

**Velagapudi Ramakrishna Siddhartha**  
**Engineering College**

**Department of Computer Science & Engineering**

REPORT ON

**One Week Training Program**

ON

**“Applied Data Sciences”**

17<sup>th</sup> – 22<sup>nd</sup> May 2021

The Department of Computer Science and Engineering has successfully organized a One Week Training Program on “Applied Data Sciences” from 17 May 2021 to 22 May 2021 in association with Cluster IT Solutions Pvt Ltd . A total of 181 students have attended this training program from department of CSE.

The main objectives for organizing this training program are:

- To understand key components of any ML model
- To apply machine learning algorithm types for real-world applications.
- To understand the statistical principles that forms the foundation of ML
- To develop basic Supervised learning models
- To develop basic Unsupervised learning models
- Apply various methods for evaluating the machine learning training models
- Explore the opportunities and challenges the Artificial Intelligence and Machine Learning offer and insights into exciting advancements in technology.

*Training Program Speakers:*

S.No	Speakers
1.	B.Sunil Kumar, Data Scientist Cluster IT Solutions Pvt Ltd
2	Mrs.Dharna Ahuja Data Scientist Cluster IT Solutions Pvt Ltd
3	Mr A.Vivek Data Scientist Harman Connected Services Technologies Pvt Ltd

## Day1: 17<sup>th</sup> May 2021

**Speaker:** Ms.Dharna Ahuja  
Data Scientist  
Cluster IT Solutions Pvt Ltd

### **Topic: Libraries of Python used in machine learning**

This session was handled by Mrs.Dharna Ahuja madam, she handled the session on teaching the libraries of python like Numpy, Pandas, Seaborn, Matplotlib.

**Numpy:** NumPy is a Python library used for working with arrays.

It also has functions for working in domain of linear algebra, fourier transform, and matrices.

**Pandas:** Pandas is an open-source, Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

**Matplotlib:** Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

**Seaborn:** Seaborn is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions.

The screenshot shows a Cisco Webex Meeting interface. The main content is a presentation slide titled "df.replace(to\_replace='Same', value=10000)". The slide displays a table with columns "Position", "level", and "Salary". The table contains 10 rows of data. The interface also shows a list of participants on the right, including "VRSEC", "Dharna ahuja", and several other users with IDs. A chat window at the bottom right shows a message: "9/9/201804/ Thank you madam".

Position	level	Salary
0. Business Analyst	1	45000
1. Junior Consultant	2	50000
2. Senior Consultant	3	60000
3. Manager	4	80000
4. Group Manager	5	100000
5. Project Manager	6	150000
6. (A)Level	7	200000
7. Senior Partner	8	300000
8. Partner	9	500000
9. (S)EC	10	1000000

Cisco Webex Meetings Meeting Info Connected

Dharna ahuja 188W1A05A7\_TANKA...

Participants (174)

YRSEC H041, 118

Dharna ahuja

1A 188W1A05A8\_TRIPAR...

1 188W1A0562

0A 0554 Abhishek

1D 178W1A05A3-SEC-B P...

1S 188W1A0501-Akunuri...

1S 188W1A0502-Arvapafl...

1A 188W1A0503\_VYBHAV...

Mute All Unmute All

Chat

9793018047

From YRSEC to Dharna ahuja (Private) 17:01 PM

Thank you madam

Send to Everyone

Unmute Start video Share Record

Present Chat

Viewing Dharna ahuja's...

Untitled2.ipynb

```
df.replace(to_replace="Partner",value="A-Level")
```

	Position	Level	Salary
0	Business Analyst	1	45000
1	Junior Consultant	2	50000
2	Senior Consultant	3	60000
3	Manager	4	80000
4	Country Manager	5	110000
5	Region Manager	6	150000
6	A-Level	7	200000
7	Senior Partner	8	300000
8	Chief	9	500000
9	CEO	10	1000000

**Day2: 18<sup>th</sup> May2021:**

**Speaker: Mr A.Vivek**

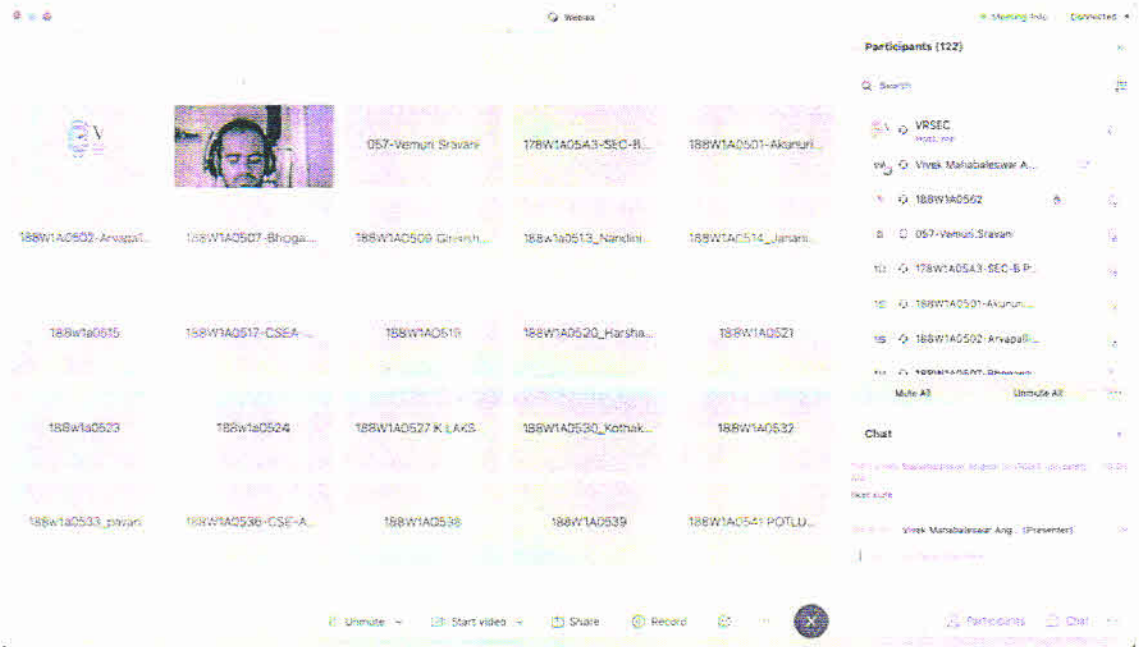
**Data Scientist**

**Harman Connected Services Technologies Pvt Ltd**

**Topic: Supervised Learning and UnSupervised Learning**

In this session Mr.A.Vivek explained the concepts of supervised and unsupervised learning .The classification, clustering and regression techniques have been discussed in the session.





**Day3: 19<sup>th</sup> May2021:**

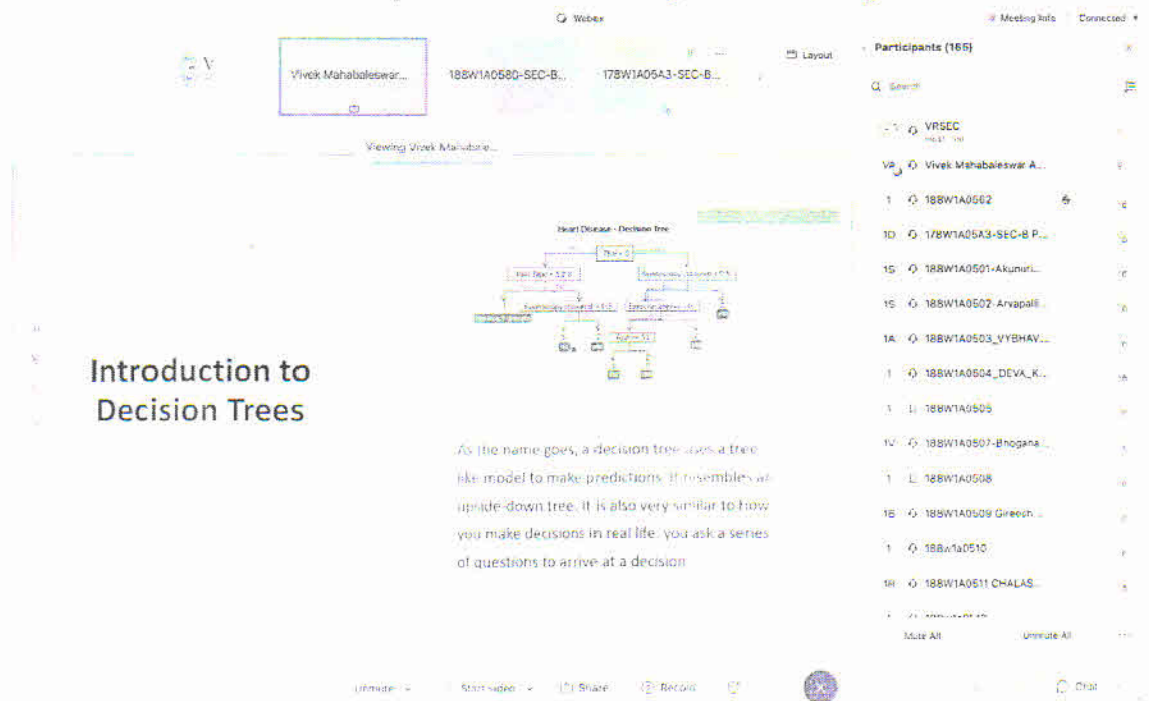
**Speaker: Mr A.Vivek**

**Data Scientist**

**Harman Connected Services Technologies Pvt Ltd**

**Topic: Decision trees**

In this session Mr.A.Vivek explained how to classify the data using decision tree model.



Day4: 20<sup>th</sup> May2021:

