YEDİTEPE UNIVERSITY
Following Atatürk's Renaissance

COMPUTER ENGINEERING DEPARTMENT

# CSE464 Term Project

## INTRODUCTION TO DATA SCIENCE & BIG DATA ANALYTICS (AUTUMN SEMESTER 2024)

## GROUP 8

**MEHMET ÖNAL**
**20200702081**

**ONAT RİBAR**
**20210702099**

# 1. Background Information

## i. Business Information

The global garment industry is a significant contributor to the world's economy, with a market value estimated at $1.79 trillion in 2024, which accounts for approximately 1.63% of the global GDP. This sector encapsulates a wide range of activities, from textile production to the design, manufacturing, and distribution of clothing and apparel. In 2022, the apparel manufacturing industry in the United States employed approximately 93,000 workers, thus highlighting its role as a major employer.

The industry is characterized by its labor-intensive nature, with millions of workers involved worldwide. For instance, the textiles and garment sectors employed approximately 91 million workers globally in 2022. Despite its vast scale, the industry faces challenges such as overproduction, which can hinder creativity and devalue quality craftsmanship. Additionally, issues related to labor conditions and sustainability are prevalent, which prompts calls for more ethical and transparent practices.

Technological advancements and shifting consumer preferences are driving changes within the industry. Furthermore, the COVID-19 pandemic has impacted global supply chains, leading to a reevaluation of manufacturing and distribution strategies.

## ii. Organization Challenge

There exist several challenges within the industry due to its labor-intensive nature and reliance on manual processes. Meeting global demand requires precise production scheduling and optimized team performance, yet productivity often fluctuates due to factors like idle time, frequent style changes, and excessive overtime. Additionally, managing financial incentives effectively to motivate workers without increasing costs is a never ending issue. Tracking individual and team-level productivity is further complicated by the manual nature of operations, which makes it difficult to identify inefficiencies.

The company is a leading player in the global garment manufacturing industry. As a highly labor-intensive organization, the company depends on skilled workers and efficient production systems to meet the increasing global demand for garment products. With manual processes forming a significant part of its operations, employee performance is a critical factor in ensuring timely delivery and maintaining high-quality standards for customers. The company’s inability to proactively address potential bottlenecks is hindering its ability to operate efficiently and meet delivery deadlines.

## 2. Problem Summary/Definition

The garment industry is inherently reliant on its workforce, as manual processes form the backbone of its production system. This dependency on blue-collar workers introduces significant vulnerabilities, as team and individual productivity directly impact overall success and efficiency of operations.

The challenges are reflected by the limited availability of data-driven tools available to the business administrators. Reliance on manual tracking methods leaves managers unable to address systematic inefficiencies or accurately measure the contribution of individual workers and teams. This absence of business insights make it difficult to incentivize high-performing employees, identify underperforming teams, or determine when administrative actions are necessary. Moreover, the inability to predict performance creates challenges in areas such as resource and workforce allocation, prediction of product delivery dates, and even recruitment planning. Current analytical tools are insufficient to ensure sustainable and consistent production.

Weak and inconsistent production cycles are a potentially recurring issue, which would stem from factors such as idle time, mismanagement of financial incentives (e.g., limited compensation), and unbalanced workload across teams and individuals. Without a clear understanding of the relationships between such metrics, efforts to boost productivity will result in diminishing returns. For instance, while financial incentives may motivate workers, the lack of tangible data linking incentives to productivity outcomes may lead to wasted resources and misguided policies.

To address these challenges, it is crucial to adopt data-reliant solutions that provide actionable insights into worker performance and team dynamics. Ultimately, the company requires the utilization of historical data for descriptive and predictive analysis to achieve proactive management of resource allocation, workforce productivity, and performance bottlenecks. Implementing these solutions will enable the company to build a more resilient and effective production system, resulting in improved business value, operational KPIs, and a competitive advantage within the global garment industry.

## 2.1 Problem Analysis

```
data['diff']= data['actual_productivity'] - data['targeted_productivity']
data['diff'].describe()
```

| | diff |
|---|---|
| count | 1160.000000 |
| mean | -0.001914 |
| std | 0.156802 |
| min | -0.561958 |
| 25% | -0.023394 |
| 50% | 0.000431 |
| 75% | 0.073319 |
| max | 0.644375 |

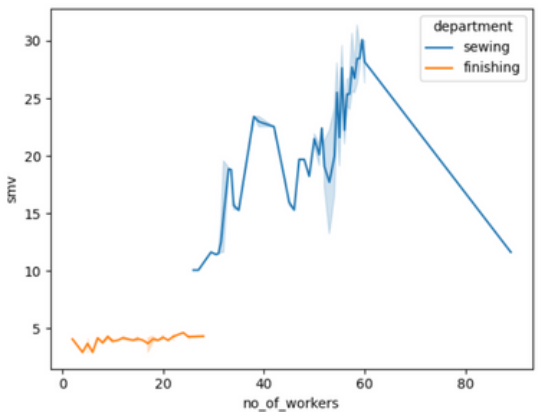
The purpose of taking this difference is to see the deviation between the actual and targeted productivity and to prepare the data for use in modeling this deviation.
The data['diff'] column represents the difference between the productivity target and the actual productivity, and the analysis of this difference is critical to identify improvement opportunities in the production process. This value can be an important indicator in future forecasts, analysis or performance evaluations.

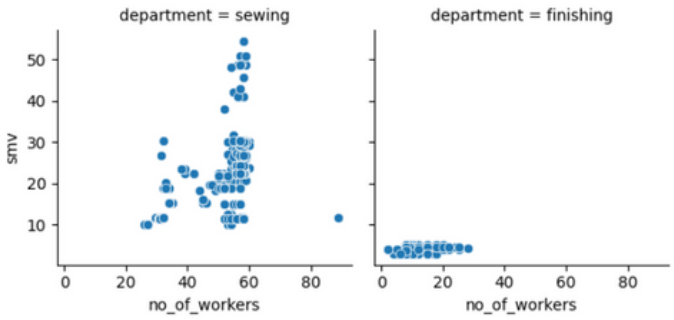
- mean: -0.001914 The mean value is -0.001914, meaning that the difference between actual productivity and targeted productivity is slightly negative on average. This generally indicates that productivity is falling short of target, but the difference is not very large. Since the mean difference is close to zero, it may indicate that there is no significant deviation from the productivity target.

- min: -0.561958 The smallest difference is -0.561958. This indicates that in some cases, actual productivity is well behind target (around 56%). This indicates that productivity is seriously low in some production processes.

- 25% (1st Quartile): -0.023394 It indicates that 25% of the data is below this value. That is, in the first quartile (low values), the difference between actual productivity and targeted productivity is negative and slightly below target with a small deviation on average.

