MSc in Industry 4.0
IND5003 Data Analysis for Sense Making
Lecturer: Prof. Vik Gopal

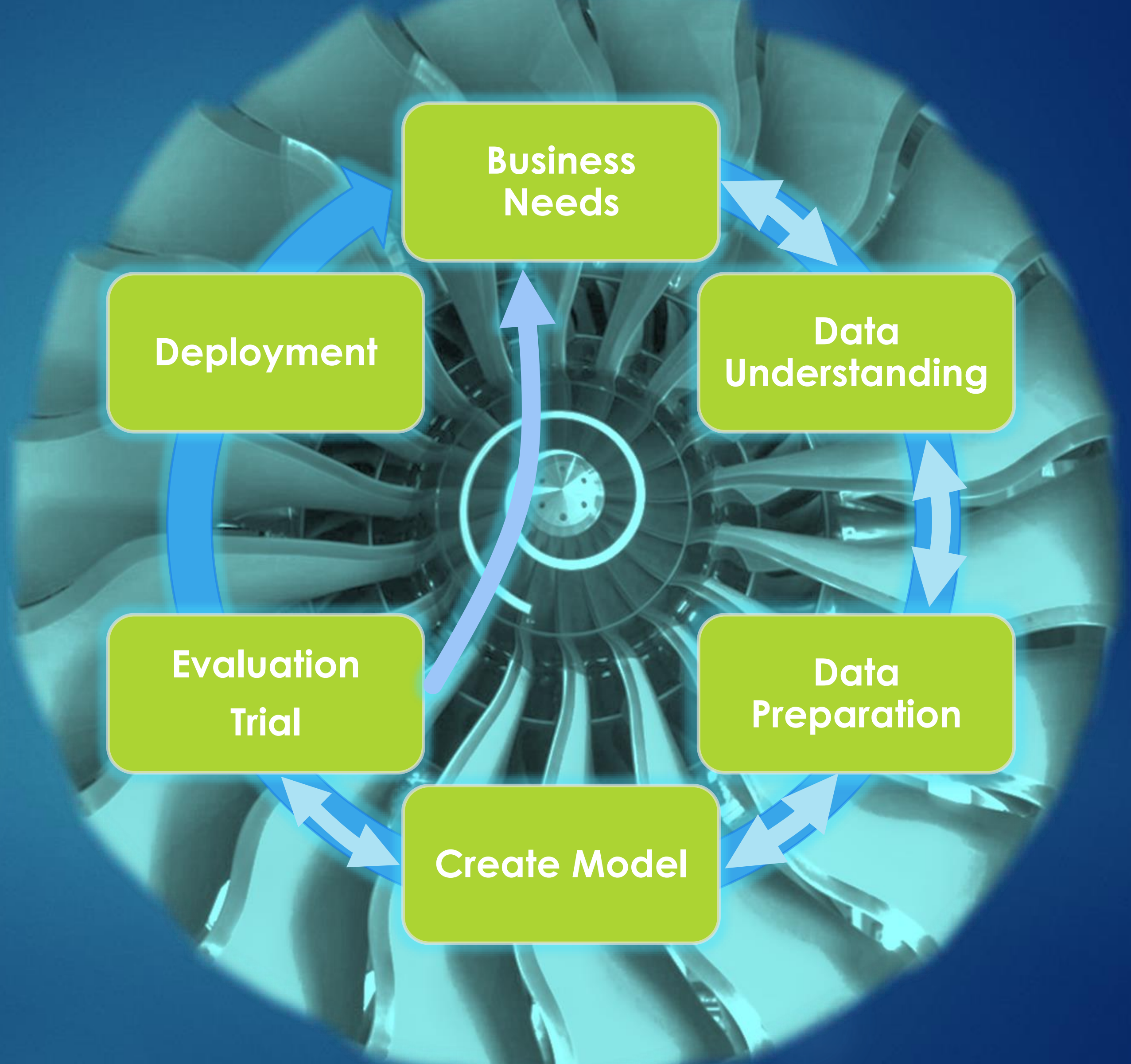
Group 5 Project :
Airplane Engine Vibration

Group member:

1) Elvis Ayroso (A0074996H)
2) Roger Lim (A0545092E)
3) Stanley Poh Wee Chin (A0013649E)
4) Wong Sin Wee (A0225653E)

Methodology Overview

1. **Business Needs** - create a common business understanding of current situation and project objectives.
2. **Data Understanding** - took the data for further exploration, and by collaborating with the subject matter experts and understanding the build process.
3. **Data Preparation** - Data was cleaned and formatted, several hundreds of features were extracted and final data sets for the analysis were selected.
4. **Create Models** - after initial models, the team iterated between the third and fourth step to further refine the models that most robust for the process.
5. **Evaluation and Trial** - against the most robust model for the build environment.
6. **Deployment Proposal** - visualizations and layouts that help the stakeholders understand how the results can be operationalized in practice



Business Needs

Problems: Poor Production Yield

1. Rejects at Test Bed:

- **Test bed pass rate is about 80%.**
- **On each engine been rejected, the rework lead time is 20 days and direct cost more than 100k.**
- **The engine is build to order, any reject will delay the delivery to airplane factory and cause the airplane delivery delay to airline.**
- **The top failure modes are HP and IP vibration.**
- **By improving the quality of the HP and IP modules build by 50%, the overall test bed pass rate can be improved from 80 to 91%.**
- **Estimated 200 engines produced a year in a plant, that is 22 reworks can be avoided.**

2. Rejects at Build:

- **High Pressure(HP) Module h**aving highest rework for build, pass rate is 70%. Due to the flatness and concentricity measurement out of drawing specification.

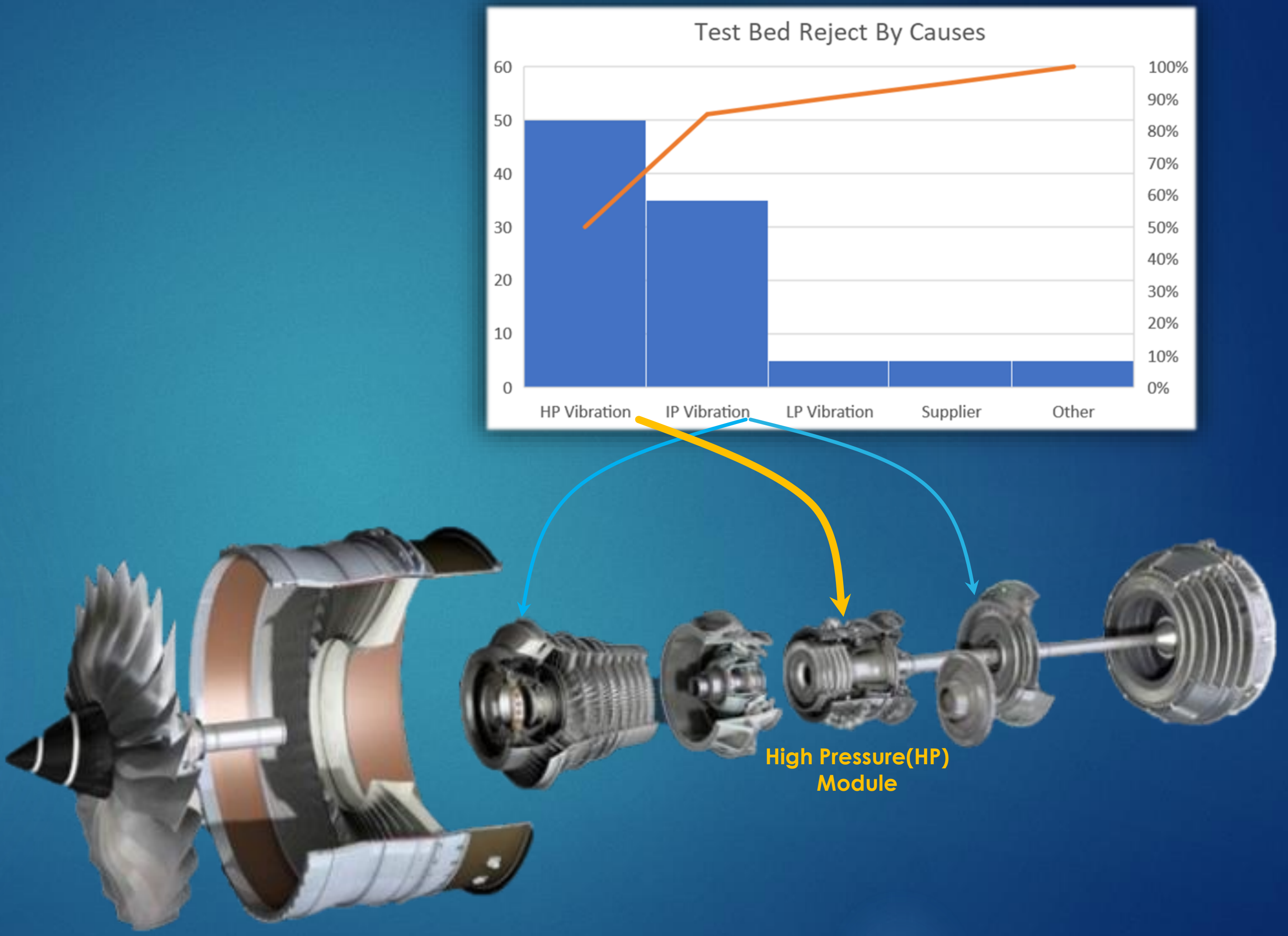


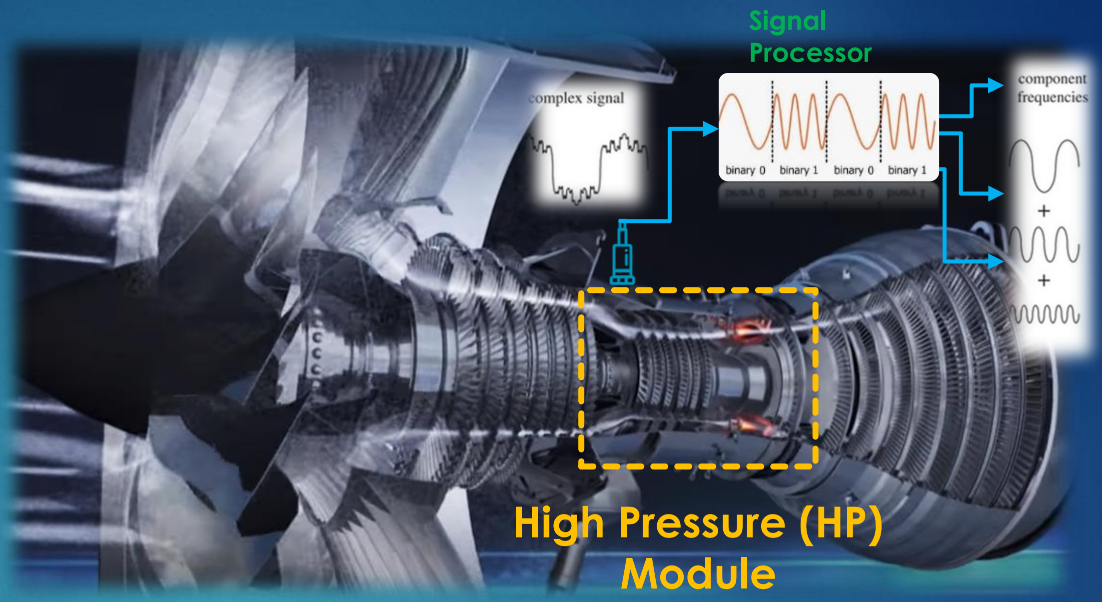
Figure from Rolls-Royce XWB apps

Business Needs

Objective for This Project

Using Machine Learning to:

- Improve the build quality of HP module to increase the pass rate at Test.
- Explore opportunities to improve the yield for the HP module build.



https://www.youtube.com/watch?v=JxkJ-FwFeVI (How jet engine works)

https://www.youtube.com/watch?v=m0X0o8aBrxk (3D modelling)

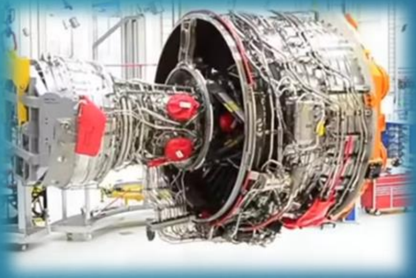
Figure from Rolls-Royce 1000 apps

Data Understanding

Engine build process:

Engine Test Where The Vibration signals to be recorded.

See The Jet Engine page 269 to 270.
HP Module build during balancing.



https://www.youtube.com/watch?v=K2R6NTgvEV4
(How jet engine been assembled and test)

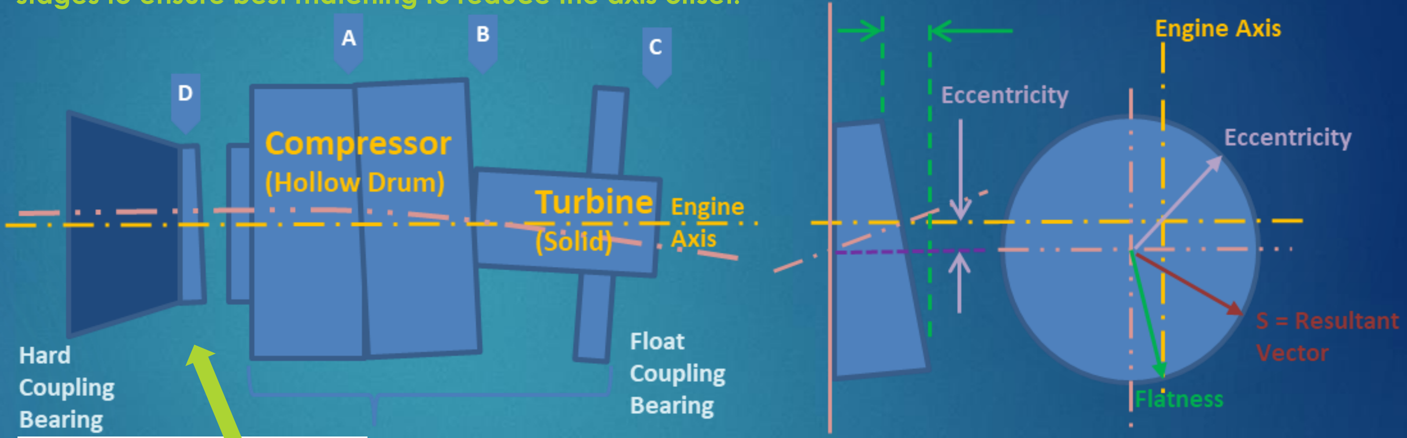
Data Understanding

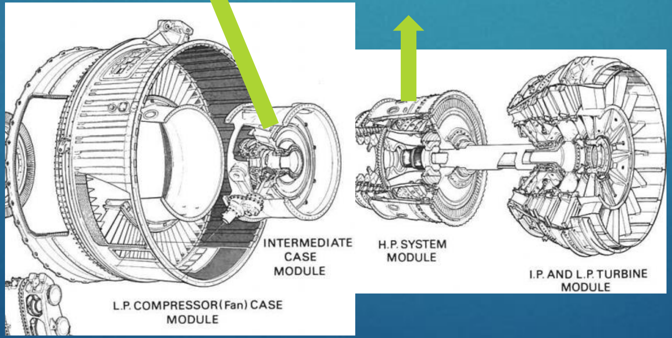
Deep dive to the HPS build:

- The HPS is the power generation source of the engine. The HPS rotate at 12,000 rpm at the power of 60 F1 cars for a large commercial wide body airplane engine.
- Vibration will damage the component and caused early fatigue.
- Vibration caused by the off axis and unbalance of the rotating components.
- Measurements are carried out along the staking of the components to ensure the off-axis within a specification.
- The measurements data can be pulled out from the MES.
- The vibration data is recorded at the test bed for the complex performance test curve of the engine. For the evaluation of this project, the maximum vibration is used.
- 70% of the total rejected at HP module due to C1 measurement out of specification.

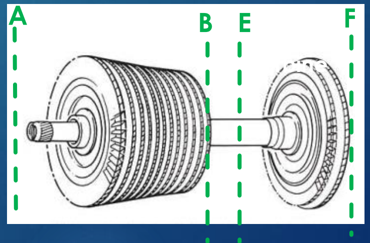
			Flatness Measurement								Balancing Measurement							
	Label	Label	A1	A2	B1	B2	C1	C3	D1	D3	Ax	Ay	Bx	By	Ex	Ey	Fx	Fy
Index	Pass/Fail	Vibration	um	um	um	um	um	Deg	um	Deg	um	Deg	um	Deg	um	Deg	um	Deg
1	Pass	0.23	5	25	12	20	42	154	10	100	39	227	85	63	38	312	60	221
2	Reject	0.11	25	21	16	6	76	156	8	121	70	98	109	73	51	131	66	262

Components will never be manufactured to perfect flat and balanced. Multiple of components will be stacking together during the assembly. Hence, the component will be measure at stages to ensure best matching to reduce the axis offset.





