**Predicting high potential employees and employees at risk**

**1.1 Introduction**

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.

Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include:

* Speech recognition
* Learning
* Planning
* Problem solving

**1.2 Objectives of Research**

The main objective of the project is predicting the employees who are high potential and who are at risk in earlier stage (jobs) to promote the growth of their business, which is of importance to either an organization or a company and this project is done by using machine learning techniques.

**1.3 Problem Statement**

* It is important to solve the problem because company’s growth revolves around the work/projects done by the employees.

**2. Review of literature**

* Employees with high potential should be predicted/preferred initially so that companies can recruit them at earlier stages without any risk and they can be given higher salaries rather than giving to the employees who are not working.
* So if we predict an employee with high potential in the earlier stage itself we can solve the problem of job termination.

**3. Data Collection**

The data collection sources contain the above data:

1. **Satisfaction level**: Based on the satisfaction level i.e. (his/her effort or work done) employees are predicted whether they are high potential or at risk in earlier stages.

2**. Last evaluation**: Based on the last project done by the employee it gives the data of an employee and decides whether the employee is at high potential or at risk.

3. **Number projects**: It gives the number of projects done by the employee.

4. **Average monthly hours**: It gives the average data of an employees working hours per month.

5. **Time spent**: It gives the data on time spent by an employee.

6. **Work accident**: In case of any accidents occur during working hours

7. **Left**: Employees who leave the company.

8. **Promotion**: Employees promotion is based on their work experience and the accuracy on their projects.

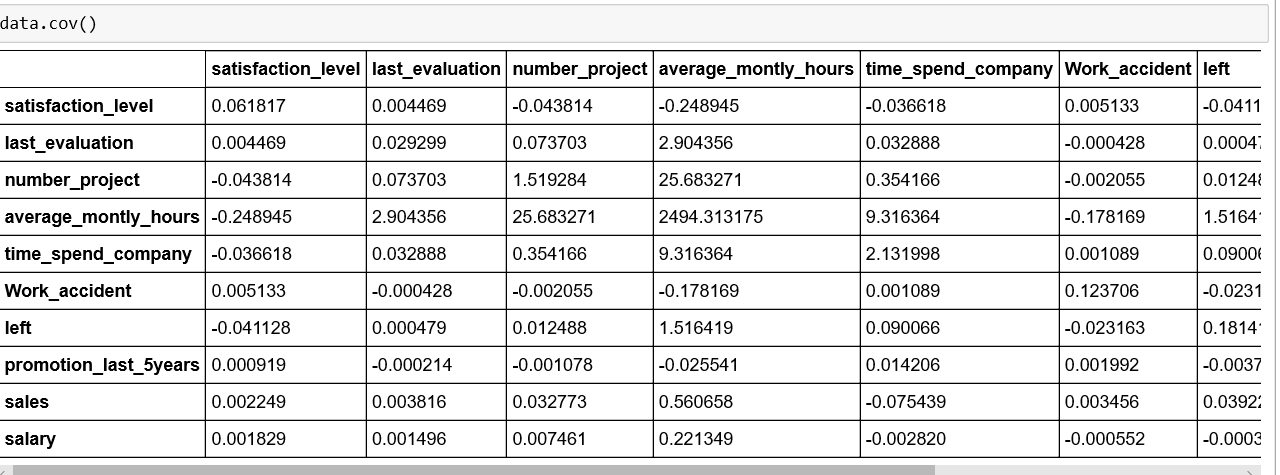
9. **Sales**: It deals with the employee where he/she belongs to.

10. **Salary**: It depends upon working experience and promotion of an employee.

**4. Methodology**

**4.1 Exploratory Data Analysis**

**4.1.1 Figures and Tables**

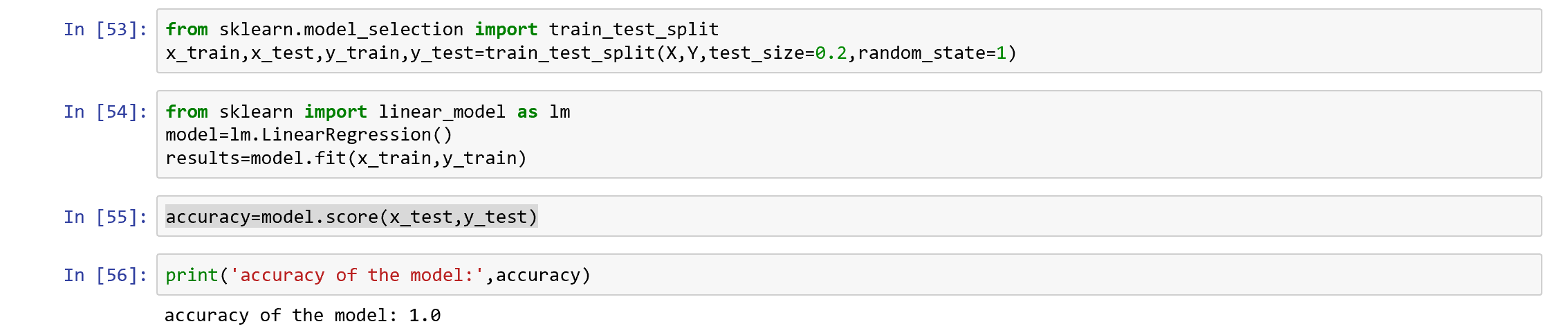
Covariance: 

