**Machine Learning Using Python**

*An Internship report submitted in partial fulfillment of the requirements*

*For the award of the Degree of*

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE ENGINEERING

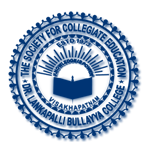
By

D. Manasa

(Regd. No.: 320136410034)

K. Aiswarya Jahnavi

(Regd. No.: 320136410053)



**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**

**Dr. L. BULLAYYA COLLEGE OF ENGINEERING**

(Permanently Affiliated to Andhra University, Visakhapatnam)

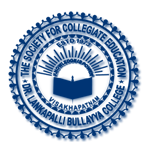
New Resapuvanipalem,Visakhapatnam-530013

Year of submission:2023

**Dr. L. BULLAYYA COLLEGE OF ENGINEERING**

**New Resapuvanipalem, Visakhapatnam-530013**

**Department of Computer Science Engineering**



**Bonafide certificate**

This is to certify that Miss **D. Manasa, and K. Aiswarya Jahnavi** bearing register numbers **320136410034, 320136410053** students of final year B.Tech in Computer Science Engineering, have carried out the internship work titled ”**Machine Learning Using Python”** at Inspire Softech Solutions, Chennai during the academic year **2022-23.**

**Head of the Department**

Dr. D. Madhavi

Professor

Dept. of Computer Science Engineering

**Industry Certificate Scan copy**





**ACKNOWLEDGEMENT**

I thank **Prof. D. Deepak Chowdary**, Principal of Dr. L. Bullayya College of Engineering for permitting us to carry out this Internship.

I would like to thank **Prof. D. Madhavi***,* HOD, Dept. of C.S.E, Dr.L.Bullayya College of Engineering, Visakhapatnam, for her guidance and assistance in producing this work.

I am deeply indebted to my Internship guide **Mr. Saravana Kumar**, Technical Educator, Inspire Softech Solutions, for guiding me throughout the project in spite of his busy schedule.

Also deserving of thanks are my family and friends, for their support and their confidence in my achievements.

**D. Manasa**

**(320136410034)**

**K. Aiswarya Jahnavi**

**(320136410053)**

**CONTENTS OF THE REPORT Page No.**

1. Abstract. 6
2. About the Organization. 7
3. **CHAPTERS**

Learning Objectives. 8

Prerequisite Basic Skills. 8

Learning Process. 9

Learning Outcomes. 16

Executive Summary. 17

1. Appendices. 18

**Abstract**

This will help to grasp the fundamental concepts of machine learning, including supervised and unsupervised learning, regression, classification, clustering, and model evaluation.

An emphasis will be placed on utilizing Python libraries such as Numpy, Matplotlib, TensorFlow, Scikit-learn, and Pandas to implement machine learning algorithms effectively. To gain proficiency in data manipulation, preprocessing, and model development. The internship focuses on a variety of machine learning algorithms such as linear regression, decision trees, support vector machines, neural networks, and ensemble methods. The learning objective is to understand the principles behind these algorithms and their practical applications.

Understanding the importance of model evaluation, hyperparameter tuning, and optimizing machine learning models for improved performance and generalization are crucial parts of this internship.

* Through collaborative discussions, peer-to-peer learning, and guidance from experienced mentors, encouraged to explore innovative solutions and gain insights into best practices in machine learning.

**Organization Information**

Inspire has highly qualified professionals trained in the latest cutting-edge technology

to train and support, corporations, universities, colleges, schools, and government departments,

to enhance individual skills for better performance.

They assure us that all training for skill enhancement we provide will have a positive impact on the performance driven by competition in the industry. Their training motivates every individual to achieve their goals and throttle their full potential thereby ensuring performance delivery.

They assure that all training for skill enhancement they provide will have a positive impact on the performance driven by competition in the industry. Their inspiring training motivates every individual to achieve their goals and throttle their full potential thereby ensuring performance delivery. They bring instant solutions in training to increase human performance through life and business skills. They link their training skills to our organizational initiatives and professionally deliver to meet our Learning and development (L&D) requirements. Swapping 'multi-caps', they perform different roles like a coach, facilitator, trainer, teacher, consultant, architect, and SME (Subject Matter Experts). They provide Career, Professional, and Personal Skills through their training services, workshops, seminars, and lectures.

