Interim Report

Project Title: Advanced HR Analytics: Workforce Performance Prediction
Student Name: Neeraj Suresh
Roll No.: AA.SC.P2MCA2401434
Date: November 30, 2025

Abstract

This project aims to design and implement an advanced HR analytics system that leverages machine learning to predict employee performance using multi-source HR datasets. By integrating data such as attendance, training scores, recruitment sources, and previous performance, the system identifies key factors influencing workforce productivity. The model enables proactive HR decision-making in recruitment, training allocation, and retention strategies. Tools like Python (scikit-learn, Pandas) and Power BI are used to build, evaluate, and visualize predictive models, ultimately supporting data-driven workforce management.

Objectives

- Develop a predictive analytics framework to forecast employee performance and retention likelihood.
- Identify key HR factors influencing workforce productivity through feature selection and explainability techniques.
- Implement machine learning models using historical multi-source HR data.
- Deliver actionable insights via interactive Power BI dashboards for HR stakeholders.

Scope

This study focuses on building a predictive system utilizing comprehensive HR data covering attendance records, training completion, recruitment channels, and past performance metrics. It will involve data cleaning, feature engineering, model development, evaluation, and visualization stages. The project aims to provide insights that improve recruitment quality, training effectiveness, and employee retention. While privacy considerations are respected by using anonymized or synthetic datasets, the scope does not extend to real-time employee monitoring or psychological assessments.

Introduction

Background

Many organizations rely on reactive HR practices based on experience rather than data-driven insights, leading to inefficiencies like high turnover and poor workforce planning. With the availability of large HR datasets, predictive analytics offers a method to forecast employee outcomes and optimize HR interventions. This project explores the integration of machine learning techniques in HR analytics to enhance decision-making, focusing on workforce performance prediction to drive strategic recruitment, training, and retention.

Related Work

Several platforms and studies demonstrate the value of predictive HR analytics. Deloitte and IBM Watson Analytics have developed frameworks utilizing logistic regression, random forests, and support vector machines to predict turnover and performance. Research highlights attendance, training, and recruitment source as critical predictors. This project extends prior work by combining multiple HR data sources into a unified predictive model enhanced with explainable AI for transparency.

Problem Statement

HR departments often face challenges in identifying which employees will perform well or remain long-term, relying on reactive, anecdotal methods. The aim is to bridge this gap by developing an analytics system capable of accurately predicting employee performance, thereby enabling proactive workforce decisions and reducing attrition-related costs.

Methods and Algorithms

Module 1: Data Collection and Cleaning
Collect multi-source HR data using Python's Pandas, handle missing and inconsistent data, normalize fields, and remove duplicates to prepare a clean dataset.

Module 2: Feature Engineering
Derive new metrics relevant to performance, reduce dimensionality using Principal Component Analysis (PCA), and select features that contribute most to prediction accuracy.

Module 3: Model Development
Implement and compare machine learning models including Random Forest, Logistic

Regression, and Gradient Boosting, using train-test splits and evaluating metrics such as accuracy, precision, recall, and F1 score.

Module 4: Model Explainability
Use SHAP values and feature importance plots to provide interpretability of model predictions, helping HR understand key factors driving employee outcomes.

Module 5: Dashboard and Reporting
Create Power BI dashboards to visualize predictions, feature trends, and provide actionable insights for HR teams.

Expected Outcome

The project will deliver a functional HR analytics system capable of predicting employee performance categories with high accuracy. It will highlight top influencing factors and present this information through interactive dashboards, enabling HR decision-makers to optimize recruitment, training, and retention strategies. This will lead to more efficient workforce management and reduced attrition.

Student Name: Neeraj Suresh
Roll No: AA.SC.P2MCA2401434
Date: November 30, 2025
Signature: