PYTHON FOR MACHINE LEARNING INTERNSHIP

Submitted to
RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA
BHOPAL (M.P)

INTERNSHIP REPORT

Submitted by
Chetan Choukade [0176CS233D07]

Submitted to:

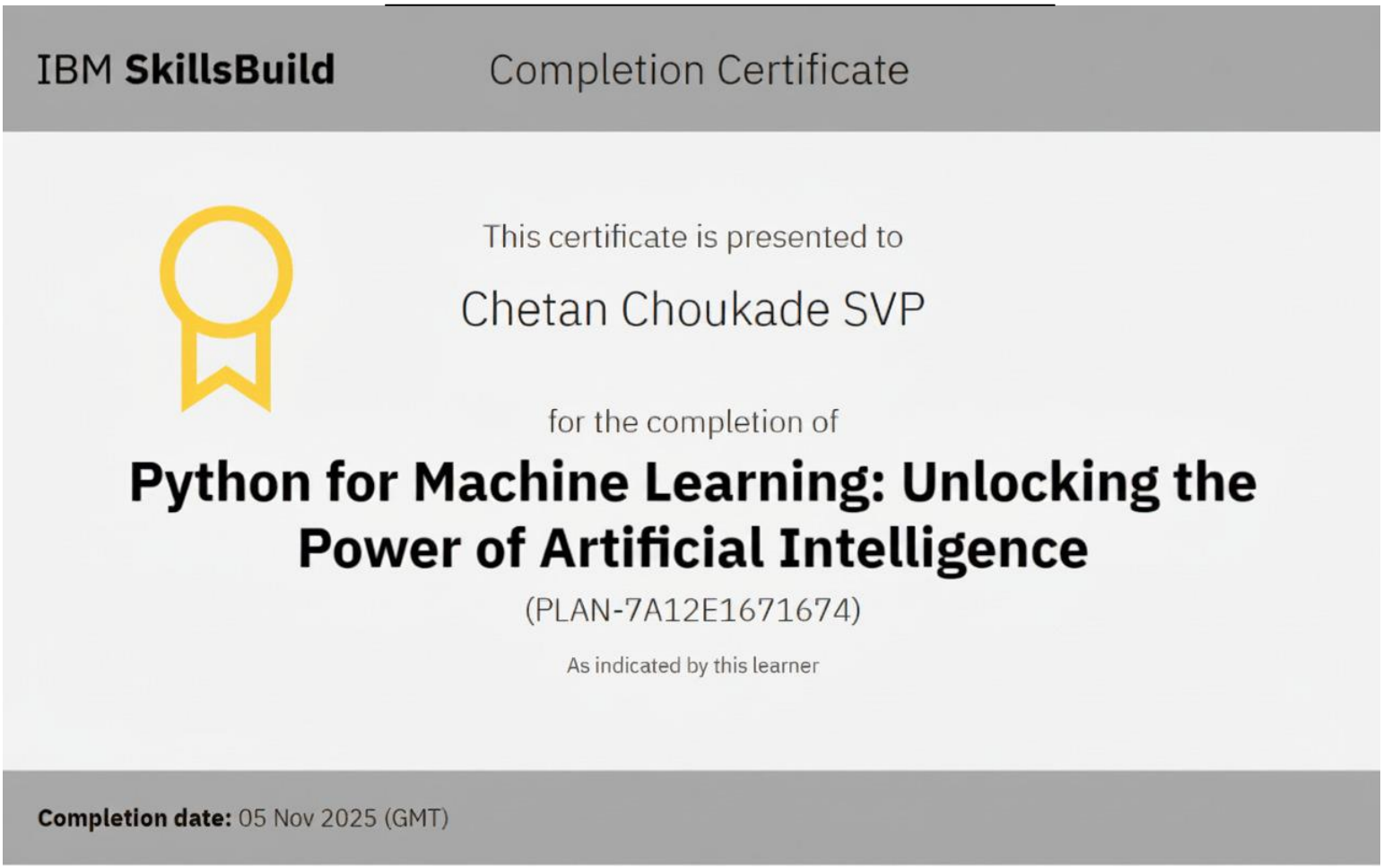
Prof. Alesh Sharma

Department of Computer Science & Engineering (CSE)

Lakshmi Narain College of Technology, Bhopal (M.P.)

