Project Documentation

Date	09 November 2024
Team ID	592545
Project Name	Garment Worker Productivity Prediction using Machine Learning

Introduction:

Project Overview:

The garment industry is one of the largest industries in the world, and garment worker productivity is a crucial factor in determining the success and profitability of a company. In this project, we aim to develop a machine learning model that predicts the productivity of garment workers based on a given set of features. Our dataset contains information on various attributes of garment production, including the quarter, department, day, team number, time allocated, unfinished items, over time, incentive, idle time, idle men, style change, number of workers, and actual productivity. We will use this dataset to train and evaluate our predictive model.

The development of an accurate garment worker productivity prediction model using machine learning can have significant implications in various domains, including manufacturing, human resources, and supply chain management. This model can help companies identify the factors that affect worker productivity and take corrective actions to improve efficiency, reduce costs, and enhance their competitiveness in the market.

Purpose:

The purpose of the project is to develop a machine learning model that predicts the productivity of garment workers. This project aims to achieve several specific goals:

1. Model Development: The basic aim is the happening of a predictive machine intelligence model namely worthy providing correct forecasts of costume worker output. This model will be devised, prepared, and proven using the dataset's rich array of attributes.
2. Productivity Enhancement: A center purpose concerning this project search out educate the determinants influencing laborer output inside the clothing industry. By recognizing and understanding these elements, our aim search out offer valuable visions to associations, permissive bureaucracy to enact calculated mediations that help trained workers efficiency.
3. Cost Reduction: Improved output, stopping from a deeper understanding of the determinants at play, can bring about solid cost reductions in the garment

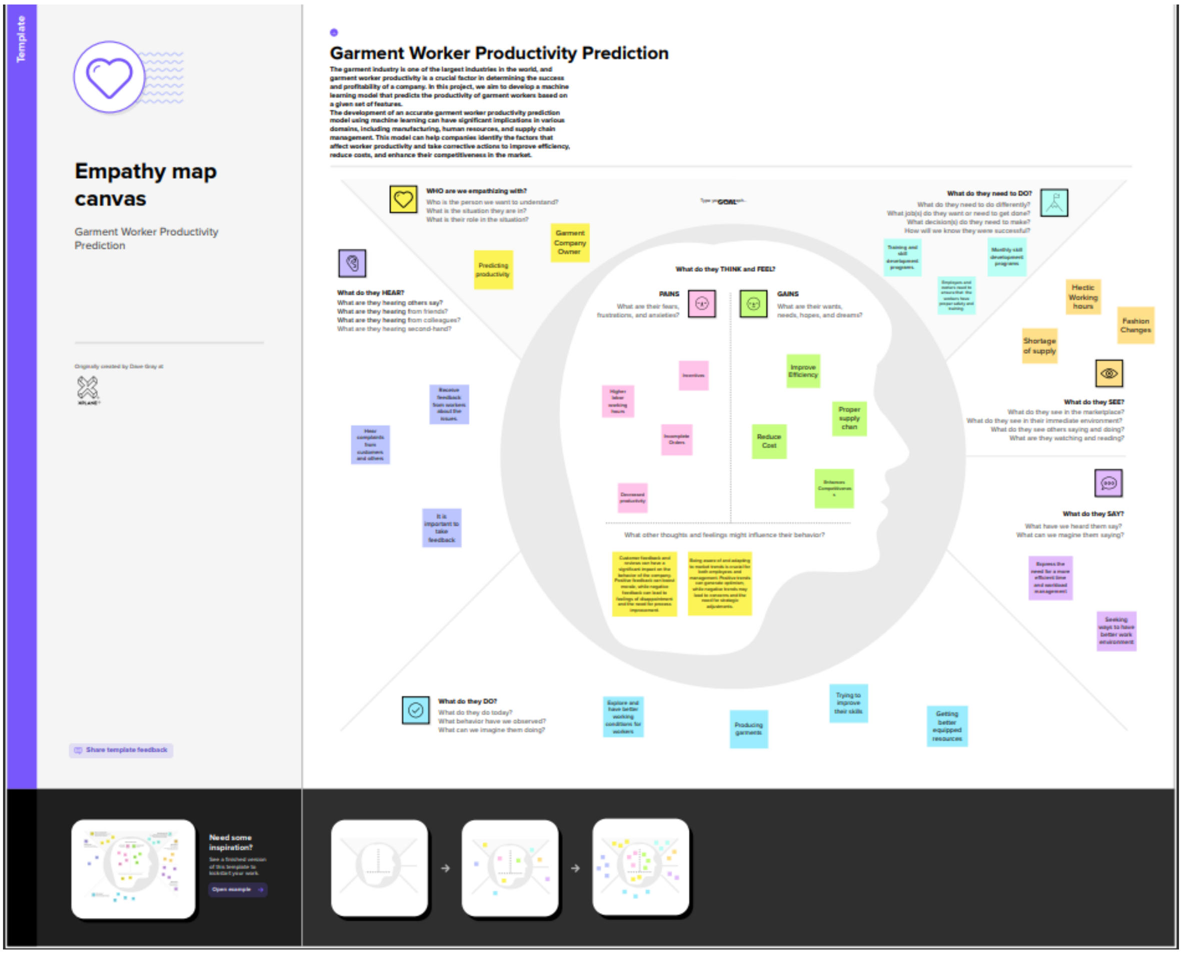
result process. Optimizing trader depiction has the potential to weaken the money required to achieve a particularized level of gain.

4. Enhanced Competitiveness: By helping output and trimming functional costs, this project aspires to uplift the ambitious importance of guests in the garment industry. Efficiency gains can turn to more ambitious valuing and enhanced delivery periods, through fortifying their position marketing.
5. Cross-Industry Impact: The judgments accumulate from this project are not confined to the robe subdivision unique. The communication well-informed about managing and optimizing a trained workers have widespread relevance, touch differing other energies, containing production, workforce, and supply chain management.

