Indian Institute of Technology
Bombay

2022 April 24

Yield Prediction in Semiconductor Manufacturing Process

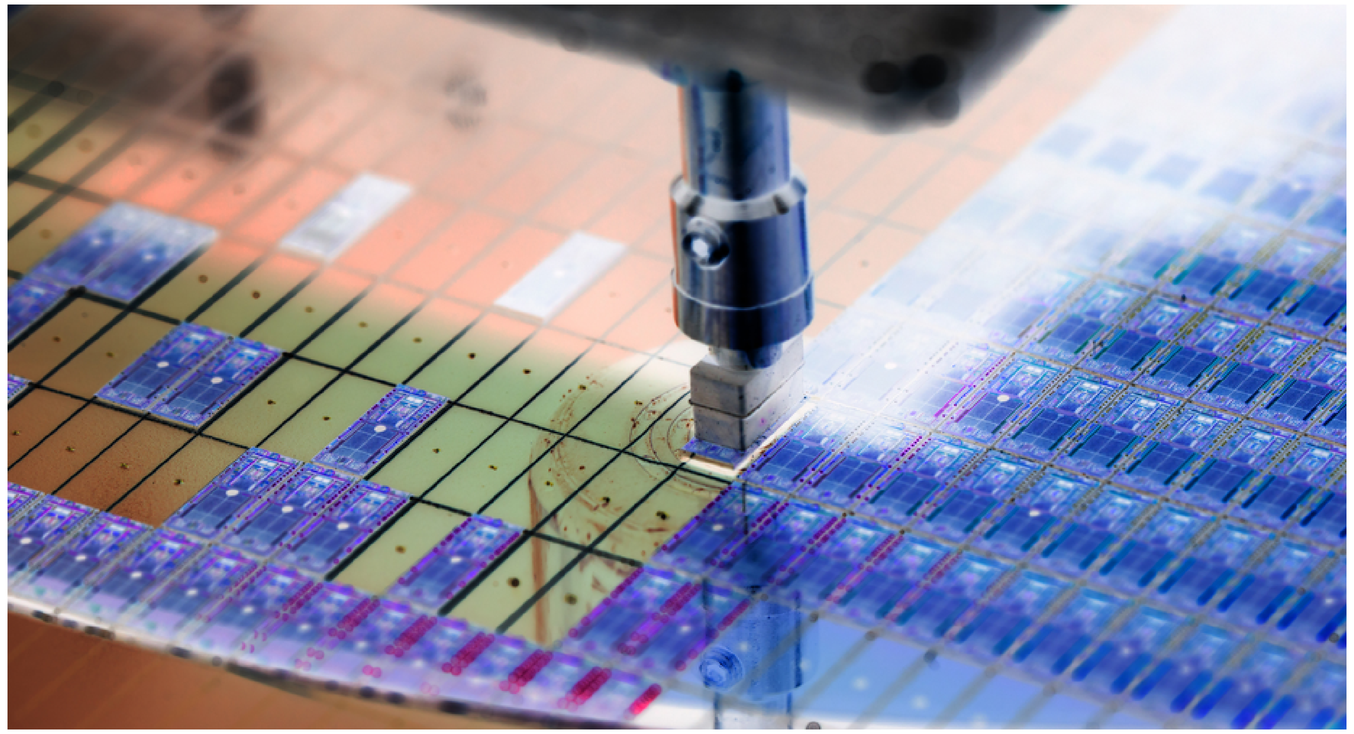
SUPERVISED PROJECT EXPOSITION

PRATHAM SEHGAL,TANMAY BARHARTE,
SHIVPRASAD KATHANE

GUIDE: PROF ALANKAR ALANKAR

INTRODUCTION

- The project focuses on predicting and optimizing yield in semiconductor manufacturing.
- The UCI COM dataset provides a valuable source of data for analyzing factors that impact yield.
- Machine learning techniques will be applied to the dataset to predict yield and identify key contributing factors.
- The insights gained from this analysis can be used to improve efficiency and reduce costs.
- Ultimately, this project aims to enhance product quality and maintain a competitive edge in the semiconductor manufacturing industry.



