



BROCHALIGN

A Multi-Axis Regression Approach for ITP Aero Manufacturing Data

Dante Schrantz and Miguel Diaz

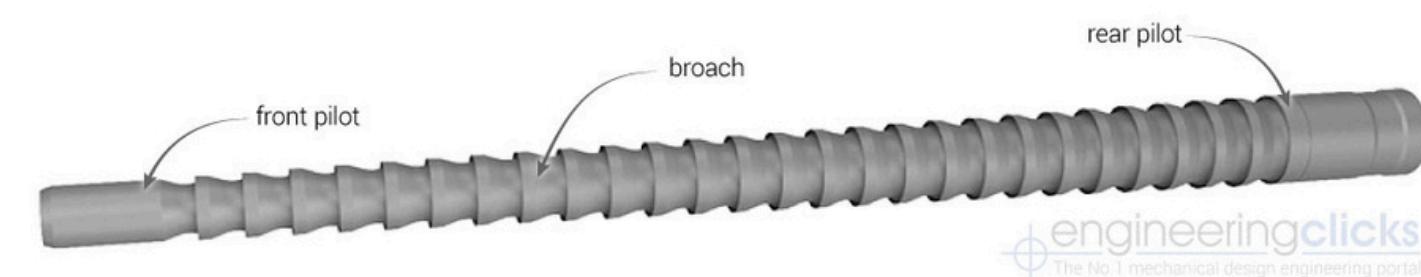
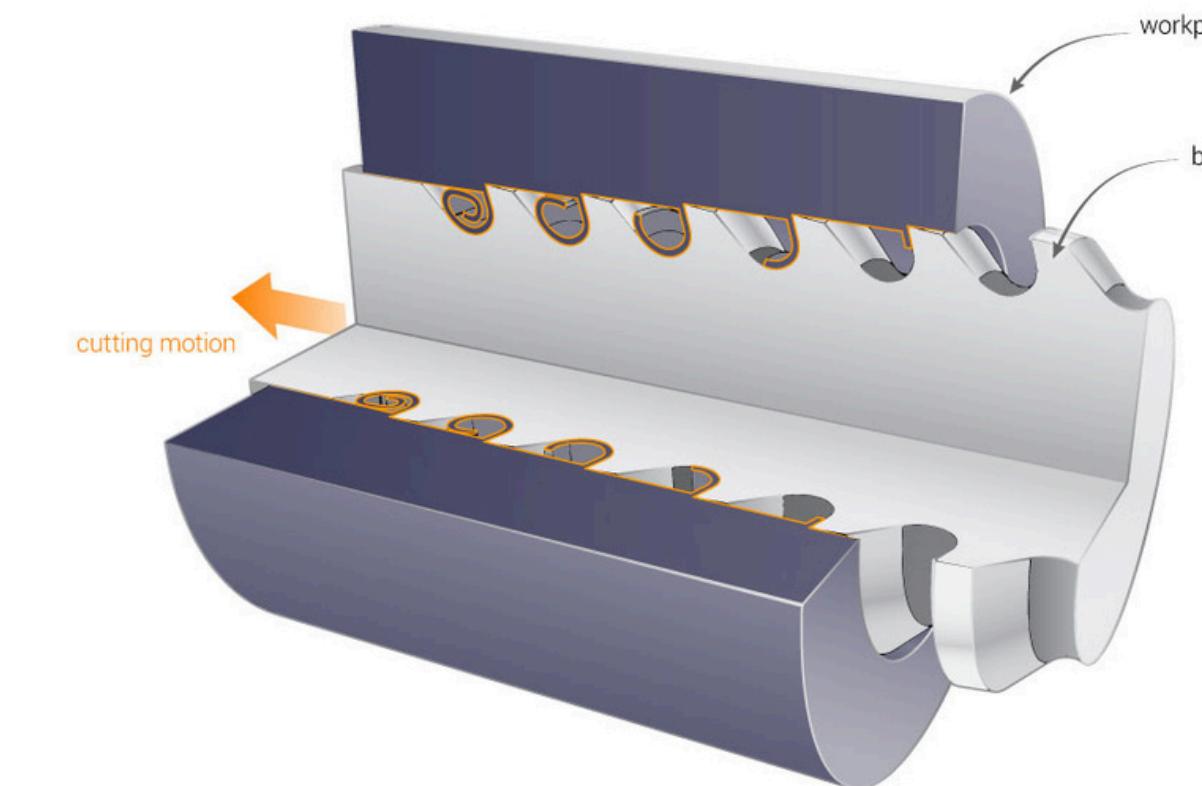
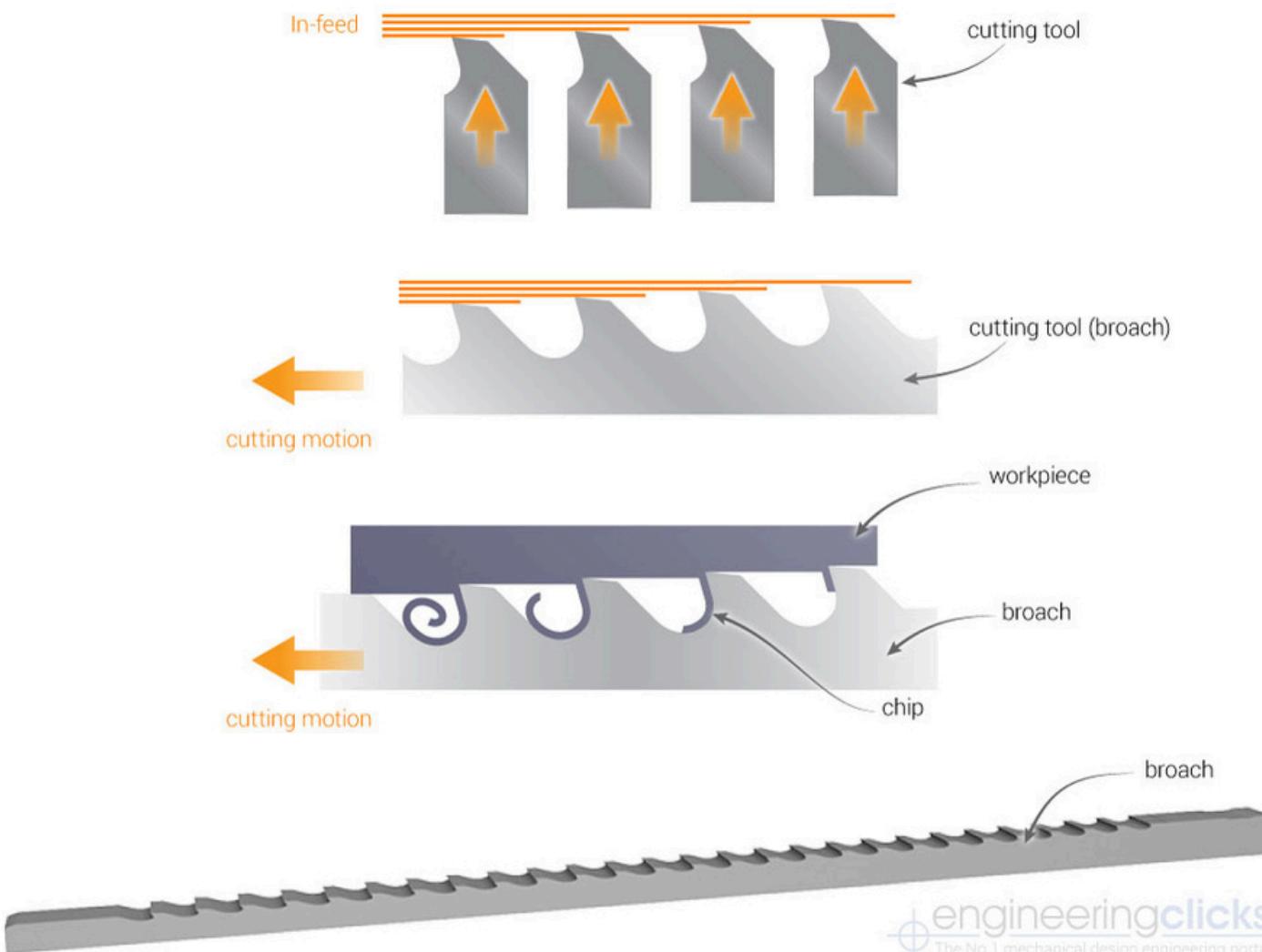
Copyright: All aircraft photographs in this presentation are the property of Dante Shrantz.
Their use, reproduction, or distribution without authorization is strictly prohibited.



