PROJECT PRODCO

TEAM PURPLE

RALUCA JOAN GRACE KATIE PREETY

Challenges

SKU	VALUE	PASS	TOTAL VALUE
A	300	11015	3,304,500
B	250	6532	1,633,000
C	400	22696	9,078,400
X	1000	5354	5,354,000
Z	600	9096	5,457,600
			24,827,500

SKU	VALUE	DEFECT	TOTAL VALUE LOST to DEFECT
A	300	6545	1,963,500
B	250	1826	456,500
C	400	8850	3,540,000
X	1000	1462	1,462,000
Z	600	2335	1,401,000
			8,823,000

New Production Line

A new integrated production line installed last year. However it is experiencing high defect rate which prevented PRODCO reaching their production target.

High Costs

The installation of new production line costs 100M. Value lost to defect amount to 8M with additional personnel costs for inspection.

Complaints and Backlog

Unable to detect defects in some shipped production batches resulted in high volumes of compliant claims. Additional inspection time is also causing backlogs on orders.

Objectives

To build an analytical model based on the 75K records of production data from November 19’ to January 20’. Using Production Process information, such as Position, Humidity, Temperature, Duration in each zones and the Block to create the certain product to explore the root cause of defect.

Root Cause

Develop a defect detection model will enable engineering team to pinpoint the location of issue.

Cost-saving

Using this model as Proof of Concept to understand the cause of defect can help reduce future personnel costs.

Decision-Making

Make recommendation based on the model and understanding the business value; help PRODCO to make data-driven decisions.

Our approach

	Result_Type				
SKU	Defect_1	Defect_2	Defect_3	Defect_4	Defect_5
A001	55.16% 3,610	1.83% 120	24.52% 1,605	9.23% 604	9.26% 606
B003	2.96% 54	2.68% 49	54.11% 988	20.87% 381	19.39% 354
C005	2.59% 229	42.01% 3,718	26.23% 2,321	15.27% 1,351	13.91% 1,231
X007	4.24% 62	3.49% 51	49.25% 720	20.79% 304	22.23% 325
Z009	3.38% 79	3.81% 89	50.24% 1,173	21.80% 509	20.77% 485

Data Cleaning

Missing values (<1%) inputted

Duplicates and redundant removed e.g. Date, Index

Statistical Tool

Python & R

Data Exploratory

Histogram, Scatter plot, Correlation Matrix

Models

Decision Tree, Random Forest, Multinomial Regression, Gradient Boosting

Model Selection

Decision Tree

a) Comprehensive - traces each path to a conclusion;

b) Specific - assigns values to each decision path and outcome;

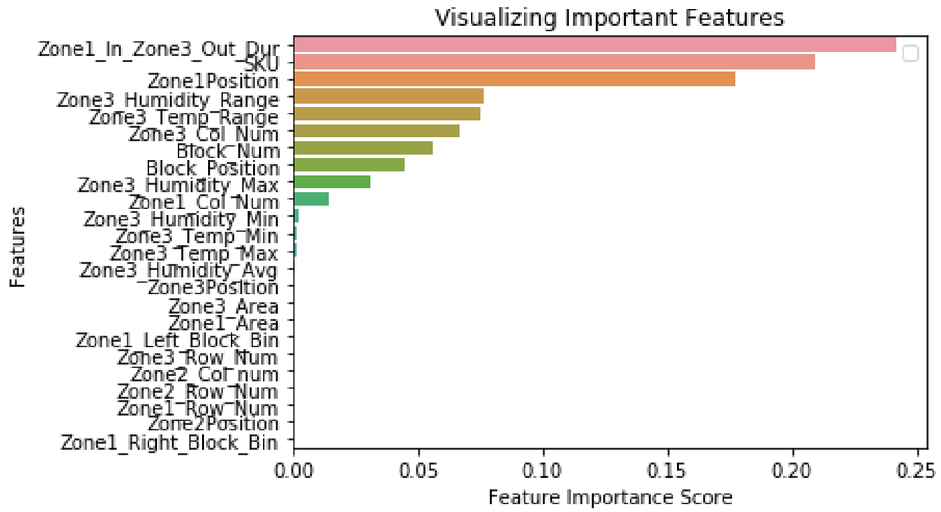
c) Versatile - applicable to our business problems;

