Text

Description automatically generated

**Activity based**

**Project Report on**

**Software Engineering**

**Project Module - I**

**Submitted to Vishwakarma University, Pune**

**Under the Initiative of**

**Contemporary Curriculum, Pedagogy, and Practice (C2P2)**

**By**

**Maaz Sameer Shaikh**

**SRN No : 202101332**

**Roll No : 29**

**Div : E**

**Third Year Engineering**

**Faculty Incharge:- Pratibha Mahajan**

**Date Of Project 1:- 15/2/24**

**Department of Computer Engineering**

**Faculty of Science and Technology**

**Academic Year**

**2023-2024 Term-II**

**Machine Learning : Project I**

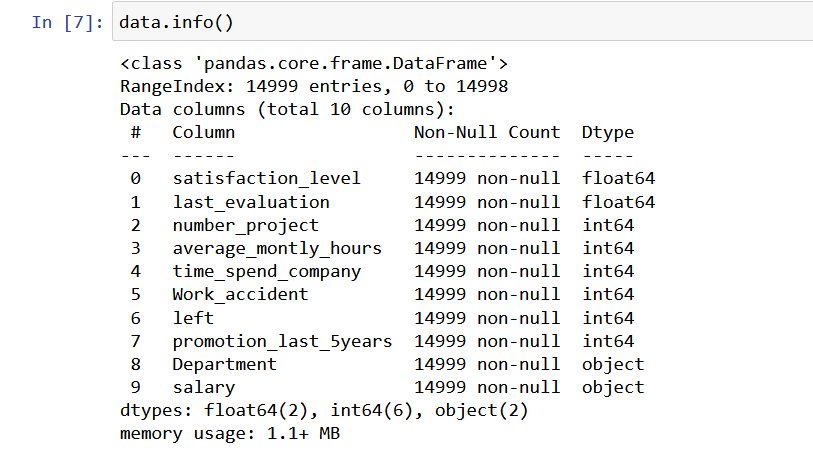
**Project Name: Forecast the demand for human resources in an organization based on business growth, project pipeline, and historical hiring patterns.**

