LIVERPOOL JOHN MOORES UNIVERSITY

International Institute of Information Technology Bangalore

upGrad

Master of Science in

Data Science

Now integrated with **Generative AI**

Table of Contents

The Era Of Generative AI

Usually, this first page is reserved for “About upGrad”. But the world is at the cusp of Generative AI rapidly changing the world as we know it. At upGrad, we’ve always believed in imparting learners the skills necessary to thrive in the fast-evolving world of technology. We are hence quite thrilled to pioneer Generative AI as an elective in the Master of Science in Data Science.

With this key inclusion of Generative AI, learners will delve deeper into the fascinating realm of using Data Science to build practical applications like conversational AI chat bots, image creators, and content recommenders amongst others, to solve real-world challenges. So dive into this brave new world of Generative AI and Large Language Models with us, and watch yourself transform into a 10x Data Scientist.

“IIIT Bangalore prides itself in constantly updating cutting-edge topics to its curriculum. Our faculty has shaped this exciting Generative AI elective along with upGrad’s industry experts, thus ensuring both academic rigour as well as incorporating the latest advancements in tech.”

Dr. V. Sridhar,
Head-Faculty, IIITB

“As an organisation that asks professionals to stay updated with the latest skills, we had to be one of the first to teach Generative AI. With this move, we are excited to witness the impact that Generative AI will have on the future, as well as the value our learners will bring to the field with this essential skill.”

Mayank Kumar, Co-founder & MD
upGrad

About upGrad

upGrad has delivered over 20 million hours of learning, delivering programs by collaborating with universities across the world including Liverpool John Moores University, IIIIT Bangalore and Deakin Business School among others.

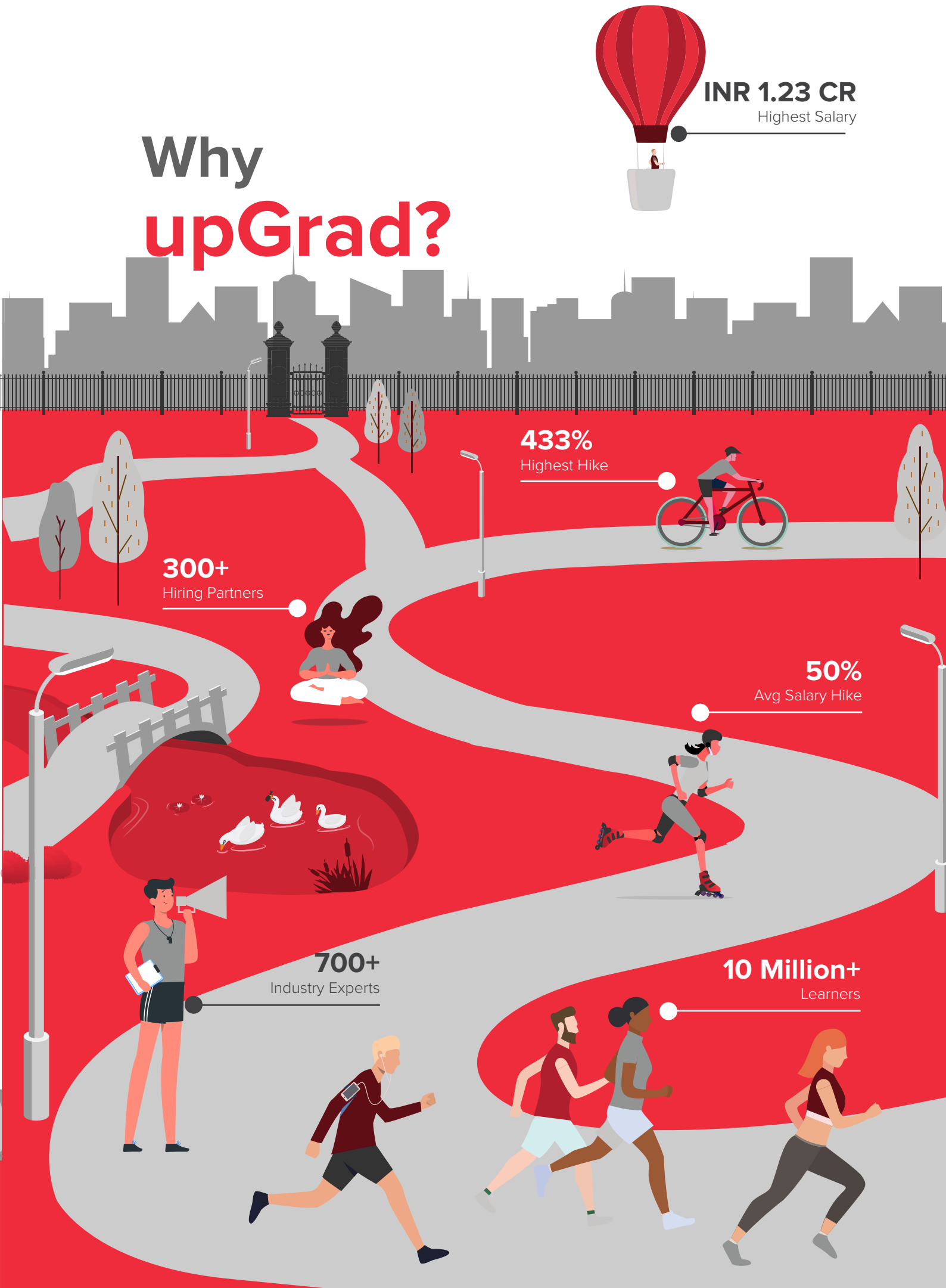
Online education is a fundamental disruption that will have a far-reaching impact. upGrad was founded taking this into consideration. upGrad is an online education platform to help individuals develop their professional potential in the most engaging learning environment.

Since its inception, upGrad has delivered over 20 million hours of learning, delivering programs by collaborating with universities across the world, including LJMU, IIIT Bangalore and Deakin Business School among others. And it doesn't end there.

The faculty includes an average of 15+ years of experience. The faculty covers the conceptual depths of topics such as Data Science, Machine Learning and AI, and Big Data Analytics. These will be complemented by industry-relevant case studies from major industry verticals by industry leaders with 8+ years of experience from upGrad's industry network.

Furthermore, our strong placement network, industry mentorship and the credibility of a Master's Degree will provide you with just the right push to accelerate your career in Data Science!

Why upGrad?


INR 1.23 CR
Highest Salary
433%
Highest Hike
300+
Hiring Partners
50%
Avg Salary Hike
700+
Industry Experts
10 Million+
Learners

Program Highlights

Dual Accreditation and Alumni Status
Get certified by IIITB and LJMU, UK and gain dual alumni status on successful completion of the program along with access to LJMU's digital library.

Programming Language & Tools
Learn 5+ Programming Languages and Tools like Python, Tableau, MySQL and more. Optional modules for further upskilling.building, career fairs, industry mentors and much more.

For the Industry, by the Industry
Learn from 60+ case studies and industry experts who mentor you throughout the program.

5 Specialisations
Choose from 5 specialisation-son the basis of your back-ground and career aspirations and get the learning you want.

Live Classroom Session
Live Classroom hour with Dr Manoj Jayabalan, Post-Doctoral Fellow at LJMU, to solve queries related to dissertation.

Global Access to Jobs
With 360-degree career support and dual alumni status, gain global access to jobs.

Faculty and Industry Experts

Dr. Debabrata Das
Director, IIITB

Dr. Debabrata Das is Director of IIITB. He has received his PhD from IIT-KGP. His main areas of research are IoT and Wireless Access Network.

Chandrashekar Ramanathan
Dean Academics, IIITB

Prof. Chandrashekar has a PhD from Mississippi State University and experience of over 10 years in several multinational organisations.

S. Anand
CEO, Gramener

A gold medallist from IIM Bangalore, an alumnus of IIT Madras and London Business School, Anand is among the top 10 data scientists in India with 20 years of experience.

Tricha Anjali
Ex-Associate Dean, IIIT-B

Prof Tricha has a Ph.D from Georgia Tech as well as an integrated M.Tech. from IIT Bombay. Her research interests include computer networks.

Behzad Ahmadi
Data Scientist Walmart Labs

An M. Tech graduate and PhD from Jersey Institute of Technology, Behzad possesses tremendous years of experience in Data Science and ML.

Anshuman Gupta
Director - Data Science, Pitney Bowes

He has a PhD (Dual) from Penn State University as well as a BTech Degree from IIT Bombay.

Prof. G. Srinivasaraghavan
Professor, IIITB

Prof. Srinivasaraghavan has a PhD in Computer Science from IIT-K and 18 years of experience with Infosys and several other MNCs.

Mirza Rahim Baig
Ex- Lead Analyst, Flipkart

Mirza is a veteran professional with 10+ years of experience in applications of data science, machine learning in e-commerce and healthcare.

Sajan Kedia
Ex- Data Science Lead, Myntra

Sajan graduated from IIT, BHU and has tons of experience in Data Science, Big Data, Spark, Machine Learning and Natural Language Processing.

Rajesh Sabapathy
Sr Director, Data Science, UHG Group

Rajesh has 10+ years of experience leading Data Science teams in various domains solving complex problems using Deep Learning & ML technique.

Prof. Dhiya Al-Jumeily
The Head and Professor - AI, LJMU

A Senior Member of the IEEE and a Chartered IT Professional. He is a fellow of the UK Higher Education Academy.

Bijoy Kumar Khandelwal
COO, Actify Data Labs

Bijoy comes with a deep understanding of the private and cloud architectures and has helped numerous companies make the transition.

Ujjyaini Mitra
Head of Analytics, Zee5

An alumnus of McKinsey and Co, Flipkart and Bharati Airtel with over 11 years of experience.

Ankit Jain
ML Engineering Manager, Meta

An alumnus of IIT Bombay, UCB, and HBS with over 9 years of experience. Ankit has been recognised as 40 Under40 Data Scientist for 2022.

Dr. Atif Waraich
Faculty - Computer Science,
LJMU

A Senior Faculty of Engineering and Technology at LJMU who has multiple publications in the healthcare domain.

Prof. Paulo Lisboa
Head of Dept - Applied Mathematics,
LJMU - Retired

Studied Mathematical Physics at LU and was the chairman of Industrial Mathematics at LJMU in 1996 and Head of Graduate School in 2002.

Dr Gabriela Czanner
Faculty - Engineering and Technology,
LJMU

A Senior Lecturer in Statistics and Data Science at the Department of Applied Mathematics at LJMU. Her research focus is Advanced Statistics for Decision Support.

Dr. Manoj jayabalan
Faculty of Engineering and Technology,
Liverpool John Moores University

Dr. Ahmed Kaky
Faculty of Engineering and Technology,
Liverpool John Moores University

upGrad Learning Experience

Student Support Team

- We have a dedicated/ Student Support Team for handling your queries via email or call-back requests
- This support team is available 7 days a week, 24 hours a day

Industry Mentors

- Receive unparalleled guidance from industry mentors, teaching assistants and graders
- Receive one-on-one feedback on sub-missions and personalised feedbacks on improvement

Industry Networking

- Live sessions by experts on various industry topics
- One-on-one discussion and feedback sessions with industry mentors

upGrad BaseCamp

- Fun-packed, informative and career building workshop sessions by industry professionals and professors
- Group activities with your peers and alumni

Expert Feedback

- Personalised expert feedback on assignments and projects
- Regular live sessions by experts to clarify concept-related doubts

Q&A Forum

- Timely doubt resolution by industry experts and peers
- 100% expert-verified responses to ensure quality learning

New Additions

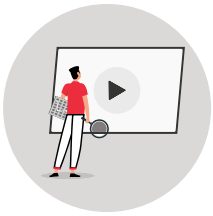
Career Essential Soft-skills Program

- Excel your personal & professional life with upGrad's Soft Skills Program
- Study Three fundamental Skills - Interview & Job Search, Corporate & Business Communication and Problem Solving
- Get access to 40+ learner hours of soft skills content delivered by the best faculty & Industry experts

30-Hour Programming Bootcamp for Non-tech Learners

- Non-tech background? No need to fear Programming anymore
- A 30-hour Python Programming bootcamp, focusing on developing Basic + Intermediate Python Programming Concepts to assist non-tech learners
- A blended learning experience delivered via Interactive live sessions and assessments

Industry Projects



IMDb Movie Analysis

Uber Supply-Demand Gap

Lead Scoring

Fraud Detection

Creditworthiness of Customers

Speech Recognition

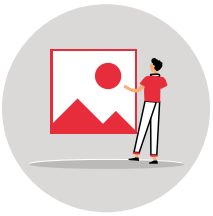
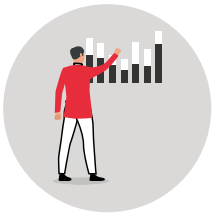


Image Captioning

Social Media Listening

Telecom Churn



Interactive Market Campaign Analysis

Retail Giant Sales Forecasting

And many more!

Learning Path

Preparatory Course
0 week

Data Toolkit
12 weeks

Machine Learning
10 weeks

Choose any of the 5 Specialisations
22 weeks (with 4 weeks of Capstone)

Natural Language Processing
Tools: ChatGPT, OpenAI API, Dall-E, Midjourney, Copilot, Flask

Deep Learning
Tools: ChatGPT, OpenAI API, Dall-E, Midjourney, Copilot, Flask

Business Analytics
Tools: ChatGPT, OpenAI API, Dall-E, Midjourney, Copilot, Flask

Business Intelligence/ Data Analyics
Tools: Python, Power BI, Excel, mySQL, MongoDB, Shiny, Tableau

Data Engineering
Tools: Hadoop, HBase, Sqoop, Hive, Flume, PySpark, Spark, Airflow

Executive PG Programme in Data Science (Natural Language Processing)

Executive PG Programme in Data Science (Deep Learning)

Executive PG Programme in Data Science (Business Analytics)

Executive PG Programme in Data Science (Business Intelligence/ Data Analytics)

Executive PG Programme in Data Science (Data Engineering)

Research Methodology

Dissertation

MSc - LJMU (Natural Language Processing)

MSc - LJMU (Deep Learning)

MSc - LJMU (Business Analytics)

MSc - LJMU (Business Intelligence/ Data Analytics)

MSc - LJMU (Data Engineering)

Executive PG Programme in Data Science

COMMON CURRICULUM

PRE-PROGRAM PREPARATORY CONTENT

1. DATA ANALYSIS IN EXCEL

1. **INTRODUCTION TO EXCEL**
2. **DATA ANALYSIS IN EXCEL - I: FUNCTIONS, FORMULAE, AND CHARTS**
3. **DATA ANALYSIS IN EXCEL - II: PIVOTS AND LOOKUPS**

Taught by one of the most renowned data scientists in the country (S.Anand, CEO, Gramener), this module takes you from a beginner-level Excel user to an almost professional user.

---

2. ANALYTICS PROBLEM SOLVING

1. **THE CRISP-DM FRAMEWORK - BUSINESS AND DATA UNDERSTANDING**
2. **CRISP-DM FRAMEWORK - DATA PREPARATION, MODELLING, EVALUATION AND DEPLOYMENT**

This module covers concepts of the CRISP-DM framework for business problem-solving.

COURSE 1: DATA TOOLKIT

1. INTRODUCTION TO PYTHON

1. **UNDERSTANDING THE UPGRAD CODING CONSOLE**
2. **BASICS OF PYTHON**
3. **DATA STRUCTURES IN PYTHON**
4. **CONTROL STRUCTURE AND FUNCTIONS IN PYTHON**
5. **OOP IN PYTHON**

Build a foundation for the most in-demand programming language of the 21st century.

2 WEEKS

**The Curriculum is subject to change as per the inputs from university or industry experts*

2. PROGRAMMING IN PYTHON

1. **LOGIC AND SYNTAX BUILDING**
2. **DATA STRUCTURES: LISTS, STRINGS, DICTIONARIES, AND STACKS**
3. **TIME COMPLEXITY**
4. **SEARCHING AND SORTING**
5. **TWO POINTERS**
6. **RECURSION**

Learn how to approach and solve logical problems using programming.

1 WEEK

---

3. PYTHON FOR DATA SCIENCE

1. **INTRODUCTION TO NUMPY**
2. **INTRODUCTION TO MATPLOTLIB**
3. **INTRODUCTION TO PANDAS**
4. **GETTING AND CLEANING DATA**

Learn how to manipulate datasets in Python using Pandas which is the most powerful library for data preparation and analysis.

1 WEEK

---

4. DATA VISUALISATION IN PYTHON

1. **INTRODUCTION TO DATA VISUALISATION**
2. **DATA VISUALISATION USING SEABORN**

Humans are visual learners, and hence no task related to data is complete without visualisation. Learn to plot and interpret various graphs in Python and observe how they make data analysis and drawing insights easier.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

5. EXPLORATORY DATA ANALYSIS

1. **DATA SOURCING**
2. **DATA CLEANING**
3. **UNIVARIATE ANALYSIS**
4. **BIVARIATE ANALYSIS AND MULTIVARIATE ANALYSIS**

Learn how to find and analyse the patterns in the data to draw actionable insights.

1 WEEK

---

6. CREDIT EDA CASE STUDY

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Solve a real industry problem through the concepts learnt in exploratory data analysis.

1 WEEK

---

7. INFERENTIAL STATISTICS

1. **BASICS OF PROBABILITY**
2. **DISCRETE PROBABILITY DISTRIBUTIONS**
3. **CONTINUOUS PROBABILITY DISTRIBUTIONS**
4. **CENTRAL LIMIT THEOREM**

Build a strong statistical foundation and learn how to 'infer' insights from a huge population using a small sample.

1 WEEK

---

8. HYPOTHESIS TESTING

1. **CONCEPTS OF HYPOTHESIS TESTING - I: NULL AND ALTERNATE HYPOTHESIS, MAKING A DECISION, AND CRITICAL VALUE METHOD**
2. **CONCEPTS OF HYPOTHESIS TESTING - II: P-VALUE METHOD AND TYPES OF ERRORS**
3. **INDUSTRY DEMONSTRATION OF HYPOTHESIS TESTING: TWO-SAMPLE MEAN AND PROPORTION TEST, A/B TESTING**

Understand how to formulate and validate hypotheses for a population to solve real-life business problems.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

9. DATA ANALYSIS USING SQL

1. **DATABASE DESIGN**
2. **DATABASE CREATION IN MYSQL WORKBENCH**
3. **QUERYING IN MYSQL**
4. **JOINS AND SET OPERATIONS**

Data in companies is definitely not stored in excel sheets! Learn the fundamentals of databases and extract information from RDBMS using the structured query language.

1 WEEK

---

10. ADVANCED SQL & BEST PRACTICES

1. **WINDOW FUNCTIONS**
2. **CASE STATEMENTS, STORED ROUTINES AND CURSORS**
3. **QUERY OPTIMISATION AND BEST PRACTICES**
4. **PROBLEM-SOLVING USING SQL**

Apply advanced SQL concepts like windowing and procedures to derive insights from data and answer pertinent business questions.

