ที่ ศธ 0604/๓๓๕

สำนักนโยบายและแผนการอาชีวศึกษา
สำนักงานคณะกรรมการการอาชีวศึกษา
กระทรวงศึกษาธิการ กทม. 10300

๒๖ ตุลาคม 2564

เรื่อง   ทุนฝึกอบรมออนไลน์หลักสูตร Data Science and Big Data Analytics

เรียน   ผู้อำนวยการสถาบันการอาชีวศึกษาและผู้อำนวยการสถานศึกษาในสังกัดทุกแห่ง

สิ่งที่ส่งมาด้วย   รายละเอียดทุนฝึกอบรมออนไลน์หลักสูตร Data Science and Big Data Analytics
จำนวน 1 ฉบับ

        ด้วยรัฐบาลสาธารณรัฐอินเดียจะจัดการฝึกอบรมออนไลน์หลักสูตร Data Science and Big Data
Analytics ภายใต้ e-ITEC Programme ระหว่างวันที่ 7 กุมภาพันธ์-18 มีนาคม 2565 ซึ่งทุนฝึกอบรมออนไลน์
หลักสูตรดังกล่าวเป็นการอบรมเกี่ยวกับ Big Data Technologies ได้แก่ ระบบปฏิบัติการลีนุกซ์ (Linux)
ข้อมูลมหัต (Big Data) และระบบที่จัดเก็บและวิเคราะห์ข้อมูลขนาดใหญ่ (Hadoop) รายละเอียดตามสิ่งที่ส่งมาด้วย

        สำนักนโยบายและแผนการอาชีวศึกษาพิจารณาแล้วเห็นว่า ทุนฝึกอบรมดังกล่าวข้างต้น
จะเป็นประโยชน์ต่อการพัฒนาความรู้ความสามารถด้าน Big Data Technologies ของบุคลากรในสังกัด
สำนักงานคณะกรรมการการอาชีวศึกษา จึงขอประชาสัมพันธ์ให้สถาบันการอาชีวศึกษาและสถานศึกษา
ในสังกัดทุกแห่งแจ้งให้ผู้สนใจทราบต่อไป โดยผู้สนใจสามารถสมัครได้ที่เว็บไซต์ https://www.itecgoi.in/e-itec และ
ส่งสำเนาใบสมัครออนไลน์พร้อมหนังสือเสนอชื่อ มาที่ระบบงานสารบรรณอิเล็กทรอนิกส์สำนักนโยบายและ
แผนการอาชีวศึกษา ภายในวันที่ 27 ธันวาคม 2564

        จึงเรียนมาเพื่อทราบและประชาสัมพันธ์ให้ผู้สนใจทราบด้วย

ขอแสดงความนับถือ

(นายวิทวัต ปัญจมะวัต)
ผู้อำนวยการสำนักนโยบายและแผนการอาชีวศึกษา

กลุ่มยุทธศาสตร์การอาชีวศึกษาต่างประเทศและวิเทศสัมพันธ์
โทร. 0 2026 5555 ต่อ 3012
โทรสาร 0 2280 1252

e-ITEC Programme on Data Science and Big Data Analytics

1. Big Data Technologies

Introduction to Linux

The Evolution of Linux operating system, The Architecture and Structure of Linux, Installation, Introduction to Linux File system, File processing commands, Text Processing Commands, Backup and recovery commands, Network commands, Basic of I/O commands, Inter Process communication, Introduction to Users and Groups, Essentials of Effective User, Group, and Password Management, understanding permissions, Access control list and chmod command, chown and chgrp commands

Introduction to Big data and Hadoop

Introduction to big data platform, Big data challenges, Big Data Applications, Types of Big Data Technologies, Limitations and Solution of Big data Architecture, What is Hadoop, Brief History and Evolution of Hadoop, Hadoop Distributions and Vendors.

