Internship Report
On
Salary Prediction Model

Submitted by
[Student Name 1: RAJAMMAGARI SHAIK MAHAMMAD SAHIL]
[University Roll No: 218R1A0543]
[College Name: SAI RAJESWARI INSTITUTE OF TECHNOLOGY]

[Student Name 2: SYED MAHAMMAD RAFI SADDAM]
[University Roll No: 218R1A0556]
[College Name: SAI RAJESWARI INSTITUTE OF TECHNOLOGY]

Submitted to
Ms. Mallika Srivastava
Head, Training Delivery
EISystems Services

&

Mayur Dev Sewak
Head, Internships & Trainings
EISystems Services

Student's Declaration

I, **Rajammagari Shaik Mahammad Sahil,** a student **B. Tech** program,
Roll No. **218R1A0543** of the Department of **CSE**, **Sai Rajeswari Institute of Technology**
College do hereby declare that I have completed the mandatory internship in **EiSystems**
Technologies under the faculty guideship of **Mr. Sabari Rajan,** Department **CSE**, **Sai**
Rajeswari Institute of Technology.

R.S. Mahammad Sahil

Endorsements

SIGNATURE
[Faculty Guide Name: Mr. Sabari Rajan]
[Faculty Guide Department: Computer Science and Engineering]
[College Name: Sai Rajeswari Institute of Technology]

SIGNATURE
[Head of Department Name: Dr. Y. Subba Reddy]
[Department Name: Computer Science and Engineering]
[College Name: Sai Rajeswari Institute of Technology]

Table of Content

List of Figures

Executive Summary

During my two-month internship at **EiSystems**, I had the unique opportunity to immerse myself in the field of data science with Python. This internship was part of the Data Science with Python program, and it provided me with a robust foundation in both theoretical and practical aspects of data science. The primary focus of my internship was to develop a linear regression model to predict employee salaries based on their years of experience, which was a part of a broader curriculum designed to enhance my skills in Python programming and data science.

The internship began with foundational topics in Python, including variables, constants, naming conventions, and basic data types such as numbers and strings. As the program progressed, I learned about more complex data structures like lists, tuples, dictionaries, and sets. These foundational skills were crucial as they formed the basis for more advanced topics.

Key topics covered during the internship included:

- **Control Statements and Loops:** Understanding how to control the flow of a program using if-else statements, loops, and other control structures. This knowledge was essential for writing efficient and effective code.

- **File Handling:** Learning how to read from and write to files, which is essential for data processing tasks. This included understanding different file formats and how to manipulate them using Python.

- **Functions, Packages, and Modules:** Developing reusable code through functions and organizing code using packages and modules. This helped in writing modular and maintainable code.

- **Exception Handling:** Implementing error handling to make programs more robust and reliable. This involved learning how to anticipate and manage errors gracefully.

- **Object-Oriented Programming (OOP):** Understanding the principles of OOP, including classes, objects, inheritance, and polymorphism. This was crucial for designing complex systems and applications.

Page 4

In addition to these foundational topics, the internship also covered essential data science libraries such as NumPy, Pandas, and Matplotlib. These libraries are critical for data manipulation, analysis, and visualization. I learned how to use NumPy for numerical operations, Pandas for data manipulation, and Matplotlib for creating visualizations. These skills were instrumental in handling and analyzing large datasets efficiently.

One of the key projects during the internship was a Quiz Game, which helped reinforce my understanding of Python basics and control structures. This project involved creating a simple game where users could answer questions and receive scores based on their responses. It was a fun and engaging way to apply the concepts learned and to practice coding in a real-world scenario.

The culmination of the internship was the development and deployment of a linear regression model to predict employee salaries based on their years of experience. This project involved several steps, including data collection and preprocessing, model development, and deployment using Streamlit and Pickle libraries. The project provided hands-on experience in building and deploying machine learning models, which is a critical skill in data science.

Key achievements of my internship include:

- Successfully developing and deploying a linear regression model with a high degree of accuracy. The model's performance was evaluated using metrics such as Mean Absolute Error (MAE) and R-squared, which indicated strong predictive capabilities.

- Creating an interactive web application that allows users to predict salaries based on experience. This application was designed to be user-friendly, with a clean interface and straightforward functionality.

- Enhancing my understanding of machine learning algorithms and their practical applications. This included not only the development of the model but also the fine-tuning of its parameters to improve accuracy.

- Gaining experience in using Python libraries for data analysis and model deployment. This involved extensive use of Pandas for data manipulation, NumPy for numerical operations, and Scikit-learn for model development and evaluation.

This internship has significantly contributed to my professional growth by providing practical experience in data science and machine learning. It has also reinforced my interest in pursuing a career in this field. The skills and

knowledge I have acquired will be invaluable as I continue to develop my expertise in data science. Furthermore, the experience of working on real-world projects has given me a deeper understanding of the challenges and rewards of this field.

In addition to the technical skills, this internship has also helped me develop important soft skills such as problem-solving, critical thinking, and effective communication. Working on projects that required both technical proficiency and user-centric design has taught me the importance of balancing these aspects to create successful products. I have also learned the value of collaboration and feedback, as working with my mentors and peers provided me with new perspectives and insights that enhanced the quality of my work.