Model Tuning

1) Features selection

Boruta (Python) – 53 ➔ 25

Regression Trees – 25 ➔ 10



Model Tuning

2) Performance

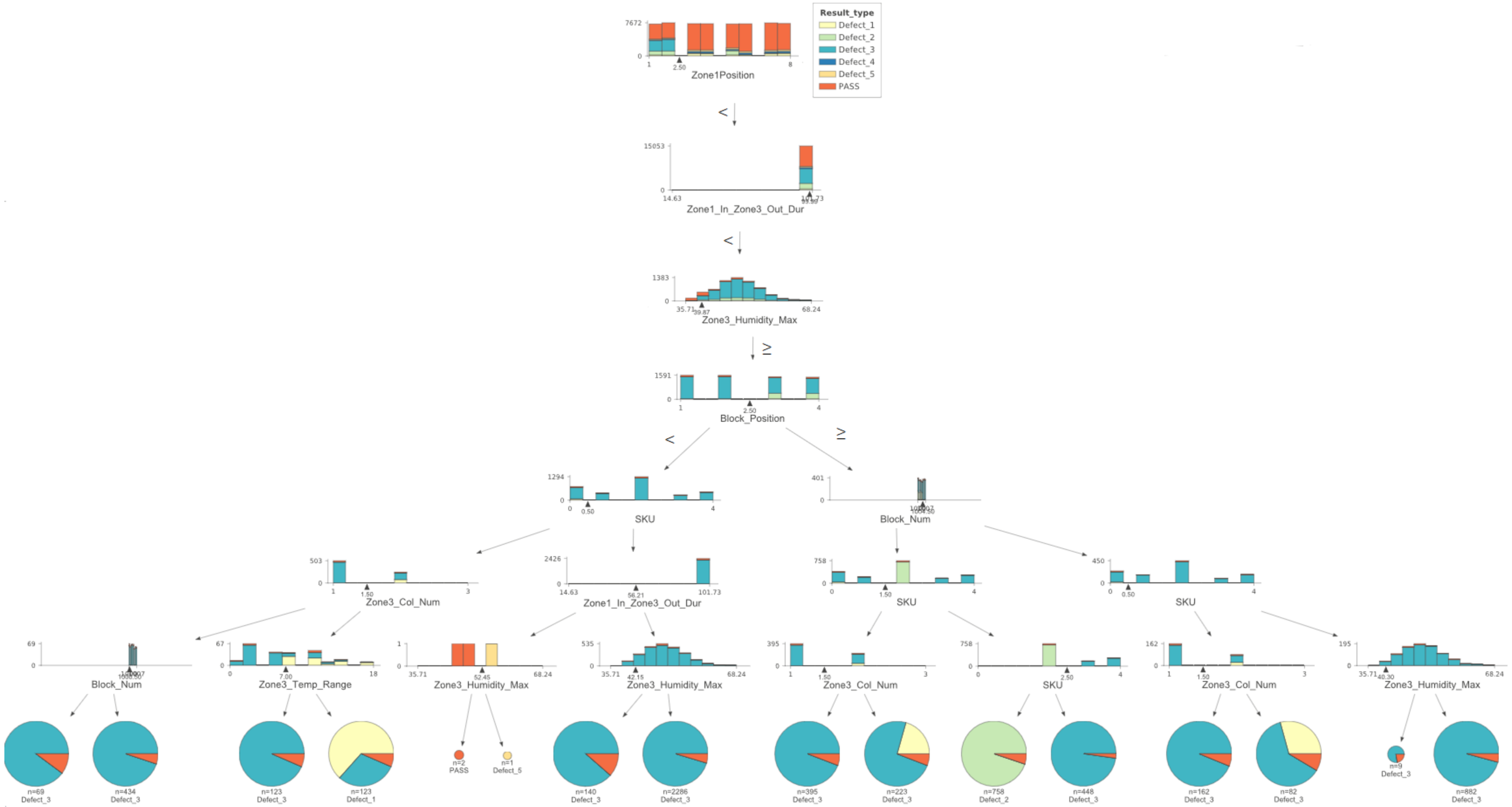
Max depth	*Accuracy*	*MAE (°)*
1	0.7214	0.87
2	0.792533	0.61
3	0.796933	0.61
4	0.8244	0.46
5	0.8436	0.38
6	0.867267	0.3
7	0.881133	0.26
8	0.881267	0.26
9	0.880133	0.26
10	0.8786	0.26
11	0.873733	0.27
12	0.8676	0.29
13	0.861267	0.3
14	0.852467	0.32
15	0.844867	0.34

