

Model Curriculum

AI – Data Scientist

(Elective – Model Risk Assessment / Model
Business Performance / Visualizations)

SECTOR: IT-ITeS
SUB-SECTOR: FUTURE SKILLS
OCCUPATION: ARTIFICIAL INTELLIGENCE & BIG DATA ANALYTICS
REF ID: SSC/Q8104, V1.0
NSQF LEVEL: 8

 Skill India कौशल भारत - कुशल भारत	 IT - ITeS SSC NASSCOM	 N-S-D-C National Skill Development Corporation Transforming the skill landscape
<h2>Certificate</h2>		
COMPLIANCE TO QUALIFICATION PACK – NATIONAL OCCUPATIONAL STANDARDS		
is hereby issued by the IT-ITeS Sector Skills Council NASSCOM		
for MODEL CURRICULUM		
Complying to the National Occupational Standards of Job Role / Qualification Pack: <u>'AI – Data Scientist'</u> QP No. <u>'SSC/Q8104 NSQF Level 8'</u>		
Date of Issuance: October 17 th 2018		 Authorised Signatory (IT-ITeS Sector Skills Council NASSCOM)
Valid Upto *: October 17 th 2019		
* Valid up to the next review date of the Qualitification Pack		

TABLE OF CONTENTS

1. Curriculum	01
2. Trainer Prerequisites	18
3. Assessment Criteria	19

AI – Data Scientist

CURRICULUM / SYLLABUS

This program is aimed at training candidates for the job of a “AI – Data Scientist”, in the “IT-ITeS” Sector/Industry and aims at building the following key competencies amongst the learner

Program Name	AI – Data Scientist		
Qualification Pack Name and Reference ID.	SSC/Q8104, V1.0		
Version No.	1.0	Version Update Date	17/10/2018
Pre-requisites to Training	Bachelor's Degree in Engineering / Technology / Statistics / Mathematics / Computer Science		
Training Outcomes	<p>After completing this programme, participants will be able to:</p> <p>Compulsory:</p> <ul style="list-style-type: none"> • Explain the nature of work across the IT-ITeS sector, the various sub sectors and their evolution. • Elaborate the various occupations under the Future Skills sub sector and the impact of these on organizations and businesses. • Discuss the growing importance of AI and Big Data Analytics and its impact on the society. • 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, preprocess 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 • 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. • Design a structured plan for self learning and development. 		

	<ul style="list-style-type: none">• 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. <p>Electives:</p> <ul style="list-style-type: none">• 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.
--	--

This course encompasses 11 out of 11 Compulsory National Occupational Standards (NOS), 3 out of 3 Electives of “AI – Data Scientist” Qualification Pack issued by “IT-ITeS Sector Skills Council”.

COMPULSORY NOS

Sr. No.	Module	Key Learning Outcomes	Equipment Required
1	<p>IT-ITeS Sector – An Introduction</p> <p>Theory Duration (hh:mm) 06:00</p> <p>Practical Duration (hh:mm) 03:00</p> <p>Corresponding NOS Code Bridge Module</p>	<ul style="list-style-type: none"> • Explain the relevance of the IT-ITeS sector • State the various sub-sectors in the IT-ITeS sector • Detail the nature of work performed across the sub-sectors • Identify and list organizations in the sector • Discuss the evolution of the sub sectors and the way forward • Explain the disruptions happening across the IT-ITeS sector 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD Projector and Laptop for presentations • Lab equipped with the following: - <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) • Chart paper and sketch pens
2	<p>Future Skills – An Introduction</p> <p>Theory Duration (hh:mm) 01:00</p> <p>Practical Duration (hh:mm) 01:00</p> <p>Corresponding NOS Code Bridge Module</p>	<ul style="list-style-type: none"> • Define the general overview of the Future Skills sub-sector • Describe the profile of the Future Skills sub-sector • Explain the various occupations under this sub-sector • List key trends across the occupations in this sub-sector • List various roles in the Future Skills sub-sector 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD Projector and Laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)