Moreover, the internship provided opportunities to engage in continuous learning and improvement. Throughout the program, I encountered various challenges that required innovative solutions. This iterative process of problem-solving and learning from mistakes has been instrumental in building my resilience and adaptability. I also had the chance to participate in discussions and workshops that broadened my understanding of current trends and best practices in data science.

Conclusion :

Overall, this internship has been a transformative experience that has equipped me with the skills and confidence to pursue a career in data science. I am grateful for the opportunity to apply my knowledge in a practical setting and to contribute to meaningful projects. The lessons learned and the skills acquired during this internship will undoubtedly serve as a strong foundation for my future endeavors in the field of data science. The comprehensive training and hands-on experience have not only enhanced my technical abilities but also prepared me to tackle complex challenges in my future career.

Overview of Organization

India's leader in workshops & trainings at IITs, NITs & top engineering colleges:

EISystems Services is a leading Indian technology identity with operations across India. EISystems (We call it EISys) offers trainings in Cybersecurity, Machine Learning, Automobiles, Internet of Things, Robotics and Socialmedia for enterprises and student community. Till date we have trained approximately 50000 students and impacted around 2 lakhs students through our various outreach initiatives since our founding.

Project Summary

Project Title: Predicting Employee Salaries Using Linear Regression

Objective:

The primary objective of this project was to develop a linear regression model to predict employee salaries based on their years of experience. The goal was to create an accurate and reliable predictive model that could be deployed as a web application for easy user interaction. This project aimed to demonstrate the practical application of machine learning techniques in solving real-world problems and to provide a hands-on experience in model development and deployment.

Key Points:

- Check weather all the required documents are present it the directory.
- If all the documents are present open the directory with the help of command prompt.
- When you see the directory location in the command prompt terminal type the command "streamlit run filename.py".
- The project file name is Project.py so type "streamlit run Project.py".
- A new Browser page will pop up while asking to enter the year of experience of the employee.
- After mentioning the year, the then we should have to click on start button.
- It will predict the salary of the employee based on the input given to it

Methodology:

1. **Data Collection and Preprocessing:**
   - **Data Collection**: The dataset used for this project was sourced from a reliable database and included various attributes such as years of experience, salary, and other relevant factors. The dataset was chosen for its relevance and completeness, ensuring that it provided a solid foundation for model development.
   - **Data Cleaning**: The initial step involved cleaning the dataset to handle missing values and outliers. Missing values were imputed using appropriate statistical methods, while outliers were identified and treated to prevent them from skewing the model's results.
   - **Data Transformation**: The data was then transformed to ensure it was in a suitable format for analysis. This included normalizing numerical features and encoding categorical variables if any were present.
   - **Data Splitting**: The cleaned and transformed data was split into training and testing sets. The training set was used to develop the model, while the testing set was used to evaluate its performance. A typical split ratio of 80:20 was used to ensure that the model had sufficient data for training while retaining enough data for a robust evaluation.

2. **Model Development:**
   o **Feature Selection**: The first step in model development was selecting the appropriate features for the linear regression model. In this case, the primary feature was the years of experience, which was used to predict the salary.
   o **Model Training**: The linear regression model was developed using the Scikit-learn library. The training process involved fitting the model to the training data and adjusting the model parameters to minimize the error between the predicted and actual salaries.
   o **Model Evaluation**: The model's performance was evaluated using the testing dataset. Key metrics such as Mean Absolute Error (MAE) and R-squared were calculated to assess the model's accuracy and predictive capabilities. The MAE provided a measure of the average error in the predictions, while the R-squared value indicated the proportion of variance in the salary that could be explained by the years of experience.
   o **Hyperparameter Tuning**: To improve the model's performance, hyperparameter tuning was performed. This involved experimenting with different model parameters and selecting the combination that resulted in the best performance. Techniques such as cross-validation were used to ensure that the model was not overfitting to the training data.

3. **Model Deployment:**
   o **Model Saving**: Once the model was trained and evaluated, it was saved using the Pickle library. Pickle is a Python library that allows for the serialization and deserialization of Python objects, making it easy to save the trained model and load it later for predictions. This ensured that the model could be efficiently stored and reused without retraining.
   o **Web Application Development**: A web application was developed using Streamlit, a powerful and easy-to-use framework for creating interactive web applications in Python. Streamlit allowed for the rapid development of a user-friendly interface where users could input their years of experience and receive a predicted salary. The application was designed to be intuitive, with a clean layout and responsive design.
   o **Integration**: The saved model was integrated into the web application for real-time predictions. This involved loading the model using Pickle and using it to generate predictions based on user input. The application also included features such as input validation and error handling to ensure a smooth user experience. Streamlit's capabilities made it possible to create an interactive and dynamic web application that provided immediate feedback to users.

Results:

- **Model Performance**:

The linear regression model demonstrated a high degree of accuracy in predicting employee salaries based on experience. The model's performance was evaluated using metrics such as Mean Absolute Error (MAE) and R-squared, indicating strong predictive capabilities. The MAE was found to be low, suggesting that the model's predictions were close to the actual salaries. The R-squared value was high, indicating that a significant proportion of the variance in salaries could be explained by the years of experience.

- **Web Application**:

The web application provided an intuitive and user-friendly interface for users to interact with the model. Users could input their years of experience and receive an estimated salary in real-time. The application was designed to be responsive and accessible, ensuring that users could easily access it from different devices. Streamlit's interactive widgets and real-time updates made the application engaging and easy to use.

