MACHINE LEARNING- WEEK-2

DSBA CURRICULUM DESIGN

FOUNDATIONS

Data Science Using Python

Statistical Methods for Decision Making

CORE COURSES

Advanced Statistics

Data Mining

Predictive Modelling

Machine Learning(Week-2/5)

Time Series Forecasting

Data Visualization

SQL

DOMAIN APPLICATIONS

Financial Risk Analytics

Marketing Retail Analytics

LEARNING OBJECTIVE OF THIS MODULE

- Supervised Learning : KNN & Naïve Bayes

- Ensemble Techniques: Bagging, Boosting, Cross-validation and SMOTE

- Text Mining & Sentiment Analysis

LEARNING OBJECTIVES OF THIS SESSION

- Ensemble Techniques – Bagging and Boosting

- Under sampling and Over sampling (SMOTE)

- Cross-Validation
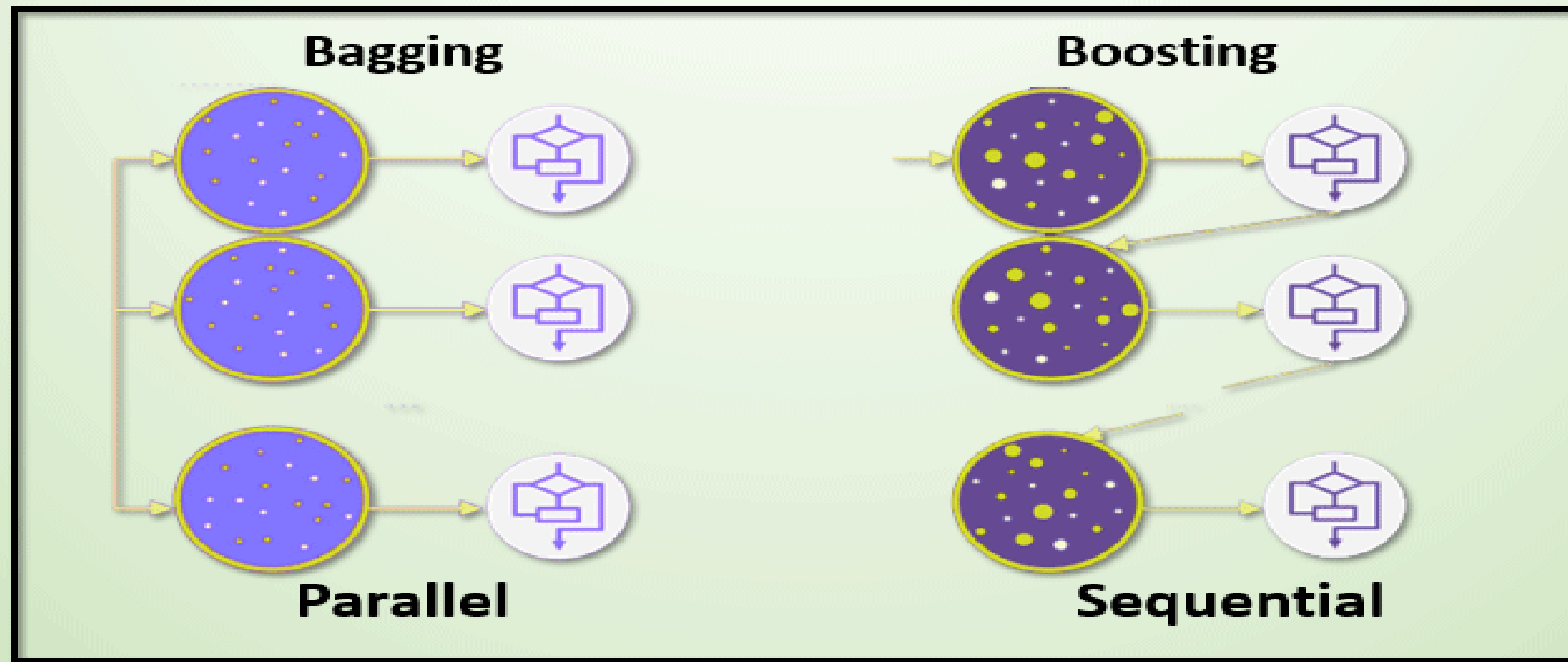
greatlearning
Learning for Life

TRY ANSWERING THE FOLLOWING

- Name any algorithm based on ensemble techniques.

