Training Parameters

Sector	**IT-ITeS**
Sub-Sector	**Future Skills**
Occupation	**Artificial Intelligence & Big Data Analytics**
Country	**India**
NSQF Level	**6**
Aligned to NCO/ISCO/ISIC Code	**NCO-2015/NIL**
Minimum Educational Qualification and Experience	***Relevant Experience in job roles related to IT/AI/Big Data Analytics/Programming** **The relevant experience would include work, internship, and apprenticeship after completing relevant educational qualifications.** **** PG or UG or diploma with courses related to Engg./ Science** **Pursuing first year of 2-year PG Program after completing 3-year UG** degree** **OR** **Pursuing 1-year PG diploma after 3-year UG** degree** **OR** **Completed 4th year UG (in case of 4-year UG**with honours/ honours with research)** **OR** **Pursuing 4th year UG (in case of 4-year UG**with honours/ honours with research) and continuing education** **OR** **Completed 3-Year UG** Degree with 1 year of relevant experience*** **OR** **Previous Relevant qualification of NSQF level 5 with 3 years of relevant experience***
Pre-Requisite License or Training	**NA**
Minimum Job Entry Age	**21 years (not mandatory)**
Last Reviewed On	**22 Sep 2020**
Next Review Date	**22 Sep 2025**
NSQC Approval Date	**22 Sep 2020**

QP Version	**3.0**
Model Curriculum Creation Date	**22 Sep 2020**
Model Curriculum Valid Up to Date	**22 Sep 2025**
Model Curriculum Version	**3.0**
Minimum Duration of the Course	**690 hours**
Maximum Duration of the Course	**780 hours**

Program Overview

This section summarizes the end objectives of the program along with its duration.

Training Outcomes

At the end of the program, the learner should have acquired the listed knowledge and skills.

Compulsory:

- Describe the use cases of AI & Big Data Analytics in various industries and define the various roles under this occupation.
- Apply basic and advanced statistical concepts used for data sciences such as Bayesian concepts, Conditional probability, Prior and Posterior probabilities etc.
- Apply different methods to import, pre-process and explore data such as importing data from different formats, cleaning data, and summarizing data, dimension reduction and defining correlations.
- Conduct research and design on different algorithms for a variety of data formats such as graphs or strings.
- Design complex algorithms such as deep neural networks, convolutional neural networks and recurrent neural networks for use cases such as image and speech recognition.
- Use statistical tools such as statistical integrated development environments (IDEs), or software packages, libraries and frameworks for importing, preprocessing, exploring data and designing models
- Apply suitable algorithms to implement various use cases such as chatbot development, image annotation, speech recognition etc.
- Plan their schedules and timelines based on the nature of work.
- Demonstrate how to communicate and work effectively with colleagues.
- Use different approaches to effectively manage and share data and information.
- Develop strong relationships at the workplace through effective communication and conflict management.
- Interpret client requirements clearly and deliver them to improve client satisfaction.
- Apply the principles of persuasive communication for negotiations and discussions.
- Identify best practices to maintain an inclusive, environmentally sustainable workplace

Electives:

- Predict model risk by identifying the risk factors and define mitigation measures
- Apply different methods to optimize model performance such as mini-batch gradient descent, RMSprop and Adam.
- Assess the most appropriate way to report data such as by identifying the right audience, creating a narrative and selecting suitable visualizations.

Compulsory Modules

The table lists the modules and their duration corresponding to the Compulsory NOS of the QP.

NOS and Module Details	Theory Duration (In Hours)	Practical Duration (In Hours)	On-the-Job Training Duration (Mandatory)	On-the-Job Training Duration (Recommended)	Total Duration (In Hours)
Module 1 (*Bridge Module*): *Artificial Intelligence & Big Data Analytics – An Introduction*	04:00	08:00	00:00	00:00	12:00
Module 2 (*Bridge Module*): *Basic Statistical Concepts*	08:00	16:00	00:00	00:00	24:00
Module 3 (*Bridge Module*): *Advanced Statistical Concepts*	10:00	20:00	00:00	00:00	30:00
Module 4 (*Bridge Module*): *Statistical Tools and Usage*	08:00	16:00	00:00	00:00	24:00
SSC/N8101 – Import data as per specifications NOS Version No. 2 NSQF Level 5	**40:00**	**50:00**	**00:00**	**00:00**	**90:00**
Module 5: Importing Data	40:00	50:00	00:00	00:00	90:00
SSC/N8102 – Preprocess data as per specifications NOS Version No. 2 NSQF Level 5	**50:00**	**100:00**	**00:00**	**00:00**	**150:00**
Module 6: Preprocessing Data	50:00	100:00	00:00	00:00	60:00
SSC/N8103 – Perform exploratory data analysis as per specifications. NOS Version No. 2 NSQF Level 5	**10:00**	**20:00**	**00:00**	**00:00**	**30:00**
Module 7: Exploring Data	10:00	20:00	00:00	00:00	30:00
SSC/N8104 – Perform research and design of algorithmic models NOS Version No. 3 NSQF Level 6	**30:00**	**60:00**	**00:00**	**00:00**	**90:00**
Module 8: Data Structures and Algorithms	06:00	12:00	00:00	00:00	18:00
Module 9: Graph Algorithms	06:00	12:00	00:00	00:00	18:00
Module 10: String Algorithms	06:00	12:00	00:00	00:00	18:00
Module 11: Neural Networks	06:00	12:00	00:00	00:00	18:00
Module 12: Programming for Data Science	06:00	12:00	00:00	00:00	18:00
SSC/N8105 – Apply pre-designed algorithmic models to specified use cases NOS Version No. 2 NSQF Level 6	**10:00**	**20:00**	**00:00**	**00:00**	**30:00**

Module 13: Applications of pre-designed Algorithms	10:00	20:00	00:00	00:00	30:00
SSC/N9014 Maintain an inclusive, environmentally sustainable workplace NOS Version No. 1 NSQF Level 5	**10:00**	**20:00**	**00:00**	**00:00**	**30:00**
Module 14: Inclusive and environmentally sustainable workplaces	10:00	20:00	00:00	00:00	30:00
DGT/VSQ/N0102 Employability Skill 60 Hours NOS Version No. 1 NSQF Level 4	**24:00**	**36:00**	**00:00**	**00:00**	**60:00**
Module 15: Introduction to Employability Skills	00:30	01:00	00:00	00:00	01.50
Module 16: Constitutional values - Citizenship	00:30	01:00	00:00	00:00	01.50
Module 17: Becoming a Professional in the 21st Century	01:00	01:30	00:00	00:00	02.50
Module 18: Basic English Skills	04:00	06:00	00:00	00:00	10.00
Module 19: Career Development & Goal Setting	01:00	01:00	00:00	00:00	02.00
Module 20: Communication Skills	02:00	03:00	00:00	00:00	05.00
Module 21: Diversity & Inclusion	01:00	01:30	00:00	00:00	02.50
Module 22: Financial and Legal Literacy	02:00	03:00	00:00	00:00	05.00
Module 23: Essential Digital Skills	04:00	06:00	00:00	00:00	10.00
Module 24: Entrepreneurship	03:00	04:00	00:00	00:00	07.00
Module 25: Customer Service	02:00	03:00	00:00	00:00	05.00
Module 26 : Getting ready for apprenticeship & Jobs	03:00	05:00	00:00	00:00	08.00
OJT	**00:00**	**00:00**	**120:00**	**00:00**	**120:00**
Total Duration	**204:00**	**366:00**	**120:00**	**00:00**	**690:00**

Elective Modules

Elective 1: Model Risk Assessment

NOS and Module Details	Theory Duration (In Hours)	Practical Duration (In Hours)	On-the-Job Training Duration (Mandatory)	On-the-Job Training Duration (Recommended)	Total Duration (In Hours)
SSC/N8106 – Evaluate risk of deploying algorithmic models NOS Version No. 2 NSQF Level 6	10:00	20:00	00:00	00:00	30:00
Elective 1: Identifying Model Risk	10:00	20:00	00:00	00:00	30:00
Total Duration	**10:00**	**20:00**	**00:00**	**00:00**	**30:00**

Elective 2: Model Business Performance

NOS and Module Details	Theory Duration (In Hours)	Practical Duration (In Hours)	On-the-Job Training Duration (Mandatory)	On-the-Job Training Duration (Recommended)	Total Duration (In Hours)
SSC/N8107 – Evaluate business performance of algorithmic models NOS Version No. 2 NSQF Level 6	10:00	20:00	00:00	00:00	30:00
Elective 2: Measuring Model Performance	10:00	20:00	00:00	00:00	30:00
Total Duration	**10:00**	**20:00**	**00:00**	**00:00**	**30:00**

Elective 3: Visualizations

NOS and Module Details	Theory Duration (In Hours)	Practical Duration (In Hours)	On-the-Job Training Duration (Mandatory)	On-the-Job Training Duration (Recommended)	Total Duration (In Hours)
SSC/N8108 – Define business outcomes and create visualizations from results of the analysis NOS Version No. 2 NSQF Level 5	10:00	20:00	00:00	00:00	30:00
Elective 3: Create Visualizations	10:00	20:00	00:00	00:00	30:00
Total Duration	**10:00**	**20:00**	**00:00**	**00:00**	**30:00**

Module Details

Module 1: Artificial Intelligence & Big Data Analytics – An Introduction

Bridge Module

Terminal Outcomes:

- Explain fundamental use cases of AI/Bigdata, types of AI systems and types of roles under this occupation

Duration (In Hours): *04:00*	**Duration (In Hours)**: *08:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Explain the relevance of AI & Big Data Analytics for the society • Explain the various use-cases of AI & Big Data in the industry • Define "general" and "narrow" AI • Describe the fields of AI such as image processing, computer vision, robotics, NLP, etc.	• Outline a career map for roles in AI & Big Data Analytics • Analyse the differences between key terms such as Supervised Learning, Unsupervised Learning and Deep Learning
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)	

Module 2: Basic Statistical Concepts

Bridge Module

Terminal Outcomes:

- Distinguish between various basic statistical concepts
- Apply different statistical techniques

Duration (In Hours): *8:00*	**Duration (In Hours):** *16:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Distinguish between different probability distributions such as Normal, Poisson, Exponential, Bernoulli, etc. • Identify correlation between variables using scatterplots and other graphical techniques	• Apply basics of descriptive statistics including measures of central tendency such as mean, median and mode • Apply different correlation techniques such as Pearson's Correlation Coefficient, Methods of Least Squares etc. • Apply different techniques for regression analysis including linear, logistic, ridge, lasso, etc. • Use hypothesis testing to draw inferences and measure statistical significance
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs such as RStudio, Jupyter Notebooks, etc. • Software libraries such as NumPy and Pandas, etc. • Statistical software tool such as IBM SPSS Statistics, SataCorp Stata, Mathworks, etc.	

Module 3: Advanced Statistical Concepts

Bridge Module

Terminal Outcomes:

- Distinguish between the applications of different statistical machine learning models

Duration (In Hours): *10:00*	**Duration (In Hours)**: *20:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Distinguish between prior and posterior measures of probability • Comprehend the difference between supervised and unsupervised learning • Describe the suitable statistical models for structured and unstructured data	• Use maximum likelihood estimation to estimate the parameters of a statistical model • Apply the concepts of conditional probability including Bayes theorem
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs such as RStudio, Jupyter Notebooks, etc. • Software libraries such as NumPy and Pandas, etc. • Statistical software tool such as IBM SPSS Statistics, SataCorp Stata, Mathworks, etc.	

Module 4: Statistical Tools and Usage

Bridge Module

Terminal Outcomes:

- Assess the various types of statistical tools available.
- Assess the various types of statistical packages available.

Duration (In Hours): *08:00*	**Duration (In Hours)**: *16:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Discover the basics of using statistical software packages and IDEs such as RStudio, Jupyter Notebooks	• Apply basic functions and libraries present in statistical software packages and IDEs • Make use of statistical packages, frameworks and libraries such as NumPy and Pandas in developing applications
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs such as RStudio, Jupyter Notebooks, etc. • Software libraries such as NumPy and Pandas, etc. • Statistical software tool such as IBM SPSS Statistics, SataCorp Stata, Mathworks, etc.	