**Learning Objectives**

Initially, we aim to solidify our Python programming skills, delving into basic data structures and syntax. This project guides us in grasping the fundamental concepts of machine learning and distinguishing between supervised and unsupervised learning. Through hands-on exploration and preprocessing of datasets with Python libraries like Pandas and visualization tools like Matplotlib and Seaborn, we learn to handle missing data and outliers. Implementation of core machine learning algorithms, including linear regression, decision trees, and k-nearest neighbors, is achieved using popular libraries like Scikit-learn. Feature engineering becomes a focal point, emphasizing the importance of creating and transforming features. The project guides us through model evaluation using metrics such as accuracy and precision, delving into hyperparameter tuning for optimization. Introduction to neural networks and deep learning, deployment strategies, integration into applications, project management using version control (e.g., Git), and ethical considerations add depth to the learning experience. Communication and documentation skills are honed to effectively convey findings to both technical and non-technical stakeholders. The project fosters a mindset of continuous learning, encouraging exploration of advanced topics and complex machine learning models in this dynamic field.

**Learning Process**

**Feasibility Study:**

**The technical feasibility** of this internship involves an evaluation of the technological aspects required for its successful implementation. This includes assessing the compatibility of the curriculum with the latest versions of relevant programming languages, libraries, and frameworks. It ensures that participants have access to up-to-date tools commonly used in the field of machine learning.

**The economic feasibility** of this internship involves assessing the financial aspects associated with its implementation. This includes a cost-benefit analysis to determine whether the investment in terms of time, money, and resources is justified by the potential outcomes and benefits.

**Requirement Specifications:**

The requirement specifications for this internship typically encompass the following aspects:

1. Prerequisites:

Basic knowledge of programming, preferably in Python.

Understanding of fundamental mathematical concepts, especially those related to statistics.

Familiarity with data structures and algorithms.

2. Software and Tools:

Proficiency in Python programming language.

Knowledge of relevant libraries and frameworks for machine learning, such as NumPy.

Experience with Jupyter Notebooks or other relevant coding environments.

3. Hardware Requirements:

A computer or laptop with sufficient processing power to handle machine learning tasks.

Adequate storage space for datasets and machine learning models.

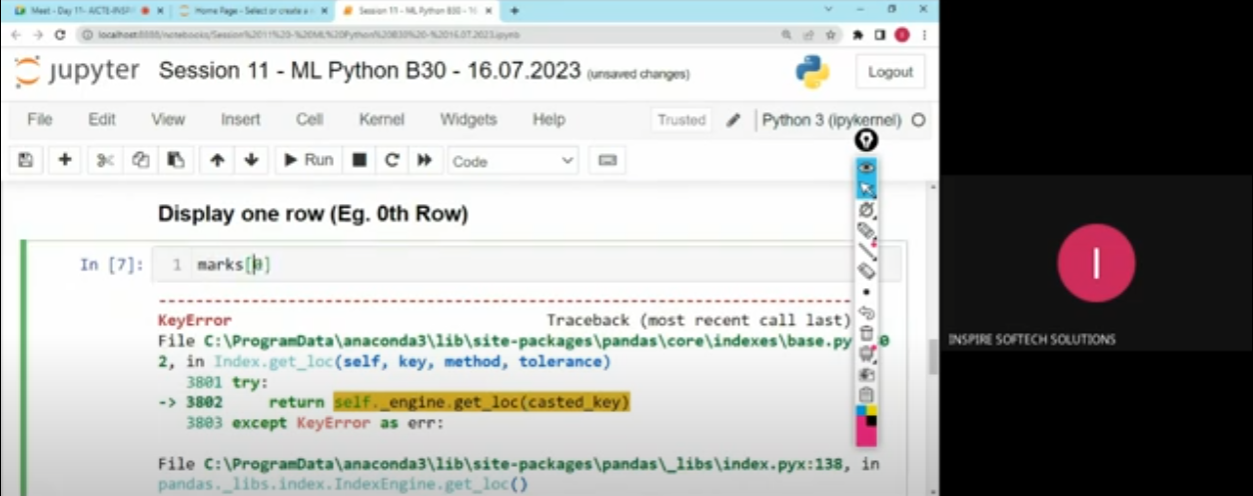
4. Project Work:

Willingness and ability to work on practical projects to apply the learned concepts.