- Is Bagging and Boosting same. If not, what is the main difference.
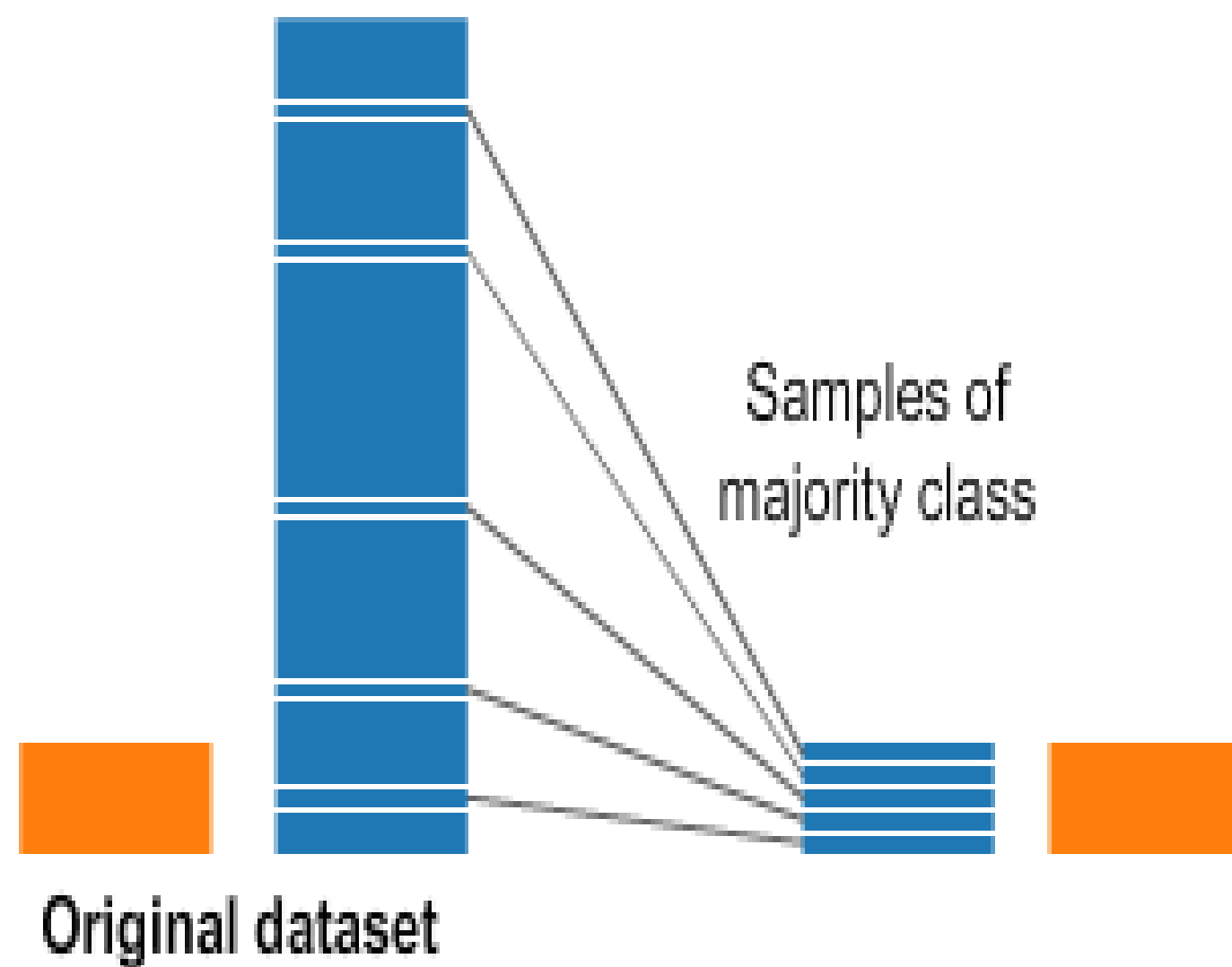
- Name any over sampling technique.

BROAD OVERVIEW- Bagging and Boosting
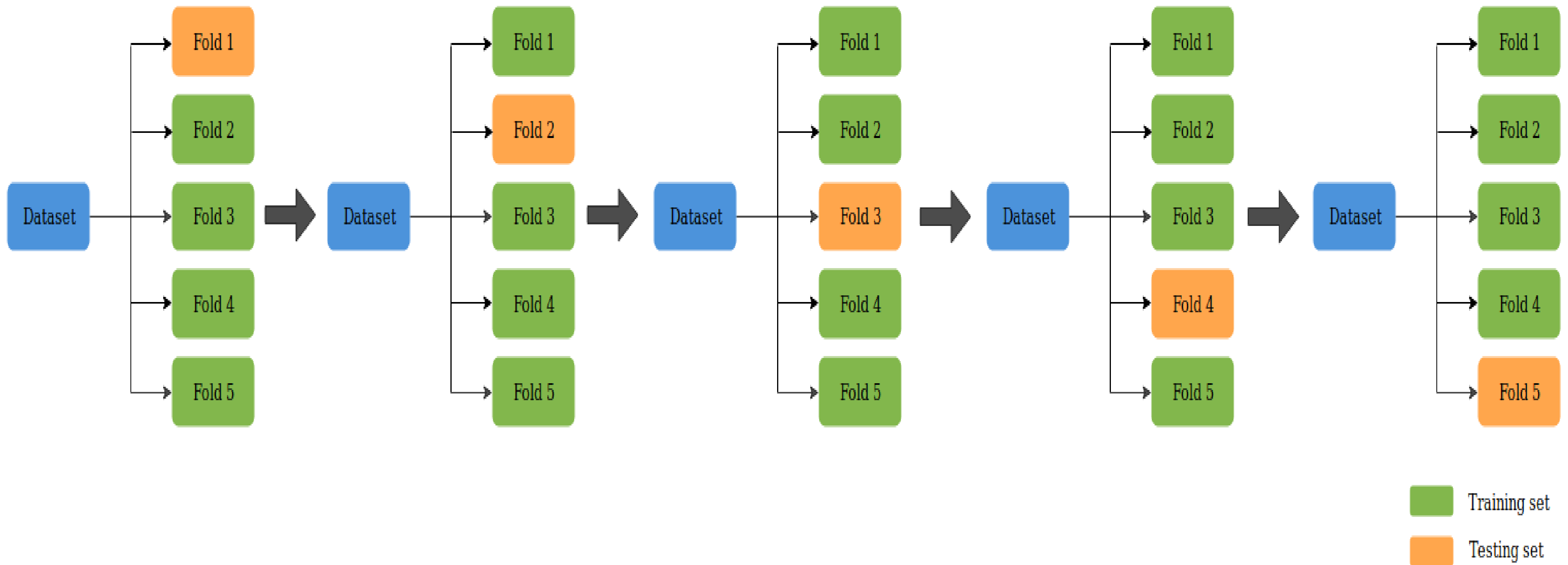
BROAD OVERVIEW- under and over sampling