Session
2025-26

CERTIFICATE



LAKSHMI NARAIN COLLEGE OF TECHNOLOGY, BHOPAL

DEPARTMENT OF CSE

ACKNOWLEDGEMET

I, **Chetan Choukade (0176CS233D07)**, would like to express my sincere gratitude to the **Head of the Department, Department of Computer Science & Engineering, Lakshmi Narain College of Technology Excellence, Bhopal**, for providing me the opportunity to undertake the **Internship Program from Sep 2025 – Nov 2025**. This internship has allowed me to enhance my technical skills and gain valuable exposure to industry-aligned learning.

I extend my heartfelt thanks to **IBM SkillsBuild** for offering the learning pathway **"Python for Machine Learning: Unlocking the Power of Artificial Intelligence"**, which served as a comprehensive and practical platform to develop essential skills in Python programming, data science, and machine learning.

Finally, I would like to express my warm appreciation to my faculty members, friends, and everyone who supported me during the course of this internship.

(Signature)
Chetan Choukade (0176CS233D07)

INDEX

CHAPTER 1

INTRODUCTION

The “Python for Machine Learning: Unlocking the Power of Artificial Intelligence” program offered by IBM SkillsBuild is a highly specialized learning pathway designed to provide students with a strong foundation in Python programming, data analysis, and machine learning techniques. IBM SkillsBuild is a globally recognized platform that focuses on building essential technical and professional capabilities required in modern industries. Through this comprehensive training program, I gained a deep understanding of how machine learning models are created, trained, evaluated, and deployed to solve real-world problems.

The internship-style learning experience began with mastering the fundamentals of Python, including data types, functions, modules, file handling, and object-oriented programming. Python was chosen as the central language because of its simplicity and the extensive ecosystem of libraries that support machine learning and artificial intelligence. I worked extensively with libraries such as NumPy for numerical computing, Pandas for data manipulation, Matplotlib and Seaborn for data visualization, and Scikit-learn for implementing machine learning algorithms.

The course introduced the foundational concepts of machine learning, beginning with supervised and unsupervised learning techniques. I learned how datasets are collected, cleaned, transformed, and normalized before being used to train models. Concepts such as feature engineering, outlier detection, missing value imputation, and data preprocessing were covered in detail, giving me practical insight into preparing real-world data for analysis.

One of the major highlights of the program was the hands-on experience with various machine learning algorithms. I practiced implementing models such as Linear Regression, Logistic Regression, K-Nearest Neighbors, Decision Trees, Random Forests, Support Vector Machines, and K-Means Clustering. Through these exercises, I learned how each algorithm works internally, when to use them, and how to fine-tune their parameters to improve performance. I also learned about evaluation metrics like accuracy, precision, recall, F1-score, mean squared error (MSE), confusion matrix, and ROC curves—metrics that are essential for analyzing model effectiveness.

The program also emphasized the importance of visualization and explained how graphical analysis can uncover patterns, relationships, and insights within large datasets. Creating plots, charts, and heatmaps helped me interpret data more efficiently and make decisions during the model-building process. These visual skills strengthened my ability to communicate findings clearly and professionally.

In addition to technical implementation, the course provided an understanding of real-world machine learning workflows, including how data pipelines are created, how models are deployed, and how businesses integrate AI solutions to enhance decision-making. I also learned about ethical considerations such as data privacy, fairness in AI, algorithmic bias, and responsible development practices.

A key learning experience involved building and training a complete machine learning model from scratch. This included loading datasets, cleaning and preprocessing them, training multiple algorithms, comparing their performance, tuning hyperparameters, and selecting the best model. This practical approach helped me translate theoretical concepts into working solutions that are commonly used in industries such as healthcare, finance, e-commerce, cybersecurity, and automation.

Throughout the learning journey, IBM's SkillBuild platform provided high-quality lectures, project exercises, assessments, and real-time examples that strengthened my conceptual clarity. By completing the modules and assessments, I developed the confidence to build machine learning applications independently and apply them to real-world problems.