3	<p>Artificial Intelligence & Big Data Analytics – An Introduction</p> <p>Theory Duration (hh:mm) 04:00</p> <p>Practical Duration (hh:mm) 02:00</p> <p>Corresponding NOS Code Bridge Module</p>	<ul style="list-style-type: none"> • Explain the relevance of AI & Big Data Analytics for the society • Explain a general overview of AI & Big Data Analytics and the various roles under this occupation • Define a career map for roles in AI & Big Data Analytics • Explain the role of a Data Scientist and his/her key responsibilities • List the range of skills and behavior, expected from a Data Scientist • State the growth opportunities for a Data Scientist 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD Projector and Laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)
4	<p>Basic Statistical Concepts</p> <p>Theory Duration (hh:mm) 08:00</p> <p>Practical Duration (hh:mm) 12:00</p> <p>Corresponding NOS Code Bridge Module</p>	<ul style="list-style-type: none"> • 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. • Use scatterplots and other graphical techniques to identify correlation between variables • Apply different techniques for regression analysis including linear, logistic, ridge, lasso, etc. • Distinguish between different probability distributions such as Normal, Poisson, Exponential, Bernoulli, etc. • Use hypothesis testing to draw inferences and 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD Projector and Laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)

		measure statistical significance	
5	<p>Advanced Statistical Concepts</p> <p>Theory Duration (hh:mm) 13:00</p> <p>Practical Duration (hh:mm) 17:00</p> <p>Corresponding NOS Code Bridge Module</p>	<ul style="list-style-type: none"> Apply the concepts of conditional probability including Bayes theorem Distinguish between prior and posterior measures of probability Use maximum likelihood estimation to estimate the parameters of a statistical model Distinguish between the applications of different statistical machine learning models Comprehend the difference between supervised and unsupervised learning Select the most suitable models based on whether the data is structured or unstructured 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Lab equipped with the following: <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated)
6	<p>Statistical Tools and Usage</p> <p>Theory Duration (hh:mm) 08:00</p> <p>Practical Duration (hh:mm) 12:00</p> <p>Corresponding NOS Code Bridge Module</p>	<ul style="list-style-type: none"> Distinguish between the different type of statistical tools and software packages 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 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Lab equipped with the following: <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated) Latest versions of statistical software packages and IDEs

7	<p>Importing Data</p> <p>Theory Duration (hh:mm) 05:00</p> <p>Practical Duration (hh:mm) 20:00</p> <p>Corresponding NOS Code SSC/N8101</p>	<ul style="list-style-type: none"> Identify various commonly known open source and paid data sources Discuss the uses and characteristics of different open source and paid data sources Develop knowledge on capturing various types of data such as enterprise data, consumer data etc. Demonstrate how to read data from various file formats and import it Describe the purpose of metadata Develop knowledge on how to organize and map metadata as per the needs of the analysis Use different tools to import data from both public and private databases or data stores and store it in datasets or dataframes Distinguish between different types of data such as numerical, categorical, etc. 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Lab equipped with the following: <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated)
8	<p>Preprocessing Data</p> <p>Theory Duration (hh:mm) 15:00</p> <p>Practical Duration (hh:mm) 35:00</p> <p>Corresponding NOS Code SSC/N8102</p>	<ul style="list-style-type: none"> Explain the difference between unprocessed and processed data Describe the various anomalies that may be found in unprocessed data Comprehend the impact of unprocessed data on subsequent analytical operations Describe the properties of different tools that can be used to preprocess data Analyze unprocessed data to discover anomalies such as missing values, incorrect data types, etc. 	<ul style="list-style-type: none"> Whiteboard and Markers LCD projector and laptop for presentations Lab equipped with the following: <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated)

		<ul style="list-style-type: none"> 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. 	
9	<p>Exploring Data</p> <p>Theory Duration (hh:mm) 15:00</p> <p>Practical Duration (hh:mm) 35:00</p> <p>Corresponding NOS Code SSC/N8103</p>	<ul style="list-style-type: none"> Explain the limitations found in exploring data of different types such as numerical or categorical Describe the properties of various tools that can be used to explore data Select the right tool to explore the data based on its characteristics Apply different functions used to summarize data including mean, median, mode, range, variance, frequency 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 Demonstrate the principles of hypothesis testing to draw inferences from the results of a data analysis Categorize the various types of prescriptive actions that can be recommended from the results of a data analysis 	<ul style="list-style-type: none"> Whiteboard and Markers LCD projector and laptop for presentations Lab equipped with the following: <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated)



