Topic: K-Nearest Neighbor

1.) Prepare a model for glass classification using KNN. The dataset looks like this:

	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
1	1.52101	13.64	4.49	1.10	71.78	0.06	8.75	0.00	0.00	1
2	1.51761	13.89	3.60	1.36	72.73	0.48	7.83	0.00	0.00	1
3	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0.00	0.00	1
4	1.51766	13.21	3.69	1.29	72.61	0.57	8.22	0.00	0.00	1
5	1.51742	13.27	3.62	1.24	73.08	0.55	8.07	0.00	0.00	1
6	1.51596	12.79	3.61	1.62	72.97	0.64	8.07	0.00	0.26	1
7	1.51743	13.30	3.60	1.14	73.09	0.58	8.17	0.00	0.00	1
8	1.51756	13.15	3.61	1.05	73.24	0.57	8.24	0.00	0.00	1
9	1.51918	14.04	3.58	1.37	72.08	0.56	8.30	0.00	0.00	1
10	1.51755	13.00	3.60	1.36	72.99	0.57	8.40	0.00	0.11	1
11	1.51571	12.72	3.46	1.56	73.20	0.67	8.09	0.00	0.24	1
12	1.51763	12.80	3.66	1.27	73.01	0.60	8.56	0.00	0.00	1
13	1.51589	12.88	3.43	1.40	73.28	0.69	8.05	0.00	0.24	1
14	1.51748	12.86	3.56	1.27	73.21	0.54	8.38	0.00	0.17	1
15	1.51763	12.61	3.59	1.31	73.29	0.58	8.50	0.00	0.00	1
16	1.51761	12.81	3.54	1.23	73.24	0.58	8.39	0.00	0.00	1
17	1.51784	12.68	3.67	1.16	73.11	0.61	8.70	0.00	0.00	1

2.) Implement a KNN model to classify the animals in categories.

	animal.name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic
1	aardvark	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0
2	antelope	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0
3	bass	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0
4	bear	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0
5	boar	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0
6	buffalo	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0
7	calf	1	0	0	1	0	0	0	1	1	1	0	0	4	1	1
8	carp	0	0	1	0	0	1	0	1	1	0	0	1	0	1	1
9	catfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0
10	cavy	1	0	0	1	0	0	0	1	1	1	0	0	4	0	1
11	cheetah	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0
12	chicken	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1
13	chub	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0
14	clam	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
15	crab	0	0	1	0	0	1	1	0	0	0	0	0	4	0	0
16	crayfish	0	0	1	0	0	1	1	0	0	0	0	0	6	0	0
17	crow	0	1	1	0	1	0	1	0	1	1	0	0	2	1	0
18	deer	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0
19	dogfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0

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: Feature Engineering, Hyper Parameter tuning etc.

3. All the codes (executable programs) are running without errors

4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here.