See The Jet Engine page 230 & 231.



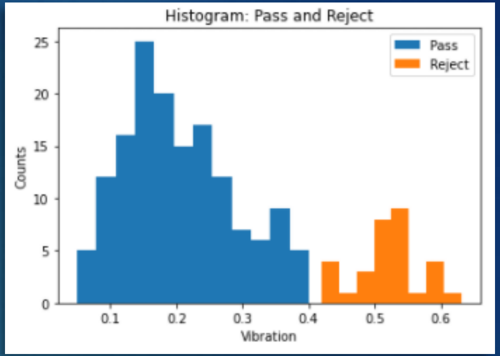
See The Jet Engine page 269 for balancing couple planes.

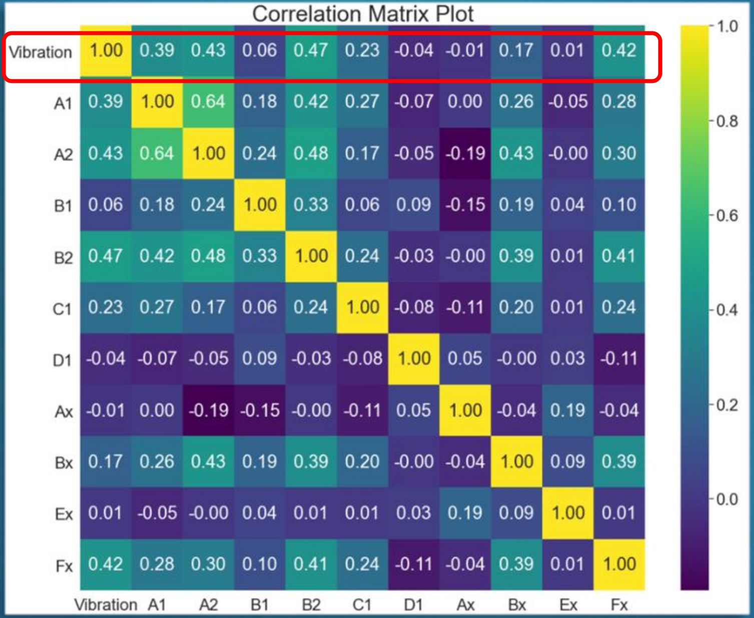
Data Understanding

EDA:

- There are total 180 engine data set, 149 Pass, 31 Reject.
- The Pass and Reject kind of 2 distinctly near to normal distribution groups.
- As from the subject expert observation, the measurements data not having strong relationship to the Vibration reading. Which as demonstrated in the scatter matrix plot.
- Performed regression analysis and drop influencer point not helping. The residual plot too

	Vibration							
	count	mean	std	min	25%	50%	75%	max
PassFail								
Pass	149.0	0.203020	0.084787	0.05	0.140	0.19	0.260	0.40
Reject	31.0	0.519032	0.052240	0.42	0.495	0.52	0.545	0.63





	Vibration	A1	A2	B1	B2	C1	D1	Ax	Bx	Ex	Fx
Vibration	1.00	0.39	0.43	0.06	0.47	0.23	-0.04	-0.01	0.17	0.01	0.42
A1	0.39	1.00	0.64	0.18	0.42	0.27	-0.07	0.00	0.26	-0.05	0.28
A2	0.43	0.64	1.00	0.24	0.48	0.17	-0.05	-0.19	0.43	-0.00	0.30
B1	0.06	0.18	0.24	1.00	0.33	0.06	0.09	-0.15	0.19	0.04	0.10
B2	0.47	0.42	0.48	0.33	1.00	0.24	-0.03	-0.00	0.39	0.01	0.41
C1	0.23	0.27	0.17	0.06	0.24	1.00	-0.08	-0.11	0.20	0.01	0.24
D1	-0.04	-0.07	-0.05	0.09	-0.03	-0.08	1.00	0.05	-0.00	0.03	-0.11
Ax	-0.01	0.00	-0.19	-0.15	-0.00	-0.11	0.05	1.00	-0.04	0.19	-0.04
Bx	0.17	0.26	0.43	0.19	0.39	0.20	-0.00	-0.04	1.00	0.09	0.39
Ex	0.01	-0.05	-0.00	0.04	0.01	0.01	0.03	0.19	0.09	1.00	0.01
Fx	0.42	0.28	0.30	0.10	0.41	0.24	-0.11	-0.04	0.39	0.01	1.00

OLS Regression Results

Dep. Variable:	Vibration	R-squared:	0.330
Model:	OLS	Adj. R-squared:	0.310
Method:	Least Squares	F-statistic:	17.02
Date:	Thu, 22 Oct 2020	Prob (F-statistic):	1.13e-13
Time:	23:20:15	Log-Likelihood:	131.35
No. Observations:	179	AIC:	-250.7
Df Residuals:	173	BIC:	-231.6
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1216	0.020	6.071	0.000	0.082	0.161
A1	0.0004	0.002	0.178	0.859	-0.004	0.005
A2	0.0034	0.002	2.253	0.025	0.000	0.006
B2	0.0056	0.002	3.374	0.001	0.002	0.009
C1	0.0010	0.001	0.935	0.351	-0.001	0.003
Fx	0.0004	0.000	3.588	0.000	0.000	0.001

Omnibus:	4.322	Durbin-Watson:	2.149
Prob(Omnibus):	0.115	Jarque-Bera (JB):	4.306
Skew:	0.343	Prob(JB):	0.116
Kurtosis:	2.673	Cond. No.	311.

