**Business Case and Domain Analysis**

**Project Overview**

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 concern 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.

**Problem Statement**

The objective of this project is to create a predictive model that assists the which can predict the employee performance based on factors as inputs. This will be used to hire employees. The project aims to leverage predictive analytics to enhance the understanding of important Factors effecting employee performance and create department wise performances analysis and recommendations to improve the employee performance based on insights from analysis.

The organization faces challenges in accurately assessing employee performance, leading to inefficiencies in hiring processes and suboptimal workforce productivity. To address these challenges, the project aims to leverage data analysis and predictive modeling techniques to gain insights into factors influencing employee performance and develop solutions for more effective performance management and hiring.

The project will focus on four key objectives:

1. **Department-wise Performances Analysis:** Analysing the performance of different departments within the organization to identify strengths, weaknesses, and areas for improvement.
2. **Identification of Top 3 Important Factors Affecting Employee Performance:** Identifying the key factors that significantly influence employee performance through data analysis and statistical modeling.
3. **Development of a Trained Predictive Model:** Building a predictive model that can forecast employee performance based on relevant input factors. This model will be utilized for hiring purposes to select candidates who are likely to perform well in their roles.
4. **Recommendations for Performance Improvement:** Providing actionable recommendations to enhance employee performance based on insights gained from the analysis and predictive modeling.

**Analysis Methodology**

Here the which employee performance is being decrease by comparing columns.

* Employee between the age group 30-35 is the majority.
* 70% of the people travel rarely, 20% travel frequently rest do not travel.
* Majority of the employees belong to research and development.
* Almost 35% of the people are nearer to the office i.e. the distance from their home is lesser than or equal to 10.
* More than 60% of the people have educational qualification of 2 and 4.
* Majority (40%) of the people are from life science field and 30% are from medical field.
* 60% of the people are almost satisfied with environment condition of the office with more than 3 and 4 ratings.
* Gender count: 60% male 40% female.
* Almost 40% of the people have partial involvement in job and 20% have good involvement.
* More than 45% employees seem to be satisfied with their job.
* 50% of the people are married, 30% single and the rest are divorced.
* 40% of the employee have worked experience less than 10 year
* 15% of the people have worked for less than 1 company which implies they are freshers.
* 30% of the people have worked for more than 5 companies.
* 80% of the people have average work rating.
* People who travel more are more expected to leave the job.
* People who do not do overtime do not leave the job.
* Singles are expected to quit the job.
* People from Development, Senior developer and sales executive are more probably leaving their job.

**Recommendation**

All three models achieved very similar high validation accuracy scores, indicating strong predictive power. However, the choice of the model may depend on other factors, including resource constraints and model interpretability. Random Forest: Recommended if computational resources are available. It offers robust performance and can handle complex data. XGBoost: A strong alternative to Random Forest, efficient and highly accurate, but may require fine-tuning. Logistic Regression: Suitable as a simple, interpretable baseline model.

Establish measurable goals and milestones to track progress.

Clearly define performance expectations and objectives for the employee.

Collaborate with the employee to develop a personalized action plan for development.

When discussing performance improvement with an employee, it's crucial to maintain a respectful and constructive dialogue. Focus on providing support, guidance, and opportunities for growth to help the employee reach their full potential. Because

EmployeeTravel : The workers who travel alot are more likely to quit then other employees.

Department : The worker in Research & manager are more likely to stay then the workers on other departement.

EducationField : The workers with Human Resources and Technical Degree are more likely to quit then employees from other fields of educations.

Gender : The Male are more likely to quit.

JobRole : The workers in Developer, Human Resources, Senior Develo are more likely to quit the workers in other positions.

MaritalStatus : The workers who have Single marital status are more likely to quit the Married, and Divorced.

OverTime :Over time rate is almost equal

**Predictive Model:**

This model will be a useful tool for making decisions and forecasts in real-time for Recommendations for performance improvement. And a trained model that can predict employee performance based on factors as inputs. This will be used to hire employees

Conduct regular performance reviews to assess progress and identify areas for improvement. Encourage open communication and address any concerns or challenges proactively. Offer constructive feedback on strengths and areas needing improvement.

**Model Performance Evaluation :**

**Decision Tree:**

Validation Accuracy: 1%

Strengths: It produces accurate and interpretable models with relatively little user intervention. It is fast, both at build and apply time, and supports parallel execution.

Considering: Additional processing power might be needed.

**Random Forest:**

Validation Accuracy: 90.55%

Strengths: Highly accurate and powerful, highly resistant to overfitting, and can estimate missing data.

Considering: Additional processing power might be needed.

**Support Vector Machine:**

Validation Accuracy: 72.50%

Strengths: Effective in High-Dimensional spaces, effective in spaces where number of dimensions is greater than the number of spaces and is memory-efficient.

Considering: Additional processing power might be needed.

**Gradient boosting**

Validation Accuracy: 94.16

Strengths: Training speed is fast, especially on large datasets, generally more accurate compared to other modules, and it can handle categorical features without encoding.

Considering: Additional processing power might be needed.

**K\_Nearest Neighbors:**

Validation Accuracy: 72.50%

Strengths: Simplicity in implementation, it makes prediction based on the nearest neighbours in the dataset, allowing it to handle complex decision boundaries.

Considering: Additional processing power might be needed.

**Artificial Nueral Network (ANN):**

Validation Accuracy: 72.22%

Strengths: Capable of handling high-dimensional data, strong against overfitting, and high accuracy.

Considering: Additional processing power might be needed.