Purpose



IMPLEMENT A MACHINE LEARNING MODEL TO PREDICT WHEN BROACHING COMPONENTS NEED TO BE REPLACED, REDUCING COSTS AND OPTIMIZING PRODUCTION LINE EFFICIENCY.



Metrics and Requirements

THE MAXIMUM ABSOLUTE ERROR OF THE MODEL ON THE VALIDATION DATASET MUST BE LESS THAN 1 FOR THE CORRECTOR ON THE X-AXIS AND 0.15 FOR THE OTHER AXES. ADDITIONALLY, THE RMSE MUST BE LESS THAN 0.25 FOR THE CORRECTOR ON THE X-AXIS AND 0.025 FOR THE OTHER AXES.



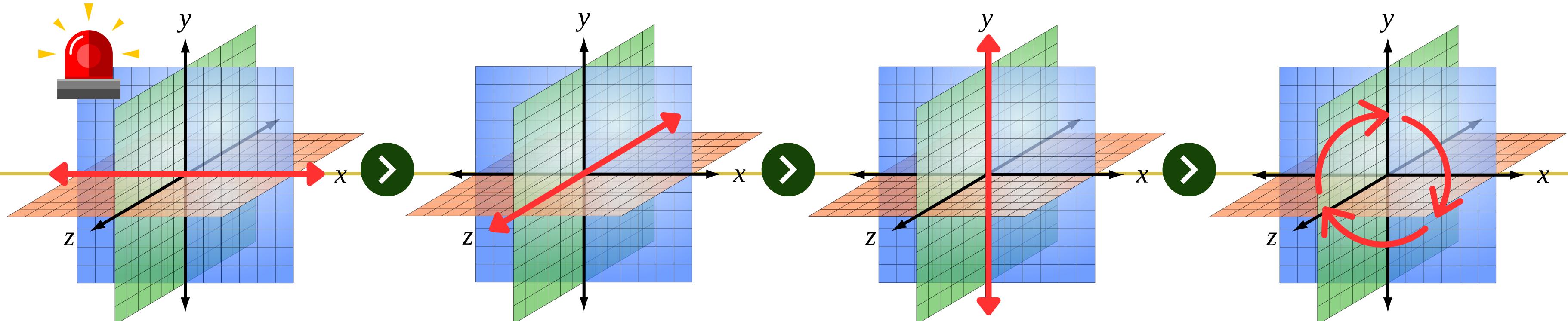
Objective

Improve the quality of products manufactured by ITP Aero and reduce their production costs.



Target Variables - Correctors

WE HAVE 4 TARGET VARIABLES CORRESPONDING TO **EACH AXIS OF THE BROACHING PROCESS**



XCMM

RIGHT TO LEFT

ZCMM

FORWARD TO
BACKWARD

CCMM

FROM TOP TO
BOTTOM

BCMM

ROTATIONAL



Actual Corrections

VALUES DIRECTLY APPLIED BY OPERATORS OR SYSTEMS TO ATTEMPT TO COMPENSATE FOR DEVIATIONS.



CMM Corrections

OUR MODEL PREDICTS MACHINE-MEASURED DEVIATIONS (XCMM, ZCMM, BCMM, AND CCMM) BASED ON HISTORICAL SYSTEM DATA.

Final Corrections:

XC + XCMM ↳ Final correction on X-axis.

ZC + ZCMM ↳ Final correction on Z-axis.

BC + BCMM ↳ Final correction on B-axis.

CC + CCMM ↳ Final correction on C-axis.



BROCHALIGN



Explanatory Variables

c("Brocha", "PartNumber",
"OrdenFabricacion",
"FBrochado", "Utilaje",
"Maquina",
"TpolIndexador", "NUso",
"NDisco")

9

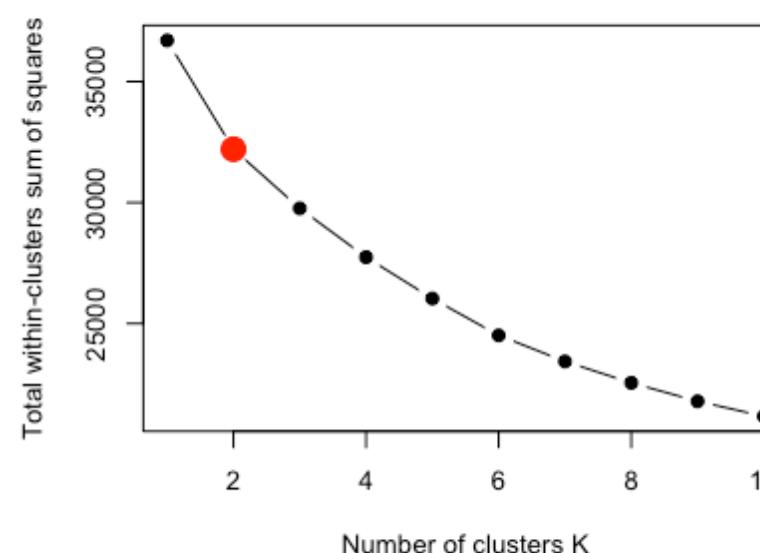
• • •

Unsupervised Clustering

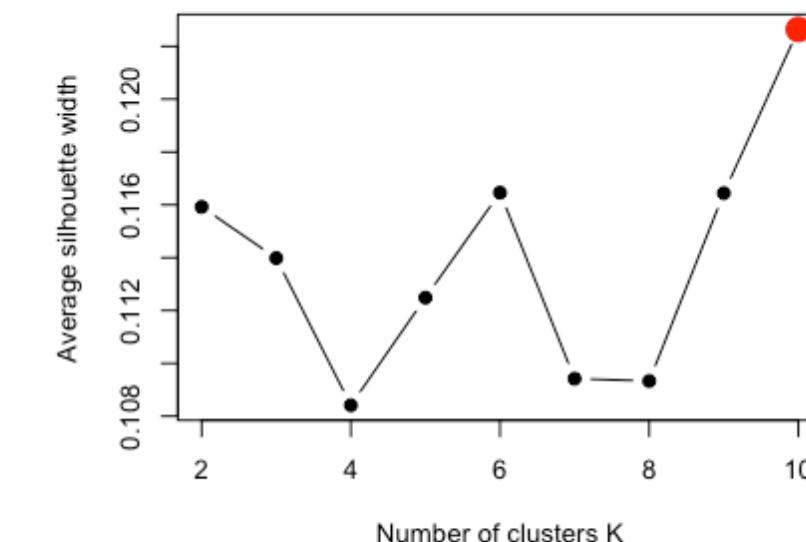
Total time taken: 2.06 mins

HAVING ONLY 2 CLUSTERS IS THE MOST OPTIMAL FOR OUR DATA. IN FACT, THE CLUSTERS ARE WELL-SEPARATED, AND THERE IS NO OVERLAPPING.