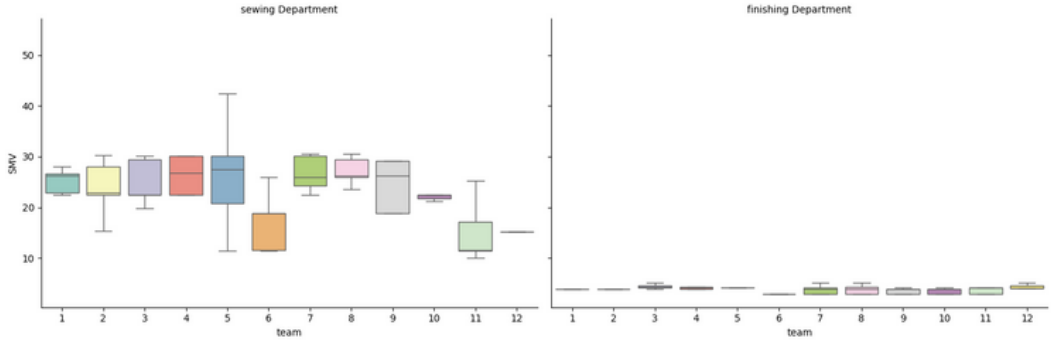
---



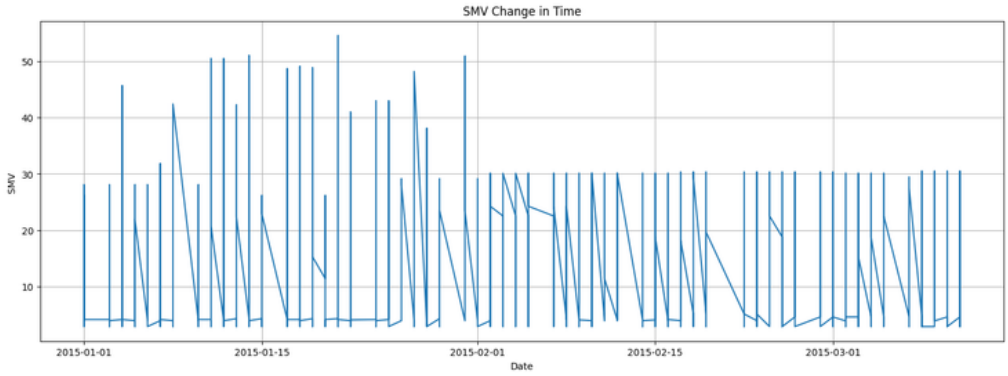

The garment production process exhibits significant variability in operational efficiency across different departments, particularly in how the number of workers correlates with the Standard Minute Value (SMV). The analysis reveals that in the sewing department, an increase in the number of workers is associated with a corresponding rise in the SMV, indicating that the complexity or scale of tasks increases with the workforce size. This can be seen in the line plot that showcases this trend, with a substantial increase in the SMV as the number of workers reaches the 60 mark. Furthermore, the scatter plot reinforces this pattern, showing concentrated data points with higher SMV values among teams with more workers. This trend implies that the sewing department struggles to maintain efficiency as team sizes expand, possibly due to logistical challenges or difficulties in coordination.

## 2.1 Problem Analysis



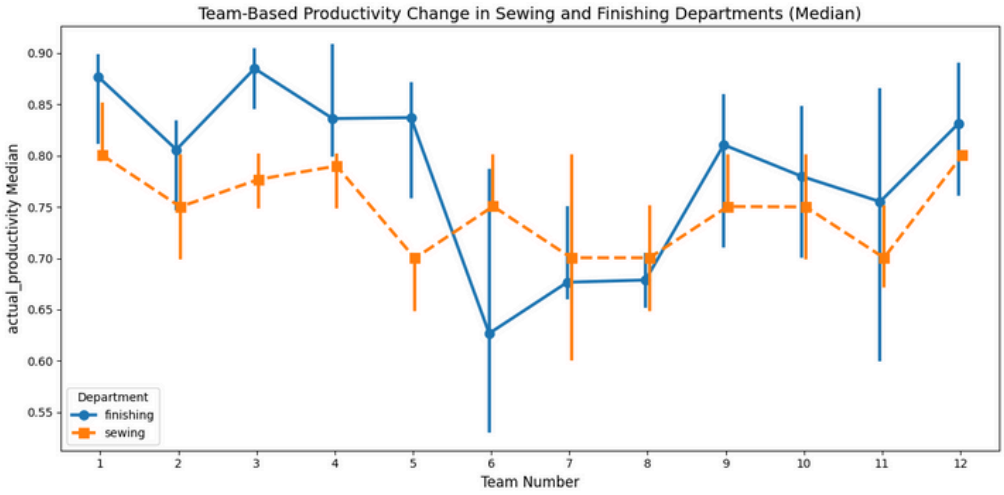

The box plots show that the time it takes for teams to make things (SMV) is very different, especially in the sewing part of the factory. Some sewing teams take much longer than others. It's like some teams take way more time to do the same thing, and some teams are not as consistent with their time. For example team 5 takes more time to make an item than team 12. This shows that in the sewing area, some teams have problems being consistent with their production. But, in the finishing part, all the teams are pretty much the same, and don't take much time to finish the work. This difference shows us that the sewing teams need help being more consistent. This company needs to figure out why some sewing teams take longer and need a system to help them be more regular, so they all take about the same time no matter which team is doing the job.

