International Journal of Engineering and Extended Technologies Research (IJEETR)
Sunil Patil Vol2, Issue 2, PP. 234-239 2022

First Pass Yield Improvement by Process Capability Analysis and Predicting type of Assembly fit by building XG boost Machine Learning model using Python

Sunil Patil
Production Professional
Email ID- 77sunil@gmail.com

Abstract

First pass yield [1] is a good measure of the effectiveness of a process and the elimination of waste from that process. Many other measures of productivity and efficiency fail to account for the cost of rework and in many industries, rework can be a significant portion of the time and value added to final production.

Goal is to perform process capability analysis [2] and predicting future quality levels to improve first pass yield and reduce rework.

Here in this project, shaft and hole component of manufacturing assembly are considered separately with their specification, designed to meet the Assembly fitment class, here fitment required is Transition fit.

Assembly of component is by random selection of both shaft and hole component.

Random measurements are generated for both components within the design specification, with available measurement Process capability analysis is done using python. Both Process capability results were at acceptable range.

Class of Assembly fitment is generated by pairing components. Measurement of components and assembly fitment class data is taken to build machine learning model [3].

XG boost model [4] is build and trained on available data, initial model accuracy found 96% and is further improved to 98% for larger sample data.

Prediction accuracy for determining class of fit found 99.5%.

Keywords: First Pass Yield, Manufacturing assembly, Process capability, Python, Machine learning, XG boost.

1. INTRODUCTION

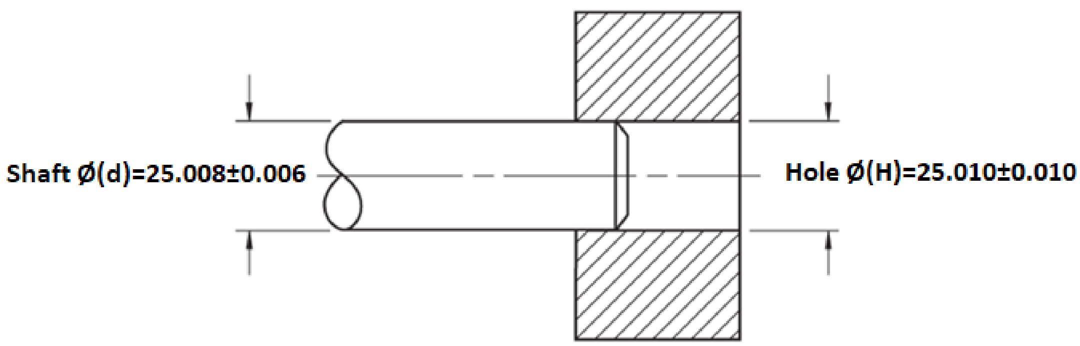


Figure-Assembly

Manufacturing industry often require two parts to be assemble, mostly they are Hole and shaft type assemblies.

Huge amount of time, cost and efforts are wasted to tackle assembly fitment problems.

Although, Design specification satisfy intent of fitment, it is due to process variation, that affects the assembly process and fitment of components.

The underlying problem is considered as our business problem and our business objective is to: -

1.Minimize Rejections and Repair

2.Maximize First Pass yield.

Here, we have considered shaft and hole assembly, as shown in above figure.

2. LITERATURE REVIEW

A pragmatic view on process capability studies in recent years an increasing number of organizations use process capability studies on a regular basis.

Contemporaneous with the increasing number of organizations using process capability studies, warnings have been launched that imprudent use of numerical measures of capability, the so-called process capability indices, might lead the user to make erroneous decisions.

As a result, many practitioners of today are left with somewhat ambivalent attitude towards process capability studies.

Mats Deleryd [5] states that Process capability studies are used for monitoring the capability of a process.

This implies that it has to be based on some sort of collection of data from the process. In order to get a fair picture of the capability of the process, it has to be stable when the data is collected. After the collection of data from a stable process, the data may be assessed in several ways.