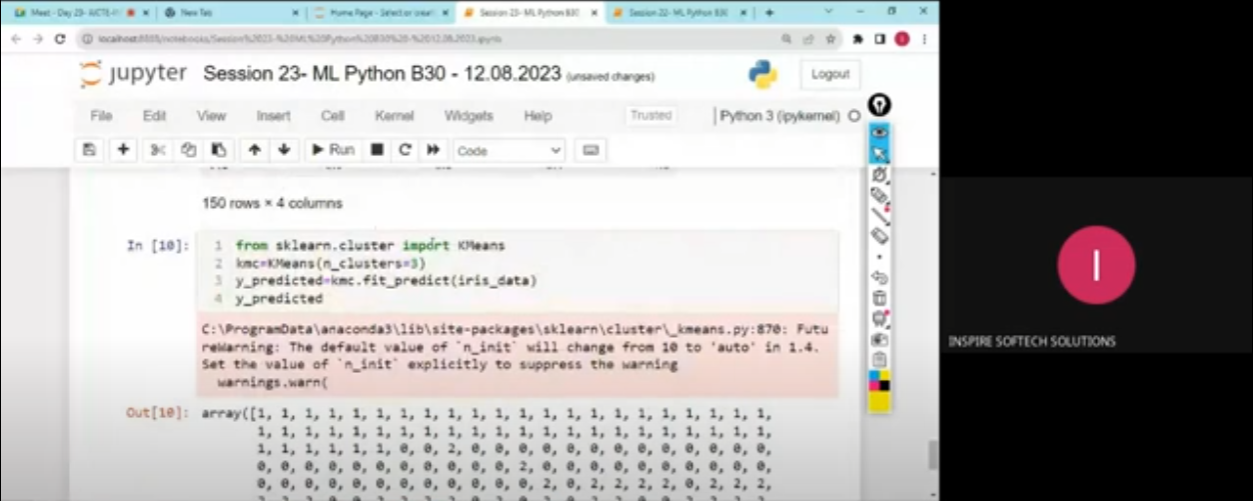
**Work Schedule:**

| **S.No** | **Week** | **Tasks/Outputs** |
| --- | --- | --- |
| 1. | Week-1 | Basics for Machine Learning and Python Programming. |
| 2. | Week- 2 | Python Libraries like math and numpy for basic computations. |
| 3. | Week-3 | Data Visualization using Python libraries like matplotlib, seaborn, and plotly.  Introduction to the pandas library. |
| 4. | Week-4 | Data handling and data analysis using pandas for cricket dataset. |
| 5. | Week-5 | Linear Regression for car price prediction. |
| 6. | Week-6 | SMS classification using Logistic Regression.  K-Means clustering on Iris dataset. |

**Meeting Pictures:**

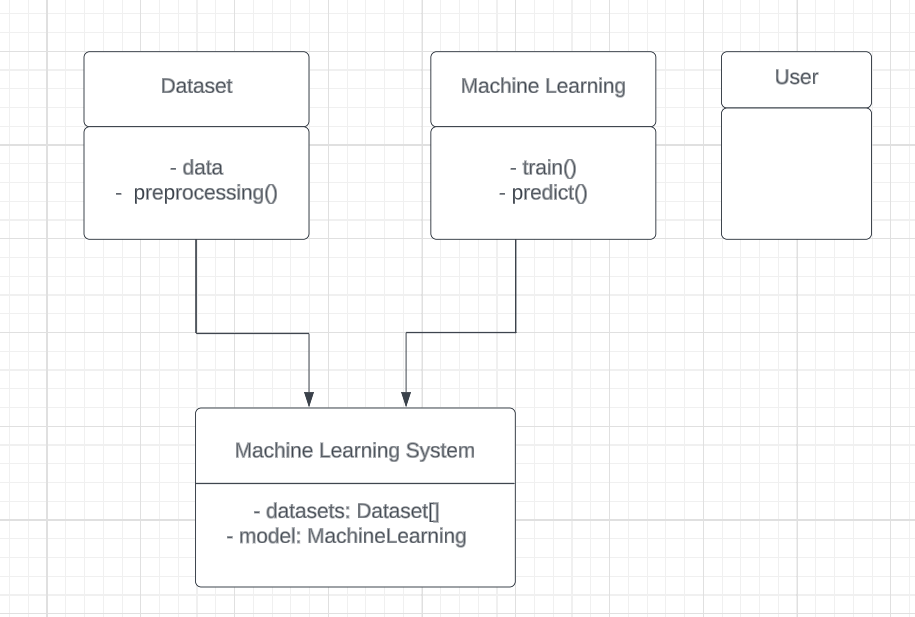
****

This was the internship session on 16th July, 2023 on the data visualization.

