[2024] Pattern Recognition Projects (CS)

The objective of the projects is to prepare you to apply different machine learning algorithms to real-world tasks. This will help you to increase your knowledge about the workflow of the machine learning tasks. You will learn how to clean your data, applying pre-processing, feature engineering, regression, and classification methods. Each project will be delivered in milestones.

- The best three teams for each project will be honored.
- Registration starts: Wednesday 3/4/2024 11:59 PM.
- Registration ends: Saturday 6/4/2024.
- Delivering Milestone 1: 23/4/2024 11:59 PM.
- Delivering Milestone 2: Practical exam.
- Minimum number of members is 4 and the maximum is 6, **however teams with 6 members will be assigned extra requirements to be announced later**.
- You must deliver a detailed report for each milestone contains all your work (feature analysis, algorithms used in each module and the achieved accuracy for each one)

Note : Each report will be graded

In the first milestone, you will apply the followings :-

Preprocessing: Before building your models, you need to make sure that the dataset is clean and ready-to-use.

Regression: Apply different regression techniques (at least two) to find the model that fit your data with minimum error.

Milestone 1: 50%

- Preprocessing, Regression.

Milestone 1 Report Must Include:

- You must explain in details the **preprocessing techniques** you needed to apply on your dataset and how you implemented them.
- Perform **analysis** on the dataset as studied and explain how the features affect and relate to each other.
- You must explain what **regression techniques** you used (at least two).
- Mention the **differences** between each model and the acquired **results** (accuracy/error and so on).
- You must clearly mention **what features** you used or discarded to create your regression models.
- Explain what the **sizes** of your training, testing and validation sets are, if exist.
- Mention any further techniques that were used to **improve** the results (if exist).
- You should include **screenshots** of the resultant(s) regression line plots if applicable.
- Finally, write a **conclusion** about this phase of the project and what intuition you had about your problem and how it was proved/disproved.

Project(1): Apartment Rent Prediction

The "Apartment Rent Prediction" project aims to leverage advanced machine learning algorithms to accurately forecast rental prices for apartments in a given area. In today's dynamic real estate market, both tenants and landlords can benefit immensely from having reliable estimates of rental prices.

Dataset Snapshots:

id	category	title	body	amenities	bathrooms	bedrooms	currency	fee	has_photo	pets_allow	price	price_displ	price_type	square_fe
5668626895	housing/rent/apa	Studio apartment 2	This unit is located at second St NE, Uhland Terrace	NE, Washington, DC 20002, V		0	USD	No	Thumbnail	None	790	$790	Monthly	101
5664597177	housing/rent/apa	Studio apartment 8	This unit is located at 814 Schutte Road, Evansville, 47712, INMonthly rental rates		1	1	USD	No	Thumbnail	None	425	$425	Monthly	106
5668626833	housing/rent/apa	Studio apartment N	This unit is located at N Scott St, 14th St N, Arlington, VA 22209,		1	0	USD	No	Thumbnail	None	1390	$1,390	Monthly	107
5659918074	housing/rent/apa	Studio apartment 1	This unit is located at 1717 12th Ave, Seattle, 98122, WAMonthl		1	0	USD	No	Thumbnail	None	925	$925	Monthly	116
5668626759	housing/rent/apa	Studio apartment V	This unit is located at Washington Blvd, N Cleveland St, Arlington, Arlington, 22201			0	USD	No	Thumbnail	None	880	$880	Monthly	125
5667891676	housing/rent/apa	0 BR in New York N	**RARE GEM WITH PRIVATE OUTDOOR TERRACE**	Dishwashe	1	0	USD	No	Thumbnail		2475	$2,475	Monthly	130
5668627426	housing/rent/apa	Studio apartment 2	This unit is located at 2432 Penmar Ave, Venice, 90291, CAMonthly rental rates ra			0	USD	No	Thumbnail	None	1800	$1,800	Monthly	132
5668626687	housing/rent/apa	Studio apartment C	This unit is located at Oak St NW, 16th St NW, Washington, DC 20010, Washingtor			0	USD	No	Thumbnail	None	840	$840	Monthly	136
5668610290	housing/rent/apa	Studio apartment 3	This unit is located at 333 Hyde St, San Francisco, 9	Refrigerato	1	0	USD	No	Thumbnail	None	1495	$1,495	Monthly	138
5668627023	housing/rent/apa	Studio apartment A	This unit is located at A St SE, 19th St SE, Washington, Washington, 20003, DCMor			0	USD	No	Thumbnail	None	890	$890	Monthly	141
5668627099	housing/rent/apa	Studio apartment 1	This unit is located at 15th St SE, Independence Ave SE, Washington DC, Washingt			0	USD	No	Thumbnail	None	990	$990	Monthly	146
5668626548	housing/rent/apa	Studio apartment A	This unit is located at Arkansas Ave NW, Varnum St NW, Washington, Washington			0	USD	No	Thumbnail	None	840	$840	Monthly	150
5664598162	housing/rent/apa	Studio apartment 2	This unit is located at 2326 N sixth Avenue, Tucson, 85705, AZMonthly rental rates			1	USD	No	Thumbnail	Cats,Dogs	1795	$1,795	Monthly	156
5668626900	housing/rent/apa	Studio apartment S	This unit is located at Salem Ln NW, 45th St NW, Washington, Washington, 20007			0	USD	No	Thumbnail	None	1090	$1,090	Monthly	178
5664571820	housing/rent/apa	Studio apartment 5	This unit is located at 57 Taylor Street, San Francisc	Refrigerato	1	0	USD	No	Thumbnail	None	1695	$1,695	Monthly	190
5659276240	housing/rent/apa	Studio Cottage 214	New Bern Studio includes : 1 beds, 1 microwave, 1 r	AC,Basket	1	1	USD	No	Thumbnail	Cats,Dogs	1560	$1,560 We	Weekly	200
5654898031	housing/rent/apa	Studio Cottage 214	New Bern Studio includes : 1 bedrooms, 1 microwa	AC,Basket	1	1	USD	No	Thumbnail	Cats,Dogs	1560	$1,560	Monthly	200

~Dataset Header Cont'd:

address	cityname	state	latitude	longitude	source	time
	Washingto	DC	38.9057	-76.9861	RentLingo	1577359415
814 Schutte Rd	Evansville	IN	37.968	-87.6621	RentLingo	1577017063
	Arlington	VA	38.891	-77.0816	RentLingo	1577359410
1717 12th Avenue	Seattle	WA	47.616	-122.328	RentLingo	1576667743
	Arlington	VA	38.8738	-77.1055	RentLingo	1577359401
350 West 50th St	Manhattar	NY	40.7629	-73.9885	Listanza	1577289784
2432 Penmar Avenue	Venice	CA	33.9932	-118.461	RentLingo	1577359461
	Washingto	DC	38.9328	-77.0297	RentLingo	1577359393
333 Hyde St	San Franci	CA	37.7599	-122.438	RentLingo	1577358313
	Washingto	DC	38.9118	-77.0132	RentLingo	1577359424
15th St SE Independence Avenue S	Washingto	DC	38.9118	-77.0132	RentLingo	1577359430
	Washingto	DC	38.9118	-77.0132	RentLingo	1577359381
2326 N 6th Ave	Tucson	AZ	32.2264	-110.944	RentLingo	1577017148
	Washingto	DC	38.9118	-77.0132	RentLingo	1577359415
57 Taylor St	San Franci	CA	37.7599	-122.438	RentLingo	1577015121
	New Bern	NC	35.0847	-77.0609	ListedBuy	1576618076

Milestone 1 tasks:

1. Apply pre-processing on the provided dataset. (You must preprocess all the features even if you won't use them later after feature selection)
2. Apply Feature Selection and Experiment with regression techniques to reduce the error on prediction of the "price_display" (Deliver at least two regression models with significant difference).
3. Finish Milestone 1 Report.

