

Data Science Syllabus:

- Intro Data Science

Python:

- Install python
- Jupyter Notebook
 - <http://jupyter.readthedocs.io/en/latest/install.html>
- Spyder
 - <https://pypi.python.org/pypi/spyder>
- Introduction to Python
- Basic Python Syntax
- Why python for machine learning
- Datatypes
 - Int
 - Float
 - Complex
 - Boolean
 - str
 - List
 - Tuple
 - Set
 - Dictionaries
- Functions
- Lambda function
- Map function
- Modules
- Files I/O
- Exceptions
- Classes/Objects
- Iterators

MySQL with Python

- Installation steps
- Installing MySQL DB Connector and creating Database
- Data Science Performing CRUD Operation on Table

Statistics

- **Descriptive Statistics**
 - Data exploration (histograms, bar chart, box plot, line graph, scatter plot)
 - Qualitative and Quantitative Data
 - Measure of Central Tendency (Mean, Median and Mode),
 - Measure of Positions (Quartiles, Deciles, Percentiles and Quantiles),
 - Measure of Dispersion (Range, Median, Absolute deviation about median, Variance and Standard deviation), Anscombe's quartet
 - Other Measures: Quartile and Percentile, Interquartile Range

Data Science Syllabus:

➤ Statistical Analysis Initial Data Analysis

- Relationship between attributes: Covariance, Correlation Coefficient, Chi Square
- Measure of Distribution (Skewness and Kurtosis), Box and Whisker Plot (Box Plot and its parts, Using Box Plots to compare distribution) and other statistical graphs Probability
- Probability (Joint, marginal and conditional probabilities)
- Probability distributions (Continuous and Discrete)
- Density Functions and Cumulative functions

Machine learning Packages

- Numpy
- Pandas
- Matplotlib
- Seaborn

EDA – Exploratory Data Analysis

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Z-Score statistics
- Probability Density

Machine algorithms

➤ Linear Regression

- Linear Regression in depth math intuition
- Regression basics: Relationship between attributes using Covariance and Correlation
- Relationship between multiple variables: Regression (Linear, Multivariate) in prediction
- Residual Analysis
- Identifying significant features, feature reduction using AIC, multi-collinearity
- Non-normality and Heteroscedasticity
- Hypothesis testing of Regression Model
- Confidence intervals of Slope
- R-square and goodness of fit
- Influential Observations – Leverage
- Ridge and Lasso
- Multicollinearity in Linear Regression
- Learning – Bias and Variance

➤ Multiple linear Regression.

- Polynomial Regression
- Regularization methods
- Lasso, Ridge and Elastic nets
- Categorical Variables in Regression
- Non-Linear Regression
- Logit function and interpretation
- Types of error measures (ROCR)
- R Square and Adjusted R square

Data Science Syllabus:

- Hypothesis Test, Type 1 Error, Type 2 Error
- P-Value, T-test, chi Square test, Anova Test
- P-Value, T- Test, correlation with python
- Chi Square test with python

- Logistic Regression in depth intuition
- Logistic Regression Multiclass classification
- Decision Tree
 - Entropy in Decision tree
 - Decision Tress information Gain
 - Gini Impurity intuition
 - Decision Tree split for numerical Feature
- Forecasting models
 - Trend analysis
 - Cyclical and Seasonal analysis
- Smoothing; Moving averages; Box-Jenkins, Holt-winters, Auto-correlation; ARIMA
- Examples: Applications of Time Series in financial markets
- Performance metrics for classification problems

Project

- **House price Prediction**
 - Advance house price prediction EDA
 - Advance house price Prediction Feature Engineering
 - Advance house price Prediction Feature Selection
 - Performance metrics (ROC, AUC curve) for classification problem

- **k nearest neighbor's Classification intuition**
 - K Nearest neighbor's Classification practical solution
 - K Nearest neighbor's with explanations
- **Ensemble: What is bagging (bootstrap Aggregation)**
 - Random Forest Classifier and Regressor
 - Handling imbalanced Dataset using python
 - Hyperparameter Optimization for Xgboost
 - What is AdaBoost (Boosting techniques)
 - Euclidean Distance
 - Manhattan Distance
- **K Means Clustering intuition**
- **Hierarchical Clustering intuition**
- **DBSCAN Clustering**
- **Silhouette Clustering**

- Curse of Dimensionality
- Dimensional Reduction
- Principal Component analysis
- Principal Component analysis (PCA) using sklearn and python
- Cross Validation
- Cross Validation Types?