Hadoop Architecture and HDFS

Hadoop Architecture, Core components of Hadoop

HDFS (Hadoop Distributed File System): What is HDFS, Core components of HDFS, Hadoop Server Roles: Name Node, Secondary Name Node, and Data Node

HDFS Architecture overview, The HDFS command line and web interfaces, Analyzing the Data with Hadoop, Scaling Out, high availability and Name Node federation, HDFS – Monitoring & Maintenance.

Hadoop Environment

Demonstration to cloudera quickstart virtual machine

How to set up Hadoop cluster and Install on Virtual Machine, Hadoop Configuration, Security in Hadoop, Administering Hadoop, common hadoop shell commands

Big data analytics with Map Reduce Framework

Hadoop Map Reduce paradigm, Map Reduce Execution Framework, Anatomy of a Map Reduce Job, Partitioners and Combiners, Input Formats (Input Splits and Records, Text Input, Binary Input, Multiple Inputs) Output Formats (Text Output, Binary Output, Multiple Outputs)

Big data analytics with PIG

Introduction to PIG, Pig Execution Modes, Basics of PIG Latin Programming Conventions, Data Types, Arithmetic and Relational Operators, UDF Statements, PIG Latin Scripting, PIG Built-In Functions ,Eval Functions, Load/Store Functions, Math Functions, String Functions, Date Time Functions, Writing a PIG UDF, Piggy Bank, PIG Macros, Real-Time Data Analytics using PIG

Big data analytics with Hive

The Hive Data-ware House, Basics of Hive Query Language, Working with Hive QL, Operators and Functions, Importing Data, Querying Data & Managing Outputs, Hive Tables (Managed Tables and Extended Tables), Partitions and Buckets, Aggregating, Joins Views, Data manipulation with Hive, User Defined Functions, Writing HQL scripts.

Big data analytics with Spark

Initializing Spark, Spark Components and Architecture, Resilient Distributed Datasets (RDDs), RDD Operations, Passing Functions to Spark, Working with Key-Value Pairs, Shuffle operations, RDD Persistence, Shared Variables, Working with Spark with Hadoop, Spark SQL, Dataframes and Datasets

Big data analytics with MongoDB

Overview of SQL (DDL, DML, TCL), Introduction to NoSQL, Difference between SQL and NoSQL, working with MongoDB (Installation, CRUD operations, Aggregation pipeline, Indexing, Data Modeling)

2. Machine learning with python programming

Introduction to Python Programming

Installing Python, Introduction to Python Basic Syntax, Data Types, Variables, Operators, Input/output, Python data structure, Introduction to Strings, Lists, Tuples, Dictionaries, Sets. Flow

of Control (Modules, Branching) If, If- else, Nested if-else Looping, For, While, Nested loops Control Structure, Uses of Break & Continue ,Functions and methods and Exception Handling, OOPs Concepts, Python classes and objects, Introduction and Installation of Machine learning packages like PANDAS, NUMPY, SKLearn, Matplotlib, Seaborn. Mathematical Computing with Python (NumPy), Data Manipulation with Pandas, Machine Learning with Scikit–Learn.
Data Visualization in Python using matplotlib, Seaborn

Machine Learning Algorithms

What is machine learning? Types of learning, Applications of Machine learning, Evaluating ML techniques, Data Pre-processing concepts and hands on practice using python packages.

Unsupervised Algorithms: Clustering, Hierarchical Clustering & K means, Evaluation of Clusters, Clustering Case Study

Association rules mining, Apriori algorithm

Supervised Algorithms: Linear Regression, Decision Trees, Decision Trees case study, Naive bayes classifier, Assigning probabilities and calculating results, Naïve Bayes case study, K-Nearest Neighbors Algorithm and case study. Ensemble Learning: Concept of Model Ensembling, Random forest, Gradient boosting Machines, Model Stacking, Support Vector Machines, Neural Network and its applications, Single layer neural Network, Constructing Neural Networks model, Overview of Feed Forward Neural Network, Back propagation, Activation Functions: Sigmoid, Hyperbolic Tangent