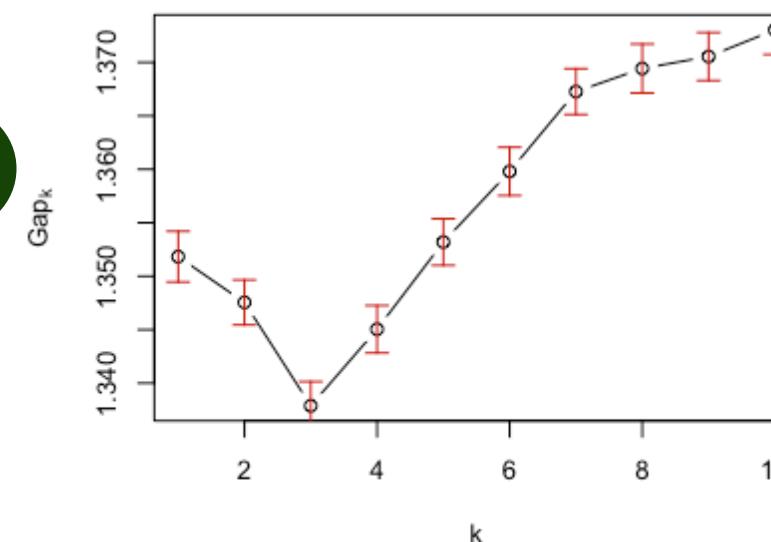
Elbow Method



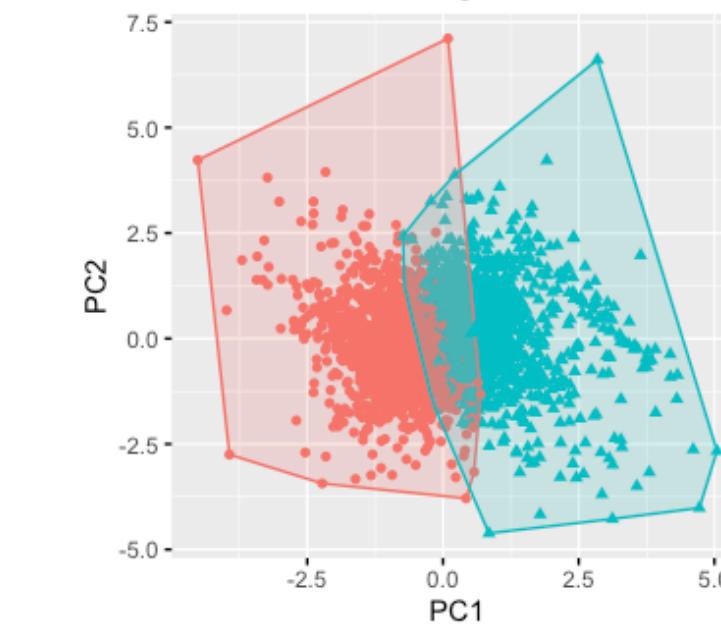
Silhouette Method



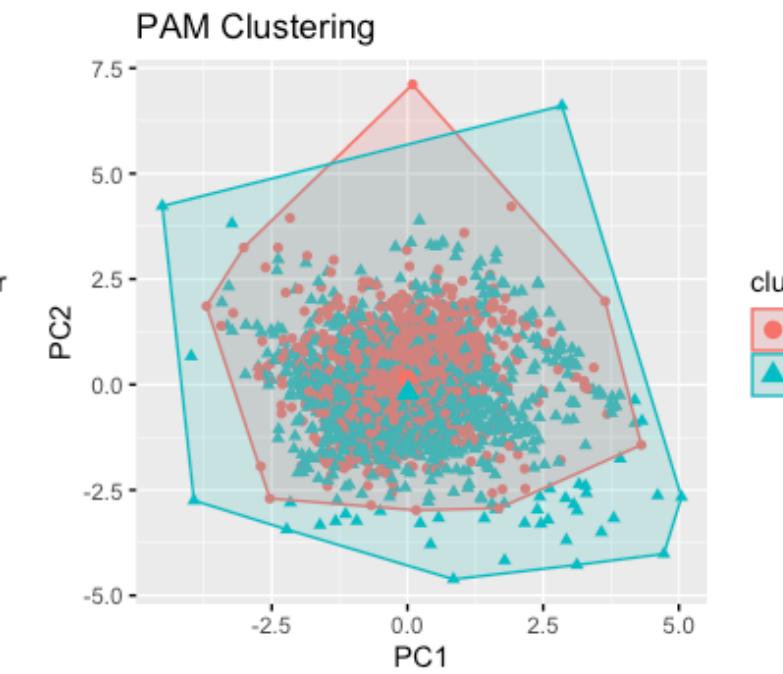
Gap Statistic Method



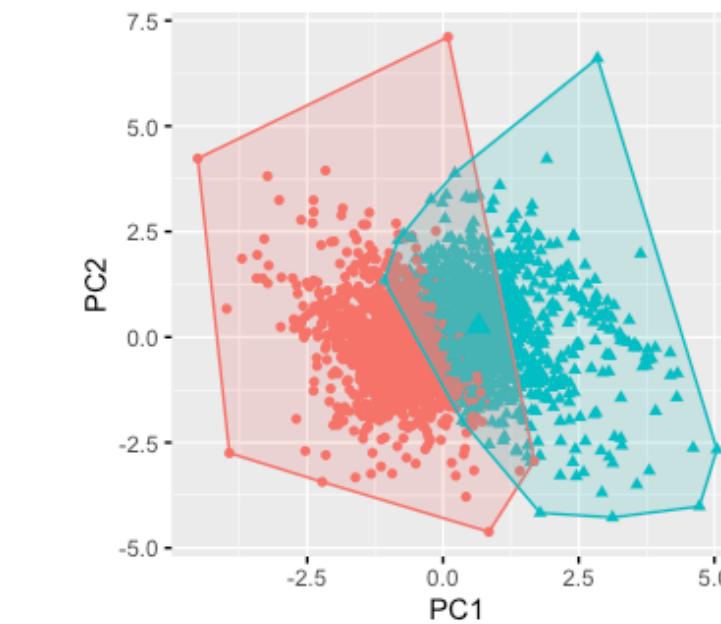
