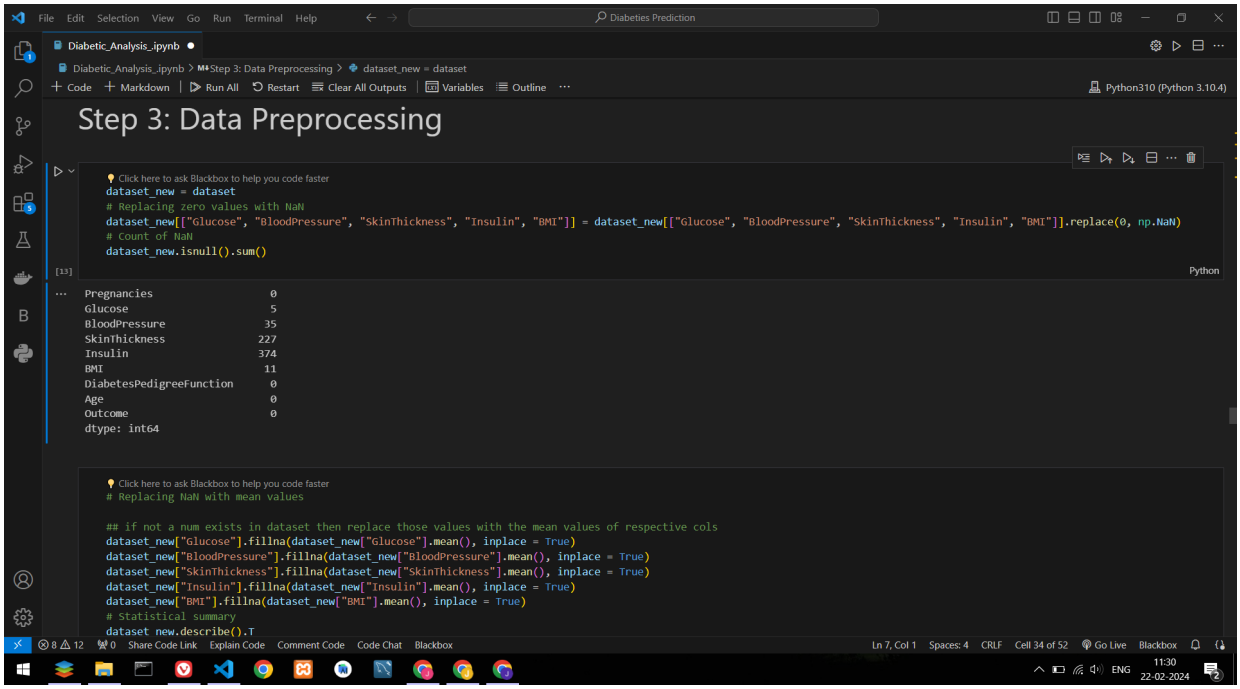


**Figure 4.5: Data Pre-Processing**

Figure 4.5 showcases the dataset pre-preprocessing, where zero values in certain features are replaced with NaN to handle missing or unrealistic entries.