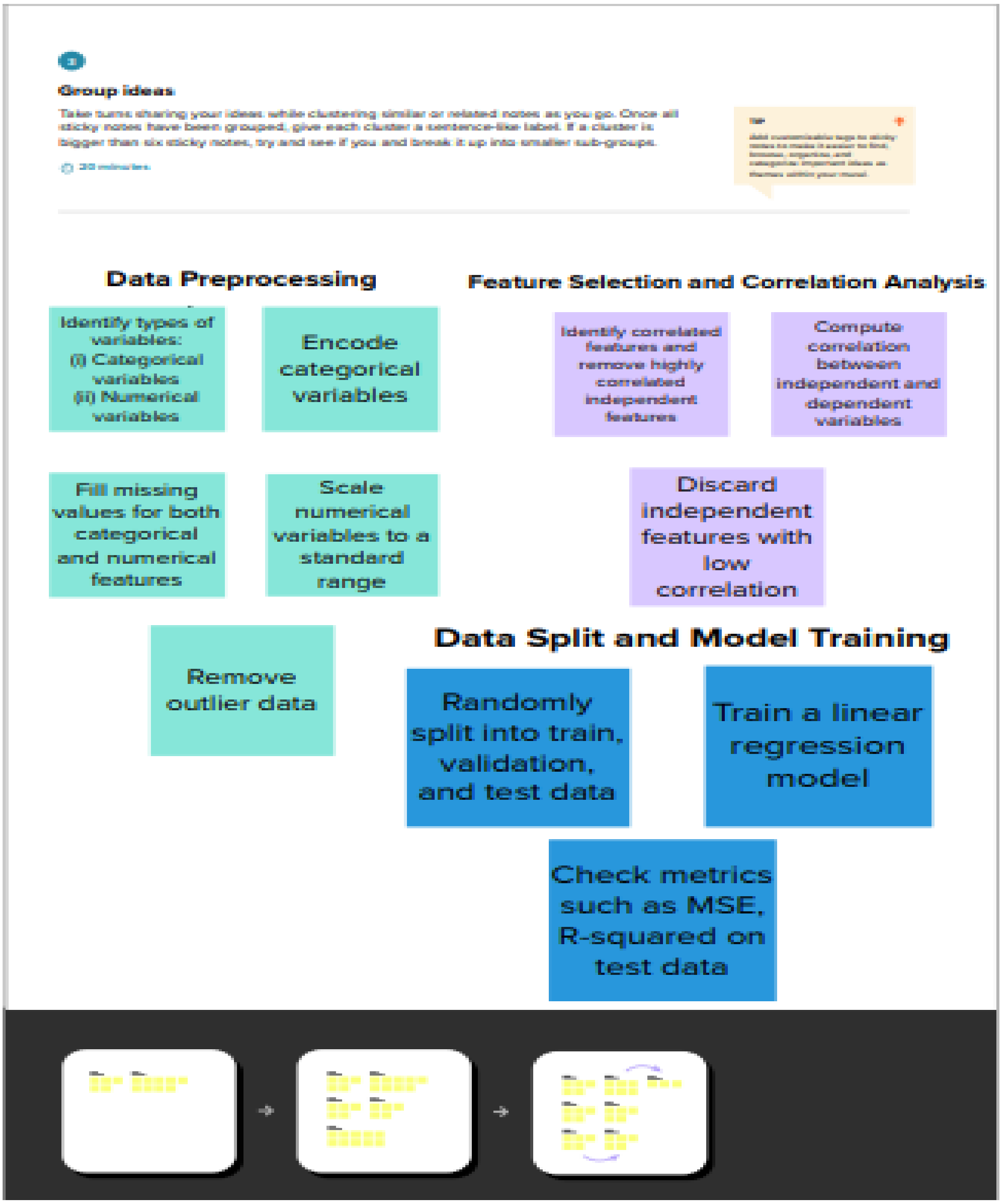
The predicting model for costume worker output project inquires to leverage the potential of machine intelligence to completely comprehend and conclude peasant conduct in the garment area. This understanding, when used, promises heightened effectiveness, weakened functional costs, and enhanced competing standing. The suggestions are not forced to the garment industry unique, as the acumens gained maintain the potential to upset and convert operations in diversified subdivisions, underscoring the meaning and universality concerning this leadership.

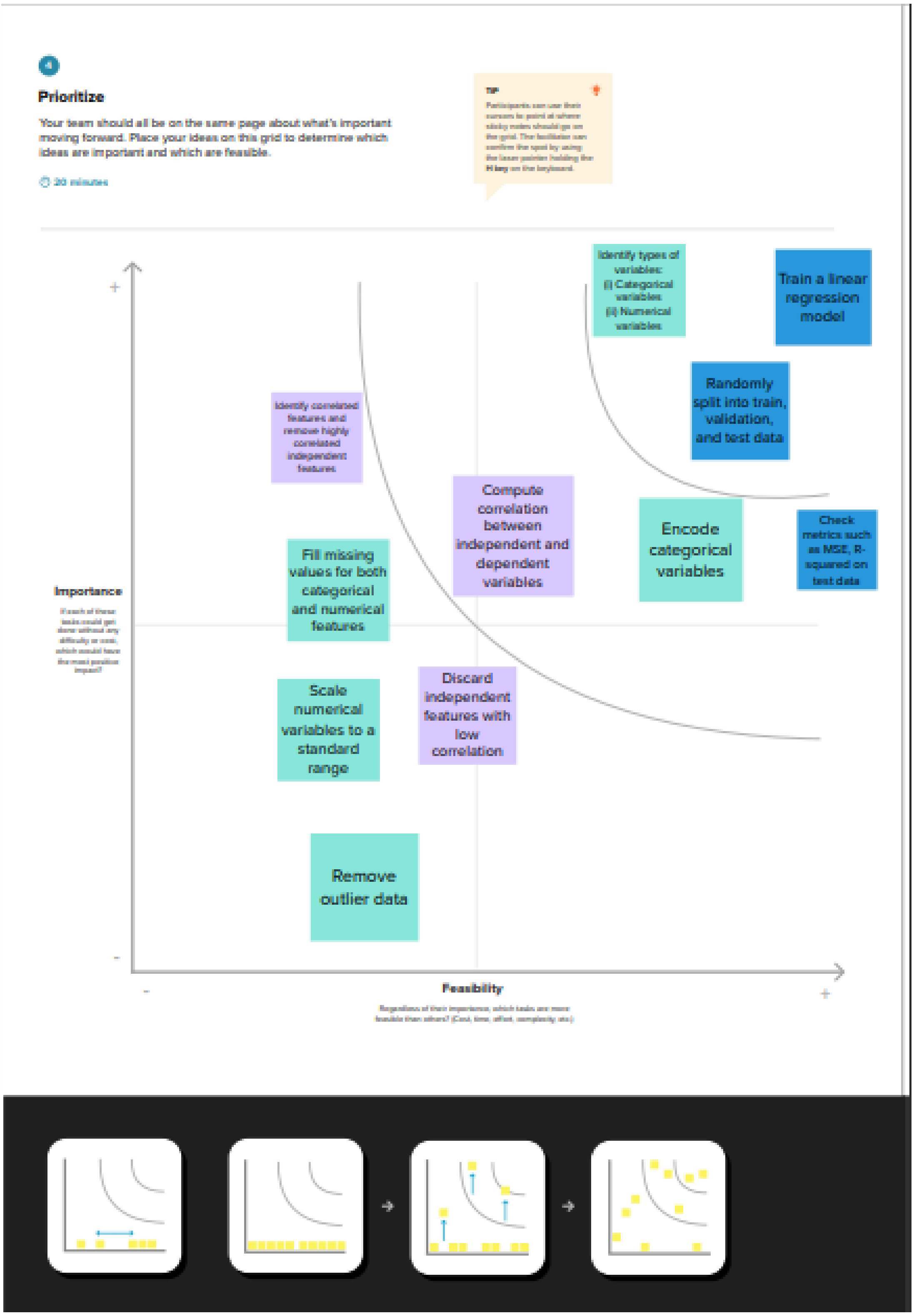
Ideation Phase:

Empathy Map Canvas



Brainstorm & Prioritize Ideas


Group ideas
Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.
20 minutes
Data Preprocessing
Identify types of variables: (i) Categorical variables (ii) Numerical variables
Encode categorical variables
Fill missing values for both categorical and numerical features
Scale numerical variables to a standard range
Remove outlier data
Feature Selection and Correlation Analysis
Identify correlated features and remove highly correlated independent features
Compute correlation between independent and dependent variables
Discard independent features with low correlation
Data Split and Model Training
Randomly split into train, validation, and test data
Train a linear regression model
Check metrics such as MSE, R-squared on test data


4
Prioritize
Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.
20 minutes
TIP
Importance
Feasibility
Identify types of variables:
(i) Categorical variables
(ii) Numerical variables
Train a linear regression model
Randomly split into train, validation, and test data
Identify correlated features and remove highly correlated independent features
Compute correlation between independent and dependent variables
Encode categorical variables
Check metrics such as MSE, R-squared on test data
Fill missing values for both categorical and numerical features
Scale numerical variables to a standard range
Discard independent features with low correlation
Remove outlier data

Project Design Phase:

Proposed Solution

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	How can we effectively employ machine learning to predict and optimize garment worker productivity in the garment industry? The textile and apparel sector is among the world's largest, with garment worker productivity being a key determinant of company success and profitability. This project aims to create a robust machine learning model for predicting worker productivity, utilizing a dataset encompassing various production attributes. The dataset includes factors like department, day, time allocation, incentives, idle time, and actual productivity. This predictive model has the potential to significantly impact manufacturing, HR, and supply chain management, enabling data-driven decisions to enhance efficiency, reduce costs, and maintain a competitive edge in the market. Moreover, it holds the promise of improving labour conditions and environmental sustainability in the industry.
2.	Idea / Solution description	The proposal aims to predict and optimize garment worker productivity in the textile and apparel industry by collecting and preprocessing data on production attributes, focusing on feature engineering, and developing machine learning models using techniques like regression, time series

		analysis, and deep learning. The models will be trained and evaluated, and a decision support system will be provided, along with training programs for workers.
3.	Novelty / Uniqueness	The proposed revenue model offers a flexible, client-cantered approach, offering various options like the Subscription Model, Pay-Per-Prediction Model, and Licensing Model. It caters to different needs and budgets, making it a versatile and customer-focused solution in the machine learning solutions market.
4.	Social Impact / Customer Satisfaction	The implementation of productivity improvement strategies in the garment industry through machine learning has a significant social impact by fostering skill development among workers. These strategies typically involve training and skill development programs, equipping the workforce with valuable competencies. As workers gain new skills, it not only enhances their immediate job performance but also has far-reaching effects on their long-term career prospects. The newfound skills and knowledge acquired in these programs can empower workers to explore opportunities beyond the garment industry, opening doors to a broader range of career options. This, in turn, can lead to higher earning potential, improved job security, and increased overall well-being for workers and their families. By investing in the skill development of garment workers, the industry not only boosts its own productivity and efficiency but also contributes to the socioeconomic development of the workforce, creating a positive ripple effect in less developed regions where the garment industry is a key economic driver.

5.	Business Model (Revenue Model)	Our business model revolves around providing a comprehensive workforce management software solution. We specialize in optimizing attendance tracking, forecasting per-worker output, and manpower requirements for

		manufacturing and service industries. Our key resources include skilled developers and analysts, supported by a robust database. We generate revenue through subscriptions and additional fees for customizable software features. Strategic partnerships with HR consultancies and tech providers bolster our expertise and market reach, ensuring accuracy and efficiency for our clients.
6.	Scalability of the Solution	The revenue model is designed for scalability, catering to businesses of different sizes. It employs a subscription-based model, pay-per-prediction option, licensing fees, and consulting services. This diverse approach ensures adaptability and long-term viability for garment manufacturing enterprises, offering them cost-effective and flexible solutions.

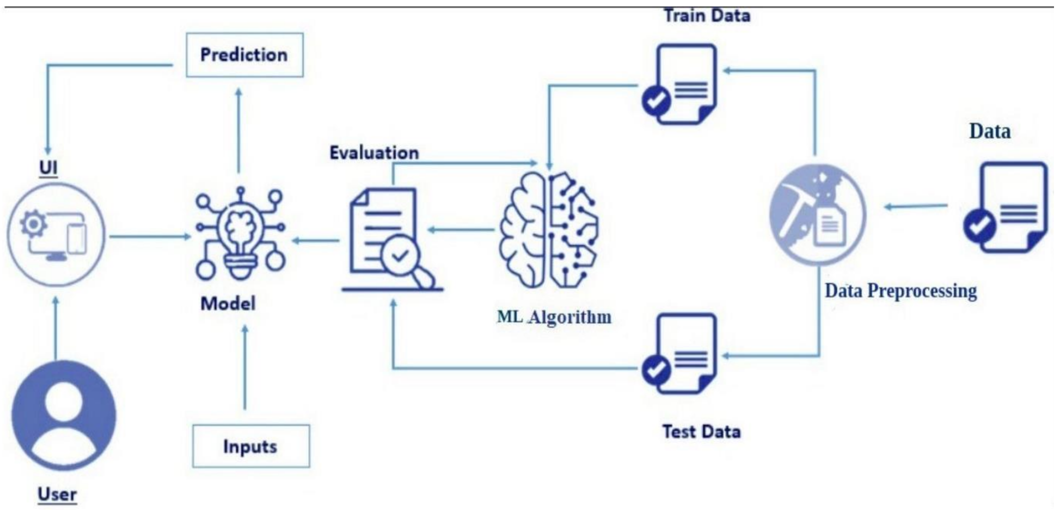
Solution Architecture

It optimizes the productivity prediction process by leveraging ANN, Linear regression and Random Forest. ANN enables complex pattern recognition and non-linear relationships to predict worker productivity, considering various interconnected factors in the production process. Linear regression is used to help identify the linear relationship between input variables and worker productivity, providing insights into how changes in specific variables impact productivity. Random Forest offers an ensemble learning method that leverages multiple decision trees, allowing for the analysis of various factors simultaneously and providing robust predictions for worker productivity.

