ANNEXURE A: Syllabus

Details of the Syllabus/model curriculum for the Associate Data Analyst Certification course at Level 5.0

Associate Data Analyst			
COURSE TYPE:	**STT – Upskilling, Dual-Flexi Qualification, Future Skills**	**NSQF Level:**	**5.0**

Course Objectives:

- Develop a strong command of Python programming, including its syntax, data types, and control structures, to create efficient and maintainable code for engineering applications.
- Acquire the skills to collect, clean, preprocess, and analyze diverse datasets, extracting valuable insights and engineering features that enhance the accuracy and performance of machine learning models.
- Master a broad range of machine learning algorithms, deep learning methodologies, and computational statistics, enabling the effective application of these techniques to solve complex engineering problems.
- Apply computer vision and natural language processing techniques to interpret visual and textual data, designing and implementing AI solutions that address real-world engineering challenges.
- Implement MLOps best practices, ensuring smooth deployment, monitoring, and maintenance of AI models in production environments, and demonstrating expertise in building scalable and reliable AI solutions.

Learning Outcomes:

LO1: Demonstrate a comprehensive understanding of Python programming concepts and object-oriented programming principles, effectively applying them to create structured and reusable code for engineering applications.

LO2: Analyze, preprocess, and wrangle diverse datasets, showcasing the ability to extract meaningful insights from raw data and engineer relevant features to enhance machine learning model performance.

LO3: Develop proficiency in a wide range of machine learning algorithms, including deep learning techniques, and confidently apply them to solve complex engineering problems, considering their strengths and limitations.

LO4: Utilize computational statistics to critically evaluate data and draw meaningful conclusions, employing hypothesis testing and statistical techniques to make informed decisions in engineering contexts.

LO5: Apply computer vision and natural language processing techniques to process visual and textual data, exhibiting the capability to design AI solutions for a variety of engineering challenges.

LO6: Implement MLOps best practices to deploy, monitor, and maintain AI models in real-world environments, showcasing expertise in building robust and scalable AI solutions for engineering applications.

CONTENT & TEACHING UNITS: CORE (INCLUDING INTERNSHIP)

Unit	Content	HRS
GN/IT/0001: Bridge Module: Python for Data Science	**Python Fundamental:** - Learn to install and configure Python IDE and use collaborative cloud interfaces. - Gain a strong grasp of Python language basics, including syntax and data types. - Develop coding skills to write, debug, and troubleshoot Python programs. - Acquire the ability to design and implement functions using both top-down and bottom-up approaches.	Th – 15 Pr – 35 OJT - 10 ---------- Total - 60

	- Understand the advantages of modular programming through the creation and utilization of functions. - Explore various parameter-passing techniques and differentiate between them. **Python for Data:** - Understand file operations, including opening, closing, reading, and writing files using Python. - Explore the open() function and understand file object attributes. - Develop proficiency in working with different file formats and operations. - Master data handling using the pandas library, including loading, manipulating, and saving data. - Gain practical experience in working with data formats such as CSV and JSON. - Apply data analysis techniques using pandas to extract insights from datasets. **Working with Database:** - Acquire hands-on experience in creating and managing relational databases. - Develop a solid foundation in relational data management using SQL. - Get introduced to the principles of NoSQL data management. - Learn practical data analysis techniques within the context of relational databases. - Understand how to work with structured data effectively for real-world applications.	
GN/IT/000 2: Data Analytics and Statistical Methods	**Python package for data science:** **numpy:** - Acquire proficiency in utilizing numpy for array operations and manipulations. - Understand the nuances of numpy arrays compared to traditional lists. - Master the creation of N-dimensional arrays, including 1D, 2D, and 3D arrays. - Learn to generate specialized matrices like zeros, ones, and identity matrices. - Comprehend random number generation and integration with numpy arrays. - Gain practical insights into treating RGB images as numpy arrays **pandas:** - Gain expertise in working with pandas Dataframes. - Learn to load and analyze datasets from diverse file formats. - Develop the ability to merge, sort, and filter Dataframes. - Handle missing data effectively within Dataframes. - Master data access and manipulation using loc and iloc functions.	Th – 25 Pr – 55 OJT - 10 --- Total - 90

