**Data Science and Machine Learning with Python**

**Day 01 – Thu 20 May 2021**

Today’s Agenda –

1. Introduction to DS, AI, ML, DL and importance in the real world
2. Types of ML and few basics of Python

Reference Notes –

1. <https://www.goeduhub.com/11391/science-career-science-science-scientist-responsibilities?show=11391#q11391>
2. <https://www.goeduhub.com/762/what-is-python-and-its-basic-introduction?show=762#q762>
3. <https://www.goeduhub.com/10370/machine-learning-vs-deep-learning?show=10370#q10370>

Conclusion - GO\_STP\_8480 - Kowshik: Today we learnt DS, AI, ML, DL importance in the real world and that types of the ML and few basics of Python. Today is so informative. I enjoyed a lot Thank you

**Day 02 – Fri 21 May 2021**

Today’s Agenda –

1. Mutable vs Immutable Objects in Python
2. Hands on Decision making in python (if-else condition)
3. Loops – for loop and while loop
4. Python Strings
5. Accessing String elements
6. Basic Operation on String
7. String Function and Methods
8. Python List

Reference Notes -

1. <https://www.goeduhub.com/108/loop-concepts-in-python-with-break-and-continue-statement?show=108#q108>
2. <https://www.goeduhub.com/3686/string-in-python?show=3686#q3686>
3. <https://www.goeduhub.com/270/describe-concept-of-python-list?show=270#q270>
4. <https://www.goeduhub.com/329/list-slicing-selecting-parts-of-lists>

Conclusion - GO\_STP\_8480 – Kowshik: I have a great session right here today. Today I learnt few basics from python such as strings, lists, if and else statements and the various control statements too. I learnt a lot and I have been waiting for the next session. Thank you

**Day 03 – Mon 24 May 2021**

Today’s Agenda –

1. Understanding of DS in python – Lists, Sets, Tuples, Dictionaries
2. Methods of each data structure in python

Conclusion - GO\_STP\_8480 - Kowshik: Today I've learnt about various data structures in python such as lists, sets, tuples, and dictionaries and the various methods of it. Today I enjoyed alot and I will be waiting for the next lecture...

**Day 04 – Tue 25 May 2021**

Today’s Agenda –

1. What are NumPy arrays?
2. Where is NumPy used?
3. NumPy arrays vs List?
4. Why NumPy is used in python?
5. NumPy operations
6. NumPy special functions
7. Introduction to pandas

Reference Notes –

1. <https://www.goeduhub.com/599/what-is-pandas-library-of-python-what-is-its-significance>
2. <https://www.goeduhub.com/592/what-numpy-science-python-numpy-tutorial-install-numpy-python>

Conclusion - GO\_STP\_8480 – Kowshik: We have learnt about numpy arrays Numpy array v/s list Numpy operations Numpy special functions And basics about pandas like how to install pandas and some basic functionalities

**Day 05 – Wed 26 May 2021**

Today’s Agenda –

1. Pandas Data Frames – Data Science Library
2. Pandas Slicing
3. Pandas Aggregate Functions
4. A Case Study on Pandas

Reference Notes –

1. <https://www.goeduhub.com/603/what-is-data-frame-in-pandas-how-do-you-create-a-data-frame>
2. <https://www.goeduhub.com/631/what-is-dataframe-slicing-in-pandas>
3. <https://www.goeduhub.com/636/aggregation-functions-in-pandas>

Conclusion - GO\_STP\_8480 - Kowshik: Today I've learnt about the pandas’ data frames such as sorting, slicing, filtering and few aggregate functions. Grouping and some statistical functions such as mean, median. Today I enjoyed a lot and I am waiting for next session

**Day 06 – Thu 27 May 2021**

Today’s Agenda –

1. What is Python Matplotlib?
2. What is Matplotlib used for?
3. Different Types of Plots in Matplotlib?
   1. Bar Graph
   2. Histogram
   3. Scatter Plot
   4. Area Plot
   5. Pie Chart

Reference Notes –

1. <https://www.goeduhub.com/639/what-is-matplotlib-in-data-science>

Conclusion - GO\_STP\_8480 – Kowshik: Today I've learnt about the matplotlib module and explored through various packages such as line plot, area plot, pie chart, bar graph, scatter plot, histogram and multiple plots and also 2D and 3D plots. Today I enjoyed a lot with learning and bit confused and I will wait for the next session. Thank you

**Day 07 – Fri 28 May 2021**

Today’s Agenda –

1. Linear Regression using python
2. Implementation of Linear Regression, Sklearn Linear Regression

Reference Notes –

1. <https://www.goeduhub.com/753/supervised-learning-learning-different-supervised-learning>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about ML and I came across different types of ML. Today I have enjoyed a lot and I am waiting for another session

**Day 08 – Mon 31 May 2021**

Today’s Agenda –

1. Concepts of
   1. Gradient descent
   2. Cost function
   3. Mean square error
   4. Learning rate
   5. Save model using Joblib and Pickle

Reference Notes –

1. <https://www.goeduhub.com/10105/describe-gradient-descents-and-its-types>
2. <https://colab.research.google.com/drive/1YS72UJg_TqmugEiqyDvB9xx3xqPbKgnZ?usp=sharing>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about gradient descent and basic mathematics behind ML and various plots, mean square error. I enjoyed a lot today and I will wait for the next session. Thank you

**Day 09 – Tue 01 Jun 2021**

Today’s Agenda –

1. Multiple Linear Regression using python
2. Implementation of Multiple Linear Regression
3. Boston Housing Price Prediction

Reference Notes –

1. <https://www.goeduhub.com/756/predicting-boston-house-price-with-regression>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concepts of using multi linear regression and before that missingno. in pandas and the visualization of multi linear regression, Boston housing price prediction. I am waiting for another session. Thank you

**Day 10 – Thu 03 Jun 2021**

Today’s Agenda –

1. What is Classification?
2. Classification vs Regression
3. What is Regression?
4. Logistic Regression in Python
5. Logistic Regression vs Linear Regression
6. Logistic Regression and Iris flowers classification problem
7. Confusion matrix and accuracy

Reference Notes –

1. <https://www.goeduhub.com/758/logistic-regression-iris-flowers-classification-problem>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the difference between classification and regression and the contents of logistics regression and its difference with linear regression and using iris dataset, and confusion matrix and scaling data. I am waiting for another session. Thank you

**Day 11 – Fri 04 Jun 2021**

Today’s Agenda –

1. Dummy Variables
2. One hot Encoding
3. Introduction to Unsupervised Machine Learning
4. Introduction to K-Means

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the topics of dummy variables in data science and the advance concepts of clustering and the advanced concepts of encoding techniques... I will wait for the next session