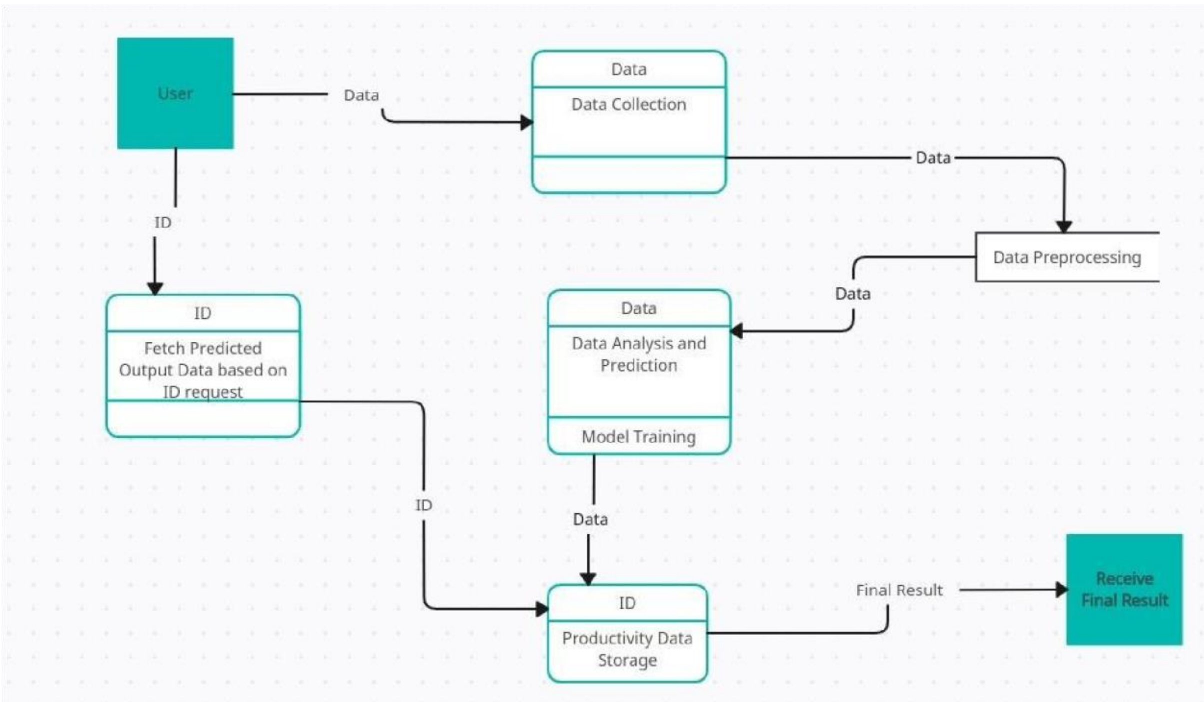
Our solution leverages Artificial Neural Networks (ANN),Linear regression and Random Forest to address the garment worker productivity problem effectively.

- Data Gathering
- Data Pre-processing
- Model Building
- Garment Worker Productivity Prediction
- Real Time Analysis

Solution Architecture Diagram



Data Flow Diagrams



User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register & access the dashboard with Gmail Login	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can register & access the dashboard with email and password	High	Sprint-1
Customer (Web user)	Registration	USN-6	As a web user, I can register for the application by entering my email, password, and confirming my password	I can enter my email, password, and confirm my password.	High	Sprint-1
		USN-7	As a web user, I will receive a confirmation email once I have registered for the application.	I can receive a confirmation email.I can click on the confirmation link in the email.	Medium	Sprint-1
		USN-8	As a web user, I can register for the application through Google.	I can register for the application using my Google credentials.	Medium	Sprint-1
	Login	USN-9	As a web user, I can log into the application by entering email & password.	I can enter my email and password.	Medium	Sprint-2
		US-10	As a web user, I can reset my password if I forget it. I can request a password reset.	I received an email with a link to reset my password. I can successfully reset my password.	High	Sprint-2
Customer Care Executive	User Management	US-11	As a Customer Care Executive, I can view user profiles.	I can access user profiles with relevant information.	High	Sprint-1
		US-12	As a Customer Care Executive, I can update user information.	I can edit and save user information as needed.	High	Sprint-1
	Ticket Management	US-13	As a Customer Care Executive, I can update the status of a support ticket.	I can change the status of a support ticket (e.g., open, resolved, in progress).	High	Sprint-2
Administrator	User Management	US-14	As an Administrator, I can create a new user account.	I can create a new user account with required information.	High	Sprint-1
	Dashboard Management	US-15	As an Administrator, I can customize the dashboard layout.	I can rearrange widgets on the dashboard. - Changes are saved and reflected for all users.	High	Sprint-2
	Reporting	US-16	As an Administrator, I can generate usage reports.	I can generate reports with relevant data and export options.	High	Sprint-2

Project Planning Phase

Technology Stack

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

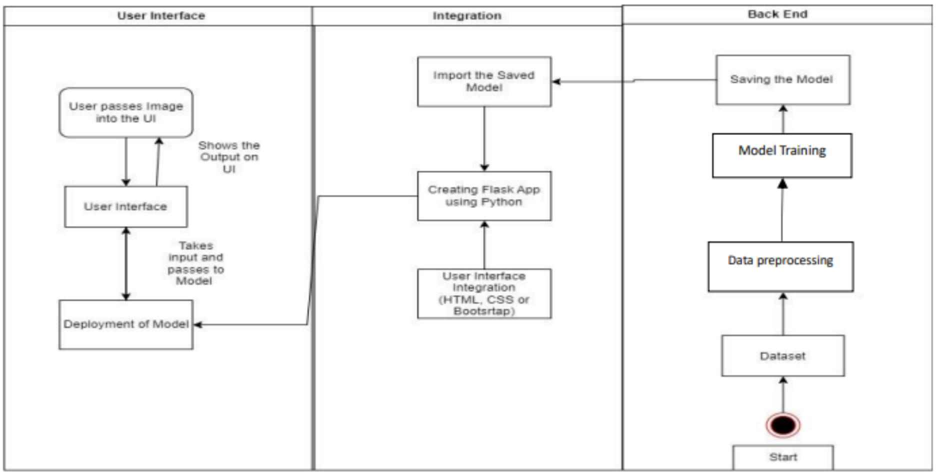


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application	HTML, CSS, JavaScript / Angular Js / React Js etc.
2.	Application Logic-1	Logic for a process in the application	Python
3.	Database	Collect the Dataset Based on the Problem Statement	File Manager, MySQL, etc.
4.	File Storage/ Data	File storage requirements for Storing the dataset	Local System, Google Drive Etc
5.	Frame Work	Used to Create a web Application, Integrating Frontend and Back End	Python Flask, etc
6.	Deep Learning Model	Purpose of Model	Linear Regression, Random Forest etc.
7.	Infrastructure	Application Deployment on Local System	Local

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Python's Flask
2.	Security Data	Encryption and secure access controls	Technology used.
3.	Scalability	Flexible 3-tier architecture and microservices	Technology used.
4.	Availability	Load balancing, disaster recovery plan	Technology used.
5.	Performance	The application is optimized for handling a substantial number of requests per second. It efficiently utilizes cache mechanisms and Content Delivery Networks (CDNs) for improved response times and reduced latency.	Technology used.

