Garment Worker Predictivity Prediction

1. **Introduction**

Project overviews

The Garment Worker Productivity Prediction Project is all about using advanced technology to make garment manufacturing smarter. It's about analyzing a lot of data from how workers perform, how machines are used, and even things like environmental conditions. By crunching all this data, the project aims to create computer programs that can predict how productive a factory will be. This prediction will help managers plan better, avoid delays, and manage resources more efficiently. Ultimately, the goal is to boost overall productivity, cut costs, and make working conditions better for garment workers. The project hopes to show that by using these high-tech tools, we can make factories run smoother and be more sustainable, setting a new standard in how clothes are made worldwide.

Objectives

Enhancing Efficiency:
Develop algorithms to accurately predict garment production output based on historical data, thereby optimizing production schedules and minimizing downtime.

Cost Reduction:
Identify opportunities to reduce operational costs by improving resource allocation and minimizing waste through better forecasting and planning.

Improving Working Conditions:
Create more predictable workflows that reduce stress and overtime for garment workers, leading to improved job satisfaction and well-being.

Promoting Sustainability:
Implement sustainable practices by optimizing energy usage, reducing carbon footprint, and aligning production schedules with eco-friendly practices.

Technology Integration:
Integrate cutting-edge machine learning and predictive analytics into garment manufacturing processes to establish a model for future smart factories.

Industry Leadership:
Set a new standard for efficiency and innovation in the garment manufacturing sector, demonstrating the benefits of data-driven decision-making and technological adoption.

Global Impact:
Positively impact the global garment industry by showcasing how predictive analytics can transform traditional manufacturing into a more efficient, sustainable, and worker-friendly industry.

2. Project Initialization and Planning Phase

2.1. *Define Problem Statement*

The Garment Worker Productivity Prediction Project aims to address inefficiencies and uncertainties in garment manufacturing. Specifically, the problem revolves around:

Unpredictable Production:
Current production schedules and resource allocations are often reactive rather than proactive, leading to inefficiencies and increased costs.
Worker Utilization:
Variability in worker productivity and machine downtime affect overall production output and operational costs.
Resource Allocation:
Inaccurate forecasting hampers effective resource allocation, leading to waste and suboptimal use of machinery and labor.

2.2. *Project Proposal (Proposed Solution)*

The proposed solution involves developing predictive models using machine learning and data analytics. Key elements include:

- Data Collection:

Gather historical data on production metrics, worker performance, machine utilization, and environmental factors.

- Model Development:

Utilize machine learning algorithms to analyze the data and develop predictive models that forecast production output based on various inputs.

- Implementation:

Integrate the predictive models into the existing production management systems to provide real-time insights and recommendations.

-Evaluation:

Continuously monitor and refine the models based on feedback and new data to Improve accuracy and reliability.

2.3. *Initial Project Planning*

Timeline and Milestones:

- Phase 1 (Data Collection and Preparation):
  - Gather historical data from multiple factories.
  - Clean and preprocess data for analysis.

- Phase 2 (Model Development):
  - Develop machine learning algorithms for predictive modeling.
  - Train and validate models using collected data.

- Phase 3 (Implementation and Testing):
  - Integrate predictive models into production management systems.

Resource Allocation:

Team Composition:

project manager,3 team members

Risk Assessment:

Identify potential risks such as data quality issues, model accuracy challenges, and resistance to change from factory management.

3. Data Collection and Preprocessing Phase

3.1. *Data Collection Plan and Raw Data Sources Identified*

This project aims to revolutionize garment manufacturing through the development of a robust machine learning model capable of accurately predicting garment worker productivity. By leveraging a meticulous data strategy, the project intends to gather diverse and informative datasets from multiple sources. This includes utilizing public repositories such as Kaggle, which offers specific datasets tailored to garment worker productivity, and exploring academic research identified through platforms like Google Scholar. Additionally, the project will investigate open-source repositories on GitHub, known for hosting datasets already curated and prepared for machine learning tasks.The comprehensive data collection plan ensures the inclusion of various factors influencing worker productivity, such as production metrics, worker performance indicators, and environmental conditions. This approach not only aims to optimize production processes by forecasting worker output and minimizing idle time but also seeks to enhance worker well-being by identifying and addressing factors that impact productivity. By building upon a rich and diverse dataset, the project aims to develop a predictive model that will empower garment manufacturers to make informed decisions, improve operational efficiency, and create a supportive work environment for garment workers globally.