<p>10</p>	<p>Data Structures and Algorithms</p> <p>Theory Duration (hh:mm) 15:00</p> <p>Practical Duration (hh:mm) 35:00</p> <p>Corresponding NOS Code SSC/N8104</p>	<ul style="list-style-type: none"> • Distinguish between different data structures such as arrays, linked lists, stacks, queues and trees • Comprehend the properties of various 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 • 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 • Distinguish between the pros and cons of efficient and naïve algorithms • Apply greedy algorithms to different use cases • Use the divide and conquer technique to solve problems involving large databases 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)
<p>11</p>	<p>Graph Algorithms</p> <p>Theory Duration (hh:mm) 15:00</p> <p>Practical Duration (hh:mm) 35:00</p> <p>Corresponding NOS Code SSC/N8104</p>	<ul style="list-style-type: none"> • 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 • Use different algorithms for decomposing graphs into parts • Use different algorithms for finding shortest paths in graphs such as breadth-first 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)



		<p>search, shortest-path-tree, Dijkstra's algorithm and Bellman-Ford algorithm</p> <ul style="list-style-type: none"> • Use greedy algorithms such as Kruskal's algorithm and Prim's algorithm to solve minimum spanning tree problems 	
12	<p>String Algorithms</p> <p>Theory Duration (hh:mm) 15:00</p> <p>Practical Duration (hh:mm) 35:00</p> <p>Corresponding NOS Code SSC/N8104</p>	<ul style="list-style-type: none"> • Use brute force approaches for pattern matching • Comprehend the concepts behind algorithms such as suffix trees that are used 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 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)
13	<p>Neural Networks</p> <p>Theory Duration (hh:mm) 25:00</p> <p>Practical Duration (hh:mm) 50:00</p> <p>Corresponding NOS Code SSC/N8104</p>	<ul style="list-style-type: none"> • Distinguish between the properties of different types of neural networks and their applications • 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 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)

		<ul style="list-style-type: none"> • Distinguish between different types of recurrent neural networks and commonly used variants such as GRUs and LSTMs • Use word vector representations and embedding layers to train recurrent neural networks • Apply attention model intuition and trigger word detection to speech recognition problems 	
14	<p>Programming for Data Science</p> <p>Theory Duration (hh:mm) 25:00</p> <p>Practical Duration (hh:mm) 50:00</p> <p>Corresponding NOS Code SSC/N8104</p>	<ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • Whiteboard and Markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) • Latest versions of statistical software packages and IDEs



15	<p>Manage Your Work to Meet Requirements</p> <p>Theory Duration (hh:mm) 12:00</p> <p>Practical Duration (hh:mm) 38:00</p> <p>Corresponding NOS Code SSC/N9001</p>	<ul style="list-style-type: none"> Define scope of work and working within limits of authority Summarize the details of the work and work environment Recognize the importance of maintaining confidentiality 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations
16	<p>Work Effectively with Colleagues</p> <p>Theory Duration (hh:mm) 12:00</p> <p>Practical Duration (hh:mm) 38:00</p> <p>Corresponding NOS Code SSC/N9002</p>	<ul style="list-style-type: none"> Use different methods and mechanisms for effective communication Recognize the importance of working effectively 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Provision to write emails and send in the lab Lab with provision for internet, email, word processor and presentation software Chart paper, markers, picture magazines and old newspapers
17	<p>Provide Data / Information in Standard Formats</p> <p>Theory Duration (hh:mm) 12:00</p> <p>Practical Duration (hh:mm) 38:00</p> <p>Corresponding NOS Code SSC/N9004</p>	<ul style="list-style-type: none"> Apply the concepts behind information and knowledge management Describe how data / information can be managed effectively Apply skills required to manage data and information effectively 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Provision for online research in the lab



18	<p>Develop Knowledge, Skills and Competence</p> <p>Theory Duration (hh:mm) 06:00</p> <p>Practical Duration (hh:mm) 19:00</p> <p>Corresponding NOS Code SSC/N9005</p>	<ul style="list-style-type: none"> Recognize the importance of self-development Identify knowledge and skills required for the job Identify avenues for self-development Create plans for self-development 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Provision for online access to all students in the lab
19	<p>Build and Maintain Relationships at the Workplace</p> <p>Theory Duration (hh:mm) 10:00</p> <p>Practical Duration (hh:mm) 15:00</p> <p>Corresponding NOS Code SSC/N9006</p>	<ul style="list-style-type: none"> Recognize the importance of open and effective communication Apply different approaches for conflict management Apply different approaches to boost recognition and motivation 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Lab with provision for internet, email, word processor and presentation software Chart paper, markers, picture magazines and old newspapers
20	<p>Build and Maintain Client Satisfaction</p> <p>Theory Duration (hh:mm) 10:00</p> <p>Practical Duration (hh:mm) 15:00</p> <p>Corresponding NOS Code SSC/N9007</p>	<ul style="list-style-type: none"> List different client requirements and use different approaches to gather them Demonstrate how to incorporate client feedback to improve quality of service 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Lab with provision for internet, email, word processor and presentation software Chart paper, markers, picture magazines and old newspapers



