Assalamualaikum warahmatullahi wabarukatuh, good morning ladies and gentlemen,

thank you for the opportunity that you have given to me. Today we are going to talk about an unique study case. Where there is a data that contain information about employees in a company for certain time. During that time some employees stay and some leave. The employees that leave, already detrimental to the company because it make the company to hire new employee, in worst case the employee were the one who have good performance., the company need to serve new budget for hiring and because there is no employee its distrup the company production So we need to analysis the significant factors that influence the employee and we need to create a model to predict employee leave or stay. This analysis needed to evaluate company environment, human resource management d optimalization of employees performance.

Okay lets start this, first we need to import the library, we do the data loading and simple data exploration. As we can there is no missing, we can say that this data is complete. Next lets go to the exploratory data analysis, okay there are the plot of each variable, from the plot we get some insight

1. in plot education, The most likely to leave the company are those with Masters education

2. The year 2012 was the year with the lowest ratio of leaving and staying employees.

3. In 2018 there was a big spike in employees leaving the company.

4. Employees from Pune city are most likely to leave from the company.

5. Most of the Payment Tier in category 2 left the company, while Payment Tier level 3 did not leave a lot

6. The number of employees is dominated by the age of 24 years to 29 years.

7. The most likely to leave the company are those with female gender

8. Employees who do not work on projects are more likely to resign from the company.

9. Most employees have 2 years of experience in their previous job.

10. There are still more employees who leave than those who leave.

For head map, there are three insight

1. The variable is significantly correlated with itself.

2. the variables that have the highest correlation are paymentTier and join year

3. The highest correlated variables are ExpreiencecurruntDomain and age

Now for feature engineering, okay we already cek for missing value, now we check the outlier,there is no outliers, no we are doing encoding, and scaling

And we want to find the important features, and the results is joining year, age and city

Now for modeling i am using, logistic regression, kneigbors clasiifier, SVC, decision tree clasifier, random forest, gausian nb, and gradient boosting clasifier. As we can see the higest score is reach by gradient boosting and random forest clasifier and decision tree

Now lets improve this model, now we already fine the best parameters for each model, now we implementing the best param model.

1. from the models we can conclude that, gradient boosing clasifier is the best model for this case

2. each model have diffrent time load, but the most take time is svc

3. , the score already the same between base model and improvement

This models have been inference to new data

Conclusion:

1. In 2018 there was a big spike in employees leaving the company.

2. Employees from Pune city are most likely to leave from the company.

3. Most of the Payment Tier in category 2 left the company, while Payment Tier level 3 did not leave a lot

4. The most likely to leave the company are those with female gender

5. Employees who do not work on projects are more likely to resign from the company.

6. the company need an anticipation for employees leaving spike, it can be evaluated on company services for employees, providing work that is in according with the salary will keep employees contribute to the company.

7. The model that is suitable for this data is the Gradient Boosting Classifier model with the base line model and the improvement model having an accuracy of 0.84

8. TThe model is able to predict new data

For deployment

We save the model with pickle and lets see the application