ANNEXE

Figure 1

Variable/stat	Variable/Rang	Commentaires
YEAR	X	
PLAYER	X	
STROKES TOTAL	RANK FINAL	Nombre de coups réalisés au total
STROKES HALF	RANK SCORE HALF	Nombre de coups réalisés après la 2ème journée
NATIONALITY	X	
HEIGHT	X	En cm
WEIGHT	X	En kg
AGE	X	
AVG DIST	RANK DIST	Distance moyenne au drive (en mètres)
PRECISION (%)	RANK PREC	Pourcentage de fairways atteints au drive
AVG APP	RANK APP	*
AVG AROUND	RANK AROUND	**
AVG PUTTING	RANK PUTTING	***
AVG PAR 3	RANK PAR 3	Nombre de coups moyen sur les par 3
AVG PAR 4	RANK PAR 4	Nombre de coups moyen sur les par 4
AVG PAR 5	RANK PAR 5	Nombre de coups moyen sur les par 5

* "coups gagnés" d'un golfeur par rapport aux données historiques sur les "coups de parcours"

** "coups gagnés" d'un golfeur par rapport aux données historiques sur les coups d'approches (autour du green)

*** "coups gagnés" d'un golfeur par rapport aux données historiques sur les putts

Ces 3 statistiques sont des mesures de l'efficacité d'un golfeur par rapport aux données historiques pour différents types de coups.

Partie exploration de données et visualisation

Table 1

	AVG DIST	HEIGHT	WEIGHT	AGE
AVG DIST	1.000000	0.303595	0.306938	-0.275547
HEIGHT	0.303595	1.000000	0.591518	0.088785
WEIGHT	0.306938	0.591518	1.000000	0.086807
AGE	-0.275547	0.088785	0.086807	1.000000

Figure 2

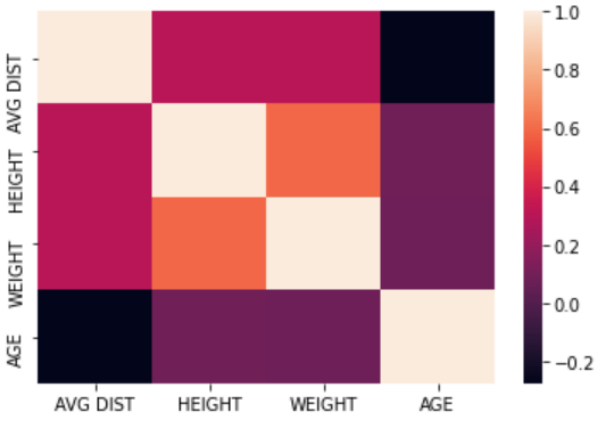


Figure 3

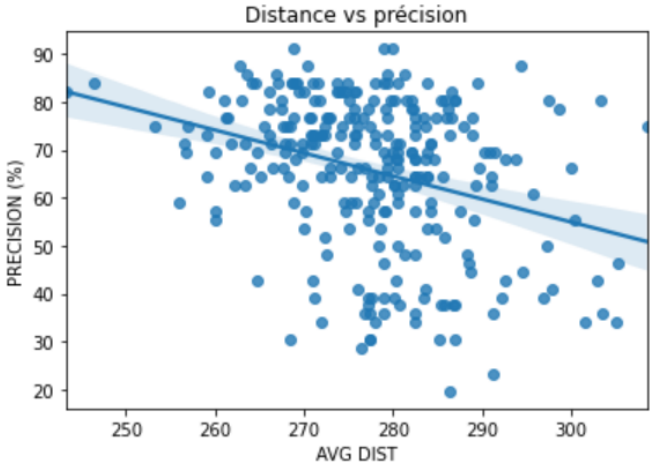


Figure 4 (2020)