**Day 12 – Mon 07 Jun 2021**

Today’s Agenda –

1. K-Means Implementation
2. Seaborn Library

Reference Notes –

1. <https://www.goeduhub.com/760/what-is-clustering-in-unsupervised-learning?show=760#q760>
2. <https://www.goeduhub.com/662/what-seaborn-library-python-data-science-seaborn-used-python?show=662#q662>
3. <https://www.goeduhub.com/9442/what-is-seaborn-heatmap?show=9442#q9442>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concepts of k means learning and elbow method and the extra library in data analysis which has seaborn and it is used for data visualization. I will wait for the next session. Thank you

**Day 13 – Tue 08 Jun 2021**

Today’s Agenda –

1. Topics of Decision Tree
2. Decision Tree Mathematical Intuition
3. Decision Tree Implementation

Reference Notes –

1. <https://www.goeduhub.com/3127/demonstrate-the-working-the-decision-tree-based-algorithm?show=3127#q3127>
2. <https://www.goeduhub.com/11055/decision-tree-on-iris-datasets-machine-learning>
3. <https://www.goeduhub.com/1173/what-is-decision-tree?show=1173#q1173>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the topics of decision tree and its mathematical intuition and its implementation. I will wait for the next session thank you

**Day 14 – Wed 09 Jun 2021**

Today’s Agenda –

1. Python Modules
2. Python Classes and Objects

Reference Notes –

1. <https://www.goeduhub.com/3706/python-modules-and-file-handling?show=3706#q3706>
2. <https://www.goeduhub.com/782/what-is-classes-and-objects-in-python?show=782#q782>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concepts of python such as creation of modules and importing of modules which as user defined and built-in and the concepts of OOPS such as classes and objects. I will wait for the next session thank you....

**Day 15 – Thu 10 Jun 2021**

Today’s Agenda –

1. Support Vector Machine
2. Implementation of Support Vector Machine
3. Python List Comprehension

Reference Notes –

1. <https://www.goeduhub.com/3487/demonstrate-and-implement-support-vector-machine?show=3487#q3487>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concepts of support vector machine and the implementation of support vector machine and the importance of supervised machine learning and the concepts of List Comprehension. I will wait for the next session thank you....

**Day 16 – Fri 11 Jun 2021**

Today’s Agenda –

1. What is Feature Engineering?
2. What is an Outlier?
3. Outlier Detection and removal using std deviation
4. What is Z score

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concepts of feature engineering and the techniques of feature learning like percentiles and z-score and the usage of outlier used for detecting and removing the outlier and the small concepts of correlation between features. Today I learnt a lot and I am waiting for the next session

**Day 17 – Mon 14 Jun 2021**

Today’s Agenda –

1. Exploratory Data Analysis

Reference Notes –

1. <https://www.goeduhub.com/11433/perform-exploratory-analysis-haberman-breast-cancer-dataset?show=11433#q11433>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the topics of Seaborn library and the understanding of plots such as boxplot, pair plot, etc., and the visualization of various methods. I will wait for the next session. Thank you

**Day 18 – Tue 15 Jun 2021**

Today’s Agenda –

1. KNN Algorithm – K-Nearest Neighbor Algorithm
2. Introduction to Web Scraping

Reference Notes –

1. <https://www.goeduhub.com/3237/demonstrate-and-implement-nearest-neighbor-knn-algorithm?show=3237#q3237>
2. <https://www.goeduhub.com/4747/what-is-web-scraping?show=4747#q4747>
3. <https://www.goeduhub.com/3237/demonstrate-and-implement-nearest-neighbor-knn-algorithm?show=3237#q3237>

Conclusion - GO\_STP\_8480 - Kowshik: I learnt about the topics of KNN algorithm and the importance of KNN algorithm and its implementation. I will wait for the next session. Thank you

**Day 19 – Wed 16 Jun 2021**

Today’s Agenda –

1. Naïve Bayes Classifier
   1. Introduction to Naïve bayes classifier
   2. Basics of probability
   3. Conditional probability
   4. Bayes theorem and implementation of Naïve bayes classifier
2. Demonstration and implementation of the Naïve Bayesian Classifier for Text Classification

Reference Notes –

1. <https://www.goeduhub.com/3426/demonstrate-and-implement-the-naive-bayesian-classifier?show=3426#q3426>
2. <https://www.goeduhub.com/3188/demonstrate-implement-bayesian-classifier-classification>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the topics of naive Bayes classification and the types of naive Bayes and the importance of probability in AI and the concepts of probability such as the definition and the importance of probability and conditional probability along with the implementation of naive Bayes. I will wait for the next session thank you....

**Day 20 – Thu 17 Jun 2021**

Today’s Agenda –

1. Python Web Scraping

Reference Notes –

1. <https://www.goeduhub.com/11393/example-scraping-flipkart-scraping-python-beautifulsoup?show=11393#q11393>
2. <https://www.goeduhub.com/4747/what-is-web-scraping?show=4747#q4747>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the topics of web scraping and the usage of web scraping and its implementation. Web scraping from Flipkart webpage and loading the data using beautiful soup and the using the code of python. I will wait for the next session thank you

**Day 21 – Fri 18 Jun 2021**

Today’s Agenda –

1. Introduction to Flask Web Framework

Reference Notes –

1. <https://www.goeduhub.com/11441/develop-a-basic-crud-application-tutorial-on-flask?show=11441#q11441>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concept of flask and the implementation of flask framework using python and this framework is mostly used in web development. Thank you and I will wait for the next session.

**Day 22 – Mon 21 Jun 2021**

Today’s Agenda –

1. Building URL Dynamically in Flask Web Framework
2. Integrating HTML, CSS and JavaScript with Flask Web Framework
3. Understanding Jinja2 Template Engine in Flask Web Framework

Reference Notes –

1. <https://github.com/shardagodara/Flask-Application>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the topics of flask web framework implementation and the basics of HTML, CSS and the creation of webpage using HTML, CSS and the framework using Flask. Thanks for it and I will wait for the next session.

**Day 23 – Wed 23 Jun 2021**

Today’s Agenda –

1. ML Model Deployment with Flask on Heroku
2. How to Deploy Machine Learning Model with Flask

Reference Notes –

1. <https://github.com/shardagodara/Heroku-class>

Conclusion – GO\_STP\_8480 - Kowshik: Today I have learnt ML model and the deployment in Heroku and the activation and sign-up account in Heroku and the deployment of ML model using Flask in Heroku platform. Thanks for today and I will wait for the next session

**Day 24 – Thu 24 Jun 2021**

Today’s Agenda –

1. GUI and TKinter in python!
2. Create a Calculator using Python GUI (TKinter)!
3. How to run Flask App on Google Colab

Reference Notes –

1. <https://www.goeduhub.com/1206/gui-and-tkinter-in-python?show=1206#q1206>
2. <https://www.goeduhub.com/2706/create-a-calculator-using-python-gui-tkinter?show=2706#q2706>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the advanced topics of python library named as TKinter and the explanation of GUI and the development of calculator project using TKinter and the deployment of flask model in Google Colab. Thanks for today class and I will wait for the next session.