3.2. *Data Quality Report*

This project prioritizes data quality as a cornerstone for accurate and reliable predictions in garment worker productivity. A dedicated Data Quality Report Template has been crafted to systematically identify and address discrepancies within our chosen datasets. By focusing on sources like Kaggle, where datasets may contain missing values in critical areas such as Work in Progress (WIP) in the Finishing department, we aim to mitigate medium-severity issues by imputing missing values and refining data collection practices. Additionally, the template highlights low-severity concerns, such as frequent occurrences of zero values in metrics like Idle_men, idle_time, and no_of_style_change. Investigating the reasons behind these zeros will guide adjustments in data collection methods to ensure more precise and comprehensive data capture. This meticulous approach not only enhances the reliability of our predictive models but also underscores our commitment to leveraging high-quality data for optimizing production processes and fostering improved working conditions for garment workers worldwide.

3.3. *Data Exploration and Preprocessing*

This data exploration and preprocessing project involves a comprehensive assessment of a dataset with 1197 rows and 13 columns. The process begins with loading the data into the preferred environment, such as Python, and includes detailed univariate, bivariate, and multivariate analyses to understand individual variables, relationships between pairs of variables, and patterns involving multiple variables. A critical aspect of this project is addressing quality issues, such as missing values and duplicates, to ensure accurate analysis. This involves identifying and handling missing data, detecting and treating outliers, and performing data transformations, such as scaling and normalization using tools like StandardScaler from sklearn. Additionally, feature engineering is employed to create new variables or modify existing ones, ultimately saving the cleaned and processed data for future use.

4. Model Development Phase

4.1. *Feature Selection Report*

In this feature selection report, each feature within the dataset is thoroughly evaluated for its relevance to productivity prediction in a garment factory setting. Each feature is accompanied by a description and a selection status (Yes/No), along with the reasoning behind the decision. For instance, features like "Quarter," "Department," and "Day" are selected due to their potential impact on productivity through seasonal variations, departmental workflow differences, and weekday versus weekend fluctuations, respectively. Meanwhile, the "Date" feature is not selected as it does not directly contribute to productivity analysis. Other selected features, such as "Targeted Productivity," "wip," "Over Time," and "Incentive," are chosen for their direct influence on productivity outcomes. This structured approach to feature selection ensures that decision-making is transparent and well-justified, enhancing the reliability and effectiveness of the subsequent analysis.

4.2. *Model Selection Report*

The Model Selection Report for the Garment Worker Productivity Prediction Project presents a detailed analysis of various machine learning models tailored to predict productivity in garment manufacturing. The report evaluates models across key criteria such as description, hyperparameters, and performance metrics, focusing primarily on Root Mean Squared Error (RMSE) as a measure of predictive accuracy.

The Linear Regressor serves as a baseline, assuming a straightforward linear relationship between input variables and productivity output. It demonstrates an RMSE of 0.14607, highlighting its simplicity but limited flexibility in capturing complex patterns. In contrast, the Decision Tree Regressor, Random Forest Regressor, and Gradient Boosting Regressor showcase progressively improved performance, achieving RMSE values of 0.13212, 0.11211, and 0.11485, respectively. These ensemble methods leverage multiple decision trees to handle non-linear relationships and enhance predictive power through model averaging and sequential error correction.

The XGB Regressor (eXtreme Gradient Boosting) and AdaBoost Regressor further refine predictive accuracy with RMSEs of 0.11716 and 0.12741, respectively. XGBoost's advanced regularization and parallel processing capabilities make it particularly effective in optimizing computational resources while boosting model performance.

