

A Project Report
On
Employee Performance Analysis

Submitted by
Ishan Gupta, 09 (B.Tech(Hons.))
Harsh Gupta, 08 (B.Tech(Hons.))

Supervisor
Bhavin Soni
Data Science Trainer

Department of Computer Engineering and Applications
GLA University



GLA University, Mathura - 281406

Date of Submission -12/10/2023

Employee Performance Analysis

Problem Statement

The aim of this project is to develop a trained model which can predict the employee performance. This will be used to hire employees and will provide recommendations to improve the employee performance based on insights from analysis.

Dataset information

- **EmpNumber:** An alphanumeric identifier for each employee, serving as distinct identifier.
- **Gender:** The gender of the employee.
- **EducationalBackground:** The educational background of the employee.
- **MartialStatus:** Employee is married or not.
- **EmpDepartment:** The department in which the employee works.
- **EmpJobRole:** The specific job role or position of the employee.
- **BusinessTravelFrequency:** Number of times employee did Business Travel.
- **OverTime:** Whether the employee works overtime.
- **Attrition:** Indicates whether the employee has left the company or not.

- **Age:** The age of the employee.
- **DistanceFromHome:** The distance of employee's residence from the workplace.
- **EmpHourlyRate:** The hourly rate of the employee.
- **NumCompaniesWorked:** The number of companies the employee has worked for.
- **EmpLastSalaryHikePercent:** The percentage of the employee's last salary hike.

- **TotalWorkExperienceInYears:** The total work experience of the employee in years.
- **TrainingTimesLastYear:** The number of times the employee received training last year.
- **ExperienceYearsAtThisCompany:** The number of years employee has worked at the current company.
- **ExperienceYearsInCurrentRole:** The number of years employee has spent in the current job role.
- **YearsSinceLastPromotion:** The number of years since the employee's last promotion.
- **YearsWithCurrManager:** The number of years the employee has been working with the current manager.

- **EmpEducationLevel:** The employee's level of education, typically an ordinal scale.
- **EmpEnvironmentSatisfaction:** Employee's satisfaction with the work environment, typically ordinal.
- **EmpJobInvolment:** Employee's job involment level, typically ordinal.
- **EmpJobLevel:** The job level of the employee, usually on an ordinal scale.
- **EmpJobSatisfaction:** Employee's job satisfaction level, typically ordinal.
- **EmpWorkLifeBalance:** Employee's work-life balance satisfaction, typically ordinal.
- **EmpRelationshipSatisfaction:** Employee's satisfaction with work relationships, typically ordinal.

- **PerformanceRating:** The performance rating of the employee, which is the variable we aim to predict.

Abstract

This project makes use of a dataset containing 1,200 workers from the fictional company INX Future Inc., which has a solid reputation in the automation and data analytics fields. The dataset contains a wide variety of employee variables, such as performance ratings, job-related factors, and demographics. Creating a prediction model for employee performance is the main goal, with the ordinal PerformanceRating serving as the target variable. Preprocessing the data, correlation analysis, and machine learning techniques are all part of the analysis. Additionally, the project's insights are used to generate practical suggestions for improving worker performance. The primary determinants impacting employee performance are thoroughly investigated using the dataset's categorical, numerical, and ordinal features. The findings provide HR and management with insightful advice on how to promote a more contented and productive work environment.

Introduction

The act of gathering, evaluating, and interpreting data regarding an employee's performance in order to pinpoint problem areas and enhance managerial choices is known as employee performance analysis. This project shows how to analyze employee performance data using data visualization tools and create suggestions for raising employee performance. The research makes use of a dataset of 1200 people and 28 attributes, such as performance ratings, job characteristics, and employee demographics. The research determines the most crucial elements influencing employee performance through a range of data visualization approaches, such as feature importance, correlation analysis, and categorization. The project's conclusions imply that the most crucial elements influencing worker performance are work-life balance, job involvement, and employee satisfaction. Additionally, the study formulates some suggestions for enhancing.

METHODOLOGY

The proposed system can be divided into three main modules. The modules and their functions are defined as follows: -

- a) **Data Collection:** The data was given from the IABAC for this project where the collected source is IABAC™. The dataset consists of Categorical data and Numerical data. The Target variable consist of ordinal data, so this is a classification problem. The dataset that we will be using will be batch i.e., static data and not dynamic.
- b) **Data cleaning and data engineering:** Here we are removed unnecessary features, dropped unimportant information and finally converting all the categorical data to numerical data.
- c) **Visualization and Interpretation:** We plotted various graphs by using matplotlib, seaborn and plotly. Analysis and interpretation of data by using these graphs.

Packages explanation

Pandas

Pandas is a Python package made especially for analyzing and manipulating data. Pandas make it easier to import, clean, and explore your dataset in EDA. It offers data structures for efficient data organization and analysis, such as DataFrames. Pandas allows you to visualize data distributions, linkages, and correlations through univariate and bivariate analysis, as well as analyze data statistics, handle missing values, filter and choose particular columns or rows.

Seaborn

Based on Matplotlib, Seaborn is a Python data visualization package. It is perfect for EDA since it makes the process of creating visually appealing and educational representations easier. A high-level interface is provided by Seaborn to create a range of plots, such as scatter plots, bar plots, box plots, histograms, and heatmaps. You can investigate data distributions, variable relationships, and correlations using these visualizations, which might offer insightful information when your project is in its EDA phase.

Matplotlib

An array of static, animated, and interactive visualizations can be produced with the robust Python package Matplotlib. For creating unique, highly customizable plots and charts, it is a vital tool in EDA. Line plots, bar charts, scatter plots, histograms, and other visualization styles can be created with Matplotlib and are essential for analyzing data distributions, patterns, and correlations.

Plotly

A Python package called Plotly offers interactive data visualization capabilities for exploratory data analysis (EDA). With its help, you may generate interactive plots, charts, and graphs for visual data exploration and analysis. Many visualization choices are available with Plotly, such as scatter plots, line charts, bar charts, histograms, heatmaps, and more. Using Plotly, you can make interactive graphs that let you highlight particular data points, zoom in and out, and hover over data points to get more details.