Orange Malaysia Sectoral

Data collection : Malaysia_Energy_Statistics_Handbook_20201 page 47

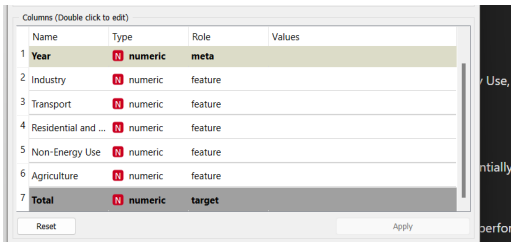
https://www.st.gov.my/en/contents/files/download/116/Malaysia_Energy_Statistics_Handbook_20201.pdf

Unit: ktoe

Year	Final Energy Consumption by Sector					
	Industry	Transport	Residential and Commercial	Non-Energy Use	Agriculture	Total
1990	5,300	5,386	1,622	838	-	13,146
1991	5,809	5,806	1,721	1,071	130	14,537
1992	6,455	6,226	1,867	1,222	391	16,161
1993	7,012	6,558	2,055	2,027	62	17,714
1994	7,283	7,262	2,300	1,817	422	19,084
1995	8,060	7,827	2,556	2,994	446	21,883
1996	9,443	8,951	3,162	1,744	486	23,786
1997	10,106	10,201	3,072	2,298	490	26,167
1998	10,121	9,793	3,313	2,023	307	25,557
1999	10,277	11,393	3,653	1,799	106	27,228
2000	11,406	12,071	3,868	2,250	104	29,699
2001	11,852	13,137	4,048	2,378	98	31,513
2002	12,854	13,442	4,387	2,511	96	33,290
2003	13,472	14,271	4,399	2,345	98	34,585
2004	14,914	15,385	4,754	2,183	87	37,323
2005	15,492	15,384	5,134	2,173	101	38,284
2006	15,248	14,819	5,424	2,819	258	38,567
2007	16,454	15,717	6,197	2,957	281	41,606
2008	16,205	16,395	6,205	2,876	287	41,968
2009	14,312	16,119	6,336	3,868	211	40,846
2010	12,928	16,828	6,951	3,696	1,074	41,477
2011	12,100	17,070	6,993	6,377	916	43,456
2012	13,919	19,757	7,065	7,497	1,053	49,291
2013	13,496	22,357	7,403	7,277	1,051	51,584
2014	13,162	24,327	7,458	6,217	1,045	52,209
2015	13,989	23,435	7,559	5,928	895	51,806
2016	16,019	24,004	8,049	8,729	415	57,218
2017	17,463	24,039	7,796	12,517	674	62,489
2018	19,046	23,555	7,774	13,262	1,021	64,658

ENERGY CONSUMPTION

47

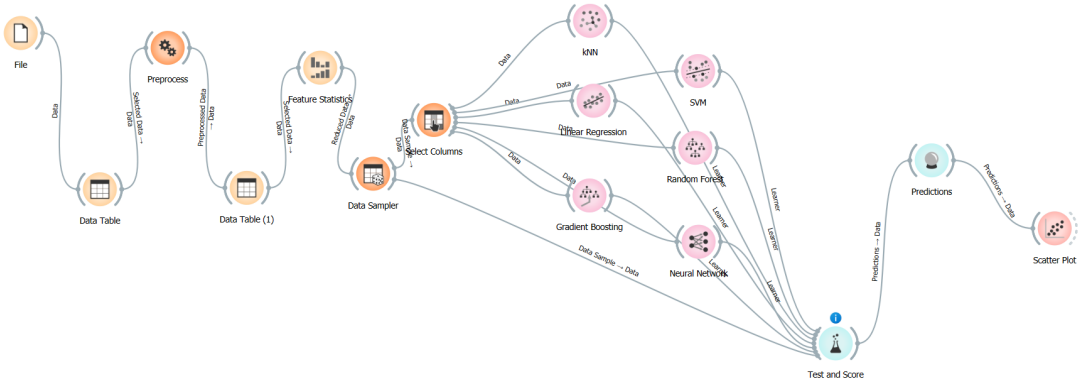


meta: not using it for predictions but as an identifier (eg: gender)

target: the **total** wor

feature: **predictor variables**

- We only need for **imputation** as only 1 data is missing
- No need for **normalization** as we are using more towards decision trees and random forest



testing out which model are the best to choose for top 3

Model	MSE	RMSE	MAE	R2
kNN	15735351.973	3966.781	3139.924	0.928
Linear Regression	13974.601	118.214	69.081	1.000
SVM	249159684.256	15784.793	14016.236	-0.141
Random Forest	12562120.747	3544.308	2492.422	0.942
Gradient Boosting	5751151.259	2398.156	1656.296	0.974
Neuro Network Tuning	1459031542....	38197.271	35233.937	-5.684