Data Preparation

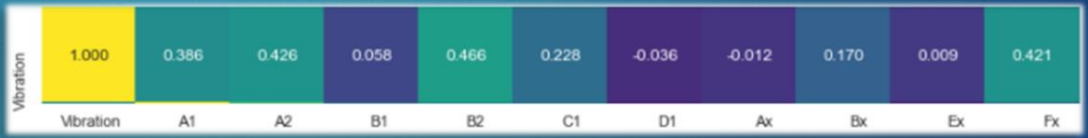
- **Raw Data Collection**

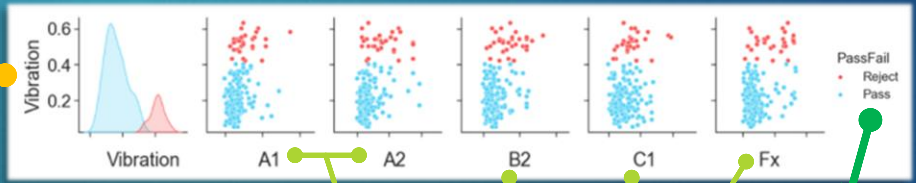
	Label	Label	Flatness Measurement								Balancing Measurement							
			A1	A2	B1	B2	C1	C3	D1	D3	Ax	Ay	Bx	By	Ex	Ey	Fx	Fy
Index	Pass/Fail	Vibration	um	um	um	um	um	Deg	um	Deg	um	Deg	um	Deg	um	Deg	um	Deg
1	Pass	0.23	5	25	12	20	42	154	10	100	39	227	85	63	38	312	60	221
2	Reject	0.11	25	21	16	6	76	156	8	121	70	98	109	73	51	131	66	262

- **Features in "Deg" to be dropped as the turbine is rotating based on the understanding of the operation of HP rotor.**

	Label	Label	Flatness Measurement						Balancing Measurement			
Index	Pass/Fail	Vibration	A1	A2	B1	B2	C1	D1	Ax	Bx	Ex	Fx
1	Pass	0.23	5	25	12	20	42	10	39	85	38	60

- **EDA**
- **Scatter Matrix: Low R-squared**
- **Regression Analysis**



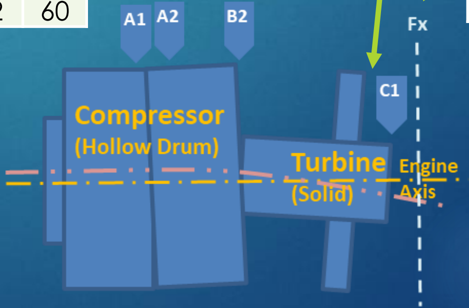


- **Select Features**
- **Divided in to 2 sets: Regression and Categorical**

	Label	Regression Features				
Index	Vibration	A1	A2	B2	C1	Fx
1	0.23	5	25	20	42	60

	Label	Classification Features				
Index	Pass/Fail	A1	A2	B2	C1	Fx
1	Pass	5	25	20	42	60

- **Select Features**
- **The features selected are good physical representation of the full HP module build.**



Create Models

Regression Model

- **Although from the EDA and regression analysis it is expected not gain result for the ML regression model.**
- **Nevertheless, the regression models are created to prepare for the future works. When new features are to be evaluated in regression models.**
- **Overall based on the current set of data, the regression cross validation score at about 0.2. Which unable to predict the Vibration confidently.**

	Label	Regression Features				
Index	Vibration	A1	A2	B2	C1	Fx
1	0.23	5	25	20	42	60

Out[34]:

	Parameters	Train_Accuracy	Test_Accuracy	4-Fold_CV_Score
Multiple Linear Regression	fit_intercept=False	0.235514	-0.007822	0.019539
Random Forest Regression	n_estimators=50, max_depth=4	0.697416	0.120070	0.197553
AdaBoost Regression	n_estimators=50	0.704885	0.182087	0.177160
Support Vector Regression	Kernel='rbf'	0.450199	0.121575	0.227106

Test our best regression on some new measurement data

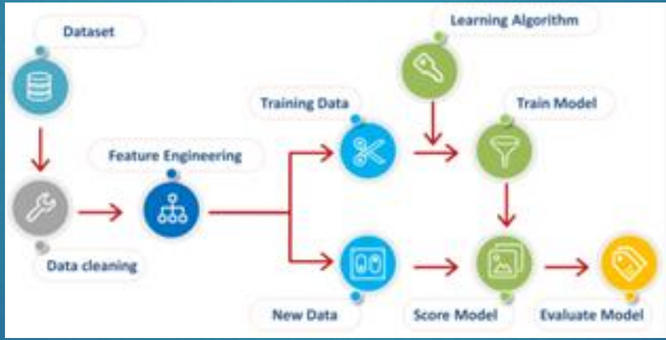
In [51]:

```python
# Measurement data input from a new build HPS:
A1=25   #<25
A2=25   #<25
B2=25  #<25
C1=25   #<25
Fx=250  #<250

measurement = pd.DataFrame({'A1': [A1],'A2': [A2],'B2': [B2],'C1': [C1],'Fx': [Fx]})
measurement = sc.transform(measurement)

#perdict the Vibration of the measurements
prediction=SVR_reg.predict(measurement)
print(f'The Vibration predicted by Support Vector is: {prediction[0]:0.3f}')
```

```
The Vibration predicted by Support Vector is: 0.205
```



Create Models

Classification Model

- **The classification model prediction scores for the small dataset quite well especially on the "Reject" class.**
- **Among the 4 classification models the Logistic Regression scores well as the most reliable model.**

