PROJECT REPORT

A predictive model to identify factors affecting defect rates in Additive Manufacturing and forecast defects accurately.

Made By: Shreyansh Verma

AIM:

This project aims to create a predictive model to identify factors affecting defect rates in **Additive Manufacturing** and forecast defects accurately. By using advanced data analysis and machine learning, it provides actionable insights for optimizing production processes, enhancing product quality, and minimizing operational inefficiencies.

- *Source of the Dataset:* https://www.kaggle.com/datasets/rabieelkharoua/predicting-manufacturing-defects-dataset
- *Credits:* ***Rabie El Kharoua****, this dataset is original and has never been shared before. It is* ***made available under the CC BY 4.0 license****, allowing anyone to use the dataset in any form as long as proper citation is given to the author.*

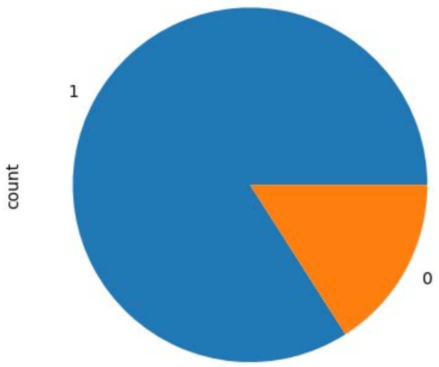
Importance of the Project:

- **Better Print Quality:** Defects like warping or weak layer adhesion in 3D printing can ruin parts. Predicting these issues helps adjust settings and improve quality.
- **Saves Money and Resources:** Materials and energy are expensive in 3D printing. By reducing waste and downtime, this project cuts costs significantly.
- **Faster Prototyping:** Custom and prototype designs can meet specs the first time, saving time and speeding up product development.
- **Energy Efficiency:** This project can help identify ways to use energy more efficiently, which is crucial for sustainable 3D printing.
- **Scalable Production:** Ensuring consistent quality across mass production in 3D printing becomes easier with defect prediction.
- **Better Use of Advanced Materials:** Predicting issues helps make the most of specialized, often expensive, materials like composites or biomaterials.
- **Meets Industry Standards:** High-quality, reliable parts are critical in sectors like aerospace and healthcare. Fewer defects mean meeting strict regulations with ease.

- **Boosts Innovation:** Companies using predictive analytics in 3D printing can deliver better products faster, staying ahead in the market.

Snapshot of Dataset:

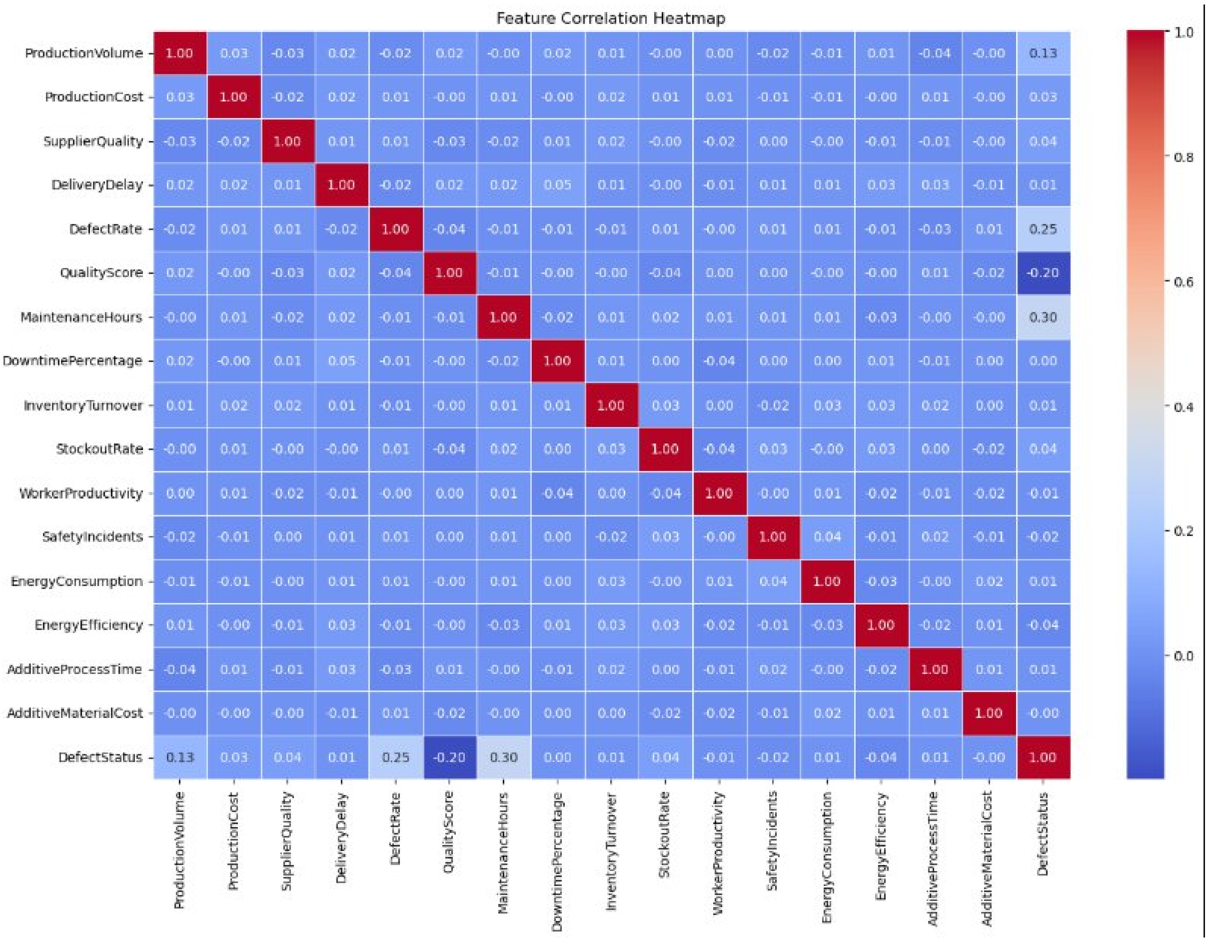
- **Production Volume** : Represents the total units produced per batch.
- **Production Cost**: Cost associated with production.
- **Supplier Quality**: Rating of supplier quality (0-100 scale).
- **Delivery Delay**: Delivery delays in days.
- **Defect Rate**: Percentage of defective items.
- **Quality Score**: A composite measure of production quality (0-100 scale).
- **Maintenance Hours**: Hours spent on maintenance.
- **Downtime Percentage**: Percentage of downtime during production.
- **Inventory Turnover**: Inventory turnover ratio.
- **Stockout Rate**: Stockout percentage.
- **Worker Productivity**: Worker productivity score (0-100 scale).
- **Safety Incidents**: Count of safety incidents.
- **Energy Consumption**: Energy used in kWh.
- **Energy Efficiency**: Efficiency score for energy usage.
- **Additive Process Time**: Time spent in additive manufacturing processes.
- **Additive Material Cost**: Cost of materials used in additive processes.
- **Defect Status**: Target variable, indicating whether defects occurred (1) or not (0).