Module 5: Importing Data

Mapped to SSC/N8101(Version No. 2)

Terminal Outcomes:

- Demonstrate reading and importing data from various file formats
- Perform data importing from various internal and external data sources
- Analyze the various uses of metadata

Duration (In Hours): *40:00*	**Duration (In Hours)**: *50:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Identify the type of data, volume of data, and variables required for the analysis • Distinguish between different types of data such as numerical, categorical, etc. • Identify common open and paid data sources • Discuss the uses and characteristics of different open source and paid data sources • Describe the purpose of metadata • Describe various Data validation tools and processes	• Demonstrate the process of capturing various types of data such as enterprise data, consumer data etc. from various data sources • Conduct the process of importing data from both public and private databases or data stores and store it in datasets or data frames • Organize and map metadata as per the needs of the analysis • Perform data profiling for data quality assessment and validation
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • Data Management tools such as MS-Excel, MySQL, etc. • Analytical and BI software such as Oracle Business Intelligence (Enterprise Edition), IBM Cognos, SAS, IBM SPSS Statistics, etc.	

Module 6: Pre-processing Data

Mapped to SSC/N8102 (Version No. 2)

Terminal Outcomes:

- Explain the fundamentals of pre-processing data
- Demonstrate the analysis of unprocessed data
- Evaluate different techniques to process data

Duration (In Hours): *50:00*	**Duration (In Hours)**: *100:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Differentiate the unprocessed and processed data • Explain the impact of unprocessed data on subsequent analytical operations • Describe the various anomalies that may be found in unprocessed data (e.g. missing values, incorrect data types, and redundant data) • Explain the Data Normalization techniques and concepts • Describe the properties of different tools that can be used to validate the pre-processed data	• Analyze unprocessed data to discover anomalies such as missing values, incorrect data types, etc. • Apply different techniques and functions to clean unprocessed data including removing missing values, transforming incorrect data types, etc. • Apply different approaches to normalize datasets such as feature scaling etc. • Apply appropriate tools and techniques to perform pre-processed data validation
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • Data Management tools such as MS-Excel, MySQL, etc. • Analytical and BI software such as Oracle Business Intelligence (Enterprise Edition), IBM Cognos, SAS, IBM SPSS Statistics, etc.	

Module 7: Exploring Data

Mapped to SSC/N8103 (Version No. 2)

Terminal Outcomes:

- Demonstrate the use of various tools and their functions to summarize data

Duration (In Hours): *10:00*	**Duration (In Hours):** *20:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Identify the key variables and data types required for modelling or analysis • Explain the limitations in exploring data of different types • Discuss the process of dimension reduction to optimize the variables in the dataset and define correlation factors • Categorize the various types of prescriptive actions that can be recommended from the results of a data analysis • Describe the principles of hypothesis testing to draw inferences from the results of a data analysis	• Apply different functions used to summarize data including mean, median, mode, range, variance, frequency • Select the right tool to explore the data based on its characteristics • Apply different approaches to perform dimension reduction on a dataset such as Principal Component Analysis, Linear Discriminant Analysis or Non-negative Matrix Factorization • Use graphical techniques such as scatterplots or clustering to evaluate correlations between different data points
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • Data Management tools such as MS-Excel, MySQL, etc. • Analytical and BI software such as Oracle Business Intelligence (Enterprise Edition), IBM Cognos, SAS, IBM SPSS Statistics, etc.	

Module 8: Data Structures and Algorithms

Mapped to NOS/N8104 (Version No. 3)

Terminal Outcomes:

- Describe the properties of various data structures such as arrays, linked lists, stacks, queues and trees
- Apply algorithms using various data structures to different problems

Duration (In Hours): *06:00*	**Duration (In Hours)**: *12:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Identify the data requirements of the analysis and suitable libraries, packages, frameworks, applications to address the objective • Distinguish between different data structures such as arrays, linked lists, stacks, queues and trees • Compare the differences in adding, removing and editing data from different types of data structures • Distinguish between the pros and cons of efficient and naïve algorithms	• Apply advanced concepts such as dynamic arrays, priority queues, disjoint sets and binary search trees to different types of problems • Use hash tables to store and modify sets of objects and mappings from one type of objects to another • Use the divide and conquer technique to solve problems involving large databases
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc.	

Module 9: Graph Algorithms

Mapped to SSC/N8104 (Version No. 3)

Terminal Outcomes:

- Apply the fundamentals of various types of graph algorithms

Duration (In Hours): *06:00*	**Duration (In Hours)**: *12:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Evaluate different algorithms for decomposing graphs into parts • Evaluate different algorithms for finding shortest paths in graphs such as breadth-first search, shortest-path-tree, Dijkstra's algorithm and Bellman-Ford algorithm • Evaluate greedy algorithms such as Kruskal's algorithm and Prim's algorithm to solve minimum spanning tree problems	• Apply the basics behind undirected graphs such as representing and exploring graphs, previsit and postvisit orderings etc. • Apply the basics behind directed graphs such as acyclic graphs, topological sorting and computing strongly connected components
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc.	

Module 10: String Algorithms

Mapped to SSC/N8104 (Version No. 3)

Terminal Outcomes:

- Evaluate the fundamentals of string algorithms such as pattern matching, brute force, and suffix trees and arrays

Duration (In Hours): *06:00*	**Duration (In Hours)**: *12:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Summarize the concepts behind algorithms such as suffix trees that are used for pattern matching	• Use brute force approaches for pattern matching • Use algorithms such as suffix arrays and Burrows-Wheeler Transform for approximate pattern matching • Use algorithms such as Knutt-Morris-Pratt for exact pattern matching • Apply different techniques to construct suffix trees and arrays
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc.	

