PROJECT SUMMARY:

The project was done with the purpose of finding out factors which affected the Performance of the employees, training a model which accurately predicts the Performance Rating of the employee, analyzing the data to provide recommendations to improve the performance and gain insights from the analysis.

The following steps were carried out:

1. Import the data provided, find out the predictor & target variables and look for missing values.
2. Analysis of Department wise performance as asked.
3. Label Encoding the ordinal columns.
4. Calculate correlation coefficient to find out the relationship between variables and then select the important features for analysis.
5. Standardizing the data and splitting it into test and train.
6. Training the data using algorithms like Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor, Gradient Boosting Classifier and Artificial Neural Network and LSTM Neural network checking the accuracy to find out which algorithm is the best.
7. Exporting the model with highest accuracy.

Summarizing the most important aspects of your model and analysis:

1. The algorithm and training method(s) you used (Such as SVM, Neural Network, etc.)

The different algorithms I used in the project are:

- **Logistic Regression**
- **Random Forest Classifier**
- **SVM**
- **Naive Bayes Classifier**
- **Decision Tree Classifier**
- **Gradient Boosting Classifier**
- **K- Nearest Neighbors**
- **Artificial neural Network**
- **LSTM neural network model**

2. The most important features selected for analysis and why? (Whether techniques such as PCA Factorization used.

- The shape of the Data Set was **1200 rows 28 columns**
- As there were only 28 features, I didn't try PCA. Instead I used feature selection method using Random Forest model, and got feature importance scores. Based on that removed irrevalant features.
- I also removed constant feature such as EmpNumber. I came to that conclusion with domain knowledge.
- The target feature was performanceRating, and the 3 features that affected the target were EmpLastSalaryHikePercent, EmpEnvironmentSatisfaction, and YearsSinceLastPromotion
- The heatmap of the dataset was also plotted to see the correlation between each features.

3. Other techniques and tools used in the project.

- The basic checks have been done.
- Conducted Expolratory Data Analysis using histplot and countplot, Violin plot, barplot
- Used Sweetviz report also for the same.
- Data preprocessing steps:
- Handled missing values adn outliers
- Categorical features where encoded using Manual encoding technique as they are nominal values.
- Scaling is done for logistic regression only, for other models we used non scaled datas.

4. FEATURE SELECTION / ENGINEERING:

1. Did you make any important feature transformations

I didn't had to use feature transformation such as log, square, square root, reciprocal,etc. as it was yeild less model score. I used Standard scaling, but the effect

was same.

2. Correlation or interactions among the features selected and how it is considered?

Correlation between the features were checked using heatmap and feature selection was done using Random Forest model

5. RESULTS, ANALYSIS AND INSIGHTS:

Results:

- Random Forest with GridSearchCV gives 93% accuracy. The features that are positively correlated are Environment Satisfaction, Last Salary Hike Percent & Worklife Balance. This means that if these factors increases, Performance Rating will increase.

- On the other hand, the features that are negatively correlated are Years Since Last Promotion, Experience Years at this Company, Experience years in Current Role & Years with Current Manager. This means that if these factors increases, Performance Rating will go down.

1. Did you find any interesting relationships in the data that don't fit in the sections above?

- Highly experienced females employees have high performance rating with greater than 20% salary hike whereas experienced male employees are able to have great performance rating even with 10% salary hike.

- For the roles Finance manager, Laboratory Technician, Human resources, Manufacturing Director and Seniour Manaer R&D, the performance rating is low. As the experience increases the performances is reducing in the Finance, Human Resources and Research and Development departments. Reason could be not receiving recognition for their work.

2. What is most important technique you used in this project

- For me it would be Data preprocessing. I tried creating a pipeline for numerical and categorical data separately which would do the preprocessing steps, but the encoding was a bit challenging for me there.

3. Provide clear answers to the business problems mentioned in the project on basis of analysis.

Department wise performances :

From the graph we can understand that employees from **Development** department has the highest performance rating, and the **Finance** department has got the lowest performance rating.

Top 3 features effecting employee performances are:

- Employee Environment Satisfaction
- Employee Last Salary Hike Percent
- Years since last promotion

A trained model which can predict the employee performance based on factors as inputs. This will be used to hire employees :

Trained the dataset on multiple models and found that Random forest, Gradient Boosting Forest are best performing generalized model

Recommendation:

- From the results, we can conclude that the company should provide a better environment as it increases the performance drastically.
- The company should increase the salary of the employee from time to time and help them maintain a worklife balance.
- On the other hand, shuffling the manager from time to time will also affect performance.

4. More business insights you gain from the analysis.

- Male employees are more satisfied in the job than female employees
- For female employees, as their experience increases, their average performance decreases whereas for male employees, their average performance increases with experience.
- Highly experienced females employees have high performance rating with greater than 20% salary hike whereas experienced male employees are able to have great performance rating even with 10% salary hike.
- For highly experienced employees, hourly rate is not significantly affecting the performance rating.