**CHAPTER 1**

**INTRODUCTION**

**1.1 BACKGROUND OF PYTHON:**

Python is a high-level programming language that Guido van Rossum first published in 1991. It was initially created as a scripting language for automating system functions, however it has since developed into a general-purpose computer language with numerous uses. Python has gained a lot of fame in recent years, in part because of its simplicity and usability as well as its powerful libraries and frameworks. Python is the perfect language for rapid prototyping and development because it can be used to run code without the need for compilation. Web creation, data analysis, artificial intelligence, machine learning, scientific computing, and many other areas use Python today. Building applications quickly and effectively is simple with Python due to its sizable and vibrant developer community, which contributes to a wide range of libraries and frameworks.

**1.2 IMPORTANCE OF PYTHON:**

Python is a flexible, high-level programming language that has become increasingly well-liked over time as a result of its clarity, readability, and usability. Here are some of the main justifications for why Python is significant:

Simple to learn: Python is an approachable language that is simple to learn and comprehend. It is the perfect language for beginners to start with due to its straightforward syntax and organised structure.

Large and vibrant community: Python has a vibrant and large developer community that actively contributes to the language by building tools, libraries, and frameworks. This makes it simple for developers to locate solutions to issues and seek assistance when required. Python has a wide range of applications, including web development, data science, machine learning, artificial intelligence, and many others.

Versatility: Python is a high-level language that abstracts away many of the difficulties of low-level languages, including memory management and pointer arithmetic. This makes writing and reading code simpler. Python is an interpreted language, which means that it executes code line by line, making it simple to test and debug programmes.

Open source: Python is a language that is available for free use and distribution. This has helped it become well-liked and widely used. Overall, Python is a significant language that is used extensively across numerous industries and has a promising future as more developers choose to use it.

**1.3 BACKGROUND OF MACHINE LEARNING:**

Artificial intelligence (AI) has a subfield called machine learning that uses models and statistical algorithms to help computers learn from data and get better at a task without having to be explicitly programmed. Early computer scientists like Arthur Samuel and Frank Rosenblatt explored the idea of building machines that could learn from data in the 1950s, which is when machine learning first emerged. Machine learning did not start to gain widespread attention or become a useful tool for solving problems in the real world, however, until the development of more sophisticated algorithms and computational power in the 1990s.

A branch of artificial intelligence known as machine learning uses models and statistical methods to help computers learn from data and execute tasks better without explicit programming. With applications in a variety of domains, including picture and audio recognition, natural language processing, autonomous cars, fraud detection, and targeted marketing, machine learning has grown in importance as a subject of study and development.

**1.4 IMPORTANCE OF MACHINE LEARNING:**

Machine learning (ML), a subset of artificial intelligence (AI), is the branch of computational science that is concerned with the analysis and interpretation of patterns and structures in data to enable learning, reasoning, and decision-making without the involvement of a human. Simply put, machine learning enables users to feed massive amounts of data into computer algorithms, which then analyse, recommend, and decide using only the input data. The algorithm can use the information to improve its decision-making in the future if any corrections are found. Due to its ability to analyse vast amounts of data and make predictions or identify patterns that may be challenging or impossible for humans to detect, machine learning has become more and more significant in recent years. Several of the main justifications for machine learning's significance are listed such as:

Automation: A variety of tasks, including data analysis, image recognition, natural language processing, and others, can be automated using machine learning algorithms. By streamlining procedures that were previously carried out manually, this can save businesses and individuals time and money.

Predictive Analytics: Machine learning models have the ability to analyse vast amounts of data and forecast future trends and outcomes. This capability can be very helpful in industries like finance, healthcare, and marketing.

**1.5 BACKGROUND OF AI:**

AI termed as Artificial Intelligence is the study of intelligent machines that can carry out tasks that ordinarily require human intelligence, such as speech recognition, decision-making, and language translation, is known as artificial intelligence (AI). The beginnings of artificial intelligence can be found in the middle of the 20th century, when computer scientists started to create models and algorithms that could mimic human thought processes. One of the earliest AI programs was developed in the 1950s by John McCarthy, Marvin Minsky, and other researchers at the Massachusetts Institute of Technology (MIT).

AI research grew to accommodate fields of study like expert systems, computer vision, and natural language processing in the 1960s and 1970s. Expert systems were computer programmes that could simulate how human decision-makers in a particular field, like engineering or medicine, would make decisions. However, due to limited computing power and a lack of data, AI research progressed slowly in the decades that followed. AI did not begin to see a resurgence until the twenty-first century, with the introduction of big data and developments in machine learning. Speech recognition, image and video analysis, autonomous vehicles, and natural language processing are just a few of the applications that use AI today. Predictive models are also being created using it for the healthcare, financial, and other industries.

**1.6 IMPORTANCE OF AI:**

Artificial intelligence (AI) is the simulation of human intelligence in machines that are designed to learn and solve issues on their own, without the assistance of a human. AI is extremely important and has the potential to completely transform many facets of our lives. The field of healthcare is one of the most significant areas for AI applications. AI has the potential to lower healthcare costs and improve patient outcomes. AI-powered systems can aid in disease diagnosis, the selection of appropriate treatments, and outcome prediction.

The field of transportation is one where artificial intelligence is used extensively. Self-driving cars with AI capabilities can make transportation systems more effective and decrease the number of accidents brought on by human error. Additionally, self-driving cars can lessen emissions, ease traffic congestion, and improve accessibility for those who are unable to drive. AI can also be used to make manufacturing processes more effective. Artificial intelligence (AI)-enabled systems can examine production data to pinpoint areas for improvement and optimise production schedules. This can aid businesses in cutting waste, enhancing quality, and boosting output. AI can also be used to anticipate equipment breakdowns, giving businesses the opportunity to perform preventative maintenance and reduce expensive downtime. Many facets of our lives could be revolutionised by artificial intelligence. It can enhance cybersecurity, optimise manufacturing processes, give insights into financial markets, transform education, and improve healthcare outcomes. It can also increase construction safety, make transportation safer, and make it more efficient.

**CHAPTER 2**

**FUNDAMENTALS OF PROGRAMMING USING PYTHON, ML AND AI**

**2.1 FUNDAMENTALS OF PYTHON:**

Python is a robust and adaptable language that can be applied to a variety of applications. Python offers a user-friendly and flexible environment for developing your programs, regardless of your programming experience.

Variables and Data Types: In Python, a variable's data type does not have to be stated explicitly. Python will choose the appropriate data type for you when you simply assign a value to a variable. In Python, floats, strings, Booleans, and integers are the most prevalent data types.

Control Structures: If-else statements, loops, and functions are just a few of the control structures available in Python. These structures give you the ability to manage the flow of your code and base decisions on specific circumstances.

Functions: Functions are code segments that can be applied repeatedly throughout your programme. They can accept parameters as input and return a value as output, and they are defined using the "def" keyword.

Lists and dictionaries: One of the most popular data structures in Python is the list. They can be altered while still in use and are used to store a variety of items. On the other hand, key-value pairs are kept in dictionaries.

Modules and packages: You can import modules and packages into your programme from Python's sizable library. You can use the extra functionality that these modules offer to increase the effectiveness and power of your code.

Comments: Comments are used to describe and document your code. Comments in Python begin with the # symbol and can appear at the beginning or end of a line of code.

Virtual Environments: Python enables you to build isolated environments called virtual environments, each of which has its own Python interpreter and set of packages. When working on numerous projects with various dependencies, this is helpful.

Libraries: The Python programming language has a large and vibrant community that has produced numerous libraries and frameworks for a variety of uses. For data analysis, popular libraries include NumPy and Pandas; for data visualization, Matplotlib; for web development, Django and Flask; and for machine learning, TensorFlow and PyTorch.

**2.2 FUNDAMENTALS OF MACHINE LEARNING:**

Machine learning field gives systems the ability to automatically learn from their experiences and get better over time without having to be explicitly programmed. The following are some machine learning basics:

Data: In order to learn from machine learning, data is needed. Text, images, audio, and sensor data are just a few examples of the different types of sources that this data may be, both structured and unstructured.

Algorithms: Models that can predict the future or make decisions based on data are trained using machine learning algorithms. These algorithms can use decision trees, neural networks, support vector machines, and clustering algorithms. They can also be supervised, unsupervised, or semi-supervised.

Training: The machine learning algorithm learns patterns and relationships in the data to make predictions or decisions. This process involves feeding data into the algorithm. Enhancing the model's performance on a given task is the aim of training.

Evaluation: After the model has been trained, it must be examined in order to gauge its effectiveness. To determine how well the model generalises to new data, it must first be tested on a different set of data.

Deployment: After the model has been trained and assessed, it can be used in real-world applications to make predictions or decisions. Deployment may entail integrating the model into current systems, creating new applications based on the model, or hosting the model using cloud-based services.

**2.3 FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE:**

Machine learning is a branch of artificial intelligence that entails teaching algorithms to make judgements or predictions based on information. In order to enable machines to recognise patterns and relationships in data without explicit programming, statistical techniques are used. Making machines capable of performing tasks that would otherwise require human intelligence is one of the main objectives of artificial intelligence. This includes activities like identifying trends in data, making choices based on in-depth knowledge, and interacting with the real world.

Natural Language Processing: The ability of machines to comprehend and process human language is known as natural language processing, or NLP. It involves activities like sentiment analysis, language translation, and speech recognition.

Robotics: A subfield of artificial intelligence that deals with creating and programming robots to carry out specific tasks. From easy tasks like picking up and placing objects to more difficult tasks like autonomous navigation, these tasks can be of any complexity.

Expert Systems: AI programmes that mimic the judgement of a human expert in a given field are known as expert systems. To make judgements or offer suggestions, they rely on a body of knowledge and a set of regulations.

The capacity of machines to interpret and comprehend visual data from their surroundings is known as computer vision. This calls for activities like image segmentation, object recognition, and facial recognition.

Neural Networks: A type of machine learning algorithm, neural networks are loosely based on the composition and operation of the human brain. For tasks like speech recognition, natural language processing, and image recognition, they are employed.

**CHAPTER 3**

**PROBLEM STATEMENT**

AI modelling is the creation, training, and deployment of machine learning algorithms that emulate logical decision-making based on available data. AI models provide a foundation to support advanced intelligence methodologies such as real-time analytics, predictive analytics, and augmented analytics.

This project focuses on the development and deployment of predictive model of employee salary prediction based on the years of experiences of the employees. The software used in development of the model is Jupyter Notebook and Spyder. The software used in deployment of the model is Streamlit. The dataset consists of two columns; the first column is the ‘Years of Experience’ and the second column is the ‘Salary’. The concept of Linear Regression is used to develop the model. The output can be viewed in Google Chrome or any other web browser.

**CHAPTER 4**

**IMPLEMENTATION AND TESTING**

* Project Code: P15
* Project Name: Development & Deployment of Predictive Model of Employee Salary Prediction
* The software used in development of the model is Jupyter Notebook and Spyder
* The software used in deployment of the model is Streamlit

**4.1 TASKS PERFORMED:**

Machine Learning Model Building Steps:

* Load data set
* Check for null values and perform data cleaning
* Check features correlation
* Convert categorical data into numerical data if present
* Extract features(X) and labels(Y) from data set
* Split the data into train and test, then do feature scaling
* Train the model using suitable ML models
* Check accuracy and perform validation
* Deploy the model

General steps:

* Extracting the data form data set.
* Analysis of the data.
* Performing the basic operations.
* Developing the predictive model.