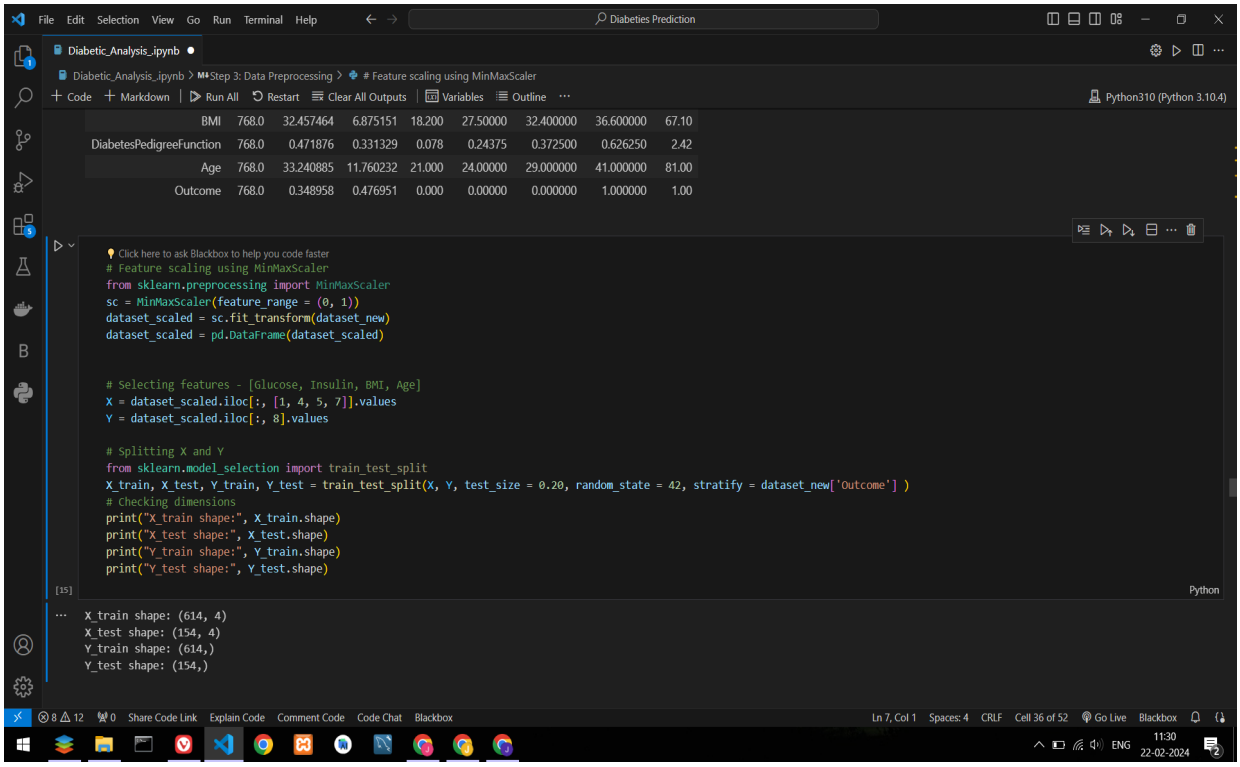


**Figure 4.6: Feature Scalling and Splitting the Data**

Figure 4.6 demonstrates the process of feature scaling applied to the dataset, ensuring uniformity in feature ranges for effective modeling.

