Deloitte.

October 4 to November 3

Deloitte Al Academy Bootcamp Playbook



Program Agenda

Module	Courses	Duration in Hrs
Core Data Fundamentals	Coding Fundamentals	6
AI and Machine Learning	Programming in Python	12
Core Data Fundamentals	Introduction to DBMS and SQL	16
Core Data Fundamentals	Statistics and Mathematics for Machine Learning	12
AI and Machine Learning	Data Visualization and Exploratory Data Analysis	16
Core Data Fundamentals	Visualization & Insights	16
AI and Machine Learning	Machine Learning Algorithms and Applications	40
AI and Machine Learning	Introduction to Tensorflow and ANN	10
AI and Machine Learning	Artificial Neural Networks	6
AI and Machine Learning	Computer Vision	4
AI and Machine Learning	Deep Neural networks	4
AI and Machine Learning	Conversational AI	16
Core Data Fundamentals	Data Engineering and ETL	48
Core Data Fundamentals	Cloud Technologies and Data Warehouse with AWS	24
Core Data Fundamentals	Deployment of ML Algorithms	24
	Integrated Capstone Project Review and Presentations	16
	Total Duration in Hrs	270

3	# Module	Course Title	Topics & Subtopics	Learning Outcomes and Assessment	Assessm ents	Tools & Technol ogies	Durati on (hours)
	1 Core Data Fundam entals	Coding Fundam entals	Code Hygiene concepts , GitHub, Leading Practices	 Illustrate the importance of basics of coding Define and describe coding best practices and the role of code hygiene Explain usage of naming conventions in coding, discuss the role of GitHub 	MCQ	None	6
i	2 Al and Machine Learning	Program ming in Python	Recap on Python Fundamentals and Data Structures Intro to Numpy: Numpy Arrays, nd arrays, Array operations like reshape, shape, matrix operations, arithmetic operations on numpy arrays "Introduction to Pandas: Reading data from .csv or excel sheets, row and column operations on pandas series and data frames. Data Manipulation: Filtering, Sorting, Selecting, Working with dates, joining data frames, Grouping, merging, pivoting, concat etc.Building Framework, Writing Libraries etc."	 Recap on pre-learning topics Describe the fundamentals of NumPy arrays Apply appropriate operations like arithmetic operations, relational operations, array shape manipulations operations, etc. 	Coding	Anacond a/ Jupyter Noteboo k, Python	12

Capstone: Briefing and Teaming

#	Module	Course Title	Topics & Subtopics	Learning Outcomes and Assessment	Assessme nts	Tools & Technolo gies	Durati on (hours)
3	Core Data Fundame ntals	Introducti on to DBMS and SQL	"Data Models, Schemas and instances Classification of database management systems Conceptual Database Design Entity – Relationship Model: Relationship: Degree of relationship, Cardinality, Participation, Key features of E-R Model: Generalization and Aggregation, E-R Diagrams - Case Studies. Database Implementation" "Introduction to SQL: DDL, DML operations SQL querying to do operations such as identifying nulls, special characters, blank rows/columns, and run distributions, run data summaries, merge tables, get unique counts SQL Joins, Aggregate functions and GROUP BY, Nested queries and sub queries. GROUP BY CLAUSE along basic aggregations such as SUM, COUNT, AVGIntroduction to Pivot & Unpivot, Aliasing pivot columns, Pivoting multiple columns CREATE, Removing an index, DROP INDEX statement. Unique, Function-based index "	 Understand the need for DBMS, storage of data in the form of relational database, design principles of the data, concept of primary key, foreign key, normal forms, ER Model etc. Performing various data manipulation operations on the data, understanding query language, extracting various information from the data stored 	Coding	Tools: Oracle/M ySQL	16

