Data Science Curriculum

Week	Hours	Topic	Content to be covered
Week 1	2	Introduction to the course	Installing and Downloading Anaconda, Overview of Jupyter Notebook, Overview of Google Colab, Introduction to the course and projects, Understanding various roles in the field of Machine Learning
	2	Python Introduction	Applications of Python, Identifiers, Comments, print() and input(), Operators in Python - Unary, binary and ternary operators.
	2	Data Structures	Premitive Data Types - int, float, complex, bool, str Data Structures - Lists, Tuples, Sets and Dictionaries
	4	Flow Control	Decision Statements - if else Loops - for, while, for else, while else, break and continue
		Assignment	
Week 2	3	Functions and Modules	Introduction to functions, Define a function, Function calling, Local and Global Variables, Modules
	4	Numpy & Pandas	Applications of numpy, proof of efficiency, creating a numpy array, slicing and indexing, numpy maths and statistics Introduction to pandas, series and dataframes, creating a series and dataframes, data accessing using indexing, dataframe functionalities, working with .csv files
	3	Visualising Data with Matplotlib and Seaborn	Visualizing the data with Matplotlib and Seaborn. Univariate, Bivariate, Multivariate Analysis. Data Cleaning, Handling Missing values and Outliers.
		Mini Project - (Data Analysis Mini Project)	
Week 3	4	Advance Pandas	Cross Tabs, loc vs iloc, Merge, Groupby, Advance Pandas Manipulations, Datetime and String Manipulations
	4	Application Development with Streamlit	Introduction to Client Server Architecture, Request and Response, Write and Magic Commands, Text Elements, Data Display Elements, Chart Elements, Input Widget, Media Elements, Layouts and Containers, Status Elements, Control flow
	2	Cloud Deployment	Introduction to Cloud, Introduction to Deployment of Applications, Some common git commands, Deploying a Streamlit app on Heroku
		Project - (Building Visualisation Dashboard with Streamlit)	
Week 4	2	Moving to Machine Learning	Why learn AI/ML? AI vs ML vs DL, Supervised vs Unsupervised learning, Classification vs Regression
	3	Linear Algebra and Statistics	Introduction and why Linear Algebra, Fundamentals of Vectors and Matrices, Unit Vector, Matrices operations, Dot Product of vectors, Angle between two vectors, Projection of a vector onto another vector, Length of projection Introduction and Fundamentals of Statistics, Measure of Central Tendency - Mean, Median and Mode, Measure of Spread – Range, Variance, Standard Deviation and IQR, Covariance and Correlation
	2	Data Preprocessing	Introducing sklearn module, Splitting data into train and test, Preprocessing Numerical Data, StandardScaler, Preprocessing Categorical Data, Nominal Encoding, Ordinal Encoding, Building Models, Making Prediction and Evaluation using sklearn, Case Study
	3	KNN Algorithm	Introducing the first ML Algorithm, Understanding the Model Building Phase, Intuition behind KNN, Developing the algorithm for KNN, Solving Classification and Regression Problems, Code Implementation for KNN
		ML Mini Project (Diamond Price Prediction)	
Week 5	4	Linear Regression and Logistic Regression	Understanding equation of a Line and Hyperplane, Building a cost function for Regression and Classification Problem, Mathematical Formulation of Linear Regression, Calculating the errors, Intuition behind Linear Regression, Python Implementation, Geometric Intuition behing Logistic Regression, Sigmoid Function, Code Sample
	2	Performance Metrics	Accuracy, Confusion matrix, Precision and Recall, F1 Score, ROC AUC, Log Loss, MAE, MSE, RMSE, R square, Adjusted R square
	2	Decision Trees	Introduction to rule based approach to solve classification and regression problem, Building Decision Trees, ID3, C4.5, C5.0 algorithms to build decision trees, Entropy, Information Gain, Gini Impurity, Python Code Sample

	2	Working with Image and Text Data	Image Data - Understanding Images, RGB channels, Images as 3D numpy arrays, working with PIL module, Image Manipulation using PIL, Cleaning Image Data Text Data - Introduction to NLP, Why Text data is hard to work with, Cleaning Text Data, Tokenisation, Stop Words, Lemmatization, Stemming, Bag of Word Representation
		Project - (Text Data Case Study)	
Week 6	6	Model Selection	Occam's Razor, What is model complexity, Overfitting and Underfitting, Bias and Variance, Model Complexity vs Train Test Error, Cross Validation, Hyper Parameter Tuning, Case Study
	4	Ensembles	Introducing Model Ensembles, Voting Ensembles, Stacking, Bagging - RandomForestClassifier and RandomForestRegression, Cascading, Boosting - XGBoostClassifier and XGBoostRegression
		Project - (Image Data Case Study)	
Week 7	4	Introduction to Unsupervised Learning	Undestanding Unsupervised Learning, Clustering, Applications of Clustering, K-Means Algorithm, Python Code Sample
	2	Case Study	Solving Customer Segmentation using Clustering
	4	Dimensioanlity Reduction	Introduction to Dimensionality Reduction, Introduction to Principle Component Analysis, Eigen Values and Eigen Vectors, Othogonality, Transforming Eigen Values, Proportion of Variance Explained in PCA, Code Implementation
		Final Project Discussion	
Week 8	4	Model Productionisation	Serialisation and Deserialisation of objects, import pickel, Serialising the machine learning models, .pkl files, dump function, Building an app with model integration, Deseriaising the machine learning model, load function, Structuring your application
	4	Model Deployment	Model Deployment on Cloud
	2	What's next?	Introduction to Deep Learning, Data Engineering and MLOps
		Final Project	