**Class 7 Mid Term Test**

Review các kiến thức đã học:

Thuật toán:

* Supervised learning :

+ Linear Regression

+ Non-Linear Regression

+ Logistic Regression

+ Decision Tree & Random Forest

+ SVM & SVR

+ KNN

+ Bayesian Classification

* Data Processing Techniques :

+ PCA

+ Encoder ( one hot – categorical )

+ Fill n/a

+ IQR

+ Outlier checking

+ Bin data

* Data evaluation techniques :

+ K fold cross validation

+ Loss function validation (gradient descent )

+ SME , SMA , Rsquare

+ Loss function

Các bạn review lại kiến thức

Nội dung test : Các thuật toán Supervised ( Regression,Decision Tree, Random Forest , SVM, KNN) đã học.

Test 1 : Predict Used Car price (Linear – Polynomial Regression )

Các bạn download dataset sau và chúng ta cùng tiến hành EDA & preprocessing

#### Attribute Information:

Attribute: Attribute Range

1. symboling: -3, -2, -1, 0, 1, 2, 3.

2. normalized-losses: continuous from 65 to 256.

3. make:

alfa-romero, audi, bmw, chevrolet, dodge, honda,

isuzu, jaguar, mazda, mercedes-benz, mercury,

mitsubishi, nissan, peugot, plymouth, porsche,

renault, saab, subaru, toyota, volkswagen, volvo

4. fuel-type: diesel, gas.

5. aspiration: std, turbo.

6. num-of-doors: four, two.

7. body-style: hardtop, wagon, sedan, hatchback, convertible.

8. drive-wheels: 4wd, fwd, rwd.

9. engine-location: front, rear.

10. wheel-base: continuous from 86.6 120.9.

11. length: continuous from 141.1 to 208.1.

12. width: continuous from 60.3 to 72.3.

13. height: continuous from 47.8 to 59.8.

14. curb-weight: continuous from 1488 to 4066.

15. engine-type: dohc, dohcv, l, ohc, ohcf, ohcv, rotor.

16. num-of-cylinders: eight, five, four, six, three, twelve, two.

17. engine-size: continuous from 61 to 326.

18. fuel-system: 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.

19. bore: continuous from 2.54 to 3.94.

20. stroke: continuous from 2.07 to 4.17.

21. compression-ratio: continuous from 7 to 23.

22. horsepower: continuous from 48 to 288.

23. peak-rpm: continuous from 4150 to 6600.

24. city-mpg: continuous from 13 to 49.

25. highway-mpg: continuous from 16 to 54.

26. **\*\*price: continuous from 5118 to 45400.\*\***

Link : [Link](https://archive.ics.uci.edu/ml/machine-learning-databases/autos/imports-85.data)

Test 2 : Predict Loan Pay Off ( Logistic Regression )

Graphical user interface, text, application

Description automatically generated

Các bạn download dataset sau và chúng ta cùng tiến hành EDA & Preprocessing

Link : <https://raw.githubusercontent.com/maxtran1997gmail/Files/main/loan_train.csv>

Choose one of the 2 tests :

Test 3 : Predict Heart Disease using any Supervised Method.

Requirements : Using at Least 2 methods and compare.

Text

Description automatically generated

Link : [Link](https://raw.githubusercontent.com/maxtran1997gmail/Files/main/dataset_2190_cholesterol.csv)

Test 4 : Wine Classification

Link : [Link](http://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.data)

Graphical user interface, text, application

Description automatically generated

Extra :

Predict Cooling and Heating Loads using Multivariate Regression

Info : [Link](https://archive.ics.uci.edu/ml/datasets/Energy+efficiency)

A picture containing text, person

Description automatically generated

Link : [Link](https://archive.ics.uci.edu/ml/machine-learning-databases/00242/ENB2012_data.xlsx)