1 WEEK

---

11. SQL ASSIGNMENT: RSVP MOVIES

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

In this assignment, you will work on a movies dataset using SQL to extract exciting insights.

1 WEEK

COURSE 2 - MACHINE LEARNING I

1. LINEAR REGRESSION

1. **SIMPLE LINEAR REGRESSION**
2. **SIMPLE LINEAR REGRESSION IN PYTHON**
3. **MULTIPLE LINEAR REGRESSION**
4. **MULTIPLE LINEAR REGRESSION IN PYTHON**
5. **INDUSTRY RELEVANCE OF LINEAR REGRESSION**

Venture into the machine learning community by learning how one variable can be predicted using several other variables through a housing dataset where you will predict the prices of houses based on various factors.

2 WEEKS

**The Curriculum is subject to change as per the inputs from university or industry experts*

2. LINEAR REGRESSION ASSIGNMENT

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Build a model to understand the factors on which the demand for bike-sharing systems vary on and help a company optimise its revenue.

1 WEEK

---

3. LOGISTIC REGRESSION

1. **UNIVARIATE LOGISTIC REGRESSION**
2. **MULTIVARIATE LOGISTIC REGRESSION: MODEL BUILDING AND EVALUATION**
3. **LOGISTIC REGRESSION: INDUSTRY APPLICATIONS**

Learn your first binary classification technique by determining which telecom operator customers are likely to churn versus those who are not to help the business retain customers.

2 WEEKS

---

4. CLASSIFICATION USING DECISION TREES

1. **INTRODUCTION TO DECISION TREES**
2. **ALGORITHMS FOR DECISION TREES CONSTRUCTION**
3. **HYPERPARAMETER TUNING IN DECISION TREES**

Learn how the human decision-making process can be replicated using a decision tree and tune it to suit your needs.

1 WEEK

---

5. UNSUPERVISED LEARNING: CLUSTERING

1. **INTRODUCTION TO CLUSTERING**
2. **K-MEANS CLUSTERING**
3. **HIERARCHICAL CLUSTERING**
4. **OTHER FORMS OF CLUSTERING: K-MODE, K-PROTOTYPE, DB SCAN**

Learn how to group elements into different clusters when you don't have any pre-defined labels to segregate them through K-means clustering, hierarchical clustering, and more.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

6. BASICS OF NLP AND TEXT MINING

1. **REGEX AND INTRODUCTION TO NLP**
2. **BASIC LEXICAL PROCESSING**
3. **ADVANCED LEXICAL PROCESSING**

Do you get annoyed by the constant spam in your mailbox? Wouldn't it be nice if we had a program to check your spelling? In this module learn how to build a spell checker & spam detector using techniques like phonetic hashing, bag-of-words, TF-IDF, etc.

---

5. BUSINESS PROBLEM SOLVING

1. **INTRODUCTION TO BUSINESS PROBLEM SOLVING**
2. **BUSINESS PROBLEM SOLVING: CASE STUDY DEMONSTRATIONS**

Learn how to approach open-ended real-world problems using data as a lever to draw actionable insights.

---

7. CASE STUDY: LEAD SCORING

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Help the Sales team of your company identify which leads are worth pursuing through this classification case study.

**The Curriculum is subject to change as per the inputs from university or industry experts*

SPECIALISATION: DEEP LEARNING

COURSE 3 - MACHINE LEARNING II

1. BAGGING & RANDOM FOREST

1. **POPULAR ENSEMBLES**
2. **INTRODUCTION TO RANDOM FORESTS**
3. **FEATURE IMPORTANCE IN RANDOM FORESTS**
4. **RANDOM FORESTS IN PYTHON**

Learn how powerful ensemble algorithms can improve your classification models by building random forests from decision trees.

1 WEEK

---

2. BOOSTING

1. **INTRODUCTION TO BOOSTING AND ADABOOST**
2. **GRADIENT BOOSTING**

Learn about ensemble modelling through bagging and boosting and, understand how weak algorithms can be transformed into stronger ones.

1 WEEK

---

3. MODEL SELECTION & GENERAL ML TECHNIQUES

1. **PRINCIPLES OF MODEL SELECTION**
2. **MODEL EVALUATION**
3. **MODEL SELECTION: BEST PRACTICES**

Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, along with general machine learning techniques like feature engineering, model evaluation, and many more.

1 WEEK

---

4. PRINCIPAL COMPONENT ANALYSIS

1. **PRINCIPAL COMPONENT ANALYSIS AND SINGULAR VALUE DECOMPOSITION**
2. **PRINCIPAL COMPONENT ANALYSIS IN PYTHON**

Understand important concepts related to dimensionality reduction, the basic idea and the learning algorithm of PCA, and its practical applications on supervised and unsupervised problems.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

5. ADVANCED REGRESSION

1. GENERALISED LINEAR REGRESSION
2. REGULARISED REGRESSION

In this module, take a more advanced look at regression models and learn the concepts related to regularisation.

1 WEEK

---

6. TIME SERIES FORECASTING (OPTIONAL)

1. INTRODUCTION TO TIME SERIES AND ITS COMPONENTS
2. WORKING WITH STATIONARY TIME SERIES
3. END-TO-END ANALYSIS OF TIME SERIES

In this module, you will learn how to analyse and forecast a series that varies with time.

0 WEEK

---

7. ADVANCED ML CASE STUDY

1 WEEK

1. PROBLEM STATEMENT
2. EVALUATION RUBRIC
3. FINAL SUBMISSION
4. SOLUTION

Build a regularized regression model to understand the most important variables to predict house prices in Australia.

COURSE 4 - ADVANCED MACHINE LEARNING AND DEEP LEARNING

1. INTRODUCTION TO NEURAL NETWORKS AND ANN

1. STRUCTURE OF NEURAL NETWORKS
2. FEED FORWARD IN NEURAL NETWORKS
3. BACKPROPAGATION IN NEURAL NETWORKS
4. MODIFICATIONS TO NEURAL NETWORKS
5. HYPERPARAMETER TUNING IN NEURAL NETWORKS

Learn the most sophisticated and cutting-edge technique in machine learning - Artificial Neural Networks or ANNs.

2 WEEKS

**The Curriculum is subject to change as per the inputs from university or industry experts*

2. CONVOLUTIONAL NEURAL NETWORKS

1. **INTRODUCTION TO CONVOLUTIONAL NEURAL NETWORKS**
2. **BUILDING CNNS WITH PYTHON AND KERAS**
3. **CNN ARCHITECTURES AND TRANSFER LEARNING**
4. **STYLE TRANSFER AND OBJECT DETECTION**

Learn the basics of CNN and OpenCV and how to classify image data using various architectures which you will then implement using Python and Keras.

1 WEEK

---

3. CONVOLUTIONAL NEURAL NETWORKS -INDUSTRY APPLICATIONS

1. **INDUSTRY DEMONSTRATION: USING CNNS WITH FLOWERS IMAGES**
2. **INDUSTRY DEMONSTRATION: USING CNNS WITH X-RAY IMAGES**

Apply CNNs to Computer Vision tasks like detecting anomalies in chest X-Ray scans.

1 WEEK

---

4. OBJECT DETECTION & IMAGE SEGMENTATION

1. **FUNDAMENTALS OF OBJECT DETECTION**
2. **REGION-BASED DETECTORS**
3. **ONE-SHOT DETECTORS**
4. **CUSTOM OBJECT DETECTION**
5. **SEMANTIC SEGMENTATION**

Learn the applications of DL in computer vision through industry-relevant detection algorithms such as RCNNs, YOLO and SSD.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

5. RECURRENT NEURAL NETWORKS (OPTIONAL)

1. **WHAT MAKES A NEURAL NETWORK RECURRENT**
2. **VARIANTS OF RNNS: BIDIRECTIONAL RNNS AND LSTMS**
3. **BUILDING RNNS IN PYTHON**

Ever wondered what goes behind machine translation, sentiment analysis, and speech recognition? Learn how RNN helps in areas having sequential data like text, speech, videos, and a lot more.

1 WEEK

---

6. GESTURE RECOGNITION

1. **TWO ARCHITECTURES: 3D CONVS AND CNN-RNN STACK**
2. **UNDERSTANDING GENERATORS**
3. **STARTER CODE WALKTHROUGH**
4. **PROBLEM STATEMENT AND FINAL SUBMISSION**

Make a Smart TV system which can control the TV with the user's hand gestures as the remote control

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 5 - GENERATIVE AI

1. FUNDAMENTALS OF TRANSFORMERS ARCHITECTURE, GENERATIVE AI, CHATGPT & PROMPT ENGINEERING USING NON REASONING, CHAIN OF THOUGHT & ADVANCED TECHNIQUES	1 WEEK
2. PRODUCT DEVELOPMENT USING OPENAI APIS, FINE TUNING USING STAR TECHNIQUE IN PYTHON	1 WEEK
3. INTEGRATING SPEECH USING WHISPER API AND APPLICATION DEPLOYMENT USING FLASK	1 WEEK
4. FUNDAMENTALS OF DESIGN, PHOTOGRAPHY, PRODUCT DEVELOPMENT USING STABLE DIFFUSION IN PYTHON & CREATE PIXXELCRAFT AI TO ENABLE FAST-TRACK DIGITISATION FOR OFFLINE E-COMMERCE BUSINESSES BY GENERATING HIGH-QUALITY IMAGES AI FOR A LARGE PRODUCT PORTFOLIO	1 WEEK
5. APPLICATIONS OF LLMS IN DATA SCIENCE PROJECTS & AUTOMATING NEWS RECOMMENDATION USING GPT3 AND COPILOT POWERED MACHINE LEARNING APPLICATIONS OF LLMS	1 WEEK
6. INTERVIEW GYNIE AI: CHATBOT DEVELOPMENT PROJECT	1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 6 - CAPSTONE PROJECT

CAPSTONE PROJECT

1. **AN OVERVIEW OF THE DOMAIN AND ASSOCIATED CONCEPTS**
2. **PROBLEM STATEMENT**
3. **EVALUATION RUBRIC**
4. **MID SUBMISSION**
5. **FINAL SUBMISSION**
6. **SOLUTION**

Choose from a range of real-world industry-woven projects on advanced topics like Recommendation Systems, Fraud Detection, Emotion Detection from faces, Social Media Listening, and Speech Recognition among many others.

