Machine Learning Approach for Employee Performance Prediction (Garment Worker Productivity)

Table of Contents

1. Introduction

1.1 Project Overview

This project focuses on developing a machine learning-based solution to predict garment worker productivity using historical shop-floor data. By analyzing key factors such as team, SMV, overtime, idle time, and departmental information, the system aims to forecast actual productivity before the workday begins. The goal is to support better planning, improve operational efficiency, and provide HR and managers with actionable insights for performance tracking and workforce optimization.

1.2 Objectives

- Forecast actual productivity of garment workers based on past production data
- Help HR and supervisors identify underperforming or high-performing teams
- Enable data-driven decisions for staffing, training, and incentive planning
- Deploy the predictive model through a simple and accessible web application

2. Project Initialization and Planning Phase

2.1 Define Problem Statement

Problem Statement: HR needs a quick way to predict each employee's productivity. Accurate forecasts let managers adjust staffing, overtime, and incentives early, cutting delays and labour costs. The task is challenging because productivity depends on a noisy mix of categorical factors (quarter, department, day) and numeric ones (SMV, overtime, idle time).
Problem Statement Report: Click Here

2.2 Project Proposal (Proposed Solution)

The proposed project, “Employee Performance Predictor,” uses machine-learning to forecast each employee's performance. By training a model on a rich dataset—quarter, department, weekday, team ID, SMV, overtime, idle time, incentives, and other shop-floor metrics—we will generate a 0-to-1 productivity score that managers can view in a simple web app. The tool supports the manager’s goal of smarter staffing and incentive decisions, cuts overtime and delay risk, and streamlines daily planning, leading to higher efficiency and lower labour costs.
Project Proposal Report: Click Here

2.3 Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for the garment worker productivity prediction system. It includes setting timelines and shaping the overall project strategy. In this phase, the team builds a clear understanding of the dataset, sets targets for accurate productivity forecasting, and plans the steps for data cleaning, feature selection, and model development. Proper planning ensures a smooth workflow, supports integration into a web application, and helps deliver a practical tool for HR and production managers.
Project Planning Report: Click Here

3. Data Collection and Preprocessing Phase

3.1 Data Collection Plan, Raw Data Sources Identification Report

The dataset for “Employee Performance Prediction” is sourced from Kaggle and contains details such as department, day, team, SMV, overtime, idle time, and actual productivity. Data quality is ensured through cleaning steps like handling missing values, correcting inconsistent labels, and dropping unusable columns. These steps provide a reliable and ethical foundation for building accurate predictive models.
Data Collection Report: Click Here

3.2 Data Quality Report

The dataset for “Employee Performance Prediction” is sourced from Kaggle. It includes production-related features such as department, day, team, targeted productivity, SMV, overtime, idle time, and more. Data quality is maintained through thorough verification, handling missing values (e.g., dropping the wip column), and standardizing categorical entries, establishing a reliable foundation for predictive modeling.
Data Quality Report: Click Here

3.3 Data Exploration and Preprocessing

Data Exploration involves analyzing the garment worker productivity dataset to identify patterns, feature distributions, and potential outliers. Preprocessing steps include dropping columns with excessive missing values, converting the date column to month, merging duplicate category labels, and applying multi-column label encoding to categorical features. These essential steps improve data quality and consistency, ensuring the accuracy and reliability of the machine learning models used in the productivity prediction project.
Data Exploration and Preprocessing Report: Click Here

4. Model Development Phase

4.1 Feature Selection Report

The Feature Selection Report outlines the reasoning behind choosing specific features (e.g., department, day, team, SMV, overtime, idle time) for the productivity prediction model. Each feature was evaluated for its relevance and contribution to the target variable, actual_productivity.
Feature Selection Report: Click Here

4.2 Model Selection Report

The Model Selection Report explains the choice of Linear Regression, Random Forest, and XGBoost models for predicting garment worker productivity. These models were selected for their ability to handle both categorical and numerical features, model complex relationships, and deliver reliable predictions.
Model Selection Report: Click Here

4.3 Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code applies selected algorithms to the garment productivity dataset, laying the groundwork for accurate prediction. The Model Validation and Evaluation Report then tests how well the models perform using metrics like R^2 score, MAE, and MSE to ensure reliability and effectiveness in predicting actual worker productivity.
Model Development Phase Template: Click Here

5. Model Optimization and Tuning Phase

5.1 Hyperparameter Tuning Documentation

Basic hyperparameter tuning was performed on the XGBoost model using simple adjustments to parameters like **n_estimators** and **max_depth**. While slight improvements in performance were observed during testing, the gains were minimal. To maintain simplicity and avoid unnecessary complexity, these tuned versions were not implemented in the final deployment. The default settings already provided reliable and consistent results aligned with the project goals.

5.2 Performance Metrics Comparison Report

Model	MAE	R^2 Score
Linear Regression	~0.1039	~0.3025
Random Forest	~0.0677	~0.6082
XGBoost	**~0.0673**	**~0.6263**

5.3 Final Model Selection Justification

The Final Model Selection Justification explains the choice of XGBoost as the final model. Its high accuracy, ability to handle complex patterns, and strong performance align well with the project goals, making it the best option for predicting garment worker productivity.

Model Optimization and Tuning Phase Report: Click Here

6. Results

6.1 Output Screenshots

Drive: [Click Here]

7. Advantages & Disadvantages

Advantages

1. Helps HR & Managers to make data driven decisions
2. Helps in sorting employees based on performance
3. Easy to use web interface

Disdvantages

1. It isn't intergrated to the HR databse to get real-time data
2. Missing Values in dataset required cleaning
3. It is specific to Garment Factory Domain

8. Conclusion

The project successfully showcases the use of machine learning to predict garment worker productivity in a factory setting. It highlights the potential for improving daily planning and resource allocation. The final web application provides a quick and user-friendly way to input worker data and receive accurate productivity predictions.

9. Future Scope

- Host the productivity prediction app on cloud platforms for wider accessibility
- Integrate the system with live factory or HR databases to automate input collection
- Add smart alerts to notify managers when predicted productivity falls below target
- Explore advanced techniques like neural networks or ensemble stacking to boost model accuracy
- Develop a mobile or tablet interface to help floor managers access predictions in real time during operations

10. Project Demonstration

10.1 Source Code

Source code: Click Here

10.2 GitHub & Project Demo Link

Github Repo: Click Here
Drive: Click Here