

Multiple Linear Regression (Module -7)

Q1) Prepare a prediction model for profit of 50_startups data.

Note: Do transformations for getting better predictions of profit and make a table containing R^2 value for each prepared model.

	R.D.Spend	Administration	Marketing.Spend	State	Profit
1	165349.20	136897.80	471784.10	New York	192261.83
2	162597.70	151377.59	443898.53	California	191792.06
3	153441.51	101145.55	407934.54	Florida	191050.39
4	144372.41	118671.85	383199.62	New York	182901.99
5	142107.34	91391.77	366168.42	Florida	166187.94
6	131876.90	99814.71	362861.36	New York	156991.12
7	134615.46	147198.87	127716.82	California	156122.51
8	130298.13	145530.06	323876.68	Florida	155752.60
9	120542.52	148718.95	311613.29	New York	152211.77
10	123334.88	108679.17	304981.62	California	149759.96

Q2 Predict the sales of the computer

	X	price	speed	hd	ram	screen	cd	multi	premium	ads	trend
1	1	1499	25	80	4	14	no	no	yes	94	1
2	2	1795	33	85	2	14	no	no	yes	94	1
3	3	1595	25	170	4	15	no	no	yes	94	1
4	4	1849	25	170	8	14	no	no	no	94	1
5	5	3295	33	340	16	14	no	no	yes	94	1
6	6	3695	66	340	16	14	no	no	yes	94	1
7	7	1720	25	170	4	14	yes	no	yes	94	1
8	8	1995	50	85	2	14	no	no	yes	94	1
9	9	2225	50	210	8	14	no	no	yes	94	1
10	10	2575	50	210	4	15	no	no	yes	94	1
11	11	2105	33	170	8	15	no	no	yes	94	1

Q3) Consider only the below columns and prepare a prediction model for predicting Price.

Corolla < Corolla[c("Price", "Age_08_04", "KM", "HP", "cc", "Doors", "Gears", "Quarterly_Tax", "Weight")]

	Id	Model	Price	Age_08_04	Mfg_Month	Mfg_Year	KM	Fuel_Type	HP	Met_Color	Color	Automatic	cc	Doors	Cylinder
1	1	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	13500	23	10	2002	46986	Diesel	90	1	Blue	0	2000	3	4
2	2	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	13750	23	10	2002	72937	Diesel	90	1	Silver	0	2000	3	4
3	3	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	13950	24	9	2002	41711	Diesel	90	1	Blue	0	2000	3	4
4	4	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	14950	26	7	2002	48000	Diesel	90	0	Black	0	2000	3	4
5	5	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	13750	30	3	2002	38500	Diesel	90	0	Black	0	2000	3	4
6	6	TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors	12950	32	1	2002	61000	Diesel	90	0	White	0	2000	3	4
7	7	TOYOTA Corolla 2.0 D4D 90 3DR TERRA 2/3-Doors	16900	27	6	2002	94612	Diesel	90	1	Grey	0	2000	3	4
8	8	TOYOTA Corolla 2.0 D4D 90 3DR TERRA 2/3-Doors	18600	30	3	2002	75889	Diesel	90	1	Grey	0	2000	3	4
9	9	TOYOTA Corolla 1800 T SPORT VVT-i 2/3-Doors	21500	27	6	2002	19700	Petrol	192	0	Red	0	1800	3	4
10	10	TOYOTA Corolla 1.9 D HATCHB TERRA 2/3-Doors	12950	23	10	2002	71138	Diesel	69	0	Blue	0	1900	3	4
11	11	TOYOTA Corolla 1.8 VVTL-i T-Sport 2/3-Doors	20950	25	8	2002	31461	Petrol	192	0	Silver	0	1800	3	4
12	12	TOYOTA Corolla 1.8 16V VVT-i 3DR T SPORT BNS 2/3-Doors	19950	22	11	2002	43610	Petrol	192	0	Red	0	1800	3	4
13	13	TOYOTA Corolla 1.8 16V VVT-i 3DR T SPORT 2/3-Doors	19600	25	8	2002	32189	Petrol	192	0	Red	0	1800	3	4
14	14	TOYOTA Corolla 1.8 16V VVT-i 3DR T SPORT 2/3-Doors	21500	31	2	2002	23000	Petrol	192	1	Black	0	1800	3	4
15	15	TOYOTA Corolla 1.8 16V VVT-i 3DR T SPORT 2/3-Doors	22500	32	1	2002	34131	Petrol	192	1	Grey	0	1800	3	4
16	16	TOYOTA Corolla 1.8 16V VVT-i 3DR T SPORT 2/3-Doors	22000	28	5	2002	18739	Petrol	192	0	Grey	0	1800	3	4
17	17	TOYOTA Corolla 1.8 16V VVT-i 3DR T SPORT 2/3-Doors	22750	30	3	2002	34000	Petrol	192	1	Grey	0	1800	3	4

Hints:

1. Business Problem
 - 1.1. Objective
 - 1.2. Constraints (if any)
2. Data Pre-processing
 - 2.1 Data cleaning, Feature Engineering, EDA etc.
3. Model Building
 - 3.1 Partition the dataset
 - 3.2 Model(s) - Reasons to choose any algorithm
 - 3.3 Model(s) Improvement steps
 - 3.4 Model Evaluation
 - 3.5 Python and R codes
4. Deployment
 - 4.1 Deploy solutions using R shiny and Python Flask.
5. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

Note:

1. For each assignment the solution should be submitted in the format
2. Research and Perform all possible steps for improving the model(s) accuracy.
Ex: Transformations, Feature Engineering, Hyper Parameter tuning, Outlier treatment, etc.
3. All the codes (executable programs) are running without error
4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here.