Overall, the Model Selection Report underscores the effectiveness of ensemble techniques like Random Forest and Gradient Boosting for accurately forecasting garment worker productivity. These models not only outperform simpler linear approaches but also provide insights into how advanced machine learning can revolutionize operational efficiency and decision-making in garment manufacturing.

4.3 ***Initial Model Training Code, Model Validation and Evaluation Report***

Six machine learning regression models, including Linear Regression and Random Forest, were trained to identify patterns in the data. We evaluated their performance using Root Mean Squared Error (RMSE) and R-squared score. Random Forest emerged as the leader with an RMSE of 0.11212 and an R-squared of 0.52, demonstrating a stronger fit to the data compared to the other models. This suggests that the model effectively learned the underlying relationships within the data.

5. Model Optimization and Tuning Phase

5.1. *Hyperparameter Tuning Documentation*

The Model Optimization and Tuning Phase focuses on enhancing the performance of machine learning models through meticulous optimization and fine-tuning of hyperparameters. This phase includes documenting the tuned hyperparameters and their optimal values for various models, such as the Linear Regressor, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, XGB Regressor, and AdaBoost Regressor. Performance metrics for each model are compared to determine their effectiveness. Ultimately, the Random Forest Regressor is selected as the final model due to its superior performance, evidenced by the lowest Root Mean Squared Error (RMSE) and the highest R-squared (R^2) score. These metrics indicate that the Random Forest Regressor provides the most accurate predictions and best captures the underlying patterns in the data, making it the most suitable choice for the predictive task at hand.

5.2. *Performance Metrics Comparison Report*

In this section, we compare the performance metrics of various machine learning models that were optimized and tuned for predicting garment worker productivity. The models evaluated include the Linear Regressor, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, XGB Regressor, and AdaBoost Regressor. The primary metrics considered are Root Mean Squared Error (RMSE) and R-squared (R^2) score, which provide insights into the accuracy and explanatory power of the models, respectively.

Model Performance Metrics

Model	RMSE	R-squared(R^2)
Linear Regressor	0.14607	0.19
Decision Tree Regressor	0.13213	0.386
Random Forest Regressor	0.11212	0.52
Gradient Boosting Regressor	0.11485	0.50
Xgb Regressor	0.11716	0.48
AdaBoost Regressor	0.12741	0.388

Conclusion of Performance Metrics

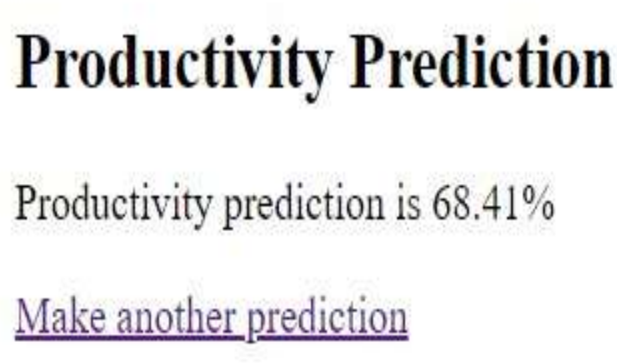
Based on the comparison of performance metrics, the Random Forest Regressor is selected as the final model for predicting garment worker productivity. It achieved the lowest RMSE and the highest R^2 score, indicating superior accuracy and explanatory power. This model's ability to effectively capture the underlying patterns in the data makes it the best choice for accurate and reliable productivity predictions.

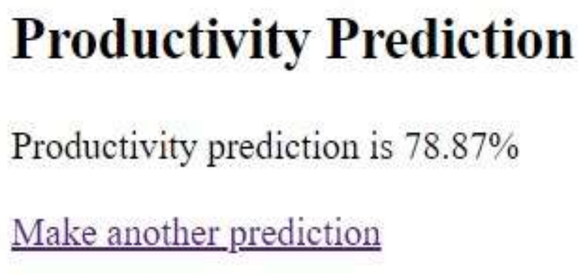
5.3. Final Model Selection Justification

