BUSINESS CASE

INX Future Inc Employee Performance

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 the clear indicators of non-performing employees, so that any penalization of non-performing employee, if required, may not significantly affect other employee morals.

PROJECT OVERVIEW:

The goal of this project is to analyse employee performance within the organization and develop predictive models to enhance hiring processes and improve overall performance.

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.

Problem Statement:

- The organization faces challenges in accurately assessing employee performance, leading to inefficiencies in hiring processes and suboptimal workforce productivity.

- To address these challenges, the aim is to develop a data science project which identifies the root causes of declining employee performance indices at INX Future Inc. and provide clear indicators for non-performing employees without negatively impacting overall employee morale or the company's reputation as a top employer.

DOMAIN ANALYSIS:

1) EmpNumber: Unique identifier for each employee in the dataset.

2) Age: Age of the employee, providing insight into workforce demographics and potential correlations with attrition.

3) Gender: Gender of the employee, which may impact workplace dynamics and attrition patterns.

4) EducationBackground: The educational background of the employee, influencing skillset and career trajectory.

5) MaritalStatus: Marital status of the employee, potentially affecting work-life balance and job satisfaction.

6) EmpDepartment: Department in which the employee works, indicating job role and organizational structure.

7) EmpJobRole: Specific job role of the employee within their department, reflecting responsibilities and career path.

8) BusinessTravelFrequency: Frequency of business travel for the employee, impacting lifestyle and job satisfaction.

9) DistanceFromHome: Distance of employee's residence from the workplace, influencing commuting stress and retention.

10) EmpEducationLevel: Level of education attained by the employee, reflecting qualifications and potential for advancement.

11) EmpEnvironmentSatisfaction: Employee satisfaction with the work environment, affecting morale and turnover.

12) EmpHourlyRate: Hourly wage of the employee, a factor in compensation satisfaction and retention.

13) EmpJobInvolvement: Level of involvement and engagement in the job role, affecting performance and attrition risk.

14) EmpJobLevel: Level of hierarchy within the organization, indicating seniority and career progression.

15) EmpJobSatisfaction: Satisfaction level with the job role, impacting employee morale and retention.

16) NumCompaniesWorked: Number of companies the employee has previously worked for, indicating job stability and turnover risk.

17) OverTime: Whether the employee works overtime, influencing work-life balance and burnout.

18) EmpLastSalaryHikePercent: Percentage of the employee's last salary hike, affecting compensation satisfaction and retention.

19) EmpRelationshipSatisfaction: Satisfaction with relationships at work, influencing job satisfaction and likelihood of turnover.

20) TotalWorkExperienceInYears: Total work experience of the employee, influencing skill level and career trajectory.

21) TrainingTimesLastYear: Number of training sessions attended by the employee last year, indicating investment in skill development and career growth.

22) EmpWorkLifeBalance: Employee's perceived balance between work and personal life, affecting job satisfaction and retention.

23) ExperienceYearsAtThisCompany: Years of experience at the current company, indicating loyalty and potential for promotion.

24) ExperienceYearsInCurrentRole: Years of experience in the current job role, influencing expertise and potential for advancement.

25) YearsSinceLastPromotion: Time since the employee's last promotion, impacting career progression and job satisfaction.

26) YearsWithCurrManager: Years of tenure with the current manager, affecting job satisfaction and retention.

27) Attrition: indicates whether the employee has left the company or not.

28) PerformanceRating: Target variable for the given problem. this is the performance rating assigned to the employee, influencing career development and potential for retention.

ANALYSIS METHODOLOGY

Both visual analysis and statistical methods are carried out for data exploration.

Visual analysis methods:

- Univariate analysis
- Bivariate analysis
- Multivariate analysis

Statistical methods used:

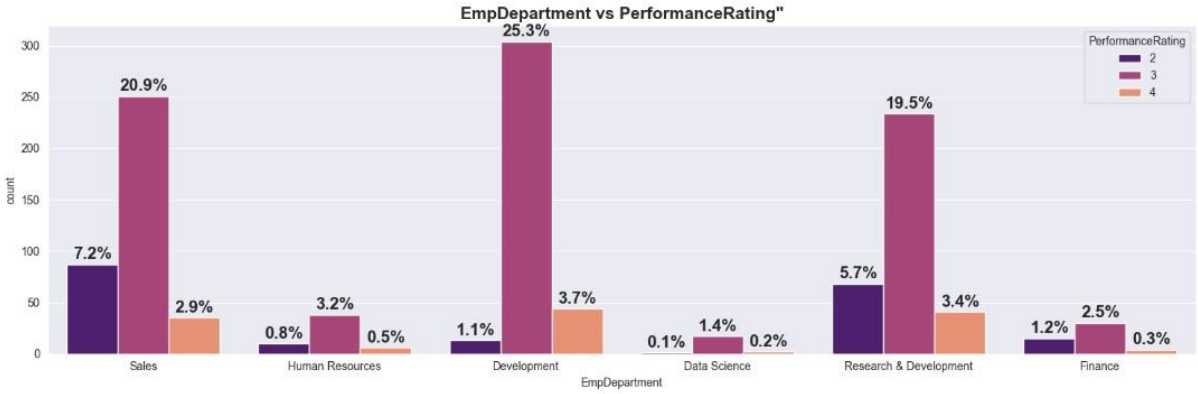
- Correlation coefficient
- ANOVA test
- Chi-square test

