Anaconda pass : hasan242Mh2

What is DATA?

DDDM - Data-driven decision-making (DDDM) is defined as **making decisions based on hard data as opposed to intuition, observation, or guesswork**. The value of data-driven decisions is dependent on the quality of the data and its analysis and interpretation.

Automation & Data science – Automation a intelligence nai, command only but Data science a intelligence ace.

Machine learning – 1 statistical 2 nural network

Two different types of machine learning model:

* Supervised
* Unsupervised

Simple Linear Regression:

linear mean amra Jodi kono problem k akta line r mote akta model dea solve korte chai then seta hobe linear model.

Regression mean real value, output hisabe predict korte hobe

So, linear regression mean line r model dea real value predict kora bujhay

* Model ta baka hole linear regression thakto na seta polynomial regression hoy jeto
* So linear regression holo dependent and independent variable r moddhe relation make korar process matro

In csv file ‘area’ column hocce feature & ‘price’ column hocce level

# linear regression hocce superviced machine learning processs

# feature hoy 2 dimension as input

28/03/22

We are importing 4 libraries

* Pandas
* Matplotlib.pyplot
* Numpy
* Sklearn

Scikit-learn (sklearn): <https://scikit-learn.org/stable/getting_started.html>

# What is statistical modeling?

Sklearn r moddhe linear regression ace

#task – car driving risk analysis - <https://www.kaggle.com/datasets/studymart/car-driving-risk-analysis>

For training purpose 60% data and for test purpose 40% data

Speed r basis a risk analysis korbo

Need to find –

* best fitted line
* Need to predict risk for different speed
* Score
* Accuracy
* M – coefficient
* C – intercept
* Prove y

<https://stackoverflow.com/questions/45627784/unable-to-obtain-accuracy-score-for-my-linear>

29/03/22

# how to plot graph using matplotlib.pyplot library

Plt.scatter(df[‘speed’], df[‘risk’], marker = ‘\*’, color = ‘blue’)

Plt.title(‘Graph title’)

Plt.xlebel(‘speed in km’)

Plt.ylebel(‘risk in percentage %’)

Doing the task – car driving risk analysis

Breaking the task into sequence.

Importing 3 libraries

(pandas, matplotlib, numpy)

Read the csv file

Store features and level in x & y variable

from sklearn.model\_selection import train\_test\_split

xtrain, xtest, ytrain, ytest = train\_test\_split(x,y, test\_size = .40, random\_state = 1)

from sklearn.linear\_model import LinearRegression

store LinearRegression in a variable ‘reg’

reg.fit(xtrain, ytrain)

plt.plot(df.speed, reg.predict(df[['speed']]))