4 WEEKS

SPECIALISATION: NATURAL LANGUAGE PROCESSING

COURSE 3 - MACHINE LEARNING II

1. BAGGING & RANDOM FOREST

1. **POPULAR ENSEMBLES**
2. **INTRODUCTION TO RANDOM FORESTS**
3. **FEATURE IMPORTANCE IN RANDOM FORESTS**
4. **RANDOM FORESTS IN PYTHON**

Learn how powerful ensemble algorithms can improve your classification models by building random forests from decision trees.

1 WEEK

---

2. BOOSTING

1. **INTRODUCTION TO BOOSTING AND ADABOOST**
2. **GRADIENT BOOSTING**

Learn about ensemble modelling through bagging and boosting, and understand how weak algorithms can be transformed into stronger ones.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

3. MODEL SELECTION & GENERAL ML TECHNIQUES

1. **PRINCIPLES OF MODEL SELECTION**
2. **MODEL EVALUATION**
3. **MODEL SELECTION: BEST PRACTICES**

Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, along with general machine learning techniques like feature engineering, model evaluation, and many more.

1 WEEK

---

4. PRINCIPAL COMPONENT ANALYSIS

1. **PRINCIPAL COMPONENT ANALYSIS AND SINGULAR VALUE DECOMPOSITION**
2. **PRINCIPAL COMPONENT ANALYSIS IN PYTHON**

Understand important concepts related to dimensionality reduction, the basic idea and the learning algorithm of PCA, and its practical applications on supervised and unsupervised problems.

1 WEEK

---

5. ADVANCED REGRESSION

1. **GENERALISED LINEAR REGRESSION**
2. **REGULARISED REGRESSION**

In this module, take a more advanced look at regression models and learn the concepts related to regularisation.

1 WEEK

---

6. TIME SERIES ANALYSIS (OPTIONAL)

1. **INTRODUCTION TO TIME SERIES AND ITS COMPONENTS**
2. **WORKING WITH STATIONARY TIME SERIES**
3. **END-TO-END ANALYSIS OF TIME SERIES**

In this module, you will learn how to analyse and forecast a series that varies with time.

2 WEEKS

**The Curriculum is subject to change as per the inputs from university or industry experts*

7. ADVANCED ML CASE STUDY

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Build a regularised regression model to understand the most important variables to predict house prices in Australia.

1 WEEK

COURSE 4 - ADVANCED MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

1. NEURAL NETS FOR NLP

1. **UNDERSTANDING NEURAL NETWORKS**
2. **LOSS FUNCTIONS AND BACK PROPAGATION**
3. **UNDERSTANDING TENSORFLOW**
4. **CASE STUDY: IMDB MOVIE REVIEW CLASSIFICATION**

Learn the most sophisticated and cutting-edge technique in machine learning - Artificial Neural Networks or ANNs.

1 WEEK

---

2. SYNTACTIC PROCESSING

1. **INTRODUCTION TO SYNTACTIC PROCESSING**
2. **PARSING**
3. **INFORMATION EXTRACTION**
4. **CONDITIONAL RANDOM FIELDS**

Learn how to analyse the syntax or the grammatical structure of sentences using POS tagging and Dependency parsing.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

3. SYNCTACTIC PROCESSING

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Use the techniques such as POS tagging and Dependency parsing to extract information from unstructured text data.

1 WEEK

---

4. SEMANTIC PROCESSING

1. **INTRODUCTION TO SEMANTIC PROCESSING**
2. **DISTRIBUTIONAL SEMANTICS**
3. **INDUSTRY APPLICATIONS OF DISTRBUTIONAL SEMANTICS**
4. **TOPIC MODELLING**

Learn the most interesting area in the field of NLP and understand different techniques like word-embeddings and topic modelling to build an application that extracts opinions about socially relevant issues.

2 WEEKS

---

5. APPLIED DL IN NLP

1. **INTRODUCTION TO MACHINE TRANSLATION**
2. **ATTENTION-BASED NMT MODEL**
3. **CUSTOM MODEL BUILDING IN TENSORFLOW**

Apply the concepts of DL in natural language processing problems through encoder-decoder architecture and NMTs, and implement them in TensorFlow.

1 WEEK

---

6. CASE STUDY: AUTOMATIC TICKET CLASSIFICATION

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Categorise support tickets with the help of Unsupervised learning and Topic modelling.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 5 - GENERATIVE AI

1. FUNDAMENTALS OF TRANSFORMERS ARCHITECTURE, GENERATIVE AI, CHATGPT & PROMPT ENGINEERING USING NON REASONING, CHAIN OF THOUGHT & ADVANCED TECHNIQUES	1 WEEK
2. PRODUCT DEVELOPMENT USING OPENAI APIS, FINE TUNING USING STAR TECHNIQUE IN PYTHON	1 WEEK
3. INTEGRATING SPEECH USING WHISPER API AND APPLICATION DEPLOYMENT USING FLASK	1 WEEK
4. FUNDAMENTALS OF DESIGN, PHOTOGRAPHY, PRODUCT DEVELOPMENT USING STABLE DIFFUSION IN PYTHON & CREATE PIXXELCRAFT AI TO ENABLE FAST-TRACK DIGITISATION FOR OFFLINE E-COMMERCE BUSINESSES BY GENERATING HIGH-QUALITY IMAGES AI FOR A LARGE PRODUCT PORTFOLIO	1 WEEK
5. APPLICATIONS OF LLMS IN DATA SCIENCE PROJECTS & AUTOMATING NEWS RECOMMENDATION USING GPT3 AND COPILOT POWERED MACHINE LEARNING APPLICATIONS OF LLMS	1 WEEK
6. INTERVIEW GYNIE AI: CHATBOT DEVELOPMENT PROJECT	1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 6 - CAPSTONE PROJECT

1. CAPSTONE PROJECT

1. **AN OVERVIEW OF THE DOMAIN AND ASSOCIATED CONCEPTS**
2. **PROBLEM STATEMENT**
3. **EVALUATION RUBRIC**
4. **MID SUBMISSION**

Choose from a range of real-world industry-woven projects on advanced topics like Recommendation Systems, Fraud Detection, Emotion Detection from faces, Social Media Listening, and Speech Recognition among many others.

4 WEEKS

SPECIALISATION: BUSINESS ANALYTICS

COURSE 3 - ADVANCED MACHINE LEARNING

1. BAGGING & RANDOM FOREST

1. **POPULAR ENSEMBLES**
2. **INTRODUCTION TO RANDOM FORESTS**
3. **FEATURE IMPORTANCE IN RANDOM FORESTS**
4. **RANDOM FORESTS IN PYTHON**

Learn how powerful ensemble algorithms can improve your classification models by building random forests from decision trees.

1 WEEK

---

2. MODEL SELECTION & GENERAL ML TECHNIQUES

1. **PRINCIPLES OF MODEL SELECTION**
2. **MODEL BUILDING AND EVALUATION**
3. **FEATURE ENGINEERING**
4. **CLASS IMBALANCE**

Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, along with general machine learning techniques like feature engineering, model evaluation, and many more.

2 WEEKS

**The Curriculum is subject to change as per the inputs from university or industry experts*

3. TIME SERIES FORECASTING

1. **INTRODUCTION TO TIME SERIES AND ITS COMPONENTS**
2. **SMOOTHING TECHNIQUES**
3. **INTRODUCTION TO AR MODELS**
4. **BUILDING AR MODELS**

In this module, you will learn how to analyse and forecast a series that varies with time.

2 WEEKS

---

4. MODEL SELECTION CASE STUDY

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Apply your business acumen to the newly learnt machine learning techniques, and select the right model most appropriate for a provided business scenario.

1 WEEK

COURSE 4 - DATA VISUALISATION AND STORYTELLING

1. VISUALISATION USING TABLEAU

1. **DATA EXPLORATION IN TABLEAU**
2. **VISUALISING AND ANALYSING DATA IN TABLEAU WITH BASIC PLOTS**

Learn basic visualisation techniques using the most in-demand visualisation tool in the industry.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

2. ADVANCED EXCEL

1. **EXCEL FUNCTIONS**
2. **DATA ANALYSIS IN EXCEL**
3. **ADVANCED TOOLS AND VISUALISATIONS**

Learn the advanced concepts in Excel and start to perform data analysis like a pro!

1 WEEK

---

3. VISUALISATION USING POWERBI

1. **POWERBI: INTRODUCTION AND SETUP**
2. **VISUALISING AND ANALYSING DATA IN POWERBI**
3. **DATA TRANSFORMATIONS USING POWERBI**

Take your visualisation game a step forward by understanding how to operate PowerBI.