Max depth 7
Highest Accuracy .88
Lowest Mean Absolute Error 0.26

Decision Tree Model Evaluation

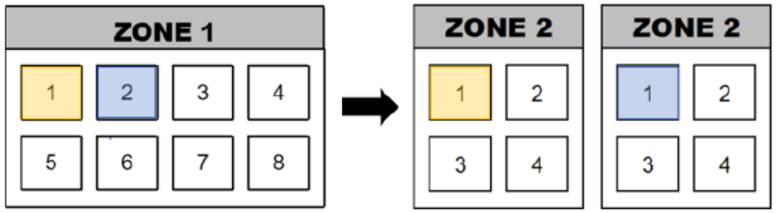
	Defect 1	Defect 2	Defect3	Defect 4	Defect 5	PASS
Defect 1	656	0	10	0	0	42
Defect 2	0	689	0	0	0	33
Defect3	16	0	1207	0	0	76
Defect 4	0	0	0	0	0	0
Defect 5	0	0	0	0	0	4
PASS	126	110	105	622	638	10666

	precision	recall	f1-score	support
Defect 1	0.926554	0.822055	0.871182	798
Defect 2	0.954294	0.862328	0.905983	799
Defect3	0.929176	0.913011	0.921023	1322
Defect 4	0	0	0	622
Defect 5	0	0	0	638
PASS	0.869487	0.985676	0.923943	10821
accuracy	0.8812			
macro avg	0.613252	0.597178	0.603688	15000
weighted avg	0.809264	0.8812	0.842311	15000

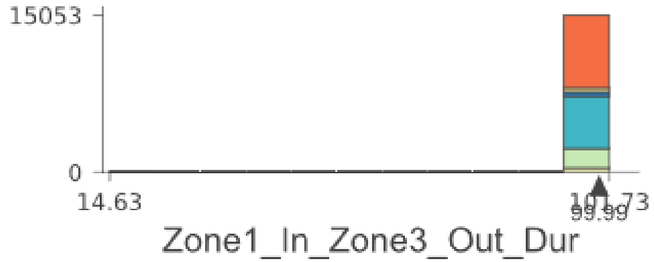

Result_type
Defect_1
Defect_2
Defect_3
Defect_4
Defect_5
PASS
Zone1Position
Zone1_In_Zone3_Out_Dur
Zone3_Humidity_Max
Block_Position
SKU
Block_Num
Zone3_Col_Num
Zone1_In_Zone3_Out_Dur
SKU
SKU
Block_Num
Zone3_Temp_Range
Zone3_Humidity_Max
Zone3_Humidity_Max
Zone3_Col_Num
SKU
Zone3_Col_Num
Zone3_Humidity_Max
n=69 Defect_3
n=434 Defect_3
n=123 Defect_3
n=123 Defect_1
n=2 PASS
n=1 Defect_5
n=140 Defect_3
n=2286 Defect_3
n=395 Defect_3
n=223 Defect_3
n=758 Defect_2
n=448 Defect_3
n=162 Defect_3
n=82 Defect_3
n=9 Defect_3
n=882 Defect_3

Main Causes of Defects

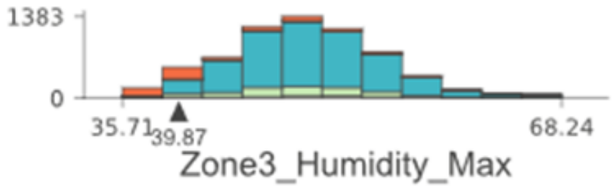
1. Position in Zone 1 (1 & 2)



2. Duration from Zone 1 to Zone 3 (<= 99.99 Seconds)



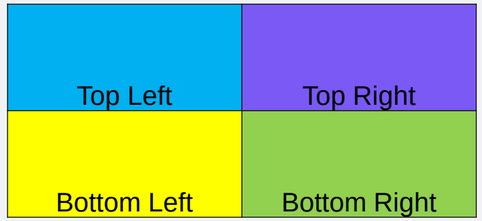
3. Humidity Max in Zone 3 (> 39.87%)



Recommendations

01

Explore the possibility of placing product in Top Right, Bottom Right, Bottom Left in Zone1 and the environmental differences between Top Left Area and the rest.



02

A further investigation on products spending shorter than 99.99 seconds between Zone 1 and Zone3.

03

Work with the engineering team to test out if ensuring humidity range lower than 39.87% in Zone3 will help reduce defect rate.

Q&A

RALUCA, JOAN, GRACE, KATIE, PREETY & Special Thanks to Vivek