****

Thus was the internship session on 12th August, 2023 on K-Means clustering algorithm.

**Analysis and Design with UML:**

****

In this diagram:

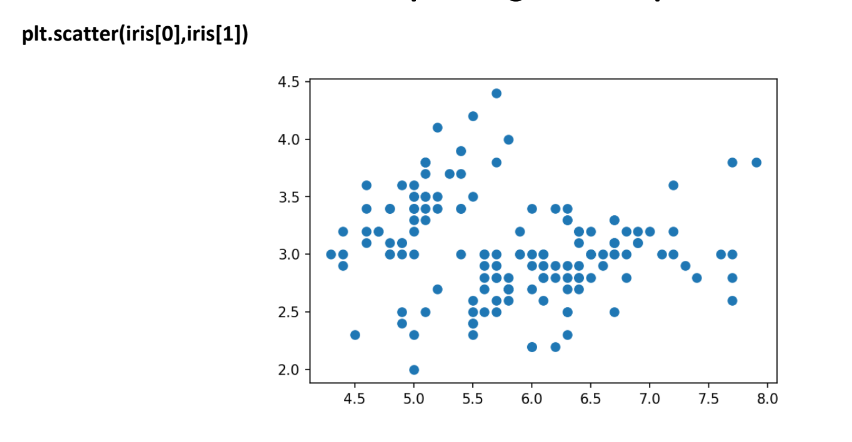
The dataset is a class representing the dataset with attributes like data and methods like preprocessing().

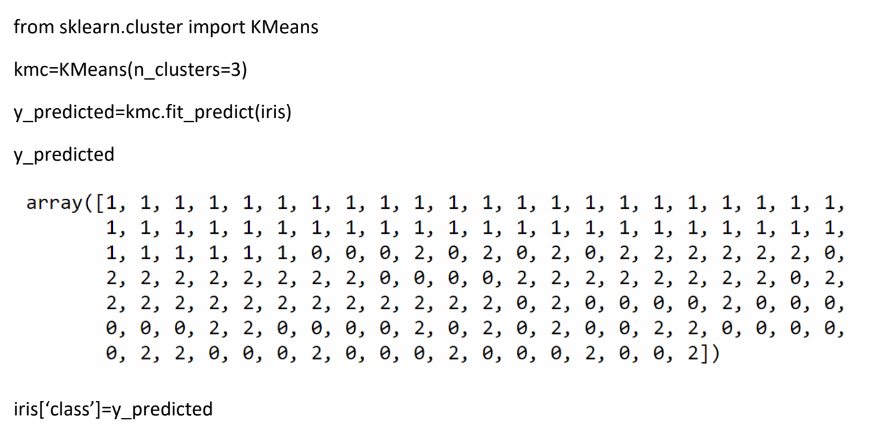
Machine Learning is a class representing the machine learning model with methods train() and predict().

User is a class representing the user interacting with the system.