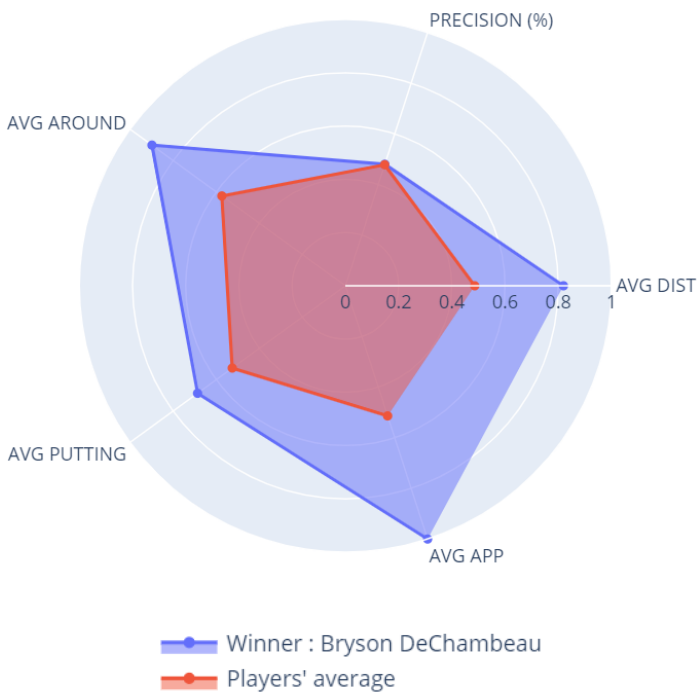


Figure 5 (2018)

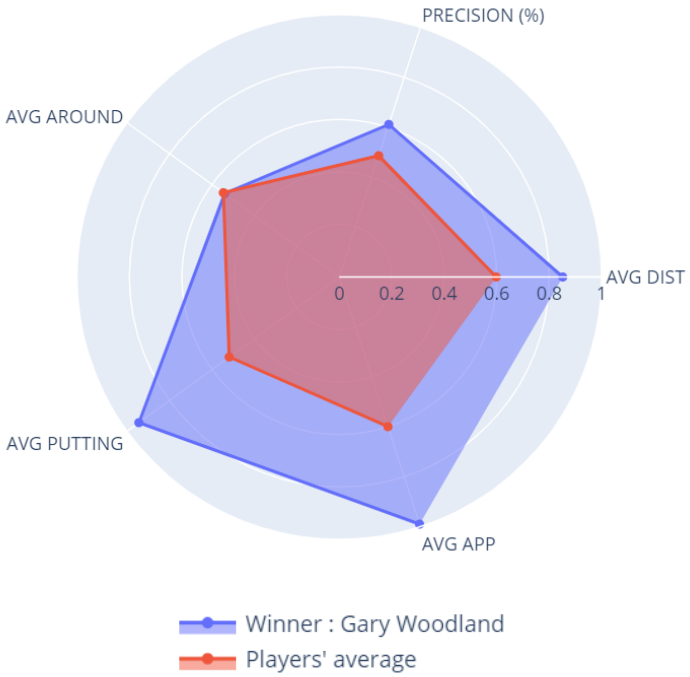


Figure 6 (2017)

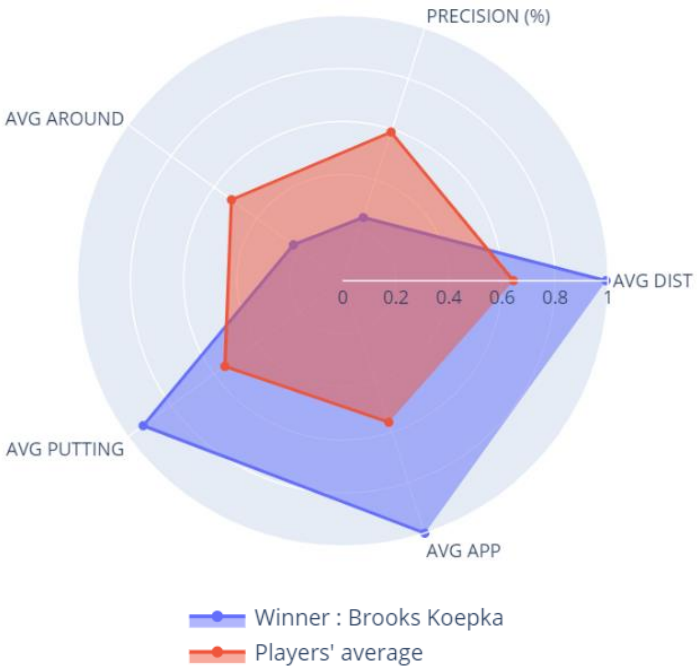


Figure 7 (2016)