Introduction to Deep Learning

Introduction to deep Learning, Why Deep Learning is taking off? Introduction to Tensorflow, Introduction to Keras, Building blocks of deep neural networks, Activation Functions, why non-linear activation functions? Computer Vision: Introduction to Convolutional Neural Network. Sequence Modeling: Recurrent Neural Network. Real world case studies for convolutional neural networks and recurrent neural network model

Project Work

Data Science and Big Data Analytics
7th Feb 2022 to 18th March 2022 (Monday - Friday)
Tentative Day wise Schedule - 13:00 HRS - 20:00 HRS (IST)

Date	Day	13:00 - 16:00	17:00 - 20:00
07-Feb-22	Monday	The Evolution of Linux operating system, The Architecture and Structure of Linux, Installation, Introduction to Linux File system, File processing commands, Text Processing Commands, Backup and recovery commands, Network commands, Basic of I/O commands	Basic of I/O commands, Inter Process communication, Introduction to Users and Groups, Essentials of Effective User, Group, and Password Management, understanding permissions, Access control list and chmod command, chown and chgrp commands
08-Feb-22	Tuesday	Introduction to big data platform, Big data challenges, Big Data Applications, Types of Big Data Technologies, Limitations and Solution of Big data Architecture	What is Hadoop, Brief History and Evolution of Hadoop, Hadoop Distributions and Vendors.
09-Feb-22	Wednesday	Hadoop Architecture, Core components of Hadoop	Hadoop ecosystem, Hadoop Yarn,Introduction to Hive,Pig,Sqoop, zookeeper, Flume, oozie,Spark, Hbase etc. Comman Hadoop Shell commands
10-Feb-22	Thursday	Distributed File System, What is HDFS, Where does HDFS fit in, Core components of HDFS, HDFS Daemons, Hadoop Server Roles: Name Node, Secondary Name Node, and Data Node HDFS Architecture, Replication, Rack Awareness	Data Pipelining, Node Failure Management.HDFS Name Node High Availability HDFS Data storage process,Anatomy of writing and reading file in HDFS, HDFS Web Interface.
11-Feb-22	Friday	Demonstration to cloudera quickstart virtual machine ,How to set up Hadoop cluster and Install on Virtual Machine	Hadoop Configuration,Security in Hadoop, Administering Hadoop, common hadoop shell commands
14-Feb-22	Monday	Hadoop Map Reduce paradigm, Map and Reduce tasks, Map Reduce Execution Framework, Anatomy of a Map Reduce Job run Partitioners and Combiners	Input Formats (Input Splits and Records, Text Input, Binary Input, Multiple Inputs), Output Formats (Text Output, Binary Output, Multiple Output).
15-Feb-22	Tuesday	Hadoop Data Types,Map Reduce program structure,Map-only program, Reduce-only program	Use of combiner and partitioner,Counters, Schedulers,Complex Map Reduce programming