**Day 25 – Fri 25 Jun 2021**

Today’s Agenda –

1. End to End Machine Learning Project – Used Car Price Prediction

Reference Notes - <https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho>

Conclusion - GO\_STP\_8480 - Kowshik: Today we discussed about the project implementation an end-to-end project of designing car price prediction using the source of ML algorithm and using Kaggle dataset the design and implementation of car price prediction is done successfully and discussed about the importance of features. Thanks for today session and I will wait for next session.

**Day 26 – Mon 28 Jun 2021**

Today’s Agenda –

1. Introduction to random forest
2. Hyper Parameter tuning

Reference Notes – <https://www.goeduhub.com/11098/random-forest-algorithm-on-iris-datasets-machine-learning?show=11098#q11098>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concept of forest algorithms and the implementation of forest algorithms and understanding the concepts of hyper-parameters and the explanation of RandomSeachCV and GridSearchCV in the hyper-parameters, and the deployment of ML model in Heroku. Thanks for today, and I will wait for the next session.

**Day 27 – Tue 29 Jun 2021**

Today’s Agenda –

1. End to End Machine Learning Project – Customer Segmentation using K-Mean Clustering

Reference Notes - <https://www.kaggle.com/arjunbhasin2013/ccdata>

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concepts of PCA (Principal component analysis) and the development of another end-to-end project named as customer segmentation and it is developed using K-means clustering. Thanks for today and I will wait for another session.

**Day 28 – Wd 30 Jun 2021**

Today’s Agenda –

1. Deploy Machine Learning Model into AWS Cloud Server EC2 Instance

Conclusion - GO\_STP\_8480 - Kowshik: Today I have learnt about the concepts of deployment of machine learning model in Amazon EC2, and the introduction of Amazon cloud services and the usage of EC2 instance in cloud services. Thanks for today and I will wait for the next session.

**Day 29 – Thu 01 Jul 2021**

Today’s Agenda –

1. Time Series Forecasting using ARIMA and SARIMA

Reference Notes –

1. <https://www.goeduhub.com/11397/what-is-hypothesis-testing-in-statistics-machine-learning>
2. <https://www.goeduhub.com/11377/correlation-covariance-statistics-machine-learning-science>

Conclusion - GO\_STP\_8480 - Kowshik: Today we learnt and understood the concepts of time series and the introduction of hypotheses and P value and developed a time series forecasting using the tools of ARIMA and SARIMA. Thanks for today and I will wait for the next session.

**Day 30 – Fri 02 Jul 2021**

Today’s Agenda –

1. Interview Questions on AI and ML

Reference Notes –

1. <https://www.goeduhub.com/11194/machine-learning-interview-question-answers-science-part1?show=11194#q11194>
2. <https://www.goeduhub.com/11274/asked-python-interview-questions-detailed-explained-answers>
3. <https://www.goeduhub.com/11363/machine-learning-interview-question-answers-science-part2?show=11363#q11363>

Conclusion – GO\_STP\_8480 - Kowshik: In today's class I learnt about the interview Q&A on ML and having a clarity and sharing the details on ML job related and some extra tips to crack job interviews and a few resume tips and the enrichment our career growth. Thank you for this course and I have enriched myself a lot during the course. Thanks for everything ma'am.

**45 Days Internship on Data Science and Machine Learning with Python - Complete Training starts from 20-05-2021 to 03-07-2021**

**Complete Set of Questions GOEDUHUB\_TECHNOLOGIES**

**Day 1**  
  
👉What is Machine Learning?  
👉What is Data Science?  
👉What is Deep Learning?  
👉Why Python in Machine Learning?  
👉What is Google Colab?  
👉Describe Life Cycle of Data Science Project?  
👉What are the Different Types of Machine Learning?  
  
GO\_STP\_8480 - Kowshik:  
  
1. Machine Learning - Visual Analysis  
2. Data Science - Knowledge of Data and we are generating the data. It represents the processing of Data  
3.  Deep Learning - Mimic human brain. It consists of neural networks ANN, CNN, RNN  
4. Python is simple to understand and easy to learn and Python is a general-purpose language, it can do a set of complex machine learning tasks.  
5. Google colab is used to execute the python code and it just acts like an interpreter and Google colab is executed only in online mode.  
6. DS cycle  
1. Business Understanding - what is data  
2. Data Collection - Kaggle  
3. Data Preparation - handling missing data  
4. Exploratory Data Analysis - Summarization of clean data  
5. Modeling - Heart of data Analysis, predictions  
6. Model evaluation Tune model  
7. Model Deployment - Using Flask Clouds  
8. Different types of ml  
9. Supervised learning  
10. Unsupervised learning  
11.Reinforced learning

**Day 2**  
  
👉What are an immutable and mutable object in Python?  
👉What is a loop statement in Python?  
👉What is indentation in Python?  
👉How to check multiple conditions in Python?  
👉What is Pass Statement in Python  
👉What are the control statements in Python?  
  
GO\_STP\_8480 - Kowshik:  
  
1. Mutable objects can be changeable and immutable objects are unchangeable like, immutable objects like str and mutable objects are list  
2.  A loop statement is used for an iteration based on a specific condition and these is used for certain times. It is simple to say as the combination of multiple conditional statements.  
3. Indentation refers to a tab space either 8 or 4 spaces  
4. Using if...elif…else statements  
5. Pass statement refers to keep the code as for future updating and whenever it overcomes it simply moves to the next line. It's simply acts as similar as comment in python  
6. Control statements can be executed based on the specific condition like if, else, break, continue

**Day 4**  
  
👉 Difference between list and tuple?  
👉What are an immutable and mutable object in Python?  
👉 How to access items in a dictionary python?  
👉 What is set and why use?  
👉 What is frozen set?  
👉 Difference between append and extend?  
👉 Difference between indexing and Slicing?  
👉 How to find the largest and lowest value in the list?  
  
GO\_STP\_8480 - Kowshik:  
  
1. Lists are mutable and whereas, tuples are immutable and cannot be modified using indexing  
2. Mutable objects are lists, sets, dictionaries and whereas immutable objects are string, tuples, float, int  
3. Using keys, we can access dictionaries i.e., dictionary values. To access items we can use key, value pair in an iterator.  
for key in dic.items() --> It returns key value pairs in form of tuples.  
4. Set is a unique collection of different data types of values. A set is a collection of both unordered and un-indexed  
5. frozenset([iterable]). Frozen set takes an iterable object as an input parameter such as list and dict which results an immutable object as output.  
6. append is used to add a list item in a list L whereas extend is used to concatenate two lists L1 and L2 in the L1 list. append takes only one argument and it has one data-type whereas extend takes multiple value as a different datatype  
7. Indexing is used to access the data in a given datatype such as list, str whereas slicing is used to access a subsequence using start and end indexes of a lust or str  
8. Largest - max(L) whereas smallest - min(L)

**Day 5**  
  
👉 What is NumPy?  
👉 Why NumPy is used in Python?  
👉 Where is NumPy used?  
👉 Difference between numpy and Python List  
👉 How to convert 1D Numpy array to 2D Numpy array  
👉 How to find the memory size of any array?  
👉 What is use of Pandas?  
👉 What is Series in Pandas?  
  