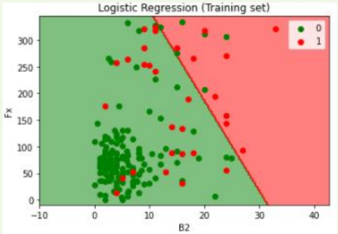
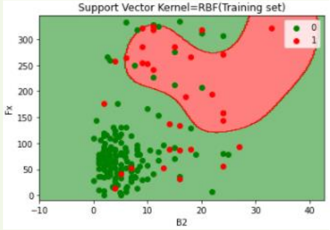
- Highest AUC

- Highest Precision-Recall

- No sign of overfitting

- Most robust model

	Label	Classification Features				
Index	Pass/Fail	A1	A2	B2	C1	Fx
1	Pass	5	25	20	42	60

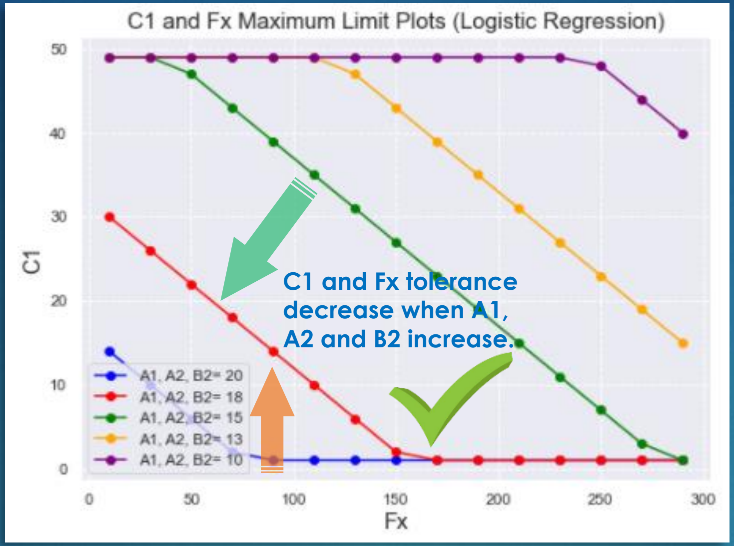
	Classification Models			
	Logistic Regression	Support Vector	K_Neighbors	Random Forest
Characteristics	- Parametric model. - Find the best fit based on the statistical probability. - Sensitive to outlier - Suitable for small dataset.	-Try to maximize the margin between the closest support vectors based on geometry distance. - Not sensitive to outlier. - Suitable for small dataset.	- Non Parametric model. - Use geometry distance to separate the cluster. - Features shall be scaled. - Suitable for large dataset.	- For classification, it use decision tree Yes or No . - Able to provide probability interpretation. - Can handle complex features.
Baseline	Bench mark: 149 Pass, 31 Reject. Accuracy predict: "Pass"=82.7%, "Reject"=17.3%			
CV Accuracy	82%	81%	78%	78%
CV F1 Score	36%	49%	41%	42%
AUC	0.908	0.865	0.810	0.884
Precision _Recall	60%	40%	55%	50%
Train	Accuracy: 83 F1 Score : 42	98 94	91 75	99 96
Test	Accuracy: 86 F1 Score : 29	86 44	86 55	83 40
How each model do the classification				

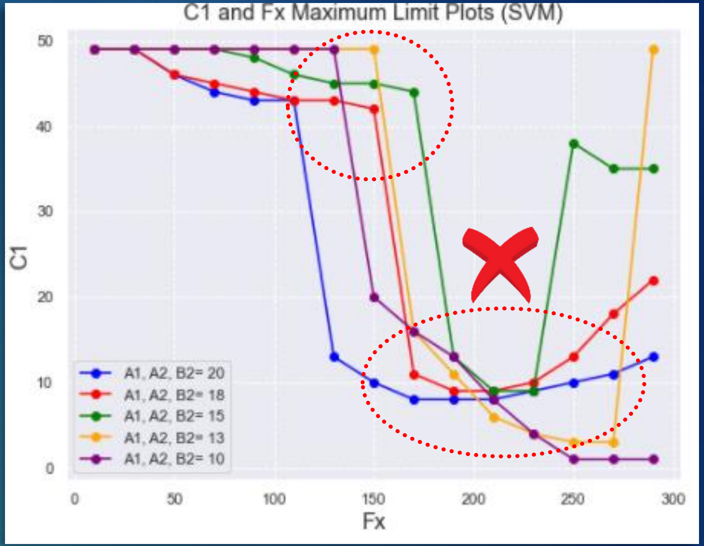
Overfitting: Train set almost perfect scores but not the case for test set.