- **Validation**:

 The model's performance was further validated through cross-validation techniques, ensuring its robustness and reliability. This involved splitting the data into multiple folds and training the model on different subsets to assess its generalizability. The cross-validation results confirmed that the model was not overfitting and could generalize well to new data.

Conclusion:

The project successfully achieved its objective of developing a predictive model for employee salaries. The linear regression model, combined with the interactive web application, provides a practical tool for estimating salaries based on experience. This project has enhanced my understanding of machine learning algorithms, data preprocessing techniques, and web application development. The skills and knowledge gained from this project will be instrumental in my future endeavors in data science.

In addition to the technical achievements, this project has also provided valuable insights into the practical challenges of deploying machine learning models. This includes considerations such as data quality, model interpretability, and user experience design. By addressing these challenges, I have gained a deeper appreciation for the complexities of real-world data science projects.

Furthermore, this project has reinforced the importance of continuous learning and improvement. Throughout the development process, I encountered various obstacles and learned to overcome them through research, experimentation, and collaboration. This iterative approach has not only improved the quality of the final product but also enhanced my problem-solving skills and resilience.

Overall, this project has been a rewarding experience that has significantly contributed to my professional growth. It has provided me with practical experience in developing and deploying machine learning models, as well as valuable insights into the challenges and rewards of data science. I am confident that the skills and knowledge gained from this project will serve as a strong foundation for my future career in data science.

Additionally, the project has highlighted the importance of effective communication and documentation. Throughout the project, I maintained detailed documentation of the steps taken, the challenges faced, and the solutions implemented. This not only helped in keeping track of the progress but also ensured that the project could be easily understood and replicated by others. Effective communication with mentors and peers was also crucial in receiving feedback and improving the project.

The project also emphasized the significance of ethical considerations in data science. Ensuring the privacy and security of the data used, as well as the fairness and transparency of the model, were important aspects that were

carefully addressed. This experience has underscored the responsibility that comes with working in the field of data science and the importance of adhering to ethical standards.

In conclusion, this project has been an invaluable learning experience that has equipped me with the skills and knowledge necessary to excel in the field of data science. The practical experience gained, combined with the theoretical knowledge acquired during the internship, has provided a solid foundation for my future career. I am excited to continue exploring the field of data science and to apply the skills and insights gained from this project to future endeavors.
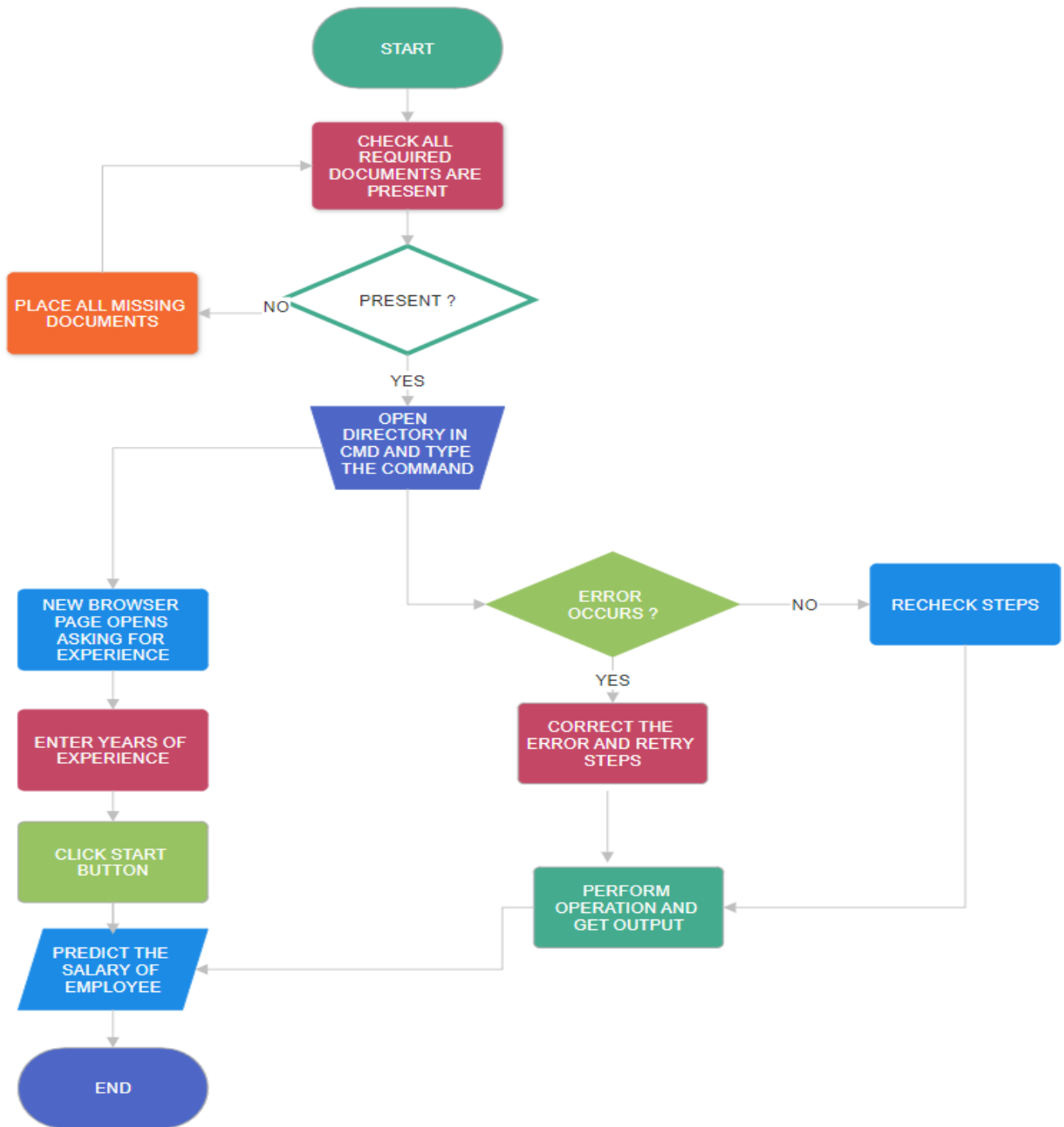
Data Flow Diagram / Process Flow