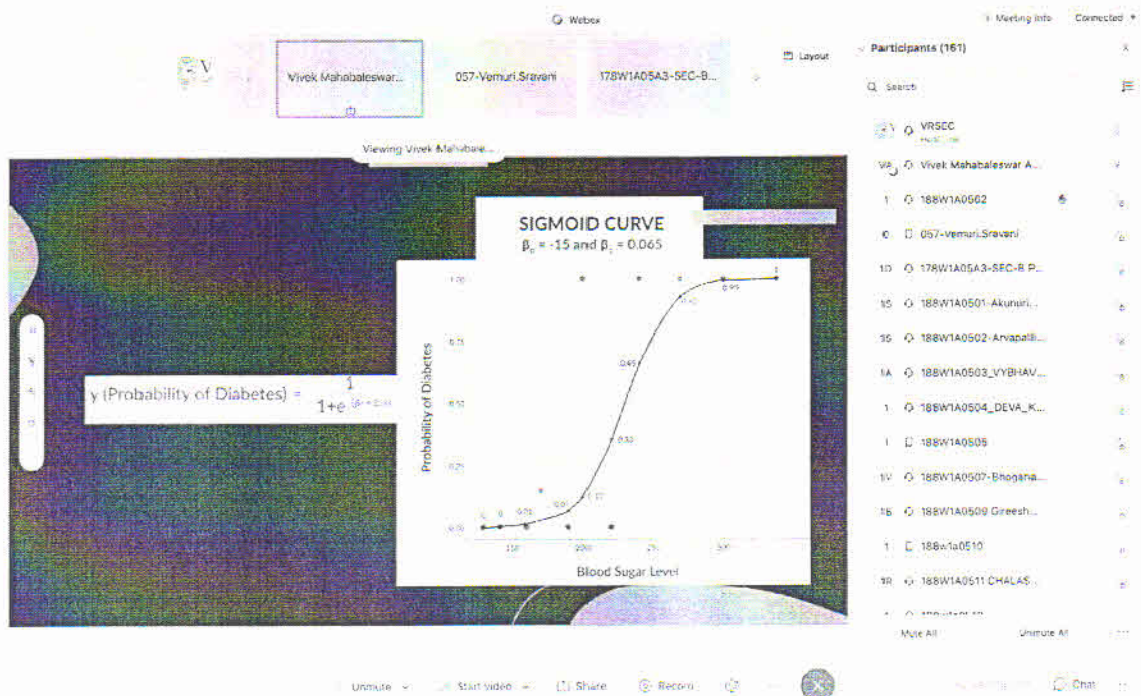
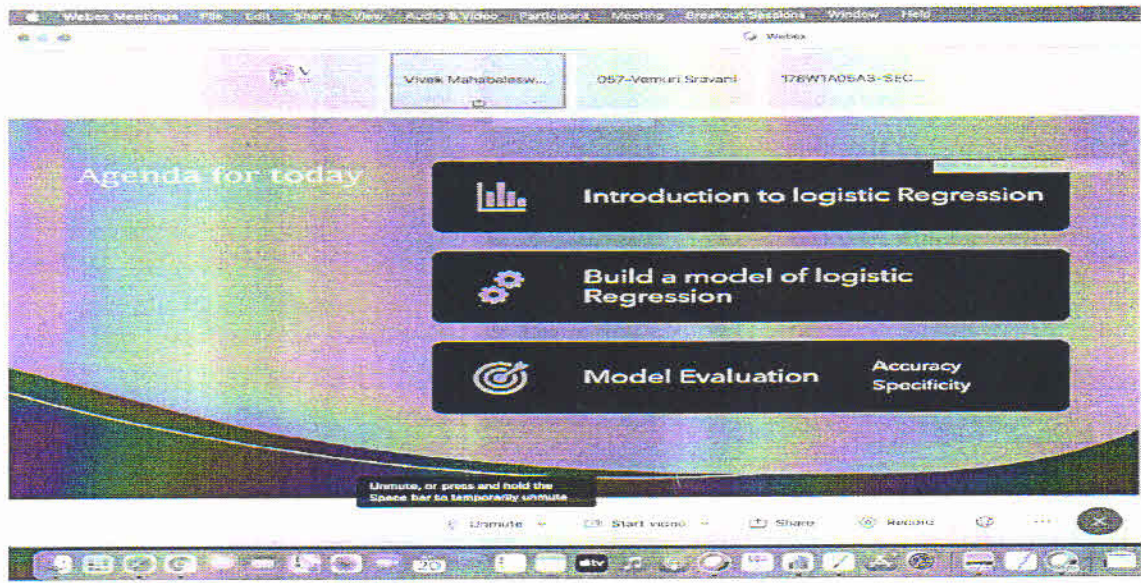
Speaker: Mr A.Vivek

Data Scientist

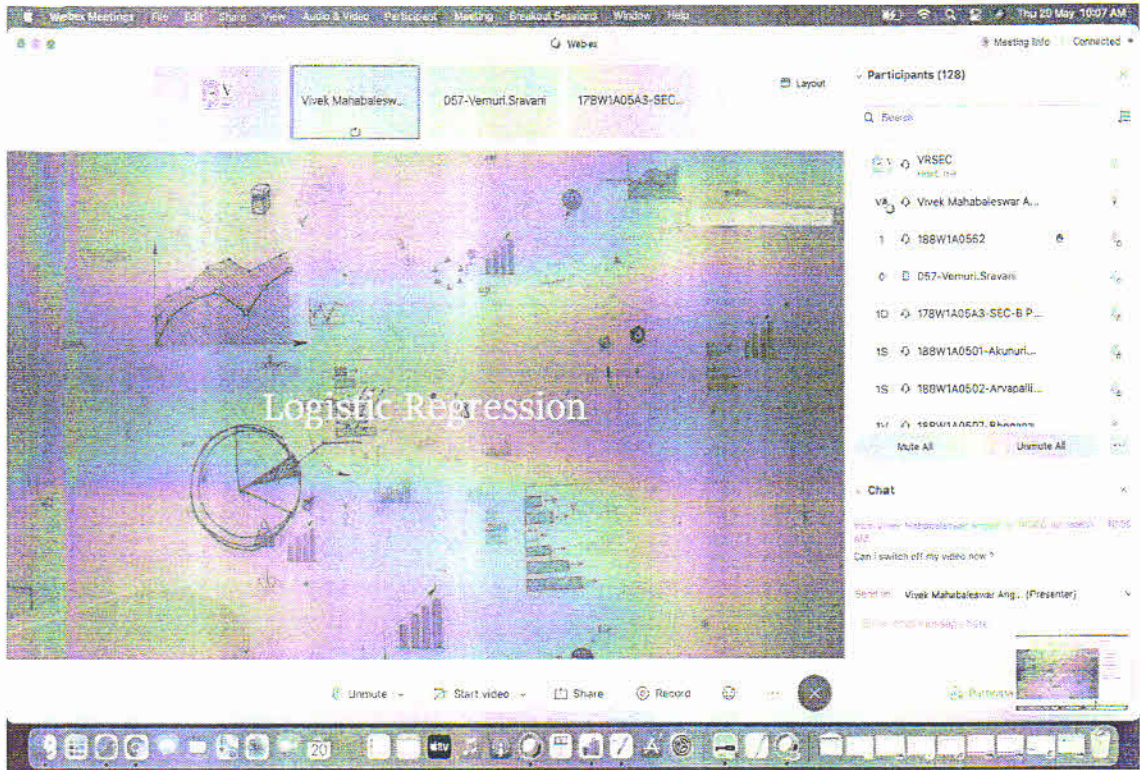
Harman Connected Services Technologies Pvt Ltd

