**PROJECT (Code: 10281): Employee Performance Analysis (INX Future Inc.)**

About Project

INX Future Inc., (referred as INX), is one of the leading data analytics and automation solutions provider with over 15 years of global business presence. INX is consistently rated as top 20 best employers past 5 years. INX human resource policies are considered as employee friendly and widely perceived as best practices in the industry. Recent years, the employee performance indexes are not healthy and this is becoming a growing concerns among the top management. There has been increased escalations on service delivery and client satisfaction levels came down by 8 percentage points. CEO, Mr. Brain, knows the issues but concerned to take any actions in penalizing non-performing employees as this would affect the employee morale of all the employees in general and may further reduce the performance. Also, the market perception best employer and thereby attracting best talents to join the company. Mr. Brain decided to initiate a data science project, which analyses the current employee data and find the core underlying causes of this performance issues. Mr. Brain, being a data scientist himself, expects the findings of this project will help him to take right course of actions. He also expects a clear indicators of non performing employees, so that any penalization of non-performing employee, if required, may not significantly affect other employee morals.

Requirement / Goal

The following insights are expected from this project.

1. Department wise performances

2. Top 3 Important Factors effecting employee performance

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

4. Recommendations to improve the employee performance based on insights from analysis.

Project Summary

Analysis

* **Data Collection**: Raw dataset was provided by INX Future Inc., which contains 1200 rows and 28 columns, with addition information to deal with the dataset and also some business problems statement which needs to be solved with the help of this project.
* **Data visualization**: Various libraries like Numpy, Pandas, Pandas Profiling, Matplotlib and Seaborn have been used to extract more information from each variables of dataset.
* **Data Processing**: Sklearn library was majorly used for Data preprocessing. Seaborn library was used for plotting outliers. One-hot encoding method was applied to encode categorical variable since all the categorical variable were nominal data. Feature Selection techniques like Correlation Coefficient, Mutual Information Gain Method, PCA, feature importance using xgboost and Random forest were used but Correlation Coefficient was more suitable for the dataset and better accuracy score.

Modelling

* Libraries such as Sklearn, Imblearn, xgboost were mostly used while modelling process. Splitting of data was done with train\_test\_split method. To balance the dataset SMOTE (Oversampling) Technique was used for better prediction.
* Hyperparameter tuning with GridSerch CV has been used to get the best possible parameters for the model to perform well.
* Model like Random Forest, XGBoost and Decision are used for modelling which was having better accuracy score compared to SVC, KNN, ANN, etc
* For evaluation, methods like confusion matrix, accuracy score, classification report have been used for all the models.

Features Selection / Engineering

* The most important features selected for analysis were EmpEnvironmentSatisfaction, EmpDepartment, EmpJobLevel, EmpLastSalaryHikePercent, TotalWorkExperienceInYears, EmpJobRole, and YearsSinceLastPromotion because these features were more affecting the performance rating, So that I could collect more of important insights from the data.
* Feature Transformation was not necessary for this use case according to my knowledge.
* EmpEnvironmentSatisfaction and EmpLastSalaryHikePercent were highly correlated with Performance Rating variable, So it is considered that if values of these variable increases Performance Rating will increase.

Insights and Results

Insights:

* **Department wise Performance and Insights**:

- There are more employees in Sales, Development and Research & Development Department.

- Sales and Research & Development Departments have high percentage of low performance rating.

- Development Department has high percentage of 3 and 4 performance rating compared to low performance rating.

- On the other hand Development, Sales and Research & Development Departments also have most number of employees with average performance rating.

* **Employment Environment Satisfaction, Employee Last Salary Hike Percent and Emp Work Life Balance the three core underlying causes of performance issues among the employees**.
* Low Performance rating is majorly seen in Sales, Research & Development and Finance Departments, specifically in Job roles like Sales Executive, Research Scientist and Manager R&D.
* Another major factors affecting the Performance Rating are Experience Years at this Company, Experience years in current role, Years Since last promotion, and years with current manager.

Results:

* **Random Forest Classifier (Accuracy Score: 96%):**

Model performed really well compared to other models. F1 score of 93%, 97% and 87% for (2, 3, 4) respectively. This model can be used to solve business problems and also can be helpful for company to understand the factors affecting Employee Performance Rating. **It also can predict the employee performance based on factors as inputs and can be used to hire employees based on that.**

* **XGBoost Classifier (Accuracy Score: 95%):**

This model also performed well with F1 score 93%, 96% and 85% respective of the target categories.

* **Decision Tree Classifier (Accuracy Score: 94%):**

This model also performed well with F1 score 90%, 96% and 86% respective of the target categories.

Recommendations

* Employees with low Performance Rating should be taking very care of since they can ruin the working atmosphere of the other employees.
* Interactive and encouraging training sections should be arranges every 6months or a year to encourage those kind of employee whose performance are going low.
* Some positive changes should be made in the working environment for employees.
* Sales and Research & Development department should follow strategies or policies followed by Development department as there are high percentage of Good Performance Rating and low percentage of poor performance rating.
* Salary Hike Policies should be changed as per convenience of company as well as employees.
* Years of salary hike should be reduced and hike percentage should be increased for employee satisfaction.
* Employee issues should be taken into consideration so that they can work peacefully and improve their performance.
* Flexibility in working hours should be granted to balance the work life to improve performance in work.
* The model will help increase the market perception of best employer and thereby attracting best talents to join the company.