Figure-1 dataflow diagram

Code / Program with Supported Screenshots

```
[ ]:   #REQUIRED MODULES AND LIBRARIES
```

```
[ ]:   import pandas as pd
       import warnings
       warnings.filterwarnings("ignore")
```

Figure-2 Required Modules

```
[ ]:   #READING DATASET

       df = pd.read_csv("Salary_dataset.csv")
```

Figure-3 Reading Dataset

```
[12]:  df.head()
```

[12]:	Unnamed: 0	YearsExperience	Salary
0	0	1.2	39344.0
1	1	1.4	46206.0
2	2	1.6	37732.0
3	3	2.1	43526.0
4	4	2.3	39892.0

Figure-4 using head ()

[16]:

	Unnamed: 0	YearsExperience	Salary
0	0	1	39344
1	1	1	46206
2	2	1	37732
3	3	2	43526
4	4	2	39892
5	5	3	56643
6	6	3	60151
7	7	3	54446
8	8	3	64446
9	9	3	57190
10	10	4	63219
11	11	4	55795
12	12	4	56958
13	13	4	57082
14	14	4	61112
15	15	5	67939
16	16	5	66030
17	17	5	83089
18	18	6	81364
19	19	6	93941
20	20	6	91739
21	21	7	98274
22	22	8	101303
23	23	8	113813
24	24	8	109432
25	25	9	105583
26	26	9	116070

Figure-5 Cleaning Data

```
[17]:  import matplotlib.pyplot as plt

[18]:  plt.title("Salary of the Employee based on Experience")
       plt.xlabel("Years of Experience")
       plt.ylabel("Salary")
       plt.scatter(x,y,color = "red")
       #plt.plot(x,color = "black")
       #plt.plot(y,color = "cyan")
       plt.grid()
       plt.show()
```

Figure-6 Code for plotting Dataset on Graph

Figure-7 Image of data plotted on a Graph

```
[19]:  from sklearn.linear_model import LinearRegression

[20]:  model = LinearRegression()

[21]:  model.fit(x,y)

[21]:      ▼    LinearRegression  ⓘ ❓

           LinearRegression()

[22]:  op = model.predict([[12]])

[23]:  op[0]

[23]:  142556.46744499326

[24]:  import pickle

[25]:  pickle.dump(model,open("Data.pkl","wb"))
```

Figure-8 importing and using linear regression model and pickle from respective packages

```
1   import pickle
2   import streamlit as st
3
4   model = pickle.load(open("Data.pkl","rb"))
5
6   def fun1():
7       st.title("Predicting the Salary of a employee based on his experience")
8       area = st.number_input("Enter the year of experience of the employee")
9       pred = st.button("Start")
10      if pred:
11          op = model.predict([[area]])
12          st.write("The Predicted Salary of the Employee is : ",op[0])
13
14
15  fun1()
```

Figure-9 importing and using pickle and streamlit libraries

	A	B	C	D	E
		YearsExpe	Salary		
	0	1.2	39344		
	1	1.4	46206		
	2	1.6	37732		
	3	2.1	43526		
	4	2.3	39892		
	5	3	56643		
	6	3.1	60151		
	7	3.3	54446		
	8	3.3	64446		
	9	3.8	57190		
	10	4	63219		
	11	4.1	55795		
	12	4.1	56958		
	13	4.2	57082		
	14	4.6	61112		
	15	5	67939		
	16	5.2	66030		
	17	5.4	83089		
	18	6	81364		
	19	6.1	93941		
	20	6.9	91739		
	21	7.2	98274		
	22	8	101303		
	23	8.3	113813		
	24	8.8	109432		
	25	9.1	105583		
	26	9.6	116970		
	27	9.7	112636		
	28	10.4	122392		
	29	10.6	121873		

Figure-10 Dataset with Irregularities

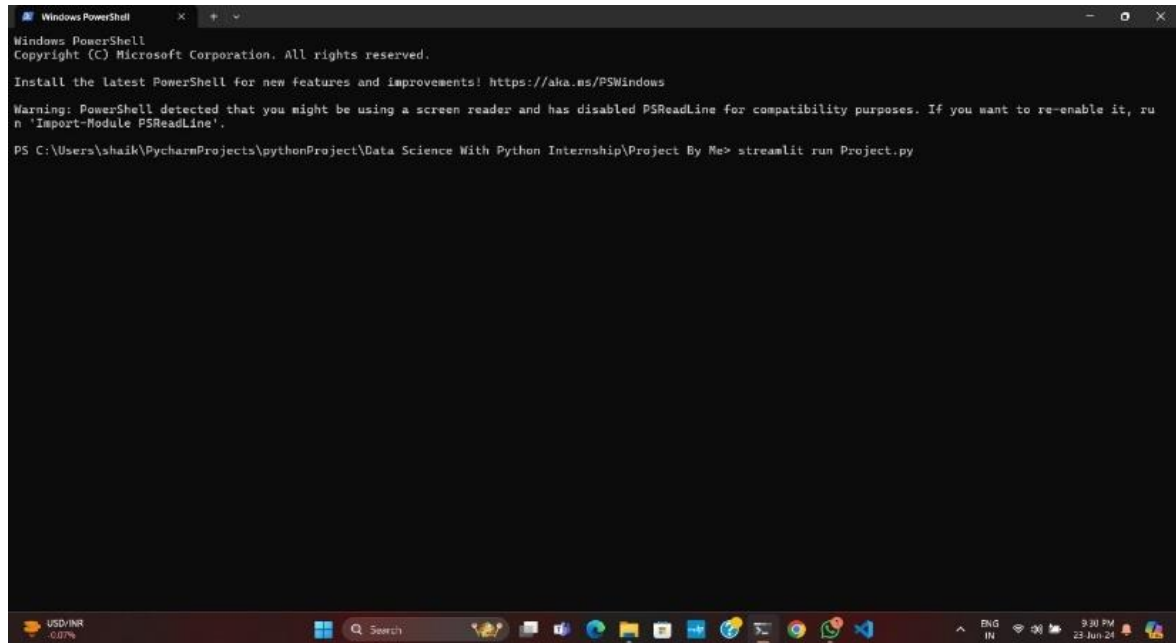# Input / Output with Datasets & Supported Screenshots