1 WEEK

---

4. STRUCTURED PROBLEM SOLVING USING FRAMEWORKS

1. **INTRODUCTION TO STRUCTURED PROBLEM SOLVING**
2. **INTERVIEWING AND FRAMEWORKS - I: 5W AND 5WHYS**
3. **INTERVIEWING AND FRAMEWORKS - II: SPIN**
4. **INDUSTRY DEMONSTRATIONS ON FRAMEWORKS**
5. **UNDERSTANDING BUSINESS MODEL CANVAS AND ISSUE TREE FRAMEWORK**
6. **INDUSTRY DEMONSTRATIONS ON ISSUE TREE FRAMEWORK**
7. **SPECIALISED FRAMEWORKS FOR BUSINESS PROBLEMS: 7PS, 5CS, ETC.**

Learn how to attack a business problem using various structured frameworks like 5W, 5WHYs, and SPIN.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

5. DATA STORYTELLING

1. **INTRODUCTION TO DATA STORYTELLING**
2. **COMPONENTS OF A GOOD STORY WITH DATA - UNDERSTANDING YOUR STAKEHOLDER AND STAKEHOLDER EMPATHY, LEVELS OF DETAILS FOR DIFFERENT STAKEHOLDERS - CXO/LEADERSHIP VS TEAM PRESENTATIONS, VISUALS, ETC.**
3. **GOLDEN RULES FOR DATA STORYTELLING**

Learn how to effectively strategise, communicate, and fine-grain your data analysis projects and understand how to optimally present your findings to technical and non-technical stakeholders and upgrade your storytelling skills.

1 WEEK

---

6. AIRBNB CASE STUDY

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Use your newly learnt UI tools skills to analyse an AirBnB dataset to make important business decisions. But the analysis is just a small part; can you also effectively present it using Data Storytelling to the right stakeholders?

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 5 - GENERATIVE AI

1. FUNDAMENTALS OF TRANSFORMERS ARCHITECTURE, GENERATIVE AI, CHATGPT & PROMPT ENGINEERING USING NON REASONING, CHAIN OF THOUGHT & ADVANCED TECHNIQUES	1 WEEK
2. PRODUCT DEVELOPMENT USING OPENAI APIS, FINE TUNING USING STAR TECHNIQUE IN PYTHON	1 WEEK
3. INTEGRATING SPEECH USING WHISPER API AND APPLICATION DEPLOYMENT USING FLASK	1 WEEK
4. FUNDAMENTALS OF DESIGN, PHOTOGRAPHY, PRODUCT DEVELOPMENT USING STABLE DIFFUSION IN PYTHON & CREATE PIXXELCRAFT AI TO ENABLE FAST-TRACK DIGITISATION FOR OFFLINE E-COMMERCE BUSINESSES BY GENERATING HIGH-QUALITY IMAGES AI FOR A LARGE PRODUCT PORTFOLIO	1 WEEK
5. APPLICATIONS OF LLMS IN DATA SCIENCE PROJECTS & AUTOMATING NEWS RECOMMENDATION USING GPT3 AND COPILOT POWERED MACHINE LEARNING APPLICATIONS OF LLMS	1 WEEK
6. INTERVIEW GYNIE AI: CHATBOT DEVELOPMENT PROJECT	1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 6 - CAPSTONE PROJECT

1. CAPSTONE PROJECT

1. **POWER BI - OPTIONAL**
2. **AN OVERVIEW OF THE DOMAIN AND ASSOCIATED CONCEPTS**
3. **PROBLEM STATEMENT**
4. **EVALUATION RUBRIC**
5. **MID SUBMISSION**
6. **FINAL SUBMISSION**
7. **SOLUTION**

Solve an end-to-end real-life industry problem from a wide variety of domains.

4 WEEKS

SPECIALISATION: BUSINESS INTELLIGENCE / DATA ANALYTICS

COURSE 3: ADVANCED DBS AND BIG DATA ANALYTICS

1. DATA MODELLING

1. **DATABASE DESIGN RECAP**
2. **BUILDING BLOCKS OF DATA MODELLING**
3. **PROBLEM SOLVING USING DATA MODELLING**
4. **DATA MODELLING: OPTIONAL ASSIGNMENT**

In this module, you will learn and use data modelling on a dataset to solve a business problem.

1 WEEK

---

2. ADVANCED SQL AND BEST PRACTICES

1. **WINDOW FUNCTIONS**
2. **CASE STATEMENTS, STORED ROUTINES, AND CURSORS**
3. **QUERY OPTIMISATION AND BEST PRACTICES**
4. **PROBLEM SOLVING USING SQL**

Apply advanced SQL concepts like windowing and procedures to derive insights from data and answer pertinent business questions.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

3. INTRODUCTION TO BIG DATA AND CLOUD

1. **BIG DATA AND CLOUD COMPUTING**
2. **AMAZON WEB SERVICES**
3. **BIG DATA STORAGE AND PROCESSING - HADOOP**
4. **EMR CLUSTER IN AWS**

Understand the basics of big data and cloud and learn to work with an EMR cluster on a cloud-based service.

1 WEEK

---

4. ANALYTICS USING SPARK

1. **EXPLORATORY DATA ANALYSIS WITH PYSPARK**
2. **PREDICTIVE ANALYSIS WITH SPARK MLLIB**

Use PySpark to do EDA and Predictive Analysis using Spark's ML library.

2 WEEKS

---

5. BIG DATA CASE STUDY

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Use your analytics skills to work on a large dataset in the cloud to solve an industry problem.

1 WEEK

COURSE 4 - DATA VISUALISATION AND STORYTELLING

1. VISUALISATION USING TABLEAU

1. **DATA EXPLORATION IN TABLEAU**
2. **VISUALISING AND ANALYSING DATA IN TABLEAU WITH BASIC PLOTS**

Learn basic visualisation techniques using the most in-demand visualisation tool in the industry.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

2. ADVANCED EXCEL

1. **EXCEL FUNCTIONS**
2. **DATA ANALYSIS IN EXCEL**
3. **ADVANCED TOOLS AND VISUALISATIONS**

Learn the advanced concepts in Excel and start to perform data analysis like a pro!

1 WEEK

---

3. VISUALISATION USING POWERBI

1. **POWERBI: INTRODUCTION AND SETUP**
2. **VISUALISING AND ANALYSING DATA IN POWERBI**
3. **DATA TRANSFORMATIONS USING POWERBI**

Take your visualisation game a step forward by understanding how to operate PowerBI.

1 WEEK

---

4. STRUCTURED PROBLEM SOLVING USING FRAMEWORKS

1. **INTRODUCTION TO STRUCTURED PROBLEM SOLVING**
2. **INTERVIEWING AND FRAMEWORKS - I: 5W AND 5WHYS**
3. **INTERVIEWING AND FRAMEWORKS - II: SPIN**
4. **INDUSTRY DEMONSTRATIONS ON FRAMEWORKS**
5. **UNDERSTANDING BUSINESS MODEL CANVAS AND ISSUE TREE FRAMEWORK**
6. **INDUSTRY DEMONSTRATIONS ON ISSUE TREE FRAMEWORK**
7. **SPECIALIZED FRAMEWORKS FOR BUSINESS PROBLEMS: 7PS, 5CS, ETC.**

Learn how to attack a business problem using various structured frameworks like 5W, 5WHYs, and SPIN.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

5. DATA STORYTELLING

1. **INTRODUCTION TO DATA STORYTELLING**
2. **COMPONENTS OF A GOOD STORY WITH DATA - UNDERSTANDING YOUR STAKEHOLDER AND STAKEHOLDER EMPATHY, LEVELS OF DETAILS FOR DIFFERENT STAKEHOLDERS - CXO/LEADERSHIP VS TEAM PRESENTATIONS, VISUALS, ETC.**
3. **GOLDEN RULES FOR DATA STORYTELLING**

Learn how to effectively strategise, communicate, and fine-grain your data analysis projects and understand how to optimally present your findings to technical and non-technical stakeholders and upgrade your storytelling skills.

1 WEEK

---

6. AIRBNB CASE STUDY

1. **PROBLEM STATEMENT**
2. **EVALUATION RUBRIC**
3. **FINAL SUBMISSION**
4. **SOLUTION**

Use your newly learnt UI tools skills to analyse an AirBnB dataset to make important business decisions. But the analysis is just a small part; can you also effectively present it using Data Storytelling to the right stakeholders?

1 WEEK

COURSE 5: ADVANCED PROBLEM SOLVING AND PROGRAMMING

1. DATA STRUCTURES - SETS, DICTIONARIES, STACKS, QUEUES

1. **IN-BUILT DATA STRUCTURES**
2. **STACK**
3. **QUEUE**
4. **TREES**

Learn user-defined data structures -Stack, Queue, and Trees in Python that help in advanced data manipulation.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

2. SEARCHING AND SORTING

1. **SEARCHING**
2. **SORTING**
3. **TWO POINTERS**

Learn most fundamental searching and sorting algorithms and design techniques

---

3. ALGORITHM ANALYSIS + RECURSION

1. **ALGORITHM ANALYSIS**
2. **TIME AND SPACE COMPLEXITY**
3. **RECURSION**

Learn how to assess the efficiency of your code using algorithm analysis techniques and learn to write recursive algorithms

---

4. ADVANCED DATABASE PROGRAMMING USING PANDAS

1. **ADVANCED DATA WRANGLING WITH PANDAS - I**
2. **ADVANCED DATA WRANGLING WITH PANDAS - II**

Learn and implement advanced wrangling functions and techniques in Pandas related to date-time, multi-columns aggregation, hierarchical indexing, and more.

---

5. PYTHON & SQL LAB

1. **SQL: TIMED TEST + ASSIGNMENT**
2. **PYTHON: TIMED TESTS I & II**
3. **VIDEO SUBMISSION**

In this competitive assignment, you will solve a variety of programming questions in both SQL and Python in a timed environment. You will also demonstrate one of the questions through a video submission to help improve your interviewing skills.