Topic: Logistic regression

In this session Mr.A.Vivek gave an introduction to logistic regression and explained how to build a model of logistic regression and evaluate the model by assessing the accuracy and specificity.







**Day5: 21<sup>th</sup> May2021:**

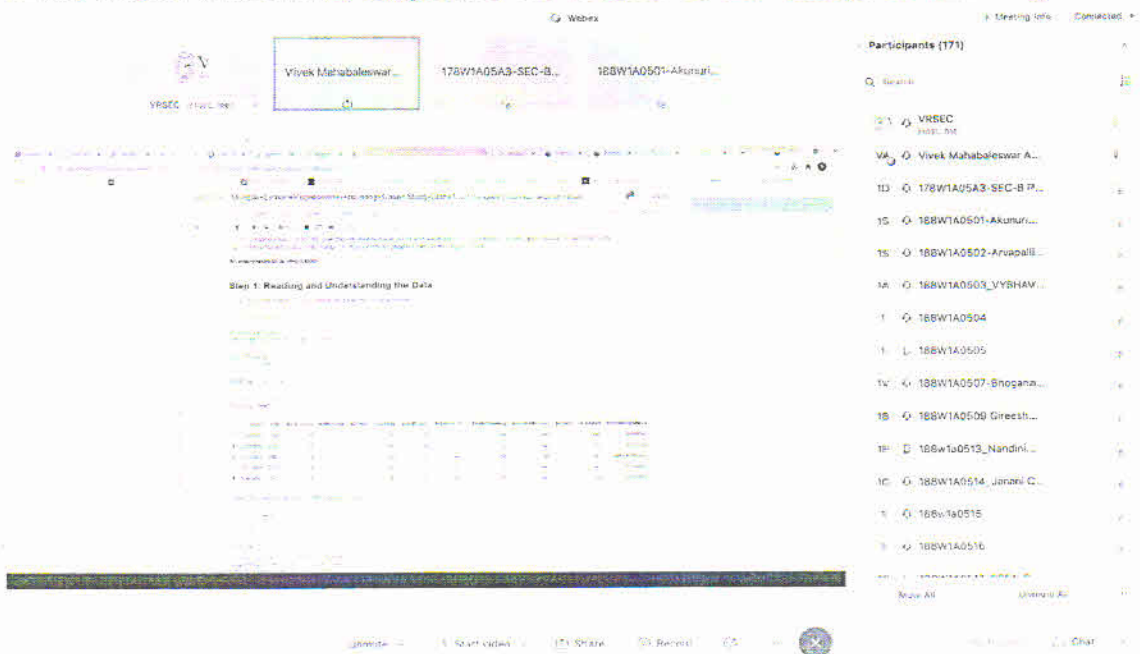
**Speaker: Mr A.Vivek**

**Data Scientist**

**Harman Connected Services Technologies Pvt Ltd**

**Topic: Linear regression**

In this session Mr.A.Vivek enlightened the students on the concept of linear regression.