**Dataset Information and Insights**

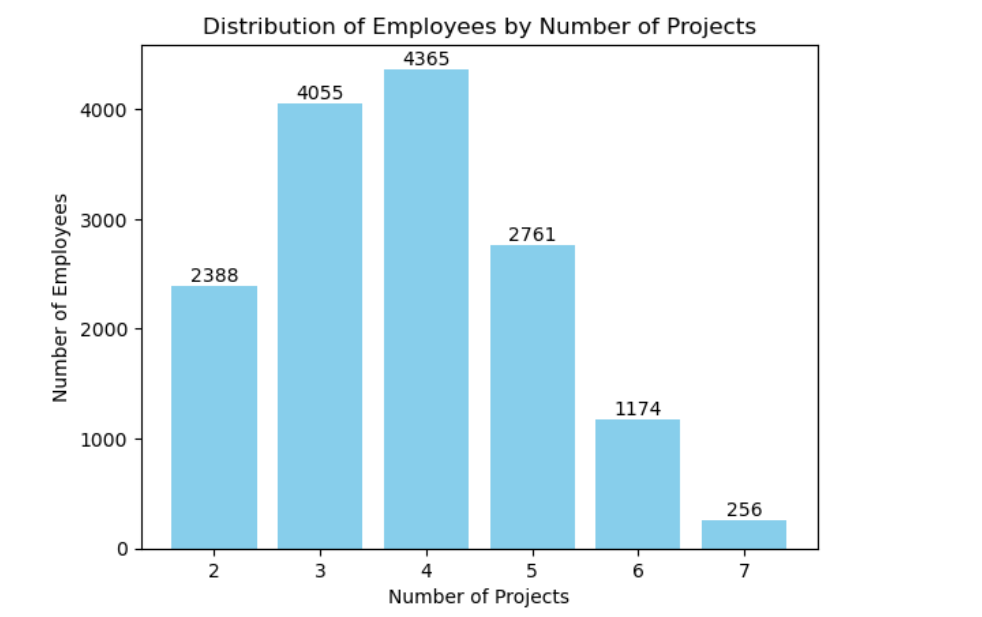
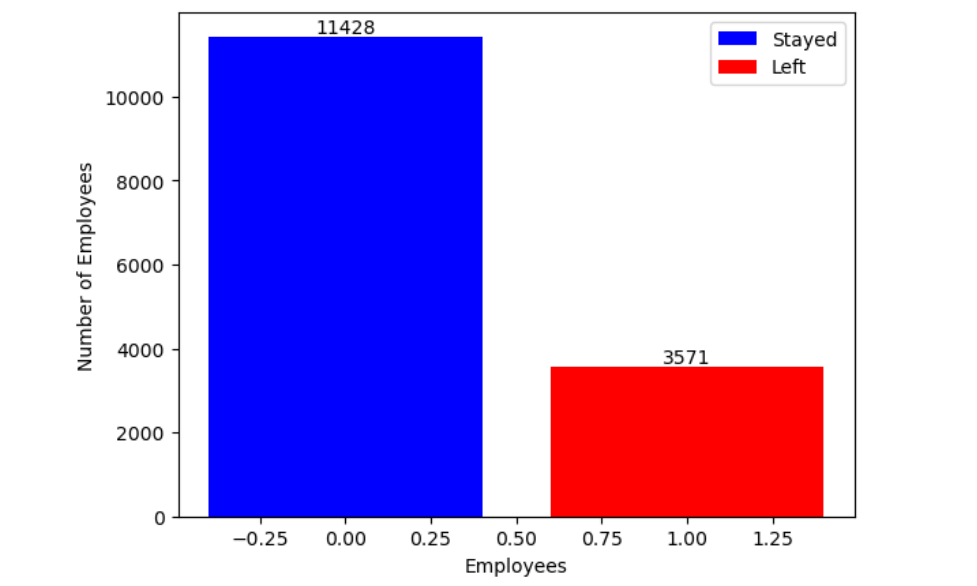
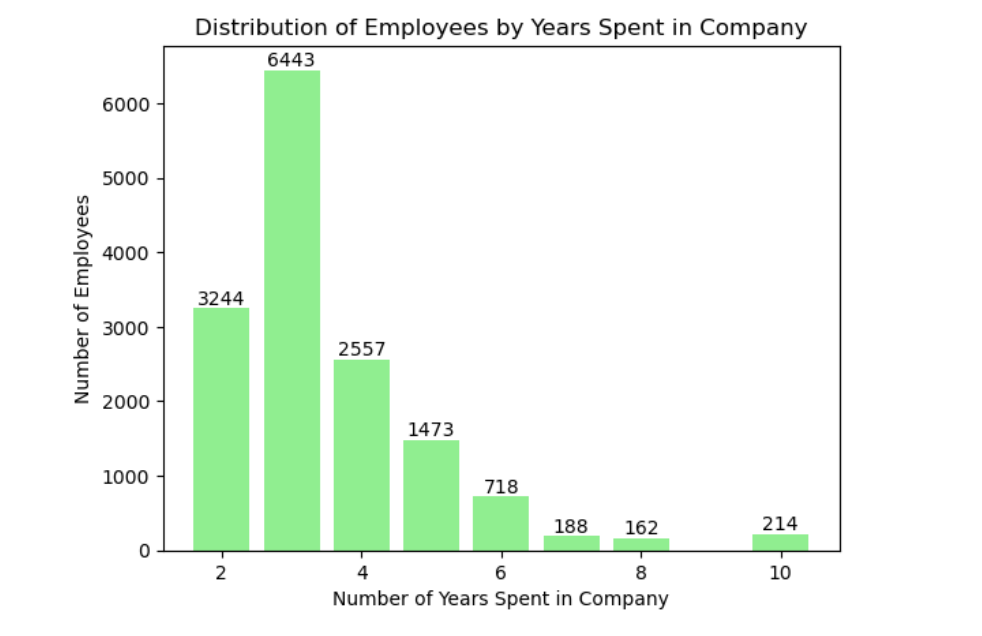
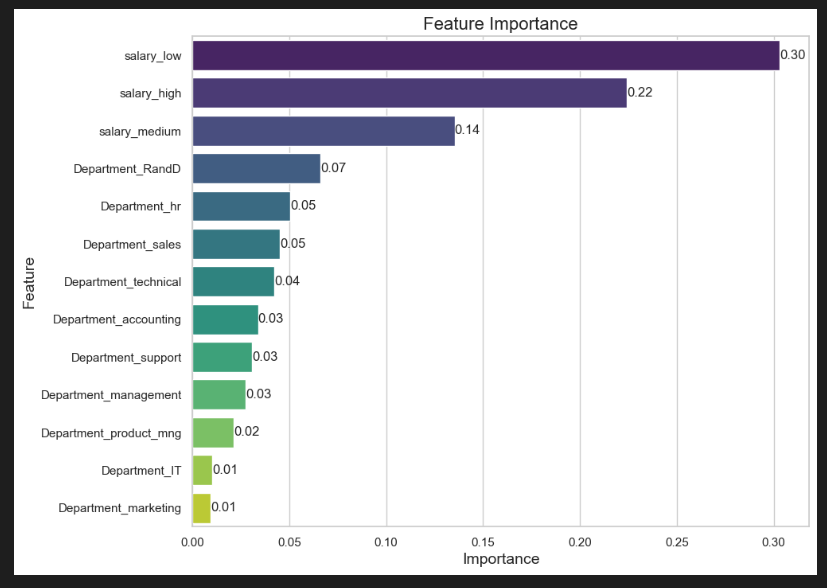
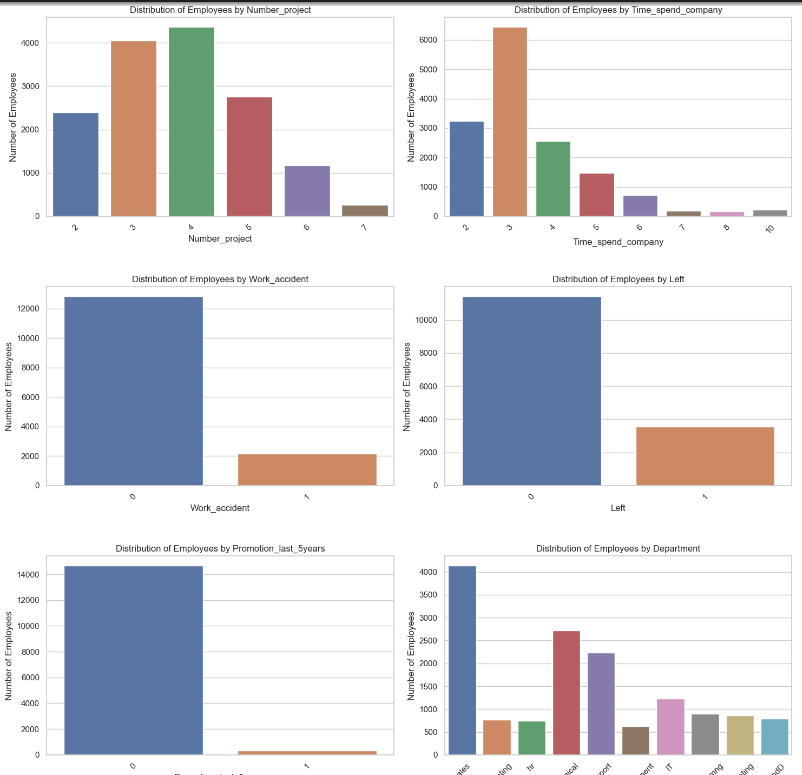