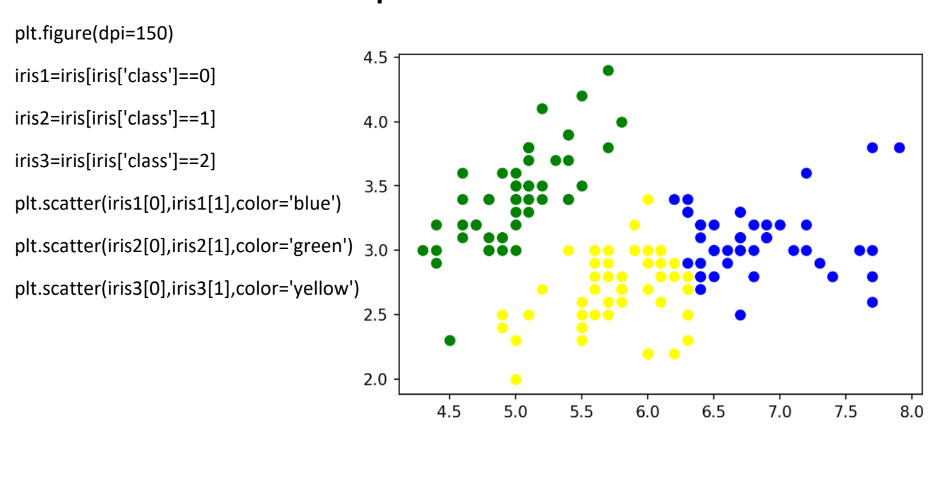
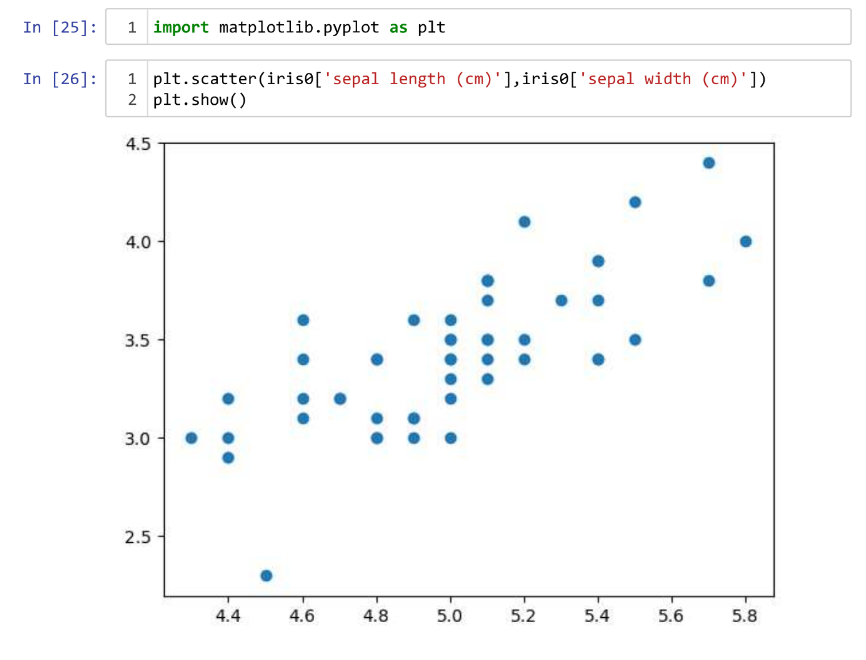
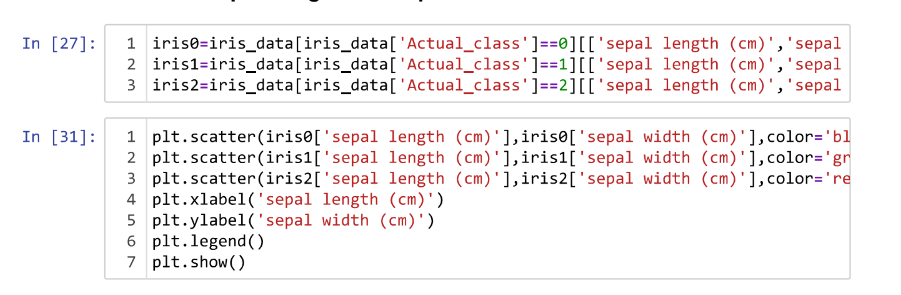
Machine Learning System is a class that aggregates the components, having attributes like datasets (an array of datasets) and models (an instance of the machine learning model).