<p>21</p>	<p>Persuasive Communication</p> <p>Theory Duration (hh:mm) 10:00</p> <p>Practical Duration (hh:mm) 15:00</p> <p>Corresponding NOS Code SSC/N9010</p>	<ul style="list-style-type: none"> Identify different requirements and how to adapt to each distinct requirement Demonstrate how to use evidences to support arguments 	<ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Lab with provision for internet, email, word processor and presentation software Chart paper, markers, picture magazines and old newspapers
	<p>COMPULSORY NOS</p> <p>Total Duration:</p> <p>Theory Duration 242:00</p> <p>Practical Duration 520:00</p>	<p>Unique Equipment Required</p> <ul style="list-style-type: none"> Whiteboard and Markers LCD Projector and Laptop for presentations Lab equipped with the following: - <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated) Chart paper and sketch pens Latest version of statistical software packages and IDEs Chart paper, markers, picture magazines and old newspapers <p>Popular Software Tools</p> <p><u>BI Software:</u> IBM Cognos Impromptu, Oracle Business Intelligence Enterprise Edition</p> <p><u>Analytical software tools:</u> IBM SPSS Statistics, SAS, StataCorp Stata, MathWorks</p> <p><u>Data mining tools:</u> IBM InfoSphere Warehouse, RapidMiner</p> <p><u>Development Software:</u> Python, R, C++, Java</p> <p><u>Development Libraries or Platforms:</u> OpenCV, TensorFlow, Theano, Knime, Scikit-learn, Torch, Keras</p> <p><u>Data Management PaaS:</u> AWS, Hortonworks, Cloudera, Azure</p>	

ELECTIVES (Mandatory to select at least one title)

ELECTIVE 1: Model Risk Assessment

Sr. No.	Module	Key Learning Outcomes	Equipment Required
1	<p>Identifying Model Risk</p> <p>Theory Duration (hh:mm) 15:00</p> <p>Practical Duration (hh:mm) 35:00</p> <p>Corresponding NOS Code SSC/N8106</p>	<ul style="list-style-type: none"> Describe the various factors that contribute to algorithmic risk such as flawed data or assumptions, coding errors, insufficient sample sizes Comprehend the impact that risk factors might have on the outcome of the algorithmic model 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 Categorize 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 Translate mitigation measures into a structured corrective action that can be communicated to the rest of the organization 	<ul style="list-style-type: none"> Whiteboard and markers LCD projector and laptop for presentations Lab equipped with the following: <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated)
	<p>ELECTIVE 1</p> <p>Total Duration:</p> <p>Theory Duration 15:00</p> <p>Practical Duration 35:00</p>	<p>Unique Equipment Required</p> <ul style="list-style-type: none"> Whiteboard and markers LCD projector and laptop for presentations Lab equipped with the following: <ul style="list-style-type: none"> PCs/Laptops Internet with Wi-Fi (Min 2 Mbps Dedicated) <p>Popular Software Tools</p>	

	<p><u>Data mining tools:</u> IBM InfoSphere Warehouse, RapidMiner</p> <p><u>Development Software:</u> Python, R, C++, Java</p> <p><u>Development Libraries or Platforms:</u> OpenCV, TensorFlow, Theano, Knime, Scikit-learn, Torch, Keras</p>
--	--

ELECTIVE 2: Model Business Performance

Sr. No.	Module	Key Learning Outcomes	Equipment Required
1	<p>Measuring Model Performance</p> <p>Theory Duration (hh:mm) 25:00</p> <p>Practical Duration (hh:mm) 50:00</p> <p>Corresponding NOS Code SSC/N8107</p>	<ul style="list-style-type: none"> • Categorize the different performance metrics for algorithms based on different business outcomes • Compute the performance of the model with regards to meeting the specified business outcome • 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. 	<ul style="list-style-type: none"> • Whiteboard and markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)
	<p>ELECTIVE 2</p> <p>Total Duration:</p> <p>Theory Duration 25:00</p> <p>Practical Duration 50:00</p>	<p>Unique Equipment Required</p> <ul style="list-style-type: none"> • Whiteboard and markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) <p>Popular Software Tools</p> <p><u>Data mining tools:</u> IBM InfoSphere Warehouse, RapidMiner</p> <p><u>Development Software:</u> Python, R, C++, Java</p> <p><u>Development Libraries or Platforms:</u> OpenCV, TensorFlow, Theano, Knime, Scikit-learn, Torch, Keras</p>	