Module 11: Neural Networks

Mapped to SSC/N8104 (Version No. 3)

Terminal Outcomes:

- Distinguish between the properties of different types of neural networks and their applications
- Build shallow and deep neural networks

Duration (In Hours): *06:00*	**Duration (In Hours)**: *12:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Distinguish between different types of recurrent neural networks and commonly used variants such as GRUs and LSTMs • Summarize the concepts behind convolutional neural networks and recurrent neural networks	• Build shallow and deep neural networks using techniques such as forward propagation and back propagation • Apply the foundational layers of convolutional neural networks such as pooling and convolutions and stack them properly in a deep network to solve multi-class image classification problems • Build convolutional neural networks and apply it to object detection problems • Use word vector representations and embedding layers to train recurrent neural networks • Apply attention model intuition and trigger word detection to speech recognition problems
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc. • Machine Learning and Neural Network frameworks and libraries such as Caffe, TensorFlow, Keras, etc.	

Module 12: Programming for Data Science

Mapped to SSC/N8104 (Version No. 3)

Terminal Outcomes:

- Evaluate various object-oriented and dynamic programming concepts
- Evaluate the fundamentals of programming with respect to developing machine learning algorithms

Duration (In Hours): *06:00*	**Duration (In Hours)**: *12:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Distinguish between the limitations of different programming, command line or scripting languages to develop machine learning algorithms • Select the most suitable programming languages to develop or optimize the statistical machine learning algorithm	• Use object-oriented programming concepts such as abstraction, encapsulation, modularity, etc. to write user defined functions and classes • Apply dynamic programming concepts to solve complex optimization problems • Use the streaming model to compute real-time or large amounts of data that cannot be stored in the memory
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc. • Machine Learning and Neural Network frameworks and libraries such as Caffe, TensorFlow, Keras, etc.	

Module 13: Applications of pre-designed Algorithms

Mapped to SSC/N8105 (Version No. 2)

Terminal Outcomes:

- Evaluate the various applied science use cases and apply algorithms to those use cases

Duration (In Hours): *10:00*	**Duration (In Hours):** *20:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Identify the objective of the analysis and define the suitable predesigned algorithms, libraries, packages, frameworks, applications to address the objective • Categorize the different use cases for algorithms such as recommender engines, chatbots, image annotation • Distinguish between the pros and cons of different open source libraries and packages to satisfy specific use cases	• Apply algorithms to specific use cases and scenarios such as vision, text recognition, image recognition, natural language processing based on suitable open source libraries selected • Validate the models implemented using appropriate tools and processes • Optimize and iterate algorithm models to achieve quality of outputs
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc. • Machine Learning and Neural Network frameworks and libraries such as Caffe, TensorFlow, Keras, etc.	

Module 14: Inclusive and environmentally sustainable workplaces

Mapped to SSC/N9014 (Version No. 1)

Terminal Outcomes:

- Illustrate sustainable practices at workplace for energy efficiency and waste management
- Apply different approaches to maintain gender equality and increase inclusiveness for PwD

Duration (In Hours): *10:00*	**Duration (In Hours)**: *20:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Describe different approaches for efficient energy resource utilisation and waste management • Describe the importance of following the diversity policies • Identify stereotypes and prejudices associated with people with disabilities and the negative consequences of prejudice and stereotypes • Discuss the importance of promoting, sharing and implementing gender equality and PwD sensitivity guidelines at organization level	• Practice the segregation of recyclable, non-recyclable and hazardous waste generated • Demonstrate different methods of energy resource use optimization and conservation • Demonstrate essential communication methods in line with gender inclusiveness and PwD sensitivity
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)	

Module 15: Introduction to Employability Skills

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Discuss the Employability Skills required for jobs in various industries
- List different learning and employability related GOI and private portals and their usage

Duration :1.5 Hours (0.5 Theory + 1 Practical)

Module 16: Constitutional values - Citizenship

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Explain the constitutional values, including civic rights and duties, citizenship, responsibility towards society and personal values and ethics such as honesty, integrity, caring and respecting others that are required to become a responsible citizen
- Show how to practice different environmentally sustainable practices

Duration:1.5 Hours (0.5 Theory + 1 Practical)

Module 17: Becoming a Professional in the 21st Century

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Discuss importance of relevant 21st century skills.
- Exhibit 21st century skills like Self-Awareness, Behaviour Skills, time management, critical and adaptive thinking, problem-solving, creative thinking, social and cultural awareness, emotional awareness, learning to learn etc. in personal or professional life.
- Describe the benefits of continuous learning

Duration:2.5 Hours (2.5 Theory + 0 Practical)

Module 18: Basic English Skills

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Show how to use basic English sentences for everyday conversation in different contexts, in person and over the telephone
- Read and interpret text written in basic English
- Write a short note/paragraph / letter/e -mail using basic English

Duration: 10 Hours (5 Theory + 5 Practical)

Module 19: Career Development and Goal Setting

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Create a career development plan with well-defined short- and long-term goals

Duration: 2 Hours (1 Theory + 1 Practical)

Module 20: Communication skills

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Demonstrate how to communicate effectively using verbal and nonverbal communication etiquette.
- Explain the importance of active listening for effective communication
- Discuss the significance of working collaboratively with others in a team

Duration: 5 Hours (2 Theory + 3 Practical)

Module 21: Diversity and Inclusion

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Demonstrate how to behave, communicate, and conduct oneself appropriately with all genders and PwD
- Discuss the significance of escalating sexual harassment issues as per POSH

Duration: 2.5 Hours (2.5 Theory+ 0 Practical)

Module 22: Financial and Digital Literacy

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Outline the importance of selecting the right financial institution, product, and service
- Demonstrate how to carry out offline and online financial transactions, safely and securely

Duration: 5 Hours (2 Theory+ 3 Practical)

Module 23: Essential Digital Skills

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Describe the role of digital technology in today's life
- Demonstrate how to operate digital devices and use the associated applications and features, safely and securely
- Discuss the significance of displaying responsible online behaviour while browsing, using various social media platforms, e-mails, etc., safely and securely
- Create sample word documents, excel sheets and presentations using basic features
- utilize virtual collaboration tools to work effectively

Duration: 10 Hours (4 Theory+ 6 Practical)

Module 24: Entrepreneurship

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Explain the types of entrepreneurship and enterprises
- Discuss how to identify opportunities for potential business, sources of funding and associated financial and legal risks with its mitigation plan
- Describe the 4Ps of Marketing-Product, Price, Place and Promotion and apply them as per requirement
- Create a sample business plan, for the selected business opportunity