One way to do the assessment is to use process capability indices, which provide numerical measures of the capability. Based on the assessment, improvement efforts can be initiated. The fact that, theoretically, there is no limit of how capable a process can become, implies that there is no end to the improvement initiatives aimed at achieving more and more capable processes.

3.METHODLOGY

The relationship between two mating components is known as "fit", and pertains to how tight or loose the items should be when joined together.

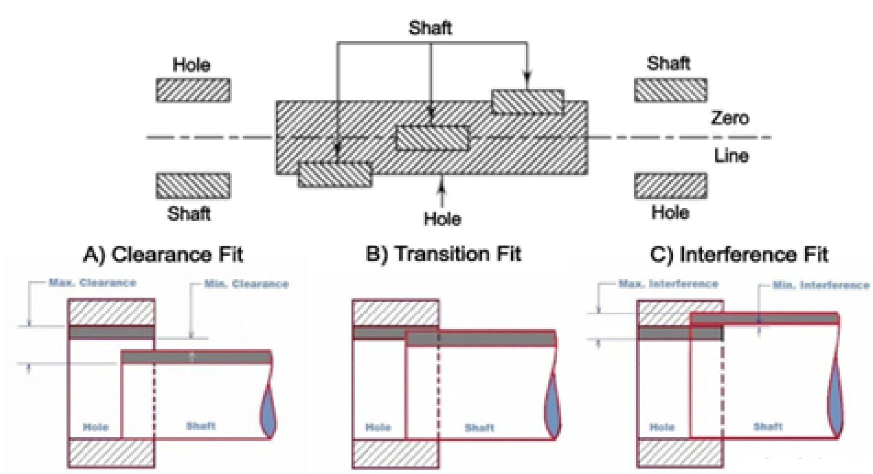
There are three types of fit commonly referenced in manufacturing.

1.Clearance Fit-allow for loose mating

2.An interference fit will be much tighter than a clearance fit

3. A transition fit would fall between a clearance and interference fit

Better understandable with below figure



In our case, with mentioned dimension, Transition fit is our required Fit.

Now, referring to American National Standard Preferred Hole Basis Metric Clearance Fits [6] (ANSI B4.2–1978, R1984). e.g.

Basic Size		Loose Running Hole H11	Loose Running Shaft c11	Loose Running Fit†	Free Running Hole H9	Free Running Shaft d9	Free Running Fit†	Close Running Hole H8	Close Running Shaft f7	Close Running Fit†	Sliding Hole H7	Sliding Shaft g6	Sliding Fit†	Locational Clearance Hole H7	Locational Clearance Shaft h6	Locational Clearance Fit†
25	Max	25.130	24.890	0.370	25.052	24.935	0.169	25.033	24.980	0.074	25.021	24.993	0.041	25.021	25.000	0.034
	Min	25.000	24.760	0.110	25.000	24.883	0.065	25.000	24.959	0.010	25.000	24.980	0.007	25.000	24.987	0.000
30	Max	30.130	29.890	0.370	30.052	29.935	0.169	30.033	29.980	0.074	30.021	29.993	0.041	30.021	30.000	0.034

Basic Size		Locational Transition Hole H7	Locational Transition Shaft k6	Locational Transition Fit†	Locational Transition Hole H7	Locational Transition Shaft n6	Locational Transition Fit†	Locational Interference Hole H7	Locational Interference Shaft p6	Locational Interference Fit†	Medium Drive Hole H7	Medium Drive Shaft s6	Medium Drive Fit†	Force Hole H7	Force Shaft u6	Force Fit†
25	Max	25.021	25.015	+0.019	25.021	25.028	+0.006	25.021	25.035	−0.001	25.021	25.048	−0.014	25.021	25.061	−0.027
	Min	25.000	25.002	−0.015	25.000	25.015	−0.028	25.000	25.022	−0.035	25.000	25.035	−0.048	25.000	25.048	−0.061

We came to know the maximum and minimum diameter differences between shaft and hole, for various fits.

Summarising this in below table,

	Lead to Rejection			Required Fit		Lead to Repair	
	CLERANCE FIT			TRNSITION FIT		INTEREFERENCE FIT	
FIT	CLOSE RUNNING(H8f7)	SLIDING(H7g6)	LOCATION CLEARANCE(H7h6)	Similar Transition(H7k6)	Fixed Transition(H7n6)	Press fit(H7p6)	Driving Fit(H7s6)
MAX (Diameter Difference)	0.074	0.041	0.034	0.019	0.006	-0.001	-0.014
MIN (Diameter Difference)	0.010	0.007	0.000	-0.015	-0.028	-0.035	-0.048
Force	Loose	Slide	Light Push	Mallet_push	Light force	Press	Drive

e.g., fit clearance calculation for various cases and type of fit.

1. { (hole diameter H1) 25.020 - (Shaft diameter d1) 25.002 } = 0.018 , which falls in clearance and transition fit.

2. { (hole diameter H1) 25.000 - (Shaft diameter d1) 25.015 } = - 0.015 , which falls in Interference and transition fit.

Above explained phenomenon of, having different fits due to various fit clearances, which are caused, mainly due to process variation, is our business problem.

As here you can see, we can have three different fits, in Machine learning language, this problem is called as, three class classification problem.

Now, let us first understand the cause of this effect i.e., Process variation

By doing Process Capability Analysis-

Process capability analysis represents a significant component of the *Measure* phase from the DMAIC[7] (Define, Measure, Analysis, Improve, Control) cycle during a Six Sigma project. This analysis measures how a process performance fits the customer's requirements, which are translated into specification limits for the characteristics of the product to be manufactured or produced. The results from this analysis may help to identify variation within a process and develop further action plans that lead to better yield, lower variation, and fewer defects.

Specifications

Specifications are the voice of the customer. Every process should be capable of fulfilling the customer's requirements, which must be quantified to be attainable. Specification limits are the numerical expressions of the customer requirements. Due to natural variations within the process, specifications usually are a range with upper and lower bounds. *USL* (Upper Specification Limit) is a value above which the process performance is unacceptable, while *LSL* (Lower Specification Limit) is a value below which the process performance is unacceptable.

Process Performance -Process performance is the voice of the process. A process can be considered right when it is approximating to the *target*, with as little *variation* as possible. In the Six Sigma approach, the most common process performance measures are:

- Yield (*Y*): the number of good products or items produced by the process. It can be assessed once the process is finished, counting the items that fit the specifications:

$$Y = \frac{total - defects}{total}$$

- First-time yield (*FTY*): takes into consideration the rework in the middle of the process. Thus, regardless of the number of correct items at the end of the process, counts the correct items as “first time” correct items:

$$FTY = \frac{total - rework - defects}{total}$$

- Defects per opportunity (*DPU*): number of nonconformities per unit. Defects are the complement of the yield:

$$DPU = 1 - FTY = \frac{defects}{total}$$

- Defects per million opportunities (*DPMO*): number of nonconformities per million opportunities. It is mainly used as a long-term performance measure of a process:

$$DPMO = \frac{number\ of\ defects}{number\ of\ opportunities} \times 10^6$$

Process vs. Specifications

International Journal of Engineering and Extended Technologies Research (IJEETR)
Sunil Patil Vol2, Issue 2, PP. 234-239 2022

The *sigma score* of a process (Z) is a simple number that conveys how a process fits the customer specifications. Processes that reach a *sigma* level of 6 may be considered as "almost perfectly" (i.e. with almost zero defects) designed processes. A *sigma* value of 6 implies that less than 3.4 DPMO (defects per million opportunities) will be attained. The *sigma* is the number of standard deviations that fit between the specification limit and the mean of a process. It is calculated using the formula:

$$Z = \min\left\{\frac{(USL - \bar{x})}{\sigma}, \frac{(\bar{x} - LSL)}{\sigma}\right\}$$

where:

- USL = Upper Specification Limit
- LSL = Lower Specification Limit
- $\bar{x}$ = process mean
- σ = standard deviation

DPMO through sigma scores-

Z value	DPMO
1	690,000
2	308,000
3	66,800
4	6,210
5	233
6	3.4

Capability Indices

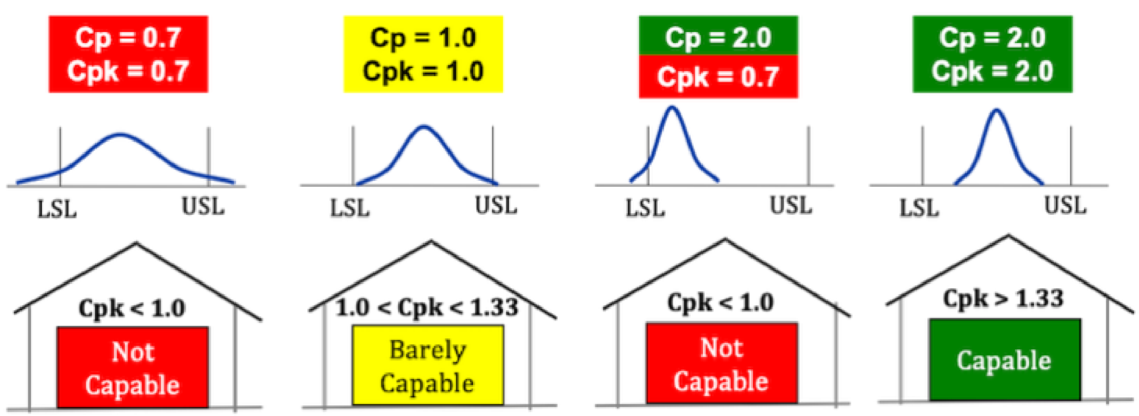
Capability indices directly compare the customer specifications with the performance of the process. They are based on the fact that the *natural limits* or *effective limits* of a process are those between the mean and +/- 3 standard deviations (i.e. where 99.7% of the data is contained). The capability of a process (C*p*) is calculated using the formula:

$$C_p = \frac{USL - LSL}{6\sigma}$$

However, this formula does not allow to validate whether the process is centered in the mean (which is desirable). To deal with this issue, the adjusted capability index (C*pk*) is calculated using the formula:

$$C_{pk} = \min\{\frac{USL - \bar{x}}{3\sigma}, \frac{\bar{x} - LSL}{3\sigma}\}$$

