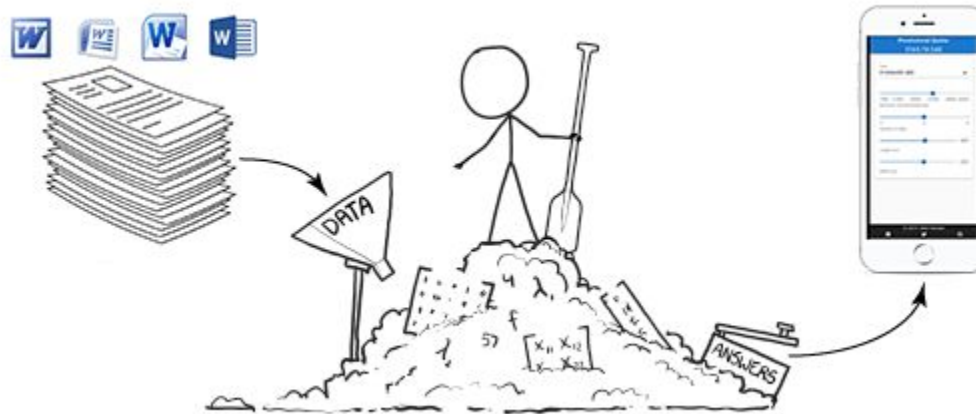


Sales Quotes: from 15 minutes to 15 milliseconds using Machine Learning



Bart Fransen

Data Science - Spring 2019

CONTEXT



BCW B.C.M. N.V.
 KMO's-avenue Belgislaan 14
 b. 2500-Brabant
 Tel: +32 (0)3 1 483 0770
 Fax: +32 (0)3 1 483 0778
 1979 Molenbeek BE-1046-146-197
 B.A.N. 84 85 701 002-4786-197

OFFERTE

aanhangwagens **aggrats** **concrete** **concrete** **concrete** **concrete**

[illegible]

GEACHT.

NAAR AANLEIDING VAN DE PRELIMINAIRE KENNEN WIL DE VOLGENDE OFFERTE AANBODEN

AXLES, BRAKE		AXLES, LOWERED		TYRE, SPARE		DIMENSIONS, LENGTH, CM		DIMENSIONS, WIDTH, CM		TYPE, CATEGORY, CAR		TYPE, CATEGORY, CLOSURE		- MAKE, YEAR	
0	1	1	0	0			500.0		200.0		1				
1	0	1	1	1			250.0		140.0		0				
2	0	0	0	1			300.0		140.0		0				
3	1	0	0	1			450.0		200.0		1				
4	1	1	0	1			350.0		160.0		0				
5	1	0	0	1			300.0		180.0		0				
6	1	1	0	0			400.0		180.0		0				
7	1	0	0	0			600.0		250.0		0				
8	1	1	0	0			250.0		160.0		0				
9	1	0	0	1			300.0		180.0		0				
10	1	1	0	0			510.0		230.0		0				
11	1	0	0	0			310.0		200.0		0				
12	1	1	1	1			350.0		160.0		0				
13	0	0	1	1			250.0		140.0		0				
14	1	0	0	1			600.0		210.0		1				
15	1	1	0	0			1250.0		420.0		0				
17	0	0	0	0			260.0		140.0		0				
18	1	1	0	0			400.0		180.0		0				
19	1	1	0	0			310.0		200.0		0				
20	1	0	0	1			300.0		180.0		0				