Duration: 7 Hours (3 Theory+ 4 Practical)

Module 25: Customer Service

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Describe the significance of analysing different types and needs of customers
- Explain the significance of identifying customer needs and responding to them in a professional manner.
- Discuss the significance of maintaining hygiene and dressing appropriately

Duration: 5 Hours (2 Theory+ 3 Practical)

Module 26: Getting Ready for Apprenticeship and Jobs

Mapped to NOS DGT/VSQ/N0102 (Version No. 1)

Key Learning Outcomes:

- Create a professional Curriculum Vitae (CV)
- Use various offline and online job search sources such as employment exchanges, recruitment agencies, and job portals respectively
- Discuss the significance of maintaining hygiene and confidence during an interview
- Perform a mock interview
- List the steps for searching and registering for apprenticeship opportunities

Duration: 8 Hours (3 Theory+ 5 Practical)

Elective 1: Identifying Model Risk

Mapped to SSC/N8106 (Version No.2)

Terminal Outcomes:

- Evaluate the various factors that contribute to algorithmic risk
- Apply different techniques to estimate model risks

Duration (In Hours): *10:00*	**Duration (In Hours)**: *20:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Define the purpose and metrics of the algorithm model • Describe the various factors that contribute to algorithmic risk such as flawed data or assumptions, coding errors, insufficient sample sizes • Analyze the impact that risk factors might have on the outcome of the algorithmic model • Categorize and document the various mitigation measures that can be introduced to counter each type of model risk • Select suitable checks and mitigation measures to counter the risk	• Compute deviation from expected outcomes of model by testing it with multiple inputs • Apply different techniques to estimate the risks involved when the model deviates from expected outcomes • Translate mitigation measures into a structured corrective action that can be communicated to the rest of the organization • Evaluate the model for all possible scenarios
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc. • Machine Learning and Neural Network frameworks and libraries such as Caffe, TensorFlow, Keras, etc.	

Elective 2: Measuring Model Performance

Mapped to SSC/N8107 (Version No. 2)

Terminal Outcomes:

- Identify various metrics to measure model performance
- Computer model performance based on these metrics

Duration (In Hours): *10:00*	**Duration (In Hours)**: *20:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Identify the objective of the model used for the analysis • Categorize the different performance metrics for algorithms based on different business outcomes • Develop methodological approaches for identifying model hyperparameters • Describe different hyperparameters that can maximize model performance	• Apply different techniques to identify hyperparameters such as grid search, random search, Bayesian optimization • Use different optimization algorithms such as mini-batch gradient descent, RMSprop, Adam etc. • Apply the concepts behind hyperparameter tuning, batch normalization etc. • Compute the performance of the model with regards to meeting the specified business outcome
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • IDEs for programming languages for data science (e.g., C++, Python, R, C#, Java, etc.) such as Jupyter Notebook, RStudio, Visual Studio, etc. • Machine Learning and Neural Network frameworks and libraries such as Caffe, TensorFlow, Keras, etc.	

Elective 3: Create Visualizations

Mapped to SSC/N8108 (Version No. 2)

Terminal Outcomes:

- Explain the importance of results of an analysis that contribute to meeting business outcomes
- Identify the different visualizations that can be used to support the reporting of analysis results
-

Duration (In Hours): *10:00*	**Duration (In Hours):** *20:00*
Theory – Key Learning Outcomes	**Practical – Key Learning Outcomes**
• Identify the right target audience to report the results of a data analysis • Define different delivery modes and format to report the results of a data analysis • Distinguish between the pros and cons of using a specific visualization to represent certain types of data • Evaluate business impact and disseminate relevant information to the concerned person	• Summarize the results of a data analysis into a clear narrative • Represent outcomes through visualizations using standard templates and tools such as Tableau, QlikView, d3js etc • Perform version control and maintain reports in a knowledge base
Classroom Aids:	
• Whiteboard and Markers • Chart paper and sketch pens • LCD Projector and Laptop for presentations	
Tools, Equipment and Other Requirements :	
Labs equipped with the following: • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) Tools and Programming Languages: • Data Visualization software such as IBM Cognos Impromptu, Oracle Business Intelligence Enterprise Edition, QlikView, Tableau, PowerBI, etc.	

Annexure

Trainer Requirements

1.	**Trainer's Qualification and experience in the relevant sector (in years)** *(as per NCVET guidelines)*	**Educational Qualification:** Graduate in any discipline, preferably Science/Computer Science/Electronics and Engineering /Information Technology **Industry Experience:** Minimum 2 years of relevant experience in data analysis, data science, big data, or AI **Training Experience:** Minimum 1 year of full-time training experience in data analysis, data science, big data, or AI **Certification:** "Trainer" mapped to the Qualification Pack "MEP/Q2601, V2.0" Minimum accepted score is 80% aggregate
2.	**Master Trainer's Qualification and experience in the relevant sector (in years)** *(as per NCVET guidelines)*	**Educational Qualification:** Graduate in any discipline, preferably Science/Computer Science/Electronics and Engineering /Information Technology **Industry Experience:** Minimum 2 years of relevant experience in data analysis, data science, big data, or AI **Training Experience:** Minimum 1 year of full-time training experience in data analysis, data science, big data, or AI **Certification:** " Trainer" mapped to the Qualification Pack "MEP/Q2602, V2.0" Minimum accepted score is 90% aggregate
3.	**Tools and Equipment Required for the Training**	☒Yes ☐No *(If "Yes", details to be provided in Annexure)*
4.	**In Case of Revised Qualification, details of Any Upskilling Required for Trainer**	NA

Assessor Requirements

1.	**Assessor's Qualification and experience in relevant sector (in years)** *(as per NCVET guidelines)*	**Educational Qualification:** Graduate in any discipline, preferably Science/Computer Science/Electronics and Engineering /Information Technology **Industry Experience:** Minimum 2 years of relevant experience in data analysis, data science, big data, or AI **Training Experience:** Minimum 1 year of full-time training experience in data analysis, data science, big data, or AI **Certification:** "Assessor" mapped to the Qualification Pack "MEP/Q2701, V2.0" Minimum accepted score is 80% aggregate