- *The dataset focuses on defect instances more because they do not occur often. However, non-defect instances were added too. For this reason the dataset is imbalanced, we are balancing it before proceeding with machine learning techniques.*

<u>Weight Balancing:</u> $\text{Weight for class } i = \dfrac{\text{Total number of samples}}{\text{Number of classes} * \text{count}(i)}$

<u>Data Analysis:</u>



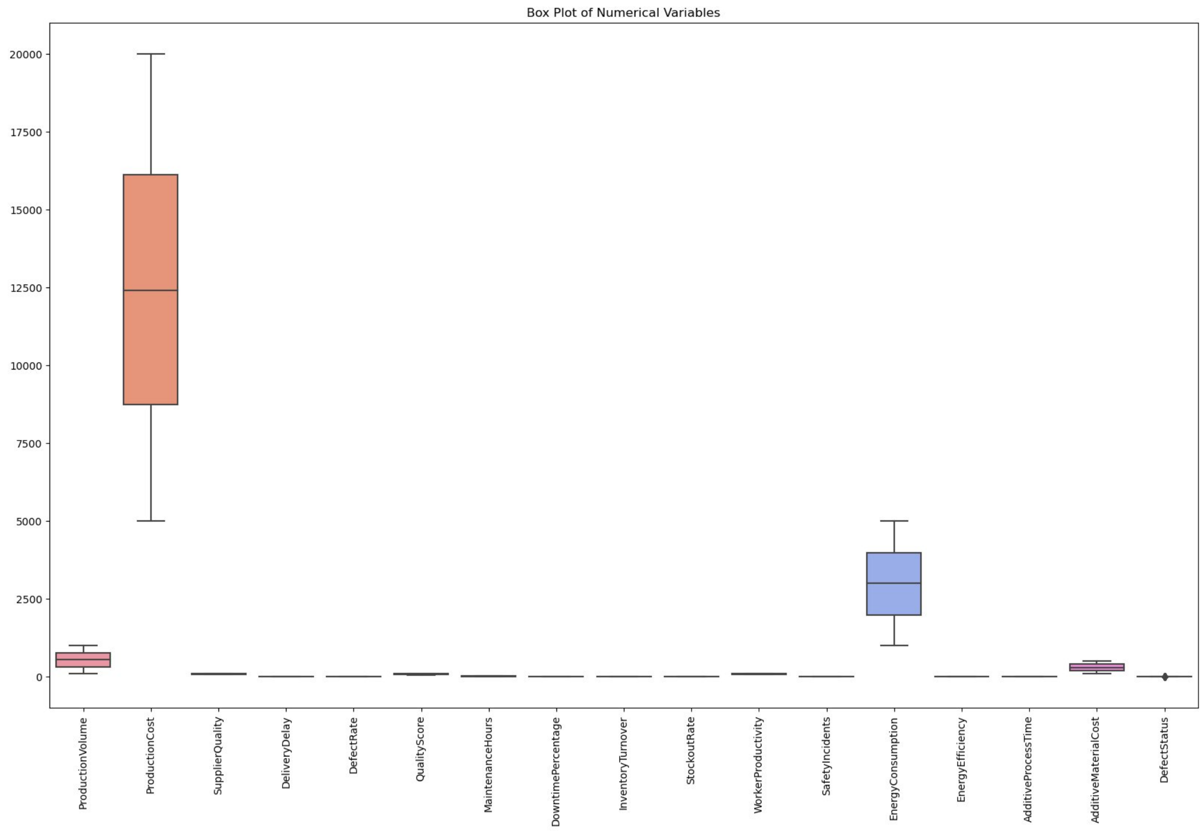
Most Important Features for Defect Status (High Correlation):