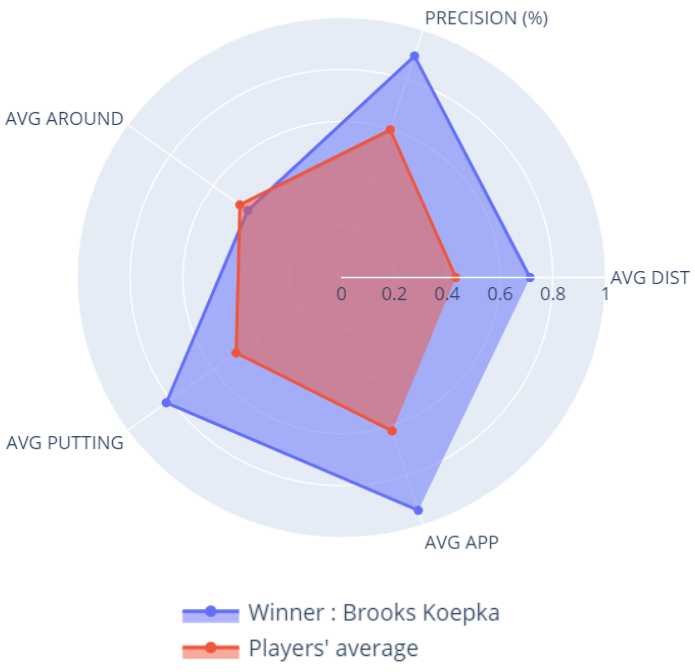


Figure 8

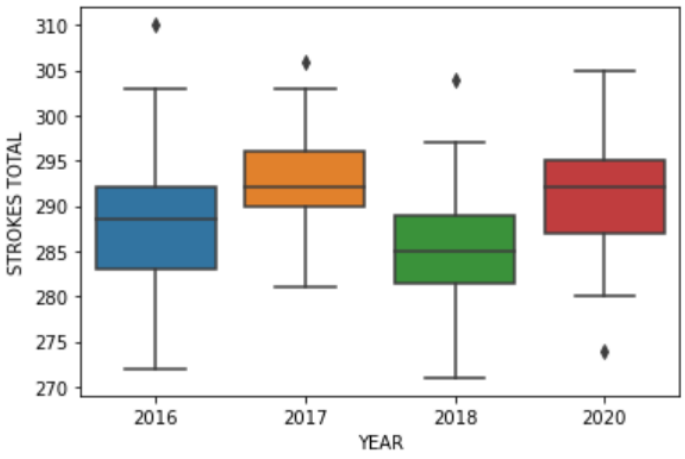


Figure 9

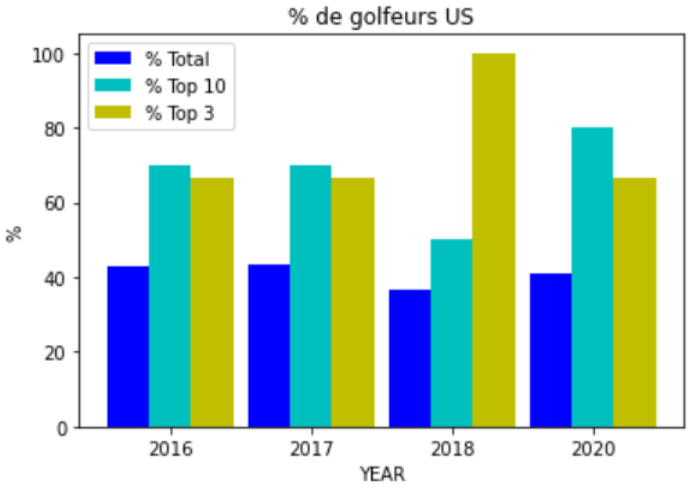
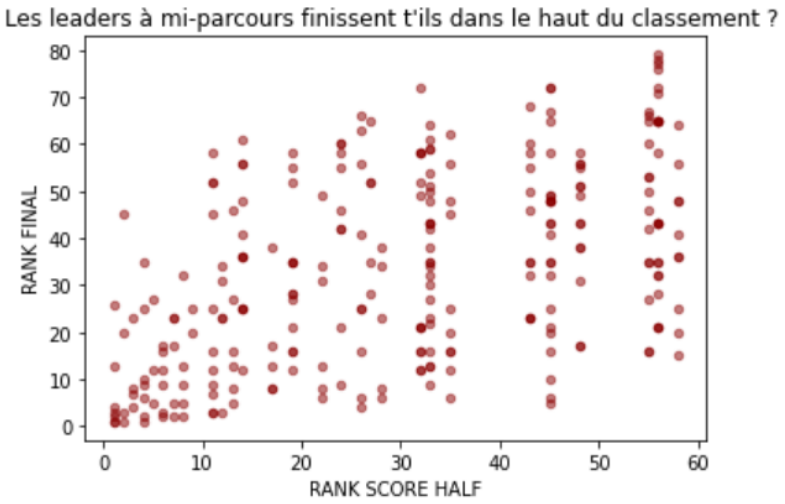