#	Module	Course Title		Learning Outcomes and Assessment	Assessment s	Tools & Technologi es Required for the session	Durati on (hours)
2	Core Data Fundame ntals	Statistics and Mathema tics for Machine Learning	 Introduction to Descriptive Statistics: Mean, Median, Mode, Variance, Correlation. "Introduction to Probability: Random Variables, Expectation, Probability, Theorems on Probability. Conditional probability, Bayes theorem" Introduction to inferential statistics, probability distributions binomial, Poisson, normal "Hypothesis Testing: Sampling Techniques, Central Limit Theorem, Z-test, t-test, F-test. ANOVA Nonparametric tests: Mann-Whitney, Chi-square test" Hands-on to implement hypothesis testing and other statistical concepts on a relevant dataset 	 Understand measures of central tendency, deviation from mean, relationship among variables, understanding dependency, Discrete and continuous probability distributions and their significance in real time data conditional probabilities, random variables and random outcomes, Calculate mean and variance of random variables Understanding K-fold, random sampling, stratified sampling etc 	NA	Tools: Anaconda/ Jupyter Notebook, Python Packages: numpy, pandas, stats, statmodels	12
5	Al and Machine Learning	Data Visualizati on and Explorato ry Data Analysis	 "Data Visualization: using matplotlib and seabornUnivariate and Bivariate plots: barplot, histogram, kde plot, pie-chart, donut chart, scatter plot, boxplot, pairplot, heatmap etc." "Exploratory Data Analysis: handling missing values, imputing, dealing with outliers, binning continuous variables, Data Encoding Techniques, Data Normalization, Feature Engineering - PCAOverview of Linear Algebra: Matrices, Rank of a Matrix, Eigen Values and Eigen Vectors" Hands-on on visualizing the data and cleaning the data - includes tasks like outlier treatment, missing value imputation, feature engineering, data transformation, normalization etc. 	 Create static plots with matplotlib and seaborn. Extract inferences out of various plots, understanding missing values, distribution of the attributes, dependencies between the attributes etc. Understand the importance of Linear Algebra in ML, dependency of attributes and rows in the dataset, feature engineering using PCA 	NA	Tools: Anaconda/J upyter Notebook, Python Packages: numpy, pandas, sklearn, matplotlib, seaborn	16

Capstone: Check point 1 pulse check

#	Module	Course Title		Learning Outcomes and Assessment	Assessme nts	Tools & Technolo gies Required for the session	Durat ion (hour s)
6 a	Core Data Fundam entals	Visualiz ation & Insights	 Introduction to Tableau, Connecting to Standard Data sources, Hierarchies, Sorting, Grouping, Sets and Filters, Parameters, Introduction to Maps, Color and Other Visual Cues, Formatting Your Visualizations, Dashboards and Story Features 	 Learners will be able to understand PowerBI as a visualization tool, connecting the data sources to the tool, retrieving the data. 	NA	PowerBl	6

