***Introduction To Machine Learning – Mercan Karacabey***

***TASK-1 What is machine learning?***

General Definition : Gives computers the ability to learn without being explicitly programmed.

ArtificiaI Intelligence-> Machine Learning ->Deep Learning

It is learned by machines based on some algorithms by utilizing its composition and values from the density of data.(Without human programming and rule based programming

***Machine Learning :***

***Explains with 3 categories:***

* Supervised Learning : Classification & Regression – Labels Known
* Unsupervised Learning : Clustering(k-means / dbscan) & Dimension Reduction – Labels not known
* Reinforcement Learning : Deep Learning – It used by enhancing power of learning.

Dimension Reduction 🡪 It is critic and important for performance and visualization.

***Knowledge Driven versus Data Driven->***

Knowledge Driven : Human based approach – Condition Based

Data Driven : Based on data composition , future extraction

***Black Box – White Box 🡪***

If we know the content of code defined as White Box , against of this condition defined as Black Box

***Black Winter of AI 🡪***

Between the 1970s and 1980s, it was not enough knowledge driven models for learning capacity and expected results.

Explicit Programming is similar to White Box Programming.

***TASK -2 Supervised versus Unsupervised Learning***

Evaluation - Labels

Supervised uses labels.

Unsupervised does not use labels.

*Categories*

* Supervised Learning : Classification & Regression – Labels Known
* Unsupervised Learning : Clustering(k-means / dbscan) & Dimension Reduction – Labels not known

*Supervised Learning Examples:*

E-Mail Filtering : Spam or not?

Supervised Learning , because label is known. Regression or decision tree can be use.

Pearson Test of English

Realty Classification

Unsupervised Learning Examples:

Clustering Problems – Define Customer Interest - Anomaly Detection

Some Image Processing Progress

Fraud Detection

***TASK 3 : Data Science Project Progress:***

*What are the common steps in data preprocessing ?*

🡪Define research goals

🡪Find data and retrieve data

🡪Data Preparation // Cleaning // Outlier Detection // Data Scaling(MIN-MAX Method-Z-Score Method)

🡪Data Augmentation // Exploration

🡪Data Modeling – Select algorithm -- Learning Model – Model Run & Execution

🡪Presentation and automation

*\_\_Data Prepation Step : Anomaly Detection Basics\_\_*

*Outlier Detection*

*Clean Noise*

*Find Novelties*

*Deviations*

*Exceptions*

For min-max standard scaling algorithms : In the preprocessing steps, for missing values we use statistical methods.One of these statistical method is min-max step. Or I replace the median/mean value and missing value.(Has Gaussion Curve – Normal Distributed Situation)

Min-max : Fill the values with 0-1 related column.

MIN-MAX – Standard Scaling – Using Functions

import pandas as pd  
df = pd.read\_csv("iris.csv")  
df.head()  
  
df.tail()  
  
df["sepal.length\_new"] = df["sepal.length"]  
  
df = df.drop("sepal.length",1)  
  
  
print(df)  
  
from sklearn import preprocessing  
  
x = df.values  
min\_max\_scaler = preprocessing.MinMaxScaler()  
x\_scaled = min\_max\_scaler.fit\_transform(x)  
df = pd.DataFrame(x\_scaled)  
  
  
from sklearn.preprocessing import StandardScaler  
  
scaler = StandardScaler()  
print(scaler.fit(df))  
StandardScaler(copy=True, with\_mean=True, with\_std=True)  
print(scaler.mean\_)  
  
print(scaler.transform(df))  
  
print(scaler.transform(df))

MIN-MAX – STANDARD SCALING – NOT USING FUNCTIONS

df["sepal.length\_new"].min()  
df["sepal.length\_new"].max()  
  
min\_max\_scaler\_value = []  
for i in df["sepal.length\_new"]:  
 min\_max\_scaler\_value.append(i - min(df["sepal.length\_new"]) / max(df["sepal.length\_new"])-min(df["sepal.length\_new"]))  
   
   
z\_score\_calculation = []  
for i in df["sepal.length\_new"]:  
 z\_score\_calculation.append(i - (df["sepal.length\_new"]).mean / df["sepal.length\_new"].std / math.sqrt(len(df["sepal.length\_new"])))