- Maintenance Hours (0.30)
- Defect Rate (0.25)
- Quality Score (-0.20) : i.e as defect decreases as quality score increases
- Production Volume (0.13)

Least Important Features for Defect Status (Low Correlation):

- ➢ Additive Process Time (0.01)
- ➢ Delivery Delay (0.01)
- ➢ Energy Consumption (0.01)
- ➢ Product Cost (0.03) and so on……

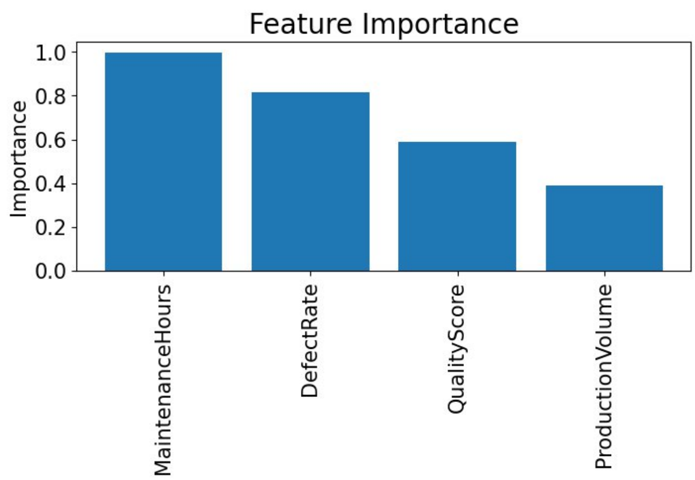
These less important features may not significantly affect the performance of the model.
Hence we dropped the features with correlation < |0.1|.



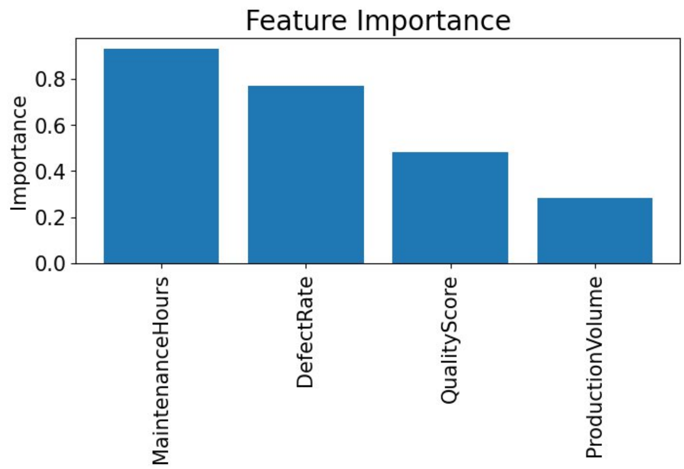
From observing these box plots, we can conclude that there are no outliers present in the data which needs to be removed.

Feature Importance:

- **For Logistic Regression:**



- **For SVM:**



Results:

- **For Logistic Regression:**

Class	Precision	Recall	F1-Score	Support
0	0.39	0.79	0.53	102
1	0.95	0.77	0.85	546

➳ **Accuracy:** 0.77

- **For SVM:**

Class	Precision	Recall	F1-Score	Support
0	0.41	0.78	0.54	102
1	0.95	0.79	0.86	546

➳ **Accuracy:** 0.79

Inference:

Both Logistic Regression and SVM gives similar results as basic working of both models is similar.

- **For Class = 0 (Defect Status = 0):** This means that product is defect free.
  - Since we can tolerate classification of defect free product into defect class.
  - Precision can be low. (Here approx. 0.40)
- **For Class = 1 (Defect Status = 1):** This means that product has defects.
  - Since we can't tolerate classification of defected product into defect free class as may be this leads to failure of some part of machine which is produced by additive manufacturing resulting in catastrophic events.
  - Recall has to be higher. (Here approx. 0.79)

Conclusion:

As observed above, we can predict the outcome significantly based on only 4 features out of 18**: Maintenance Hours, Defect Rate, Quality Score and Production Volume.** Using these features, we can predict whether the product has defects or not.

There are various parts of machines which are being made by additive manufacturing (3-D printing) technique. Therefore, it is essential to discard the defected products as if they fail during the use of machine, it can be disastrous. So, using these models, we can predict whether the product has defects or not based on the 4 features selected.

Furthermore, we can also optimize these features to generate more defect free products.