Date	Day	13:00 - 16:00	17:00 - 20:00
16-Feb-22	Wednesday	Introduction to PIG:Introduction to PIG, PIG vs. Map Reduce,Pig Execution Modes,Running PIG, PIG Latin Statements	Basics of Pig Latin Programming: Conventions, Data Types, Arithmetic and Relational Operators, UDF Statements, PIG Latin Scripting
17-Feb-22	Thursday	Pig built in functions:Eval Functions, Load/Store Functions,Math Functions,String functions, Date Time Functions, Tuple, Bag, Map Functions.	UDFs (user defined functions), Control Structures, Commands: Writing a PIG UDF,Piggy Bank ,Data Fu, PIG Macros,Parameter Substitution,Shell and Utility Commands,Real-Time Data Analytics using PIG
18-Feb-22	Friday	Introduction to HIVE:Introduction to Hive,Comparison with Traditional Database, Basics of Hive Query Language. Datatypes, Operators and Functions, Hive Tables (Managed Tables and Extended Tables)	Working with Hive QL: Partitions and Buckets,Storage Formats,Importing data,Altering and Dropping Tables. Querying Data-Sorting,Aggregating,Map Reduce Scripts,Joins and Sub queries,Views
21-Feb-22	Monday	Querying Data-Sorting,Aggregating,Map Reduce Scripts,Joins and Sub queries,Views	More on Hive QL:Data manipulation with Hive,UDFs,Appending data into existing Hive table,Writing HQL scripts
22-Feb-22	Tuesday	Apache Spark: Overview, Linking with Spark, Initializing Spark, Resilient Distributed Datasets (RDDs), External Datasets, RDD Operations	Passing Functions to Spark, Working with Key-Value Pairs, Shuffle operations, RDD Persistence, Removing Data, Shared Variables
23-Feb-22	Wednesday	Spark SQL: Spark SQL	Working with Spark with Hadoop, DataFrames and Datasets
24-Feb-22	Thursday	Introduction to NOSQL: Overview of SQL (DDL, DML, TCL)	Introduction to NoSQL, Difference between SQL and NoSQL
25-Feb-22	Friday	Working with MongoDB: Installation, CRUD operations	Aggregation pipeline, Indexing, Data Modeling
28-Feb-22	Monday	Installing Python, Introduction to Python Basic Syntax, Data Types, Variables, Operators, Input/output, Python data structure, Introduction to Strings, Lists, Tuples, Dictionaries, Sets.	If, If- else, Nested if-else Looping, For, While, Nested loops Control Structure, Uses of Break & Continue
01-Mar-22	Tuesday	Functions and methods and Exception Handling, OOPs Concepts, Python classes and objects	Introduction to Numpy
02-Mar-22	Wednesday	Working with data- Pandas	Data Visualization using matplotlib

Date	Day	13:00 - 16:00	17:00 - 20:00
03-Mar-22	Thursday	What is machine learning? Types of learning, Applications of Machine learning, Evaluating ML techniques	Data preprocessing using Python
04-Mar-22	Friday	Unsupervised Algorithms: Clustering, Hierarchical Clustering & K means, Evaluation of Clusters, Clustering Case Study	Association rules mining, Apriori algorithm
07-Mar-22	Monday	Supervised Algorithms: Decision Trees, Decision Trees case study	Naive bayes classifier, Assigning probabilities and calculating results, Naïve Bayes case study
08-Mar-22	Tuesday	Linear Regression, Simple Linear Regression, Multiple linear regression	K Nearest Algorithm
09-Mar-22	Wednesday	Ensemble Learning: Concept of Model Ensembling, Random forest, Gradient boosting Machines, Model Stacking, Support Vector Machines	Neural Network and its applications, Single layer neural Network, Constructing Neural Networks model, Overview of Feed Forward Neural Network, Back propagation, Activation Functions: Sigmoid, Hyperbolic Tangent
10-Mar-22	Thursday	Introduction to deep Learning, Why Deep Learning is taking off? Introduction to Tensorflow, Introduction to Keras, Sequence Modeling: Recurrent Neural Network.	Building blocks of deep neural networks, Activation Functions
11-Mar-22	Friday	why non-linear activation functions? Computer Vision: Introduction to Convolutional Neural Network.	case study CNN
14-Mar-22	Monday	Sequence Modeling: Recurrent Neural Network.	Case Study RNN
15-Mar-22	Tuesday	Real world case studies for convolutional neural networks and recurrent neural network model	Real world case studies for convolutional neural networks and recurrent neural network model
16-Mar-22	Wednesday	Real world case studies for convolutional neural networks and recurrent neural network model	Real world case studies for convolutional neural networks and recurrent neural network model
17-Mar-22	Thursday	Project Work	Project Work
18-Mar-22	Friday	Project Work	Project Work