ELECTIVE 3: Visualizations

Sr. No.	Module	Key Learning Outcomes	Equipment Required
1	<p>Create Visualizations</p> <p>Theory Duration (hh:mm) 15:00</p> <p>Practical Duration (hh:mm) 35:00</p> <p>Corresponding NOS Code SSC/N8108</p>	<ul style="list-style-type: none"> • Explain how the results of an analysis can contribute to meeting business outcomes • Categorize the different business outcomes that can be met from the results of a data analysis • Identify the right target audience to report the results of a data analysis • Identify the right delivery mode and format to report the results of a data analysis • Comprehend how content might change based on the target audience • Summarize the results of a data analysis into a clear narrative • Identify the different visualizations that can be used to support the reporting of analysis results • Distinguish between the pros and cons of using a specific visualization to represent certain types of data • Select the right tool to create the visualizations • Comprehend the importance of version control and uploading the report in a knowledge base 	<ul style="list-style-type: none"> • Whiteboard and markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated)



<p>ELECTIVE 3 Total Duration:</p> <p>Theory Duration 15:00</p> <p>Practical Duration 35:00</p>	<p>Unique Equipment Required</p> <ul style="list-style-type: none"> • Whiteboard and markers • LCD projector and laptop for presentations • Lab equipped with the following: <ul style="list-style-type: none"> • PCs/Laptops • Internet with Wi-Fi (Min 2 Mbps Dedicated) <p>Popular Software Tools</p> <p><u>Visualization tools:</u> QlikView, Tableau, Power BI</p>
--	---

<p>GRAND Total Duration</p> <p>Minimum Duration for the QP = <u>812 hrs</u></p> <p>Theory: <u>257 hrs</u> Practical: <u>555 hrs</u></p> <p>Maximum Duration for the QP= <u>937 hrs</u></p> <p>Theory: <u>297 hrs</u> Practical: <u>640 hrs</u></p>	<p>Unique Equipment Required for the QP:</p> <ul style="list-style-type: none"> • Whiteboard and Markers • LCD Projector and Laptop for presentations • Lab equipped with the following: - <ul style="list-style-type: none"> • PCs/Laptops • Internet with WiFi (Min 2 Mbps Dedicated) • Chart paper and sketch pens • Latest version of statistical software packages and IDEs • Chart paper, markers, picture magazines and old newspapers <p>Popular Software Tools</p> <p><u>BI Software:</u> IBM Cognos Impromptu, Oracle Business Intelligence Enterprise Edition</p> <p><u>Analytical software tools:</u> IBM SPSS Statistics, SAS, StataCorp Stata, MathWorks</p> <p><u>Data mining tools:</u> IBM InfoSphere Warehouse, RapidMiner</p> <p><u>Development Software:</u> Python, R, C++, Java</p> <p><u>Development Libraries or Platforms:</u> OpenCV, TensorFlow, Theano, Knime, Scikit-learn, Torch, Keras</p> <p><u>Data Management PaaS:</u> AWS, Hortonworks, Cloudera, Azure</p> <p><u>Visualization tools:</u> QlikView, Tableau, Power BI</p>
---	---

(This syllabus/ curriculum has been approved by **SSC: IT- ITeS Sector Skills Council NASSCOM**)

Trainer Prerequisites for Job role: “AI – Data Scientist” mapped to Qualification Pack: “SSC/Q8104, V1.0”

Sr. No.	Area	Details
1	Description	To deliver accredited training service, mapping to the curriculum detailed above, in accordance with the Qualification Pack <u>SSC/Q8104, V1.0</u>
2	Personal Attributes	This job may require the individual to work independently and take decisions for his/her own area of work. The individual should have a high level of analytical thinking ability, passion for artificial intelligence and big data analytics, and attention for detail, should be ethical, compliance and result oriented, should also be able to demonstrate interpersonal skills, along with willingness to undertake desk-based job with long working hours.
3	Minimum Educational Qualifications	Graduate in any discipline preferably Science/Computer Science/Electronics and Engineering /Information Technology
4a	Domain Certification	Certified for Job Role: “ <u>AI – Data Scientist</u> ” mapped to QP: “ <u>SSC/Q8104, V1.0</u> ”. Minimum accepted score is 80%
4b	Platform Certification	Recommended that the trainer is certified for the Job role “Trainer” mapped to the Qualification Pack “ <u>MEP/Q0102</u> ”. Minimum accepted score is 80% aggregate
5	Experience	10+ years of work experience/internship in Data Scientist or related roles that involve data analysis and modelling