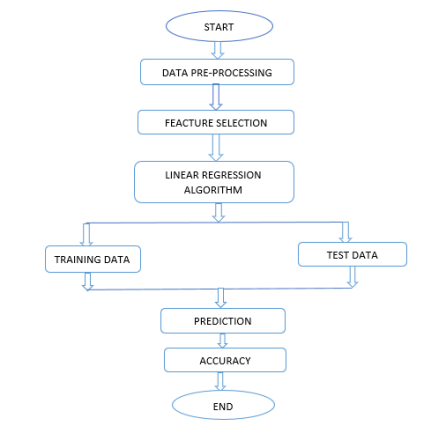
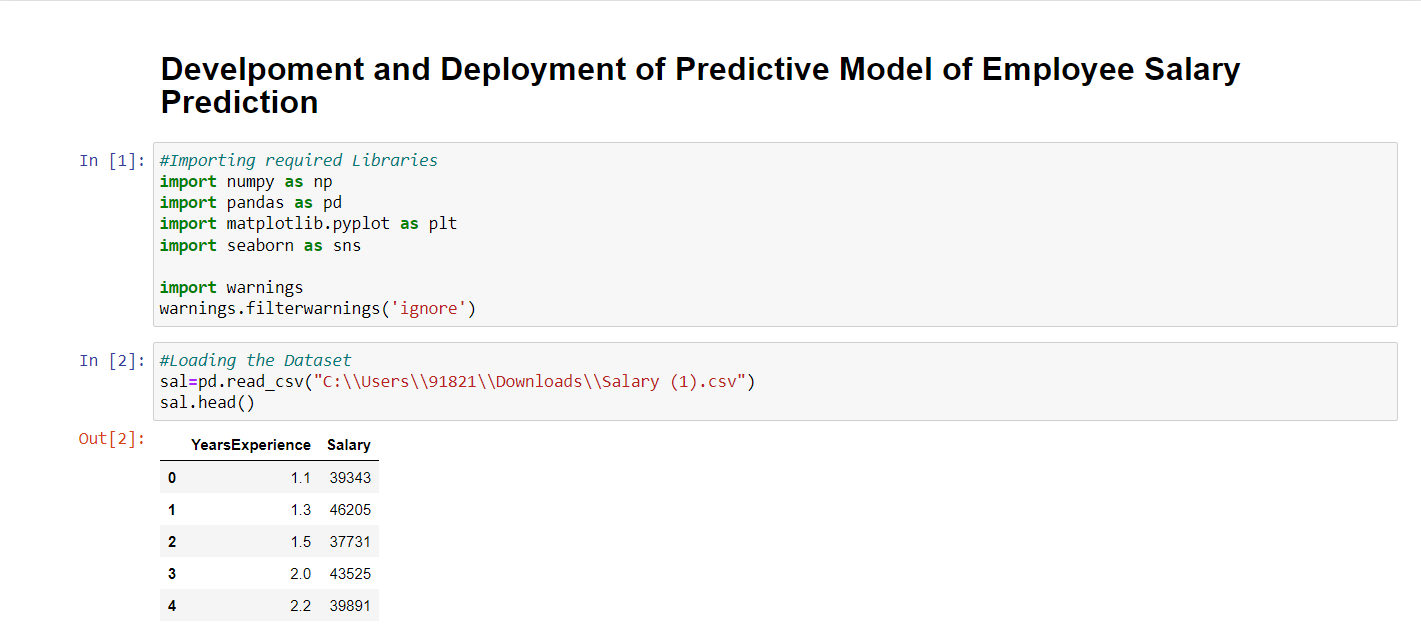
**4.2 DESIGN:**

