1. **Introduction:**

The dataset comprises employee-level data from INX Future Inc., covering demographic details, job roles, departmental affiliations, education, experience, and satisfaction metrics. It includes the target variable Performance Rating, enabling analysis of employee performance patterns and predictive modelling to identify performance influencers across various roles and departments. INX Future Inc. faces declining employee performance and client satisfaction. To address this, the objective is to analyse employee data, identify performance drivers, and build a predictive model for employee performance. The aim is to support strategic decisions for hiring, departmental interventions, and morale-friendly corrective actions.

**2. Dataset Overview:**

1. The dataset includes 28 explanatory variables describing different aspects of employee performance.
2. Performance Rating is the target variable, which represents the performance of the employees.

**Attribute Information:**

* + - 1. Emp Number – Unique identifier assigned to each employee.
      2. Age – Employee’s age in years.
      3. Gender – Gender of the employee (e.g., Male or Female).
      4. Education Background – Field of study or educational stream.
      5. Marital Status – Marital status of the employee (e.g., Single, Married, Divorced).
      6. Emp Department – Department in which the employee works.
      7. Emp Job Role – Specific job position or role of the employee.
      8. Business Travel Frequency – How frequently the employee travels for work.
      9. Distance From Home – Distance (in km) from home to workplace.
      10. Emp Education Level – Highest level of education (1=Low, 5=High).
      11. Emp Environment Satisfaction – Satisfaction with workplace environment (scale of 1 to 4).
      12. Emp Hourly Rate – Hourly wage of the employee.
      13. Emp Job Involvement – Level of involvement or engagement in their job (1 to 4).
      14. Emp Job Level – Hierarchical level or grade of the employee in the organization.
      15. Emp Job Satisfaction – Satisfaction with their job role (scale of 1 to 4).
      16. Num Companies Worked – Number of companies the employee has worked for.
      17. Over Time – Indicates if the employee works overtime (Yes/No).
      18. Emp Last Salary Hike Percent – Percentage increase in last salary hike.
      19. Emp Relationship Satisfaction – Satisfaction with relationships at work (1 to 4 scale).
      20. Total Work Experience in Years – Total professional experience in years.
      21. Training Times Last Year – Number of training programs attended last year.
      22. Emp Work Life Balance – Balance between personal and work life (1 to 4 scale).
      23. Experience Years at This Company – Total years spent at the current company.
      24. Experience Years in Current Role – Number of years in the current role.
      25. Years Since Last Promotion – Time elapsed since the last promotion.
      26. Years With Curr Manager – Number of years with the current manager.
      27. Attrition – Indicates if the employee has left the company (Yes/No).
      28. Performance Rating – Overall performance rating of the employee (target variable).

**3. Data** **Summary**:

Number of records: 1200

Number of features: 28

**Data Types:**

Numerical Features - 19

Categorical Features - 9

**Null Values:**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Null Values** | **Type** |
| Emp Num | 0 | Categorical |
| Age | 0 | Numerical |
| Gender | 0 | Categorical |
| Education Background | 0 | Categorical |
| Marital Status | 0 | Categorical |
| Emp Department | 0 | Categorical |
| Emp Job Role | 0 | Categorical |
| Business Travel Frequency | 0 | Categorical |
| Distance From Home | 0 | Numerical |
| Emp Education Level | 0 | Numerical |
| Emp Environment Satisfaction | 0 | Numerical |
| Emp Hourly Rate | 0 | Numerical |
| Emp Job Involvement Level | 0 | Numerical |
| Emp Job Level | 0 | Numerical |
| Emp Job Satisfaction | 0 | Numerical |
| Num Companies Worked | 0 | Numerical |
| Over Time | 0 | Categorical |
| Emp Last Salary Hike Percent | 0 | Numerical |
| Emp Relationship Satisfaction | 0 | Numerical |
| Total Work Experience in Years | 0 | Numerical |
| Training Times Last Year | 0 | Numerical |
| Emp Work Life Balance | 0 | Numerical |
| Experience Years at This Company | 0 | Numerical |
| Experience Years in Current Role | 0 | Numerical |
| Years Since Last Promotion | 0 | Numerical |
| Years With Curr Manager | 0 | Numerical |
| Attrition | 0 | Categorical |
| Performance Rating | 0 | Numerical |

**4. Data Cleaning and Preprocessing:**

1. Performed outlier handling to ensure data quality and improve model robustness.
2. Applied label encoding and mapping techniques to convert categorical variables into numerical form for model compatibility.