**4.2 Data Modelling**

Here we have used two models :

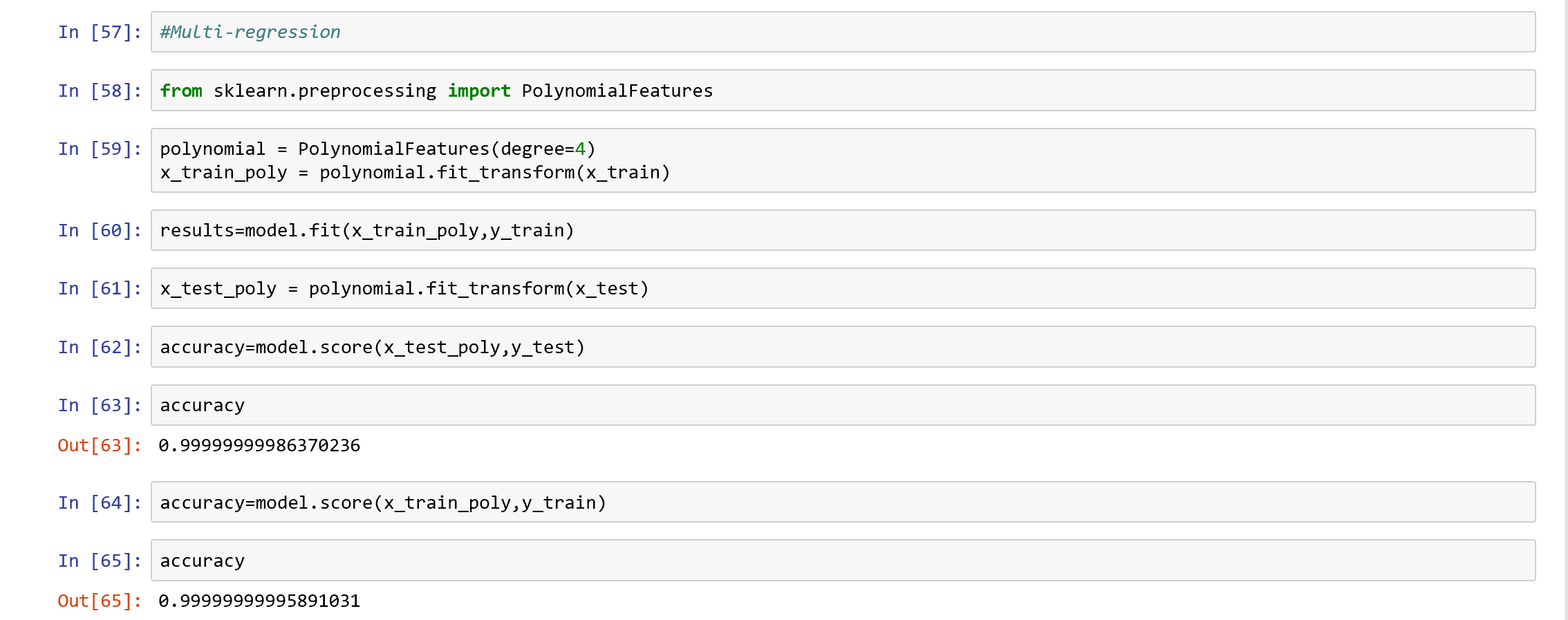
1**.Linear Regression** – **Linear Regression** is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

Output:



2. **Multi linear Regression-** Multiple Linear Regression attempts to model the Relationship between two or more features and a response by fitting a linear equation to observed data. The steps to perform multiple linear Regression are almost similar to that of linear Regression.

Output:



**5. Findings and Suggestions**

High potential employees and employees at risk is predicted by satisfaction level.

**6. Conclusion**

* We developed this project using python3.
* By taking the dataset we evaluated with different models and checked the accuracy for which model we can get maximum percentage so that particular model can be sustained and taken as good model.