Capstone: Check point 1 pulse check

#	Module	Course Title		Learning Outcomes and Assessment	Assessment s	Tools & Technologi es	Duration (hours)
6 b	Core Data Fundament als	Visualizatio n & Insights (continued)	"Overview of PBI and exploring the interface, Comparison (basic) with other popular tools, Power BI Data source connection - Building Power Queries, Connecting to CSV, Excel etc, Data Manipulation using Power Query - Sorting, Merging, Appending, Index column, Conditional column, Rows and Columns selection, Aggregation and Statistics functions for Numeric, Text Formatting functions, Grouping data, Date and time manipulation functions Data loading and transformation -Eliminating duplicates, filter, groups, replacing values, and custom columns or measures, hierarchies, Joins and Relationships - Cross directional filters, Cardinality, Active & Inactive relationships, Model view Introduction to DAX, Calculated Columns, Implicit and Explicit Measures, Using variables in DAX expressions and Aggregation Functions creating visuals and reports, slicers, Drill through filters, Univariate analysis, Bivariate analysis, Dashboard in Power BI"	 Various operations on data using PowerBl, applying aggregation. Data Analyiss expressions (DAX), implementing them in the tool Creating different types of charts, reports, Univariate and bivariate analysis of the data and building different charts with multiple variables, creation of dashboard 	Assessment: Case study using PowerBI (peer- evaluated)	PowerBI	10
7	Al and Machine Learning	Machine Learning Algorithms and Applications	 "Introduction to Machine Learning: Supervised and Unsupervised Learning, Linear Regression: Predicting continuous variable, assumptions of Linear Model, constructing a regression model in Python, Model evaluation using loss functions, RMSE, R-Square" "Logistic Regression in Python: Logistic Regression: Predicting a binary variable, interpreting model output, using Python to create a logistic model Checking model diagnostics, computing accuracy metrics, ROC, AUC, doing kfold cross validation" "Decision Trees using Python: understanding purity metrics, Gini, entropy, information gain, creating a classification tree. Regression Trees Tree based Ensembles: Bagging and Boosting Techniques. Random Forest, Ad boost, Gradient Boosting, XGBoostHyper-parameter tuning in these techniques. Gridsearch algorithm for hyper-parameter tuning" 	 "Understand the need for Machine Learning, Predictive models. Be able to interpret model results, Be able to check model assumptions, Be able gauge the out of sample model performance" "Be able to understand how a linear classifier works. Be able to gauge the performance of a classifier out of sample Dealing with practical issues like imbalanced class problem, understanding various performance metrics and choosing the right measure to evaluate the model" Be able to build a tree based classifier and extract rules. Be able to build a tree based regressor and extract rules. Be able to find out the variable importance, meaning of oob score in bagging trees, appropriate usage of hyper parameters and their importance, Tuning the parameters for all tree-based models 		Tools: Anaconda/ Jupyter Notebook, Python Packages: numpy, pandas, sklearn, statmodels	24

Capstone: Check point 1 submission

#	Module	Course Title	Topics	Learning Outcomes and Assessment	Assessments	Tools & Technologies	Duration (hours)
7	Al and Machine Learning	Machine Learning Algorithms and Applications	 Unsupervised Learning: Clustering techniques - K-means and hierarchical clustering, elbow method, silhouette score, profiling Recommendation Systems: Collaborative and content-based filtering, Nearest Neighbor, SVD, NMF "Time Series - Univariate Analysis: Moving average, weighted moving averages, exponential smoothing etcMultivariate modelling, ARIMA, Holt-winters, SARIMA" Hands-on on ML concepts learnt so far 	 Understanding the concepts of unsupervised learning, clustering and segmenting the data into groups, infer the insights from the clusters Be able to understand the types of recommendation algorithms, implementation of recommendation engines "Understanding the components of time series like seasonality, cyclic, trend etc, apply simple algorithms like moving average. Understand the concepts like exponential smoothing, auto regression, differencing, etc by implementing algorithms like ARIMA, holt-winters, SARIMA etc." 	Assessment: Case study, including Stats and Mathematics for ML and EDA (evaluated by instructor)	Tools: Anaconda/ Jupyter Notebook, Python Packages: numpy, pandas, sklearn, statmodels	16
8	Al and Machine Learning	Introduction to TensorFlow and ANN	Introduction to TensorFlow, Tensors, placeholders, operations on tensors. Installing TensorFlow, Working with/without GPU, The Architecture of TensorFlow, Operations on TensorFlow "Introduction to Neural Networks: What is neuron, perceptron. Activation Functions (Sigmoid, tanh, Relu etc.), Optimization Techniques: Gradient Descent, Batch Gradient, Mini-batch Gradient and Stochastic Gradient, RMSProp, ADAM etc. Loss Functions: Cross-entropy loss, MSE concepts of FFNN and Backpropagation algorithms, MLP Architecture"	 Participants will be able to understand the need for AI, Role of TensorFlow, syntax and functionalities fo TensorFlow, Fundamentals of TensorFlow variables, storage structure, manipulation of tensors etc. "Be able to understand the concept of neural networks, finding minima in a loss function, gradient descent, effect of change in learning rate etc. Math behind all these concepts" 	MCQ	Tools: Google Colab Packages: TensorFlow, Keras	10
9	Al and Machine Learning	Artificial Neural Networks	"Linear Regression model using TensorFlow: Implementation and tracking of training v/s validation loss, Running the model for multiple epochs Regularization techniques like L1, L2, dropout, Earlystopping etc." Hands-on on Classification and Regression using MLP Architecture	 Participants will get a flavor of implementing and building predictive models using TensorFlow, performance evaluation in Linear Regression 		Tools: Google Colab Packages: TensorFlow, Keras	6