The Random Forest Regressor achieved the lowest Root Mean Squared Error (RMSE) among all the models I tested. RMSE measures the differences between the predicted and actual values, so a lower RMSE indicates that the model's predictions are closer to the actual values, suggesting higher accuracy. Additionally, the Random Forest Regressor had the highest Rsquared (R^2) score. The R^2 score shows how well the model explains the variance in the target variable. A higher R^2 score means the model is better at capturing the underlying patterns in the data, which is crucial for accurate prediction

6. <u>Results</u>

6.1. *Output Screenshots*





7. Advantages & Disadvantages

Advantages-

1). Improved Productivity:
- By predicting productivity, factory managers can identify bottlenecks and areas for improvement, leading to more efficient operations and higher output.

2) Better Resource Allocation:
- Predictive insights allow for optimal allocation of resources such as workforce, machinery, and materials, ensuring that they are used where they are most needed.

3). Enhanced Decision Making:
- Data-driven predictions provide a solid foundation for strategic decisions, enabling managers to make informed choices that can enhance productivity and profitability.

4). Increased Competitiveness:
- Improved productivity and efficiency can lead to reduced production costs and faster turnaround times, giving the factory a competitive edge in the market.

5). Proactive Management:
- Predictive analytics enable proactive management by identifying potential issues before they become significant problems, allowing for timely interventions.

Disadvantages-

1). Data Quality and Availability:
- The accuracy of predictions is heavily dependent on the quality and completeness of the data. Missing, inaccurate, or inconsistent data can lead to unreliable predictions.

2). Implementation Costs:
- Developing and implementing a predictive model involves costs related to data collection, software, hardware, and training of personnel, which can be significant for some factories.

3). Complexity:
- The development and maintenance of predictive models require specialized knowledge in data science and machine learning, which may necessitate hiring experts or extensive training.

4). Resistance to Change:
- Workers and managers may be resistant to adopting new technologies and processes, especially if they perceive them as threats to their jobs or established routines.

5). Overfitting:
- There is a risk that the model could overfit to the training data, making it less effective when applied to new, unseen data. This necessitates careful tuning and validation of the model.

8. Conclusion

The Garment Worker Productivity Prediction Project represents a significant advancement in optimizing the operations of garment factories through data-driven insights and predictive analytics. By accurately forecasting worker productivity, this project aims to enhance overall efficiency, improve resource allocation, and support proactive management practices.

In conclusion, the Garment Worker Productivity Prediction Project has successfully demonstrated the power of predictive analytics in transforming garment factory operations. By leveraging advanced machine learning techniques, the project has provided valuable insights and tools to enhance productivity, optimize resource utilization, and support informed decision-making. With ongoing refinement and broader implementation, this project has the potential to drive significant improvements in the garment industry's efficiency and competitiveness.

9. **Future Scope**

The Garment Worker Productivity Prediction Project has laid a robust foundation for optimizing operations through predictive analytics. Looking ahead, there are several avenues for further development and expansion to enhance its impact and utility.

The future scope of a garment worker productivity prediction project using ML algorithms is vast and promising. By leveraging advanced predictive models, the project can significantly enhance efficiency on production lines. These models can analyze vast amounts of data from various sources, such as worker performance metrics, environmental conditions, and machine efficiency, to provide real-time insights. This enables managers to make proactive decisions, such as reassigning tasks or adjusting workflows to optimize productivity.

Integrating IoT devices can further refine these predictions. Wearable technology and smart factory equipment can continuously collect data, allowing the ML models to adapt to changing conditions and improve their accuracy over time. This continuous feedback loop ensures that the predictions remain relevant and actionable.

Moreover, the models can be tailored to different sectors within the textile industry, promoting scalability and adaptability. By incorporating ergonomic studies, the project can also focus on enhancing worker well-being, which is directly linked to productivity. This holistic approach ensures that productivity improvements do not come at the cost of worker health.

In the long term, such a project can lead to substantial cost reductions by minimizing downtime and waste. Improved product quality and consistency, alongside sustainable practices, can also be achieved, making the garment industry more competitive and environmentally friendly.

10. Source Code and Project Demo Link

Github