

Approach

1. First I imported the necessary libraries.
2. Read and viewed the data.
3. Checked for missing values.
4. Did variable Typecasting.
5. Calculated Total business value generated by an employee by using groupby empid and used it as a feature.
6. Kept only 1 entry per employee which is the last reporting date.
7. Generated features like Emp_Promotion whether the employee got any promotion, Emp_Experience overall experience of the employee once they have joined the company, Attrition whether the employee has left the company or not based on last working date and used it as a target variable.
8. Dropped the variables which are of no use.
9. Encoded the categorical variables.
10. Merged the test dataset with train based on empid.
11. Build a decision tree model.
12. Did predictions and computed the f1 score.
13. Did hyperparameter tuning.
14. Again build a new decision tree model with new hyperparameter values.
15. Did predictions and computed the f1 score for the new decision tree model.
16. At last, converted the test data frame to a CSV file.