Train - Test Split

Test

80 / 20

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

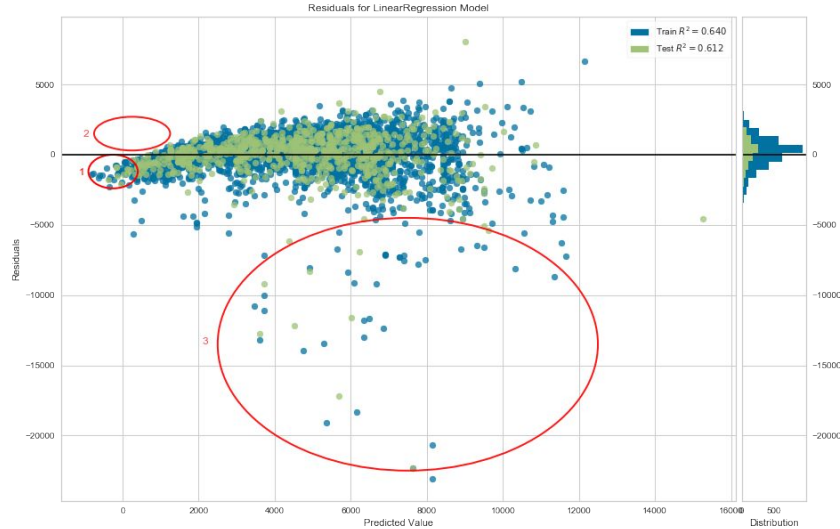


A line chart illustrating the reduction in the number of words in the dataset across four stages. The y-axis represents the number of words, ranging from 0 to 8,000 in increments of 2,000. The x-axis lists the stages: Word, Text, Pandas, and Split. The data points are connected by a dark blue line, and each point is labeled with its corresponding value.

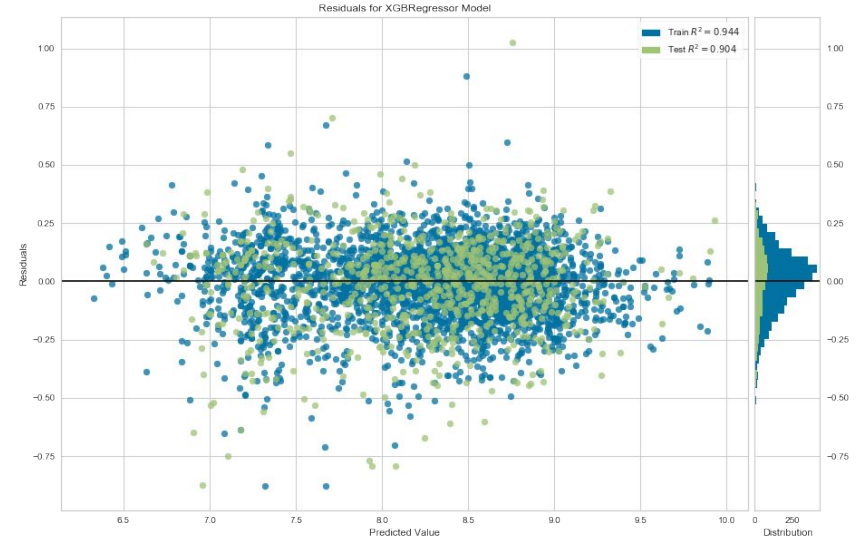
Stage	Number of Words
Word	5,340
Text	4,721
Pandas	4,370
Split	3,496

