# Python for Data Science and Machine Learning

- by Nitish Vishwakarma

## Python libraries and Algorithms/Functions :

1. **Scikit Learn**

    There are following ML algorithms covered:

    a. Linear Model- Regression
    b. Logistic Regression
    c. K- NN
    d. Decision Tree
        1. CHAID
        2. CART
    d. Random Forest
    e. Bagging with Random Forest
    f. Boosting with Random Forest
    g. K - Means Clustering
    h.Hierarchical Clustering
    I. Principal Component Analysis.
    j. Factor Analysis

2. **Scipy**
    a. Statistical Test
    b. Skewness
    c. Kurtosis
    d. Mean , Median and Harmonic Mean
3. **Keras**
    a. ANN
    b. CNN
    c. Perceptron
4. **Tensorflow**
    a. CNN
    b. Perceptron
    c. Deep Learning Project on Image Processing (Feed forward, back propagation, Epoc, batch, activation function etc.)

**Performance Evaluation Metrics covered with Library:**
**These are covered in (Excel Calculations , SAS ,Python)**

**Python Library : Sklearn- Metric**

1. **Accuracy Score**
2. **F1 Score**
3. **Confusion Matrix (Techniques on how to trade-off sensitivity and specificity)**
    a. **True Positive**
    b. **True Negatives**
    c. **False Positive**
    d. **False Negative**
4. **False Positive Rate**
5. **False Negative Rate**
6. **Precision**
7. **Recall**
8. **Prevalence**
9. **Mean Absolute Error**
10. **Mean Squared Error**
11. **F1 Score**

**SAS Modeling with Algorithm covered:**

1. **Linear Regression -** Proc Reg, Proc GLM
2. **Multiple Linear Regression -** Proc Reg, Proc GLM
3. **Binary Logistic Regression** : Proc Logistic, Excel implementation with Confusion Matrix calculation
4. **Rank Ordering** using Proc Rank , Proc Univariate
5. **K- Means- Clustering -** Proc Varclus, Proc cluster
6. **Decision Tree**- Proc Hpsplit
7. **Factor Analysis**- Proc Factor

**Major key machine learning concepts,techniques and best practice included in this course:**

**These points are important from the perspective of :**

- Industry Relevance
- They can be direct, and are critical questions in Interview rounds and answer to which makes a lasting impression?
- They can be an answer to real world challenges during a job/project.
- Some of them are ground rules and some are advanced practices.
- ***Students will learn these here with best words/terminology to use and when , also which/when not to use?**

1. Difference between Random Observation and Statistical Evidence and how we do it?
2. Descriptive Statistics & Inferential Statistics? How do we perform them?
3. Variables/Features : Types and difference?
4. Sampling & Population: SRS and stratified sampling
5. Feature Engineering techniques
6. Feature Selection through Lasso etc.
7. Feature Selection with forward/backward/stepwise
8. Standardization & Normalization
9. One- Hot Encoding / Design Variable : Same or different ?
10. Bias-Variance concepts and Generalization concept explanation.
11. Overfit- Underfit Problem and resolutions.
12. Difference between Correlation and Covariance with cases
13. Pearson or Spearman Correlation. Which & when to use ?
14. Observation and Performance window and its implication on Model Testing & Training
15. Train Test Split and importanceModel Serialization and Model deserialization (with Pickle file implementation)
16. Hyperparameter and Parameter estimates- Difference?
17. Multicollinearity and best practices available.
18. Effect of Un-Tidy data in feature engineering.
19. What is Loss in Loss function? And what is the function?
20. Gradient Descent and +ve / -ve Slope?
21. Why do we say OLS to OLS?
22. Can we minimize Maximum likelihood MLE? Or Always Maximize?
23. Confidence Interval - 95% or 99% - Layman definition
24. The mystery of P-Values and its relevance in industry.
25. P values- Type1,Type-2, Type-3. Which one to choose?
26. Pure Split and impure Split . How to choose the best split?
27. Sigmoid Function/Relu Function? Which one to choose
28. Probability and difference with Odds/Odds Ratio
29. Why do we use transform data with log?
30. Target based binning. How to choose the optimal bin?