Project Planning Phase

Product Backlog, Sprint Schedule, and Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Project setup & Infrastructure	USN-1	Set up the development environment with the required tools and frameworks to start the garment worker productivity prediction using machine learning project.	10	Medium	Kashish
Sprint-2	development environment	USN-2	Gather a diverse dataset for training the model.	10	Medium	Shreya
Sprint-3	Data collection	USN-3	Preprocess the collected dataset by splitting it into training and validation sets.	1	High	Shivani
Sprint-3	data preprocessing	USN-4	Explore and evaluate different architectures to select the most suitable model for garment worker productivity prediction.	1	High	Kashish
Sprint-3	model development	USN-5	Train the selected model using the preprocessed dataset and monitor its performance on the validation set.	1	High	Rohan
Sprint-3	Training	USN-6	Implement data augmentation techniques to improve the model's robustness and accuracy.	2	High	Shreya
Sprint-4	model deployment & Integration	USN-7	Deploy the trained model as an API or web service to make it accessible for garment worker productivity prediction. integrate the model's API into a user-friendly web interface for prediction of garment worker productivity.	15	medium	Shivani
Sprint-5	Testing & quality assurance	USN-8	Conduct thorough testing of the model and web interface to identify and report any issues or bugs. fine-tune the model hyperparameters and optimize its performance based on user feedback and testing results.	15	medium	Rohan

Project Tracker, Velocity & Burndown Chart

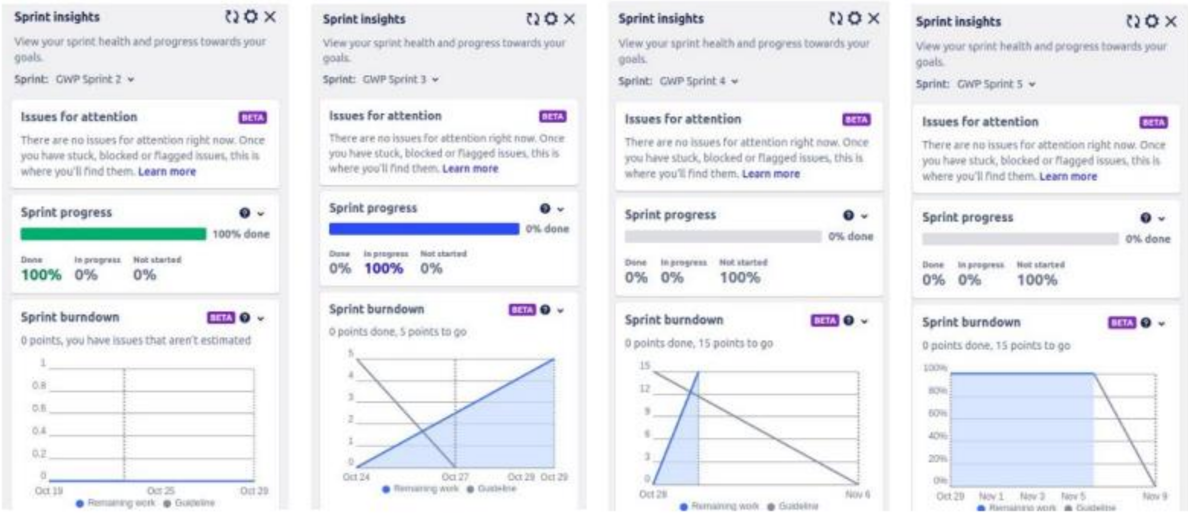
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Sprint Release Date (Actual)
Sprint-1	10	4 Days	15 Oct 2023	18 Oct 2023	15 Oct 2023
Sprint-2	10	5 Days	19 Oct 2023	23 Oct 2023	19 Oct 2023
Sprint-3	5	4 Days	24 Oct 2023	27 Oct 2023	26 Oct 2023
Sprint-4	15	10 Days	28 Oct 2023	06 Nov 2023	28 Oct 2023
Sprint-5	15	4 Days	06 Nov 2023	09 Sep 2023	06 Nov 2023

Velocity:

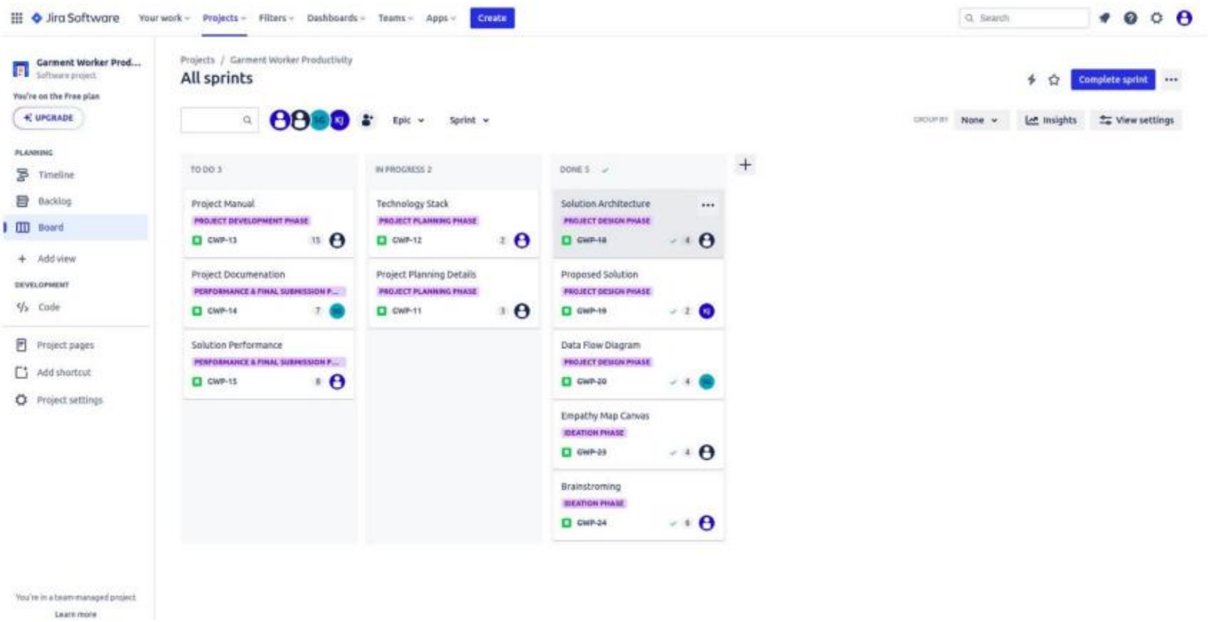
AV = sprint duration / velocity

AV = 25/11 = 2.27

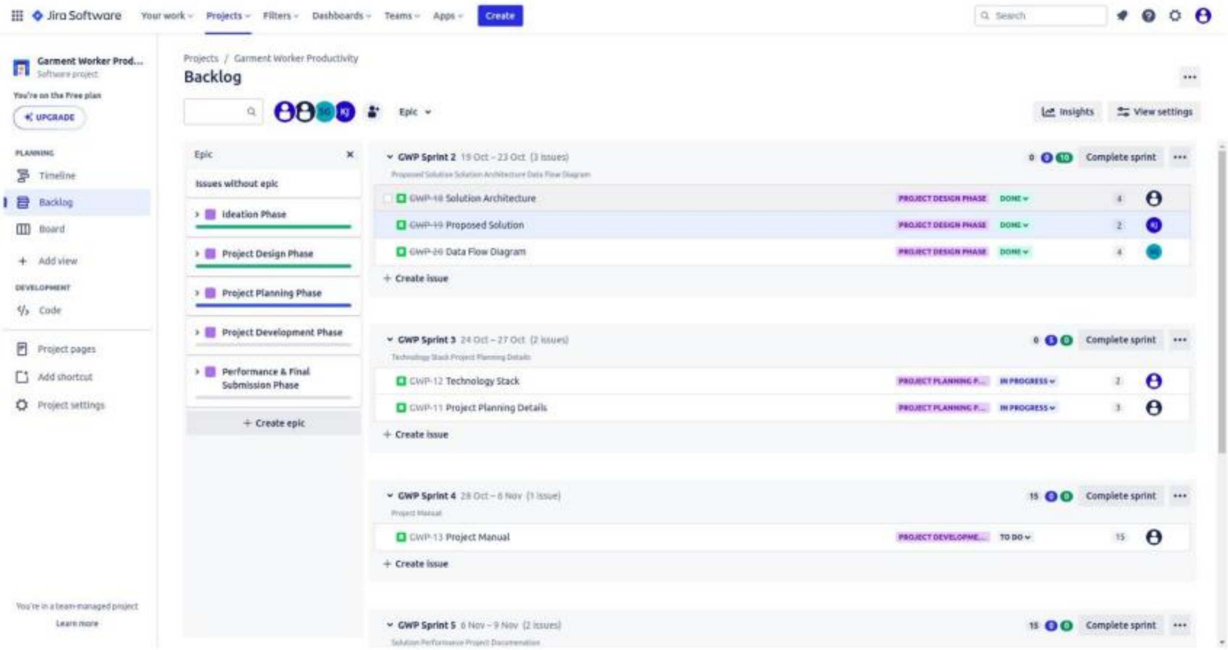
Burndown Chart:

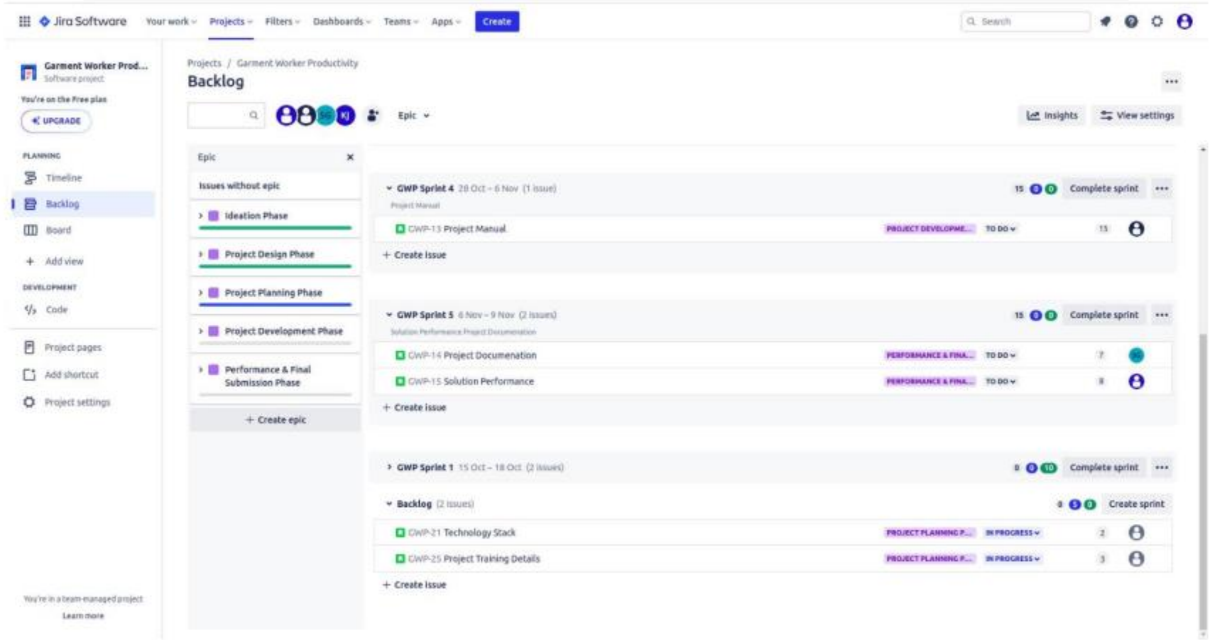


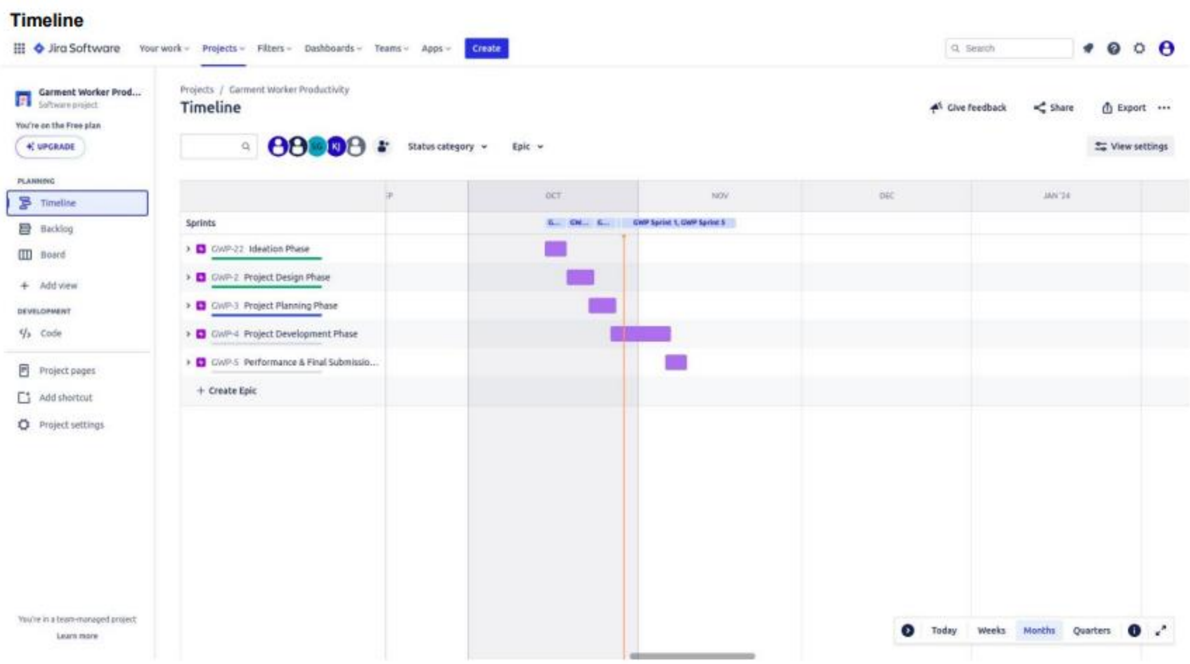
Board section



Backlog section







Project Development Phase

Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Values	Screenshot

1.	Metrics	**Regression Model:**
		RMSE score
			Linear Regression Model
			`#training score`
			`predict_train = lr.predict(X_train)`
			`mse = mean_squared_error(y_train, predict_train)`
			`rmse_lr_train = np.sqrt(mse)`
			`print('Root Mean Squared Error:', rmse_lr_train)`
			
			`#testing score`
			`predict_test = lr.predict(X_test)`
			`mse = mean_squared_error(y_test, predict_test)`
			`rmse_lr_test = np.sqrt(mse)`
			`print('Root Mean Squared Error:', rmse_lr_test)`
			
			Decision Tree Regressor Model
			`#training score`
			`predict_train_dtr = dtr.predict(X_train)`
			`mse = mean_squared_error(y_train, predict_train_dtr)`
			`rmse_dtr_train = np.sqrt(mse)`
			`print('Root Mean Squared Error:', rmse_dtr_train)`
			
			`#testing score`
			`predict_test_dtr = dtr.predict(X_test)`
			`mse = mean_squared_error(y_test, predict_test_dtr)`
			`rmse_dtr_test = np.sqrt(mse)`
			`print('Root Mean Squared Error:', rmse_dtr_test)`

Random Forest Regressor Model

```
#training score
predict_train_rfr = rfr.predict(X_train)
mse = mean_squared_error(y_train, predict_train_rfr)
rmse_rfr_train = np.sqrt(mse)
print('Root Mean Squared Error:', rmse_rfr_train)
```

```
#testing score
predict_test_rfr = rfr.predict(X_test)
mse = mean_squared_error(y_test, predict_test_rfr)
rmse_rfr_test = np.sqrt(mse)
print('Root Mean Squared Error:', rmse_rfr_test)
```

Gradient Boosting Regressor Model

```
#training score
predict_train_gbr = gbr.predict(X_train)
mse = mean_squared_error(y_train, predict_train_gbr)
rmse_gbr_train = np.sqrt(mse)
print('Root Mean Squared Error:', rmse_gbr_train)
```

```
#testing score
predict_test_gbr = gbr.predict(X_test)
mse = mean_squared_error(y_test, predict_test_gbr)
rmse_gbr_test = np.sqrt(mse)
print('Root Mean Squared Error:', rmse_gbr_test)
```

Extreme Gradient Boost Regressor Model

```
#training score
predict_train_xgb = xgb.predict(X_train)
mse = mean_squared_error(y_train, predict_train_xgb)
rmse_xgb_train = np.sqrt(mse)
print('Root Mean Squared Error:', rmse_xgb_train)
# 0.0575744...
```

```
#testing score
predict_test_xgb = xgb.predict(X_test)
mse = mean_squared_error(y_test, predict_test_xgb)
rmse_xgb_test = np.sqrt(mse)
print('Root Mean Squared Error:', rmse_xgb_test)
```

Bagging Regressor Model

```
#Evaluate performance
y_train_pred = bagging_reg.predict(X_train)
y_test_pred = bagging_reg.predict(X_test)
train_rmse_b = np.sqrt(mean_squared_error(y_train, y_train_pred))
test_rmse_b = np.sqrt(mean_squared_error(y_test, y_test_pred))

print("Bagging regressor:")
print(f"Training RMSE: {train_rmse_b}")
print(f"Testing RMSE: {test_rmse_b}")
```