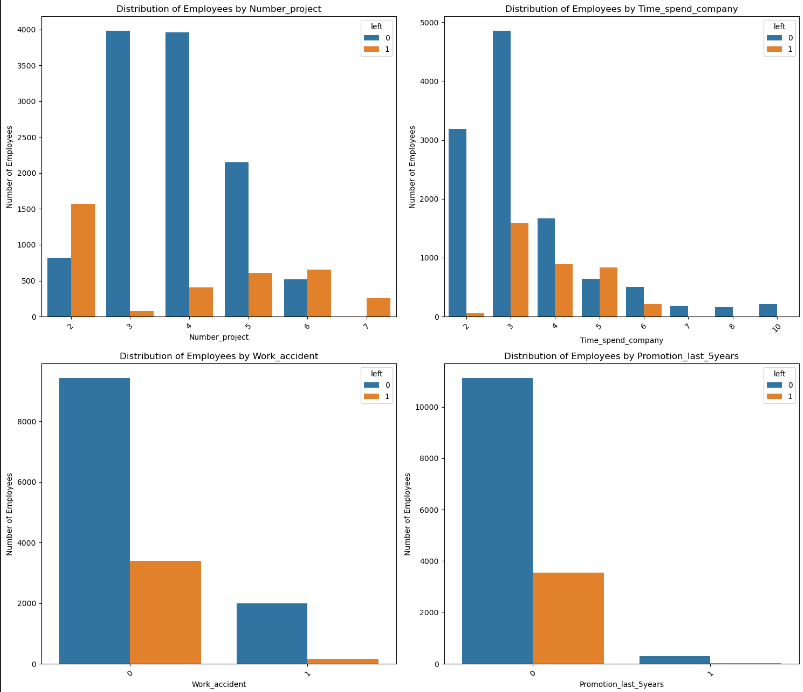
* The dataset contains information about employees, including features such as satisfaction level, last evaluation, number of projects, average monthly hours, time spent in the company, work accident, promotion in the last 5 years, department, and salary.
* It consists of both numerical and categorical variables.
* The target variable is 'left', indicating whether an employee has left the company (1) or not (0).