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 6 - CAPSTONE PROJECT

1. CAPSTONE PROJECT

1. **AN OVERVIEW OF THE DOMAIN AND ASSOCIATED CONCEPTS**
2. **PROBLEM STATEMENT**
3. **EVALUATION RUBRIC**
4. **MID SUBMISSION**
5. **FINAL SUBMISSION**
6. **SOLUTION**

Solve an end-to-end real-life industry problem from a wide variety of domains.

4 WEEKS

SPECIALISATION: DATA ENGINEERING

COURSE 3: DATA ENGINEERING - I

1. DATA MANAGEMENT AND RELATIONAL DATABASE MODELLING

1. **ENTERPRISE DATA MANAGEMENT**
2. **RELATIONAL DATABASE MODELLING**
3. **NORMAL FORMS AND ER DIAGRAMS**

Understand the concepts of Data Management and learn to model data from a Relational Database.

1 WEEK

---

2. INTRODUCTION TO BIG DATA(OPTIONAL)

1. **4VS OF BIG DATA**
2. **BIG DATA: INDUSTRY CASE STUDIES**

This module you will learn what big data is, its various characteristics, and its determining factors. You will also get an idea of the various sources of big data and the wide range of big data applications in different industries such as retail, healthcare, and finance.

0 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

3. INTRODUCTION TO CLOUD AND AWS SETUP

1. **INTRODUCTION TO CLOUD**
2. **AWS SETUP**

Understand what is cloud and setup your AWS account which will be required during the program.

1 WEEK

---

4. INTRODUCTION TO HADOOP AND MAPREDUCE PROGRAMMING

1. **CONCEPTS RETAILED TO DISTRIBUTED COMPUTING**
2. **HADOOP DISTRIBUTED FILE SYSTEM**
3. **MAPREDUCE PROGRAMMING IN PYTHON**

Understand the world of distributed data processing and storage with Hadoop. Learn to write MapReduce jobs in Python.

1 WEEK

---

5. ASSIGNMENT (OPTIONAL)

1. **INTRODUCTION, PROBLEM STATEMENT AND GRADING RUBRICS**

Solve an assignment to brush up on the skills learnt so far.

0 WEEK

---

6. NOSQL DATABASES AND APACHE HBASE NOSQL DATABASES AND MONGODB (OPTIONAL)

1. **CONCEPTS OF NOSQL DATABASES**
2. **INTRODUCTION TO APACHE HBASE**
3. **HBASE PYTHON API**
4. **COMPARISON OF NOSQL DATABASES**

Learn the concepts of NoSQL databases. Understand the working of Apache HBase.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

7. DATA WAREHOUSING (OPTIONAL)

1. **INTRODUCTION TO DATA WAREHOUSE AND DATA LAKES**
2. **DESIGNING DATA WAREHOUSING FOR AN ETL DATA PIPELINE**
3. **DESIGNING DATA LAKE FOR AN ETL DATA PIPELINE**

Understand the intricacies behind designing a data warehouse and a data lake for use case(s).

0 WEEK

---

8. DATA INGESTION WITH APACHE SQOOP AND APACHE FLUME

1. **INTRODUCTION TO DATA INGESTION**
2. **STRUCTURED DATA INGESTION WITH SQOOP**
3. **UNSTRUCTURED DATA INGESTION WITH FLUME**

Get familiar with the challenges involved in data ingestion. Use Sqoop and Flume to ingest structured and unstructured data into Hadoop.

1 WEEK

---

9. MAPREDUCE PROGRAMMING ASSIGNMENT

1. **PROBLEM STATEMENT AND SAMPLE DATASET**
2. **SOLUTION**

Practise MapReduce Programming on a Big Dataset.

1 WEEK

COURSE 4 - DATA ENGINEERING - II

1. HIVE & QUERYING

1. **FUNDAMENTALS OF APACHE HIVE**
2. **WRITING HQL FOR DATA ANALYSIS**
3. **PARTITIONING AND BUCKETING WITH HIVE**

Manage and query a data warehouse with Apache Hive. Learn to write optimised HQL for large-scale data analysis.

2 WEEKS

**The Curriculum is subject to change as per the inputs from university or industry experts*

2. ASSIGNMENT (OPTIONAL)

1. **INTRODUCTION, PROBLEM STATEMENT AND GRADING RUBRICS**	Solve an assignment to brush up the skills learnt so far.	**0 WEEK**

3. AMAZON REDSHIFT

1. **DATA WAREHOUSING WITH REDSHIFT** 2. **ANALYSE DATA WITH REDSHIFT**	Learn to deploy a Redshift cluster and use it for querying data.	**1 WEEK**

4. INTRODUCTION TO APACHE SPARK

1. **SPARK ARCHITECTURE** 2. **RDD, DATAFRAME API, SPARK SQL**	Get introduced to Apache Spark, a lighting fast big data processing engine.	**1 WEEK**

5. PROJECT: ETL DATA PIPELINE

1. **INTRODUCTION AND PROBLEM STATEMENT** 2. **GRADING RUBRICS AND SUBMISSION**	Make use of Sqoop, Redshift & Spark to design an ETL data pipeline.	**2 WEEKS**

6. AWS CLOUD INFRASTRUCTURE (OPTIONAL)

1. **THE AWS CLOUD PLATFORM** 2. **BUILDING AND DEPLOYING VIRTUAL MACHINES** 3. **AWS CLOUD STORAGE SOLUTIONS** 4. **APPLICATION DEPLOYMENT** 5. **CLOUD ADMINISTRATION AND SECURITY** 6. **LOAD BALANCING AND BACKUP STRATEGIES** 7. **CLOUD AUTOMATION**	Do a deep dive into AWS Cloud.	**0 WEEK**

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 5 - DATA ENGINEERING - III

1. OPTIMISING SPARK FOR LARGE-SCALE DATA PROCESSING

1. **RUNNING SPARK ON MULTINODE CLUSTER**
2. **SPARK MEMORY & DISK OPTIMISATION**
3. **OPTIMISING SPARK CLUSTER ENVIRONMENT**

Use PySpark to create large-scale data processing applications.

1 WEEK

---

2. APACHE FLINK(OPTIONAL)

1. **INTRODUCTION TO APACHE FLINK**
2. **BATCH DATA PROCESSING WITH FLINK**
3. **STREAM PROCESSING WITH APACHE FLINK**
4. **SQL API**

Get Introduced to Apache Flink and learn query batch data.

Use DataStream API to create a stream processing application.

0 WEEK

---

3. REAL-TIME DATA STREAMING WITH APACHE KAFKA

1. **INTRO TO REAL-TIME DATA PROCESSING ARCHITECTURES**
2. **FUNDAMENTALS OF APACHE KAFKA**
3. **SETTING UP KAFKA PRODUCER AND CONSUMER**
4. **KAFKA CONNECT API & KAFKA STREAMS**

Understand the producer-consumer architecture of Apache Kafka. Learn to set up a Kafka cluster for managing real-time data.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

4. REAL-TIME DATA PROCESSING USING SPARK STREAMING

1. **SPARK STREAMING ARCHITECTURE**
2. **SPARK STREAMING APIS**
3. **BUILDING STREAM PROCESSING APPLICATION WITH SPARK**
4. **COMPARISION BETWEEN SPARK STREAMING AND FLINK**

Learn about the real-time data processing architecture of Apache Spark. Build Spark Streaming applications to process data in real-time.

1 WEEK

---

5. ASSIGNMENT (OPTIONAL)

1. **INTRODUCTION, PROBLEM STATEMENT AND GRADING RUBRICS**

Solve an assignment to brush up on the skills learnt so far.

0 WEEK

---

6. BUILDING AUTOMATED DATA PIPELINES WITH AIRFLOW

1. **FUNDAMENTS OF AIRFLOW**
2. **WORKFLOW MANAGEMENT WITH AIRFLOW**
3. **AUTOMATING AN ENTIRE DATA PIPELINE WITH AIRFLOW**

Automate Data Pipelines with Airflow.

1 WEEK

---

7. ANALYTICS USING PYSPARK

1. **EXPLORATORY DATA ANALYSIS WITH PYSPARK**
2. **PREDICTIVE ANALYSIS WITH SPARK MLLIB**

Use PySpark to do EDA and Predictive Analysis using Spark's ML library.

1 WEEK

---

8. PROJECT: REAL-TIME DATA PROCESSING

1. **INTRODUCTION AND PROBLEM STATEMENT**
2. **GRADING RUBRICS AND SUBMISSION**

Build an end-to-end real-time data processing application using Spark Streaming and Kafka.