Capstone: Check Point 2 pulse check

#	Module	Course Title	Topics	Learning Outcomes and Assessment	Assessment	Tools & Technologi es	Duratio n (hours)
1 0	Al and Machine Learning	Computer Vision	Introduction to Image data: Concepts of pixels, color maps, types of images like gray-scale, binary and RGB. Basics of image operations like erosion, dilation, filtering, segmentation	Be able to understand the concept of Image data, its storage structure, basic operations on images		Tools: Google Colab Packages: TensorFlow, Keras	4
1	Al and Machine Learning	Deep Neural networks	Convolutional Neural Networks: Meaning of convolution operation, convolution layers, kernals/filters, pooling and padding techniques, feature maps etc. Hands-on on Image Classification using CNN	 Understand the working of CNN models Understand pooling and padding techniques, meaning of filters, feature maps 	Assessment: Case study, including Artificial Neural Networks and Computer Vision (peer-evaluated)	Tools: Google Colab Packages: TensorFlow, Keras	4
1 2	Al and Machine Learning	Conversati onal Al	Text Data: Cleaning text, removing stopwords, punctuation, stemming and lemmatizing. Creating count matrix and TFIDF matrix Recurrent Neural Networks with Tensorflow: Working with text data and Natural Language processing using RNN Introduction to Speech recognition using RNN Handson on NLP using RNN and LSTM	 Understand about text data, clean the text data, text data for analysis, frequency and inverse document frequency Use RNN to solve case studies on natural language processing Understand the basics of speech recognition 	Assessment: Case study (peer-evaluated)	Tools: Google Colab Packages: TensorFlow, Keras	16
1 3 a	Core Data Fundament als	Data Engineerin g and ETL	"Introduction to Big Data and HadoopBig Data Characteristics, Significance, Limitations of existing technologies for Big Data AnalyticsOverview of Hadoop - Architecture, Components, Hadoop Distributed File System (HDFS) operations, Hands-on, YARN, MapReduce Concepts"	"By the end of this module the participants will: - Know what is Big Data and why is important for enterprises, What is Hadoop, Hadoop Distributed File System and MapReduce - Have working knowledge of how to use them - Be able to work with HDFS and perfoem the basic file system activities"		Tools: Bigdata platform, Hadoop, Hive, Kafka, PySpark	8