Note: You must preprocess all features, but model and feature selection can be done after that (i.e You can drop a feature only after preprocessing and with valid reason)

Project(2): Electronic Device Rating Prediction

In today's rapidly evolving technology landscape, consumers are often faced with a multitude of options when purchasing electronic devices such as smartphones, laptops, tablets, and wearables. This project seeks to provide consumers, retailers, and manufacturers with valuable insights into the pricing dynamics of electronic devices, enabling informed decision-making and strategic planning.

Dataset Snapshots:

brand	processor_	processor_	processor_gn	ram_gb	ram_type	ssd	hdd	os	graphic_c
ASUS	Intel	Core i3	10th	4 GB	DDR4	0 GB	1024 GB	64-bit Windows	0 GB
Lenovo	Intel	Core i3	10th	4 GB	DDR4	0 GB	1024 GB	64-bit Windows	0 GB
Lenovo	Intel	Core i3	10th	4 GB	DDR4	0 GB	1024 GB	64-bit Windows	0 GB
ASUS	Intel	Core i5	10th	8 GB	DDR4	512 GB	0 GB	32-bit Windows	2 GB
ASUS	Intel	Celeron D	Not Available	4 GB	DDR4	0 GB	512 GB	64-bit Windows	0 GB
ASUS	Intel	Celeron D	Not Available	4 GB	DDR4	0 GB	512 GB	64-bit Windows	0 GB
ASUS	Intel	Celeron D	Not Available	4 GB	DDR4	0 GB	512 GB	32-bit Windows	0 GB
ASUS	Intel	Core i5	10th	8 GB	DDR4	0 GB	1024 GB	32-bit Windows	2 GB
Lenovo	Intel	Core i5	10th	4 GB	DDR4	0 GB	1024 GB	32-bit Windows	0 GB
acer	AMD	Ryzen 5	10th	4 GB	DDR4	0 GB	512 GB	32-bit Windows	4 GB
acer	AMD	Ryzen 5	10th	4 GB	DDR4	0 GB	512 GB	32-bit Windows	6 GB
acer	AMD	Ryzen 5	10th	4 GB	DDR4	0 GB	512 GB	32-bit Windows	4 GB
Avita	Intel	Celeron D	Not Available	4 GB	DDR4	0 GB	512 GB	64-bit Windows	0 GB

~Dataset Header Cont'd:

weight	warranty	Touchscre	msoffice	Price	rating	Number of	Number of Reviews
Casual	No warranty	No	No	34649	2 stars	3	0
Casual	No warranty	No	No	38999	3 stars	65	5
Casual	No warranty	No	No	39999	3 stars	8	1
Casual	No warranty	No	No	69990	3 stars	0	0
Casual	No warranty	No	No	26990	3 stars	0	0
Casual	No warranty	No	No	22990	3 stars	0	0
Casual	No warranty	No	No	21990	3 stars	31	3
Casual	No warranty	No	No	58799	3 stars	0	0
Casual	No warranty	No	No	49999	3 stars	0	0
Casual	No warranty	No	No	59990	4 stars	1946	240
Casual	No warranty	No	No	93700	4 stars	259	47
Casual	No warranty	No	No	72990	4 stars	425	67
Casual	No warranty	No	No	17490	3 stars	1120	196

Milestone 1 tasks:

1. Apply pre-processing on the provided dataset. (You must preprocess all the features even if you won't use them later after feature selection). **You must use One hot encoding for at least one feature.**

2. Apply Feature Selection and Experiment with regression techniques to reduce the error on prediction of the “rating” (Deliver at least two regression models with significant difference).
3. Finish Milestone 1 Report.

Note: You must preprocess all features, but model and feature selection can be done after that (i.e You can drop a feature only after preprocessing and with valid reason)