K-means Clustering



PAM Clustering



Fuzzy C-means Clustering



BROCHALIGN

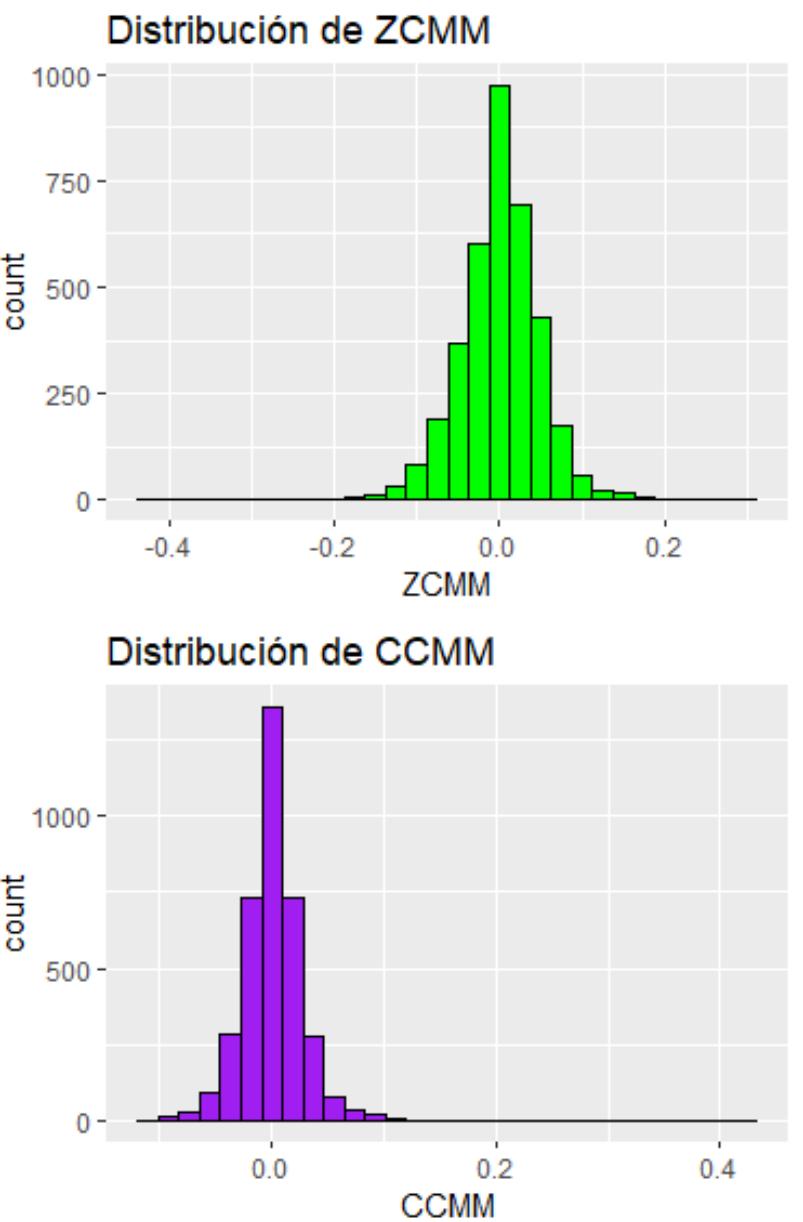
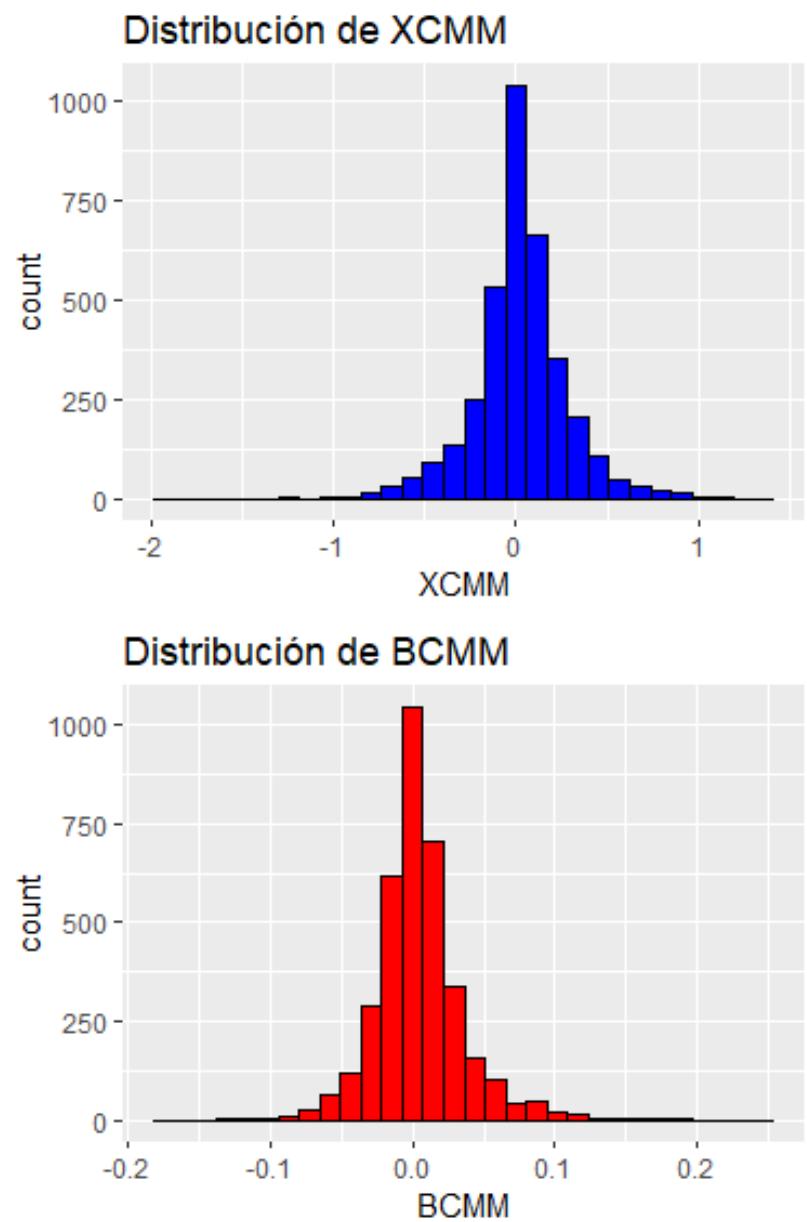
Supervised Regression