	- Skillfully implement advanced data analysis methods like nlargest() and nsmallest(). - Understand how to add or remove attributes from Dataframes for tailored analysis. **Data Visualization with Matplotlib:** - Develop a deep understanding of data visualization with Matplotlib and Seaborn. - Acquire the skills to craft line charts with embellishments using key parameters. - Master diverse plot types like bar charts, pie charts, and box plots. - Learn to customize graph size, save images, and enhance appearance. - Gain proficiency in creating appealing visualizations using Matplotlib. - Explore Seaborn for producing advanced visualizations, including box plots. **Power BI:** - Cultivate proficiency in utilizing Power BI for data representation. - Acquire skills to transform and summarize data efficiently. - Master the art of combining and linking data tables for comprehensive analysis. - Learn to visualize data using a range of charts, from column and bar charts to line charts. - Skillfully present data using funnel, waterfall, TreeMap charts, and scatter plots. - Develop the ability to employ visual data slicers for interactive data representation. **Creating Maps and Visualizing Geospatial Data:** - Gain hands-on experience in utilizing Folium for creating maps and geospatial visualizations. - Learn to mark specific locations on maps and generate choropleth maps for data representation. **Statistical Analysis and Probability** - Develop a sound understanding of statistical concepts and methods. - Acquire the ability to perform descriptive statistics like mean, median, and mode. - Learn correlation techniques such as Pearson's Correlation Coefficient. - Apply graphical tools like scatterplots to uncover correlations between variables. - Grasp the nuances of different probability distributions like Normal, Poisson, Exponential, and Bernoulli.	

	- Master hypothesis testing for drawing inferences and measuring statistical significance. - Gain a comprehensive understanding of conditional probability and Bayes' theorem. - Develop expertise in parameter estimation using maximum likelihood estimation. - Differentiate between various statistical machine learning models, including supervised and unsupervised learning.	
GN/IT/000 3: Machine Learning	**Introduction:** - Develop a clear grasp of the foundational concepts of machine learning. - Distinguish between various types of machine learning approaches: supervised, unsupervised, and reinforcement learning. - Understand the critical role of data preprocessing and the trade-off between bias and variance. **Regression:** - Acquire proficiency in applying linear and non-linear regression methodologies. - Apply multivariate regression to real-world data analysis scenarios. - Evaluate the effectiveness of regression models using R-square and goodness of fit metrics. - Utilize techniques like subset selection and regularization for enhanced model performance. - Comprehend the significance of regularization methods, including Lasso, Ridge, and Elastic Net. **Classification:** - Develop expertise in employing supervised classification techniques. - Apply feature engineering strategies to improve classification outcomes. - Master the art of model validation through techniques like cross-validation. - Understand and use classification metrics like precision, recall, F1-measure, accuracy, and AUC. - Grasp the theoretical foundation of decision theory, Naive Bayes, Bayesian networks, Decision Trees, Random Forests, k-NN, SVMs, and Neural Networks. **Unsupervised Learning: Clustering:** - Cultivate skills in unsupervised learning, particularly focusing on clustering techniques.	Th – 35 Pr – 65 OJT - 20 --- Total - 120