`kNN`: decent R2, only explains 92.8% of the variance. RMSE n MAE relatively high

`Linear`: capturing 100% of the variance (almost perfect fit). very low MSE, RMSE and MAE support this too. High R2 could indicate overfitting

`SVM` : poor performance (neg R2)

`Random Forest` : High R2, other error metrics also lower than those `kNN` and `SVM`

`Grad Boost` : High R2, other error metric also relatively low

`Neural Network` : Extremely poor

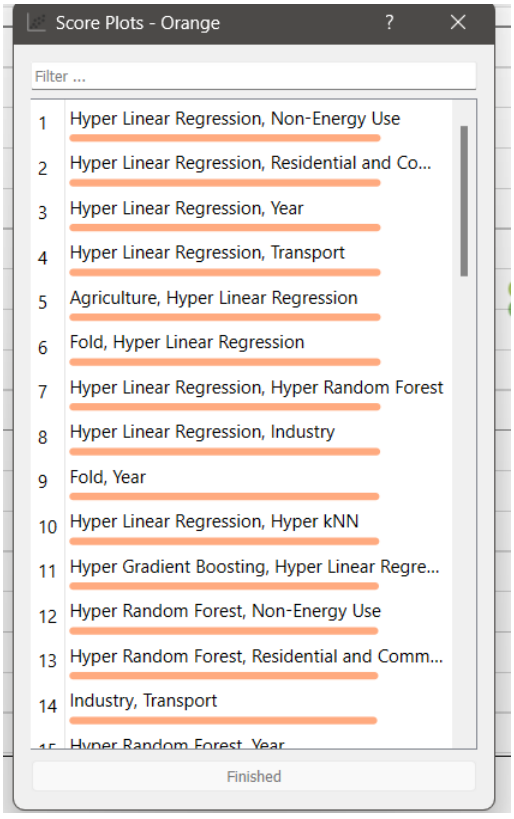
Results shows kNN, Linear, Random For and Grad boosts are the best

