**PROJECT REPORT**

(Employee Performance Analysis)

**PROJECT SUMMARY:-**

1. I used Random Forest algorithm to find out employee performance.

2. Random Forest is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of classes or mean prediction of the individual trees.

3. It can be used to rank the importance of variables in a regression or classification problem in a natural way.

4. The algorithm is often used in scientific works because of its advantages.

Ex: - It can be used for quality assessment.

5. Random Forest algorithm applications:-

* Banking
* Medicine
* Stock Market
* E-Commerce

**REQUIREMENT:-**

We require data of employee performance from a company.

* Software:-
* Anacondas with Jupyter Notebook.
* Hardware:-
  + Intel processor with 3 GB or higher RAM.
  + 1 TB HDD.

**DATA PROCESSING:-**

* It refers to the transformations applied to our data before feeding it to the algorithm.
* It is technique used to convert the raw data into a clean data set.
* In the given dataset for Employee Performance, I checked its shape, data types, duplicates values, null values, its full information, etc.
* For achieving better result from the applied model the format of data has to be in a proper manner so the data processing is needed.
* Another aspect is that data set should be formatted in a such a way that more than one Machine Learning algorithms are executed in one dataset, and best out of them is chosen.

**ANALYSIS:-**

1. Exploratory Data Analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods.
2. For Department wise performance using mean, I use ‘group by’ method here.
3. And the top 3 factors which affects the employee performance is determined by correlation method.
4. The terms which affects the employee performance with their correlation are:-

* EmpEnvironmentSatisfaction (0.395561)
* EmpLastSalaryHikePercent (0.333722)
* EmpWorkLifeBalance (0.124429)

1. I also remove the Outliers by checking it through ‘box plot’. The features which contains outliers are:-

* ExperienceYearsAtThisCompany
* YearsSinceLastPromotion
* TotalWorkExperienceYears
* ExperienceYearsCurrentRole

**VISUALIZATION:-**

Visualization is any technique for creating images, diagrams, or animations to communicate a message. Visualization through visual imagery has been an effective way to communicate both abstract and concrete ideas since the dawn of humanity.

For visualisation of employee performance we use ‘matplotlib.pyplot’ it gives a bar chart of employee’s performance of every department.

Visualization is also use in recommendation of employee’s performance.

For Visualization, I use Matplotlib and Seaborn both the packages for better Visual. In which the missing data is shown by the help of ‘Heatmap’ and the target data is shown by the help of ‘Barplot’.

And the Recommendation System is also done through Visualization in which I observed that

In sales department more employee are working as compared to all Department and the rest are shown in the ‘Jupyter File’

**MODEL:-**

Data divides into two parts:

Train model and

Test model

Train model has already trained and based on training model we predict test model.

I have taken 30% Test data and 70% Train data.

**References:-**

* Visualization (To find out highest performance)
* Group By (To find out department wise performance)
* Correlation (find out important factors to improve employee’s performance)
* Random Forest (Algorithm to predict a model of employee’s performance)
* Scaling (To find standardization and normalization)
* Recommendation system (With the help of bar chart graph)

**Amit Singh**

**(amitsingh.hw@gmail.com)**

**From 27th May Batch**

**--------------------------------------------------------------------------------------------**