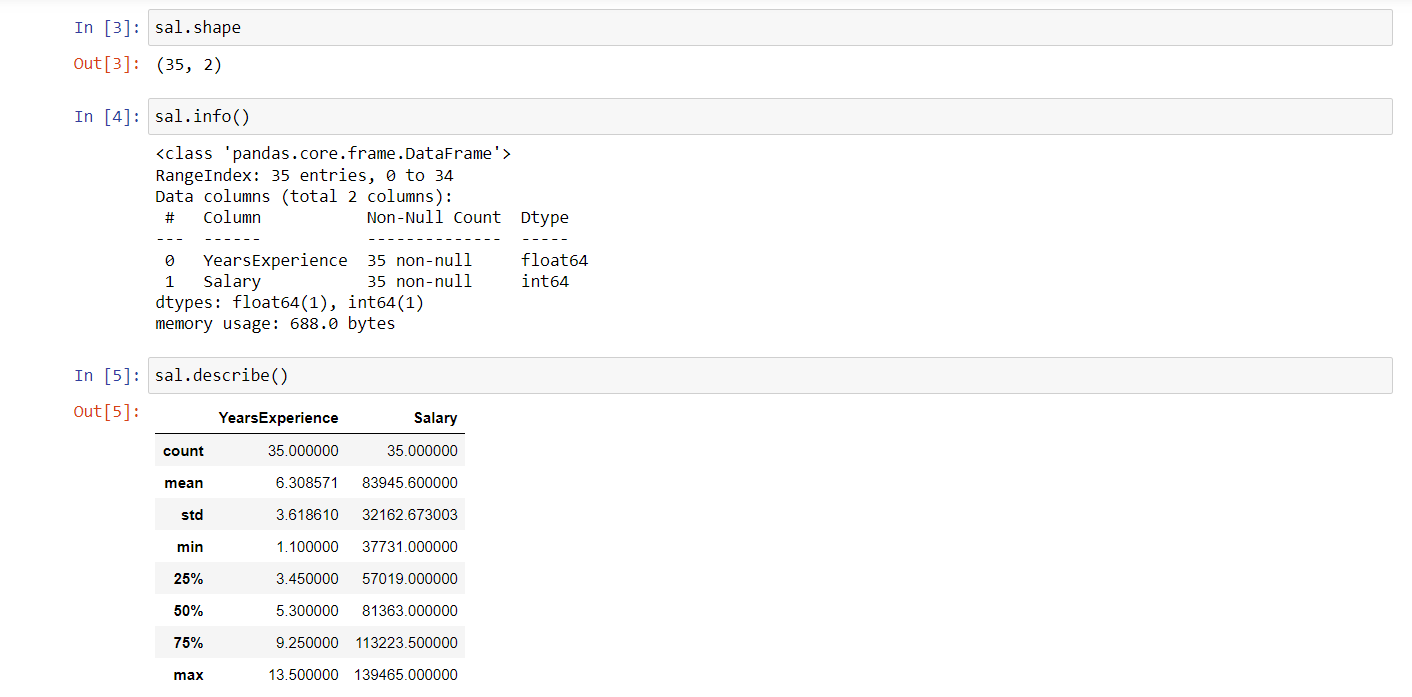
Fig 4.1: Data flow Diagram

The machine learning algorithm is started by importing the dataset. Then the dataset is processed. The processing includes operations like checking for null values, dropping unwanted rows or columns, converting categorical values into numerical values or vice versa, etc. If null values are present, then they must be removed. Next step is differentiating the features and labels. This process becomes simple by plotting a scatter graph or heatmap and checking the feature correlation. Then, the data is split into train and test data. The model is trained using Linear Regression and tested for accuracy.

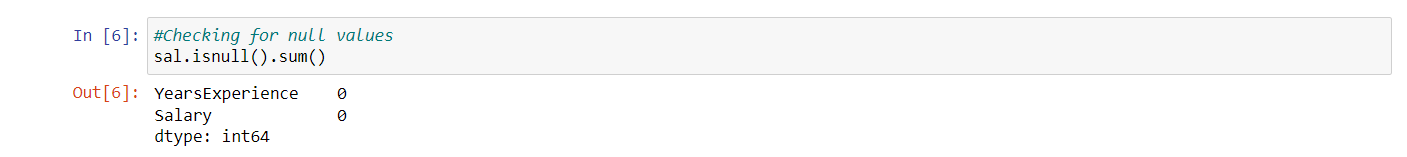
**4.3 DEVELOPMENT:**

1.Load data set





The first step in developing a predictive model is to import the required libraries using import() function and load the dataset.

2.Check for null values

Then, we check for null values, if any present using the isnull() function. As the result shows, there were no null values in the data so no need of data cleaning.

3.Check features correlation

The next step is to check the feature correlation by plotting a scatter plot and heat map from the seaborn library.

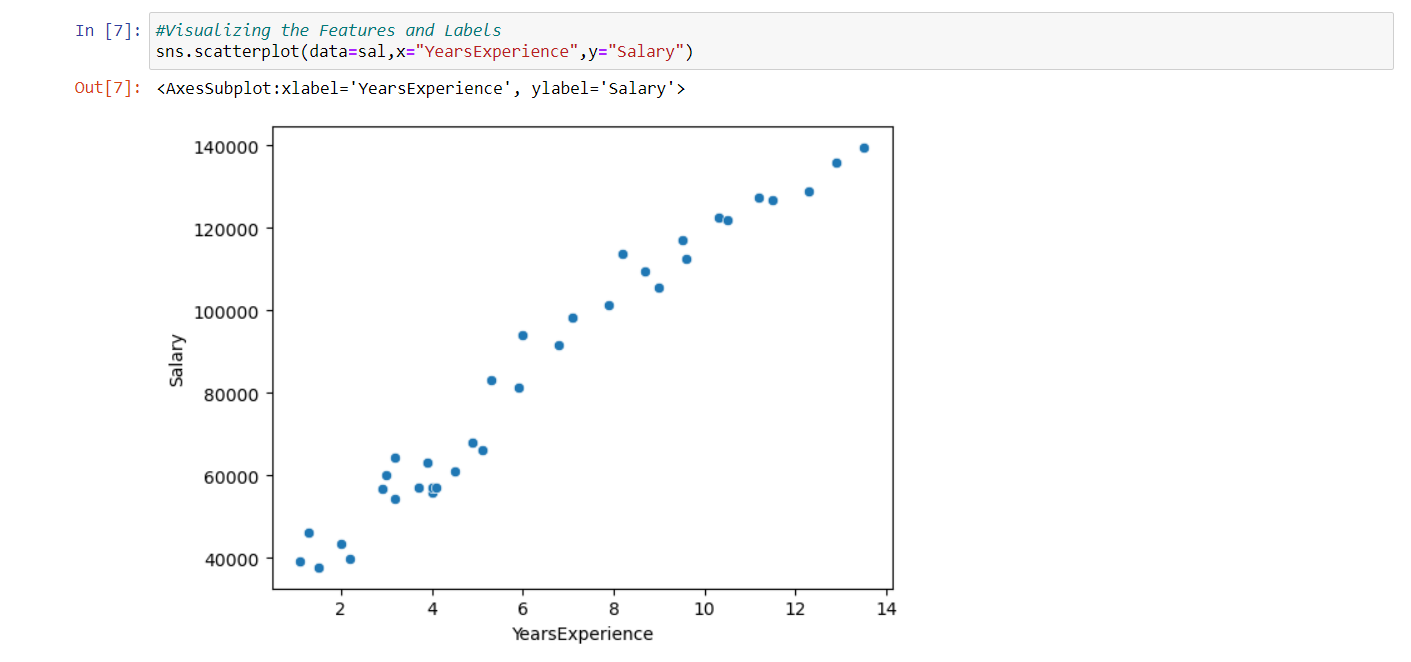


Fig 4.2: Analysis of Salary v/s Years of experience

The above graph is the scatter plot which is a bivariate analysis of salary v/s years of experience (numerical v/s numerical data).