GO\_STP\_8480 - Kowshik:  
  
1. numpy is a multi-dimensional array i.e., It is an n dimensional array and it consists of various methods to modify the state.  
2. numpy can be a best substitute for python lists, where in numpy there are various methods to control the numpy arrays and these are fast in execution.  
3. numpy can be used in advanced concepts such as data science, Machine Learning, statistics, Linear Algebra, Advanced Calculus  
4. numpy is far better than lists in python. Where, in numpy the indexing and accessing is fast and it takes less memory to allocate.  
5. reshape() method can be used to convert from 1D numpy array to 2D numpy array  
6. size() method can be used to determine the size of any dimensional array  
7. Pandas can be used for data analysis, data manipulation, data cleaning, tabular formation, and statistical data  
8. series is a one-dimensional array i.e., It is a list in pandas. Where it can support int, str, float.

**Day 6**  
  
👉 Define Python Pandas  
👉 What Are The Different Types Of Data Structures In Pandas?  
👉 Define Dataframe In Pandas  
👉 How Can You Create An Empty Dataframe In Pandas?  
👉 What Are The Most Important Features Of The Pandas Library?  
👉 What Is Group by Function In Pandas?  
  
  
GO\_STP\_8480 - Kowshik:  
  
1. Pandas is a open module which consists of several packages used for the analysis of data and also for manipulation of data  
2. Series, Data-frames, Panel  
3. Data-frames is a collection of a 2D array similar as 2D list in python and it is a size mutable with structured tabular form consists of rows and columns  
4. dataframe = pandas.DataFrame() # which has empty columns  
5. Features - Renaming, Reshaping, plotting, grouping  
6. pd.groupby() used to split the data based on some conditions and various criteria.

**Day 7**  
  
👉 What is Matplotlib?  
👉 Why we need data visualization?  
👉 What do you mean by xlabel(), ylabel() and title() in Matplotlib?  
👉 How to create Legend in Matplotlib?  
👉 What do you mean by bar() function, xticks() function, barh()function?  
👉 what is Histograms?  
  
  
GO\_STP\_8480 - Kowshik:  
  
1. Matplotlib is a python library which is used for plotting the graphs i.e., To visualize the data  
2. To understand the topic and to predict the data for the next instance and to correct the data of any error occurs in it by changing the function.  
3. xlabel() is used for heading in x-axis or horizontal axis and ylabel() is used for heading in y-axis or vertical axis and title() is used for heading of the plot  
4. pyplot.legend(x,y,label=["Numbers"],["blue"],loc="lower left")  
5. bar() function is used to plot bar graph and xticks() is used for plotting the bar graph and barh() is used to plot horizontal bars in plotting  
6. Histograms are used to represent the data in both horizontal and vertical axis. These are continuous or cumulative data

**Day 8**  
  
👉 What is Supervised Machine Learning  
👉 What is Linear Regression?  
👉 What does train\_test\_split function?  
👉 What is test\_size and random\_state?  
👉 How does random\_state manipulates machine learning model?  
👉 What is scikit Learn?  
  
  
GO\_STP\_8480 - Kowshik:

1. Supervised ML is a subcategory of machine learning and it is defined by its use of label-ed datasets to train algorithms that to classify data or predict outcomes accurately.  
2. Linear regression is basically depending on supervised learning, where the predicted output is continuous and it has constant slope. It's used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories (e.g. cat, dog).  
3. train\_test\_split function has training data and testing data and it is a Sklearn model for splitting data arrays into two subsets.  
4. test\_size specifies the size of the testing dataset. For default training size we have 0.25 test\_size and random\_state determines the randomness of splits.  
5. random\_state parameter is used for initializing the internal random number generator, which will decide the splitting of data into train and test indices. It uses the same range of data every time we run it so that we will get the same answer every time.  
6. Scikit-learn is the most important library for ML in python. The Sklearn library contains a lot of efficient tools for ML and statistical modeling includes classification, clustering, dimensionality, and reduction.

**Day 9**  
  
👉What is MSE?  
👉What is gradient descent?  
👉What is joblib?  
👉What is pickle?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. MSE -Mean Square Error is used to calculate the error for both predicted and actual values. ie., It is an average square difference between the actual and predicted values.  
2. Gradient descent is used to minimise the linear line equation y=ax+b where to minimise both coefficient and an intercept to provide a minimised linear equation.  
3. Joblib is a python library used to avoid the usage of repetitive functions. It provides a lightweight pipelining.  
4. Pickle is a replacement of joblib. Compared to pickle joblib is more faster and better to use rather than numpy arrays

**Day 10**  
  
👉 What are the parameters of pandas fillna function?  
👉 How does the dropna ( ) function in pandas work?  
👉 What is Multi Linear Regression?  
👉 How to drop a column in Pandas?  
👉 What does random\_number in data split?  
  
GO\_STP\_8480 - Kowshik :  
  
1. fillna() function is used to fill nan values not a number in a given dataset.  
2. dropna() is used to drop the nan values in a given dataset  
3. Multi linear regression is similar to linear regression but in mult linear regression it contains various coefficients and with various independent variables such as multi features for a single label value.  
4. To drop the column - drop(value, axis=1) results the deletion of column from dataset.  
5. random\_number in data split can be done by using the random\_state in train\_test\_split. Where it is none it is randomly initialised.

**Day 11**  
  
👉 What is Classification?  
👉 What is Regression?  
👉 What is the difference between logistic regression and linear regression?  
👉 What is Confusion Matrix in Machine Learning?  
👉 Which is a binary problem in logistic regression?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. Classification - It is a process bof categorising the data into classes and this can be performed on both structured and unstructured data and it refers to predict the modeling problem  
2. Regression - To analyse the relationship between dependent and independent variables such as label and features. Types of regression is linear, multi linear, logistics  
3. Linear regression has a linear function y=ax+b which states that the prediction can be done in linear whereas, in logistic regression the prediction may be logarithmic log(p/1-p) and this can be evaluated using conditional probability.  
4. Confusion matrix - It is a classification model of order n x n where, this matrix is a table used to describe the performance of classification model and it is relatively simple to understand.  
5. Logistic regression is basically a relationship between dependent and independent variables and the inputs are with 2 levels categorical and the output with 2 or more values. It can be done on conditional probability either 0 or 2 classification.