Figure 10



Partie Machine Learning

Tables 2 et 3

	AVG DIST	PRECISION (%)	AVG APP	AVG AROUND	AVG PUTTING	AVG PAR 3	AVG PAR 4	AVG PAR 5	PLAYER	PLAYING STYLE
0	305.2	46.43	1.833	0.453	0.097	3.31	4.04	4.75	Dustin Johnson	0
1	305.0	33.93	1.779	0.870	0.643	3.13	4.08	4.25	Matthew Wolff	0
2	303.5	35.71	-0.398	0.611	0.303	3.19	4.19	4.75	Taylor Pendrith	0
3	302.9	42.86	0.605	-0.270	0.255	3.13	4.17	5.00	Jon Rahm	0
4	301.6	33.93	-0.464	0.004	1.772	3.13	4.23	4.63	Joaquin Niemann	0

	AVG DIST	PRECISION (%)	AVG APP	AVG AROUND	AVG PUTTING	AVG PAR 3	AVG PAR 4	AVG PAR 5	PLAYER	PLAYING STYLE
45	277.5	64.29	0.423	-0.142	1.612	3.25	4.21	4.50	Brendon Todd	1
61	290.9	69.64	0.173	0.888	-0.032	3.00	3.93	4.83	Rory McIlroy	1
62	289.7	66.07	1.390	0.347	-1.521	3.06	4.09	4.67	Dustin Johnson	1
63	289.2	62.50	0.850	-0.274	-0.631	3.06	4.14	4.83	Adri Arnaus	1
64	284.5	67.86	0.398	-0.073	1.424	2.94	3.93	4.75	Xander Schauffele	1

Tables 4 et 5

	AVG DIST	PRECISION (%)	AVG APP	AVG AROUND	AVG PUTTING	AVG PAR 3	AVG PAR 4	AVG PAR 5	PLAYING STYLE
count	71.000000	71.000000	71.000000	71.000000	71.000000	71.000000	71.000000	71.000000	71.0
mean	283.808451	43.033239	0.206648	0.250254	0.439789	3.205352	4.199155	4.801690	0.0
std	9.125189	9.420465	0.960336	0.674538	0.953358	0.182630	0.108031	0.240303	0.0
min	264.700000	19.640000	-2.258000	-1.439000	-1.684000	2.810000	3.980000	4.250000	0.0
25%	277.400000	35.710000	-0.431000	-0.209000	-0.229500	3.130000	4.150000	4.630000	0.0
50%	282.500000	41.070000	0.143000	0.351000	0.504000	3.190000	4.190000	4.750000	0.0
75%	288.400000	50.895000	0.973000	0.661000	1.226500	3.310000	4.270000	4.940000	0.0
max	305.200000	60.710000	1.888000	1.545000	2.403000	3.690000	4.440000	5.380000	0.0