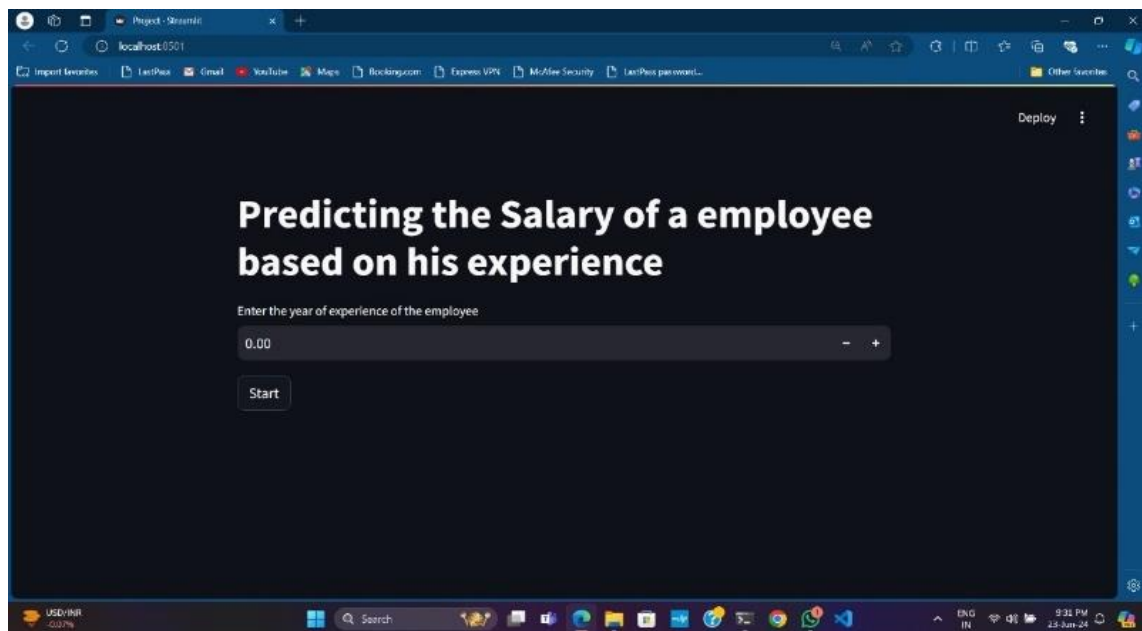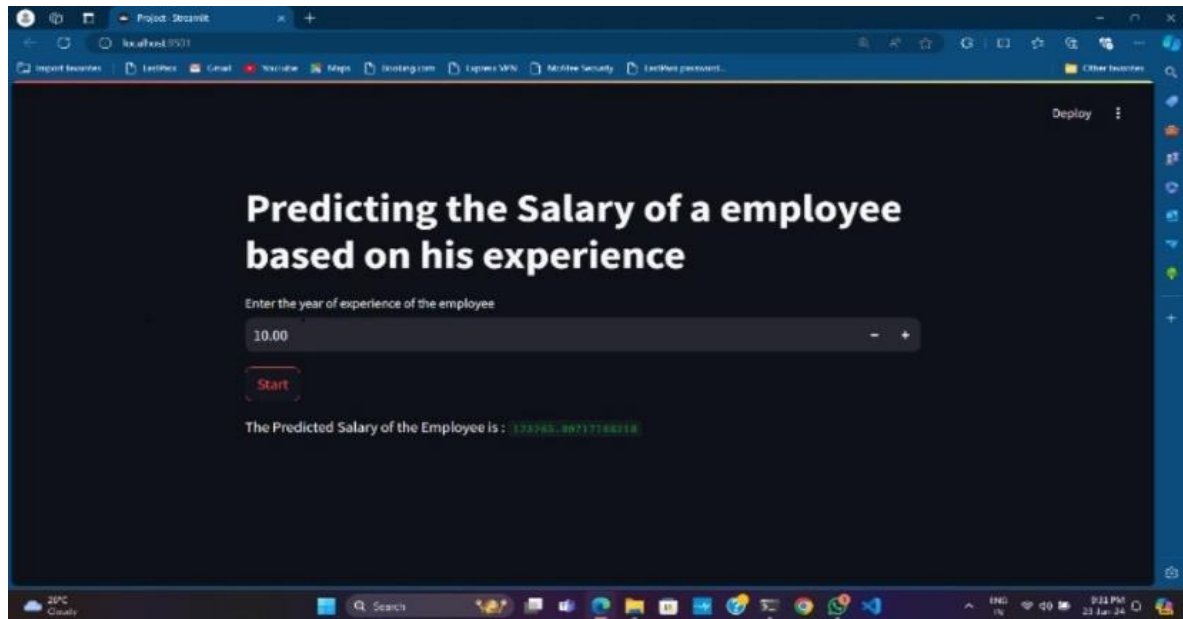

Figure-11 command to run the project

Figure-12 giving input to the project

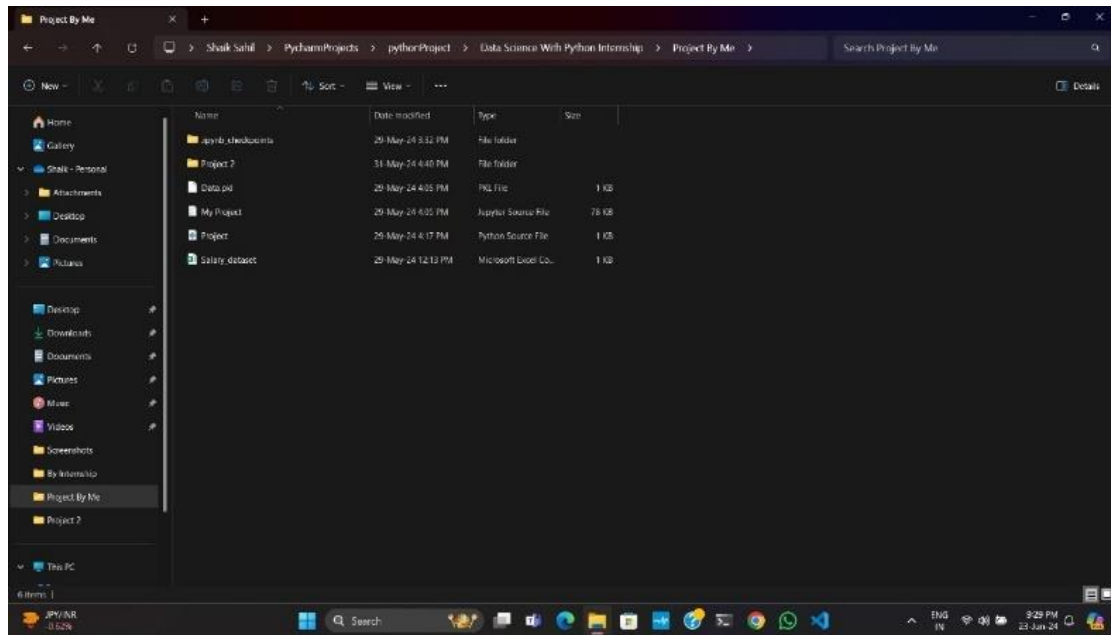**Figure-13 output predicted by the Project**