**Day 12**  
👉What is pandas dummy variable?  
👉What is the difference between OneHotEncoder and Get\_dummies?  
👉What is Sklearn LabelEncoder?  
👉What is the difference between LabelEncoder and Get\_dummies?  
👉What is unsupervised learning?  
👉Where is unsupervised learning used?  
👉What is unsupervised learning example?  
👉What is the difference between supervised and unsupervised learning?  
👉What is meant by K-means clustering?  
👉What are the basic steps for K-means clustering?  
  
GO\_STP\_8480 - Kowshik :  
1. Pandas in dummy variable is nothing but a binary variable and these are used in multiple regression  
2. One hot encoding represents the categorical data in more expressive and these data is a combination of only 0s and 1s ie., Binary data. Whereas Get\_dummies is used to convert categorical data into dummy data  
3. Label encoder is used to convert the labels into numeric form ie., Machine readable format  
4. LabelEncoder is used to normalise the labels and converts non numerical data into numerical data whereas, Get\_Dummies is used to convert categorical data into dummy data  
5. Unsupervised learning is a machine learning technique, ie., We do not need to supervise the model. It helps you all kinds of unknown patterns  
6. It is used in clustering and anomaly detection.  
7. Used in genetics and clustering the DNA samples.  
8. Supervised learning is used to collect the data and train the model consists of data and tests the model. Whereas unsupervised learning need not to supervise the model ie., No training and no collection of data.  
9. It is a type of unsupervised learning which you have a un-labeled data that aims to partition of n observations with k clusters and each cluster is equals to nearest mean.  
10. Select the cluster centers and calculate the distance between each point and cluster centers and assign the data point to the cluster center and distance should be minimum and recalculate the new cluster with value ci represents ith cluster and recalculate the distance between each data point and obtain new clusters again. If no data was reassigned then, stop otherwise start again with the new cluster center.

**Day 13**  
  
👉 What are the most practical applications of k-means?  
👉 What is elbow method?  
👉 How is the elbow method used in clustering?  
👉 How to plot WCSS against an increasing k?  
👉 What is Seaborn used for?  
👉 What is difference between Matplotlib and Seaborn?  
👉 What is hue in Seaborn plots?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. It can be used in various applications document clustering, identifying crime prone areas, insurance fraud detection, and it is used in image technology such as image compression and image segmentation  
2. Selects the optimal number of clusters by fitting the model with the range of values of k. By default the distortion score is computed  
3. We can calculate wcss where, it means within cluster sum of squares. Wcss is the sum of squares of distances of each data point in all clusters to their respective centroids to reduce the sum.  
5. Seaborn is a python module and it is mostly used in data analysis, for visualisation of data and it is a high level interface for drawing attract and informative stastical graphs  
6. 6. Matplotlib and Seaborn is used for data visualization. But, matplotlib is used for making the basic graphs and Seaborn is usd for data visualization with various colors and themes.  
7. hue is a parameter used in Seaborn which is used to for color encoding.

**Day 14**  
  
  
👉 What is Decision tree in Machine Learning  with example?  
👉 Why  Decision tree is used in Machine Learning?  
👉 What is entropy in decision tree?  
👉 What is information gain and Gini index?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. Decision tree is basically tree shaped diagram, which shows graphical representation of a problem with all the possible solution to a decision. Example of using decision tree regression problems, classification or logistic regression problems can be solved.  
2. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).  
3. Entropy:-Entropy is a measure of the randomness or unpredictability in the data-set. Entropy basically measure error of the model.  
4. Information Gain:- The information gain is based on the decrease in entropy after a data-set is split on an attribute.  
5. Gini index:- is a measure of statistical dispersion intended to represent the income inequality or wealth inequality within a nation or any other group of people. Gini\_index can be also called as Gini coefficient or Gini ratio

**Day 15**  
  
👉  What is module in Python with example?  
👉  What is difference between module package and library?  
👉  How do I create a Python module?  
👉  What is class and object in Python?  
👉  What is Self In \_\_ init \_\_ Python?  
  
GO\_STP\_8480 - Kowshik :  
  
1. A module allows you to logically organize your Python code and  can define functions, classes and variables. A module can also include a runnable code.  
2. A module is a collection of various functions and the python code whereas a package is a collection of modules, and library is a collection of packages  
3. Create a module named as example.py and store it in a package as example and this package is stored in local system path of python modules and you can use this module by simply writing this >>> from example import example.py  
4. A class in python is the blueprint or template  from which specific objects are created. Whereas, The object is an entity that has a state and behavior associated with it.  
5. using \_\_inti\_\_ method we can create constructor. Whereas self is used to reference the class to object parameters

**Day 16**  
  
👉 What is support vector machines with example?  
👉 What does Hyperplane in SVM mean?  
👉 Write some Support Vector Machine (SVM) Usecases  
👉 What is kernel in SVC?  
👉 What is List comprehension?  
👉 When do you need list comprehension in Python?  
👉 How does the condition work in Python list comprehension?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. Support vector machine is supervised learning algorithm for classification problems. Example ::> Support Vector Machine draws a hyperplane in n dimensional space such that it maximizes the margin between classification groups.  
2. hyperplane is a flat affine subspace of dimension N-1. Visually, in a 2D space and maximal margin hyperplane, so this means that our optimal hyperplane will be the one who has  maximum margin.  
3. It uses in face detection, handwriting recognition, and hypertext categorisation, bio informatics.  
4. kernel function is used to convert data from low dimensions to high dimensions.  
5. List comprehension in Python is an easy and compact syntax for creating a list from a string or another list. List comprehension is a compact version of using iterator ie., for loop in python  
6. List comprehension is a statement used for creating lists and returns the lists.  
7. Condition ::> [x if x>y else y]

**Day 17**  
  
👉 What is a Feature?  
👉 What is feature Engineering?  
👉 What is an outlier in machine learning?  
👉 How does machine learning deal with outliers?  
👉 What is a heat map used for?  
👉 What is Z score in data science?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. Feature is an independent variable used to predict and analysis the data.  
2. Feature engineering is the process of using domain knowledge to extract features from raw data. A feature is a property shared by independent units on which analysis or prediction is to be done.  
3. An outlier is an object that deviates significantly from the rest of the objects. They can be caused by measurement or execution error. The analysis of outlier data is referred to as outlier analysis or outlier mining.  
4. Outliers deals with ML  
- Univariate method. One of the simplest methods for detecting outliers is the use of box plots.  
- Multivariate method. Outliers do not need to be extreme values.- Minkowski error.  
5. A heat map is a data visualization technique that shows magnitude of a phenomenon as color in two dimensions. Similar to matplotlib but in advanced  
6. z-score is is z = (x-μ)/σ, where x is the raw score, μ is the population mean, and σ is the population standard deviation

**Day 18**  
  
👉 What is Exploratory Data Analysis (EDA) ?  
👉 Why do we need exploratory data analysis ( EDA )?  
👉 What are the steps in exploratory data analysis ( EDA )?  
👉 What Is Bivariate Analysis?  
👉 What Is Univariate Analysis?  
👉 What is box plots in seaborn?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. Exploratory Data Analysis is the process of exploring data and deriving some meaningful conclusions from it like finding patterns, hypothesis testing, dimensionality reduction, handling missing values and many more.  
2. EDA in Python uses data visualization to draw meaningful patterns and insights.  
3.  
- Description of data.  
- Handling missing data.  
- Handling outliers.  
- Understanding relationships and new insights through plots.  
4. Analysis of two variable is known as bivariate analysis. We can plot 2D scatter plots for plotting graph between two variables.  
5. Analyze each variable separately. Univariate analysis can be done using the following graphs.  
6. Box plot - depicts lower to upper quartile values of the data, with a line at the median.