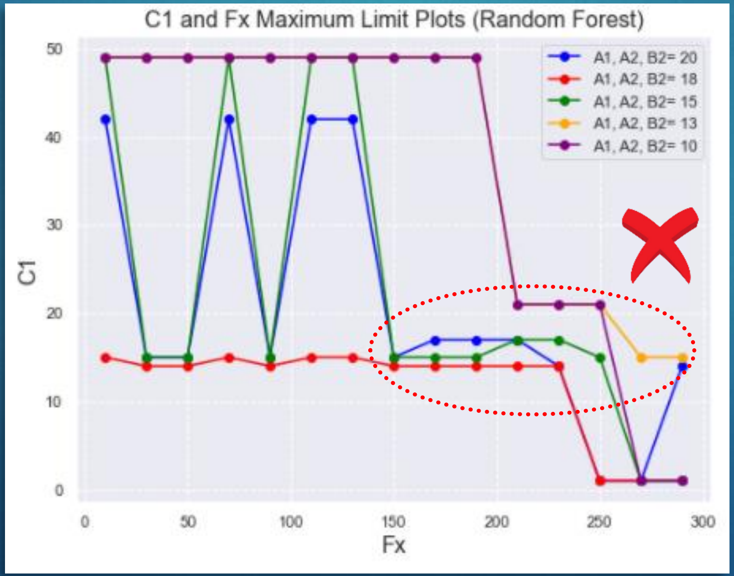
Evaluation

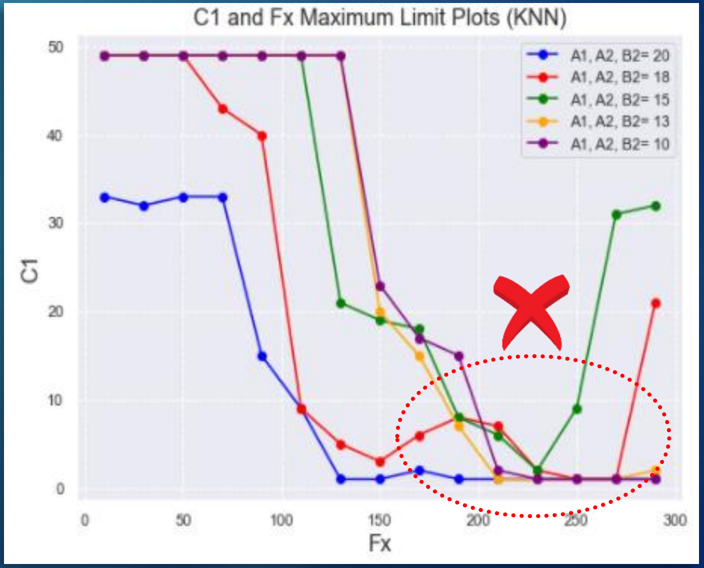
Test the model

- A simulation of the models been carried out by using Compressor features A1, A2 and B2 to predict the Turbine feature C1.
- Rule: If the Compressor features values A1, A2 and B2 increase, the available tolerance of the Turbine C1 will need to decrease in order to prevent vibration reject at engine test.
- In this case only the Logistic Regression able show consistency to the Rule.









Deployment Proposal

How the model help:

1. **Predict if the HP module will cause high vibration during Test.**
2. **At build in-progress of HP module, with the compressor measurement, the model able to predict the passing limits for Turbine C1 and Fx.**

Logistic Regression Model

Coefficient:

	Feature name	Coefficient	Odds_ratio
3	B2	0.770891	2.161691
5	Fx	0.453616	1.573994
1	A1	0.408341	1.504320
4	C1	0.237067	1.267526
2	A2	0.199571	1.220878
0	Intercept	-2.104775	0.121873

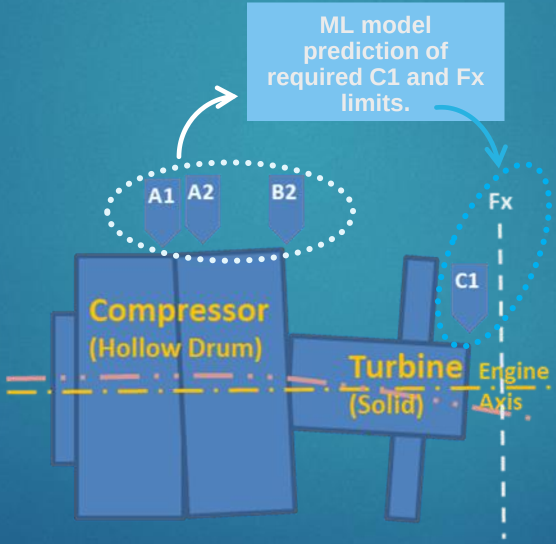
1. Sentencing the Pass or Reject with the HPS Measurements

```
# Measurement data input:
A1=25    #<25
A2=25    #<25
B2=25    #<25
C1=10    #<25
Fx=350   #<250

measurement = pd.DataFrame({'A1': [A1],'A2': [A2],'B2': [B2],'C1': [C1],'Fx': [Fx]})
measurement = sc.transform(measurement)

#perdict the Vibration of the measurements
prediction=pd.DataFrame(LR_classifier.predict(measurement))
#print(prediction)
result=prediction[0].map({0:'Pass', 1:'Reject'});
print(f'The Vibration predicted by Logistic Regression classifier is: {result[0]}')
```

```
The Vibration predicted by Logistic Regression classifier is: Reject
```

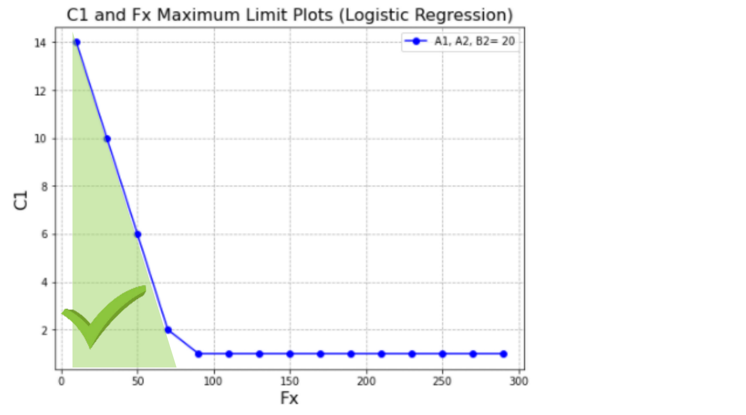


2. Predicting the required Turbine Parameters C1 and Fx with the Compressor Measurements

```
# Measurement data input from Compressor
A1=20    #<25
A2=20    #<25
B2=20    #<25

# Find the limit of the C1 and Fx on the matching Turbine
C1_list=[]
Fx_list=[]
C1_range=np.arange(1,50,1)
Fx_range=np.arange(10,300,20)
for i in Fx_range:
    for j in C1_range:
        measurement = pd.DataFrame({'A1': [A1],'A2': [A2],'B2': [B2],'C1': [j],'Fx': [i]})
        measurement = sc.transform(measurement)
        prediction=LR_classifier.predict(measurement)
        if prediction>0 :
            break
    C1_list.append(j)
    Fx_list.append(i)

# C1 and Fx predict tramline
plt.figure(figsize=(8,6))
plt.plot(Fx_list,C1_list,'o-',label='A1, A2, B2= 20', color='blue')
plt.grid(linestyle='--')
plt.xlabel('Fx',fontsize=16)
plt.ylabel('C1',fontsize=16)
plt.legend()
plt.title('C1 and Fx Maximum Limit Plots (Logistic Regression)',fontsize=16);
```