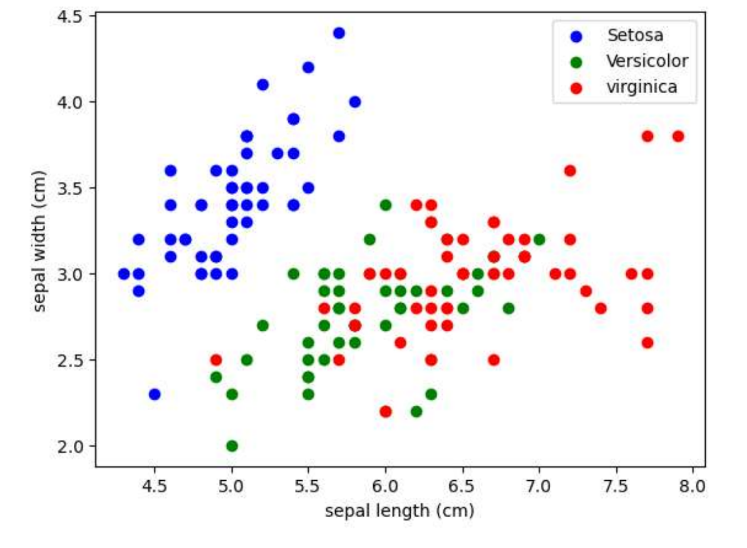
**Implementation**

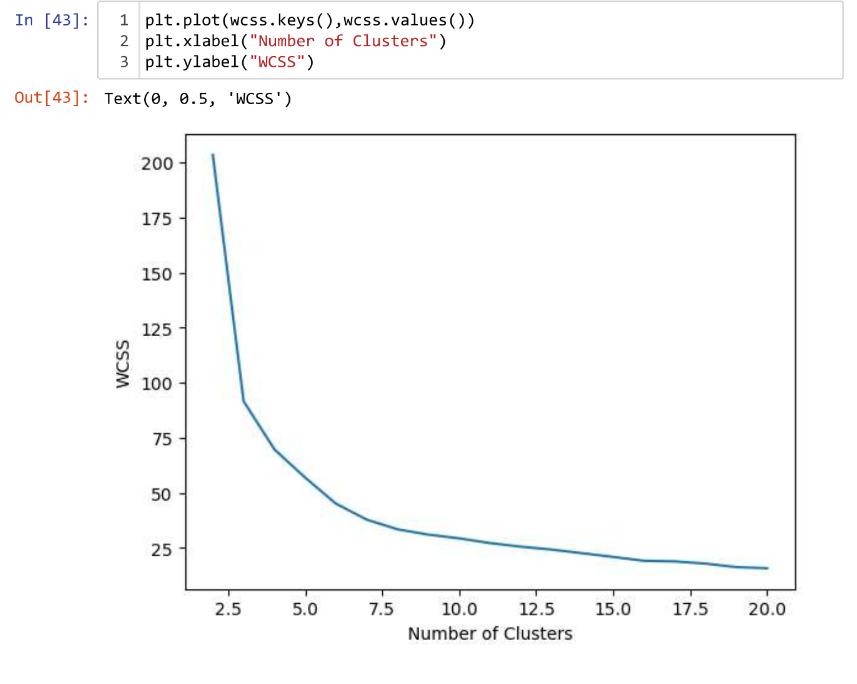




**K-Means Clustering:** K-means clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning.

**Scatter plot:** Scatter plot is a visual representation of how two variables relate to watch other.





**Learning Outcomes**

The learning outcomes of a machine learning with Python internship typically encompass a range of knowledge, skills, and practical experiences that participants are expected to acquire throughout the program.

We gain proficiency in foundational concepts such as data handling, analysis, and visualization using libraries like Pandas. We have developed a deep understanding of linear regression, logistic regression, and their applications, honing our skills in predictive modeling. Through hands-on experience in K-Means Clustering, we’ve learned the nuances of unsupervised learning, including data partitioning and cluster optimization. Exposure to hyperparameter tuning enhances our ability to fine-tune machine learning models for optimal performance. Additionally, we’ve acquired practical skills in the Python programming language, a fundamental tool in the field. This internship fosters a comprehensive grasp of data science processes, from data cleaning to model evaluation. By the end, we were equipped with a holistic skill set that prepared us for real-world applications in data-driven decision-making and predictive analytics.

**Executive Summary**

This internship provides a comprehensive learning experience in the field of data science. Participants delve into foundational concepts such as data handling, analysis, and visualization using Pandas, gaining proficiency in Python programming. The curriculum covers key machine learning techniques, including linear regression, logistic regression, and K-Means Clustering, offering practical insights into predictive modeling and unsupervised learning. This internship emphasizes hands-on projects, enabling students to apply theoretical knowledge to real-world scenarios. By the end of this internship, we’ve emerged with a well-rounded skill set, ready to tackle challenges in data science and contribute meaningfully to projects requiring machine learning expertise.

**References**

Dataset by : [www.kaggle.com](http://www.kaggle.com)

Reference book : Python Machine Learning by Sebastian Raschka

**Appendices**