BROCHALIGN

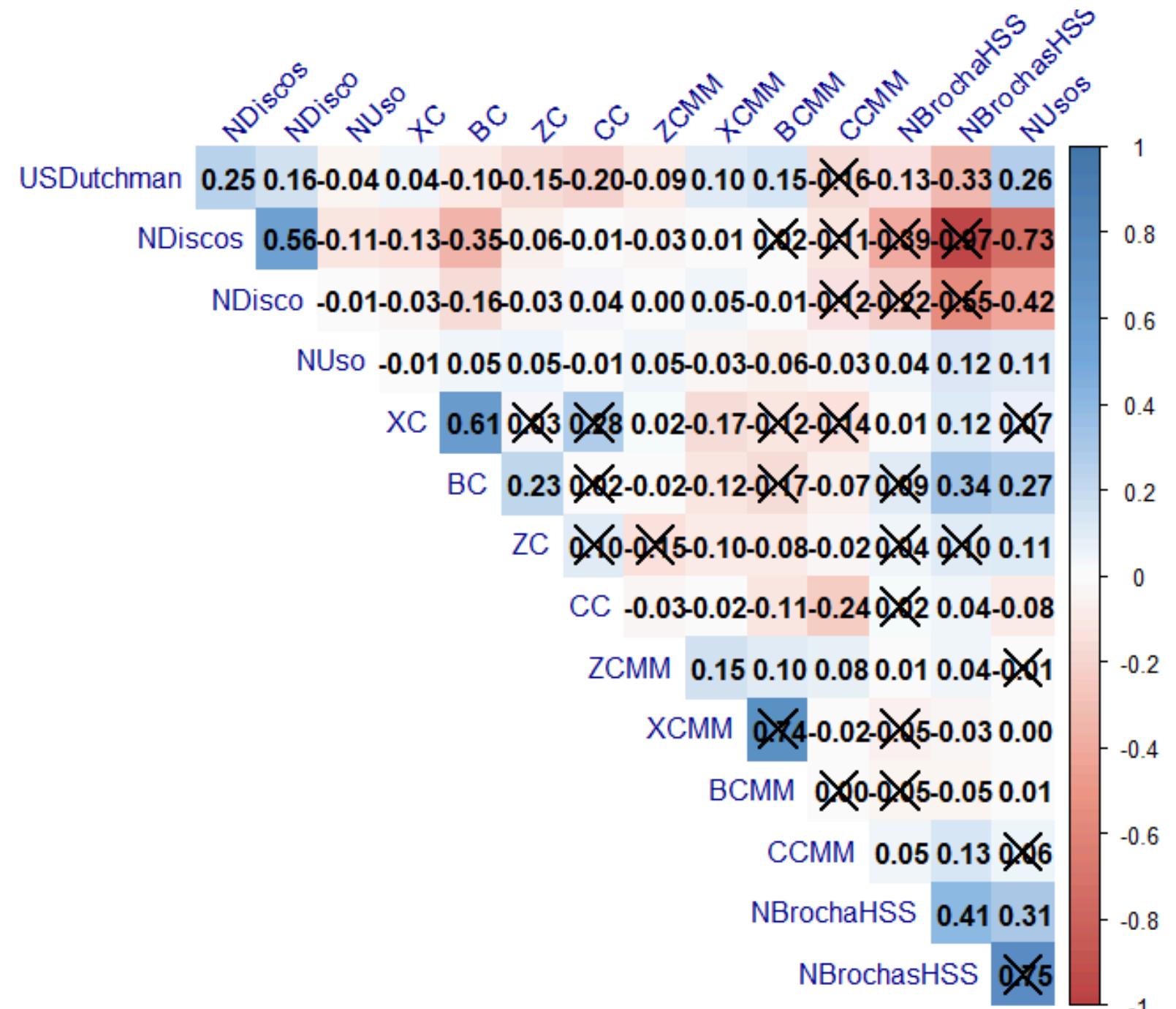
Distribution of Target Variables

ALL TARGET VARIABLES HAVE A FAIRLY NORMAL DISTRIBUTION. XC IS THE VARIABLE WITH THE HIGHEST VARIANCE AND IS THE MOST DIFFICULT TO PREDICT.



FEATURE ENGINEERING

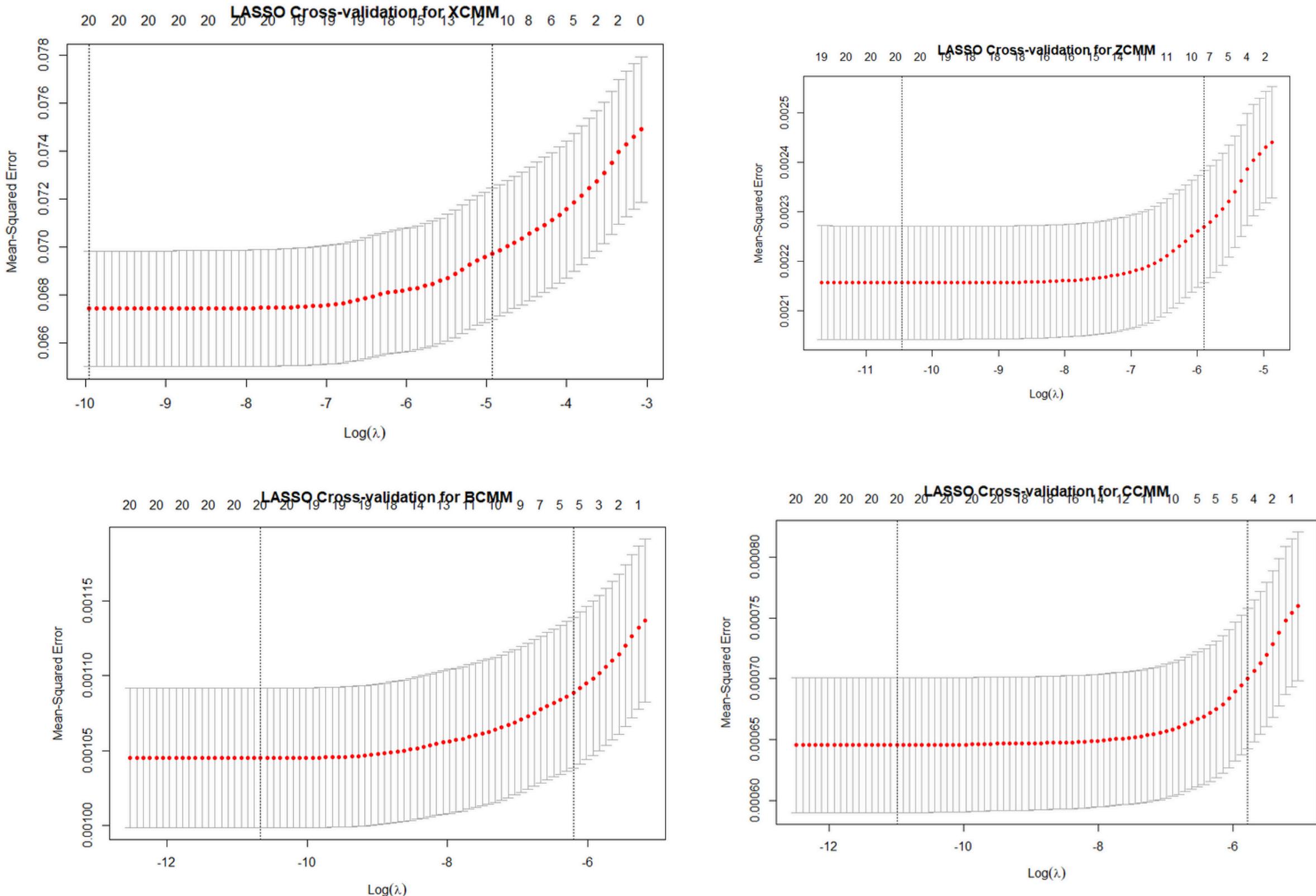
Correlation



THERE ARE CERTAIN VARIABLES THAT ARE HIGHLY CORRELATED WITH EACH OTHER. FOR EXAMPLE, THE X-AXIS CORRECTOR (XCMM) IS HIGHLY CORRELATED WITH THE B-AXIS CORRECTOR (BCMM). THERE IS ALSO A HIGH CORRELATION BETWEEN NBROCHASHSS AND NUSOS OF 0.75.

Results Lasso Regression

ALL OUR VARIABLES ARE SIGNIFICANT, SO WE CANNOT OMIT ANY VARIABLE FROM OUR MODEL.

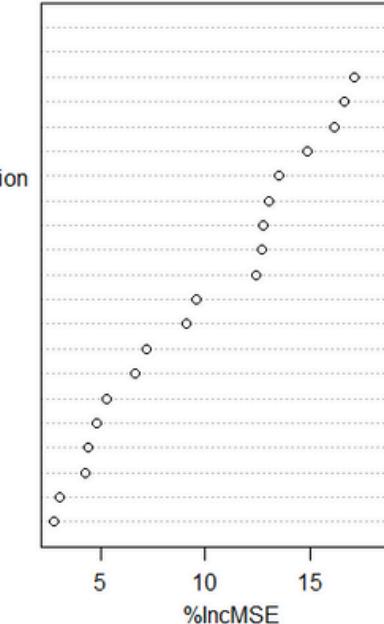


Importance of Dependent Variables

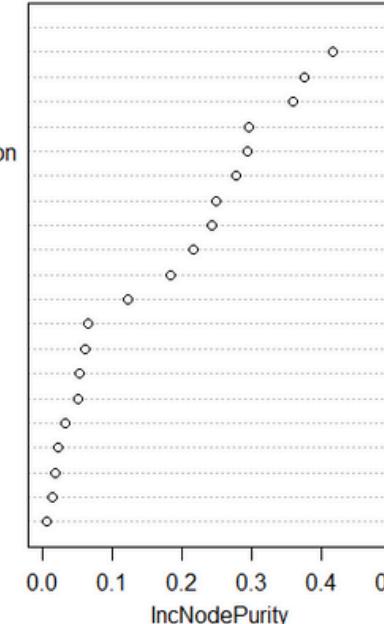
THE MOST SIGNIFICANT
DEPENDENT VARIABLES ARE
UTILAJE, MAQUINA, BROCHA,
TIPO INDEXADOR, Y
ORDENFABRICACION