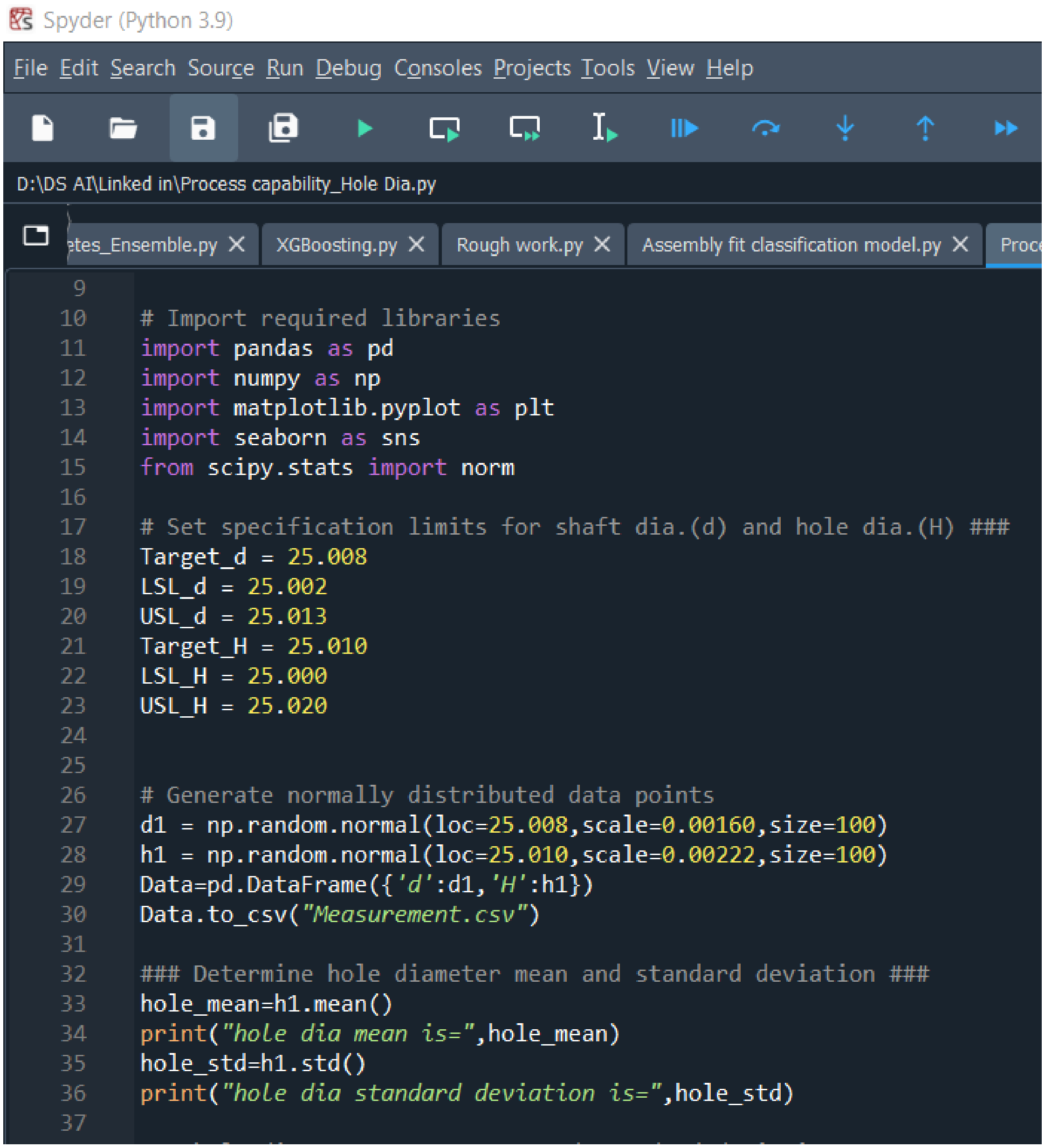
Like the sigma score, capability indices help to determine how well a process is meeting customer specifications. In general, a C*pk* of 1.33 is acceptable, but the greater its value, the better.



Now ,Let us first conduct process capability analysis on input using Python, we have two input variables 1.Shaft diameter = d ,2.Hole diameter=H and one output variable i.e. Y= fitment class.

International Journal of Engineering and Extended Technologies Research (IJEETR)
Sunil Patil Vol2, Issue 2, PP. 234-239 2022

Each input contains 100 random values with variation in base dimension. Below are the key event, plots and process capability summary



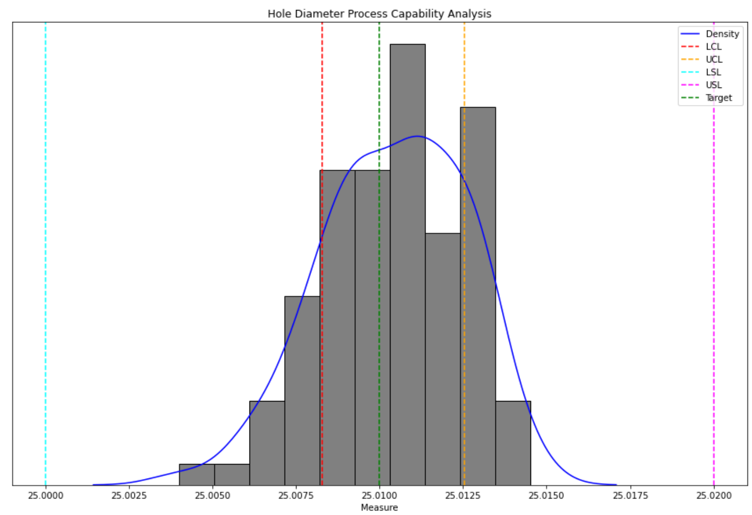
International Journal of Engineering and Extended Technologies Research (IJEETR)
Sunil Patil Vol2, Issue 2, PP. 234-239 2022

```
Hole Diameter PROCESS CAPABILITY ANALYSIS
-----------------------------------
Specifications

Taget: 25.01
LSL: 25.0
USL: 25.02
-----------------------------------
Indices

Cp: 1.565
Cpk: 1.501
z: 4.504
-----------------------------------
Summary Statistics

Number of samples: 100
Sample mean: 25.01
Sample std: 0.002
Sample max: 25.015
Sample min: 25.004
Sample median: 25.011
Percentage of data points below LSL: 0.0%
Percentage of data points above USL: 0.0%
```