Overall, this program enhanced my technical problem-solving abilities and exposed me to the rapidly growing field of Artificial Intelligence. It provided me with strong fundamentals in Python programming and machine learning methodologies, preparing me for advanced roles in data science, AI engineering, and analytical research. This training serves as a significant milestone in my academic and professional growth, equipping me with the skills required to participate in AI-driven innovation and contribute effectively to modern technological advancements.

CHAPTER 2

ABOUT THE ORGANISATION

IBM SkillsBuild is a global learning and employability platform developed by IBM with the mission to empower students, educators, and early-career professionals with future-ready technical and professional skills. It is a free and industry-recognized initiative that focuses on democratizing access to world-class training in fields such as Artificial Intelligence, Machine Learning, Cloud Computing, Cybersecurity, Data Science, and Software Development. By providing hands-on, practical, and self-paced learning experiences, IBM SkillsBuild aims to bridge the gap between academic knowledge and real-world industry requirements.

IBM (International Business Machines Corporation), the parent organization, is one of the world's oldest and most respected technology companies. Known for its contributions to computing, innovation, and enterprise solutions, IBM has played a pioneering role in the development of artificial intelligence through its iconic projects such as IBM Watson. SkillsBuild represents IBM's commitment to nurturing the next generation of tech professionals by providing them with cutting-edge educational resources, expert-designed courses, and globally recognized certifications.

Another core aspect of IBM SkillsBuild is its emphasis on practical experience. Unlike theoretical classroom-based teaching, SkillsBuild encourages learners to engage in hands-on projects, coding exercises, interactive labs, and case studies. This not only enhances technical understanding but also builds confidence in applying machine learning techniques to real datasets. Through such practice-driven learning, participants develop the ability to tackle real-world business problems using AI-driven solutions.

The platform also provides digital credentials in the form of IBM-recognized badges and certificates, which hold significant value in the job market. These credentials validate the learner's expertise and demonstrate their readiness for roles in rapidly growing fields such as data science, machine learning engineering, and analytics. The industry relevance of these certifications makes IBM SkillsBuild a trusted and respected platform among employers, academic institutions, and career development programs worldwide.

In addition to technical courses, IBM SkillsBuild actively promotes an inclusive learning environment. It offers resources for learners from diverse backgrounds, including those without prior technical experience. The curriculum is designed to be beginner-friendly while still offering advanced pathways for those aiming to specialize further. This inclusivity reflects IBM's vision of building a skilled global workforce capable of contributing to digital transformation across sectors.

Overall, IBM SkillsBuild stands as a powerful initiative in professional education, offering high-quality technical training tailored to modern industry needs. Its commitment to accessible learning, hands-on skill development, and employability makes it a valuable platform for students and professionals aspiring to build careers in Artificial Intelligence, Machine Learning, and Data Science. By completing this program, I gained essential skills, practical exposure, and a strong foundation that will support my future technological and professional growth.

CHAPTER 3

ABOUT THE PROJECTS

The “Python for Machine Learning: Unlocking the Power of Artificial Intelligence” course offered through IBM SkillsBuild is a comprehensive training program designed to provide learners with a solid foundation in the fundamentals of machine learning using Python. The course integrates theoretical understanding with practical, hands-on exercises to help learners develop the technical skills required to build, train, and evaluate machine learning models. It follows a structured learning path that enables students to progress from basic Python programming concepts to advanced machine learning algorithms and real-world applications.

The course begins with an introduction to Python programming, where learners become familiar with essential concepts such as variables, data types, loops, functions, modules, and file handling. This foundational understanding is crucial because Python serves as the primary language for most machine learning tasks. The course emphasizes clean coding practices, logical thinking, and efficient problem-solving techniques, which lay the groundwork for handling complex data structures and algorithms later in the program.

Following the basics, the course transitions into data handling and data manipulation techniques using popular Python libraries such as NumPy and Pandas. These modules teach learners how to efficiently work with arrays, apply mathematical operations, manipulate datasets, and prepare data for machine learning tasks. The practical assignments in this section involve loading datasets, cleaning missing values, transforming data formats, normalizing columns, and performing exploratory data analysis (EDA). This hands-on experience helps learners understand how raw data is converted into structured, meaningful information suitable for model building.