	AVG DIST	PRECISION (%)	AVG APP	AVG AROUND	AVG PUTTING	AVG PAR 3	AVG PAR 4	AVG PAR 5	PLAYING STYLE
count	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.0
mean	275.606863	73.468480	0.480387	0.235387	0.342373	3.092794	4.110294	4.813824	1.0
std	9.888051	7.986652	0.896340	0.586594	0.907356	0.163943	0.119967	0.225570	0.0
min	243.300000	55.360000	-2.478000	-1.200000	-1.960000	2.690000	3.830000	4.250000	1.0
25%	269.050000	67.860000	-0.116250	-0.143500	-0.248250	3.000000	4.020000	4.670000	1.0
50%	275.700000	73.210000	0.457000	0.218500	0.352000	3.060000	4.100000	4.810000	1.0
75%	282.100000	80.360000	1.155750	0.610750	0.965000	3.190000	4.210000	4.940000	1.0
max	308.500000	91.070000	2.214000	1.696000	2.523000	3.560000	4.400000	5.630000	1.0

Table 6

	YEAR	PLAYER	AVG DIST	PRECISION (%)
6	2020	Bryson DeChambeau	297.9	41.07
69	2018	Bryson DeChambeau	283.0	64.29
151	2017	Bryson DeChambeau	282.0	69.64

Table 7

OLS Regression Results

Dep. Variable:	AVG DIST	R-squared:	0.214
Model:	OLS	Adj. R-squared:	0.202
Method:	Least Squares	F-statistic:	17.66
Date:	Fri, 11 Dec 2020	Prob (F-statistic):	3.49e-10
Time:	22:06:20	Log-Likelihood:	-720.83
No. Observations:	199	AIC:	1450.
Df Residuals:	195	BIC:	1463.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	212.6300	20.791	10.227	0.000	171.626	253.634
HEIGHT	0.3457	0.133	2.609	0.010	0.084	0.607
WEIGHT	0.2406	0.089	2.694	0.008	0.064	0.417
AGE	-0.5068	0.104	-4.893	0.000	-0.711	-0.303

Omnibus:	3.775	Durbin-Watson:	0.722
Prob(Omnibus):	0.151	Jarque-Bera (JB):	4.138
Skew:	0.131	Prob(JB):	0.126
Kurtosis:	3.656	Cond. No.	6.50e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 6.5e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Table 8

OLS Regression Results

Dep. Variable:	AVG DIST	R-squared:	0.186
Model:	OLS	Adj. R-squared:	0.178
Method:	Least Squares	F-statistic:	22.43
Date:	Fri, 11 Dec 2020	Prob (F-statistic):	1.70e-09
Time:	22:11:56	Log-Likelihood:	-724.25
No. Observations:	199	AIC:	1454.
Df Residuals:	196	BIC:	1464.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	278.2799	0.658	422.923	0.000	276.982	279.578
x1	3.4043	0.660	5.154	0.000	2.102	4.707
x2	-3.1093	0.660	-4.708	0.000	-4.412	-1.807

Omnibus:	4.237	Durbin-Watson:	0.644
Prob(Omnibus):	0.120	Jarque-Bera (JB):	4.305
Skew:	0.208	Prob(JB):	0.116
Kurtosis:	3.589	Cond. No.	1.09

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 9

OLS Regression Results

Dep. Variable:	RANK FINAL	R-squared (uncentered):	0.952
Model:	OLS	Adj. R-squared (uncentered):	0.948
Method:	Least Squares	F-statistic:	227.1
Date:	Fri, 11 Dec 2020	Prob (F-statistic):	4.66e-84
Time:	22:23:13	Log-Likelihood:	-516.39
No. Observations:	149	AIC:	1057.
Df Residuals:	137	BIC:	1093.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
HEIGHT	0.2692	0.141	1.909	0.058	-0.010	0.548
WEIGHT	-0.1589	0.103	-1.541	0.126	-0.363	0.045
AGE	0.0480	0.113	0.425	0.671	-0.175	0.271
RANK SCORE HALF	0.0756	0.043	1.739	0.084	-0.010	0.162
AVG DIST	-0.3567	0.071	-4.991	0.000	-0.498	-0.215
PRECISION (%)	0.0165	0.048	0.345	0.730	-0.078	0.111
AVG APP	-10.7375	0.897	-11.976	0.000	-12.510	-8.964
AVG AROUND	-11.9316	1.137	-10.492	0.000	-14.180	-9.683
AVG PUTTING	-10.7732	0.800	-13.463	0.000	-12.355	-9.191
AVG PAR 3	-3.2733	4.315	-0.759	0.449	-11.806	5.259
AVG PAR 4	17.1956	4.932	3.487	0.001	7.444	26.948
AVG PAR 5	8.5523	2.989	2.861	0.005	2.641	14.464