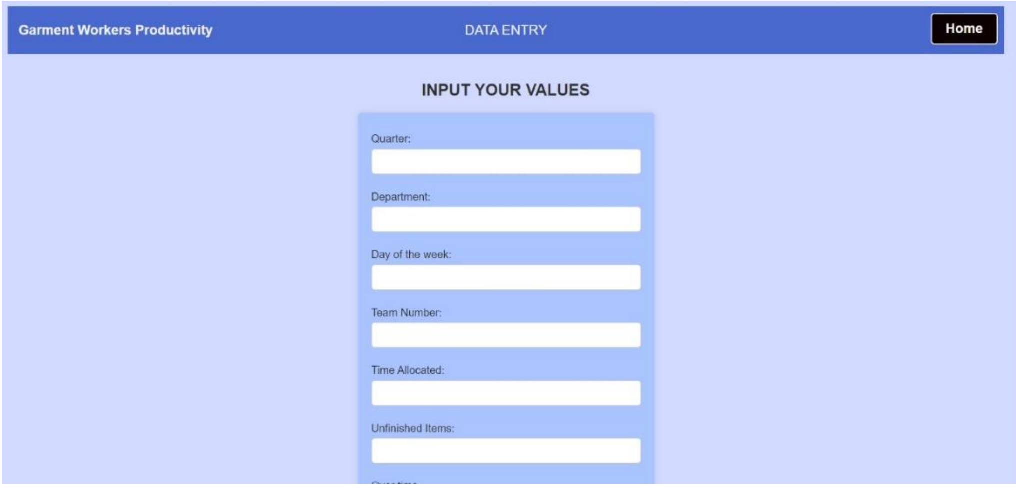
Boosting Regressor Model

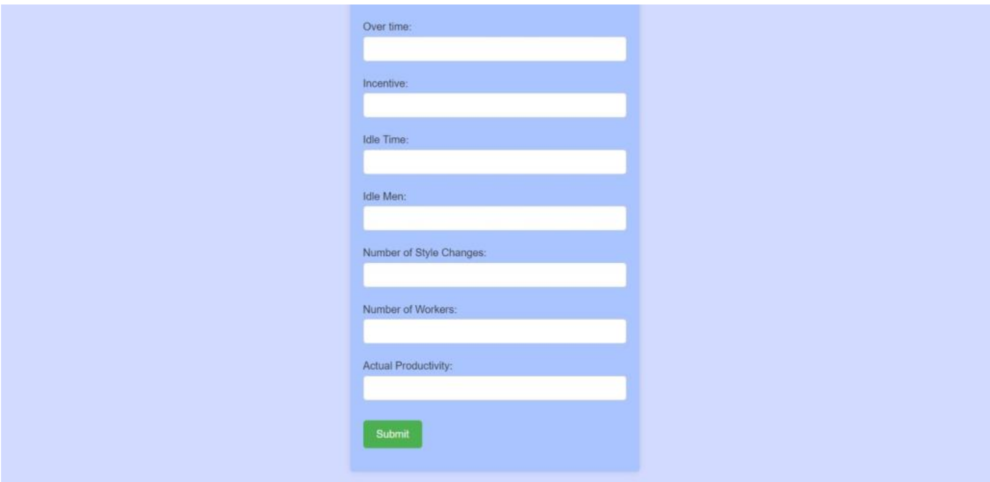
```
#Evaluate performance
y_train_pred = boosting_reg.predict(X_train)
y_test_pred = boosting_reg.predict(X_test)
train_rmse_bo = np.sqrt(mean_squared_error(y_train, y_train_pred))
test_rmse_bo = np.sqrt(mean_squared_error(y_test, y_test_pred))

print("AdaBoost Regressor:")
print(f"Training RMSE: {train_rmse_bo}")
print(f"Testing RMSE: {test_rmse_bo}")
```

Results

Output Screenshots


Garment Workers Productivity
DATA ENTRY
Home
INPUT YOUR VALUES
Quarter:
Department:
Day of the week:
Team Number:
Time Allocated:
Unfinished Items:


Over time:
Incentive:
Idle Time:
Idle Men:
Number of Style Changes:
Number of Workers:
Actual Productivity:
Submit


Garment Workers Productivity
Home
Team Details
Garment Worker Productivity Prediction
Team ID:- 592545
21BCE0440 - ROHAN RAI
21BCE0484 - SHIVANI DASH
21BCE0821 - SHREYA GUPTA
21BCE3204 - KASHISH JAIN

Advantages & Disadvantages

Advantages:

- **Improved Efficiency**: The model can optimize workforce management, reducing idle time and increasing overall productivity.
- **Cost Savings:** Higher productivity and reduced idle time can lead to lower production costs, contributing to increased profitability.
- **Quality Enhancement:** Identifying areas for process improvement can lead to better product quality, enhancing the reputation of the manufacturing company.
- **Strategic Insights:** The model provides valuable insights for manufacturers to make data-driven decisions, enhancing strategic planning and operations.
- **Employee Satisfaction:** By addressing factors affecting productivity, the model can contribute to improved working conditions and job satisfaction among garment workers.

Disadvantages:

- **Data Quality Dependency:** The accuracy of predictions relies heavily on the quality of input data. Inaccurate or biased data may lead to flawed predictions.
- **Privacy Concerns:** Collecting and using data on workers may raise privacy concerns. Ensuring compliance with privacy regulations is crucial.
- **Ethical Considerations:** Decisions based on the model's predictions could impact workers' lives. Ensuring ethical use of the model and avoiding unintended consequences is essential.
- **Implementation Costs:** Developing, deploying, and maintaining the machine learning model, along with any required infrastructure, may involve significant upfront and ongoing costs.

Conclusion

In conclusion, the development of the garment worker productivity prediction model represents a strategic initiative with far-reaching implications for both the social and business dimensions of the garment manufacturing sector. Through the application of advanced machine learning methodologies, this project endeavors to optimize workforce management, reduce idle time, and amplify overall productivity, thereby fostering a more competitive and economically robust industry.

The model's potential to discern areas for process refinement, curtail production costs, and elevate product quality underscores its pivotal role in shaping the industry's operational landscape. Nevertheless, the project acknowledges and addresses challenges such as data quality assurance, privacy compliance, and ethical considerations as integral components of its implementation strategy.

Looking ahead, the incorporation of emerging machine learning techniques, exploration of diverse data sources, and sustained collaboration among stakeholders are imperative for the model's enduring relevance and efficacy. By navigating these considerations judiciously, the project aspires not only to redefine workforce management practices in the garment industry but also to contribute to a paradigm of sustainability, ethical governance, and enduring prosperity for both enterprises and the workforce they engage.

Future Scope

The future scope of this garment worker productivity prediction project is characterized by a trajectory of continuous improvement and expansion. One avenue for advancement involves delving into advanced machine learning techniques, such as deep learning and reinforcement learning, to further refine the model's predictive accuracy and capabilities. Additionally, the integration of diverse data sources, including real-time environmental data and insights from wearable technology, presents an opportunity to enrich the model's understanding of the multifaceted factors influencing worker productivity. Establishing a system for continuous model optimization ensures its adaptability and efficacy over time. The global deployment of the model, benchmarked across various manufacturing environments, could ascertain its applicability on a broader scale. Collaborating with industry stakeholders, implementing user feedback mechanisms, and exploring extensions to related industries contribute to a holistic approach for sustained relevance. Incorporating an ethical AI framework ensures that the model aligns with evolving industry standards and maintains transparency and fairness in its predictions. In essence, the future trajectory of this project envisions not only technological advancements but also a lasting impact on workforce management practices across industries.

Appendix

In the appendix, we explore the broader impact of our garment worker productivity prediction model. Socially, the model could revolutionize workforce management, identifying ways to improve conditions, incentives, and training for garment workers. This could result in increased job satisfaction and better working conditions. On the business front, the model's potential is equally compelling, offering manufacturers opportunities to optimize workforce strategies, reduce idle time, and enhance overall productivity. The anticipated outcomes include increased profits, reduced production costs, and improved product quality. The model's knack for pinpointing areas for process improvement and refining supply chain efficiency adds an extra layer of value, making it a dual force for positive change in both individual lives and the industry landscape.