MOTIVATION

- Semiconductor manufacturing process is monitored using signals collected from sensors and measurement points.
- Feature selection can be applied to identify the most relevant signals that contribute to yield excursions downstream in the process.
- Analyzing and testing different combinations of features can identify essential signals impacting the yield type, leading to increased process efficiency and decreased production costs.

Indian Institute of Technology
Bombay

Objective

In this project,we will build a classifier to predict the Pass/Fail yield of a particular process entity and analyze whether all the features are required to build the model or not.

Indian Institute of Technology
Bombay

Unlocking Insights

Data Visualization

Indian Institute of Technology
Bombay

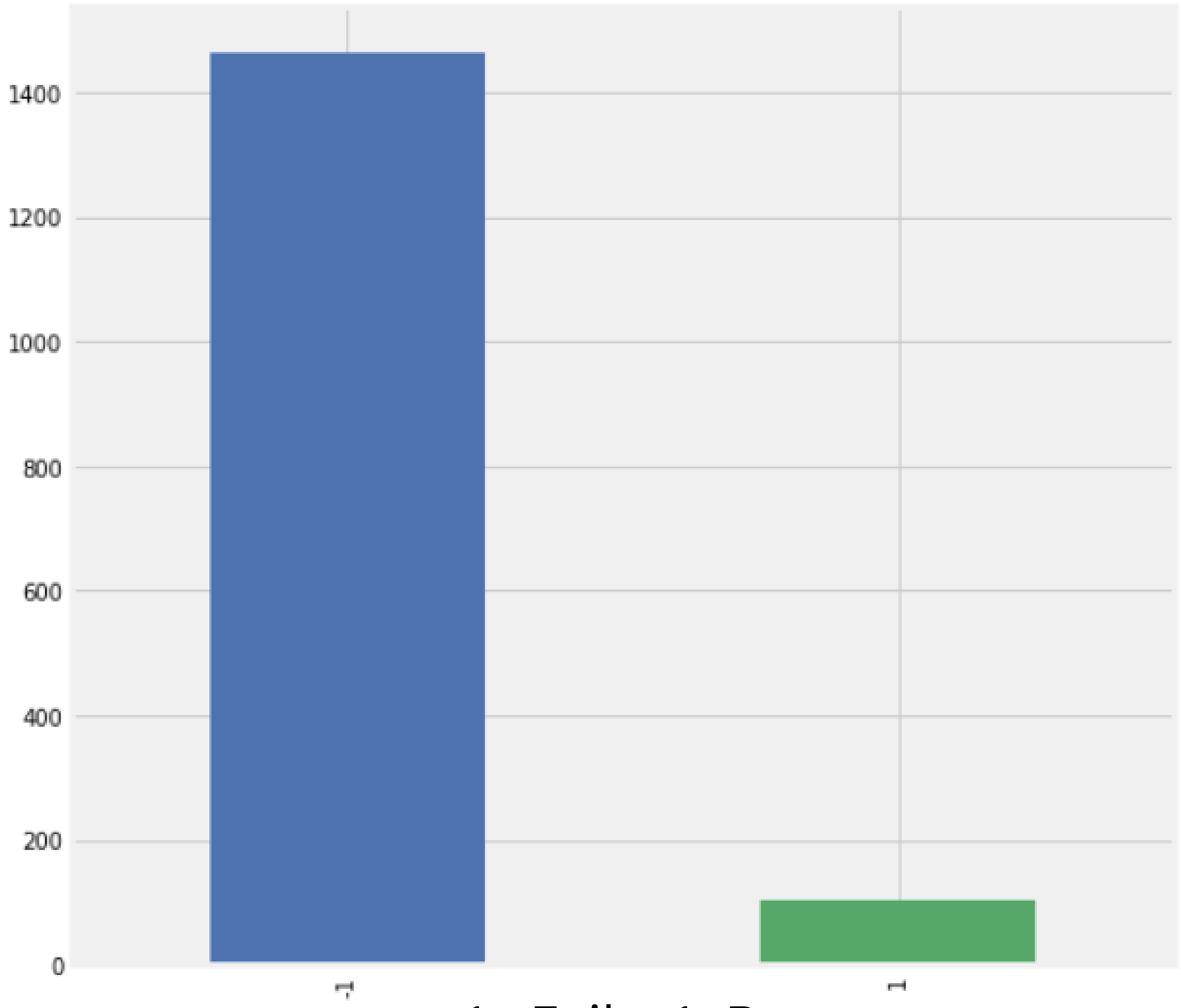
THE DATASET

```
In [ ]: 1 display(X) # features dataframe
```

