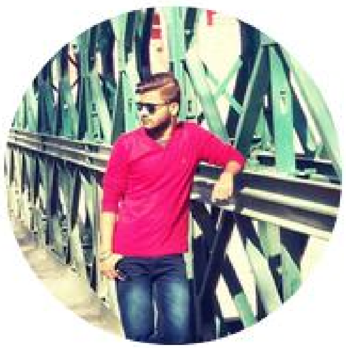
# Machine Learning Bootcamp

Evaluate Your Time

# Machine Learnining Bootcamp

## Meet The People Who made it Possible

### CORE MANAGERS



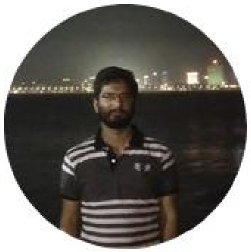
**PRAKHAR TIWARI**
Chairperson

**PRASHANT ARYA**
Vice-Chairperson

**ANSHUL CHAUHAN**
Treasurer

# Bootcamp Team

## Those Who were always Here for You!



**PRASHASY ASHOK**
Debugging Mentor

**DHRUV DHINGRA**
Data Structure Instructor

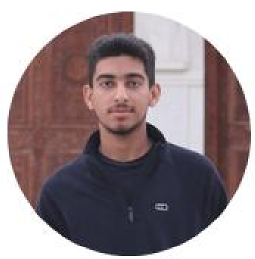
**ABHIMANYU AMAN SRIVASTAVA**
Hands on Mentor | IEEE Member

**SAJAL LAHOTI**
Python Mentor

**AYUSH PATEL**
Hands On Mentor

**CHANDRASHEKHAR CHOUDHARY**
Python Mentor

**AYUSH VERMA**
Design Committee

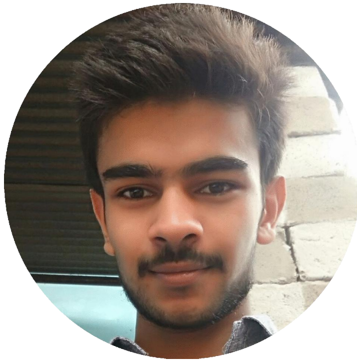


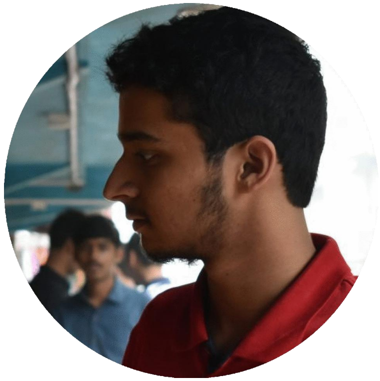
**ARNAV SHARMA**

**DEEPESH YADAV**

# Bootcamp Team

## Your Mentors!





**Abhimanyu Shan Aman Srivastava**
IEEE Member & Deep Learning Enthusiast

**Aditya Mohan**
ACM Electronics Team

# What we Learnt throughout the Bootcamp

Python Basics

Data Preprocessing

Regression

Classification

Regularization Methods

Clustering

Natural Language Preprocessing

Deep Learning : ANN and CNN

Dimensionality Reduction

Ensemble Methods

# Day 1

# Python Basics and Data Structure

Operators

Decision Statements

Loops

Strings

Methods

Functions and Lambda Expression

# Day 2

# Python Basics: Object Oriented

Classes

Objects

Errors

Exception Handling

Debugging Techniques

# Day 3
# Data Preprocessing

Statistical Operations

Python Libraries

Data Encoding

Splitting Dataset

Feature Scaling

# Day 4

# Preprocessing: Feature Scaling

Working with Dataset

Obtaining Features

Slicing of Data

Feature Scaling

Standardization

Normalization

# Day 5
# Regression

Simple Linear Regression

Multiple Linear Regression

Polynomial Regression

Support Vector Regression

Decision Tree

Random Forest Regression

# Regression:
# A comparative Analysis

| Regression Model | Pros | Cons |
|---|---|---|
| Linear Regression | Works on any size of dataset, gives informations about relevance of features | The Linear Regression Assumptions |
| Polynomial Regression | Works on any size of dataset, works very well on non linear problems | Need to choose the right polynomial degree for a good bias/variance tradeoff |
| SVR | Easily adaptable, works very well on non linear problems, not biased by outliers | Compulsory to apply feature scaling, not well known, more difficult to understand |
| Decision Tree Regression | Interpretability, no need for feature scaling, works on both linear / nonlinear problems | Poor results on too small datasets, overfitting can easily occur |
| Random Forest Regression | Powerful and accurate, good performance on many problems, including non linear | No interpretability, overfitting can easily occur, need to choose the number of trees |

# Day 6-7

# Classification

Logistic Regression

K Nearest Neighbour

Support Vector Machine

Kernel SVM

Naive Bayesian

Decision Tree Classification

Random Forest Classification

# Classification:
# A comparative Analysis

| Classification Model | Pros | Cons |
|---|---|---|
| Logistic Regression | Probabilistic approach, gives informations about statistical significance of features | The Logistic Regression Assumptions |
| K-NN | Simple to understand, fast and efficient | Need to choose the number of neighbours k |
| SVM | Performant, not biased by outliers, not sensitive to overfitting | Not appropriate for non linear problems, not the best choice for large number of features |
| Kernel SVM | High performance on nonlinear problems, not biased by outliers, not sensitive to overfitting | Not the best choice for large number of features, more complex |
| Naive Bayes | Efficient, not biased by outliers, works on nonlinear problems, probabilistic approach | Based on the assumption that features have same statistical relevance |
| Decision Tree Classification | Interpretability, no need for feature scaling, works on both linear / nonlinear problems | Poor results on too small datasets, overfitting can easily occur |
| Random Forest Classification | Powerful and accurate, good performance on many problems, including non linear | No interpretability, overfitting can easily occur, need to choose the number of trees |

acm JUIT Communitate Valemus

# Day 8

# Clustering

K Means Clustering

Hierarchical Clustering

# Clustering:
# A comparative Analysis

| Clustering Model | Pros | Cons |
|---|---|---|
| K-Means | Simple to understand, easily adaptable, works well on small or large datasets, fast, efficient and performant | Need to choose the number of clusters |
| Hierarchical Clustering | The optimal number of clusters can be obtained by the model itself, practical visualisation with the dendrogram | Not appropriate for large datasets |

# Day 9

# Natural Language Processing

Simple Linear Regression

Multiple Linear Regression

Polynomial Regression

Support Vector Regression

Decision Tree

Random Forest Regression

# Day 10
# Deep Learning

Artificial Neural Networks

Convolutional Neural Networks

# Thank You for attending

This session was presented by ACM JUIT