Undersampling

Samples of majority class

Original dataset

Oversampling

Copies of the minority class

Original dataset

BROAD OVERVIEW – Cross-validation

Industry Application - Boosting simplifies complex prediction tasks in the Energy sector

Security of bulk electric power systems is an opportunity in modern power engineering due to the continued growth in renewable energy generation.

Most energy management systems, like Siemens and ABB etc., use one or more security assessment predictors such as sensitivity matrix, security indicators, distribution factors, fast decoupled load flows etc to reduce the computational effort of the security assessment.

Methods for the assessment of security and voltage stability of electric power system shows that traditional approaches cannot be effectively applied on real time conditions because of their computation complexity. Research shows that the use of bagging and boosting models is particularly helpful in solving these challenges very efficiently.

Reference: https://arxiv.org/pdf/1601.01675.pdf

Industry Application - Boosting techniques for tumor detection

In the field of medical diagnostics detection is a crucial task. Medical imaging technique is actively developing field in engineering. MRI(Magnetic Resonance Imaging) is one those reliable imaging techniques on which medical diagnostic treatment thereof is based upon.

Manual inspection of those images is a tedious job as the amount of data and minute details are hard to recognize by the humans and often lacks calibration i.e. varies from individual to individual. For this reason automating those techniques is very crucial. The MRI deals with the complicated problem of brain tumor detection and every MRI is unique, due to its complexity and variance getting accuracy is a challenge. Using Boosting algorithms we can substantially improve our accuracy and eliminate the chance of misclassification which can save many lives.

Reference:http://www.rroij.com/open-access/detection-and-segmentation-of-brain-tumors-usingadaboost-svm.pdf

CASE STUDY- US- Heart Patient

Five million Americans are currently living with heart diseases, and the numbers are expected to rise. It is very important to understand the factors which causes Heart-attacks so that certain precaution can be taken by individuals.

In-order to understand the reasons of the Heart-attack, a data was collected from various hospitals across US which is given in US_Heart_Patients.csv. In the data set there are Heart-Att indicates whether the person suffered from Heart attack or not.

Perform EDA on the data and build various models and compare the output to predict whether the person will suffer from Heart-attack or not.

Apply **Data Science at your workplace** to gain some instant benefits:

- Get noticed by your management with your outstanding analysis backed by data science.

- Create an impact in your organization by taking up small projects/initiatives to solve critical issues using data science.

- Network with members from the data science vertical of your organization and seek opportunities to contribute in small projects.

- Share your success stories with us and the world to position yourself as a subject matter expert in data science.

ANY QUESTIONS

HAPPY LEARNING

Excelerate CareerPrep:
Career services by Great Learning.

The CareerPrep is an organized way to learn about all of the topics asked in interviews for Data Science jobs.

- Complete Excelerate career prep and learn how to crack an interview with ease.

- Understand the different kind of roles available in the data science ecosystem along with their skill set requirements

- Build a resume that has a higher probability of getting shortlisted by recruiters

- Prepare for different data science roles by reviewing 400+ questions asked in actual interviews

- Understand and implement frameworks to solve guesstimates, logical and aptitude problems often used in interviews

- Learn from the experience of alumni who have successfully transitioned

- Speed up your path towards a transition to a Data science role.

Complete Career Prep today to speed up your path towards DS transition. You'll have a 40% better chance of getting hired than other candidates.

Note: CareerPrep is mandatory step in unlocking the job board.