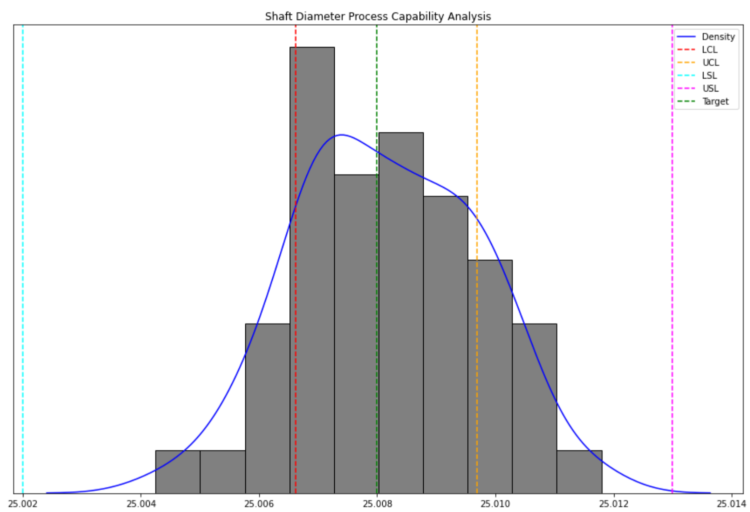



```
Shaft Diameter PROCESS CAPABILITY ANALYSIS
-----------------------------------
Specifications

Taget: 25.008
LSL: 25.002
USL: 25.013
-----------------------------------
Indices

Cp: 1.195
Cpk: 1.052
z: 1.69
-----------------------------------
Summary Statistics

Number of samples: 100
Sample mean: 25.01
Sample std: 0.002
Sample max: 25.015
Sample min: 25.004
Sample median: 25.011
Percentage of data points below LSL: 0.0%
Percentage of data points above USL: 0.0%
```



Machine Learning Model building-

Machine learning (ML) is a type of artificial intelligence (AI)that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

A machine learning model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data.

Once you have trained the model, you can use it to reason over data that it hasn't seen before, and make predictions about those data.

In our business problem we are going to build Extreme gradient boosting model.

Gradient boosting refers to a class of ensemble machine learning algorithms that can be used for classification or regression predictive modelling problems.

Ensembles are constructed from decision tree models. Trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models. This is a type of ensemble machine learning model referred to as boosting.

Models are fit using any arbitrary differentiable loss function and gradient descent optimization algorithm.
Extreme Gradient Boosting, or XGBoost for short is an efficient open-source implementation of the gradient boosting algorithm. As such, XGBoost is an algorithm, an open-source project, and a Python library.

It is designed to be both computationally efficient (e.g., fast to execute) and highly effective, perhaps more effective than other open-source implementations.

The two main reasons to use XGBoost are execution speed and model performance.

