Project report

Predicting Employee Productivity in the Garment Industry

Abstract

This report presents a comprehensive analysis aimed at predicting employee productivity in the garment manufacturing sector. Utilizing data from garment worker productivity, various machine learning models were developed and compared to identify the most effective predictor of actual productivity based on several input features.

---

1. Introduction

The garment industry, characterized by its labor-intensive nature, requires effective strategies to optimize productivity. This study explores the application of machine learning techniques to predict employee productivity, providing a tool for decision-makers to enhance operational efficiency.

---

2. Methodology

2.1 Data Preprocessing

The dataset, comprising 1197 instances and 15 attributes, was first cleansed of missing values, which were imputed using the mean of respective features to maintain data integrity. Categorical variables such as date, quarter, department, and day were transformed using label encoding to prepare for machine learning applications. Data normalization was performed using a Standard Scaler to ensure all features contributed equally to the model training process.

2.2 Feature Engineering

Key features were analyzed to understand their correlation with the target variable, actual productivity. This analysis helped in understanding the impact of each feature on employee productivity.

2.3 Data Splitting

The dataset was divided into training and test sets with a 67:33 ratio, ensuring adequate data was available for both training the models and validating their performance.

---

3. Model Development and Evaluation

3.1 Elastic Net Regression

Parameters such as alpha and l1_ratio were optimized using GridSearchCV. The best model achieved a score of 0.240, indicating moderate predictive capability. Performance metrics revealed the model may not capture all variability in the data, suggesting potential underfitting.

3.2 Support Vector Regression (SVR)

After hyperparameter tuning, the best parameters (C=10, gamma=0.1) led to an SVR model that outperformed the Elastic Net, with higher R2 and lower error metrics, indicating a better fit and higher reliability.

3.3 K-Nearest Neighbors (KNN)

Optimization of the n_neighbors parameter showed an optimal count of 10. However, the performance was inferior to SVR, highlighting the model's limitations in handling the dataset's complexity.

3.4 Ensemble Model - Random Forest

A Random Forest Regressor was tuned for parameters like n_estimators and max_depth, resulting in the most robust model with the highest R2 score and the lowest error metrics among all tested models.

---

4. Feature Importance Analysis

Both correlation analysis and feature importance from the Elastic Net model highlighted significant predictors such as SMV, number of workers, and targeted productivity, which are critical in influencing the productivity outcomes in garment manufacturing.

---

5. Conclusion and Recommendations

The Random Forest model, with its superior performance and ability to model complex nonlinear relationships, is recommended for predicting employee productivity in the garment industry. Future work should explore additional ensemble techniques and deep learning models to further enhance predictive accuracy.