Images / Video Links
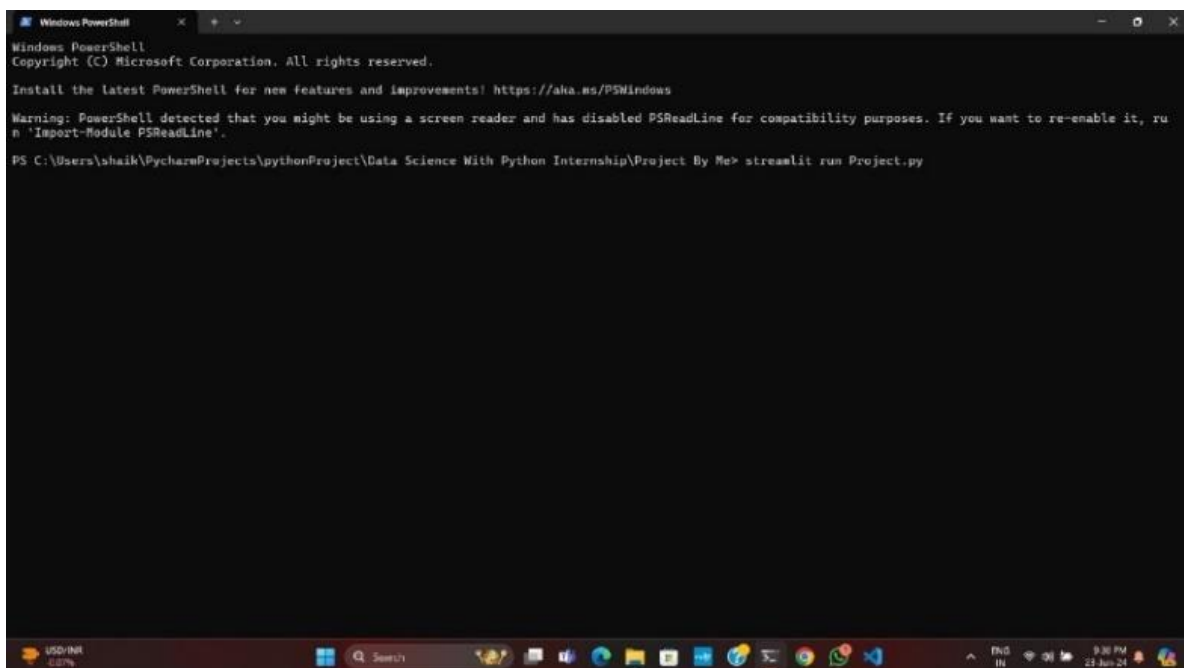

Figure-14 Documents Required for the Project

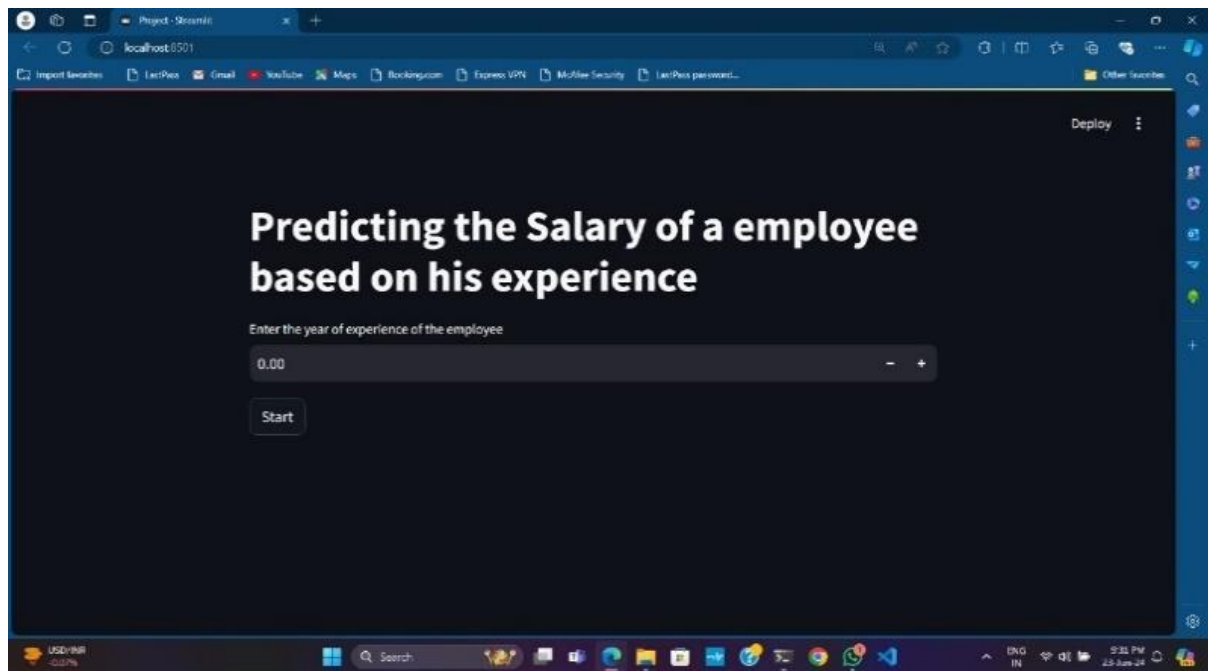**Figure-15 process to run the project**

Figure-16 Input

Figure-17 Output

Project Recording Video Link:
https://drive.google.com/file/d/18EvJiAv6Vuo-WqH_b-M7g8n7a5Lhr3Wz/view?usp=sharing

References and Requirements:

Salary Dataset :

https://www.kaggle.com/datasets/abhishek14398/salary-dataset-simple-linear-regression

Requitements:

- Python latest version

- pandas

- matplotlib

- streamlit

- scikit-learn

Project Recording Video Link:

https://drive.google.com/file/d/18EvJiAv6Vuo-WqH_b-M7g8n7a5Lhr3Wz/view?usp=sharing

Student *Self Evaluation of the Short-Term Internship*

Please rate your performance in the following areas:

		1	2	3	4	5
1)	Oral communication	1	2	3	4	5
2)	Written communication	1	2	3	4	**5**
3)	Initiative	1	2	3	4	**5**
4)	Interaction with staff	1	2	3	**4**	5
5)	Attitude	1	2	3	**4**	5
6)	Dependability	1	2	3	**4**	5
7)	Ability to learn	1	2	3	4	**5**
8)	Planning and organization	1	2	3	4	**5**
9)	Professionalism	1	2	3	**4**	5
10)	Creativity	1	2	3	**4**	5
11)	Quality of work	1	2	3	4	**5**
12)	Productivity	1	2	3	**4**	5
13)	Progress of learning	1	2	3	4	**5**
14)	Adaptability to organization's culture/policies	1	2	**3**	4	5
15)	OVERALL PERFORMANCE	1	2	3	**4**	5