**Features:**

* Satisfaction Level: Represents the level of satisfaction reported by employees.
* Last Evaluation: Indicates the score of the last performance evaluation.
* Number of Projects: Shows the number of projects an employee is involved in.
* Average Monthly Hours: Represents the average number of working hours per month.
* Time Spent in Company: Indicates the number of years an employee has spent in the company.
* Work Accident: Binary variable indicating whether an employee has had a work accident.
* Promotion Last 5 Years: Binary variable indicating whether an employee has been promoted in the last 5 years.
* Department: Categorical variable representing the department in which the employee works.
* Salary: Categorical variable indicating the salary level of the employee.

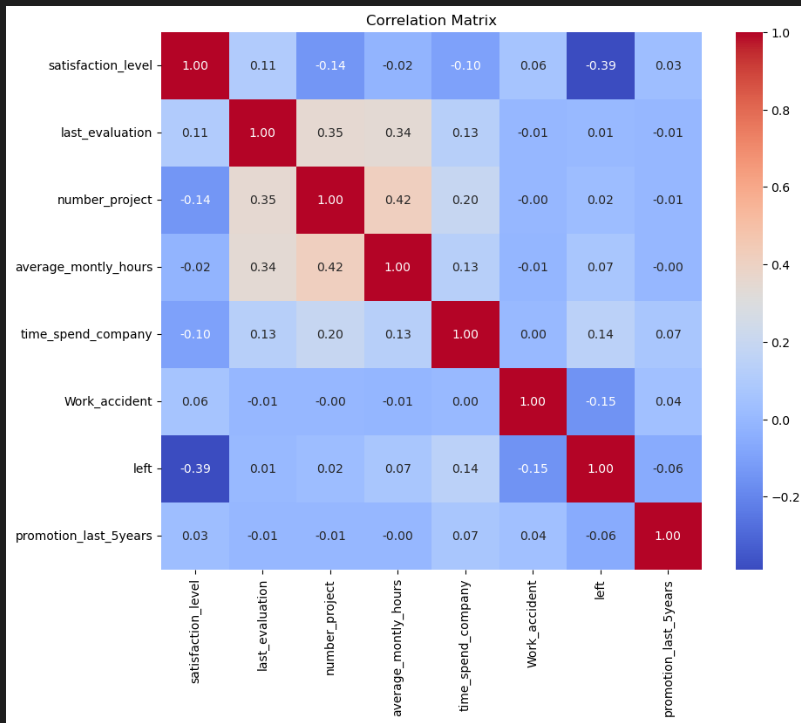


**Data Visualization and Exploratory Data Analysis (EDA)**

* The following visualizations were conducted to explore the dataset:
* Bar plot showing the number of employees who left the company versus those who stayed.
* Bar plot showing the distribution of employees based on the number of projects they are involved in.
* Bar plot showing the distribution of employees based on the number of years spent in the company.
* Count plots for various features, including the number of projects, time spent in the company, work accident, promotion last 5 years, department, and salary, both overall and segmented by employee churn status.
* 
* 
* 
* ****
* ****

**Correlation Matrix and Insights**

* The correlation matrix heatmap reveals insights into the relationships between different features and the target variable:
* The satisfaction level has a negative correlation with employee churn, indicating that lower satisfaction levels are associated with higher employee turnover.
* Other features such as the number of projects and average monthly hours also show correlations with employee churn.**Important Features for Prediction**
* Features such as satisfaction level, last evaluation, number of projects, average monthly hours, and time spent in the company are identified as important for predicting employee churn based on their impact on job satisfaction and work-life balance.

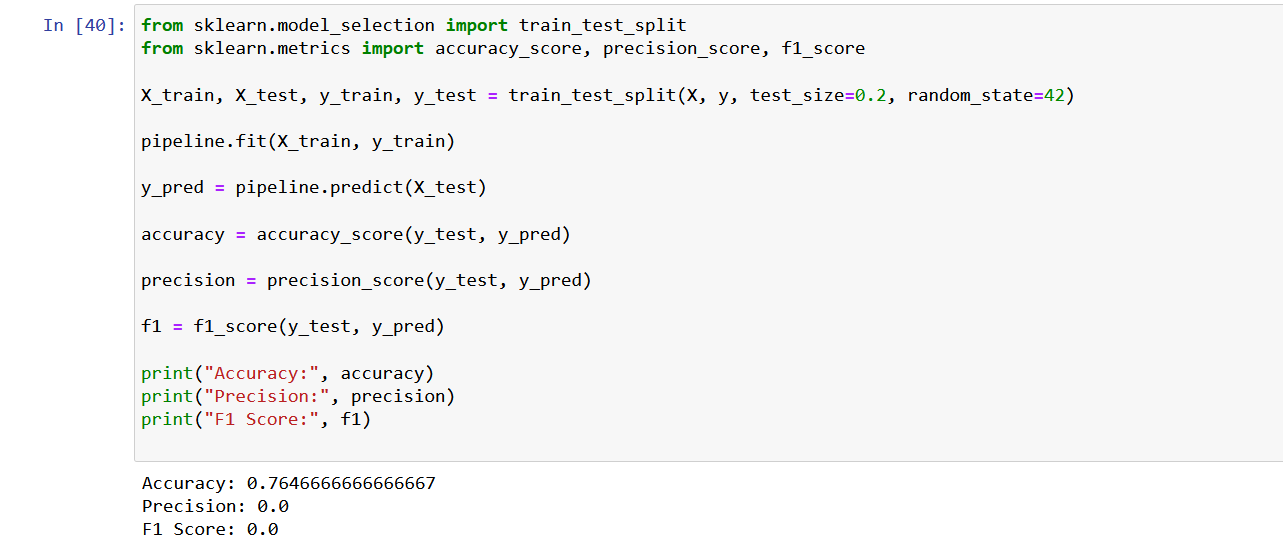


**Important Features for Prediction**

* Features such as satisfaction level, last evaluation, number of projects, average monthly hours, and time spent in the company were identified as important for predicting employee churn based on feature importance analysis.

