Data Play

Python with Machine Learning

Syllabus

CHAPTERS	TOPICS	ADVANCED TOPICS	CASE STUDIES
Introduction to Python	Python Basics Jupyter Notebook List Tuple Set Dictionary		
Control Structures	Boolean and Comparisons Conditional statements (IF-ELSE) Operator Precedence Lists - Operations and Functions Exceptional Handling	List Comprehension Generators	
Basic Probability and Terms	Events and their Probabilities Rules of Probability Conditional Probability and Independence Permutations and Combinations Bayes Theorem Descriptive Statistics Compound Probability Conditional Probability		
Probability Distributions	Types of Distributions Functions of Random Variables Probability Distribution Graphs Confidence Intervals Central Limit Theorem	Bernoulli Binomial Exponential Poisson	Practical Applications of Each Distribution Application of Exponential Distribution in forecasting Visitor volume on a website. How come agent staffing is done.
Data Transformations and Quality Analysis	Merge, Rollup, Transpose and Append Missing Analysis and Treatment Outlier Analysis and Treatment		

Exploratory Data Analysis	Summarizing and Visualizing the Important Characteristics of Data Basics of Data Visualization Crosstabs	Line Plots Bar Charts Pie Charts Histograms Scatter Plots Parallel Coordinates Correlation Scatterplots	Employee Attrition Data **(Ormae Interview Assignment)**
Hypothesis Testing	Two Sample test One Sample Test A/B Testing		
Pandas Data Frame	Numpy Array Pandas - Series and Data frames	Group By Merge Join Concatenate	
Linear Regression /Polynomial Regression	Assumptions Handling Outliers, Categorical Variables, Autocorrelation, Multicollinearity, Heteroscedasticity Interpreting P Value	How weights are calculated using Gradient Descent? Why you need second order derivative for Regression	IMDB Score Prediction For A Movie (Marax.AI Interview Assignment) Take Away: Property Price Prediction using Real Estate Data
Understanding basic machine learning concepts	Performance Measures Bias-Variance Trade-Off Overfitting & Under fitting Supervised /Unsupervised		
Logistic Regression	Need of Regularization Lasso and Ridge called L1/L2 ROC Curve Confusion Matrix		Loan Defaulter Take Away : IPL Match Winner Prediction **(Sig Tuple Interview Assignment)**
Naïve Bayes	Bayes Theorem Assumptions		Gender Prediction **([24]7.ai Interview Assignment)**
Decision Trees	Entropy Gini Index Pruning of Trees Classification and Regression Trees	Understanding where Decision Tree could perform better then Logistic Regression	Benchmarking IPL Winner Prediction using CART
Time Series	Handling Time Series Data ARIMA Model Stationary Process ACF/PACF Functions Correlation Covariance Mean Squared Error Mean Absolute Percentage Error	How can one include features like Day of Week and promotional day as features in ARIMA modelling in order to include seasonality components?	Forecasting Visitor Volumes for Hilton website. (Past 2018 **Data of Hilton** Shared from my own company experience) Time series analysis of Crop Data **(Social Cops Interview Assignment)**

Clustering	K Means DBScan	Evaluation of Clustering Algorithm Finding optimal Value of K Limitation of K Means	Iris Dataset Take Away : 1.Cars K Means (Kaggle Problem) 2.Clustering on text documents (To be done Later after reading NLP Concepts)
Support Vector Machine	Linear SVM Kernel Trick Hard Margin Soft Margin	In Case of Outliers why SVM are better. SVM over Logistic	Chat Intent Classification Models using TFIDF and SVM. (Quiet Big Datasets ~ 80K Chats) Spam Classification **(IIT Delhi Research Scholar Interview Assignment)** Auto Analyze Bank Statements using TFIDF. (Will build a model which will come up with Classes of all transactions happening)
Overview Of Text Classification	TFIDF LDA(Topic Modelling)	How to convert text documents into numerical vectors and find cosine similarity between documents.	
Ensemble Algorithm	Ensemble Techniques Bootstrap Aggregation Random Forest Boosting		
Big Data	Brief Introduction To Big Data Technologies Hadoop Map Reduce Apache Pig Apache Hive Spark		
Mock Interview /Resume Building Support	Face To Face Interview To Prepare For Actual Interview Experience	Interview questions asked by various companies.	

Why Data Play?

- Placement Support (if scored well in the Final Data Play Test which will involve case study and MCQ questions)
- Get to know the Real Time Challenges Faced In Industry
- Data Play aims to cover all the nitty-gritty of the theory.
- Hand to Hand practical implementation of the topics covered
- Get acknowledged with Questions asked in top Companies for Data Science positions
- Additional Classes if students want to deep dive into the depth of Algorithms
- Individual Attention to every student