5-Week Data Science Bootcamp

DETAILED SYLLABUS

Overview

In our endeavour to build data culture and democratize Data Science learning, we are launching a 5-week Data Science Bootcamp with the help of resources contributed by academia and industry experts. The online bootcamp will have a series of day-wise learning modules along with intuitive practice quizzes/challenges.

This is a community initiative, driven by experts and mentors, and you have the opportunity to attend it for free.

Prerequisites

- Nil, anyone with a passion for learning can make it to the finish line :)

Format

Tutors will provide learners with guided learning paths, resources and exercises to solve. The entire schedule, practical details, registration details will be put up very soon. A brief summary of the format can be found below**:**

- **Day-wise modules:** Trainers will post day-wise challenges and learning modules (mostly some of the best-curated content available on the internet that would allow you to have a structured learning path)
- **For real-time communication**, we will be using Discord. This medium will help learners to clear doubts on a real-time basis if they are stuck somewhere. In addition, this will also allow learners to interact with the mentors and fellow learners

- **Live doubt clearing and mentorship sessions** will be organized every week based on the requirements of the learners

Schedule

Week #0 - Python Crash Course and Intro to Data Science (Optional)

- Intro to Data Science - its prominence and use-cases
- Environment setup - python installation - anaconda ide
- Python for Data Science
    - Basics of Python
        - Print a string "Hello World"
        - Python basic syntax
        - Data structures and types
    - Python Lists & Strings
    - Intro to Functions
    - Brief Intro to Python Libraries for Data Science - Numpy and Pandas

Week #1 - Data Analysis and Data Visualization (Release on: 11th March)

- Dive Deep into Numpy and Pandas libraries
- Python Web Scraping
- Exploratory Data Analysis
- Intro to Data Visualization
- Graded Quiz 1 - 18th March

Week #2 - Advanced Exploratory Analysis and Data Pre-Processing (Release on: 18th March)

(Data Cleaning, Outlier detection etc.)

- Basic Statistics
- Charts and Visualization
- Outlier Analysis
- Handling Missing Values
- Handling Imbalanced datasets, Oversampling - SMOTE

- Standardization/Normalization of data - what, why and when?
- Graded Quiz 2

Week #3 - Feature Selection and Building ML Models (Release on: 25th March)

- Intro to feature extraction and feature selection - explain how they are different
- Elaborate more on Feature Extraction
- Feature selection and its importance
    - Various feature selection/engineering techniques
    - Boruta
- Building efficient and effective models
- Splitting data into test and train datasets
- ML Algorithms:
    - Linear Regression
    - Logistic Regression
    - Cost function & Gradient Descent
- Overfitting & Underfitting

Week #4 - Model tuning and ML Algorithms (Release on: 1st April)

- Other ML Algorithms
    - Tree-based models
        - Decision trees
        - Random forest
        - A brief intro to other boosting and bagging techniques/algorithms
- Model tuning
    - Hyperparameter tuning
    - Evaluation Metrics (Model evaluation)
- Project - solve real-world data science problem on Ed-tech and Fintech**.**
- Graded Assignment (Released around 8th April)

Week #5 - Applied Data Science & ML - Problem-solving (Release on: 8th April)

- HR Analytics problem - predicting employee churn
- Ed-tech customer analysis - predicting user churn
- Fraud analytics - predicting fraud detection

- Anti-money laundering analytics - predicting money laundering cases in transactions data
- Real-estate price analysis - problem
- Getting started with Data Science competitions - Kaggle