Insights driven from the analysis:

- Employees between the age group of 30-35 are the majority in number.
- 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 work experience of less than 10 years.
- 15% of the people have worked for less than 1 company which implies that 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 less likely to leave the job.
- People from job roles of sales executive, developer, sales representative, laboratory technician, research scientist and manager of R&D are more likely to leave the job.

Department wise performance analysis:

The analysis assesses the performance of employees across different departments to identify any disparities or trends.



Insights:

- Development department dominates the other departments in number.
- In terms of performance rating, employees who belong to development department perform well with the dominating 25% of rating 2.
- Hence departments such as **development, sales and Research & development** constitute the high performing employees.

Top 3 factors affecting the employee's performance:

From the data analysis and the feature engineering, there are some of the factors which proved crucial in predicting the employee's performance.

Among them the top three factors are as factors:

1. EmpEnvironmentSatisfaction
2. EmpLastSalaryHikePercent
3. EmpDepartment_Development

Insights:

- The **work environment satisfaction** has positive relationship with the target. As the work environment satisfaction increases, the performance rating of the employees also increases.
- The feature "**EmpLastSalaryHikePercent**" holds negative relationship with the target. The performance rating is quite high consistently for a fixed range of salary hike percent of 11-14%. After this range, the performance decreases. This could

attribute to other factors such as various departments and their working strength, limited top level positions etc.

- The **employee's department of development** shows high performance rating compared to other departments.

Algorithms used in this project:

- Decision Tree classifier
- Random Forest classifier
- Support Vector Machine
- Extreme Gradient Boosting

Trained models:

1. **DECISION TREE CLASSIFIER**
2. **TUNED DECISION TREE CLASSIFIER**
3. **RANDOM FOREST CLASSIFIER**
4. **TUNED RANDOM FOREST CLASSIFIER**
5. **SUPPORT VECTOR MACHINE**
6. **TUNED SUPPORT VECTOR CLASSFIER**
7. **TNED XG BOOSTING**

Model results:

	Base Decision Tree	Tuned Decision Tree	Base Random Forest	Tuned Random Forest	Base SVM	Tuned SVM	Tuned xgboost
Accuracy	0.912500	0.908333	0.937500	0.933333	0.766667	0.866667	0.933333
precision	0.866592	0.843636	0.936924	0.922979	0.663977	0.797225	0.928555
Recall	0.869841	0.887863	0.876996	0.875092	0.803175	0.781026	0.875092
F1-score	0.868072	0.864145	0.904328	0.897400	0.704754	0.788676	0.899492

- Best **Accuracy** 👉 **Base Random Forest**
- Best **Precision** 👉**Base Random Forest**
- Best **Recall** 👉**Tuned Decision Tree**
- Best **F1-score** 👉 **Base Random Forest**

The order of the ✅ Best Models from top to bottom:

	Model (Best from 1 to 7)
1	Base Random Forest ✅
2	Tuned xgboost
3	Tuned Random Forest
4	Base Decision Tree
5	Tuned Decision Tree
6	Tuned SVM
7	Base SVM

RECOMMENDATION:

Model recommendation:

- The top three would be
  - Base Random forest model
  - Tuned Random Forest model
  - Tuned XG Boost model
- These three models achieved very similar high accuracy scores of 93% 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.
- **XG Boost:** A strong alternative to Random Forest, efficient and highly accurate, but may require fine-tuning.

Recommendations for business problem:

1. Timely incentives and recognition: Since the **employees' salary hike percent** directly influences the performance, it is essential to track the progress of employees and their work is given due recognition and reasonable hike.

2. Open communication: Since the **work environment** proved to be crucial in the performance ratings of employees, addressing any concerns of employees in a proactive and empathetic way could be helpful in improving the performance and efficiency.

3. Fostering supportive and inclusive workplace culture: Team work and collaboration should be encouraged for **department-wise better performance** of all job roles.

4. Regular performance reviews: to update the employees' progress as well as to **motivate** them to improve in challenging areas.

5. Training programs: Relevant **skill development initiatives** in a regular time interval could be helpful to help them navigate this new-tech boom of AI.

CONCLUSION:

The rational yet empathetic initiatives are essential to improve the employees' performance. The above recommendations have to be implemented in a systematic and planned manner so that proper results can be expected in a regular timeframe. The top to bottom hierarchy can also have the possibility of affecting the morale of junior level employees and hence HR collaborations with employees will be helpful in overcoming this issue in a decentralized yet effective manner.