---



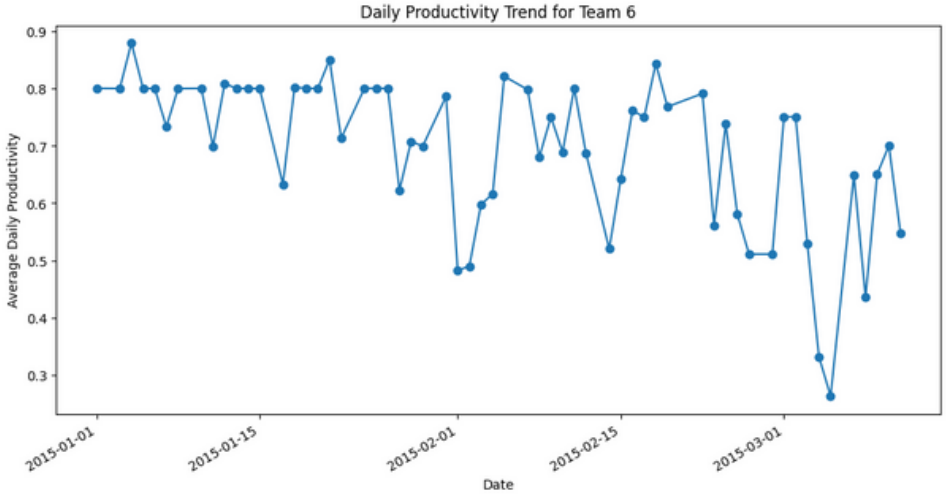

The line chart shows that the time it takes to make things (SMV) changes a lot over time. If we look at the chart before February 2015, we can see that the time goes up and down very quickly. Sometimes the time to make things is very high, and sometimes it is very low. This shows that the production process was very up and down before February 2015. Then, after February 2015, the chart shows that the time it takes to make things becomes more consistent. It stays at a more regular level. This means that the factory has a problem with making things at a stable rate and that before February 2015 there was a big issue of inconsistency. The factory needs to find out what was causing the time to change so much before February 2015, and to stop those problems from happening again. They also need a system in place that can react to fluctuations when they occur. This would allow them to prevent this instability and maintain a more steady production process.

## 2.1 Problem Analysis




The analysis of team-based productivity reveals significant differences and weaknesses in operational performance, particularly in the sewing department. While the finishing department maintains a relatively consistent average productivity across the various teams, the sewing department exhibits significant fluctuations. In particular, team 6 experiences a significant drop in average productivity followed by a rapid recovery, suggesting that some teams are particularly sensitive to performance changes. This sharp drop and the observed variability suggest that certain teams may be experiencing inconsistencies in their work processes or resource utilization. We focused on team 6, which experienced a sharp drop in productivity, because we saw little overall difference between targeted productivity and actual productivity at the beginning of our analysis.