Omnibus:	5.228	Durbin-Watson:	1.918
Prob(Omnibus):	0.073	Jarque-Bera (JB):	7.520
Skew:	-0.061	Prob(JB):	0.0233
Kurtosis:	4.094	Cond. No.	2.80e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.8e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

Table 10

OLS Regression Results

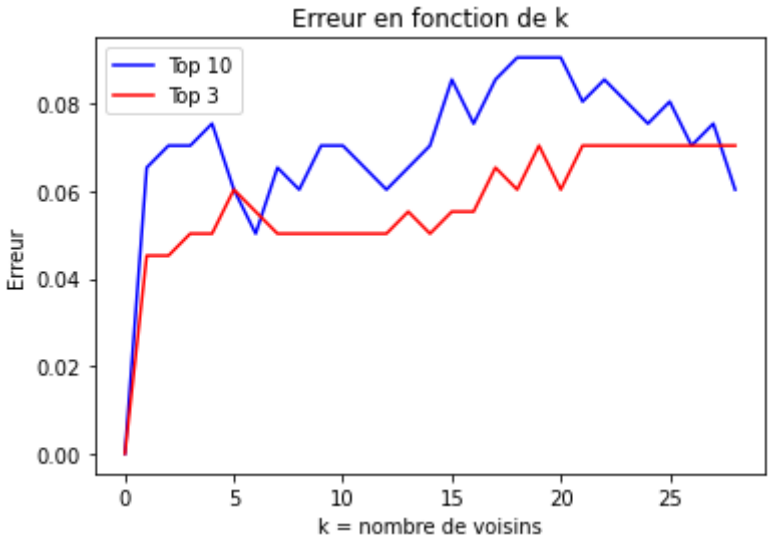
Dep. Variable:	RANK FINAL	R-squared (uncentered):	0.208
Model:	OLS	Adj. R-squared (uncentered):	0.174
Method:	Least Squares	F-statistic:	6.249
Date:	Fri, 11 Dec 2020	Prob (F-statistic):	7.52e-06
Time:	22:28:36	Log-Likelihood:	-725.48
No. Observations:	149	AIC:	1463.
Df Residuals:	143	BIC:	1481.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
x1	9.1933	1.702	5.403	0.000	5.830	12.557
x2	-3.4838	1.855	-1.878	0.062	-7.150	0.183
x3	-4.5068	2.113	-2.132	0.035	-8.684	-0.329
x4	-0.4028	2.390	-0.169	0.866	-5.126	4.321
x5	-0.2807	2.479	-0.113	0.910	-5.182	4.620
x6	1.1798	2.762	0.427	0.670	-4.281	6.640

Omnibus:	2.613	Durbin-Watson:	0.239
Prob(Omnibus):	0.271	Jarque-Bera (JB):	2.227
Skew:	0.290	Prob(JB):	0.328
Kurtosis:	3.148	Cond. No.	1.62

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 11



Figures 12 et 13

PREDICTION TOP 10 : EVALUATION DES 2 MODELES

KNN
- R2 : 0.98
- Precision : 0.94
- Recall : 0.99
- Accuracy : 0.98
- F1 : 0.96

Arbre de décision
- R2 : 0.88
- Precision : 0.73
- Recall : 0.81
- Accuracy : 0.88
- F1 : 0.76

PREDICTION TOP 3 : EVALUATION DES 2 MODELES

KNN
- R2 : 0.9
- Precision : 0.95
- Recall : 0.57
- Accuracy : 0.9
- F1 : 0.6

Arbre de décision
- R2 : 0.95
- Precision : 0.97
- Recall : 0.79
- Accuracy : 0.95
- F1 : 0.85