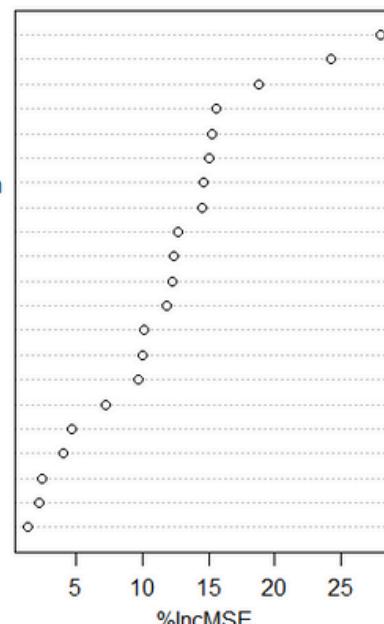
XC
BrochaSN
PartNumber
FBrochado
ZC
Utilaje
Maquina
Brocha
BC
CC
OrdenFabricacion
Tpolndexador
NUsos
NUsos
Dutchman
NDisco
NBrochaHSS
NDiscos
NBrochasHSS
DUMMY
USDutchman



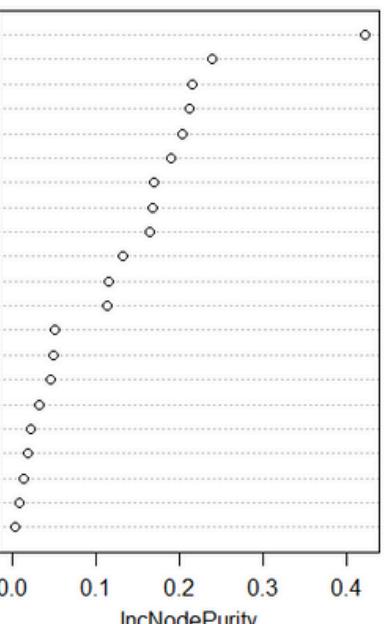
XC
FBrochado
BrochaSN
BC
OrdenFabricacion
ZC
Utilaje
PartNumber
CC
NUsos
NDisco
Brocha
Maquina
Tpolndexador
NBrochaHSS
DUMMY
Dutchman
NUsos
NDiscos
NBrochasHSS
USDutchman



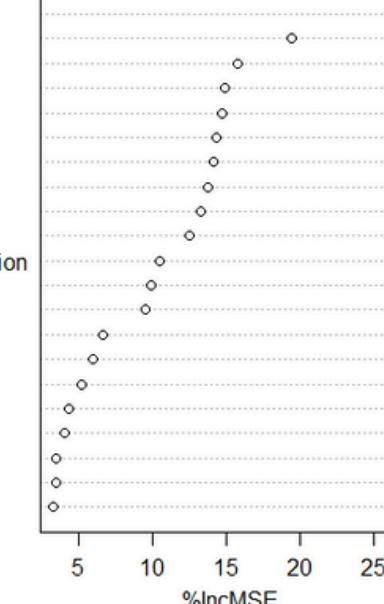
CC
NUsos
PartNumber
BrochaSN
Dutchman
BC
OrdenFabricacion
NDisco
Maquina
FBrochado
ZC
Brocha
Tpolndexador
Utilaje
XC
DUMMY
USDutchman
NUsos
NDiscos
NBrochasHSS
NBrochaHSS



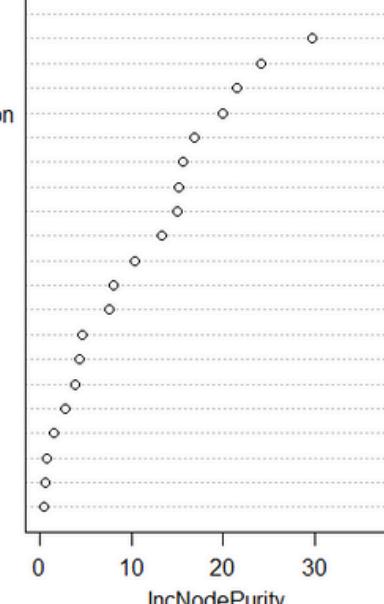
CC
OrdenFabricacion
FBrochado
XC
NUsos
PartNumber
BrochaSN
Utilaje
ZC
BC
NDisco
Brocha
Maquina
Tpolndexador
DUMMY
Dutchman
NBrochaHSS
USDutchman
NUsos
NDiscos
NBrochasHSS



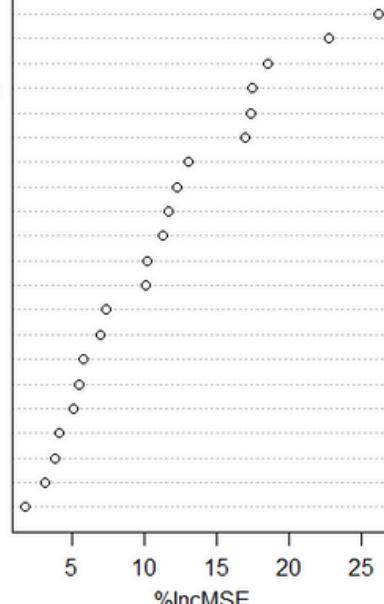
XC
BrochaSN
PartNumber
FBrochado
ZC
Utilaje
Maquina
Brocha
BC
CC
OrdenFabricacion
Tpolndexador
NUsos
NUsos
Dutchman
NDisco
NBrochaHSS
NDiscos
NBrochasHSS
DUMMY
USDutchman



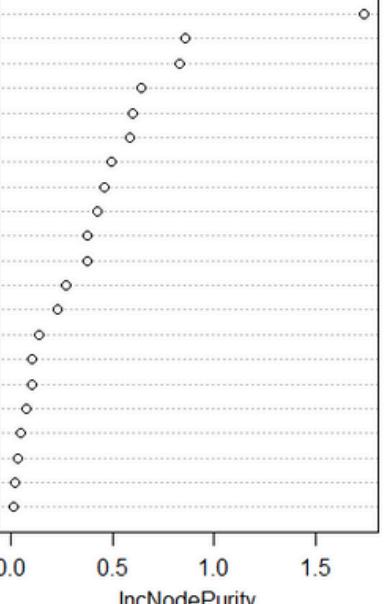
XC
FBrochado
BrochaSN
BC
OrdenFabricacion
ZC
Utilaje
PartNumber
CC
NUsos
NDisco
Brocha
Maquina
Tpolndexador
NBrochaHSS
DUMMY
Dutchman
NUsos
NDiscos
NBrochasHSS
USDutchman



FBrochado
ZC
Maquina
OrdenFabricacion
Utilaje
PartNumber
XC
Brocha
NUsos
CC
BC
BrochaSN
Tpolndexador
NUsos
DUMMY
NDisco
Dutchman
NDiscos
NBrochasHSS
USDutchman
NBrochaHSS



FBrochado
ZC
OrdenFabricacion
XC
BrochaSN
BC
PartNumber
Utilaje
Maquina
NUsos
CC
Brocha
NDisco
Tpolndexador
NBrochaHSS
DUMMY
Dutchman
NUsos
NDiscos
NBrochasHSS
USDutchman



--- Results of the **Linear Regression Model** ---

WE COMPARED ALGORITHMS USING THE PROFESSOR'S BASE CODE, MOSTLY WITHOUT MAJOR MODIFICATIONS TO ESTABLISH A BASELINE AND EVALUATE THE PERFORMANCE OF EACH MODEL FOR OUR PROBLEM.

SURPRISINGLY, THIS MODEL PERFORMS REMARKABLY WELL COMPARED TO THE OTHERS IN TERMS OF METRICS, WHICH COULD BE DUE TO THE NATURE OF THE DATA OR THE SIMPLE STRUCTURE OF THE UNDERLYING RELATIONSHIPS.