	- Attain proficiency in implementing K-means and hierarchical clustering. - Utilize techniques such as minimum spanning tree and K-nearest neighbors clustering. - Familiarize yourself with advanced techniques like BIRCH, CURE, DBSCAN, and association rule mining. - Explore practical case studies that demonstrate the practical applications of clustering techniques. **Association rule mining algorithms including apriori** **Case Studies**	
GN/IT/000 4: Deep Learning	**Introduction:** - Gain a comprehensive understanding of neural networks and their distinction from traditional machine learning approaches. - Explore the connection between biological neural networks and artificial neural networks, including the McCulloch Pitts Neuron and Thresholding Logic. **Perceptron:** - Grasp the fundamentals of the Perceptron, its architecture, weights, biases, and its role in solving Boolean functions. - Analyze error surfaces and comprehend the learning algorithm of the Perceptron. **Feed Forward Neural Network:** - Develop proficiency in constructing multi-layer feedforward neural networks. - Master the backpropagation algorithm for learning network parameters and utilizing gradient descent for optimization. - Understand the architectures and functions of various types of neural networks, including CNNs and RNNs. - Explore the nuances of activation functions and loss functions used in deep neural networks. **Convolutional Neural Networks (CNNs) Architectures:** - Acquire a deep understanding of the convolutional operation and its significance in neural networks. - Explore various CNN architectures like LeNet, AlexNet, VGGNet, GoogLeNet, and ResNet. - Grasp concepts related to overfitting, underfitting, bias, variance, and hyperparameter tuning. - Become proficient in applying techniques such as L1 and L2 regularization, dropout, early stopping, data normalization, and augmentation.	Th – 35 Pr – 70 OJT - 15 --- Total - 120

	Recurrent Neural Networks: - Develop a strong grasp of RNNs' relevance to temporal data processing. - Master the understanding of LSTMs (Long Short-Term Memory), GRUs (Gated Recurrent Units), and Bidirectional RNNs. - Apply RNNs in contexts such as sequence prediction, natural language processing, and time series analysis. **Open Ended Problem – Project** - Apply theoretical concepts to real-world projects through an open-ended problem-solving exercise.	
GN/IT/000 5: MLOps	**Introduction:** - Gain insights into the various stages of a machine learning system, from data collection to model development. - Understand the unique challenges faced in deploying machine learning models and the solutions offered by MLOps. - Comprehend the transformation from DevOps to MLOps, including the framework and software stack. **DevOps:** - Define the concept of DevOps and explore its significance in modern software development. - Explore the principles of DevOps and how they contribute to streamlined development, testing, and deployment. - Understand the core categories of DevOps tools and their role in automating various stages of the software development lifecycle. **Getting Started with AWS:** - Familiarize yourself with AWS and its components, including account registration, regions, and instance types. - Learn how to launch and connect to EC2 instances, while understanding the importance of security groups. - Develop the skills to navigate AWS, laying the foundation for deploying machine learning models in the cloud. **Containers:** - Differentiate between containers and virtual machines, discerning the benefits of containerization. - Gain proficiency in Docker through installation, container management, and custom image creation. - Explore advanced Docker techniques, including Docker Compose, networking, and creating images from Git repositories. **Docker Swarm:** - Understand Docker Swarm, its purpose, and capabilities for orchestrating containers.	Th – 35 Pr – 70 OJT - 15 --- Total - 120

	- Apply containerized service deployment strategies and implement rolling updates within a Swarm cluster. - Hone skills in Swarm management and automated deployment on AWS infrastructure. **Kubernetes:** - Develop a solid grasp of Kubernetes, its architecture, and cluster installation. - Delve into the Raft consensus algorithm and networking solutions within Kubernetes. - Explore Kubernetes objects such as Pods, Deployments, Services, and understand their roles.	
GN/IT/0102: Employability Skills	**Introduction:** - In today's competitive job market, being technically proficient is just the tip of the iceberg for a successful career in data science and machine learning. Employers are looking for candidates who possess a diverse skill set and can effectively contribute to their organizations. This chapter explores the key employability skills that are essential for data science and machine learning professionals. **Technical Proficiency:** - A strong foundation in programming languages such as Python or R, along with proficiency in relevant libraries and frameworks, is crucial. Data professionals should also be comfortable with data manipulation, cleaning, and preprocessing to ensure data quality and accuracy. **Statistical Knowledge:** - Understanding statistics and probability is fundamental for designing experiments, conducting hypothesis testing, and interpreting results. Statistical expertise is indispensable for making informed data-driven decisions. **Coding Best Practices:** - Writing clean, maintainable, and well-documented code is essential for collaboration and codebase management, ensuring that data solutions are scalable and sustainable. **Experimentation and A/B Testing:** - Experience in designing and conducting experiments to evaluate model performance and optimize outcomes is valuable for data-driven decision-making. **Resilience:** - Persistence and resilience are crucial for handling setbacks and iterating on solutions to challenging data problems, as data science projects often involve trial and error.	Th – 18 Pr – 32 OJT - 10 --- Total - 60
CONTENT & TEACHING UNITS: CORE (ELECTIVES)		

GN/IT/0007: Natural Language Processing	**Introduction to NLP:** - Develop an understanding of the various levels of natural language processing and their computational challenges. - Explore the real-life applications of NLP, including spell and grammar checkers, information extraction, question answering, and machine translation. - Grasp the significance of addressing language ambiguities and computational complexities in processing natural languages. **Text Processing:** - Learn techniques for handling language ambiguities through segmentation, stemming, and tokenization. - Gain proficiency in representing words and sentences, including word embeddings and linguistic structures like dependency parsing. - Develop the ability to preprocess and structure textual data for effective analysis and modeling. **Text Classification:** - Acquire knowledge of various text classification methods, including word window classification and neural networks. - Understand the fundamentals of N-gram language models, perplexity, and Viterbi algorithm for language modeling. - Explore recurrent neural networks, their challenges, and techniques to address gradient issues such as vanishing and exploding gradients. **Language Modelling:** - Comprehend the role and significance of language models in natural language processing. - Learn techniques for estimating model parameters and smoothing to improve language model performance. - Gain familiarity with advanced models such as LSTM, GRU, BERT, XLNet, and Transformers, and understand their applications in context modeling and generative tasks. **Machine Translation:** - Understand the evolution of machine translation approaches, including statistical and neural machine translation. - Develop proficiency in sequence-to-sequence modeling for tasks like machine translation and question-answering. - Explore attention mechanisms, natural language generation, and neural machine translation methods. **NLP Case Study**	Th – 40 Pr – 70 OJT - 10 --- Total – 120

NOS 7: Computer Vision	**Introduction:** - Gain a comprehensive understanding of computer vision and its applications in various domains. - Develop familiarity with digital image processing fundamentals, including basic mathematical operations and binary image processing techniques. - Understand the process of digital image formation and its mathematical representation, along with techniques for image transformation and filtering. **Digital Image Formation and low-level processing:** - Acquire proficiency in performing image enhancement and restoration techniques. - Learn to analyze and process images using spatial filtering and histogram processing. - Explore advanced topics like edge and corner detection, as well as Fourier Transform for frequency domain analysis. **Object Recognition:** - Understand the concepts of semantic segmentation and its applications in object recognition. - Gain insight into convolutional neural networks (CNNs) for object recognition tasks. - Learn about popular architectures such as FCN, Unet, and SegNet for semantic segmentation and object detection. **Motion and Video Analysis:** - Develop skills in motion analysis and background subtraction for video understanding. - Learn about optical flow techniques for estimating motion within video frames. - Gain familiarity with object tracking algorithms like Kalman Filter, MeanShift, and CamShift, and their applications in video stabilization and tracking. **Computer Vision for Biomedical Data Analysis:** - Explore the applications of computer vision in analyzing biomedical data such as ECG, MRI, and CT datasets. - Understand the principles behind hybrid deep learning models for biomedical signal classification. - Learn about benchmark deep learning algorithms for biomedical image segmentation, including architectures like SwinUNet, FANet, and MedT.	
TOTAL DURATION		Th – 203 Pr – 397 OJT - 90 ---

	Total – 690