THE UNIVERSITY OF ARIZONA

Master of Science in
Data Science

Curriculum

upGrad

Course - 0: Pre Course Requisites

- ## Fundamentals of Calculus and Probability

  Learn the basics of calculus and probability and build a strong foundation to learn the complex machine learning algorithms.

- ## Basics of R Programming

  Learn the industrially relevant concepts of R programming language to understand the practical aspect of data science and machine learning in order to apply those in day to day routine.

- ## Basics of Python programming

  Learn the industrially relevant concepts of Python language to understand the practical aspect of data science and machine learning in order to apply those in day to day routine.

Tools Covered: Python and R Programming

Weeks	3
Credits	0

Course - 1: Data Analysis and Visualization

- ## Functional Programming in Python

  Learn to apply some of the commonly used paradigms of functional programming in Python.

- ## Data Structures and Algorithms

  Learn various data structures like Stacks, Queues and Tree in Python, that help in advanced data manipulation. Learn Python algorithms that are particularly important from an interview perspective.

- ## Database Design and Introduction to MySQL

  Learn the fundamentals of database and extract information from RDBMS using structured query language (SQL).

- ### Tableau

  Learn advanced visualization techniques with Tableau, the most in-demand tool in the industry. It is used to visualize complex data by creating interactive plots, dashboards, and so-called 'stories'.

- ### Essential Data Science Tools: Github and Jekyll

  Learn to manage codes and projects using Git and create portfolios using Github and Jekyll.

Tools Covered: Python, MySQL, Tableau, Github and Jekyll

Weeks	10
Credits	3
Assignment	1

Course - 2: Advanced Data Visualization

- ### Libraries for Data Science: Numpy, Pandas

  Learn how to manipulate datasets in Python using Pandas, which is the most powerful library for data preparation and analysis. Learn the use of NumPy in numerical operations in data analysis.

- ### Advance Visualization

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

- ### Exploratory Data Analysis

  Learn how to find and analyse the patterns in data to draw actionable business insights. Learn the extensive use of Python in data science.

- ### Inferencing from Data with R

  Build a strong statistical foundation and learn how to infer insights from a huge population using a small sample. Practical hands-on learning using R in statistics.

- **Hypothesis Testing with R**

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

Tools Covered: R programming, Numpy, Pandas, matplotlib, seaborn and ggplot

Weeks	10
Credits	3
Assignment	1

Course - 3: Introduction to Machine Learning

- **Building Linear Regression Models from Scratch**

  How can one variable be predicted using other variables? Understand the meaning of regression and how it can be built from scratch.

- **Linear Regression Practical Applications**

  Learn to practically apply the principles learnt in linear regression.

- **Logistic Regression**

  Learn about the fundamentals of binary classification technique through logistic regression.

- **Decision Trees**

  Learn how a human decision-making process can be replicated using decision trees. Build a decision tree algorithm from scratch for various industrial applications.

- **Random Forest**

  Learn the principles of a powerful ensemble method of random forests. Apply the fundamentals of random forest on real world problems.

- **PCA**

  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.

- ### Clustering

  Learn how to group elements into different clusters without any predefined labels through K-means clustering, hierarchical clustering, and more.

- ### Bagging, Boosting and Regularization

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

- ### Model Selection and Feature Engineering

  Learn the pros and cons of quantifying model complexity, along with general machine learning techniques such as feature engineering, model evaluation, and many more.

Tools Covered: Python, sklearn and matplotlib

Weeks	**11**
Credits	
Assignment	

Course - 4: Data Base Development and Management

- ### Advance SQL

  Learn how to write optimized SQL queries that require less memory and execute in a lesser amount of time. Hands-on experience with some advanced SQL concepts such as window functions, rank and partitioning.

- ### Introduction to NoSQL

  Take your knowledge of query languages a step further by learning about MongoDB - a NoSQL database that is becoming more and more popular in the industry.

- ## Introduction to Cloud and AWS

  You will understand what cloud computing is and what technologies made cloud computing possible. Then, you will learn about various deployment and service models and explore their essential characteristics. After this, you will learn about the benefits of cloud computing by exploring its advantages over on-premises data centers. This module would end with you gaining fundamental knowledge and practical skills required to build and deliver solutions on the cloud, using the services of a major public cloud service provider, Amazon Web Services (AWS).