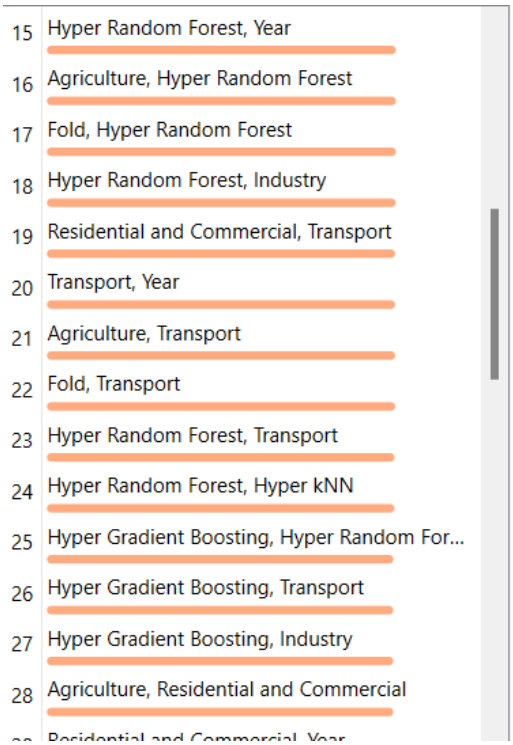
now preparing for hyper parameter

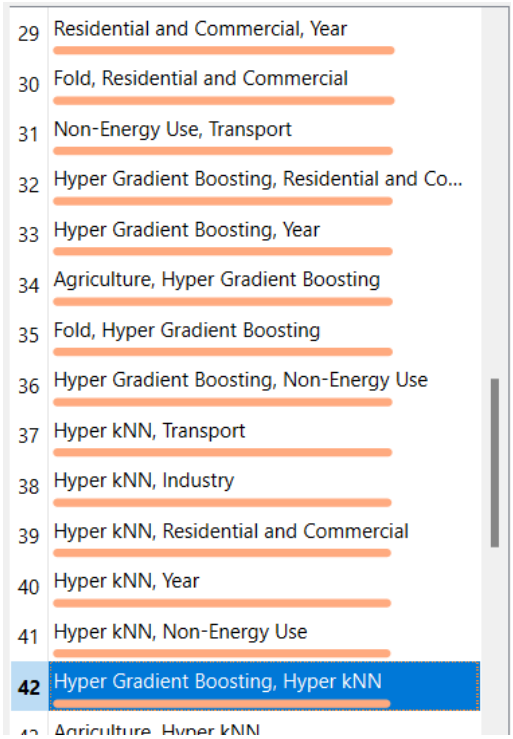
: Aims to have the closest to one of R2 and minimising other error metrics

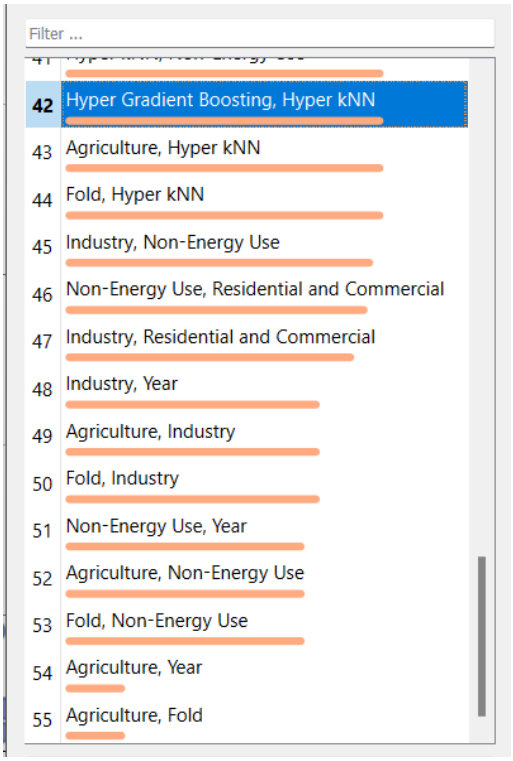
All of the changed Model will be rename Hyper + model

Model	MSE	RMSE	MAE	R2
Hyper kNN	7124158.619	2669.112	2085.286	0.967
Hyper Linear Regression	13974.601	118.214	69.081	1.000
Hyper Random Forest	6519083.850	2553.250	1830.648	0.970
Hyper Gradient Boosting	4380981.421	2093.079	1681.167	0.980




15 Hyper Random Forest, Year
16 Agriculture, Hyper Random Forest
17 Fold, Hyper Random Forest
18 Hyper Random Forest, Industry
19 Residential and Commercial, Transport
20 Transport, Year
21 Agriculture, Transport
22 Fold, Transport
23 Hyper Random Forest, Transport
24 Hyper Random Forest, Hyper kNN
25 Hyper Gradient Boosting, Hyper Random For...
26 Hyper Gradient Boosting, Transport
27 Hyper Gradient Boosting, Industry
28 Agriculture, Residential and Commercial


29 Residential and Commercial, Year
30 Fold, Residential and Commercial
31 Non-Energy Use, Transport
32 Hyper Gradient Boosting, Residential and Co...
33 Hyper Gradient Boosting, Year
34 Agriculture, Hyper Gradient Boosting
35 Fold, Hyper Gradient Boosting
36 Hyper Gradient Boosting, Non-Energy Use
37 Hyper kNN, Transport
38 Hyper kNN, Industry
39 Hyper kNN, Residential and Commercial
40 Hyper kNN, Year
41 Hyper kNN, Non-Energy Use
42 Hyper Gradient Boosting, Hyper kNN



Notes: The longer the orange bars, suggesting this combination performed well in the evaluation

All of the parameters with Hyper Linear Regression is might not good to refer as it may lead to overfitting model