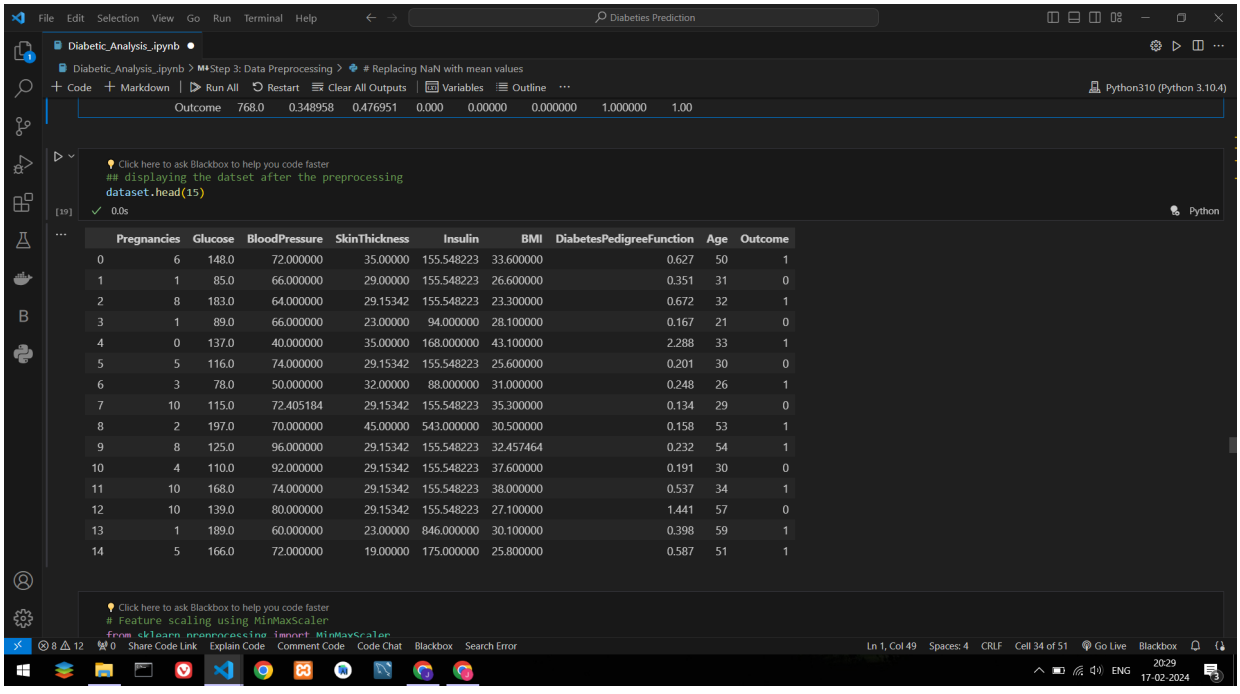


**Figure 4.7: Final Dataset after Pre-Processing**

Figure 4.7 showcases the final dataset after pre-processing, reflecting enhanced data quality for further analysis.

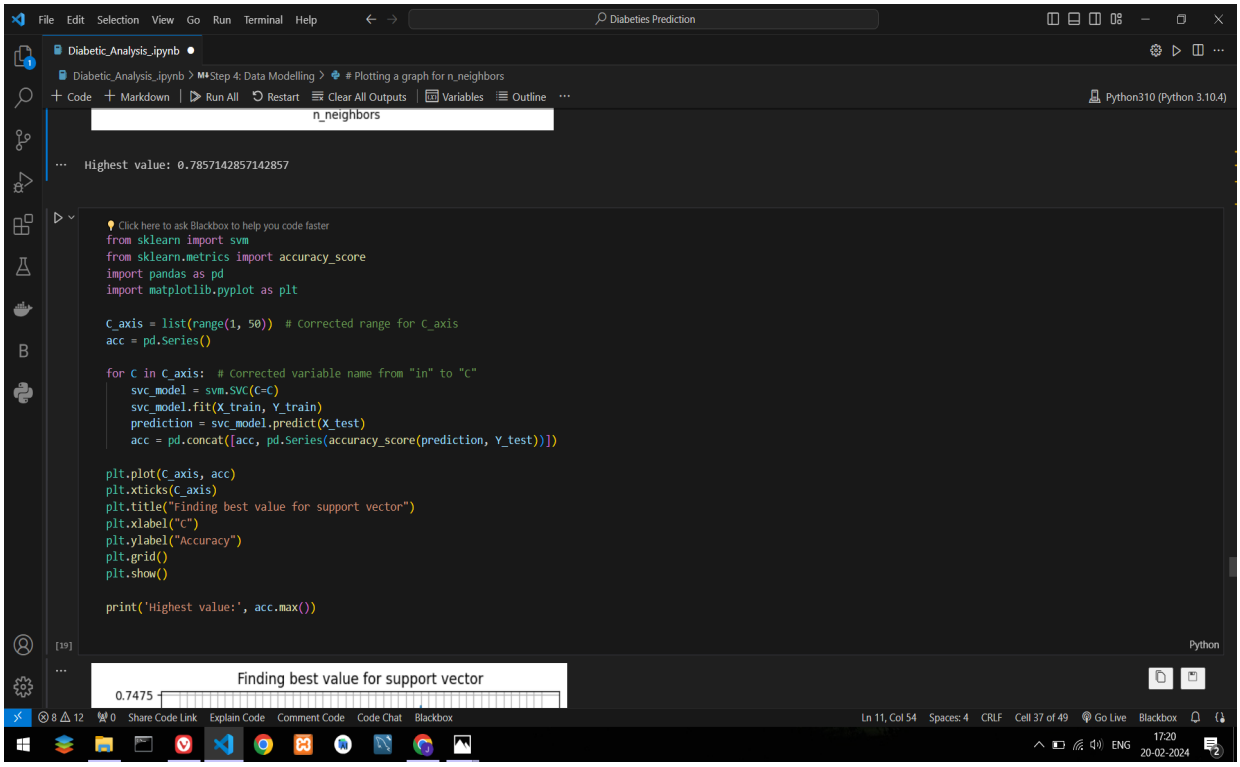


**Figure 4.8: Building Models**

Figure 4.8 presents the code snippet representing the building process of predictive models, facilitating data analysis.

