***Report on Employee Dataset Analysis***

*Group Members: Sufiya Banu, Bintou Ba, Naizak Bhot*

***Introduction:***

*In today’s dynamic business environment, understanding employee data is crucial for organizations to make informed decisions regarding recruitment, retention, and performance management. This report presents the findings of a data mining project conducted on an employee dataset, which includes information such as EEID, full name, job title, department, business unit, gender, ethnicity, age, hire date, annual salary, bonus percentage, country, city, and exit date. Through this analysis, we aim to uncover insights into employee demographics, departmental distribution, salary discrepancies, turnover rates, and other pertinent factors impacting organizational performance.*

***Background:***

*The dataset comprises information collected over a certain period, encompassing various attributes related to employees within the organization. Understanding this data can help identify patterns, trends, and correlations that offer valuable insights into workforce dynamics.*

***Findings:***

1. *Raw Data: The raw data consists of entries for each employee, including personal details, job-related information, and tenure within the organization.*
2. *Data Preparation: Prior to analysis, the dataset underwent preprocessing to handle missing values, standardize data formats, and ensure consistency across variables.*
3. *Data Cleaning: Data cleaning procedures were implemented to address inconsistencies, errors, and outliers, ensuring the integrity and accuracy of the dataset.*
4. *Exploratory Data Analysis (EDA): EDA revealed several key insights:*

*- Distribution of employees across departments and business units.*

*- Gender and ethnicity representation within the workforce.*

*- Age distribution and its impact on various metrics.*

*- Disparities in salary levels across different departments and job roles.*

*- Factors contributing to employee attrition and turnover rates.*

1. *Data Modeling: For our data analysis, we used different computer programs called machine learning algorithms to learn about and predict things about the employees. Here are the ones we used:*
   * *CatBoost: It's good at understanding information even if it's in categories, like job titles or departments. We used it because our employee data has different kinds of information, and CatBoost can handle that easily.*
   * *XGBoost: This helper is great at figuring out patterns, especially in really big sets of data. We chose it because our employee data is pretty big, and XGBoost is great at handling that.*
   * *Random Forest: It's like a group of helpers working together. They're good at making guesses based on lots of different factors. We used it because our employee data has many different things that could affect each other, and Random Forest is good at considering all those factor.*
   * *Logistic Regression: This simple helper is good at guessing between two choices, like if something will happen or not. We picked it because we wanted to understand how certain things about employees, like their age or job role, might affect whether they stay in the company or leave.*
2. *Visualization: Visual representations, including charts, graphs, and heatmaps, were utilized to illustrate trends, relationships, and distributions within the data, enhancing interpretability and communication of findings.*

***Results:***

1. *Departmental Distribution: The analysis revealed that a significant number of employees were concentrated in a particular department, possibly indicative of its strategic importance or operational focus.*
2. *Salary Discrepancies: Disparities in annual salaries across different departments were observed, influenced by factors such as job role, experience, and market demand.*
3. *Gender Representation: The dataset highlighted the gender distribution within the workforce, enabling organizations to assess gender diversity and implement measures to promote inclusivity and equality.*
4. *Turnover Rates: Analysis of exit dates and associated factors provided insights into employee turnover, including reasons for attrition such as job dissatisfaction, career advancement opportunities, and organizational culture.*

***Conclusion:***

*In conclusion, looking at the employee data gave us useful information about the people working in our company, how different departments work, how much people get paid, and why some people leave their jobs. By using special computer techniques and pictures to show the data, companies can learn a lot about their workers. They can find ways to make things better, keep employees happy, and make the company work even better overall.*