**5. Exploratory Data Analysis:**

**Correlation Analysis**

* Identified strong correlations between Total Work Experience in Years and features like Age, Emp Job Level, Experience Years at This Company, Experience Years in Current Role and Years with Curr Manager.

**Data Visualization**

* Distribution of Emp Education Level (Histplot, Boxplot)
* Feature relationships (Scatter plots, Heatmaps)
* Multivariate analysis using pair plot.

**6. Modeling & Prediction**

**Model Selection**

* Implemented various models:
  + Logistic Regression
  + Random Forest
  + Gradient Boosting Classifier
  + Support Vector Machine

**Model Evaluation**

* Performance measured using:
  + Accuracy
  + Precision
  + Recall
  + F1-Score
* Best performing model: **Random Forest Classifier** shows the highest accuracy point.

**7. Model Comparison Report**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Logistic Regression | 0.84 | 0.52 | 0.48 | 0.50 |
| **Random Forest** | **0.95** | **0.96** | **0.90** | **0.93** |
| Gradient Boosting | 0.93 | 0.89 | 0.86 | 0.88 |
| Support Vector Machine | 0.77 | 0.00 | 0.00 | 0.00 |

* **Best Model Recommendation**: Random Forest

**8. Streamlit Deployment Link**

<https://employeeperformance-56qdygjgmoqjninrzmlkmf.streamlit.app>

**Github Link**

<https://github.com/Tamanna-Goel/employee_performance>

**9. Recommendations to Improve Employee Performance**

* Enhance Job Involvement  
  → Encourage participation in decision-making, offer challenging tasks, and align roles with employee strengths.
* Improve Work-Life Balance  
  → Introduce flexible working hours, hybrid work models, and wellness programs to reduce burnout.
* Boost Environment Satisfaction  
  → Regularly assess workplace conditions, provide a positive physical and cultural environment, and address employee feedback promptly.
* Increase Opportunities for Training and Development  
  → Offer regular skill-enhancement programs, mentorship, and career growth opportunities to maintain engagement.
* Focus on Managerial Relationships  
  → Train managers in leadership and communication skills to build stronger relationships and trust with team members.
* Recognize and Reward Performance  
  → Implement transparent and fair reward systems for high-performing employees to reinforce positive behaviour.
* Monitor and Support Post-Promotion Growth  
  → Provide structured support and training after promotions to help employees adapt and perform well in new roles.
* Reduce Attrition  
  → Conduct exit interviews, improve onboarding, and create a culture of inclusion and long-term growth to retain talent.
* Tailor Department-Specific Interventions  
  → Focus improvement efforts on departments with consistently lower performance ratings.

**10. Challenges Faced and Solutions**

**Data Quality Issues**

* Presence of categorical variables like gender, department, job role, etc., which machine learning models cannot process directly.
* Outliers in features like Hourly Rate, Distance from Home, and Num CompaniesWorked potentially affecting model accuracy.
* Ensuring no missing values and maintaining data integrity throughout the pipeline.
* Choosing the most suitable model for prediction given multiple algorithm performances.
* Balancing model performance with interpretability for business decision-making.

**Solutions and Model Performance**

* Balancing model performance with interpretability for business decision-making.
* Conducted outlier detection and handling using boxplots and statistical thresholds.
* Verified and confirmed that the dataset contained no null values, ensuring smooth preprocessing.
* Evaluated models on multiple metrics (accuracy, precision, recall, F1-score) and selected **Random Forest** for its superior performance.
* Chose **Random Forest**, which not only performed well but also offered **feature importance** insights to support strategic decisions.

**11. Conclusion**

* The project aimed to analyse employee data from INX Future Inc. to understand factors affecting performance and build a predictive model.
* Exploratory Data Analysis revealed that factors like Job Involvement, Work-LifeBalance, and Environment Satisfaction have a significant impact on employee performance.
* Data preprocessing steps included:
  + Outlier detection and handling
  + Label encoding and mapping of categorical variables
* Multiple machine learning models were trained and evaluated:
  + Logistic Regression – Accuracy: 84%
  + Support Vector Machine – Accuracy: 77%
  + Gradient Boosting – Accuracy: 93%
  + **Random Forest** – **Best Accuracy: 95%**
* The **Random Forest Classifier** was chosen as the best-performing model with:
  + **Precision:** 96%
  + **Recall:** 90%
  + **F1-Score:** 93%
* The model can now be used to:
  + Predict new employee performance during hiring
  + Identify low performers for early intervention
  + Support management decisions with data-backed insights
* These results will help INX improve performance, maintain employee morale, and uphold its reputation as a top employer.