1 WEEK

**The Curriculum is subject to change as per the inputs from university or industry experts*

COURSE 6 - CAPSTONE PROJECT

CAPSTONE PROJECT

1. **AN OVERVIEW OF THE DOMAIN AND ASSOCIATED CONCEPTS** 2. **PROBLEM STATEMENT** 3. **EVALUATION RUBRIC** 4. **MID SUBMISSION** 5. **FINAL SUBMISSION** 6. **SOLUTION**	The capstone project will stitch all the components of data engineering together.	**4 WEEKS**

COURSE - RESEARCH METHODOLOGIES (11 WEEKS)

INTRODUCTION TO RESEARCH AND RESEARCH PROCESS

FAMILIARISE WITH DIFFERENT ASPECTS OF RESEARCH AND FORMULATE A RESEARCH QUESTION	• What is research, importance of reseach, what is data, what is information, what is knowledge? • Importance of research, types of originality, characteristics of research, research process • Criticism in research and its importance, Peer reviews in research and its importance • Types of research: Scientific vs Rest, Objectives of research • Structure of a research proposal: Components of the Research proposal covered over the course • Identify a research problem, formulate a research question, characteristics of a good research question

---

**The Curriculum is subject to change as per the inputs from university or industry experts*

RESEARCH DESIGN

DEVELOP AN UNDERSTANDING OF VARIOUS RESEARCH DESIGNS

- Types of research methods and pyramid of evidence
1. Study of existing researches and links between them
2. Applied and incremental
3. Discover
- Applied vs Fundamental, Quantitative vs Qualitative, Bayesian vs Frequentis, Hypothesis driven research vs Exploratory resarch
- Sample Size and Power, Precision vs accuracy trade-off, p-value vs confidence intervals using a case study

---

LITERATURE REVIEWING

LEAN HOW TO READ AND CRITIQUE A PAPER, AND HOW TO CITE A PAPER

- Intro to lit review process, what is a lit review, benefits of lit review, literature reivew process (read, analyse and cite)
- How to read and critique a paper
- Types of sources that could be cited during research, the importance of citations and how to cite
- What makes a good reference, How to use reference management software, Related scientific ethics

---

RESEARCH PROJECT MANAGEMENT

LEARN HOW TO PLAN THE PROJECT AND HOW TO ARRANGE FOR DATA

- Project management in reseach: research question, planning of the project, initiation, monitoring, closure.
- Project requirements on data: data collection, data access, data sources, availability, credibility and usability of data from different sources.
- Project requirements on analysis software: Analytical methods in Data Science, software requirement (R, Minitab, Matlab...), and data cleaning skills.
- Project requirements on time: planning, breaking the work down to tasks, Gantt Charts, Milestones identification and Deliverables, Re-planning.

**The Curriculum is subject to change as per the inputs from university or industry experts*

REPORT WRITING AND PRESENTATION SKILLS

MASTER GOOD SCIENTIFIC WRITING AND PROPER PRESENTATION SKILLS

- Art of writing a paper
- Parts of a paper
- Tools to write papers
- Publishing papers: Journals + Seminars
- Citation Methods and Rules
- Defending your thesis

SCIENTIFIC ETHICS

DEVELOP AN UNDERSTANDING OF THE ETHICAL DIMENSION IN RESEARCH

- Honor Code, Definition of Plagiarism, Type of Plagiarism, Code of good practice
- Research Claims, Professional Standard, IP, Conflict of Interest
- Legal aspects of data: Ethical Approvals for studies involving humans such as questionaire based research, Storing Primary Data,

---

COURSE - SAMPLE THESIS TOPICS (15 WEEKS)

SUBMITTING THE IN-DEPTH RESEARCH WORK IN A FINAL THESIS REPORT AND PRESENTING IT.

Sample Thesis topics to select from:

- Investigate dietary patterns and metabolite fingerprints of takeaway (fast) food consumers using PCA and clustering methods
- Investigate a diagnosis of eye diseases using imaging ophthalmic data
- Structure medical images with information geometry
- Using Social media feed to place tweets regarding natural disasters on a map
- Preventing credit card fraud through pattern recognition
- Developing a recommender system for a Media giant
- Risk modelling for Financial activities and Investment Banking

**The Curriculum is subject to change as per the inputs from university or industry experts*

Research of our learners:
A Glimpse

1 Thesis Topic

Build a prediction model to accurately detect and classify peripheral neuropathy.

Abstract

Background:

Damage to peripheral nerves causes Peripheral neuropathy (PN). Patients complain of pain, numbness and loss of balance. If not identified early and treated adequately, PN could progress rapidly and lead to fatal complications. A neurologist needs to determine the type of PN to provide differential treatment to the patient. However, defining factors to classify PN accurately has remained challenging. This research proposes a model to detect and classify PN into axonal, demyelinating, mixed and normal types from clinical and nerve conduction study (NCS) data using the Random Forest algorithm.

Data and methods:

Clinical and NCS data of 304 Indian patients, 229 affected by PN and 75 normal was collected with ethical approval from Kauvery hospital, Chennai. Exploratory data analysis and the Random Forest Algorithm was used to build a model.

Results:

Random Forest model was able to predict and classify PN with an accuracy of 96%. In axonal cases, sensory and motor nerves showed a drop in amplitudes of greater than 40% compared to normal patients. Reduced amplitude (>40%) in motor nerves of lower limbs and missing values (>90%) in sensory nerves of lower limbs identified axonal PN. Delayed onset latency (>40%) in motor nerves of upper limbs, decreased conduction velocity (>60%) in sensory nerves of upper limbs and increased onset latency (>40%) in F-waves of upper limbs delineated the demyelinating type. Median ages of patients were mixed (65), demyelinating (51) and axonal (61). Axonal (18.75% was significant in diabetic patients and demyelinating (14.8%) in non-diabetic patients. Both axonal and mixed (16.78%) types were greater in hypertensive patients, and demyelinating (17.11%) type was higher in patients without hypertension. Reflex was depressed more in mixed (17.49%) than axonal (15.51%) and demyelinating (11.89%). Mixed (37.06%) type showed more in-sensitivity to pin-prick than axonal (29.37%) and demyelinating (24.48%) types. Mixed (45%) patients tested positivefor Romberg's test more than axonal (31%) and demyelinating (21%). Mixed (34.65%) patients complained of numbness more than axonal (23.62%) and demyelinating (26.77%) types.

Conclusion:

Random forest algorithm identified and classified PN well using clinical and NCS features. Clinical features (age, diabetes, hypertension, reflex, Romberg's test, numbness and perception to pin-prick) were useful in detecting PN. Nerve conduction study features (amplitude, onset latency, conduction velocity, F-wave response and missing sensory values) were instrumental in classifying PN. Reduced amplitudes of sensory and motor nerves identified the axonal condition. Delayed onset latency and low conduction velocities along with missing and delayed F-wave responses identified the demyelinating type.

2 Thesis Topic

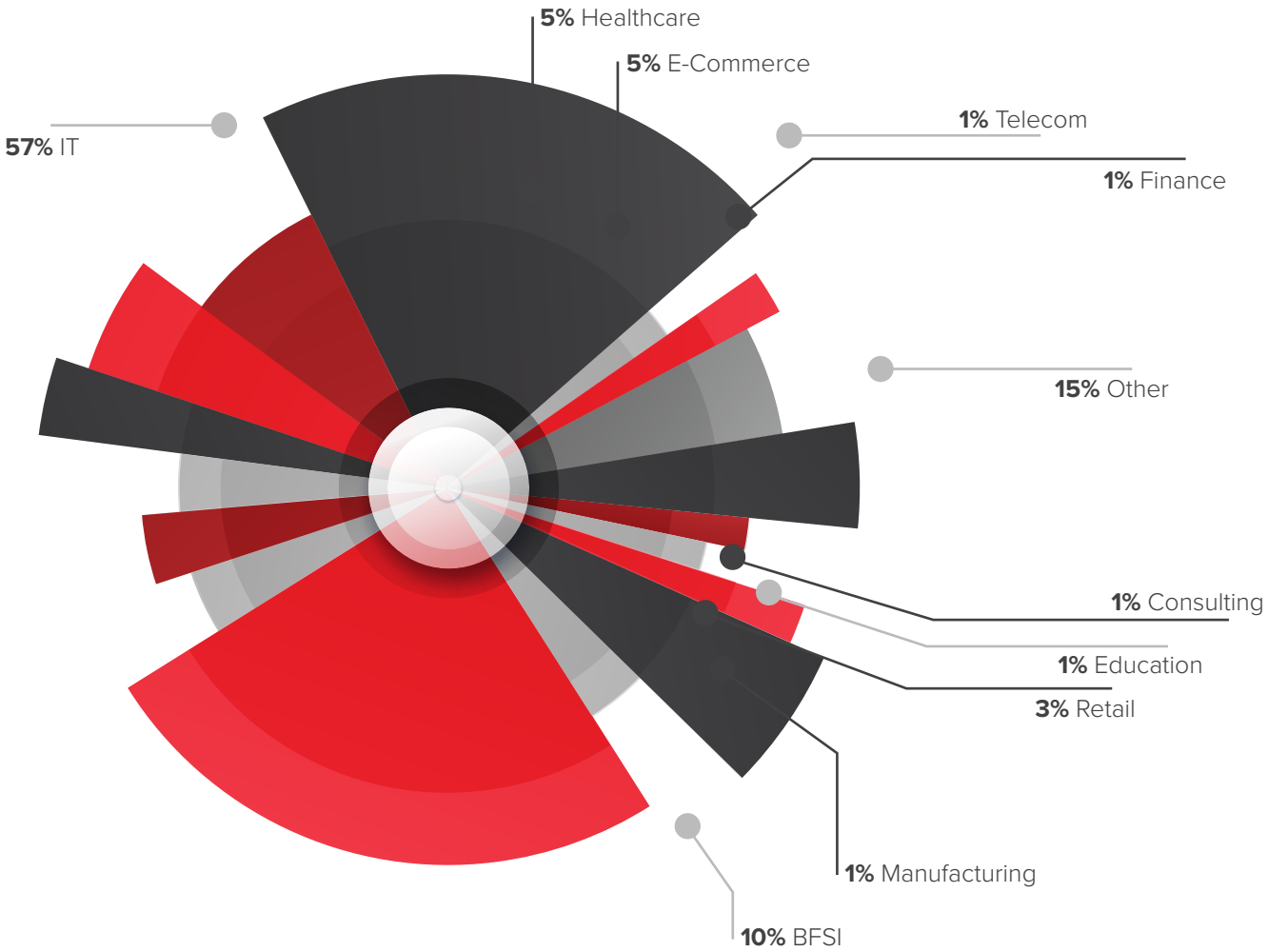
Automatic network coding of traffic junctions using machine learning.