	0	1	2	3	4	5	6	7	8	9	...	580	581	582	583	584	585	586
0	3095.78	2465.14	2230.4222	1463.6606	0.8294	100	102.3433	0.1247	1.4966	-0.0005	...	0.006	208.2045	0.5019	0.0223	0.0055	4.4447	0.0096
1	2932.61	2559.94	2186.4111	1698.0172	1.5102	100	95.4878	0.1241	1.4436	0.0041	...	0.0148	82.8602	0.4958	0.0157	0.0039	3.1745	0.0584
2	2988.72	2479.9	2199.0333	909.7926	1.3204	100	104.2367	0.1217	1.4882	-0.0124	...	0.0044	73.8432	0.499	0.0103	0.0025	2.0544	0.0202
3	3032.24	2502.87	2233.3667	1326.52	1.5334	100	100.3967	0.1235	1.5031	-0.0031	...	NaN	NaN	0.48	0.4766	0.1045	99.3032	0.0202
4	2946.25	2432.84	2233.3667	1326.52	1.5334	100	100.3967	0.1235	1.5287	0.0167	...	0.0052	44.0077	0.4949	0.0189	0.0044	3.8276	0.0342
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1561	2899.41	2464.36	2179.7333	3085.3781	1.4843	100	82.2467	0.1248	1.3424	-0.0045	...	0.0047	203.172	0.4988	0.0143	0.0039	2.8669	0.0068
1562	3052.31	2522.55	2198.5667	1124.6595	0.8763	100	98.4689	0.1205	1.4333	-0.0061	...	NaN	NaN	0.4975	0.0131	0.0036	2.6238	0.0068
1563	2978.81	2379.78	2206.3	1110.4967	0.8236	100	99.4122	0.1208	NaN	NaN	...	0.0025	43.5231	0.4987	0.0153	0.0041	3.059	0.0197
1564	2894.92	2532.01	2177.0333	1183.7287	1.5726	100	98.7978	0.1213	1.4622	-0.0072	...	0.0075	93.4941	0.5004	0.0178	0.0038	3.5662	0.0262
1565	2944.92	2450.76	2195.4444	2914.1792	1.5978	100	85.1011	0.1235	NaN	NaN	...	0.0045	137.7844	0.4987	0.0181	0.004	3.6275	0.0117