2.	**Proctor's Qualification and experience in relevant sector (in years)** *(as per NCVET*	**Educational Qualification:** Graduate in any discipline, preferably Science/Computer Science/Electronics and Engineering /Information Technology

	guidelines), (wherever applicable)	**Industry Experience:** Minimum 2 years of relevant experience in data analysis, data science, big data, or AI **Training Experience:** Minimum 1 year of full-time training experience in data analysis, data science, big data, or AI **Certification:** "Assessor" mapped to the Qualification Pack "MEP/Q2702, V2.0" Minimum accepted score is 80% aggregate
3.	**Lead Assessor's/Proctor's Qualification and experience in relevant sector (in years)** *(as per NCVET guidelines)*	**Educational Qualification:** Graduate in any discipline, preferably Science/Computer Science/Electronics and Engineering /Information Technology **Industry Experience:** Minimum 2 years of relevant experience in data analysis, data science, big data, or AI **Training Experience:** Minimum 1 year of full-time training experience in data analysis, data science, big data, or AI **Certification:** "Assessor" mapped to the Qualification Pack "MEP/Q2702, V2.0" Minimum accepted score is 90% aggregate
4.	**Assessment Mode** *(Specify the assessment mode)*	Either Online or in Classroom
5.	**Tools and Equipment Required for Assessment**	☒ Same as for training ☐ Yes ☐ No *(details to be provided in Annexure-if it is different for Assessment)*

Assessment Strategy

This section includes the processes involved in identifying, gathering and interpreting information to evaluate the learner on the required competencies of the program.

Assessment System Overview

A uniform assessment of job candidates as per industry standards facilitates progress of the industry by filtering employable individuals while simultaneously providing candidates with an analysis of personal strengths and weaknesses.

Assessment Criteria

Criteria for assessment for each Qualification Pack will be created by the Sector Skill Council. Each Performance Criteria (PC) will be assigned marks proportional to its importance in NOS. SSC will also lay down the proportion of marks for Theory and Skills Practical for each PC.

The assessment for the theory part will be based on a knowledge bank of questions created by the SSC. Assessment will be conducted for all compulsory NOS, and where applicable, on the selected elective/option NOS/set of NOS.

Guidelines for Assessment			
Testing Environment	**Tasks and Functions**	**Productivity**	**Teamwork**
• Carry out assessments under realistic work pressures that are found in the normal industry workplace (or simulated workplace). • Ensure that the range of materials, equipment and tools that learners use are current and of the type routinely found in the normal industry workplace (or simulated workplace) environments.	• Assess that all tasks and functions are completed in a way, and to a timescale, that is acceptable in the normal industry workplace. • Assign workplace (or simulated workplace) responsibilities that enable learners to meet the requirements of the NOS.	• Productivity levels must be checked to ensure that it reflects those that are found in the work situation being replicated.	• Provide situations that allow learners to interact with the range of personnel and contractors found in the normal industry workplace (or simulated workplace).

Assessment Quality Assurance framework

NASSCOM provides two assessment frameworks NAC and NAC-Tech.

NAC (NASSCOM Assessment of Competence)

NAC follows a test matrix to assess Speaking & Listening, Analytical, Quantitative, Writing, and Keyboard skills of candidates appearing for assessment.

NAC-Tech

NAC-Tech test matrix includes assessment of Communication, Reading, Analytical, Logical Reasoning, Work Management, Computer Fundamentals, Operating Systems, RDBMS, SDLC, Algorithms & Programming Fundamentals, and System Architecture skills.

Methods of Validation

To pass a QP, a trainee should score an average of 70% or more. In case of unsuccessful completion, the trainee may seek reassessment on the Qualification Pack.

Method of assessment documentation and access

The assessment agency will upload the result of assessment in the portal. The data will not be accessible for change by the assessment agency after the upload. The assessment data will be validated by SSC assessment team. After upload, only SSC can access this data.

Recommended Supplemental Readings

The learning modules covered in the Model Curriculum for AI-Data Scientist are designed to meet the expected outcomes as per the QP. While the modules aligned to NOS are focused on technical/ behavioral competencies, bridge modules cover the prerequisite/ preparatory topics that are indispensable to complete the course. However, to provide additional QP specific knowledge to the learners, the following supplemental readings on related topics are recommended. These readings will equip the learners with an understanding of advanced or ancillary concepts to take up more complex tasks as listed in the QP.

QP	Recommended Supplemental Reading
SSC/Q8104: AI- Data Scientist	1) Database Management Concepts 2) Data Visualization Techniques

References

Glossary

Term	Description
Key Learning Outcome	Key learning outcome is the statement of what a learner needs to know, understand and be able to do in order to achieve the terminal outcomes. A set of key learning outcomes will make up the training outcomes. Training outcome is specified in terms of knowledge, understanding (theory) and skills (practical application).
Training Outcome	Training outcome is a statement of what a learner will know, understand and be able to do **upon the completion of the training**.
Terminal Outcome	Terminal outcome is a statement of what a learner will know, understand and be able to do **upon the completion of a module.** A set of terminal outcomes help to achieve the training outcome.
National Occupational Standard	National Occupational Standard specify the standard of performance an individual must achieve when carrying out a function in the workplace
Persons With Disability	Persons with Disability are those who have long-term physical, mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others.
Integrated Development Environment	An integrated development environment is a software application that provides comprehensive facilities to computer programmers for software development.
Natural Language Processing	Natural Language Processing or NLP is a field of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages.

Acronyms and Abbreviations

Term	Description
QP	Qualification Pack
NSQF	National Skills Qualification Framework
NSQC	National Skills Qualification Committee
NOS	National Occupational Standards
SSC	Skill Sectors Councils
NASSCOM	National Association of Software & Service Companies
PwD	Persons with Disability
IDE	Integrated development environment
NLP	Natural Language Processing

1. Introduction to Data Science

- **Overview of Data Science**: Understanding what data science is, its applications, and its importance in today's world.
- **Roles & Responsibilities of a Data Scientist**: What data scientists do, and what skills they need.
- **Data Science Lifecycle**: Data Collection → Data Preparation → Data Analysis → Modeling → Evaluation → Deployment.
- **Key Tools and Technologies**: Introduction to key tools used in the data science field (e.g., Python, R, SQL, etc.).