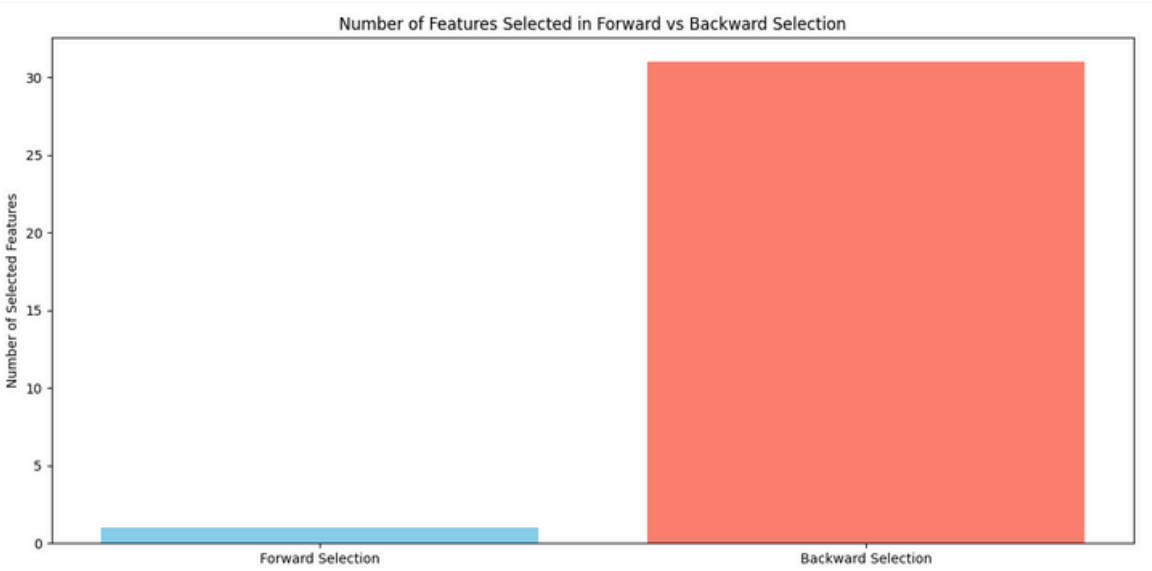
---



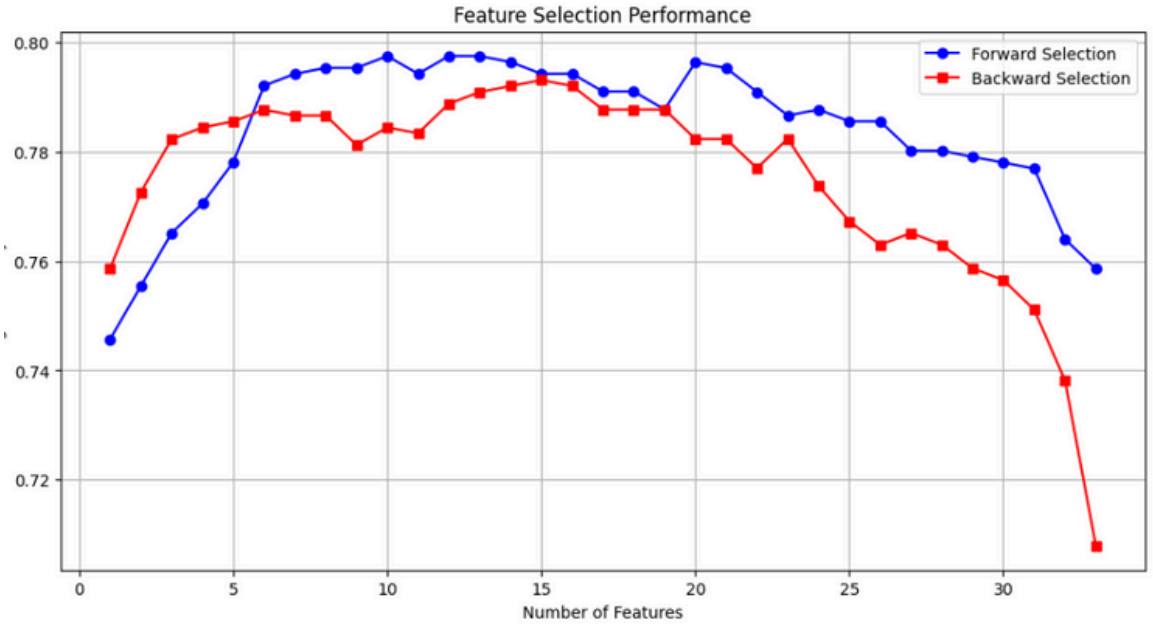

Team 6's daily productivity analysis reveals high volatility and unpredictability in production. The graph shows significant instability in daily output, with frequent dips and rebounds. There is a particularly severe drop in productivity starting around 2015-03-01. This highlights a weakness in Group 6's production systems and their inability to deliver consistent performance. This unpredictability, especially the sharp decline after 2015-03-01, makes it difficult to meet targets and manage resources effectively, and highlights the need for predictive analytics and data-driven management strategies to proactively address these fluctuations.