Criteria For Assessment Of Trainees

Job Role AI – Data Scientist

Qualification Pack SSC/Q8104, V1.0

Sector Skill Council IT-ITeS

Guidelines for Assessment

1. 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 proportion of marks for Theory and Skills Practical for each PC.
2. The assessment for the theory part will be based on knowledge bank of questions created by the SSC.
3. Assessment will be conducted for all compulsory NOS, and where applicable, on the selected elective/option NOS/set of NOS.
4. Individual assessment agencies will create unique question papers for theory part for each candidate at each examination/training center (as per assessment criteria below).
5. Individual assessment agencies will create unique evaluations for skill practical for every student at each examination/training center based on this criterion.
6. To pass a QP, a trainee should score an average of 70% across generic NOS' and a minimum of 70% for each technical NOS
7. In case of *unsuccessful completion*, the trainee may seek reassessment on the Qualification Pack.

Compulsory NOS				Marks Allocation	
Total Marks: 1100					
Assessment outcomes	Assessment Criteria for outcomes	Total Marks	Out Of	Theory	Skills Practical
1. SSC/N8101 Import data as per specifications	PC1. identify the objective of the analysis	100	10	5	5
	PC2. define the type of data to be imported		7	2	5
	PC3. define the volume of data to be imported		8	3	5
	PC4. define the key variables to be imported		15	5	10
	PC5. identify suitable sources for the data		15	5	10
	PC6. perform operations to acquire the data and store it in datasets or data frames		20	5	15
	PC7. create and populate metadata for the imported data		15	5	10
	PC8. validate imported data using appropriate tools and processes		5	0	5
	PC9. validate the desired output with the relevant stakeholders within the organization, if required		5	0	5
	Total		100	30	70
2. SSC/N8102 Preprocess data as per specifications	PC1. define the format and structure for the dataset	100	5	0	5
	PC2. define indexes and organize variables as per the defined format		5	2	3
	PC3. identify the data types for each variable of the dataset		10	3	7
	PC4. identify and fix missing values in each variable of the dataset		15	5	10
	PC5. identify and fix incorrect data types in each variable of the dataset		15	5	10

	PC6. sort the data and create subsets of the data as required		15	5	10
	PC7. perform operations to transform data types of variables as required		15	5	10
	PC8. identify and deal with data redundancy by normalizing the dataset		15	5	10
	PC9. validate preprocessed data using appropriate tools and processes		5	0	5
	Total		100	30	70
3. SSC/N8103 Perform exploratory data analysis as per specifications	PC1. identify the data types for each variable of the dataset	100	5	2	3
	PC2. identify the key variables required for modelling or analysis		8	3	5
	PC3. use statistical techniques to summarize the key variables in the dataset		15	5	10
	PC4. describe summary statistics for key variables using graphical formats		8	2	6
	PC5. perform dimension reduction to optimize the variables in the dataset, if required		8	3	5
	PC6. define the correlation factors using clustering and other techniques		8	3	5
	PC7. validate data using appropriate tools and processes		5	0	5
	PC8. repeat the analysis iteratively to arrive at optimal results		8	2	6
	PC9. validate the final output in consultation with the relevant stakeholders		5	0	5
	PC10. gain inferences from the final output of the data analysis		7	2	5

	PC11. develop a hypothesis model to explain the discovered inferences		7	2	5
	PC12. evaluate the results of the analysis and define business outcomes		8	3	5
	PC13. define prescriptive actions based on the defined business outcomes		8	3	5
	Total		100	30	70
4. SSC/N8104 Design algorithmic models as per specifications	PC1. identify the objective of the analysis	100	2	0	2
	PC2. develop a hypothesis based on the objective of the analysis		3	1	2
	PC3. identify suitable libraries, packages, frameworks, applications to address the objective		4	1	3
	PC4. identify mode of learning, i.e. supervised or unsupervised		6	2	4
	PC5. conduct research on existing statistical models to evaluate fitment with the objective		6	2	4
	PC6. depending on the use case, identify if neural networks or deep learning models can be built		6	2	4
	PC7. optimize the existing statistical models as per need		6	2	4
	PC8. identify suitable statistical models on the basis of data volumes and key variables		6	2	4
	PC9. define connectors or combinations of key variables for each statistical model		6	2	4
	PC10. determine and collect the training data		6	2	4
	PC11. design and prototype algorithmic model		9	3	6
	PC12. identify and resolve overfitting or underfitting of algorithmic model		6	2	4

