HW3: Regression Model Implementation

Design and implement your own machine learning models to predict land prices in Saudi Arabia. Experiment with different machine learning algorithms and report only the best-performing model you can get. It is highly recommended that you build this model using scikit-learn library.

Task Details

You are presented with land data in Saudi Arabia (aqardata_2.csv) and you are asked to design and implement a regression model that can predict the price of land. Experiment with different machine learning algorithms, and report only the best performing model you can get for this task. Evaluation results of your best model must be using a 5-fold cross-validation methodology on the provided data. Note that there are several ways to do cross validation in scikit-learn; however, it is highly recommended that you use `KFold` (with `shuffle=True` and `random_state=24`). More details can be found at these links:

- KFold:
    - https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html#sklearn.model_selection.KFold
- Cross-validation:
    - https://scikit-learn.org/stable/modules/cross_validation.html

Data Information:

The dataset consists of 2951 records (number of instances) and 8 features (number of attributes). This dataset has been scrapped off sa.aqar.fm to obtain land information such as price, size, street width, and locations.

To download and get more information about the data:

https://www.kaggle.com/hussainalqatari/land-market-in-saudi-arabia

Optional: You can expand your dataset in order to improve your model using other sources such as Ministry of Justice open data:

1. https://open.data.gov.sa/ar/datasets/view/ec57c751-35c1-457c-b770-1b912bb0c705/resources
2. https://open.data.gov.sa/ar/datasets/view/bb9624d4-17f1-446e-b14b-44604f9939d0/resources
3. https://open.data.gov.sa/ar/datasets/view/e0e477d7-753d-4fee-98de-a474987079e6/resources
4. https://open.data.gov.sa/ar/datasets/view/a530f315-f5b7-4693-9d12-e46c7208cb18/resources

Submission Instructions:

After you finish your work, submit **two** files in the LMS (a **PDF** report & a **ZIP** archive for the source code) in the LMS (Blackboard), and then fill in the provided form. Follow these instructions:

1. Submit **two** files in the LMS (a **PDF** report & a **ZIP** archive for the source code) as the following:
    1.1. A **PDF** report that describes all the steps you took to design and implement your models along with the following information:
        1.1.1. Your name and student ID
        1.1.2. The IDE you used
        1.1.3. The libraries you used to implement the models
        1.1.4. Description of the chosen algorithms including all the used hyperparameters
        1.1.5. Description of all data preprocessing you applied to the provided dataset
        1.1.6. Description of all other steps you have done to get your final results
        1.1.7. Evaluation results of your models using a 5-fold cross-validation methodology and report the following results (averaged across the folds):
            o   Mean Absolute Error (MAE)
            o   Root Mean Square Error (RMSE)
            o R^2 (if applicable)
        1.1.8. **Explain** the previously mentioned **evaluation metrics** (including how they are calculated using formulas) and **discuss what your results mean**.
        1.1.9. Screenshots of **all the code** & **evaluation results** from the platform you chose
    1.2. A **ZIP** compressed file containing all the source code of your models (i.e., Jupyter Notebook files)
2. Fill in this online **form** after you submit the aforementioned deliverables:
    *   https://forms.gle/3N8uGSNj8SGHqB9M6

Important Notes:

*   Don't forget to fill in the provided Google **form** after you submit the required files in the LMS.
*   Make sure to submit the two files **separately** (do not include the **PDF** report inside the **ZIP** archive).
*   The report file must be in **PDF** format. Don't use other formats such as doc, docx, odt, or txt.
*   The compressed file containing your source code must be in **ZIP** format. Don't use other formats such as rar, 7s, or tar.
*   Provide screenshots of all the code, not just some parts.