**Day 19**  
  
👉 What is K-fold cross validation?  
👉 How K-fold cross validation is implemented?  
👉 What is K Nearest Neighbor machine learning?  
👉 Why KNN is called lazy?  
👉 What is wen scraping?  
👉 Why python is good for web scraping?  
👉 Why is Web scraping used?  
👉 How do you scrape data from a website?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. It splits the data into k number of sets or sections and each section is uses for testing the data  
2. It is implemented by dividing the sections ie., Dividing the set of observations into k groups in equal size.  
3. KNN is simple supervised learning algorithm used for both regression and classification problems. KNN is  basically store all available cases and classify new cases based on similarities with stored cases.  
4. KNN is simple algorithm for classification but that's not the reason and KNN is lazy learner  because it doesn't learn a discriminative function from the training data but memorizes the training dataset instead.  
5. Extracting the data (Basically Html Files) of any website is called web scraping.This saves us a lot of time because we get a lot of data (huge amount of data) in structured form  
6. Python has a huge number of libraries and it has a beautiful soup and socket libraries used to extract the data in web pages.  
7. web scraping simplifies the process of extracting data, speeds it up by automating it and creates easy access to the scrapped data by providing it in a CSV format. In simple terms, web scraping saves you the trouble of manually downloading or copying any data and automates the whole process.  
8. Using beautiful soup and to scrap the data we use a url of the website and write a simple code on using the library beautiful soup. Run the code and extract the data.

**Day 20**  
  
  
👉 What is naive Bayes classifier in machine learning?  
👉 What are the different types of naive Bayes classifier?  
👉 Which type of naive Bayes classifier is best suited for document classification problem?  
👉 How do you use naive Bayes?  
👉 Why do we use naive Bayes classifier?  
👉 What is CountVectorizer in machine learning?  
👉 What is Tfidftransformer used for?  
  
GO\_STP\_8480 - Kowshik :  
  
1. It is supervised learning algorithm used for classification based on Bayes' Theorem.Naive Bayes Classifier is not just an algorithm, but a collection of many algorithms that work on the same concept, the Bayes' Theorem  
2. There are three type of naive bayes classifier - Gaussian - Multinomial - Bernoulli  
3. Multinomial is the best one...  
4. Import the data and make a counting of words and train the model and as well as test the model and get the accuracy.  
5. Usage of Naive Bayes Classifier -  
6. News CategorizationSpam filtering Object and face recognition.Medical Diagnosis Weather Prediction etc..  
6. It removes the punctuation marks and converts all the words to lowercase.  
7. TfidfTransformer Basically do is weight the words in the classification. For example we get "the" word multiple times in any text, So TfidfTransformer  finds out how much its contribution to the model is in the classification.

**Day 21**  
  
👉 What is Webscraping used for?  
👉 How do I use BeautifulSoup for web scraping?  
👉 What’s the best programming language for web scraping?  
👉 How does Deep Learning differ from Machine Learning?  
  
GO\_STP\_8480 - Kowshik :  
  
1. Web scraping is used for Extracting the data (Basically Html Files) of any website is called web scraping.This saves us a lot of time because we get a lot of data (huge amount of data) in structured form  
2. Use of Beautiful soup to parse the HTML content with the help of html parser.  
3. from bs4 import BeautifulSoup  
soup = BeautifulSoup(htmlcontent,"html.parser")  
3. Python is the best programming language for web scraping.  
4. Machine Learning is a superset of Deep Learning and the Deep Learning is a subset of Machine Learning.  
The data represented in Machine Learning is quite different as compared to Deep Learning as it uses structured data. The data representation is used in Deep Learning is quite different as it uses neural networks(ANN).

**Day 22**  
  
👉 What is flask framework used for?  
👉 How does a flask framework work?  
👉 What is WSGI in flask?  
👉 What is Jinja2 in flask?  
👉 What is static folder in flask?  
👉 How do I use a template in flask?  
👉 What is Route () in flask?  
  
GO\_STP\_8480 - Kowshik :  
  
1. Flask framework is used for web development and this can be implemented using with python  
2. Working of flask frameworkfrom flask import Flaskapp = Flask(\_\_name\_\_) @app.route("/")  
3. Flask is a lightweight WSGI (web server gateway interface), enabling developers to create web applications.  
4. Jinja2 is a modern day templating language for Python developers. It was made after Django's template. It is used to create HTML, XML or other markup formats that are returned to the user via an HTTP request.  
5. The static and templates directories should be present with .py/.ipynb file, not anywhere else. Inside static we can have audio/video/image files.  
6. We can render html pages using return function. It will render the home.html page from templates folder.  
7. @app.route("/about")def about():    return render\_template('about.html')  
8. route("/") is a Python decorator that Flask provides to assign URLs in our app to functions easily.

**Day 23**  
  
👉 What is the difference between Render\_template and redirect?  
👉 What is Url\_for in flask?  
👉 Is flask a Web server?  
👉 How to use css in Python flask?  
👉 What are the delimiters used in Jinja2 template?  
👉 How do you create a variable in Jinja?  
  
GO\_STP\_8480 - Kowshik :  
  
1. We can render html pages using return function. It will render the home.html page from templates folder.  
2. Flask class has a redirect() function. When called, it returns a response object and redirects the user to another target location with specified status code. location parameter is the URL where response should be redirected.  
3. def render(self):        return render\_template(user/general\_settings.html, form=self.form)    def redirect(self):        return redirect(url\_for(user.settings))  
2. If your application is placed outside the URL root, for example, in /myapplication instead of /, url\_for() properly handles that for you.  
3. @app.route(/)def index():    # ...    return redirect(url\_for(nextPage, id=DBTable.id))@app.route(/)def nextPage(id):    # ...    return render\_template(page2.html)  
3. Yes, Flask has a built-in web server  
4. To link css in python flask    Flask app       
5. Various delimiters in Jinja2 template{% ... %} for Statements{{ ... }} for Expressions to print to the template output{# ... #} for Comments not included in the template output#  ... ## for Line Statements  
6. Using {% set %} we can create a variable in Jinja2  
7. Ex:: {% set label\_cls, field\_cls = col-md-7, col-md-3 %}

**Day 24**  
  
👉 What is model deployment in ML?  
👉 Why do people deploy ML models?  
👉 How does Heroku deploy machine learning model?  
👉 What is heroku Gunicorn?  
👉 What is Procfile?  
👉 How do I deploy flask app to Heroku?  
  