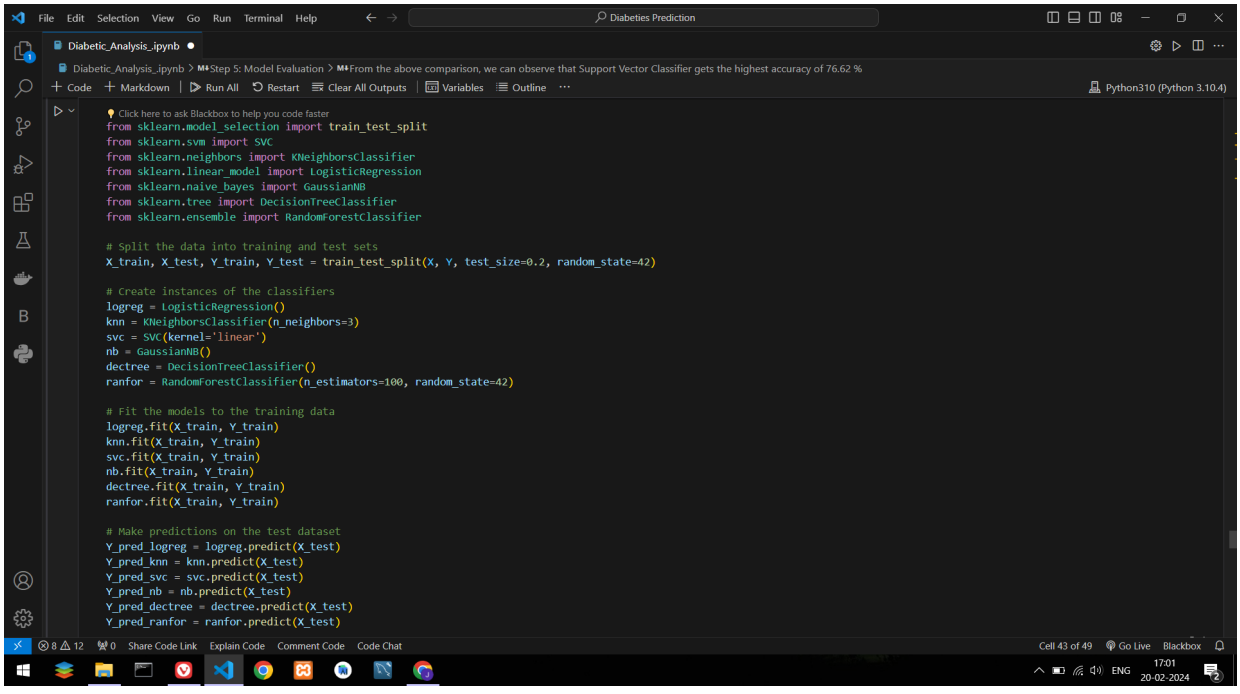


**Figure 4.9: Creating the instances of the Classifier**

Figure 4.9 Creating models for different machine learning algorithms to train and make predictions.