Resultados de predicción para la variable objetivo: XCMM

RMSE: 0.1550174

R-squared: 0.6843451

MAE: 0.04489245

Error Máximo: 1.633272

Resultados de predicción para la variable objetivo: ZCMM

RMSE: 0.02517703

R-squared: 0.7468559

MAE: 0.006653993

Error Máximo: 0.3834275

Resultados de predicción para la variable objetivo: BCMM

RMSE: 0.01738762

R-squared: 0.7255102

MAE: 0.004878562

Error Máximo: 0.2149986

Resultados de predicción para la variable objetivo: CCMM

RMSE: 0.01322425

R-squared: 0.7776438

MAE: 0.003706829

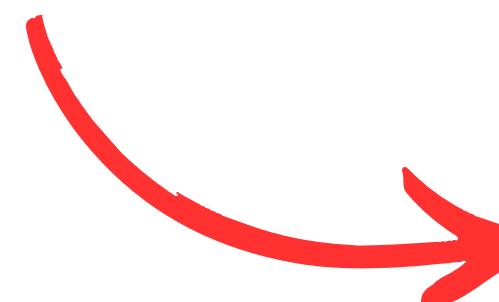
Error Máximo: 0.1905544



BROCHALIGN

--- Results of the KNN Model ---

It presented a very inferior performance, with low R-squares and high errors, this algorithm does not capture well the relationships in the data for this case.



```
Resultados de predicción para la variable objetivo: XCMM
RMSE : 0.2636578
R-squared: 0.08721923
MAE : 0.1861007
Error Máximo: 1.658

Resultados de predicción para la variable objetivo: ZCMM
RMSE : 0.04515444
R-squared: 0.185913
MAE : 0.03395983
Error Máximo: 0.3351538

Resultados de predicción para la variable objetivo: BCMM
RMSE : 0.03156794
R-squared: 0.09542559
MAE : 0.02192389
Error Máximo: 0.228

Resultados de predicción para la variable objetivo: CCMM
RMSE : 0.02633628
R-squared: 0.1182487
MAE : 0.01869477
Error Máximo: 0.4001538
```

--- Random Forest Model and XGBOOST Results ---

Although the initial versions showed improvements over KNN, with some simple tuning, these algorithms achieved exceptional performance in our final implementations.

Although Linear Regression shows acceptable results, more advanced models such as Random Forest and XGBoost provide better metrics with the adjustments made, making them our top choices for the final solution.

Resultados de predicción para la variable objetivo: XCMM
RMSE: 0.2591713
R-squared: 0.1628821
MAE: 0.18256
Error Máximo: 1.895682

Resultados de predicción para la variable objetivo: ZCMM
RMSE: 0.04607985
R-squared: 0.1542638
MAE: 0.03279023
Error Máximo: 0.3811703

Resultados de predicción para la variable objetivo: BCMM
RMSE: 0.03206198
R-squared: 0.1464778
MAE: 0.02239358
Error Máximo: 0.2153056

Resultados de predicción para la variable objetivo: CCMM
RMSE: 0.02132959
R-squared: 0.4215671
MAE: 0.01509529
Error Máximo: 0.1689157

Variable: XCMM
RMSE: 0.2622925
R-squared: 0.3012022
MAE: 0.1792448
Error Máximo: 1.873897

Variable: ZCMM
RMSE: 0.04701426
R-squared: 0.3103153
MAE: 0.03380073
Error Máximo: 0.4143787

Variable: BCMM
RMSE: 0.03079629
R-squared: 0.3396659
MAE: 0.02078967
Error Máximo: 0.2449264

Variable: CCMM
RMSE: 0.02598246
R-squared: 0.3319564
MAE: 0.01786588
Error Máximo: 0.386086



Time to get serious...

WE NEED A MORE INTRICATE MODEL IN
ORDER TO REACH OUR GOALS

Data Cleaning

We remove de NA's and convert all of our variables to factors or numeric values

Training models for XCMM (1/4)	Training models for ZCMM (2/4)	Training models for BCMM (3/4)	Training models for CCMM (4/4)
Before preprocessing: XCMM Number of rows: 3671 Number of NA values in target: 1 Range of target values: -1.872 1.411	Before preprocessing: ZCMM Number of rows: 3671 Number of NA values in target: 1 Range of target values: -0.42 0.307	Before preprocessing: BCMM Number of rows: 3671 Number of NA values in target: 1 Range of target values: -0.17 0.251	Before preprocessing: CCMM Number of rows: 3671 Number of NA values in target: 1 Range of target values: -0.105 0.43
After preprocessing: XCMM Number of rows: 3671 Number of NA values in target: 0 Range of target values: -1.872 1.411	After preprocessing: ZCMM Number of rows: 3671 Number of NA values in target: 0 Range of target values: -0.42 0.307	After preprocessing: BCMM Number of rows: 3671 Number of NA values in target: 0 Range of target values: -0.17 0.251	After preprocessing: CCMM Number of rows: 3671 Number of NA values in target: 0 Range of target values: -0.105 0.43

Brocha	NBrochasHSS	NDiscos	NUsos	USDutchman	BrochaSN	OrdenFabricacion	PartNumber	NUsos	NBrochaHSS	NDisco	FBrochado	DUMMY	Dutchman	Maquina	TpoIndexador	Uillaje	XC	ZC	BC	CC	XCMM	ZCMM	BCMM	CCMM	
011.0082.300	2	2	11	7	E0473		142	12	7	1	2	2023-10-01 22:29:34	FALSE	FALSE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0029	0.020	0.130	-0.010	-0.010	0.063	-0.024	0.020	0.032
011.0081.300	2	2	11	7	E0454		62	11	10	1	1	2023-10-02 01:20:03	FALSE	TRUE	HOFFMANN	IND-4 WALTER 900	UBR.0028	-0.450	-0.020	0.000	-0.030	0.020	-0.020	0.000	0.000
011.0070.300	2	2	11	7	E0462		146	16	11	1	1	2023-10-02 04:25:41	FALSE	TRUE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0017	0.190	0.070	0.000	0.090	-0.260	-0.020	-0.020	0.000
011.0082.300	2	2	11	7	E0473		86	9	7	2	1	2023-10-02 10:52:05	FALSE	FALSE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0031	0.100	0.120	0.020	0.020	0.017	-0.020	0.004	-0.020
011.0070.300	2	2	11	7	E0450		151	17	3	2	1	2023-10-02 16:30:07	FALSE	FALSE	HOFFMANN	IND-4 WALTER 900	UBR.0018	0.700	0.060	0.070	0.020	0.090	0.030	0.030	0.000
011.0081.300	2	2	11	7	E0454		63	11	10	1	2	2023-10-02 16:32:56	FALSE	TRUE	HOFFMANN	IND-4 WALTER 900	UBR.0028	-0.450	-0.020	0.000	-0.030	0.520	-0.070	0.060	-0.020
011.0070.300	2	2	11	7	E0462		147	16	11	1	2	2023-10-02 16:58:27	FALSE	TRUE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0017	0.190	0.070	0.000	0.090	-0.210	-0.051	-0.030	0.011
011.0070.300	2	2	11	7	E0450		212	17	3	2	2	2023-10-02 20:20:07	FALSE	FALSE	HOFFMANN	IND-4 WALTER 900	UBR.0018	0.700	0.060	0.070	0.020	0.200	0.052	0.041	0.005
011.0082.300	2	2	11	7	E0473		64	9	7	2	2	2023-10-02 23:34:52	FALSE	FALSE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0031	0.100	0.120	0.020	0.020	-0.047	0.024	0.007	-0.011
011.0081.300	2	2	11	7	E0454		61	11	10	2	1	2023-10-03 03:11:01	FALSE	TRUE	HOFFMANN	IND-4 WALTER 900	UBR.0028	0.000	-0.020	0.040	-0.030	0.308	-0.044	0.045	-0.001
011.0070.300	2	2	11	7	E0462		175	16	11	2	1	2023-10-03 04:13:50	FALSE	TRUE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0017	0.190	0.070	-0.010	0.090	0.200	0.020	0.030	-0.040
011.0070.300	2	2	11	7	E0462		205	16	11	2	2	2023-10-03 11:50:51	FALSE	TRUE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0017	0.190	0.070	-0.010	0.090	0.240	0.000	0.030	-0.020
011.0082.300	2	2	11	7	E0478		139	12	10	1	1	2023-10-04 00:01:18	FALSE	TRUE	LAPOINTE	IND-5 LAPOINTE 900	UBR.0029	0.090	0.130	0.020	-0.010	-0.400	0.006	-0.047	0.026
011.0077.300	2	2	11	7	E0388		130	20	5	1	1	2023-10-04 03:05:07	FALSE	FALSE	HOFFMANN	IND-4 WALTER 900	UBR.0014	0.500	0.070	0.090	-0.020	-0.560	-0.034	-0.056	-0.010