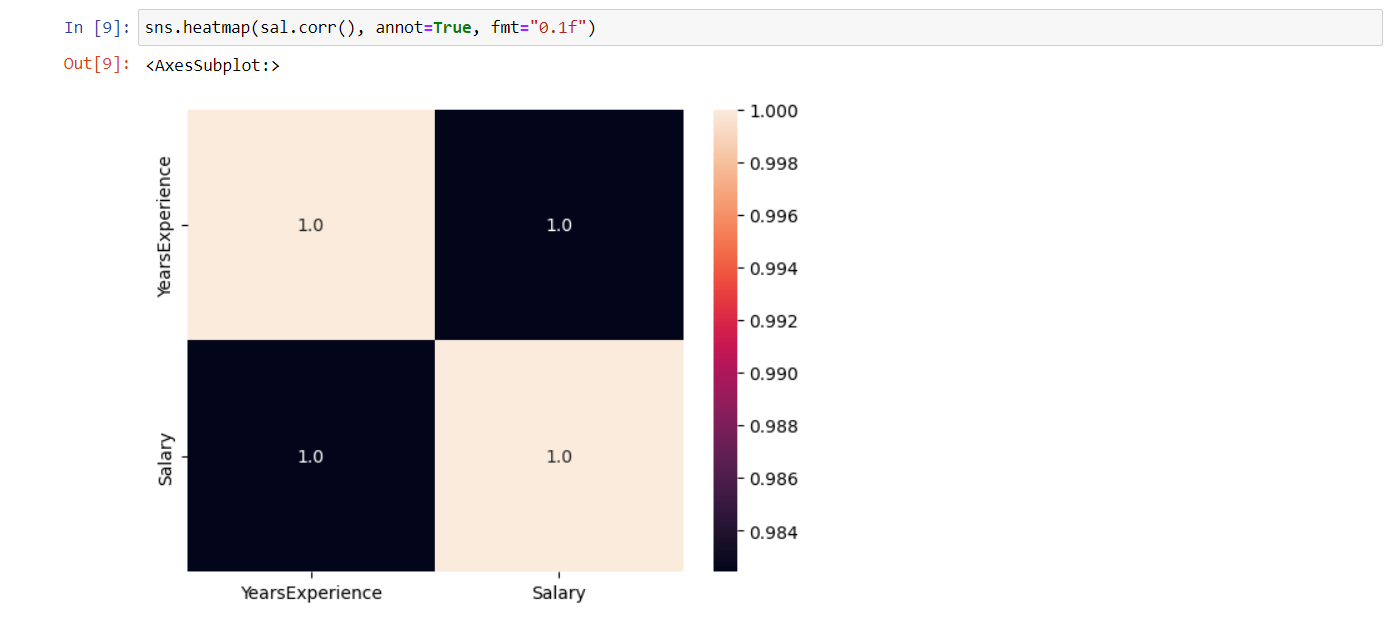
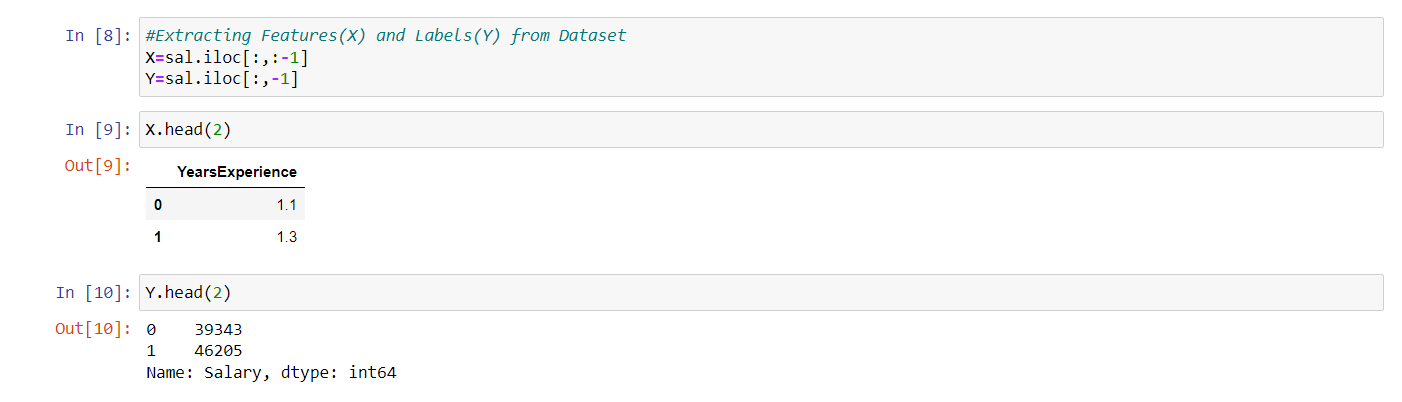


Fig 4.3: Analysis of salary v/s Years of Experience

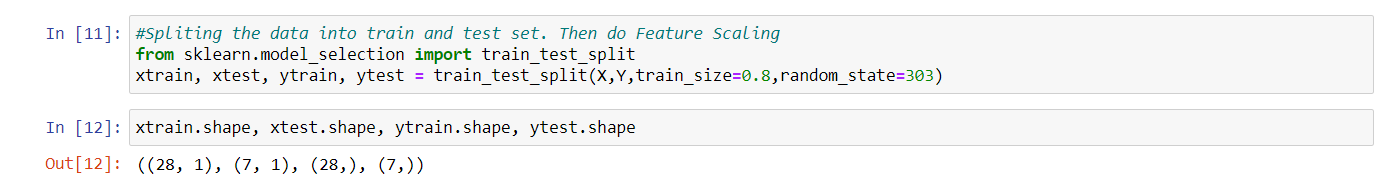
The above graph is the heatmap which is a multivariate analysis of different features. It is helpful in visualizing the correlation between different features.