1566 rows × 590 columns

It has 590 features and around 1566 rows

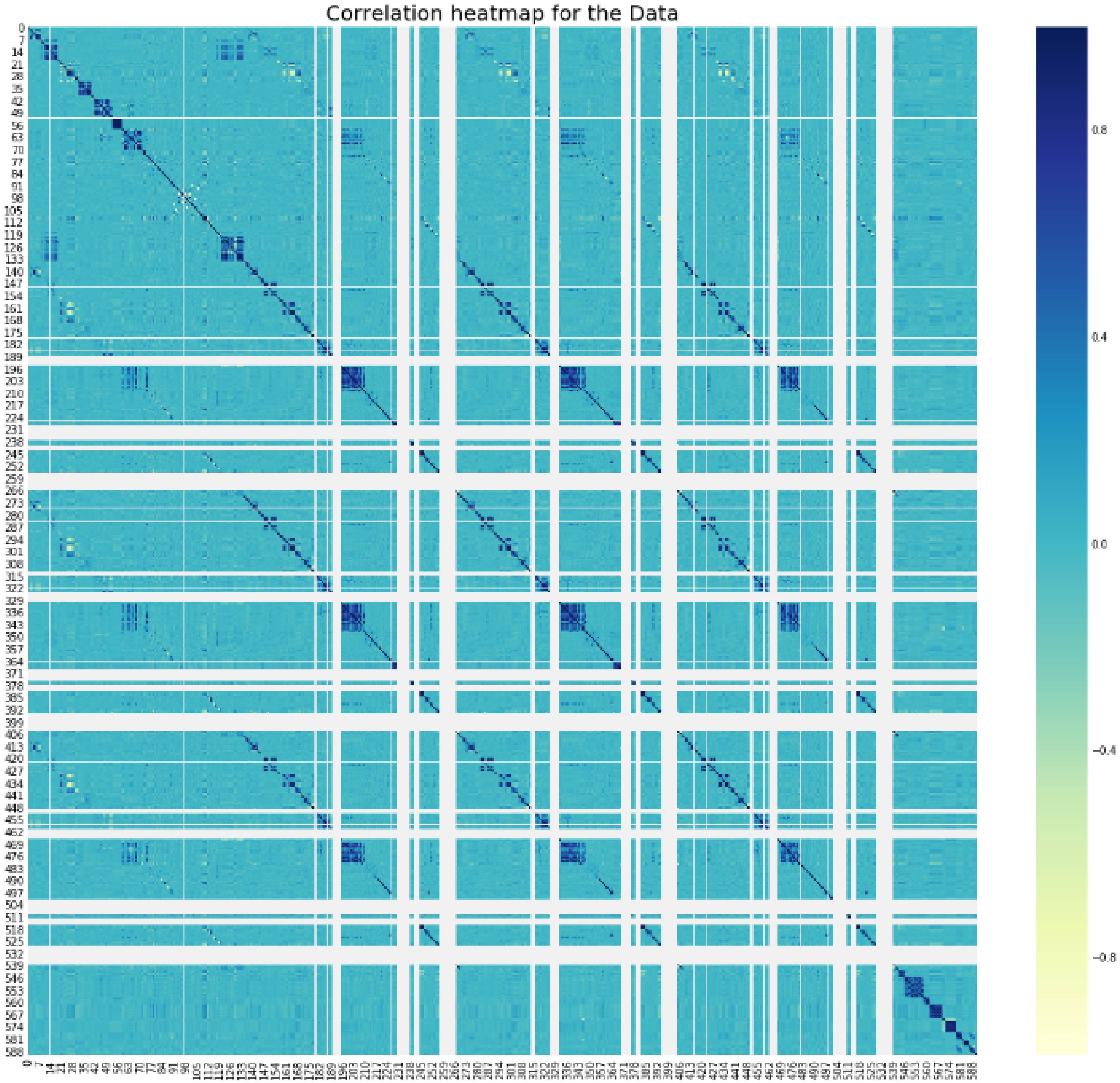


-1 : Fail :: 1: Pass

It is apparent from the plots that the dataset exhibits class imbalance.

Indian Institute of Technology
Bombay

HEATMAP

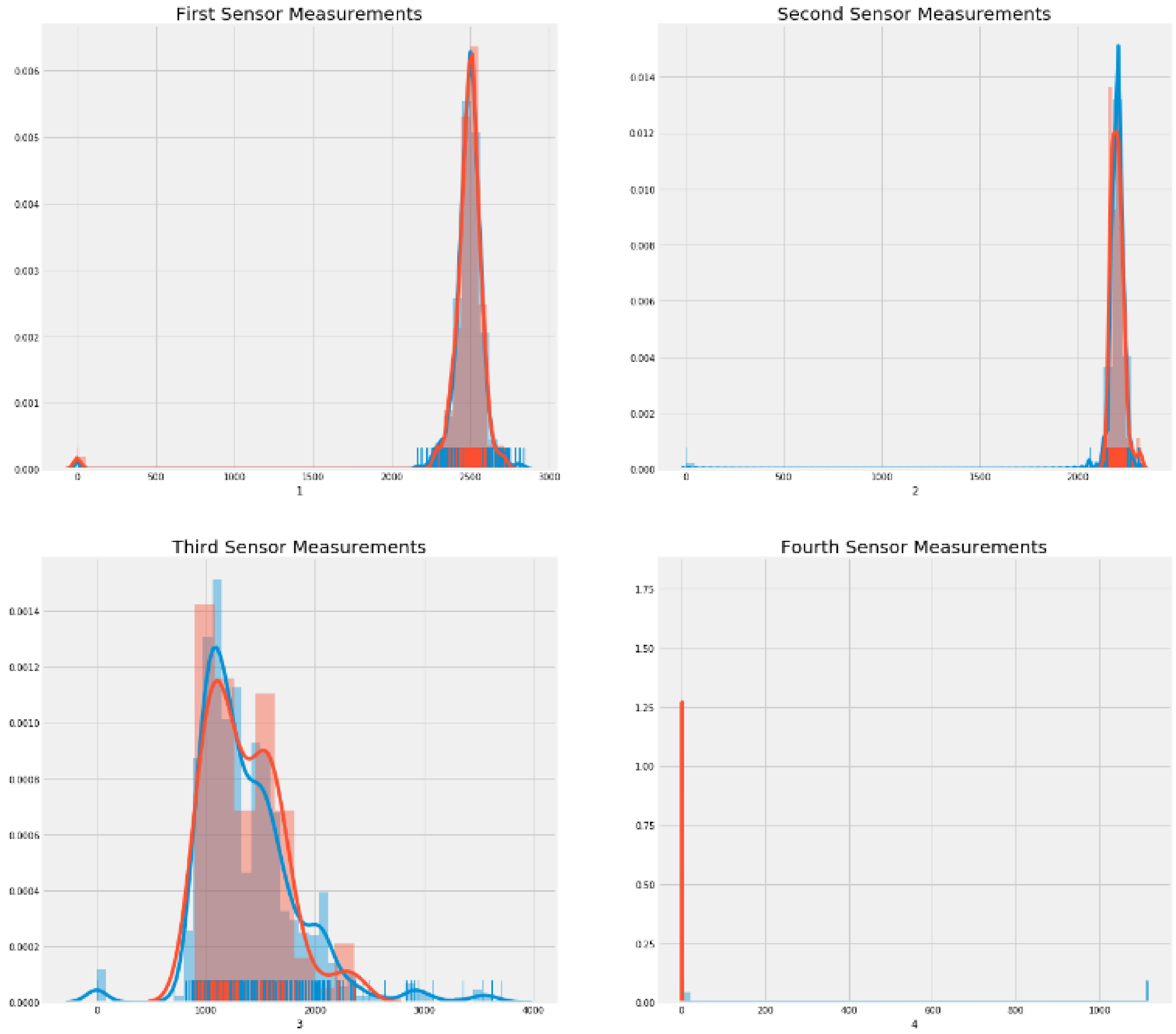


Violet (dark) regions do appear in the heatmap implying presence of **correlated features**

