Topic: Support Vector Machine (SVM)

1) Prepare a classification model using SVM for salary data

	age	workclass	education	educationno	maritalstatus	occupation	relationship	race	sex	capital
1	39	State-gov	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174
2	50	Self-emp-not-inc	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0
3	38	Private	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0
4	53	Private	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0
5	28	Private	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0
6	37	Private	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0
7	49	Private	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black	Female	0
8	52	Self-emp-not-inc	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband	White	Male	0
9	31	Private	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084
10	42	Private	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	5178
11	37	Private	Some-college	10	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0
12	30	State-gov	Bachelors	13	Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0
13	23	Private	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0
14	32	Private	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black	Male	0
15	34	Private	7th-8th	4	Married-civ-spouse	Transport-moving	Husband	Amer-Indian-Eskimo	Male	0

2.) Prepare svm model for classifying the area under fire for forest fires data

	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area	dayfri	daymon	daysat	daysun
1	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.00	1	0	0	0
2	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.00	0	0	0	0
3	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.00	0	0	1	0
4	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.00	1	0	0	0
5	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.00	0	0	0	1
6	aug	sun	92.3	85.3	488.0	14.7	22.2	29	5.4	0.0	0.00	0	0	0	1
7	aug	mon	92.3	88.9	495.6	8.5	24.1	27	3.1	0.0	0.00	0	1	0	0
8	aug	mon	91.5	145.4	608.2	10.7	8.0	86	2.2	0.0	0.00	0	1	0	0
9	sep	tue	91.0	129.5	692.6	7.0	13.1	63	5.4	0.0	0.00	0	0	0	0
10	sep	sat	92.5	88.0	698.6	7.1	22.8	40	4.0	0.0	0.00	0	0	1	0
11	sep	sat	92.5	88.0	698.6	7.1	17.8	51	7.2	0.0	0.00	0	0	1	0
12	sep	sat	92.8	73.2	713.0	22.6	19.3	38	4.0	0.0	0.00	0	0	1	0
13	aug	fri	63.5	70.8	665.3	0.8	17.0	72	6.7	0.0	0.00	1	0	0	0
14	sep	mon	90.9	126.5	686.5	7.0	21.3	42	2.2	0.0	0.00	0	1	0	0
15	sep	wed	92.9	133.3	699.6	9.2	26.4	21	4.5	0.0	0.00	0	0	0	0
16	sep	fri	93.3	141.2	713.9	13.9	22.9	44	5.4	0.0	0.00	1	0	0	0
17	mar	sat	91.7	35.8	80.8	7.8	15.1	27	5.4	0.0	0.00	0	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