2. Mathematics and Statistics for Data Science

- **Basic Statistics**: Descriptive statistics, probability distributions, measures of central tendency, and variance.
- **Probability Theory**: Bayes' Theorem, Conditional Probability, and the Law of Total Probability.
- **Statistical Inference**: Hypothesis testing, p-values, t-tests, confidence intervals.
- **Linear Algebra**: Matrix operations, eigenvalues, and eigenvectors.
- **Calculus**: Functions, derivatives, optimization (used for machine learning algorithms).

3. Data Wrangling and Preprocessing

- **Data Collection**: Methods of gathering data from various sources (web scraping, APIs, databases, CSVs, etc.).
- **Data Cleaning**: Handling missing data, data imputation, outlier detection, and removal.
- **Data Transformation**: Data normalization, standardization, scaling.
- **Data Formatting**: Structuring data, transforming categorical variables (one-hot encoding, label encoding).
- **Data Exploration**: Exploratory Data Analysis (EDA) with visualization techniques.

4. Data Visualization

- **Introduction to Visualization Tools**: Matplotlib, Seaborn (Python libraries), and Tableau/Power BI.
- **Visualizing Data**: Creating histograms, bar plots, line charts, scatter plots, box plots.
- **Advanced Visualizations**: Heatmaps, pair plots, violin plots, geographical plots.
- **Storytelling with Data**: Communicating insights from data effectively using visualizations.

5. Machine Learning Fundamentals

- **Supervised Learning**: Introduction to regression and classification algorithms.
  - **Linear Regression**, **Logistic Regression**.
  - **Decision Trees**, **Random Forests**, **Support Vector Machines (SVM)**.
  - **K-Nearest Neighbors (KNN)**, **Naive Bayes**.
- **Unsupervised Learning**: Clustering and dimensionality reduction techniques.
  - **K-Means Clustering**, **Hierarchical Clustering**.
  - **Principal Component Analysis (PCA)**, **t-SNE**.
- **Model Evaluation**: Accuracy, Precision, Recall, F1-Score, Confusion Matrix.
  - Cross-validation, ROC Curve, AUC.

6. Deep Learning & Neural Networks

- **Introduction to Neural Networks**: Structure of a neural network, activation functions, forward and backward propagation.
- **Types of Neural Networks**:
  - **Feedforward Neural Networks** (FNN).
  - **Convolutional Neural Networks (CNNs)** for image recognition.
  - **Recurrent Neural Networks (RNNs)** and Long Short-Term Memory (LSTM) for time series and text data.
- **Deep Learning Frameworks**: TensorFlow, Keras, PyTorch.
- **Training Deep Learning Models**: Gradient Descent, Backpropagation, Overfitting, Dropout.

7. Advanced Topics in Machine Learning

- **Ensemble Learning**: Random Forests, Gradient Boosting Machines (GBM), XGBoost, LightGBM, AdaBoost.
- **Model Tuning**: Hyperparameter tuning, GridSearchCV, RandomSearchCV.
- **Feature Engineering**: Creating new features to improve model performance.
- **Model Deployment**: How to deploy models for production, real-time predictions.
  - Tools: Flask, Docker, AWS, Azure, or Google Cloud.

8. Natural Language Processing (NLP)

- **Text Preprocessing**: Tokenization, stopwords removal, stemming, lemmatization.
- **Vectorization**: TF-IDF, Word2Vec, GloVe, FastText.
- **NLP Models**: Sentiment Analysis, Named Entity Recognition (NER), Text Classification.
- **Deep Learning in NLP**: Introduction to transformers and BERT (Bidirectional Encoder Representations from Transformers).

9. Big Data Technologies

- **Big Data Introduction**: Hadoop ecosystem, Spark, and their importance in data science.
- **Data Storage & Management**: HDFS, NoSQL databases (MongoDB, Cassandra).
- **Distributed Computing**: Introduction to parallel processing with Spark, PySpark.
- **Stream Processing**: Real-time data processing using Apache Kafka.

10. Capstone Project and Case Studies

- **Real-World Problem Solving**: Applying the knowledge to solve real-world data science problems (business, finance, healthcare, etc.).
- **Project Development**: Developing and presenting a comprehensive data science project that includes data wrangling, exploratory data analysis, machine learning, model evaluation, and deployment.
- **Case Studies**: Analyzing case studies of successful data science applications across different industries.

11. Industry Applications of Data Science

- **Data Science in Finance**: Fraud detection, credit scoring, algorithmic trading.
- **Data Science in Healthcare**: Predictive modeling for patient outcomes, medical imaging, drug discovery.
- **Data Science in Marketing**: Customer segmentation, recommendation systems, sentiment analysis.
- **Data Science in Supply Chain and Operations**: Demand forecasting, optimization, inventory management.

12. Career Development and Soft Skills

- **Building a Portfolio**: Showcasing projects and experience to potential employers.
- **Interview Preparation**: Data science interview techniques, common questions, and problem-solving approaches.
- **Communication Skills**: How to explain complex models and results to non-technical stakeholders.

13. Certification Exam

- **Mock Tests**: Practice exams covering all the course modules.
- **Final Assessment**: Comprehensive exam to evaluate the learner's understanding and skills in data science.
- **Certification**: Upon successful completion, candidates are awarded the NASSCOM-accredited Data Science certification.

Summary of Key Tools and Technologies Covered

- **Programming Languages**: Python, R, SQL.
- **Libraries/Frameworks**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Keras, PyTorch, NLTK.
- **Big Data Technologies**: Hadoop, Spark.
- **Cloud Platforms**: AWS, Azure, Google Cloud.

Conclusion:

This is a broad overview of what a **NASSCOM-accredited Data Science** certification course might include. The course aims to provide learners with the theoretical knowledge and practical skills required to excel in data science roles, covering key aspects from foundational mathematics and statistics to advanced machine learning and deep learning techniques.