Indian Institute of Technology
Bombay

DATA VISUALIZATION

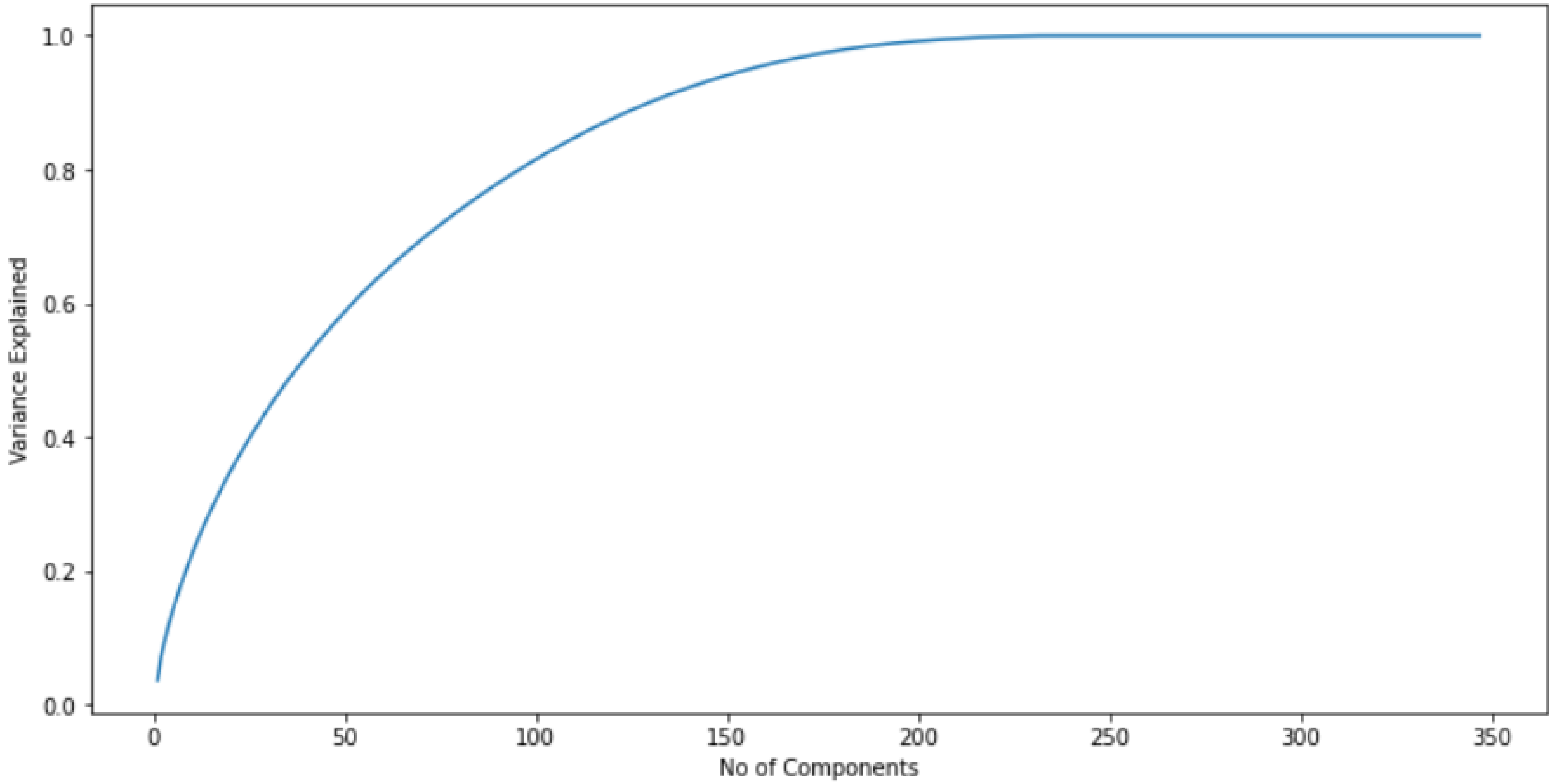
CHECKING(CLASS-WISE) DISTRIBUTION FOR FIRST 4 SENSOR MEASUREMENTS



Indian Institute of Technology
Bombay

Applying PCA and checking for optimum number of PCs that are sufficient to represent the data

THE TOP 200 PRINCIPAL COMPONENTS EXPLAIN ALMOST ALL THE VARIANCE



CONCLUSION

- In the further coming stages ,The results that we will obtain can be compared with similar studies from the existing literature to further validate the feasibility and effectiveness of the proposed models and methodologies.

- Improvement could be to explore the use of ensemble techniques such as bagging, boosting, or random forests to improve the accuracy of the fault diagnosis models. The paper mentioned that boosting techniques were used in the study, but it may be worth investigating if other ensemble methods can provide even better results.

- **Feature selection:** Different feature selection techniques can be explored to enhance the effectiveness of the PCs.

  Apply PCA-based feature selection
  Apply clustering-based feature selection -Mean difference
  FCBF (Fast Correlation-Based Filter)
  CFS (Correlation-Based Feature Selection)
  DBSCAN (Density-Based Spatial Clustering of Applications
  Laplacian Eigen maps for dimensionality reduction

QUESTIONS

Describe what activity was performed by which team member.

- The literature review and review on existing machine learning models that has been used previously was done by me(Pratham Sehgal).

- Data Cleaning Part and EDA was done by TANMAY BARHARTE.

- Applying PCA and checking the components which explain best Variance in the data was done by SHIVPRASAD KATHANE.

Overall, in your opinion, how much contribution to the project was made by each member (in %) : 2 points

- Pratham Sehgal - 35%
- Tanmay Barharte - 30%
- ShivPrasad Modi - 35%