4.Extract features and labels from data set



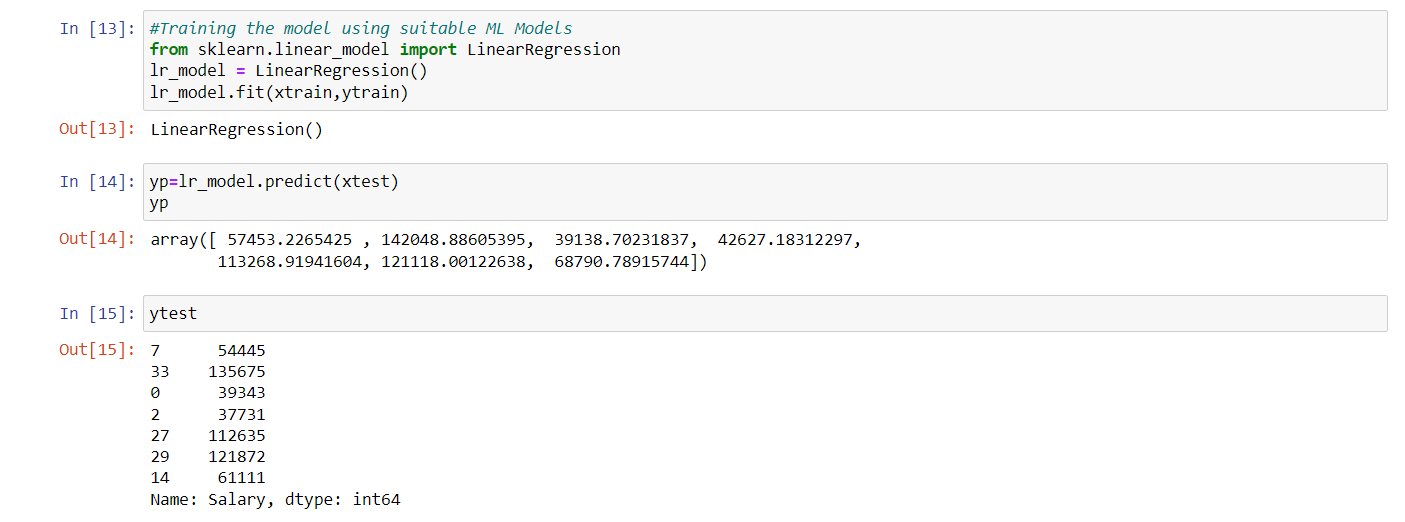
Then we extract the features(inputs) and labels(outputs) from the dataset.

5.Split the data into train and test set, then do feature scaling



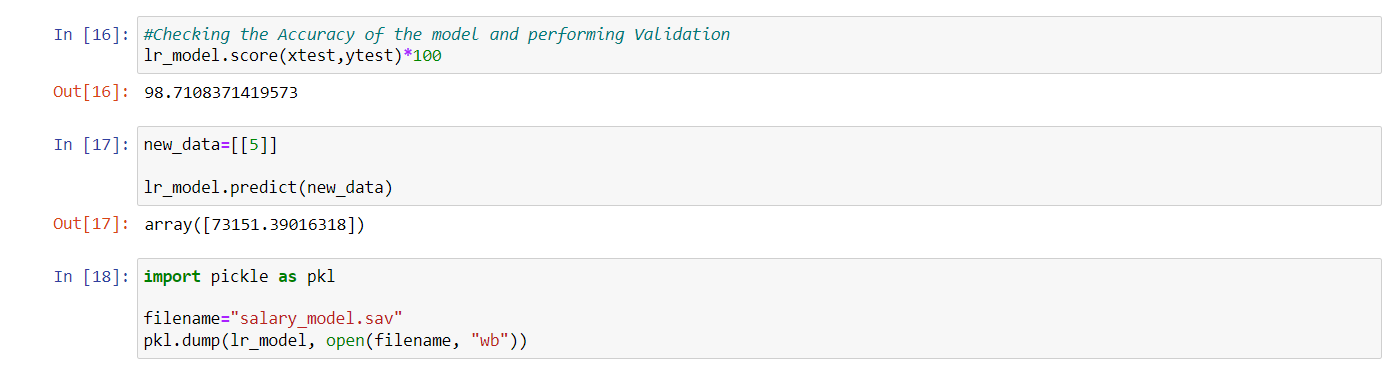
Next, we split the data into train set and test set (here, 80:20 ratio) using the train\_test\_split() function in the sklearn library.

6.Train the model using suitable ML methods



Then, we train the model using Linear Regression in the sklearn library.