One of the core sections of the course focuses on data visualization. Using libraries such as Matplotlib and Seaborn, learners create a variety of plots, histograms, scatter charts, heatmaps, and correlation graphs to gain visual insights into the data. Visualization plays a crucial role in identifying trends, patterns, and relationships within datasets, and the course places strong emphasis on helping learners interpret these visual outputs effectively.

The main highlight of the program is its detailed exploration of machine learning algorithms. The course introduces both supervised and unsupervised learning techniques using Scikit-learn, one of the most widely used machine learning libraries. Learners gain hands-on experience with algorithms such as:

- Linear Regression
- Logistic Regression
- Decision Trees
- Random Forest

Each algorithm is taught through clear explanations, coding demonstrations, and practical exercises. Learners understand when and how to use each model, how to tune hyperparameters, how to handle overfitting, and how to evaluate model performance using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and mean squared error.

Another important component of the course is model evaluation and validation. Learners explore techniques such as train-test splitting, cross-validation, performance comparison, and error analysis. This ensures a thorough understanding of how to measure the effectiveness of a machine learning model and how to choose the best-performing algorithm for a given dataset.

Throughout the course, emphasis is placed on practical, real-world problem-solving. Learners complete hands-on exercises where they build end-to-end machine learning pipelines—from loading data and preprocessing it to training models and analyzing results. These exercises help students understand the full workflow of machine learning development and prepare them for industry-level tasks.

The course also introduces students to ethical AI considerations, responsible data science practices, and the importance of fairness and transparency in machine learning models. This broader perspective helps learners understand not only the technical side of machine learning but also its societal impact.

By the end of the course, learners are equipped with the ability to independently develop, test, and deploy basic machine learning models. The combination of Python programming, data analysis, visualization, and algorithm implementation provides a strong foundation for advanced studies in artificial intelligence, data science, and deep learning.

Overall, the “Python for Machine Learning” course offered by IBM SkillsBuild provides a comprehensive and industry-aligned learning experience. It builds confidence, enhances technical capabilities, and prepares learners to apply machine learning techniques to real-world challenges across various domains such as healthcare, finance, business analytics, and automation

CHAPTER 4

Hardware/Software Environment

Hardware Environment

The hardware requirements for this course were straightforward, as most of the tasks involved Python programming, dataset manipulation, and executing machine learning algorithms. Although certain ML operations can be computationally intensive, the course content was designed to run smoothly on mid-range systems.
The ideal hardware configuration for completing the course efficiently includes:

- **Device:** Personal Laptop or Desktop Computer
- **Processor:** Intel Core i3 or higher
  (Recommended: Intel Core i5/i7 or AMD Ryzen equivalent)
- **RAM:** Minimum 4 GB
  (Recommended: 8 GB or more for handling larger datasets)
- **Storage:** 10–20 GB of free disk space for Python libraries, datasets, and notebooks
- **Internet Connection:** Stable broadband connection (2–5 Mbps or higher) for accessing IBM SkillsBuild, downloading packages, and streaming modules
- **Display:** HD Display (1366 × 768 resolution or higher) for better coding and visualization

This configuration ensured that Python IDEs, Jupyter Notebooks, NumPy, Pandas, and machine learning models operated smoothly without lag or performance issues.

Software Environment

The software environment formed the backbone of the learning experience by enabling coding, model training, data analysis, and visualization. The tools and platforms used throughout the course include the following:

---

1. Operating System

The course was compatible with multiple operating systems, making it flexible and accessible for all learners:

- Windows 10 / 11
- macOS (latest versions)
- Linux distributions (Ubuntu, Fedora, etc.)

Python and its data science libraries work seamlessly across all major platforms.

---

2. Python Programming Environment

Python served as the primary language for all machine learning and data science tasks. The following tools were used:

- **Python 3.x** (latest recommended version)
- **Jupyter Notebook / JupyterLab**
    - Ideal for writing, testing, and visualizing code
    - Suitable for step-by-step model development

Optional IDEs:

- Visual Studio Code (VS Code)
- PyCharm Community Edition
- Google Colab (cloud-based notebook environment)

These tools provided a user-friendly interface for coding, experimentation, and generating visual insights.