Compact / simplifies / Technical syllabus

Module 1: Artificial Intelligence & Big Data Analytics – An Introduction

Duration: 9 hours

- **AI Overview**: Concepts of AI, types of AI, and applications across industries.
- **Big Data Analytics**: Introduction to Big Data, technologies like Hadoop, Spark, and how AI can enhance Big Data processing.
- **AI and Big Data in Data Science**: How AI tools are used to process and analyze vast datasets in real-time.
- **Real-World Applications**: AI-powered solutions for healthcare, finance, e-commerce, etc.

Module 2: Basic Statistical Concepts

Duration: 10 hours

- **Descriptive Statistics**: Measures of central tendency, variance, and standard deviation.
- **Probability Theory**: Basic probability, conditional probability, and probability distributions (e.g., normal distribution).
- **Inferential Statistics**: Introduction to sampling techniques, confidence intervals, and hypothesis testing.
- **Visualization**: Introduction to plotting data distributions (histograms, box plots, etc.).

Module 3: Advanced Statistical Concepts

Duration: 11 hours

- **Bayesian Statistics**: Bayes' Theorem and its application in predictive modeling.
- **Advanced Probability Distributions**: Understanding gamma, chi-squared, and student-t distributions.
- **Multivariate Analysis**: Techniques like MANOVA and principal component analysis (PCA).
- **Regression Models**: Logistic regression, multinomial regression, and model assumptions.

Module 4: Statistical Tools and Usage

Duration: 12 hours

- **Statistical Software**: Hands-on training in R, Python, and advanced tools like SPSS or SAS.
- **Modeling Techniques**: Building and validating models using statistical techniques.
- **Data Visualization**: Advanced techniques using tools like Matplotlib, Seaborn (Python), ggplot2 (R).
- **Hypothesis Testing**: Practical use of t-tests, chi-square tests, and ANOVA.

Module 5: Importing Data

Duration: 13 hours

- **Data Sources**: Connecting to different data sources like APIs, web scraping, databases (SQL/NoSQL), and flat files (CSV, Excel).
- **Python/R Libraries**: Using `pandas` (Python) and `readr` (R) for data import.
- **Database Connections**: Hands-on with SQL queries and database interaction (MySQL, PostgreSQL).

- **Web Scraping**: Using BeautifulSoup and Scrapy for data collection from websites.
- **API Integration**: Working with RESTful APIs (e.g., Twitter, Google).

Module 6: Pre-processing Data

Duration: 14 hours

- **Cleaning Data**: Handling missing values, data imputation, removing duplicates.
- **Data Transformation**: Normalization, standardization, and encoding categorical variables.
- **Feature Engineering**: Creating new features, dimensionality reduction techniques.
- **Handling Outliers**: Identifying and treating outliers to improve model performance.

Module 7: Exploring Data

Duration: 15 hours

- **Exploratory Data Analysis (EDA)**: Descriptive statistics and visual inspection of data.
- **Data Visualization**: Creating plots like scatter plots, pair plots, heatmaps, and more.
- **Statistical Testing**: Identifying correlations, testing relationships between variables.
- **Identifying Patterns**: Using EDA to reveal hidden trends in data and inform model development.

Module 8: Data Structures and Algorithms

Duration: 16 hours

- **Core Data Structures**: Arrays, linked lists, stacks, queues, trees, and graphs.
- **Algorithm Design**: Introduction to algorithm complexity (Big-O notation), recursion.
- **Searching and Sorting**: QuickSort, MergeSort, and binary search algorithms.
- **Tree and Graph Algorithms**: Traversal methods (BFS, DFS), shortest path algorithms, etc.
- **Optimization**: Approaches to optimize algorithms for efficiency.

Module 9: Graph Algorithms

Duration: 17 hours

- **Graph Basics**: Introduction to graphs, types (directed, undirected, weighted, bipartite).
- **Graph Traversal**: BFS, DFS, applications like web crawling and network analysis.
- **Shortest Path**: Dijkstra's algorithm, Bellman-Ford algorithm, and Floyd-Warshall.
- **Graph Optimization**: Minimum spanning trees (Kruskal's, Prim's) and maximum flow algorithms.
- **Graph Data Structures**: Adjacency matrix, adjacency list, and edge list.

Module 10: String Algorithms

Duration: 18 hours

- **String Matching**: Naive pattern matching, Knuth-Morris-Pratt (KMP), and Rabin-Karp algorithms.
- **Text Processing**: Regular expressions, tokenization, parsing, and text cleaning.
- **Advanced String Algorithms**: Suffix trees, tries, and text search optimization.
- **String Similarity**: Levenshtein distance, Jaccard index, and cosine similarity.
- **Applications**: Search engines, spell checkers, and bioinformatics.

Module 11: Neural Networks

Duration: 19 hours

- **Introduction to Neural Networks**: Neuron models, perceptrons, and multi-layer networks.
- **Training Neural Networks**: Backpropagation, gradient descent, and optimization techniques.
- **Types of Neural Networks**: Feedforward networks, CNNs, RNNs, LSTMs.
- **Deep Learning**: Introduction to deep learning models, autoencoders, and GANs.
- **Frameworks**: Using TensorFlow, Keras, and PyTorch for building neural network models.

Module 12: Programming for Data Science

Duration: Variable

- **Programming Basics**: Python or R fundamentals, data types, loops, functions.
- **Libraries for Data Science**: Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn for Python; dplyr, ggplot2, caret for R.
- **Data Wrangling**: Using Python and R to clean, manipulate, and analyze data.
- **Model Building**: Implementing machine learning algorithms (classification, regression).
- **Best Practices**: Version control with Git, debugging, code optimization.

Module 13: Applications of Pre-designed Algorithms

Duration: Variable

- **Overview**: Understanding how pre-designed algorithms are applied to real-world problems.
- **Algorithms for Classification**: K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees.
- **Clustering Algorithms**: K-Means, Hierarchical Clustering, DBSCAN, applications in segmentation.
- **Recommendation Systems**: Using collaborative filtering, content-based filtering, and hybrid methods.
- **Optimization Algorithms**: Genetic algorithms, simulated annealing, and their use in optimization problems.
- **Real-World Case Studies**: Implementing algorithms in industries like healthcare (predictive modeling), finance (fraud detection), e-commerce (recommendations), and marketing (customer segmentation).