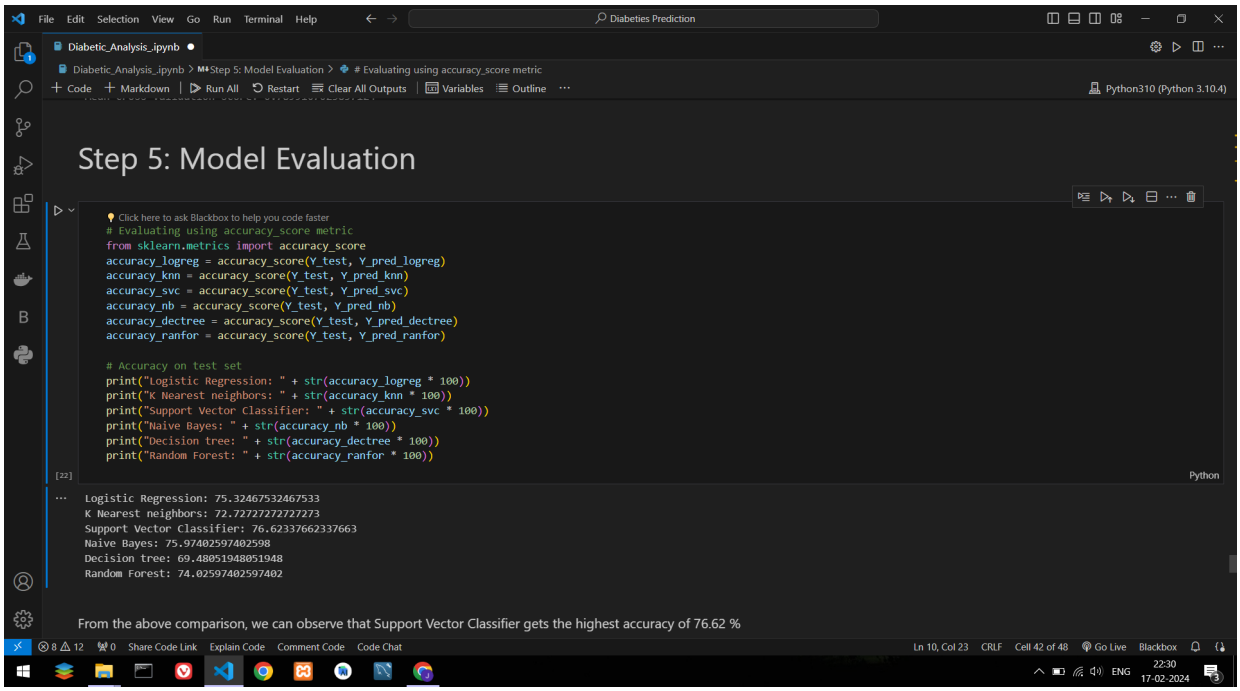


**Figure 4.10: Evaluation of the performance of Each Model**

Figure 4.10 presents the assessment of each model's performance, providing insights into their effectiveness in handling the dataset.

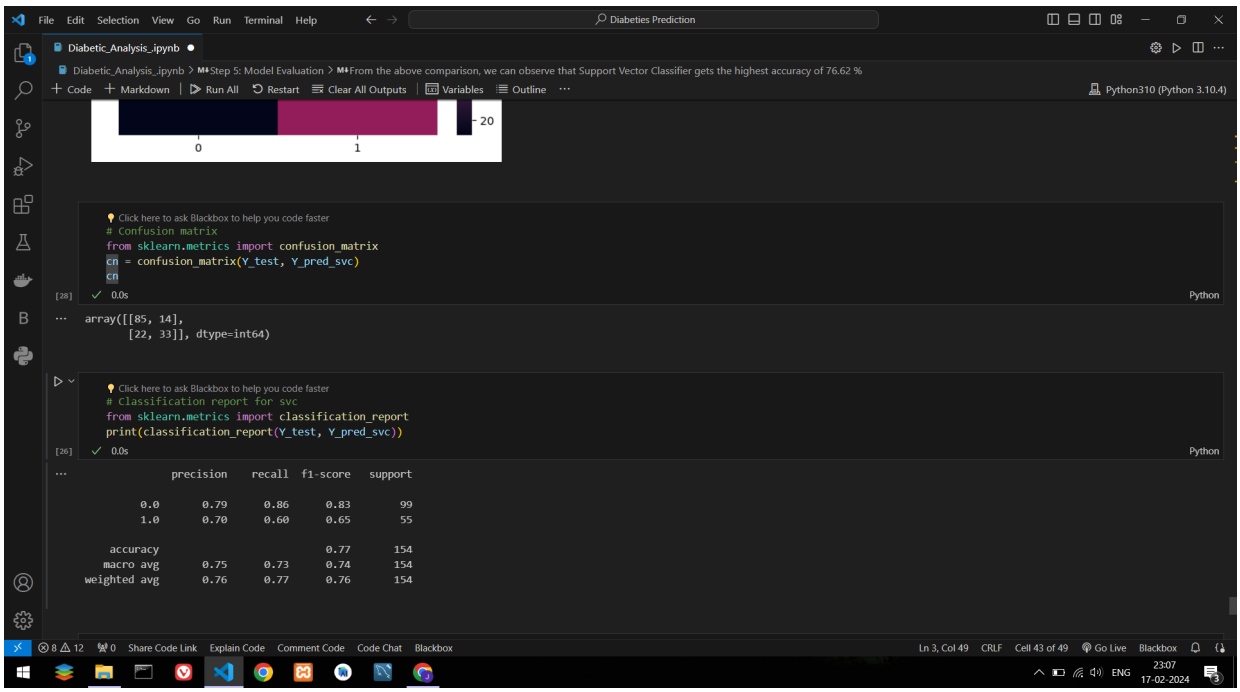


**Figure 4.11: The Accuracy scores of the models implemented**

Figure 4.11 This figure displays the confusion matrix and accuracy scores for the implemented models, offering a comprehensive evaluation of their performance.