Abstract

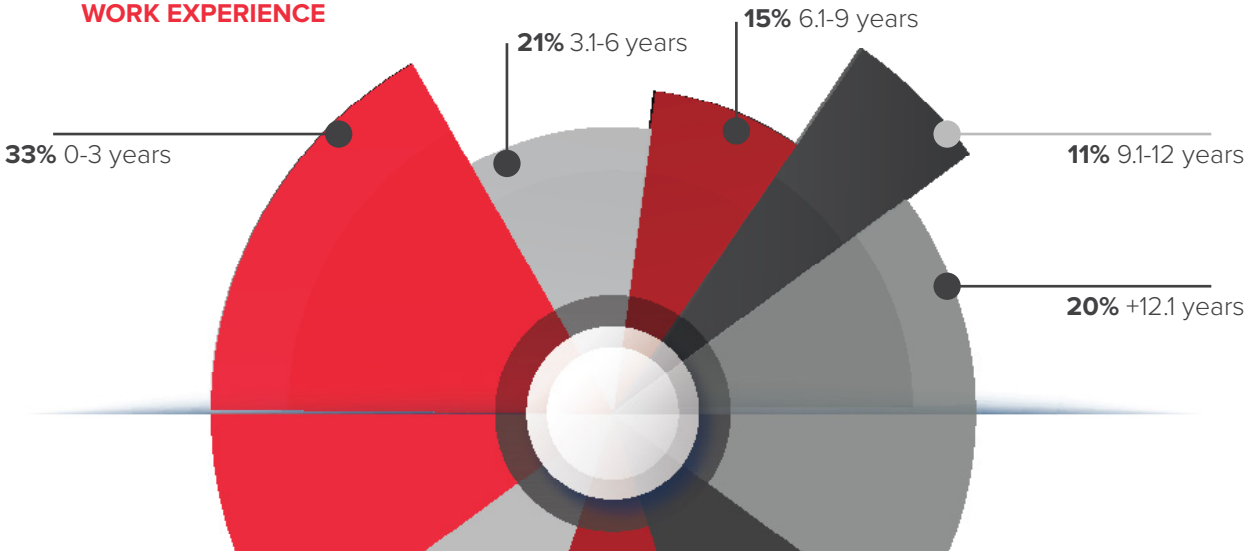
Before any traffic simulation can be performed, the network of roads and junctions is modeled. Assigning attributes to the roadway network, such as the road length and width, the junction type, number of arms, and lanes, is a crucial task while building the network. This research is an attempt to develop an efficient traffic junction classifier using machine learning and deep learning algorithms on satellite images. Three junction categories, Priority, Roundabout, and Signal, are considered for analysis. As this is a novel research idea, the required image dataset of junctions is created using the Google Maps API. By using robotic process automation, the downloading of the images is automated. Two approaches are taken to build the classifiers: a machine-learning approach and a deep-learning approach. The machine learning approach is split into two phases: the feature extraction phase and the classification phase. In the feature extraction phase, a Histogram of Oriented Gradients (HOG) descriptors is used to extract features from the images. Furthermore, in the classification phase, several classification algorithms are applied to the HOG features to build classifiers. In the deep-learning approach, taking advantage of powerful pre-trained models and transfer learning, a Convolutional Neural Network (CNN) is developed for classifying the junctions. The models are evaluated, and in the end, a comparison between the various classification models is performed. The results showed that the CNN classifier modeled had the best accuracy and AUC compared to the other models with scores of 0.81 and 0.94 respectively. Among the machine learning models that were trained on the HOG features, the Extreme Gradient Boosting model has the best accuracy of 0.62. The ultimate aim of this work is to use this junction-classifier model on real projects to aid the process of finding the type of junctions and reduce the effort and time required to model the roadway networks.

Meet the Class

INDUSTRIES OUR STUDENTS COME FROM



WORK EXPERIENCE



Elements of
Career Services

Jobs on Career Centre

Career Centre offers upGrad jobs across experience levels and CTC ranges.

- Easy apply feature for upGrad hiring partner vacancies
- Create a resume at profile builder with one click to apply for various jobs.

upGrad Elevate

- Recruitment Drive to connect you with the best talent admirers in the industry
- Get access to a wide range of opportunities and find the perfect job
- Apply your learnings to real industry problems

Interview Preparation

Pre-recorded content on topics such as:

- Profile building, communications, etc.
- Problem-solving approach
- Approaching guesstimates
- Domain-specific interview question bank and much more

Profile Builder (AI-Powered)

An easy-to-use Resume, LinkedIn and Cover letter preparation tool.

- Resume Score: AI-Driven Resume Score
- Real-time recommendations to improve.
- Match your resume to the JD and check fitment.
- LinkedIn Profile Review.
- Cover Letter creation.

Just-In-Time Interview Prep (JIT)

For upcoming job interviews JITs are conducted within 48 hours for eligible programs.

- Tailored to the job role and target domain
- Real-time feedback and tips for improvement

Personalised Industry Session

90-minute sessions over the weekend by leading industry experts.

- Session categories: Career, Technical and Communications
- Doubt resolution
- Develop proof of concepts and apply theoretical concepts in the real world
- Assess skill levels
- Peer Networking
- Classroom element
- Business communication sessions and much more

Disclaimer: Career services are subject to change. Please refer to the website or speak to our Admission Counsellor for updated details.

Experience upGrad Offline

UPGRAD BASECAMPS

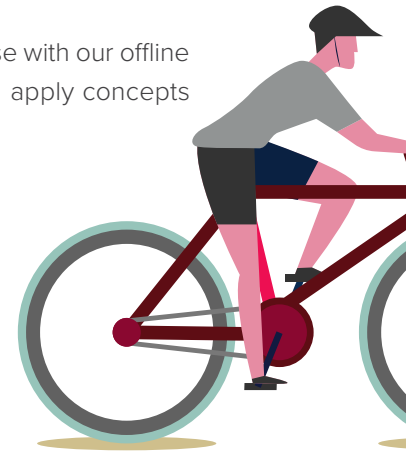
Held across all major cities in India, upGrad basecamps bring together learners, faculty and industry experts for a power-packed day of activities, career-building sessions and live group projects. Get to know your peers and faculty and hone your networking skills in an exciting environment.

CAREER FAIRS

Attend regular hiring drives in major cities across India, giving you the opportunity to interview with upGrad's 300+ hiring partners ensuring you get every opportunity you deserve.

HACKATHONS

Team up and put your learning to use with our offline Hackathons: designed to help you apply concepts and meet, network, and grow!



Hear from
Our Learners

Sachin Aggarwal, Experience: 18+ Years
"Learning with IIITB and upGrad has been an experience like no other. Being enrolled on an online program, you have your worries about how the program and teaching methods will be. My favourite part about the learning experience has been the well-designed and thoughtful content shared by IIITB professors and industry experts on upGrad platforms. Kudos to upGrad!"

Shravani Shahapure, Experience 16 Years
"For someone who really wants to pursue a career in the field of Data Science, it is worth opting for the complete course by IIITB and upGrad. IIITB and upGrad's online course on Data Science gives many opportunities and develops students for their future as they provide the best professors, thought-provoking assignments and case studies."

Savita Upadhyay, Experience: 4 Years
"It has been an amazing journey with upGrad till now. Starting with their course material to live sessions to mentor support, each helps you to always be on track and progress efficiently with the Data Science course. My sincere thanks to the entire team of upGrad and Professors of IIITB for showing me the path and direction for my dream to become a Data Analyst."

Tuhin Pal, Experience: 5 Years
"I appreciate the platform upGrad has provided and the way they have arranged modules and assignments. Modules are locked until you complete the previous one, so it feels like clearing a semester and going to the next one."

Program Details and
Admission Process

PROGRAM DURATION AND FORMAT
19 Months | Online

PROGRAM FEE
Please refer to the website for more details

PROGRAM START DATES
Please refer to the website for program start dates.
upgrad.com/data-science-masters-degree-iiitb/

ELIGIBILITY
Bachelor's Degree with minimum 50% or equivalent passing marks, and successful completion of the Executive PG Program in Data Science from IIITB with a 2.4 GPA. No coding experience required.

WEEKLY COMMITMENT (15 hours/week)

6-7 HOURS
Asynchronous learning time.

6-7 HOURS
Assignments and projects.

1 LIVE SESSION
Every two weeks.

SELECTION PROCESS

STEP 1: Selection Test
Fill out an application and take a short 17-minute online test with 11 questions.

STEP 2: Review and Shortlisting of Suitable Candidates
Our faculty will review all applications, considering the educational and professional background of an applicant and review the test scores where applicable. Following this, Offer Letters will be rolled out so you are assured of a great peer group to learn and network with.

STEP 3: Enrollment for Access to Prep Content
Make a quick block payment with assistance from our loan partners where required, receive immediate access to the prepped content and begin your upGrad journey.

FOR FURTHER INFORMATION, CONTACT

admissions@upgrad.com
1800 210 2020
We are available 24*7

Disclaimer: Program fee and payment options are subject to change. Please refer to the website for updated details or speak to our admission counsellor.

upGrad

COMPANY INFORMATION
upGrad Education Private Limited
Nishuvi, 75, Annie Besant Road,
Worli, Mumbai - 400018.