Deployment Proposal

Propose Strategy:

Drawing Specification				
A1	A2	B2	C1	Fx
<25	<25	<25	<25	<200

The Compressor and Turbine are build separately before assemble become full HP module.

1. Reduce the reject by Swap Turbines

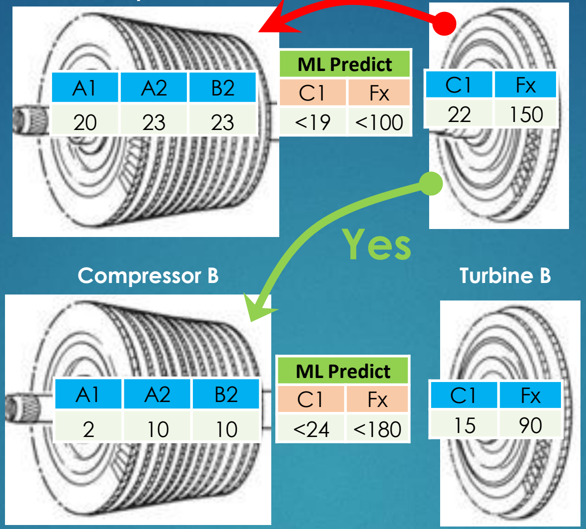
2. Reduce the reject due to drawing spec one individual measurement.

3. For the chief engine to evaluate if the out of specification can be accepted. Increase speed of decision making.

4. Batch Matching to reduce the engine reject at Test Best. For example get the best combination among 5 HP compressors and turbines sets in order to achieve best margin to vibration reject.

1. Compressor and Turbine Swap

(Improve HP Build Yield)



Compressor A:

A1	A2	B2
20	23	23

ML Predict	
C1	Fx
<19	<100

Turbine A:

C1	Fx
22	150

Compressor B:

A1	A2	B2
2	10	10

ML Predict	
C1	Fx
<24	<180

Turbine B:

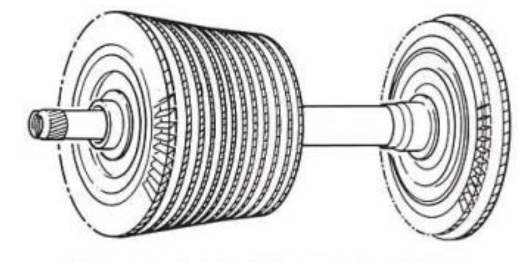
C1	Fx
15	90

2. Drawing Spec VS ML Model

(Improve HP Build Yield)

Drawing Spec: Failed				
A1	A2	B2	C1	Fx
15	15	13	28	70

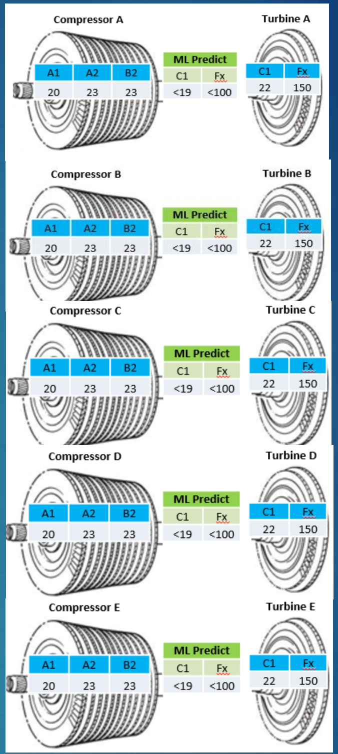
ML Predict : Pass				
A1	A2	B2	C1	Fx
15	15	13	28	70

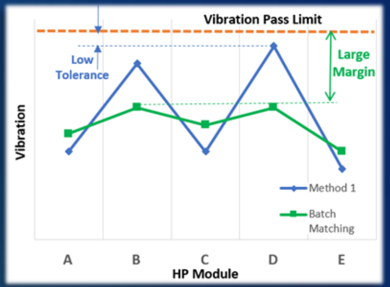


- 70% of the total rejected at HP build due to C1 measurement out of specification.

4. Batch Matching

(Improve Engine Test Yield)





Best Combination with the largest margin to vibration reject at Test.

3. Faster Approval of out of Specification Measurement

(Improve HP Build and Engine Test Yield)

Chief Design Engineer

Conclusion

The machine learning is a effective tool to improve the yield of the production which can be easily proposed to the business because little capital investment is required.

Challenges:

- **Small training dataset prone to overfitting.**
- **Yet to create a useful regression model which will be better illustrate the insight of vibration.**

Future Works:

- **Create a module of the model which can be used in standard application like Tableau to enable technician to use at production line.**
- **Create a report on how the model can help the Design Engineer to revise the drawing specification from individual measurement specification to matrix type specification.**

Lesson Learned:

- **Understanding the Business Needs, Processes and Operations from subject matter experts of relevant departments will be the keys to the success of a machine learning project.**

Thank You.