GO\_STP\_8480 - Kowshik :  
  
1. Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data. It is one of the last stages in the machine learning life cycle and can be one of the most cumbersome.  
2. Deploying a machine learning model, known as model deployment, simply means to integrate a machine learning model and integrate it into an existing production environment where it can take in an input and return an output.  
3. We created a machine learning model, trained it, created a web application to predict new data using the model and deployed it on the internet using heroku. And did all of it in python!  
4. Gunicorn is a pure-Python HTTP server for WSGI applications. It allows you to run any Python application concurrently by running multiple Python processes within a single dyno.  
5. Install gunicorn  
6. 5. Procfile is a mechanism for declaring what commands are run by your applications dynos on the Heroku platform.  
7. A Procfile declares its process types on individual lines, each with the following format: :   is an alphanumeric name for your command  
6. Process -  
$ pipenv install flask gunicorn  
$ touch Procfile  
$ touch runtime.txt  
$ mkdir app$ cd apptouch main.py  
$ cd ../$ touch wsgi.py  
$ pipenv shell  
$ git init $ git add .  
$ git commit -m Initial Commit  
heroku login  
$ heroku create eflask-app  
$ git push heroku master

**Day 25**  
  
👉 What is a Graphical User Interface (GUI)?  
👉 What are different Python Libraries to create GUIs?  
👉 What is Tkinter?  
👉 What are Tkinter Widgets?  
👉 How to create a messagebox with Tkinter in Python?  
👉 Where do I find the combobox widget In Tkinter?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. GUI is desktop application which helps a human to interact with computers, electronic devices,They are used to perform different tasks in the desktops, laptops and other electronic devices.ExamplesText-Editors (Notepad etc)Games- Sudoku , ChessBrowser- Google Chrome, Firefox  
2. Python Libraries To Create Graphical User InterfacesTkinter, Kivy, wxpython, QT python, Jpyhton  
3. Python has inbuilt Tkinter library for creating the GUI  desktop based applications.- Steps followed to create a tkinter GUI applicationImport the tkinter module.Create a main window Add widgets Enter the main event loop.  
4. There are a number of widgets.Tkinter provides various controls, used in a GUI application, Known as widgets.Button - To add various kinds of buttons in GUI appCanvas - This widget is used to shapes in GUILabel - used to display some message or information about the other widgetsEntry - used to display the single-line text field to the userFrame - Frame is used as container Menu - used to add menu items to the userMessage - Display message box to the userScrollbar - Used to scroll window up and down Text - The Text widget is used to display text in multiple linesSpinbox - It same as scrollbar but has two tabs one is and one is down  
5. Message box - messagebox.Function\_Name(title, message [, options])  
Example -  
name = Label(var,text = "Name")name.grid(row = 0, column = 0) e1 = Entry(var).grid(row = 0, column = 1)  password = Label(var,text = "Password").grid(row = 1, column = 0)  e2 = Entry(var).grid(row = 1, column = 1)   
submit = Button(var, text = "Submit", fg = "red").grid(row = 4, column=0)   
6. Combo box - combobox = ttk.Combobox(master, option=value, ...)  
Example -  
n = tk.StringVar()monthchoosen = ttk.Combobox(window, width = 27, textvariable = n)

**Day 26**  
  
👉 How do you make an end to end machine learning project?  
👉 What is feature importance in decision tree?  
👉 How do you deploy a ML project?  
👉 Write your machine learning Project ideas.  
  
GO\_STP\_8480 - Kowshik :  
  
1. Making an end to end machine learning project - can be done by using this following steps --->  
1) Pre-requisites and Resources  
2) Data Collection and Problem Statement  
3) Exploratory Data Analysis with Pandas and NumPy  
4) Selecting and Training Machine Learning Models  
5) Cross-Validation and Hyperparameter Tuning using Sklearn  
6) Saving the Trained Model  
7) Developing a web service.  
2. Decision trees are an important tool for decision making and risk analysis, and are usually represented in the form of a graph or list of rules.  
  
One of the most important features of decision trees is the ease of their application. Being visual in nature, they are readily comprehensible and applicable. Even if users are not familiar with the way that a decision tree is constructed, they can still successfully implement it.  
  
Most often decision trees are used to predict future scenarios, based on previous experience, and to support rational decision making.  
  
3. Deployment of ML project can be done - Mostly, ML projects can be hosted in websites using FTP, HTTP or HTTPS protocols and using some hosting websites such as Heroku, 000Webhost, hostinger, and GoDaddy.  
  
ML project can be deployed on any clouds. The most common way is using HTTP calls. Therefore you can deploy your machine learning model with a supported block of code for execution on the google cloud function and call the HTTP request for prediction from your web application or any other system.  
4. I have an idea to implement the detection of COVID and analysis of the spreading of COVID across all over India and in comparison with world and the prediction of COVID19 for the next 15 days. These can be implemented by using some key libraries in python such as matplotlib, Seaborn, prophet, and using pandas, and numpy to analyze the data. This is all about the project.

**Day 27**  
  
👉 What are the Hyperparameters in machine learning?  
👉 Why Hyperparameter tuning is important?  
👉 Which method is used for Hyperparameter tuning?  
👉 What is the difference between GridSearchCV and RandomSearchCV?  
👉 How do you define a random forest classifier?  
👉 Why do we use random forest classifier?  
  
  
GO\_STP\_8480 - Kowshik :  
  
1. A model hyperparameter is a configuration that is external to the model and whose value cannot be estimated from data.- They are often used in processes to help estimate model parameters.- They are often specified by the practitioner.- They can often be set using heuristics.- They are often tuned for a given predictive modeling problem.  
In machine learning, a hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are derived via training.  
  
2. Model parameters are learnt as part of training process, whereas the values of hyper-parameters are set before running the training job and they do not change during the training.  
3. Hyperparameters are crucial as they control the overall behaviour of a machine learning model. The ultimate goal is to find an optimal combination of hyper-parameters that minimizes a predefined loss function to give better results.  
3. By  using a exhaustive grid search from scikit  learn.  Using this method we can find the best set of values in the parameters search space.  
4. GridSearchCV implements the most obvious way of finding an optimal value for anything — it simply tries all the possible values (that you pass) one at a time and returns which one yielded the best model results, based on the scoring that you want, such as accuracy on the test set.  
RandomSearchCV has the same purpose of GridSearchCV: they both were designed to find the best parameters to improve your model. However, here not all parameters are tested. Rather, the search is randomized and all the other parameters are held constant while the parameters we are testing is variable.  
5. Random forest is an ensemble and supervised  machine learning algorithm which is capable of performing both regression and classification problems.  
Random forest consists of many decision trees. It is kind of forming forest of trees. Means random forest combine multiple same type of decision tree algorithm to form a random forest algorithm.  
  