- ## Introduction to Big Data

  Learn essentials for big data and understand how parallel working architectures enable big data computing. By the end of this module, you will be able to write MapReduce jobs in python MRJob to solve big data problems.

- ## Data Warehouses and Querying Using Hive

  Learn about the architecture and features of Hive. In this module, you will also learn how to write queries to derive valuable insights from the data available in the HDFS. Further, you will understand various query optimization techniques to get faster results.

- ## Working with ML Services (Optional Module)

  - You will learn how to use a monitoring service along with the machine learning (ML) and artificial intelligence (AI) services that AWS offers on its platform.
  - The ML/AI services are pre-trained solutions that can be applied to your files via the web platform. They can also be used as part of different applications buil using AWS.

Tools Covered: MySQL, NoSQL, MongoDB, AWS, EMR and Hive

Weeks	**10**
Credits	**3**
Assignment	**2**

Course - 5: Data Warehousing and Analytics in the Cloud

- ## Introduction to Apache Spark

  Learn Apache Spark framework and architecture to deal with unstructured data inside with the help of RDD. The module will help you realize the optimum usage of RDD and Paired RDD in Spark.

- ## Handling Big Data with Spark

  - Understand what is structured APIs and how these differentiate from the RDDs.
  - Learn the different types of structured APIs and how to work with them. Further, learn about Pandas API and the various optimization techniques spark offers - Catalyst optimizer and Adaptive Query Optimization.
  - Realise the optimum usage of RDD, DataFrame, and big datasets through th Apache Spark framework.

- ## SparkML

  Get hands on experience on Spark ML through different case studies in ML and understand the use of different components of Spark ML such as feature transformers, pipelines, estimators and evaluators.

- ## Recommender Systems

  - Learn how to build recommender systems using state-of-the-art techniques such as Collaborative filtering and Content-based Recsys
  - Learn how to evaluate recommender systems using offline evaluation strategies and A/B testing techniques

Tools Covered: AWS, Apache Spark, SparkML and Pyspark

Weeks	**9**
Credits	**3**
Assignment	**2**

Capstone Project- 1

- ## Best Practices for Effective Communication

  Learn how to effectively strategize, communicate, and fine tune your projects. Understand how to optimally present your findings to technical and non-technical stakeholders and upgrade your communication skills.

- ## Capstone Project

  Solve an end-to-end real-life industry problem from a wide variety of domains using the concepts of machine learning, cloud and big data.

Weeks	**5**
Credits	**1**

Course - 6: Neural Networks

- ## Introduction to Neural Networks

  Learn the most sophisticated and cutting-edge technique in machine learning - Artificial Neural Networks (ANNs). Learn and apply TensorFlow for basic classification tasks.

- ## CNN

  Learn the basics of Convolutional Neural Network (CNN) and OpenCV and apply it to computer vision tasks such as detecting anomalies in chest X-Ray scans and vehicle count and categorization to help the government ascertain the width and strength of the road.

- ## RNN

  Ever wondered what goes behind machine translation, sentiment analysis, speech recognition, etc.? Learn how Recurrent Neural Network (RNN) helps in these areas having sequential data such as text, speech and videos.

- ## Computer Vision

  Apply the concepts learned in Neural Networks to advanced computer vision tasks like Object Detection, Semantic Segmentation using YOLO, SSD, UNet, MaskRCNN.

Tools Covered: Python, OpenCV, TensorFlow, YOLO, SSD and MaskRCNN

Weeks	10
Credits	3
Assignment	1

Course - 7: Applied Natural Language Processing

- ## Lexical Processing

  Learn the industrial applications of NLP and text analysis on text corpus and perform sentiment analysis on a customer reviews dataset after performing feature engineering.

- ## Syntactic Processing

  Learn how to analyse the syntax or the grammatical structure of sentences using PoS tagging and dependency parsing. Learn to identify named entities using custom NER techniques.