	PC13. identify and resolve residual and dispersion errors with data		6	2	4
	PC14. define data flows such as human-in-the-loop constraints required to reinforce algorithmic models		4	1	3
	PC15. define and quantify success metrics for the algorithmic model		6	2	4
	PC16. create documentation on designed algorithmic models for future references and versioning		4	0	4
	PC17. retrain datasets that have been used for supervised learning on a continuous basis		6	2	4
	PC18. validate designed models using appropriate tools and processes		4	0	4
	PC19. iterate the process to fine-tune the model till the desired quality of output or performance is achieved		4	0	4
	Total		100	28	72
5. SSC/N9001 Manage your work to meet requirements	PC1. establish and agree your work requirements with appropriate people	100	6.25	0	6.25
	PC2. keep your immediate work area clean and tidy		12.5	6.25	6.25
	PC3. utilize your time effectively		12.5	6.25	6.25
	PC4. use resources correctly and efficiently		18.75	6.25	12.5
	PC5. treat confidential information correctly		6.25	0	6.25
	PC6. work in line with your organization's policies and procedures		12.5	0	12.5
	PC7. work within the limits of your job role		6.25	0	6.25

	PC8. obtain guidance from appropriate people, where necessary		6.25	0	6.25
	PC9. ensure your work meets the agreed requirements		18.75	6.25	12.5
	Total		100	25	75
6. SSC/N9002 Work effectively with colleagues	PC1. communicate with colleagues clearly, concisely and accurately	100	20	0	20
	PC2. work with colleagues to integrate your work effectively with them		10	0	10
	PC3. pass on essential information to colleagues in line with organizational requirements		10	10	0
	PC4. work in ways that show respect for colleagues		20	0	20
	PC5. carry out commitments you have made to colleagues		10	0	10
	PC6. let colleagues know in good time if you cannot carry out your commitments, explaining the reasons		10	10	0
	PC7. identify any problems you have working with colleagues and take the initiative to solve these problems		10	0	10
	PC8. follow the organization's policies and procedures for working with colleagues		10	0	10
	Total		100	20	80
7. SSC/N9004 Provide data/information in standard formats	PC1. establish and agree with appropriate people the data/information you need to provide, the formats in which you need to provide it, and when you need to provide it	100	12.5	12.5	0
	PC2. obtain the data/information from reliable sources		12.5	0	12.5
	PC3. check that the data/information is accurate, complete and up-to-date		12.5	6.25	6.25

	PC4. obtain advice or guidance from appropriate people where there are problems with the data/information		6.25	0	6.25
	PC5. carry out rule-based analysis of the data/information, if required		25	0	25
	PC6. insert the data/information into the agreed formats		12.5	0	12.5
	PC7. check the accuracy of your work, involving colleagues where required		6.25	0	6.25
	PC8. report any unresolved anomalies in the data/information to appropriate people		6.25	6.25	0
	PC9. provide complete, accurate and up-to-date data/information to the appropriate people in the required formats on time		6.25	0	6.25
	Total		100	25	75
8. SSC/N9005 Develop your knowledge, skills and competence	PC1. obtain advice and guidance from appropriate people to develop your knowledge, skills and competence	100	10	0	10
	PC2. identify accurately the knowledge and skills you need for your job role		10	0	10
	PC3. identify accurately your current level of knowledge, skills and competence and any learning and development needs		20	10	10
	PC4. agree with appropriate people a plan of learning and development activities to address your learning needs		10	0	10
	PC5. undertake learning and development activities in line with your plan		20	10	10
	PC6. apply your new knowledge and skills in the workplace, under supervision		10	0	10
	PC7. obtain feedback from appropriate people on your		10	0	10