Capstone: Checkpoint 2 Submission

#	Module	Course Title		Learning Outcomes	Assessme nts	Tools & Technolog ies	Duratio n in hours
13b	Core Data Fundame ntals	Data Engineerin g and ETL	 "Hive - Pre-processing and Data transformations on Big DataData Storage, Data Types in Hive, Hive Query Language Features, Joins in Hive, Performance optimiztion - Partitions in Hive, Handling different data formats - JSON and Parquet, Hands-on Exercises, Case study, Assignments" "Core PySparkSpark Architecture, RDD (Resilient Distributed Dataset) Operations, PySpark Applications - Deployment modes, DAG (Directed Acyclic Graph), Shared Variables, Shuffle Operations, RDD Persistence, Hands-on Exercises, Case studies" "PySpark SQLPySpark SQL - Overview, SparkSession and Dataframes, DataFrames from different data formats and sources, Hive Integration and Hive tables as data sources, DataFrames and RDDs, Performance Tuning - Caching, Broadcast Join/ Broadcast Hint, Configuration options, Hands-on Exercises, Case Studies, Assignments" "PySpark StreamingPySpark Streaming Architecture, Handling simple sources (socket streaming and file streaming), Case Study with stock market data stream, Hands-on Exercises, Use case scenario - using Spark for IOT applications with edge device data stream" 	 "From this module the learners will: - Have a good understanding of Hive - Understand Data types in Hive, Data storage in Hive - Know how to write Hive queries for all tasks from creating databases & tables to loading data, querying including joins - Perform advanced operations on tables like handling data in Parquet and JSON formats" "Have a good understanding of Spark architecture, How data is processed with RDDs in Spark, Operations on RDDs, Shared variables, Spark application deployment modes" Have good understanding of PySpark SQL and the APIs - Know how to use Spark SQL - Have an understanding of Dataframes, querying - Be able to use multiple data formats/sources" Understanding of PySpark Streaming and concepts with practical examples, case studies and hands-on exercises" 	NA	Tools: Bigdata platform, Hadoop, Hive, Kafka, PySpark	32

Training Plan Week 7: AWS track

#	Module	Course Title	Topics	Learning Outcomes	Assessments	Tools & Technologies	Duration in hours
13	Core Data Fundamentals	Data Engineering and ETL	"PySpark MLLibMLlib Packages and Utilities in PySparkMachine Learning Case Studies with large datasets - Regression, Decision Tree, Recommender Systems - Collaborative Filtering ALS, Text Analysis with TF-IDF Case studies, Hands-on Exercises"	 "Learning outcomes: - Good understanding of understanding of Spark MLlib with practical examples - Know how touse MLlib APIs in PySpark application development with hands-on exercises" 	Assessment: Case study (peer-evaluated)	Tools: Bigdata platform, Hadoop, Hive, Kafka, PySpark	8
14	Core Data Fundamentals	Cloud Technologies and Data Warehouse with AWS	AWS3: Intro to Simple Storage Service, S3 Buckets, S3 Versioning, S3 Life Cycle Management, Different storage options, Hosting a website on S3 AWS RDS: Relational Database Concepts, Database Design, Database Schema, Normalization, Constraints and Dependencies, DDL - Create, Alter Drop Databases, DML - Load and Query Data, Setting up AWS account. Introduction to AWS services, Setting up Relational Database using AWS RDS Data Warehouse Services on AWS: Setting up RedShift on AWS, Redshift - AWS Data Warehouse service, System Architecture, Getting started and using databases, Creating databases, Designing tables, Loading data and querying, Hands-on with AWS Redshift Data Lake Formation, Athena, Data Pipeline AWS Machine Learning model: AWS Machine Learning Tools introduction, AWS Sagemaker, Push dataset on AWS Sagemaker, Train and publish models using Machine Learning algorithms	 Participants will understand storage features available on AWS S3 and will deploy a website on AWS S3. Create databases, load and query data on an RDBMS Participants will know how to use RedShift - AWS data warehouse. Create databases, design tables, load and query data Participants will work on AWS Machine Learning tool Sage maker to train and predict models 		AWS Cloud Lab	24