# Chapter 5

# Reflection Notes

## 5.1 Experience:

It was a memorable experience interning at CodersCave. I learnt the tools and techniques used in Machine Learning. Working with a team full of energetic and dedicated students motivated me to pursue this course more enthusiastically. I was also able to sharpen my communication and management skills.

During this internship, I learned the Python programming language, which is specifically designed for Machine Learning. I performed all my tasks in Anaconda - Jupyter Notebook, a powerful tool well-suited for executing Python programs. Additionally, I utilized the VSCode Jupyter extension, which provided seamless support for creating, opening, and editing Jupyter Notebooks directly within VSCode. This integration enhanced my workflow and allowed me to leverage the features of VSCode while working on Machine Learning projects.

I also gained knowledge about the basic concepts of Machine Learning and how to implement them to build predictive models. As part of the internship, I worked on a project to predict if a person has Diabetes or not. Our guide at the company was friendly and supported us at every stage of our project. Their guidance was invaluable and contributed to the successful

completion of the project without any difficulties.

## 5.2 Technical Outcomes:

- **Applied Mathematics:**
  Learned the Statistical Math necessary for the pre-processing of the data used for building Machine Learning models[1]. The usage of the various Statistical distributions, and parameters was also learnt in depth.

- **Python Programming and Libraries:**
  I also got to know the importance of Python and its libraries like NumPy, Matplotlib, Scikit-Learn in Machine Learning[3]. I was able to use the IDE that is suitable to implement Machine Learning algorithms in an easy manner.

- **Data Modelling and Evaluation using ML:**
  I was able to learn the process of data modeling in Machine Learning. I also learnt the logic behind cross validating the results in Machine Learning models. Learned various Machine Learning models and was able to find the most accurate results using these models[2].

## 5.3 Non-Technical Outcomes:

- **Personal Growth:**
  A project challenges us to step out of our comfort zone, take on new responsibilities, and adapt to unfamiliar situations.These experiences contribute to personal growth and development, fostering resilience and a willingness to learn. Completing a project can boost confidence and self-esteem. It validates your abilities and gives you the reassurance that you can excel in a professional setting

- **Problem Solving Skills:**

  Developing problem-solving skills requires critical thinking, which involves analyzing situations, identifying patterns, and evaluating different solutions.These skills help us to solve real-life problems which we face while handling a project.

- **Time Management:**

  Time management refers to the way of organizing and planning how long one spends on specific activities. It may seem counter-intuitive to dedicate precious time to learning about time management, instead of using it to get on with our work.It improves the essence of the work we are engaging in and gives us time to complete more tasks thereby helping us finish our work.

- **Communication Skills:**

  Effective communication is essential for conveying ideas, discussing project requirements, and collaborating with team members. Through project work, individuals enhance their ability to articulate thoughts clearly, listen actively, and express themselves persuasively. Strong communication skills facilitate successful project outcomes and foster positive working.

- **Adaptability:**

  Project work often involves navigating unforeseen challenges, changing requirements, and evolving circumstances. Developing adaptability enables individuals to respond effectively to changes, adjust strategies as needed, and remain resilient in the face of uncertainty. Adaptability is a valuable skill in dynamic work environments and contributes to overall project success.

# References


1. Statistical Analysis: https://www.simplilearn.com/what-is-statistical-analysis-article
2. Machine Learning: https://www.geeksforgeeks.org/machine-learning
3. Python Programming: https://www.w3schools.com/python/
4. Machine Learning Algorithms For Diabetes Prediction And Diagnosis: https://www.sciencedirect.com/science/article/pii/S1877050921014629
5. Data Science LifeCycle: https://www.geeksforgeeks.org/data-science-lifecycle/