Rating Scale: 5 will be Best while 1 will be Worst

R.S. Mahammad Sahil

Signature of the Student

Annexure 1

Daily Activity Report

EISYSTEMS SERVICES
FF-110, Express Greens Plaza, Sector 1
Vaishali – Delhi NCR – India 201010
W: www.eisystems.in | E: info@eisystems.in | T: +91 92122-51000

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Introduction to Python	Ran a basic Python script	Ms. Mallika Srivastava
Day 2	Variable, Constant & Naming Convention	Understood variables and constants.	Ms. Mallika Srivastava
Day 3	Print Function & Comments	Mastered the print function.	Ms. Mallika Srivastava
Day 4	Starting with Datatypes - Number Datatypes	Understood numeric datatypes	Ms. Mallika Srivastava
Day 5	Starting with Datatypes - String Datatypes	Used built-in string methods.	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Daily Activity Report

Week No: 2

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Number Datatypes	Understood numeric datatypes.	Ms. Mallika Srivastava
Day 2	String Datatypes	Used built-in string methods effectively.	Ms. Mallika Srivastava
Day 3	List	Practiced list manipulation techniques.	Ms. Mallika Srivastava
Day 4	Tuple	Practiced tuple operations.	Ms. Mallika Srivastava
Day 5	Dictionary and Set	Understood dictionary and set operations	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Daily Activity Report

Week No: 3

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Methods	Understood method definition and usage.	Ms. Mallika Srivastava
Day 2	Tuple	Reinforced understanding of tuples.	Ms. Mallika Srivastava
Day 3	Dictionary	Practiced manipulating key-value pairs	Ms. Mallika Srivastava
Day 4	Set	Practiced using set methods effectively	Ms. Mallika Srivastava
Day 5	User Input & Typecasting	Mastered handling user input	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Daily Activity Report

Week No: <u>4</u>

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Boolean	Understood Boolean values and logic.	Ms. Mallika Srivastava
Day 2	Typecasting	Practiced converting between different datatypes.	Ms. Mallika Srivastava
Day 3	Control Statements - If, Elif, Else	Practiced controlling program flow with if, elif, and else.	Ms. Mallika Srivastava
Day 4	Control Statements - Nested If and Conditional Expressions	Practiced using nested conditions and ternary operators.	Ms. Mallika Srivastava
Day 5	Project - Implementing Control Statements	Enhanced problem-solving and debugging skills.	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Daily Activity Report

Week No: 5

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Loop in Python	Practiced implementing loops and control statements.	Ms. Mallika Srivastava
Day 2	Project 1: Quiz Game	Applied loops and control statements in a project.	Ms. Mallika Srivastava
Day 3	File Handling	Practiced reading and writing data to files.	Ms. Mallika Srivastava
Day 4	Function	Practiced creating and using functions in programs.	Ms. Mallika Srivastava
Day 5	Packages and Modules	Understood the modular approach in Python.	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Daily Activity Report

Week No: 6

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Exception Handling	Understood the necessity of exception handling.	Ms. Mallika Srivastava
Day 2	Object-Oriented Programming (OOP) Concepts	Gained a solid understanding of OOP concepts.	Ms. Mallika Srivastava
Day 3	Advanced OOP Concepts	Practiced enhancing code reusability and organization using OOP	Ms. Mallika Srivastava
Day 4	Numpy	Understood the functionality of Numpy.	Ms. Mallika Srivastava
Day 5	Pandas	Gained proficiency in using Pandas for data analysis.	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Daily Activity Report

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Matplotlib	Understood the fundamentals of data visualization with Matplotlib.	Ms. Mallika Srivastava
Day 2	Advanced Matplotlib	Practiced creating more complex and visually appealing plots.	Ms. Mallika Srivastava
Day 3	Model Demonstration	Practiced model training, testing, and evaluation.	Ms. Mallika Srivastava
Day 4	Models and Projects	Practiced implementing and comparing multiple models in a project.	Ms. Mallika Srivastava
Day 5	Model Deployment	Practiced deploying a model to a web service.	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Daily Activity Report

Day & Date	Brief Description of DailyActivity	Learning Outcome	Person In-Charge
Day 1	Review of Core Concepts	Reinforced understanding of core Python concepts.	Ms. Mallika Srivastava
Day 2	Comprehensive Project Planning	Developed a clear project plan.	Ms. Mallika Srivastava
Day 3	Data Collection and Preparation	Acquired skills in data collection and preprocessing.	Ms. Mallika Srivastava
Day 4	Model Implementation	Practiced implementing and training machine learning models	Ms. Mallika Srivastava
Day 5	Final Project Presentation and Deployment	Gained hands-on experience in deploying a machine learning model.	Ms. Mallika Srivastava
Day 6	Writing report	Enhanced report writing skills	Ms. Mallika Srivastava

Annexure 2

Weekly Progress Report

Week No: (1/2/3/4/5/6/7/8)

Week(s)	Summary of Weekly Activity
Week 1	Introduction to Python and Basic Concepts
Week 2	Datatypes and Collections
Week 3	Advanced Data Types and User Input
Week 4	Boolean, Typecasting, and Control Statements
Week 5	Loops and Project Implementation
Week 6	File Handling, Functions, and Modules
Week 7	Advanced Programming Concepts
Week 8	Data Science Libraries and Model Deployment