Capstone: Check point 3 Pulse Check

Training Plan Week 7: Azure track

#	Module	Course Title	Topics	Learning Outcomes	Assessments	Tools & Technologies	Duration in hours
13	Core Data Fundamentals	Data Engineering and ETL	"PySpark MLLibMLlib Packages and Utilities in PySparkMachine Learning Case Studies with large datasets - Regression, Decision Tree, Recommender Systems - Collaborative Filtering ALS, Text Analysis with TF-IDF Case studies, Hands-on Exercises"	 "Learning outcomes: - Good understanding of understanding of Spark MLlib with practical examples - Know how touse MLlib APIs in PySpark application development with hands-on exercises" 	Assessment: Case study (peer- evaluated)	Tools: Bigdata platform, Hadoop, Hive, Kafka, PySpark	8
14	Core Data Fundamentals	Cloud Technologies and Data Warehouse with Azure	 Azure Storage: Microsoft Azure Storage Overview, Azure Storage Account, VM Storage and Storage replication - Hands-on Lab, Azure Storage Monitoring and Access, Backup and Disaster recovery "Azure SQL Database: Relational Database Concepts, Database Design, Database Schema, Normalization, Constraints and Dependencies, DDL - Create, Alter Drop Databases, DML - Load and Query Data, Setting up relational database using Azure SQL Database Azure Data Factory, Azure Data PipeLines, Azure DataBricks, Data Lakes" Data Warehouse services on Azure: Azure Synapse Analytics, Create and connect, load data, Scale, Workload management, Pause and resume Introduction to Azure Machine Learning: Azure machine learning overview, Introduction to Azure machine learning applications -Hands-on Lab, Azure machine learning and scoring and evaluating models. Preparing Data for use with Azure Data Bricks: Introduction to Azure Data Bricks, How to perform basic operations with Azure Data Bricks SnowFlake Datawarehouse on Microsoft Azure: Introduction to SnowFlakes and Snowflake web UI, Snowflake account, provision warehouses, explore Snowflake databases, run queries 	 Participants will understand storage features available on Azure and will deploy a website on Blob Storage Understand data model concepts Create databases, load and query data on an Azure RDBMS" Use Azure Synapse data warehouse Create databases, design tables, load and query data Understand the features of Azure Machine Learning studio and its features Train and publish model using Azure Machine Learning Studio Understand Azure Data Bricks Features and significance of Azure Data Bricks Perform Operations on Azure Data Bricks Understand Snowflakes and its significance Connect to snowflake webUI on Azure 		Azure Cloud Lab	24

Capstone: Check point 3 Pulse Check

Training Plan Week 8: AWS Track

Training Sequenc e	Module	Course Title	Topics	Learning Outcomes and Assessment	Assessmen ts	Tools & Technologi es	Duration in hours
15	Core Data Fundame ntals	Deployme nt of ML Algorithm s	"MLOps, Flask, kubernetes, Need for deployment, dependencies, concepts, automating the deployment, scaling and management Sage Feature engineering1. Scaling2.Imputation3.OHC4.Histo grams Binning5.Log Transformations Understand core concepts of Feature engineering Machine Learning Basics on Cloud Learn about ML and Deep learning on SagemakerCloud Al/ML services High level understanding about Al and ML Services Cloud ML service implementation and Operations Learn about how Cloud ML models are deployed and various Production Variants with Docker Methodology"	 Be able to understand the concept od algorithm/software deployment for the business case studies. Participants will understand how to train and deploy model on AWS Sagemaker. 	Assessment: Case study, including AWS/Azure depending on track (evaluated by instructor)	Kubernetes Azure Cloud	24
16	Capstone P	roject Prepara	ations and Presentations				16

Capstone:

- Checkpoint 3
- Preparation and presentations by Participants

Training Plan Week 8: Azure Track

Trainin g Sequen ce	Module	Course Title	Topics	Learning Outcomes and Assessment	Assessment s	Tools & Technologi es	Duration in hours
15	Core Data Fundame ntals	Deployme nt of ML Algorithm s	"MLOps, Flask, Kubernetes, Need for deployment, dependencies, concepts, automating the deployment, scaling and management" Deployment of ML algorithms on Azure Machine Learning.	 Be able to understand the concept od algorithm/software deployment for the business case studies. Participants will understand how to train and deploy model on Azure Machine Learning. 	Assessment: Case study, including AWS/Azure depending on track (evaluated by instructor)	Kubernetes. Azure Cloud	24
16	Capstone Project Preparations and Presentations						16

Capstone:

- Checkpoint 3
- Preparation and presentations by Participants

Thank You