## 2.2 Feature Selection For Prediction

### Linear Regression Forward And Backward Selection

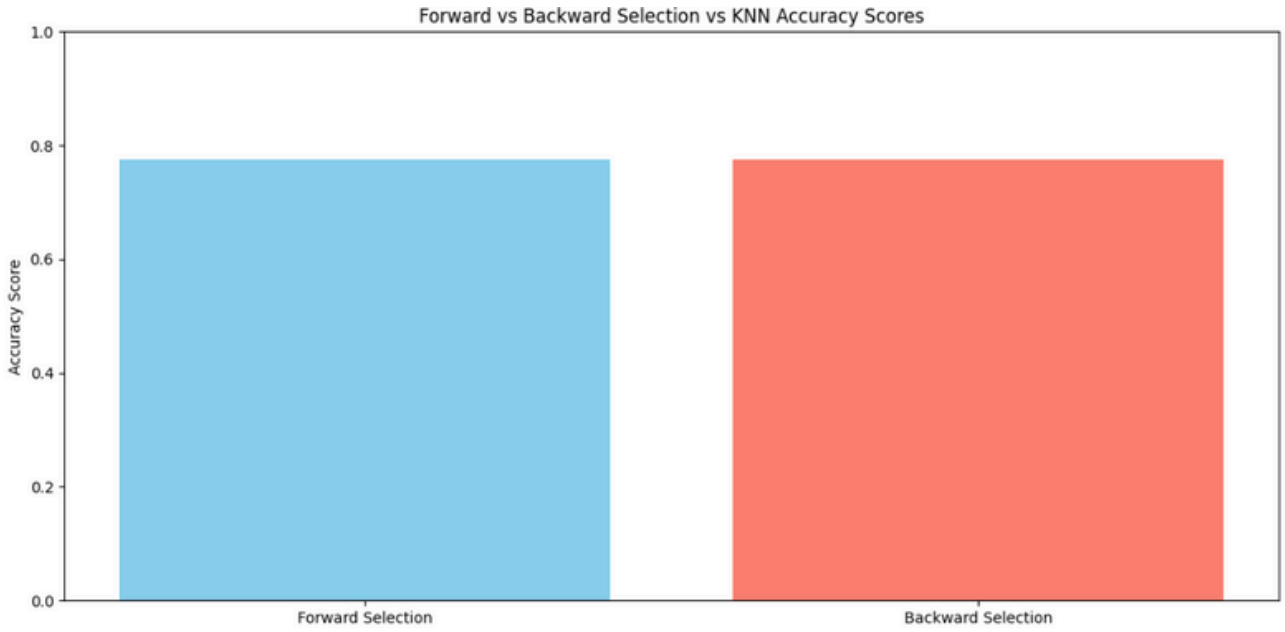



### Decision Tree Forward And Backward Selection



## 2.2 Feature Selection For Prediction

### Linear Regression Forward And Backward Selection




### Results For Three Different Classification Models

| | Model | Accuracy_Train | Accuracy_Test | Precision | Recall | F1_Score |
|---|---|---|---|---|---|---|
| 0 | Logistic Regression (Forward) | 0.728448 | 0.728448 | 0.530637 | 0.728448 | 0.614004 |
| 1 | Logistic Regression (Backward) | 0.728448 | 0.728448 | 0.530637 | 0.728448 | 0.614004 |
| 2 | Decision Tree (Forward) | 0.719828 | 0.719828 | 0.726978 | 0.719828 | 0.723071 |
| 3 | Decision Tree (Backward) | 0.758621 | 0.758621 | 0.761073 | 0.758621 | 0.759788 |
| 4 | K-Nearest Neighbors (Forward) | 0.775862 | 0.775862 | 0.758900 | 0.775862 | 0.754729 |
| 5 | K-Nearest Neighbors (Backward) | 0.775862 | 0.775862 | 0.770931 | 0.775862 | 0.772154 |

# 3. Solutions/Recommendations/Decisions

## 1. Data-Driven Productivity Improvements

- Real-Time Monitoring Systems:
  - Introduce automated tools to track team and individual productivity in real-time.
  - Provide managers with immediate insights to address problems and reduce idle time effectively.
- Historical Data Analysis:
  - Analyze historical data to identify productivity patterns and predict potential drops.
  - Proactively manage these issues to allocate resources effectively and prevent production bottlenecks.

## 2. Team-Level Optimization

- Balanced Workload Distribution:
  - Evaluate workload distribution across teams to ensure equal allocation of tasks.
  - Support teams like Team 6, which showed significant volatility, by providing additional training and resources.
- Targeted Support for Underperforming Teams:
  - Address stability issues in teams such as Team 5 through specialized training programs and improved inter-department communication.
  - Equip teams with the necessary resources to resolve challenges promptly.