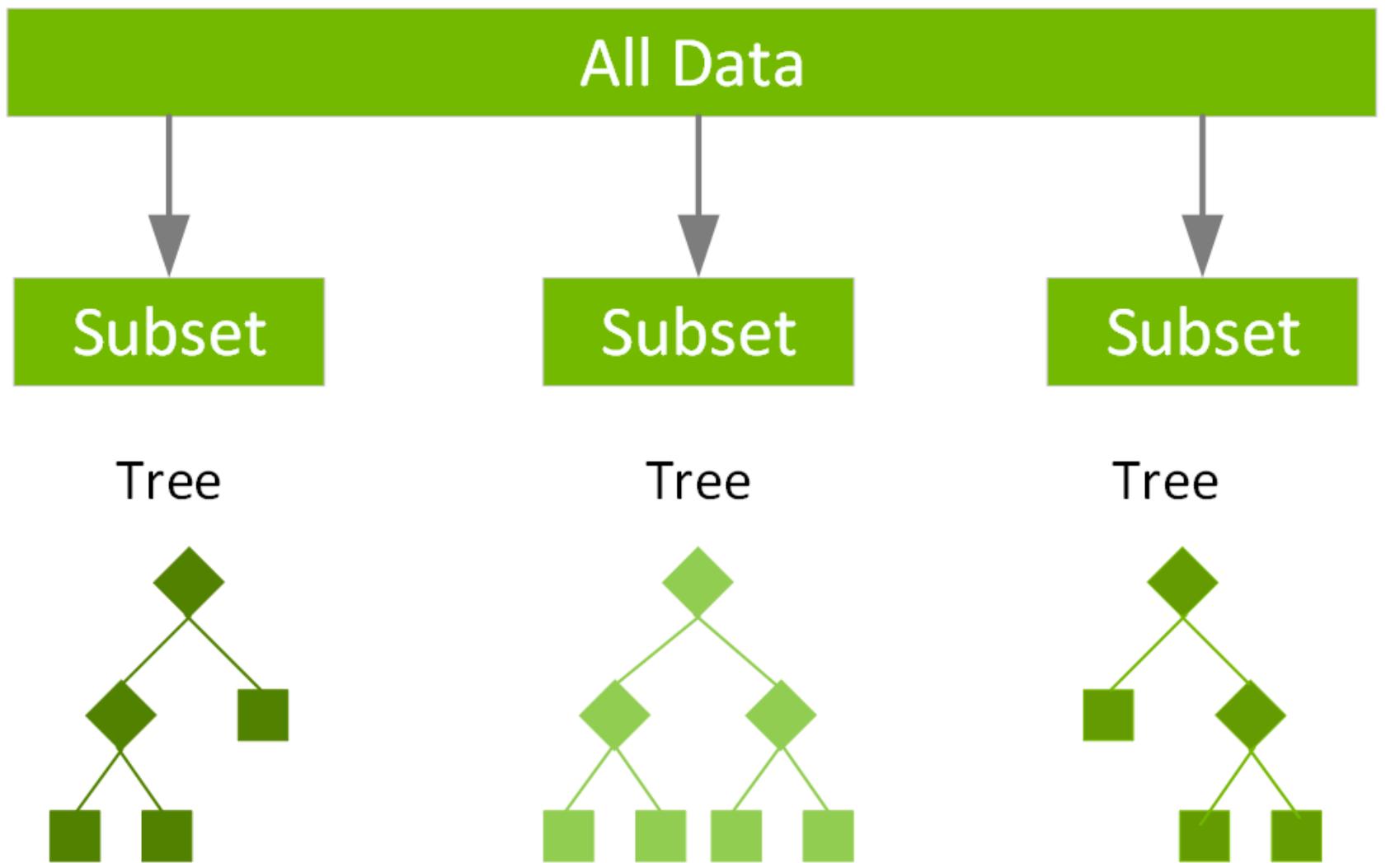
Analysis and Interpretation

FINALLY, WE EVALUATE OUR BEST MODELS AND FIND WHICH MODEL IS BEST SUITED FOR OUR DATA



BROCHALIGN

Best Model



OUR BEST MODEL UTILIZES
THE GBM LIBRARY



Results

Total time taken: 44.94 mins

ALL OF OUR MODELS ARE COMPLIANT WITH THE REQUIREMENTS FOR ITP AERO. OUR MODELS ARE NOT ONLY EFFICIENT BUT THEY ARE ALSO PRECISE



Impact

ALGORITHM THAT CORRECTLY PREDICTS THE ADJUSTMENTS NEEDED FOR THE BROACHES

XCM Performance:
RMSE: 0.08814414
MAE: 0.0501303
R-squared: 0.9001386
Max Error: 0.7921584

Best Model: XGBoost
Parameters:
eta: 0.05
max_depth: 8
min_child_weight: 3
subsample: 0.7
colsample_bytree: 0.9
gamma: 0
lambda: 0
alpha: 0
nrounds: 883

ZCM Performance:
RMSE: 0.01409206
MAE: 0.01016351
R-squared: 0.9185699
Max Error: 0.07996002

Best Model: XGBoost
Parameters:
eta: 0.05
max_depth: 8
min_child_weight: 5
subsample: 0.9
colsample_bytree: 0.9
gamma: 0
lambda: 0
alpha: 0
nrounds: 530

BCM Performance:
RMSE: 0.01161132
MAE: 0.007638338
R-squared: 0.8693849
Max Error: 0.06695456

Best Model: XGBoost
Parameters:
eta: 0.05
max_depth: 8
min_child_weight: 5
subsample: 0.9
colsample_bytree: 0.7
gamma: 0
lambda: 0
alpha: 0
nrounds: 761

CCM Performance:
RMSE: 0.007964692
MAE: 0.005251023
R-squared: 0.9085278
Max Error: 0.06480309

Best Model: XGBoost
Parameters:
eta: 0.05
max_depth: 8
min_child_weight: 5
subsample: 0.7
colsample_bytree: 0.7
gamma: 0
lambda: 0
alpha: 0
nrounds: 607



BROCHALIGN



Thank You

Note: All the code of the three R scripts can be executed entirely within the R environment. In addition, it is recommended to have a computer with sufficient resources to ensure a reasonable run time and training of the models.

Access to the code: The complete code of the three R scripts will be available in the GitHub links after December 20 deadline. You can access them by clicking on each corresponding name in the text of this presentation.

Bibliography

TIDYVERSE

CLUSTER

FPC

FACTOEXTRA

E1071

NBCLUST

GRIDEXTRA

CARET

DPLYR

XGBOOST

RANDOMFOREST