Day6: 22<sup>th</sup> May2021:

Speaker : Mr A.Vivek

Data Scientist

Harman Connected Services Technologies Pvt Ltd

Topic: Random Forests

In this session Mr.A.Vivek explained the concepts of random forests using a dataset and applying the model to the dataset by considering the important parameters. The model evaluation was done by calculating accuracy, roc curve ,precision and recall.

The screenshot shows a Zoom meeting interface. At the top, there are participant avatars for 'Vivek Mohabaleswar...', '178W1A05A3-SEC-B...', and '188W1A0501-Akumut...'. The main content is a slide titled 'Diversity and Acceptability' with the following text:

Ensembles of models are somewhat analogous to teams of individual players. If you were to choose a football team, there would be two things you'd do:

- Choose people with different skill sets, such as forwards, strikers, and a goalkeeper, to ensure **diversity**, and
- Choose good players, i.e. ensure that all players are **acceptable** from a skill set point of view (and at least better than a regular person).

**Diversity** ensures that the models serve **complementary** purposes, which means that the individual models make predictions **independent of each other**. The advantages of this are different depending on the type of ensemble:

- For example, a random forest is an ensemble with a large number of trees as individual models. Diversity ensures that even if some trees overfit, the other trees in the ensemble will mitigate the effect. The independence among the trees helps in a **lower variance** of the ensemble compared to a single tree.
- Acceptability** implies that each model is at least **better than a random model**. This is a pretty arbitrary criterion for each model to be accepted into the ensemble, and it has to be at least better than a random guesser.

At the bottom of the slide, there are controls for 'Unmute', 'Start video', 'Share', 'Record', and 'More'. On the right side, there is a 'Participants (143)' list with names like 'VRSEC', 'Vivek Mohabaleswar A...', '188W1A0562', etc.

The screenshot shows a Zoom meeting interface. At the top, there are participant avatars for 'Vivek', '188W1A0536-CSE-A...', and '178W1A05A3-SEC-B...'. The main content is a code editor window titled 'Viewing 188W1A0536-C...' containing Python code for a Random Forest model. The code includes imports for 'random', 'numpy', 'sklearn', and 'matplotlib', and defines a function 'random\_forest' that takes 'X\_train', 'y\_train', 'X\_test', and 'y\_test' as input. The function uses 'RandomForestClassifier' from 'sklearn.ensemble' to train and evaluate the model, and 'roc\_auc\_score' from 'sklearn.metrics' to calculate the ROC AUC score. The code is as follows:

```
import random
import numpy as np
import sklearn
import matplotlib.pyplot as plt

def random_forest(X_train, y_train, X_test, y_test):
    # Create a random forest classifier
    rf = RandomForestClassifier()

    # Train the classifier
    rf.fit(X_train, y_train)

    # Predict on the test set
    y_pred = rf.predict(X_test)

    # Calculate the accuracy
    accuracy = sklearn.metrics.accuracy_score(y_test, y_pred)

    # Calculate the ROC AUC score
    roc_auc = sklearn.metrics.roc_auc_score(y_test, rf.predict_proba(X_test)[:, 1])

    # Print the results
    print("Accuracy: %f" % accuracy)
    print("ROC AUC Score: %f" % roc_auc)
```

At the bottom of the code editor, there are controls for 'Unmute', 'Start video', 'Share', 'Record', and 'More'. On the right side, there is a 'Participants (178)' list with names like 'VRSEC', '188W1A0536-CSE-A...', '188W1A0567', etc.

Towards the end of the session, the coordinator Dr. M.Sobhana conveyed her vote of thanks to the speaker who enlightened the students on the scope of projects in Artificial Intelligence and Machine learning.

Dr. M.Sobhana  
25-05-2021.