FEATURE ENGINEERING

RESIDUALS PLOT



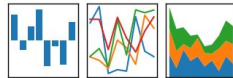
BEFORE
12 features



AFTER
20 features

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

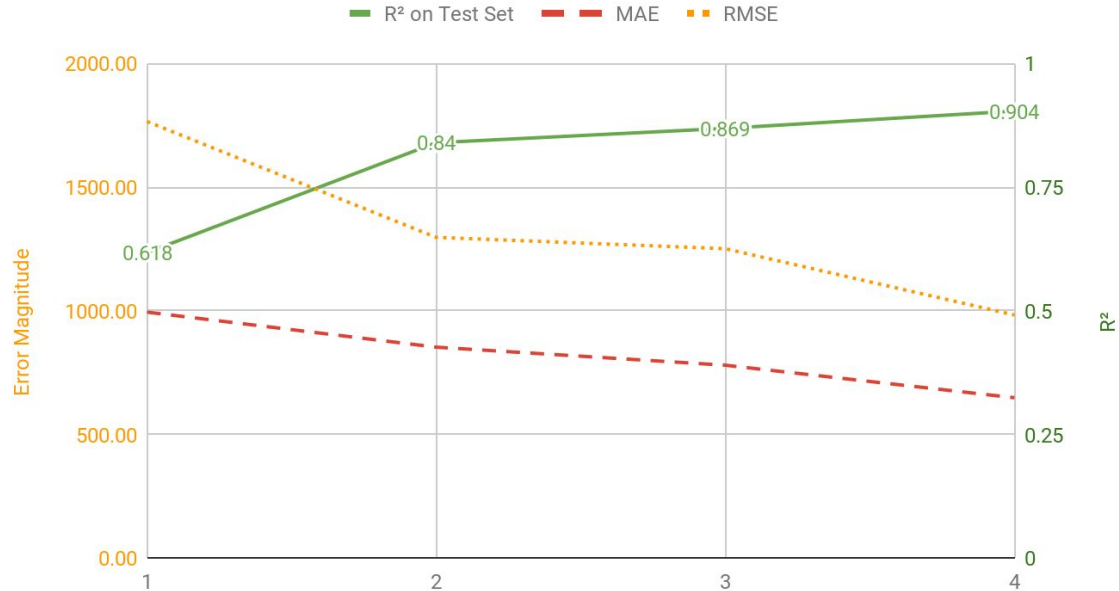


NumPy



MODELLING

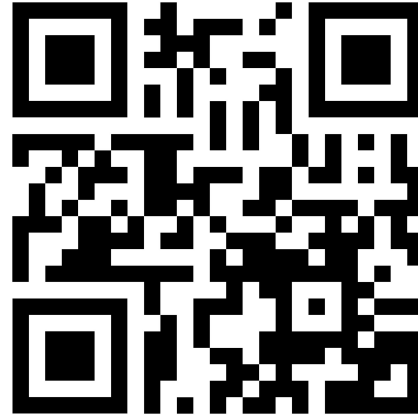
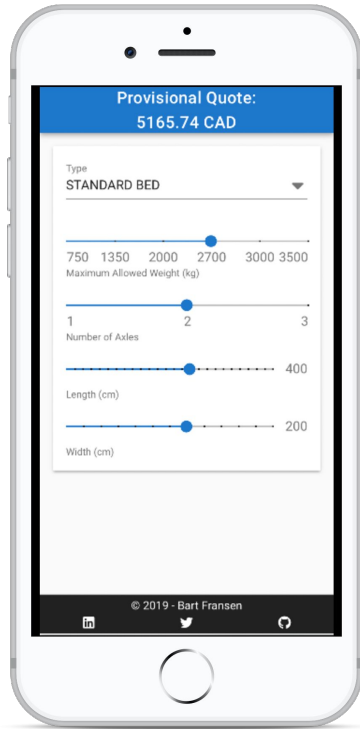
Test score, MAE, and RMSE



Best Results	
R ² on Test Set	0.904
MAE	649.04
RMSE	983.59

XGBoost

DEPLOYMENT



[HTTPS://CAPSTONE.BARTFRANSEN.COM](https://capstone.bartfransen.com)

hello@bartfransen.com



Amazon EC2



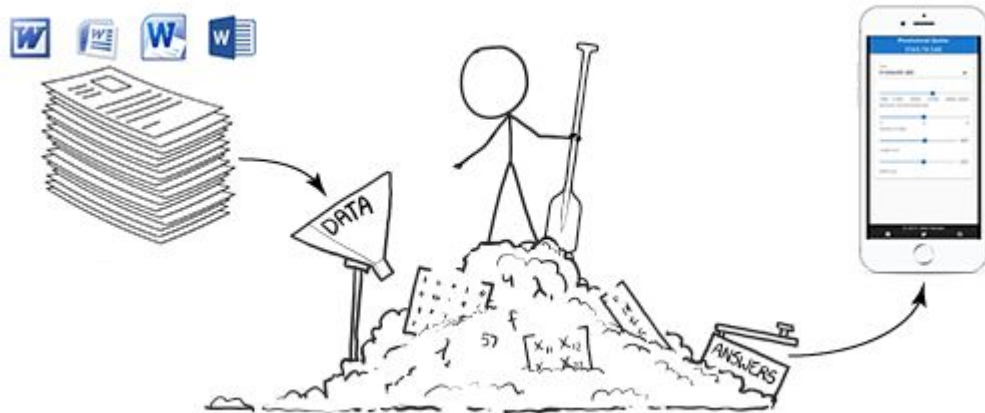
Flask



I'LL BE BACK

Sales Quotes:
from 15 minutes
to 15 milliseconds
using Machine Learning

Give it a try!



[HTTPS://CAPSTONE.BARTFRANSEN.COM](https://capstone.bartfransen.com)

hello@bartfransen.com

Bart Fransen

Data Science - Spring 2019