---

3. Python Libraries Used

Machine learning heavily depends on specialized libraries. The following libraries were used extensively during the course:

- **NumPy:** Numerical computation and array operations
- **Pandas:** Data manipulation, cleaning, preprocessing
- **Matplotlib:** Basic visualizations
- **Seaborn:** Advanced statistical and graphical visualizations
- **Scikit-learn:**
    - Machine learning algorithms
    - Model training and testing
    - Evaluation metrics and preprocessing tools

Additional concepts introduced:

- Train-test splitting
- Hyperparameter tuning
- Evaluation metrics
- Data scaling and normalization

These libraries together provided a complete ecosystem for developing and executing machine learning pipelines.

CHAPTER 5

SNAPSHOTS OF INPUT & OUTPUT

Practical 1: Loading a Dataset & Basic Analysis

```
import pandas as pd

df = pd.read_csv("iris.csv")

print(df.head())
```

```
   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      Species
0   1            5.1           3.5            1.4           0.2  Iris-setosa
1   2            4.9           3.0            1.4           0.2  Iris-setosa
2   3            4.7           3.2            1.3           0.2  Iris-setosa
3   4            4.6           3.1            1.5           0.2  Iris-setosa
4   5            5.0           3.6            1.4           0.2  Iris-setosa
```

Practical 2: Data Visualization

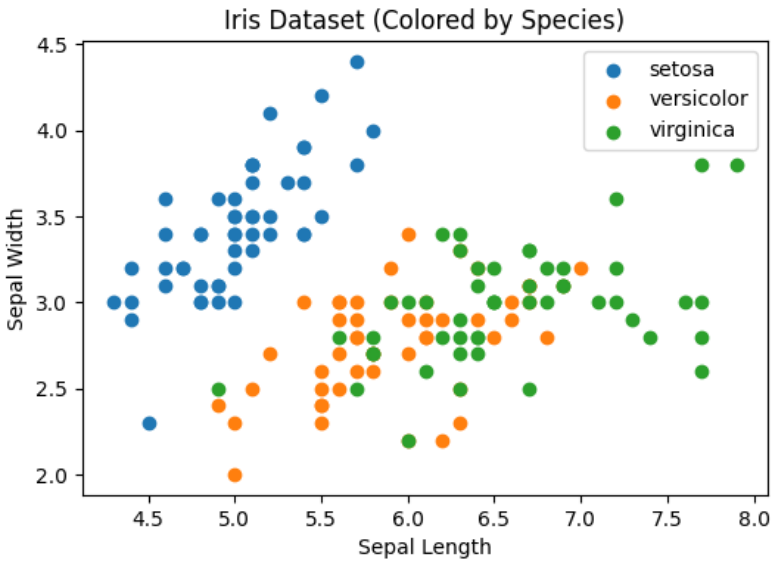
```
from sklearn.datasets import load_iris
import matplotlib.pyplot as plt

iris = load_iris()
X = iris.data
y = iris.target
species = iris.target_names

plt.figure(figsize=(6, 4))

for label in range(3):
    plt.scatter(
        X[y == label, 0],
        X[y == label, 1],
        label=species[label]
    )

plt.xlabel("Sepal Length")
plt.ylabel("Sepal Width")
plt.title("Iris Dataset (Colored by Species)")
plt.legend()
plt.show()
```



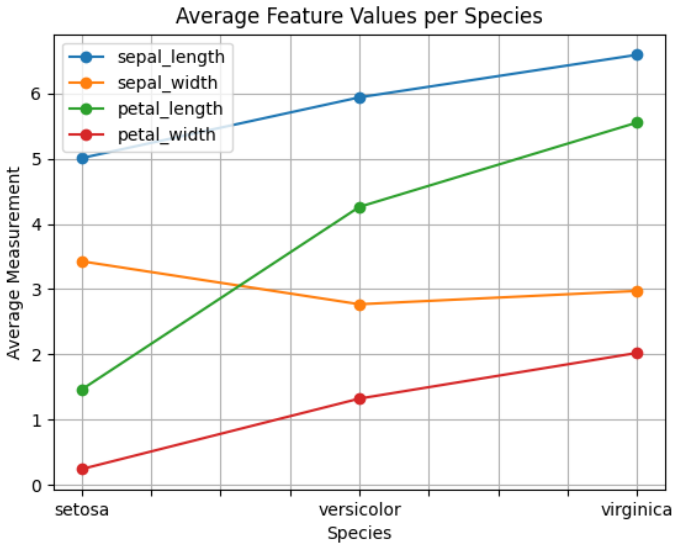
Practical 3: Line Plot of Average Measurements