**Model Trained on the Features**

* A RandomForestClassifier model was trained using the selected features and the dataset.
* The model was trained and evaluated using the training and testing sets, respectively.



**Accuracy and Other Evaluation Metrics**

* The trained model achieved the following evaluation metrics on the testing set:
* Accuracy: [0.7646666666666667]
* Precision: [0.0]
* F1 Score: [0.0]

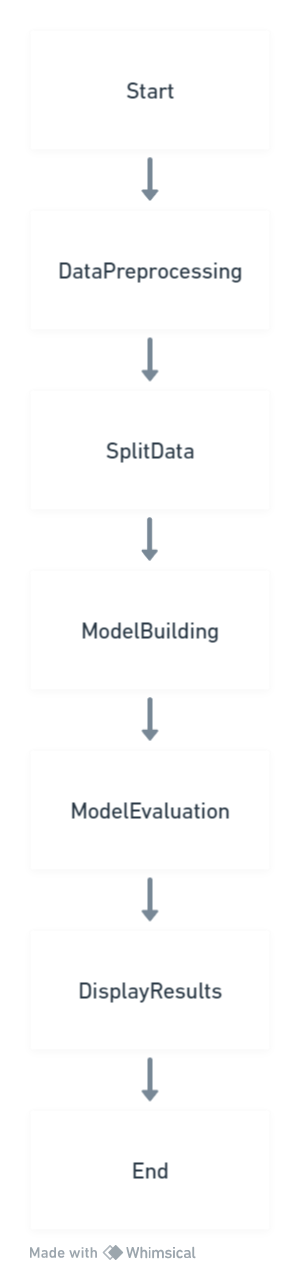
**Objectives**

* The objective of the project is to develop a predictive model to identify employees at risk of leaving the company, enabling proactive intervention to improve employee retention.
* Predictive Model Development: The primary aim is to develop a predictive model.
* Identifying At-Risk Employees: The model should accurately identify employees who are at risk of leaving the company.
* Proactive Intervention: The goal is to enable proactive intervention strategies based on the predictions of the model.
* Improving Retention: By intervening proactively, the objective is to improve employee retention rates within the organization.
* Feature Utilization: Various features such as satisfaction level, number of projects, tenure, work accidents, promotions, department, and salary will be leveraged to achieve the objective.
* Targeted Retention Strategies: The model's predictions will inform the implementation of targeted retention strategies tailored to individual employees.
* Enhancing Employee Satisfaction: Ultimately, the project aims to enhance overall employee satisfaction and organizational stability by reducing turnover rates.

**Methodology**

1. Data preprocessing: Handle missing values, encode categorical variables, and scale numerical features if necessary.
2. Feature selection: Identify important features using correlation analysis and feature importance techniques.
3. Model selection: Choose a suitable machine learning algorithm for classification tasks.
4. Model training: Train the selected model using the training data.
5. Model evaluation: Evaluate the trained model using appropriate evaluation metrics and cross-validation techniques.
6. Model deployment: Deploy the model to production for real-time predictions.

**Diagrams - Flow of the Model**

****

**Conclusion**

In conclusion,the developed predictive model demonstrates promising results in effectively identifying employees who are at risk of leaving the company. By leveraging insights gained from the analysis, the organization can implement targeted strategies to enhance employee satisfaction and retention. These proactive measures will not only mitigate turnover but also contribute to a more stable and productive work environment. Moving forward, continual refinement and adaptation of the model and strategies will be essential to sustaining high levels of employee satisfaction and retention within the organization.