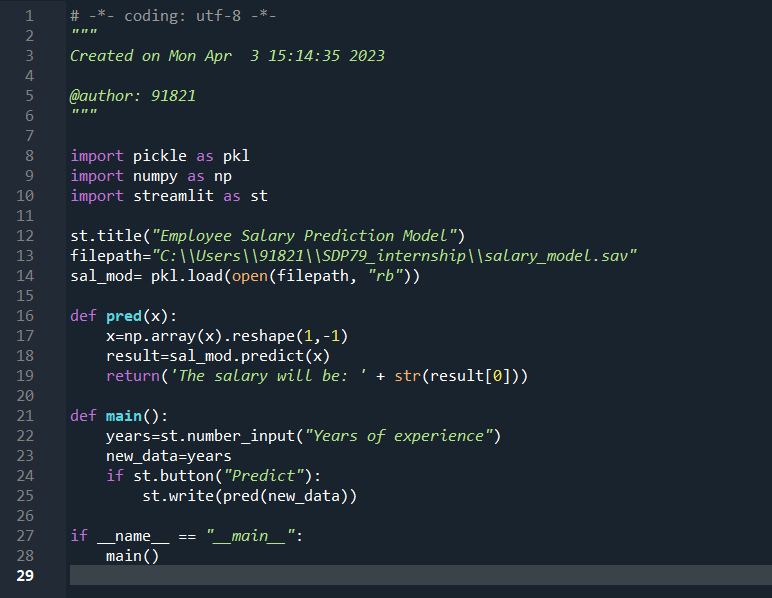
7.Check accuracy and perform validation

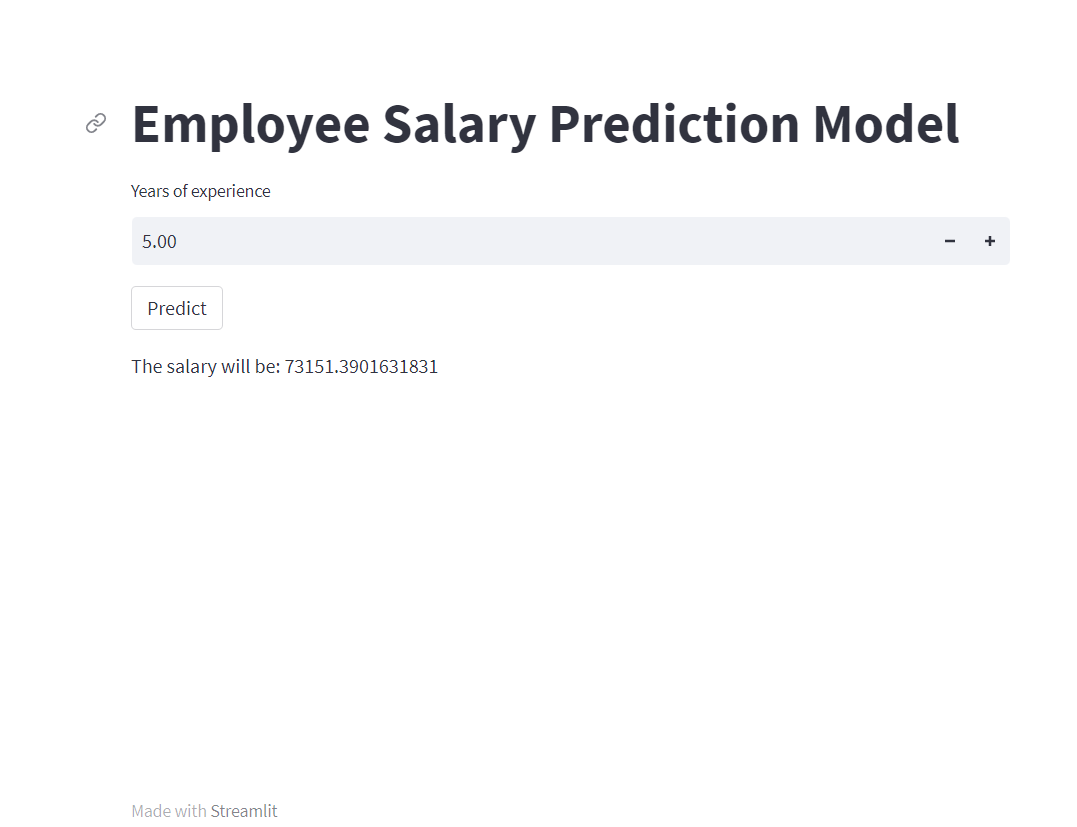


Lastly, we check for the accuracy of the model and test the model.

**4.4 DEPLOYMENT:**

For the deployment of this model, the code is written in ‘Spyder IDE’ and deployed using ‘Streamlit’ as shown below. The predicted output can be seen on any web browser.