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

df = sns.load_dataset("iris")

avg_values = df.groupby("species").mean()

plt.figure(figsize=(8,5))
avg_values.plot(marker="o")
plt.title("Average Feature Values per Species")
plt.xlabel("Species")
plt.ylabel("Average Measurement")
plt.legend(loc="best")
plt.grid(True)
plt.show()
```



CHAPTER 6
FUTURE SCOPE

The “Python for Machine Learning: Unlocking the Power of Artificial Intelligence” program has provided a strong foundation in both Python programming and machine learning concepts. As artificial intelligence continues to grow rapidly across all industries, the skills acquired during this learning plan open several promising opportunities for advanced study, research, and professional development.

The increasing adoption of AI technologies in sectors such as healthcare, finance, retail, cybersecurity, automation, and business analytics ensures a high demand for professionals who can develop intelligent data-driven solutions. With the knowledge gained through this course, I am now better equipped to understand modern data workflows, analyze datasets, and build machine learning models that can support real-world decision-making.

A major future scope lies in advancing towards more specialized domains within AI. Fields such as Deep Learning, Neural Networks, Natural Language Processing (NLP), and Computer Vision build upon machine learning fundamentals and offer opportunities to work on complex projects involving image recognition, speech analysis, chatbots, recommendation systems, and autonomous technologies. Gaining expertise in frameworks like TensorFlow, Keras, and PyTorch will further strengthen my ability to build and deploy advanced AI solutions.

Another important direction is to explore Big Data Technologies such as Apache Spark, Hadoop, and cloud-based analytics platforms. As datasets continue to grow in size and complexity, professionals with the ability to manage and process big data efficiently are highly valued. Combining machine learning knowledge with big data skills creates a powerful combination suited for modern data engineering and analytics roles.

There is also significant future scope in pursuing professional certifications to validate and deepen my expertise. Certifications such as IBM Data Science Professional Certificate, IBM Machine Learning Certification, Google TensorFlow Developer, Microsoft AI Engineer, and AWS Machine Learning Specialty can enhance career opportunities and strengthen credibility in the industry.

Beyond technical knowledge, the learning plan has also improved my problem-solving abilities, analytical thinking, and approach to structured experimentation. These skills prepare me to work effectively in data-driven environments and collaborate on interdisciplinary AI projects.

In conclusion, the future scope of this internship experience is vast and promising. The foundational skills gained through Python programming, data analysis, and machine learning model development provide a gateway to multiple advanced career paths in artificial intelligence. With continuous learning, hands-on practice, and professional certifications, I am confident that I can build a successful and impactful career in the field of AI and contribute meaningfully to the innovative technological advancements of the future

CHAPTER 7
REFERENCES

---

1. **IBM SkillsBuild – Official Learning Platform**
   https://skillsbuild.org

2. **Python Documentation – Official Python Language Reference**
   https://docs.python.org

3. **NumPy Documentation – Numerical Computing Library**
   https://numpy.org/doc

4. **Pandas Documentation – Data Analysis Library**
   https://pandas.pydata.org/docs

5. **Matplotlib Documentation – Data Visualization Library**
   https://matplotlib.org/stable/contents.html

6. **Seaborn Documentation – Statistical Visualization Library**
   https://seaborn.pydata.org

7. **Scikit-Learn Documentation – Machine Learning Framework**
   https://scikit-learn.org/stable/documentation.html

8. **IBM Developer – AI & Machine Learning Resources**
   https://developer.ibm.com/technologies/artificial-intelligence

9. **Kaggle – Datasets and Machine Learning Resources**
   https://www.kaggle.com