	knowledge and skills and how effectively you apply them				
	PC8. review your knowledge, skills and competence regularly and take appropriate action		10	0	10
	Total		100	20	80
9. SSC/N9006 Build and maintain relationships at the workplace	PC1. build rapport with appropriate people at the workplace	100	10	3	7
	PC2. develop new professional relationships		10	3	7
	PC3. build alliances to establish mutually beneficial working arrangements		10	3	7
	PC4. foster an environment where others feel respected		10	4	6
	PC5. identify and engage a diverse range of influential contacts		10	4	6
	PC6. obtain guidance from appropriate people, wherever necessary		10	3	7
	PC7. attentively listen to ideas and give constructive feedback		10	3	7
	PC8. promptly resolve conflicts between self or others		10	2	8
	PC9. work with colleagues to deliver shared goals		10	2	8
	PC10. recognize the contributions made by your colleagues		10	3	7
	Total		100	30	70
10. SSC/N9007 Build and maintain client satisfaction	PC1. gather client context and requirements	100	15	5	10
	PC2. manage fluctuating client priorities and expectations		25	10	15
	PC3. respond to client requests in a timely and accurate manner		20	5	15
	PC4. continuously improve personal service based on client feedback		15	0	15

	PC5. plan deliverables based on client needs		25	10	15
	Total		100	30	70
11. SSC/N9010 Convince others to take appropriate action in different situations	PC1. gather needs of concerned people	100	10	0	10
	PC2. adapt arguments to consider diverse needs		15	0	15
	PC3. use small wins as milestones to gain support for ideas		25	10	15
	PC4. persuade with the help of concrete examples or evidences		25	10	15
	PC5. take structured actions to reach consensus on the course of action		25	10	15
	Total		100	30	70

ELECTIVES					
Elective 1 – Model Risk Assessment				Marks Allocation	
Total Marks: 100					
Assessment outcomes	Assessment Criteria for outcomes	Total Marks	Out Of	Theory	Skills Practical
1. SSC/N8106 Evaluate risk of deploying algorithmic models	PC1. define the purpose and metrics for the algorithmic model	100	10	3	7
	PC2. define data sources used to design the model and data flows used to reinforce the model		10	3	7
	PC3. define and evaluate the assumptions used while designing the algorithmic model		10	3	7
	PC4. evaluate the range of expected outcomes of the algorithmic model		10	3	7
	PC5. test the model with different inputs and identify the factors that are		10	3	7

	creating a deviation from the expected outcomes				
	PC6. estimate the risk involved in the cases where the algorithmic model deviates from the expected outcomes		20	8	15
	PC7. introduce checks and mitigation measures for each of the potential risks to the algorithmic model		15	5	10
	PC8. create documentation on potential risks and the associated mitigation measures		5	0	5
	PC9. validate risks and mitigation measures with appropriate stakeholders		2	0	2
	PC10. recommend and implement corrective actions to the model as required		3	0	3
	PC11. evaluate the model for all possible use cases / scenarios		5	2	3
	Total		100	30	70
Elective 2 – Model Business Performance					
Total Marks: 100				Marks Allocation	
Assessment outcomes	Assessment Criteria for outcomes	Total Marks	Out Of	Theory	Skills Practical
1. SSC/N8107 Evaluate business performance of algorithmic models	PC1. identify the objective being addressed by the model	100	10	3	7
	PC2. define suitable evaluation criteria and metrics to evaluate model performance as per objective		10	3	7

	PC3. evaluate the performance of the algorithmic model		20	6	14
	PC4. identify the hyperparameters to maximize model performance		20	6	14
	PC5. test different hyperparameter configurations		20	6	14
	PC6. use best-fit hyperparameter configuration to maximize model performance		20	6	14
	Total		100	30	70
Elective 3 – Visualizations					
Total Marks: 100				Marks Allocation	
Assessment outcomes	Assessment Criteria for outcomes	Total Marks	Out Of	Theory	Skills Practical
1. SSC/N8108 Define business outcomes and create visualizations	PC1. identify the objective of the analysis	100	15	5	10
	PC2. establish the purpose, scope, and target audience to report the business outcomes		15	5	10
	PC3. define the delivery mode and format (such as excel sheets, reports, APIs) to report the business outcomes		15	5	10
	PC4. summarize the defined business outcomes into a narrative		10	2	8
	PC5. select suitable visualizations to represent the defined business outcomes		10	3	7

	PC6. represent outcomes through visualizations using standard templates and agreed language standards		25	10	15
	PC7. validate visualizations with appropriate people		5	0	5
	PC8. publish visualizations for consumption across all agreed formats		5	0	5
	Total		100	30	70