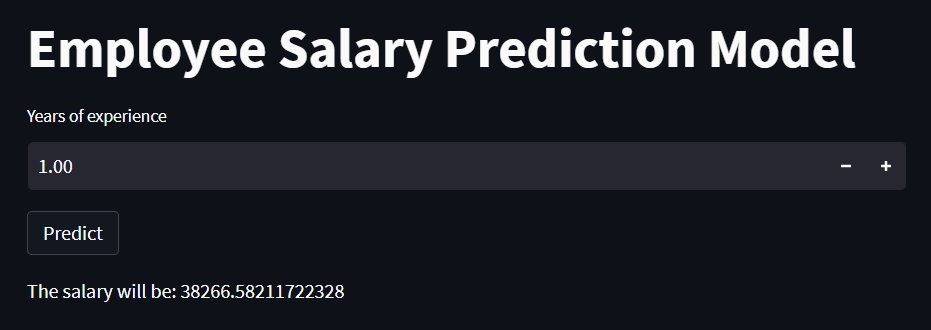


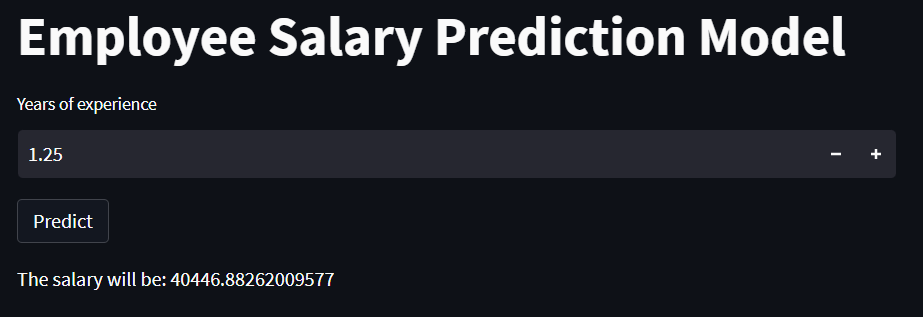


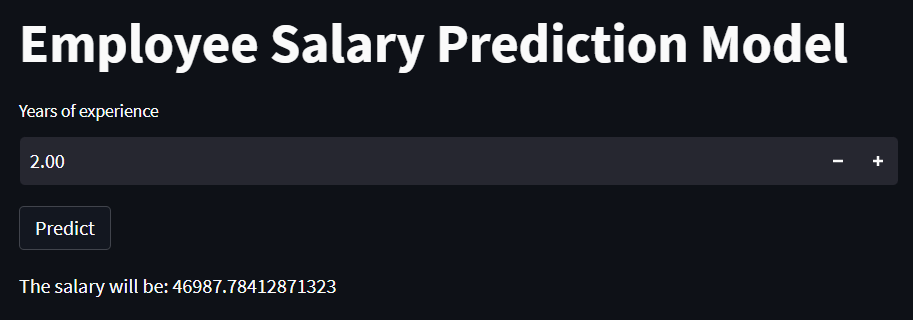
**CHAPTER 5**

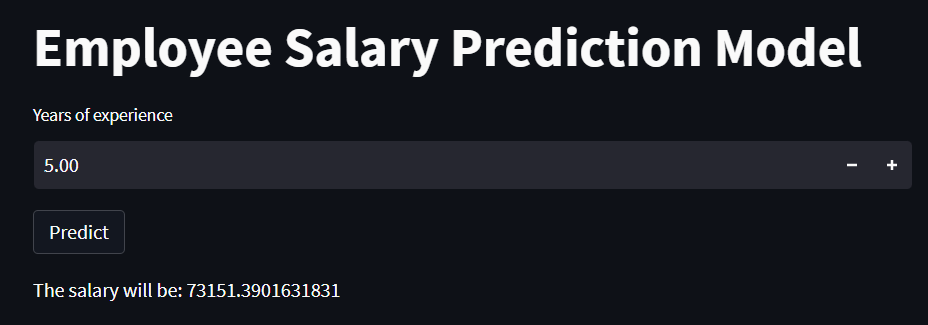
**RESULTS AND SNAPSHOTS**

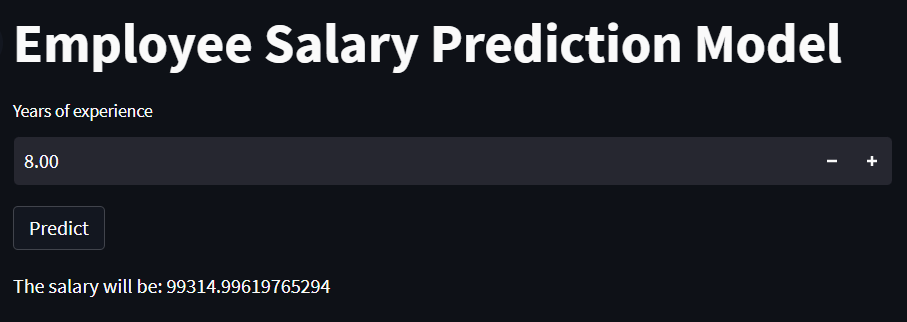
This main aim of this project is to predict the Salaries of Employees based on their years of experience. As the years of experience increases, the salary of the employees also increases.

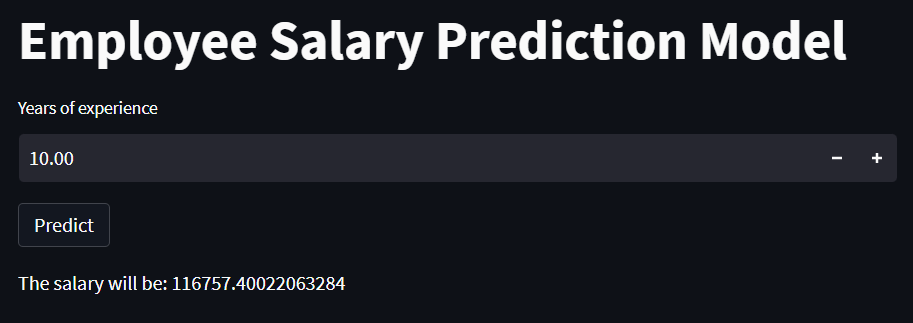


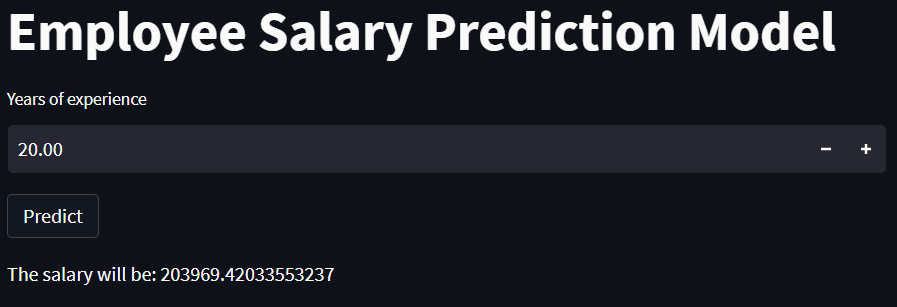












**CHAPTER 6**

**CONCLUSION**

Artificial Intelligence is everywhere, from gaming stations to maintaining complex information at work. Computer Engineers and Scientists are working hard to impart intelligent behaviour in the machines making them think and respond to real-time situations. AI is transiting from just a research topic to the early stages of enterprise adoption. Tech giants like Google and Facebook have placed huge bets on Artificial Intelligence and Machine Learning and are already using it in their products. But this is just the beginning, over the next few years, we may see AI steadily glide into one product after another.

Machine Learning can be applied to solve tough issues like credit card fraud detection, enable self-driving cars and face detection and recognition.ML uses complex algorithms that constantly iterate over large data sets, analysing the patterns in data and facilitating machines to respond different situations for which they have not been explicitly programmed. The machines learn from the history to produce reliable results. The ML algorithms use Computer Science and Statistics to predict rational outputs.

I hereby conclude my project design of employee salary prediction. Started with understanding what the project is about. Observed how the software works. Practical implementation of linear regression came in handy. After using various functions of python programming language, the data was extracted successfully and the model was trained.

In the end, the model was able to give appropriate results. Got 98% of accuracy which is conformation that the product works properly and there are no errors in building it.

Now aiming to build similar projects.

**REFERENCES**

[1] Navyashree M, Navyashree M K, Neetu M, Pooja G R, Arun Biradar “Salary Prediction in It Job Market”, International Journal of Computer Sciences and Engineering Vol.-7, May 2019.

[2] Pornthep Khongchai, Pokpong Songmuang, “Improving Students’ Motivation to Study using Salary Prediction System” 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).

[3] Phuwadol Viroonluecha, Thongchai Kaewkiriya,” Salary Predictor System for Thailand Labour Workforce using Deep Learning” The 18th International Symposium on Communications and Information Technologies (ISCIT 2018).

[4] Sananda Dutta, Airiddha Halder, Kousik Dasgupta,” Design of a novel Prediction Engine for predicting suitable salary for a job” 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN).

[5] <https://www.intel.in/content/www/in/en/analytics/data-modeling>

[6] Data Set Link: [*https://drive.google.com/drive/folders/1AEPt15EIszF59STdVKqYtkTOtOV7h\_Ho?usp=share\_link*](https://drive.google.com/drive/folders/1AEPt15EIszF59STdVKqYtkTOtOV7h_Ho?usp=share_link)