4.DISCUSSION

Process capability analysis solves the basic statistical problem in process quality controls, which is establishing a state of control over the manufacturing process, i.e., eliminating special causes of variation and then maintaining that state of control through time. Process capability analysis gives process capability chart and run chart. Process capability chart is produced by taking the measurement of

International Journal of Engineering and Extended Technologies Research (IJEETR)
Sunil Patil Vol2, Issue 2, PP. 234-239 2022

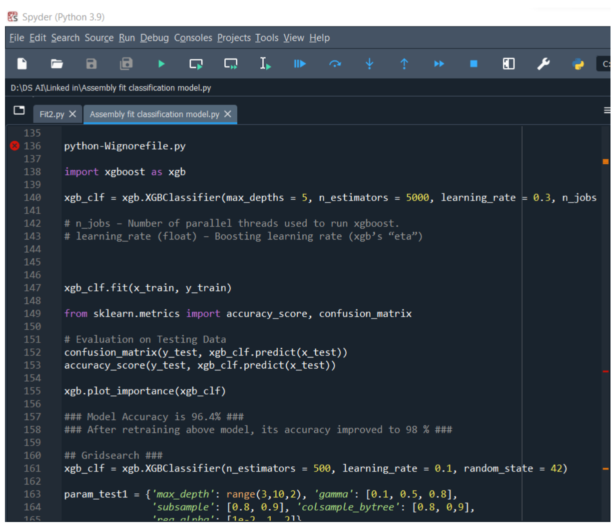
machined parts which gives the comparison of natural tolerance limits with specification limits and the natural tolerance range with the specification range.

Both Shaft and Hole process capabilities (Cp, Cpk>1.33) are above acceptable limit, implying that process is capable to produce parts within specification limit.

The developed XGBoost model is fitted to available measurement.

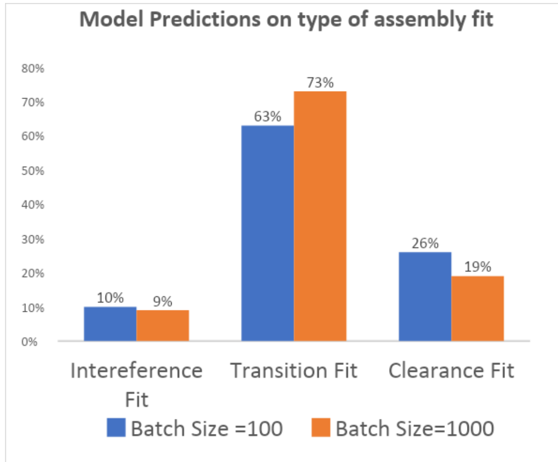
Model learns from available data and is used to predict the class of fitment for new input.

Below is the key event.



5.RESULT AND INTERPRETATION

	% of Total Assembly Fit	
	Batch Size =100	Batch Size=1000
Intereference Fit	10%	9%
Transition Fit	63%	73%
Clearance Fit	26%	19%
Model Accuracy	96%	98%
Model Prediction accuracy %	99%	99.5%



1.for evaluated shaft and hole diameter process capability, @70% assembly will fall into required fit ,20% into Clearance fit and 10% into Interference fit.

2.Based on model results, necessary cost-effective measures can be initiated at shaft or hole process to achieve desired fit

3.Entire practice of process capability analysis and predictive model building can be horizontally deployable to other manufacturing processes.

6.REFERENCES

[1] Matthew Littlefield, "Manufacturing Metrics: First Pass Yield Benchmark Data", https://blog.lnsresearch.com/bid/170419/Manufacturing-Metrics-First-Pass-Yield-Benchmark-Data, Jan 24, 2013.

[2] Roberto Salazar, "Process Capability Analysis with Python Measuring Process Performance",

https://medium.com/geekculture/process-capability-analysis-with-python,May 14,2021.

[3] E. Burns, “Machine learning,” https://www.techtarget.com/

searchenterpriseai/definition/machine-learning-ML, 2021.

[4] M. Pathak, “XGBoost,” https://www.datacamp.com/community/tutorials/ xgboost-inpython, 2019.

[5] Deleryd.M, “A strategy for mastering variation, in: Proceedings of the 51st ASQC Annual Quality

Congress,” Orlando, pp. 760- 768,1995.

[6] https://www.oreilly.com/library/view/engineering-design-graphics/9781118078884/19_appb.html

[7] Patrick Waddick, “SIX SIGMA DMAIC QUICK REFERENCE”,

https://www.isixsigma.com/new-to-six-sigma/dmaic/six-sigma-dmaic-quick-reference,2018.