Defining of random classifier - from sklearn.ensemble import RandomForestClassifier  
  
clfr= RandomForestClassifier(random\_state = 100)  
  
6. Random Forest works- Select N number of samples from datasets.- Build Decision for each sample and predict the result for each decision tree.- Voting is done for each predicted result of decision tree.- The prediction result with majority votes win.  
It can be used for both regression and classification tasks, and it's also easy to view the relative importance it assigns to the input features. If there are enough trees in the forest, the classifier won't overfit the model.

**Day 28**  
  
👉 What is Distplot Seaborn?  
👉 What is PCA in machine learning?  
👉 How do you reverse the Sklearn scale?  
👉 What is customer segmentation in machine learning?  
👉 How is segmentation done in machine learning?  
  
GO\_STP\_8480 - Kowshik :  
  
1. It is used basically for uni-variant set of observations and visualizes it through a histogram i.e. only one observation and hence we choose one particular column of the dataset.Syntax: distplot(a[, bins, hist, kde, rug, fit, ...])  
2. Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation which converts a set of correlated variables to a set of uncorrelated variables. PCA is a most widely used tool in exploratory data analysis and in machine learning for predictive models.  
3. x = scaler.inverse\_transform(x)  
4. Customer segmentation is the practice of dividing a company’s customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business.  
Before we get into the process, I will give you a brief on what kind of steps we will get.- Gather the data- Create Recency Frequency Monetary (RFM) table- Manage skewness and scale each variable- Explore the data- Cluster the data- Interpret the result  
5. Steps to follow -  
Step 1: Create a business case  
Step 2: Prepare the data  
Step 3: Use K-means clustering  
Step 4: Choosing optimal hyper-parameters  
Step 5: Visualization and interpretation

**Day 29**  
  
👉 What is Amazon EC2?  
👉 What is PuTTY and how does it work?  
👉 How do I get the key for my PuTTY generator?  
👉 What are EC2 security groups?  
👉 What is WinSCP and how does it work?  
  
GO\_STP\_8480 - Kowshik :  
  
1. Amazon Elastic Compute Cloud (EC2) is one of the services provided by Amazon Web Services and provides access to server instances on demand as a service. EC2 is a core part of AWS providing the compute facility for organisations.  
2. Like OpenSSH, PuTTY is a very versatile tool for remote access to another computer. It's probably used more often by people who want secure remote shell access to a UNIX or Linux system than for any other purpose, though that is only one of its many uses. PuTTY is more than just an SSH client.  
3. To generate a key with PuTTY, you should:- Download and start the puttygen.exe generator.  
- In the “Parameters” choose SSH2 DSA and press Generate.  
- Move your mouse randomly on the small screen to generate the key pairs.  
- Enter a key comment, which will identify the key (useful when you use several SSH keys).  
- Type in the passphrase and confirm it. The passphrase is used to protect your key. You will be asked for it when you connect via SSH.  
- Click “Save private key” to save your private key.  
- Click “Save public key” to save your public key.  
4. A security group acts as a virtual firewall for your instance to control inbound and outbound traffic. When you launch an instance in a VPC, you can assign up to five security groups to the instance. It consists of five security groups  
5. WinSCP is an open source free SFTP client, FTP client, WebDAV client, S3 client and SCP client for Windows. Its main function is file transfer between a local and a remote computer. Beyond this, WinSCP offers scripting and basic file manager functionality.

**Complete Set of Assignments GOEDUHUB\_TECHNOLOGIES**

Assignment 1 - Python Basics

<https://colab.research.google.com/drive/1fSCMDsENrtlRyAUs6XYmdO4JtupeSdpE?usp=sharing>

Assignment 2 - Python List

<https://colab.research.google.com/drive/1wAXtnP_IiuWe-JZfHKOXN52H6WKa8Gv5?usp=sharing>

Assignment 3 - Python Dictionary Sets

<https://colab.research.google.com/drive/1j1R4C9D4ngl-HylrcKhRkysaGtDbaTkL?usp=sharing>

Assignment 4 – NumPy

<https://colab.research.google.com/drive/1_haSTBme1T0ZfTpA_4q9xS9WaPLp7K9q?usp=sharing>

Assignment 5 – Pandas

<https://colab.research.google.com/drive/1mfIfOQqsECdjTebHSExj3Sg3vz4BnAKj?usp=sharing>

Assignment 6 – Matplotlib

<https://colab.research.google.com/drive/1fXM5Olpcta-ogYGI9_WgQtdBo0-4WuTH?usp=sharing>

Assignment 7 - Simple Linear Regression

<https://colab.research.google.com/drive/1OEAtg0WKYZSoGHps4GGSeb1LJ2Z67N8h?usp=sharing>

Assignment 8 - Multi Linear Regression

<https://colab.research.google.com/drive/1mBkyCy7Dai3osIhufXvAMHpD61HwPPik?usp=sharing>

Assignment 9 - Logistic Regression

<https://colab.research.google.com/drive/1lLmaWOz0UPF6_z1noeubbJh5tFst_jIW?usp=sharing>

Assignment 10 - Pandas Dummy Variables and One-hot encoding sklearn

<https://colab.research.google.com/drive/1-nJ0Bx1mH0NHaqax_HPsGgi6pXxzb6-H?usp=sharing>

Assignment 11 - Decision Tree

<https://colab.research.google.com/drive/1bhJSDEeuVaHShSu3jYj2cG2jJyst7nuA?usp=sharing>

Assignment 12 - Support Vector Machine

<https://colab.research.google.com/drive/1_CSdXshm16vaEjjVV0Yehu3NLX7yTnW3?usp=sharing>

Assignment 13 - Exploratory Data Analysis using Titanic Dataset

<https://colab.research.google.com/drive/1GhRGxymM8cddCRfAR2jhIm0Bnkni5qBg?usp=sharing>

Assignment 14 - KNN Use Case

<https://colab.research.google.com/drive/1yLpbPzhi3GtKxfHM60H1ks78JCoTh3A4?usp=sharing>

Assignment 15 - Multinomial Naive Bayes Classifier Spam or Not

<https://colab.research.google.com/drive/142YTmWsPXevTZkOvbSv9j8H5y3egi5XL?usp=sharing>

Assignment 16 - Gaussian Naive Bayes Use case Titanic Dataset

<https://colab.research.google.com/drive/1ZBlRJi4X_8iISLd0GUk9CIIL4fBJcKoy?usp=sharing>

Assignment 17 - Movie Recommendation Flask App on Heroku

<https://colab.research.google.com/drive/1Cq2-XpYQ88YwnjScLiwcDHRzeer3xU_X?usp=sharing>