Data Science Syllabus:

- How to find optimal for binary classification
- Bayes Theorem | Conditional Probability
- Naive Bayes Classifier in depth Intuition
- Support Vector Machine (SVM) Basic intuition
- Maths intuition behind Support Vector Machine
- Gradient Boosting in Depth
- Gradient Boosting complete maths in depth
- Let's Implement LSTM RNN Model for univariate Time series forecasting – Deep learning

Artificial Intelligence

Foundations for AI

- **AI: Application areas**
- **AI Basics (Divide and Conquer, Greedy, Branch and Bound, Gradient Descent)**
- **NN basics (Perceptron and MLP, FFN, Backpropagation)**
- **Convolution Neural Networks**
 - **Image classification**
 - **Text classification**
 - **Image classification and hyper-parameter tuning**
 - **Emerging NN architectures**
- **Recurrent Neural Networks**
 - **Building recurrent NN**
 - **Long Short-Term Memory**
 - **Time Series Forecasting**
- **Deep Learning**
 - **Auto-encoders and unsupervised learning**
 - **Stacked auto-encoders and semi-supervised learning**
 - **Regularization - Dropout and Batch normalization**
 - **Py torch models like Vggnet , Resnet, Densenet, Alexnet , Inception blocks Etc.**
 - **Keras models (yolo models)**
 - **Computer vision**
- **Additional Skills:**
 - **English communication skills.**
 - **Personality Development.**

Computer Vision:

- OpenCV Installation | OpenCV
- Reading, writing and displaying images with OpenCV
- Face and Eye Detection from images using HAAR Cascade Classifier
- Face and Eyes detection from the videos using HAAR Cascade Classifier
- Pedestrian Detection using OpenCV from videos
- Car Detection using HAAR Cascade and OpenCV from videos
- Face Recognition using open CV and VGG16 transfer learning
- Draw rectangle from webcam and sketch process it on a live feed

Natural Language Processing -NLP concepts

Data Science Syllabus:

- Tokenization
- Stemming and Lemmatization
- Stemming
- Lemmatization
- Bag-of-words
- TF-IDF Intuition | Text Preprocessing
- Implementing a Spam classifier
- Word2vec easily explain
- Regular Expressions
- Sentence Splitting and Tokenization
- Punctuations and Stop words, Incorrect spellings
- Properties of words and Word cloud
- Lemmatization and Term-Document TxD computation
- Sentiment Analysis (Case Study)
- Word embedding using keras
- Kaggle projects

Deployment of ML Models

- Deployment of model's premises Vs IAAS Vs PASS Cloud
- Deployment of ML Models in Heroku using Flask
- Deployment of NLP model in Heroku cloud
- Deployment of ML model in Heroku cloud
- Deployment of ML Models in AWS EC2 instance
- Deployment of movie recommendation model
- Deployment of machine learning model in Google cloud
- Deploying Machine learning models in Azure cloud

Big Data (Option):

- Introduction of Big Data
- Hadoop system Setup-big Data
- Basic of Linux and practical Implementation
- Understanding Hadoop command Linux
- Understand of basic Hadoop commands
- What is Scaling in Scaling out
- What are the problems in Scaling out
- Introduction to Hadoop
- Components of Hadoop
- Understanding the HDLC Design
- Name Node Architecture
- Hadoop 2.x yarn Architecture.

== > **Real time Projects 3 with explanation**

Data Science Syllabus: