PROJECT WRITEUP: “EMPLOYEE TURNOVER ANALYTICS”

DESCRIPTION:

Objective:

The objective is to develop and deploy machine learning models that can effectively predict and understand employee turnover, enabling the HR Department to take proactive measures to retain valuable employees. The project aims to provide actionable insights and recommendations for improving employee retention within Portobello Tech.

Problem Statement:

Portobello Tech is an app innovator that has devised an intelligent way of predicting employee turnover within the company. It periodically evaluates employees' work details including the number of projects they worked upon, average monthly working hours, time spent in the company, promotions in the last 5 years, and salary level.

Data from prior evaluations show the employee’s satisfaction at the workplace. The data could be used to identify patterns in work style and their interest to continue to work in the company.

The HR Department owns the data and uses it to predict employee turnover. Employee turnover refers to the total number of workers who leave a company over a certain time period.

Domain: As the ML Developer assigned to the HR Department, you have been asked to create ML Programs.

Analysis to be done: The problem is to identify the apps that are going to be good for Google to promote. App ratings, which are provided by the customers, is always a great indicator of the goodness of the app. The problem reduces to predict which apps will have high ratings.

Content: Dataset: (“hr_comma_sep.csv” : https://www.kaggle.com/datasets/liujiaqi/hr-comma-sepcsv)

Fields in the data –

Column Name	Description
Satisfaction_level	Satisfaction level at the job of an employee
last_evaluation	Ratings b/w 0 to 1, received by an employee at his last evaluation
number_project	Number of projects, an employee involved in
average_monthly_hours	Average number of hours in a month, spent by an employee at office
time_spend_company	Number of years spent in the company
Work_accident	0 – no accident during employee stay, 1 – accident during employee stay
Left	0 – indicates employee stays in the company 1 – indicates employee left the company
promotion_last_5years	Number of promotions in his stay
Department	Department, an employee belongs to
Salary	Salary in USD

PROJECT WRITEUP: “EMPLOYEE TURNOVER ANALYTICS”

Steps to perform:

1. Perform data quality check by checking for missing values if any.
2. Understand what factors contributed most to employee turnover by EDA.
3. Perform clustering of Employees who left based on their satisfaction and evaluation.
4. Handle the left Class Imbalance using SMOTE technique.
    a. Pre-Process the data by converting categorical columns to numerical columns.
    b. Do the split of the dataset to train and test in the ratio 80:20 with random_state=123.
    c. Up sample the train dataset using SMOTE technique from the imblearn module.
5. Perform k-fold cross-validation model training and evaluate performance.
6. Identify the best model and justify the evaluation metrics used.
7. Suggest various retention strategies for targeted employees.

Challenges and Limitations:

1. **Data Quality and Missing Values:**
    - **Challenge:** The presence of missing values in the dataset could impact the quality of the predictions. Imputing these missing values is essential for meaningful analysis.
    - **Limitation:** Imputing missing values might introduce bias if not handled carefully.

2. **Class Imbalance:**
    - **Challenge:** Class imbalance (more employees staying than leaving) affects model training, making it biased towards the majority class.
    - **Limitation:** SMOTE technique may introduce synthetic data, and its effectiveness depends on the nature of the original data distribution.

3. **Model Selection and Evaluation Metrics:**
    - **Challenge:** Choosing the right machine learning models and evaluation metrics is critical for accurate predictions.
    - **Limitation:** The selected models may not capture the complex relationships within the data, leading to suboptimal results.

4. **Feature Selection and Relevance:**
    - **Challenge:** Identifying the most relevant features contributing to turnover can be challenging.
    - **Limitation:** Irrelevant features or excluding important ones may result in a less effective model.

5. **Cross-Validation and Generalization:**
    - **Challenge:** Overfitting or underfitting during model training may impact the generalization of the model to new data.
    - **Limitation:** Cross-validation may not always represent the model's performance on unseen data accurately.

6. **Retention Strategy Effectiveness:**
    - **Challenge:** The effectiveness of retention strategies may depend on factors beyond the scope of the dataset, such as external market conditions.

PROJECT WRITEUP: "EMPLOYEE TURNOVER ANALYTICS"

- **Limitation:** Retention strategies derived solely from the dataset may not consider individual employee preferences or external factors influencing turnover.

7. **Communication of Results:**
   - **Challenge:** Communicating complex machine learning results to non-technical stakeholders, such as HR professionals, can be challenging.
   - **Limitation:** Misinterpretation of results may lead to ineffective decision-making.

Findings and Recommendations:

1) Since majority percentage of employee's turnover was from 'HR' deptt, therefore we need to bring in more employee friendly schemes particularly in HR deptt to reduce the employee's turnover.

2) It was also observed that almost 30% of the total employees who left was because of the low salary, therefore we need to increase the salaries of the employees as per current industry standards and can also implement incentive and rewards scheme for the employees.

3) It was also observed in our analysis that employees who has worked on too many projects or too less projects tends to leave the company whereas employees who have worked on avg number of projects tends to stay, therefore we need to balance out the of project distribution among the employees, we need to check that some employees shouldn't get burdened with too many projects and also it should not be the case that some employees are having very less project and their time and skills are getting wasted.

4) It was also observed that senior employee who have worked for more than 7 years in the company not to leave the company, whereas junior or mid-level employees tends to leave the company, therefore we need to create policies and schemes (like for senior level employees) which makes junior and mid-level employees of the company to remain stick to the company.

5) Also, it was observed that those employees tend to leave who got either very low evaluation score or very high evaluation score from the company, we need to bring in more clarity and make evaluation measures more employee friendly so that employees don't consider the evaluation too seriously and always strive to improve their scores.

6) I was also observed that employees leave when they are overworked or underworked, therefore we need to systematically plan and distribute the work among the employees to properly handle their workload management.

7) Also, it was observed that out of total of '14999' employees, only '319' employees got promotion in the last 5 years, therefore we need to reward more employees more frequently to retain the target employees.

Thank you!