- ## Semantic Processing

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

- ## Introduction to Attention Mechanism

  Have you ever wondered how Google translate does a perfect job in understanding the context of the sentence and translates it into any other language? Here you will learn how to build a machine translation system that is capable of translating a sentence from one language to another. You will learn about attention mechanisms, which simulate how we humans perceive textual information. The core principle of an attention mechanism is spatial understanding, which enables the machine to focus only on essential areas.

- **Advanced Language Models: BERT*, GPT***

  This module will introduce you to the evolving world of transformers and help you gain a complete understanding of how this complex model works. You will also learn how to build a transformer model from scratch and fine-tune different types of transformers such as GPT3, BERT and RoBERTa.

Tools Covered: Python, NLTK, Spacy, BERT*, GPT* and TensorFlow

Weeks	10
Credits	3
Assignment	1

Course - 8: Managing the Digital Information Environment

- **Structured Problem Solving**

  Learn how to approach a business problem using various structured frameworks, such as 5W, 5WHYs and SPIN, and apply your learnings from the course to solve a real-life business problem.

- **Data Storytelling**

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

- **Framework to DS/ML solution building strategy**

  Understand the business impact of data science and machine learning and learn a framework for building the AI and ML strategy for any business.

- **Building Data Architecture**

  Understand how to build a data architecture and identify the strengths and weaknesses of the data architecture of any organization.

- ### Introduction to MLOps

  Do you think ML ends with just deploying a ML solution? You have to monitor the performance and keep updating the model and its infrastructure from time to time. Learn how to operationalize ML models in this module.

- ### Introduction to Data and Model Lifecycle

  Learn about the various situations in which you need to tweak your model or change the data collection/generation mechanism to improve your ML solutions overtime.

Tools Covered: REST API, Python, TensorFlow, Kubeflow, Flask, Kubernetes and Docker

Weeks	**10**
Credits	**3**
Assignment	**1**

Course - 9: Ethical Issues in Information

- ### Ethical Foundations

  Learn about the birth of ethics, the role it plays in our daily lives and how to ensure ethical conduct in Data Science.

- ### Ethics in Data

  Learn how to ethically use and collect data, how to ensure privacy through masking and how to maintain anonymity and confidentiality of data and metadata. Learn about the various ethics & compliance initiatives undertaken to ensure privacy of information such as GDPR.

- ### Ethics in Modelling

  Learn the various problems faced due to algorithmic bias and how to take a data/model-driven approach to tackle this.

Weeks	9
Credits	3
Assignment	1

Capstone Project- 2

- Solve an end-to-end real-life industry problem from a wide variety of domains using the concepts of specialized machine learning topics, cloud and big data.

Weeks	5
Credits	2

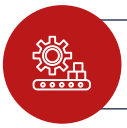
Capstones Projects

- Manufacturing: Predictive Maintenance of Machines
- Finance: Credit Card Fraud Detection, Portfolio Analysis
- Healthcare: Hospital Rating Model and service improvements
- Retail: Predicting the sales for different stores for a retail giant
- Internet: Eye for Blind
- Self Driving: Driver Drowsiness Detection
- Media: News Recommendation Engine

Industry Projects

- Credit Card Fraud Detection
- Style Transfer using GAN
- Sales Forecasting
- Eye for Blind
- Sentiment-Based Product Recommendation System
- News Recommendation System
- Stack Overflow Case Study
- Uber Case Study
- Data Warehousing Case Study
- Real-Time Twitter HashTag Analysis, using Spark Streaming-
- Recommendation System, using ALS (Alternating Least Squares)
- Market Basket Analysis, using Apriori Algorithm
- NYC Taxi Fare Prediction
- Telecom Churn Prediction
- Recommendation System using PCA and KNN
- E-Commerce Assignment
- Speech Recognition
- Gesture Recognition

Capstones Projects



Manufacturing: Predictive Maintenance of Machines

Finance: Credit Card Fraud Detection and Portfolio Analysis

Healthcare: Hospital Rating Model and Service Improvements

Retail: Predicting the Sales for Different Stores for a Retail Giant

Internet: Eye for Blind

Self Driving: Driver Drowsiness Detection

Media: News Recommendation Engine