## 3. Workforce Management and Employee Development

- Incentive Alignment with Performance:
  - Design financial incentives that reward high-performing individuals and teams.
  - Provide constructive feedback to underperformers to encourage improvement.
- Flexible Team Structures:
  - Enable the reassignment of skilled workers to teams facing immediate challenges to mitigate productivity drops.
- Continuous Training Programs:
  - Offer ongoing training to help workers adapt to new technologies and processes.
  - Ensure teams remain competitive and efficient.
- Feedback Mechanisms:
  - Create open communication channels for employees to share challenges and suggest improvements.
  - Foster a culture of collaboration and accountability.

# 4. Solutions/Recommendations/Decisions

### 4. Process and Time Efficiency

- Standardization of Task Timings:
  - Address inconsistencies in Standard Minute Values (SMV) by standardizing task timings where feasible.
  - Reduce variability and streamline operations.
- Support for Larger Teams:
  - As demonstrated by the correlation between worker numbers and SMV, provide additional training and resources to larger teams to manage increased complexity effectively.

### 5. Technological Advancements

- Predictive Models: The company should develop predictive models using historical data to anticipate when SMV might fluctuate and when productivity might drop. Based on the models, managers can proactively allocate resources to minimize disruptions.
- Choosing the Right Models: From recent model evaluations, the K-Nearest Neighbors (Forward) model performs well, with a test accuracy of 77.59%. Decision Tree (Backward) model also has a comparable performance, with a test accuracy of 75.86%. Considering the slightly higher test accuracy, the K-Nearest Neighbors (Forward) model would be a suitable choice for predictive analytics in resource allocation and productivity monitoring.
- Automation: To reduce human error and enhance productivity, the company should invest in automation for repetitive tasks. This will streamline operations and reduce the potential for mistakes.

### 6. Performance and Resource Optimization

- Regular Productivity Audits:
  - Conduct periodic audits to evaluate team performance and operational bottlenecks.
  - Use audit findings to refine strategies and implement continuous improvements.
- Demand Forecasting:
  - Utilize demand forecasting to align resources with predicted production needs, minimizing waste and overproduction risks.

# 5. Follow-Up & Evaluation Plan

1. **Implementation Review**

- Regularly review the effectiveness of implemented strategies (e.g., real-time monitoring tools, incentive programs, and workload redistribution) through monthly performance reports.
- Collect feedback from managers and employees to assess operational changes' impact on productivity and morale.

**2.Performance Metrics Tracking**

- Establish key performance indicators (KPIs) such as:
  - Productivity rate per team/individual
  - Error rates and downtime
  - Training program participation and outcomes
- Compare current productivity levels with baseline metrics to measure improvement.

**3.Predictive Model Validation**

- Periodically re-evaluate the K-Nearest Neighbors (Forward) model and Decision Tree (Backward) model to ensure consistent accuracy in predicting productivity drops.
- Update models with recent data to improve predictions and reflect changing production dynamics.

**4.Employee Engagement and Feedback Mechanisms**

- Conduct quarterly surveys to gauge employee satisfaction with the implemented changes.
- Hold focus group discussions to refine processes based on employee insights.

**5.Audit and Continuous Improvement**

- Schedule biannual productivity audits to identify new bottlenecks and operational inefficiencies.
- Leverage audit results to update strategies and incorporate best practices.

**6.Technology Integration and Monitoring**

- Monitor the impact of automation on productivity and quality.
- Conduct regular training sessions to help workers adapt to new technologies and minimize resistance to change.

**7.Resource Allocation Adjustment**

- Evaluate the distribution of resources and adjust team structures or resource allocation based on demand forecasts and productivity data.
- Address resource shortages promptly to prevent disruption in production.

**8.Long-Term Outcome Evaluation**

- Assess the overall impact of the initiatives on the company’s profitability and production efficiency over a 12-month period.
- Report findings to senior management to guide future decision-making and strategic planning.

# 6. References And Links

- **Dataset:** https://archive.ics.uci.edu/dataset/597/productivity+prediction+of+garment+employees
- **Google Colab**: https://colab.research.google.com/drive/1gniJoDi2JFq_k36ANl_JVdO69o8kMLax?usp=sharing#scrollTo=0brRZK0nEKLv