

JIS Institute of Advanced Studies and Research Kolkata
3 Year M. Tech in Data Science (Executive)
Session 2020-2021
Every Saturday (for 6 hours), no weekdays

1st Semester

Sl. No.		Subject Code	Subject Name	Class Load/Week			Credit
				L	T	P	
1	Core Courses	PGDS101	Mathematical Foundation for Data Science	3	1	0	4
2		PGDS102	Database Management System	3	1	0	4
3	Labs	PGDS191	Programming Lab (Python & R)	0	0	6	4
4		PGDS192	DBMS Lab	0	0	3	2
Total Credits							14

2nd Semester

Sl. No.		Subject Code	Subject Name	Class Load/Week			Credit
				L	T	P	
1	Core Courses	PGDS201	Big Data Analytics	3	0	0	3
2		PGDS202	Machine Learning	3	1	0	4
3	Labs	PGDS291	Big Data Analytics Lab	0	0	3	2
4		PGDS292	Machine Learning Lab	0	0	3	2
Term Project							3
Total Credits							14

3rd Semester

Sl. No.		Subject Code	Subject Name	Class Load/Week			Credit
				L	T	P	
1	Core Courses	PGDS301	Advanced Machine Learning with Python	2	0	3	4
2		PGDS302	Cloud Computing	2	0	3	4
3		PGDS303	Elective 1	3	0	0	3
4		PGDS304	Elective 2	3	0	0	3
Total Credits							14

4th Semester

Sl. No.		Subject Code	Subject Name	Class Load/Week			Credit
				L	T	P	
1	Core Courses	PGDS401	Experimental IoT	2	0	3	4
2		PGDS402	Cybersecurity	3	1	0	4
3		PGDS403	Elective 3	3	0	0	3
Term Project							3
Total Credits							14

5th Semester

Sl. No.	Subject Code	Subject Name	Class Load/Week			Credit	
			L	T	P		
1.	PGDS501	Thesis (Interim) with Seminar	-	-	-	12	
2.	PGDS502	Comprehensive Viva-voce				4	
Total Credits							16

6th Semester

Sl. No.	Subject Code	Subject Name	Class Load/Week			Credit	
			L	T	P		
1.	PGDS601	Project Dissertation with Seminar and Viva-Voce	-	-	-	16	
Total Credits							16

List of Electives	
Optimization Techniques	
Financial and Business Data Analytics	
Introduction to Speech, Image and Video Processing	
Natural Language Processing	
Social Media Analytics	
Data Warehousing and Data Mining	

Total Credits = [14+14+14+14+16+16] = 88

JIS Institute of Advanced Studies & Research Kolkata
JIS University
Centre for Data Science
3 Year M.Tech in Data Science (Executive)
Detailed Course Structure || Semester 1

PGDS101 – Mathematical Foundation for Data Science

Regression Models: Multiple Regression Analysis, Logistic Regression, Lasso Regression, Ridge Regression, Elastic Net Regression.

Linear Algebra: Basic properties of matrix and vectors: scalar multiplication, linear transformation, transpose, conjugate, rank, determinant, Inner and outer products, matrix multiplication rule and various algorithms, matrix inverse, Special matrices: square matrix, identity matrix, triangular matrix, idea about sparse and dense matrix, unit vectors, symmetric matrix, Hermitian, skew-Hermitian and unitary matrices, Matrix factorization concept/LU decomposition, Gaussian/Gauss-Jordan elimination, solving $Ax=b$ linear system of equation, Vector space, basis, span, orthogonality, orthonormality, linear least square, Eigenvalues, eigenvectors, diagonalization, singular value decomposition.

Optimization and Operational Research: Basics of optimization, how to formulate the problem, Maxima, minima, convex function, global solution, Linear programming, simplex algorithm, Integer programming, Constraint programming, knapsack problem, Randomized optimization techniques: hill climbing, simulated annealing, genetic algorithms.

Forecasting Techniques: Introduction to Time-Series Data, Components of Time-Series, Moving Average Method, Single and Double Exponential Smoothing, Holt's method, Holt-Winter Model, Auto Regressive Model, ARMA, ARIMA, Dickey Fuller Test, Ljung-box Test, Power of Forecasting Model, Theils's Coefficient

PGDS102 – Database Management System

Introduction to database management, data abstraction and system structure.

Entity relational model, entity set, relationship sets, mapping cardinalities, keys, E-R diagrams.

Relational model, database schema, relational algebra, outer join and manipulation of databases.

Tuple relational calculus: Example queries, formal definitions; SQL: Query processing and optimization, set operations, aggregate functions, data definition language and views, comparison of queries in relational algebra, SQL, tuple relation calculus.

Relational database design, various normal forms, functional dependencies, canonical cover, lossless join, dependency preservation, multi value dependency and higher normal forms; Object relational databases (ORDBMS); Object oriented models and relations – generalization, specialization, aggregation; Functions and polymorphism; Object oriented SQL; Commercial object relational database systems – MySQL, Oracle, IBMDB2.

Concurrency control and recovery: Transactions management and ACID property.

Serializability and testing for serializability, concurrency control schemes, lock-based protocols, two-phase locking protocols, graph-based protocols, time stamp-based protocols, deadlocks.

Recovery systems, log-based recovery, deferred and immediate database modification, object oriented database design.

Non-structured and semi-structured data storage: XML databases, Image and video databases, NoSQL databases – Cassandra, Hadoop and map reduced techniques.

Data warehousing, heterogeneous component systems, data scrubbing.

Data mining and knowledge discovery, basic mathematical, numerical and statistical techniques; Applications in information retrieval.

PGDS191 – Programming Lab (Python and R)

Python:

Introduction, Operators, Variables and Data Types, Conditional Statements, Looping Construct, Functions, Lists, Dictionaries, Libraries, DB – connectivity,

NumPy: NumPy Standard data types, The Basics of NumPy Arrays, Array Attributes, Array indexing, Array Slicing: accessing sub array, Reshaping Of Arrays, Concatenation & Splitting of Array, Computation on NumPy arrays, Universal Functions (optional), Aggregations (Summing the values, min, max), Array Broadcasting (optional), Comparisons, masks, Boolean logic, NumPy Sorting (ex. np.sort, np.argsort), Structured Arrays, Creation of structured arrays, Record Arrays.

Pandas: Introducing Pandas Objects, Pandas Series Object, Pandas Data Frame Object, Pandas Index Object, Data operations in Panda, Data Selection in Series, Data Selections in DataFrame, Handling Missing Data, Combining Dataset (pd.concat,pd.appended), merge and join, Aggregating and Grouping (GroupBy: Split, Apply, Combine), Pandas String, Pandas Time Series.

Advance DataFrame Methods - DataFrame.eval(), DataFrame.query()

Matplotlib: Display Plots, Histograms, Binnings, and Density, 3D plotting

Scikit Learn: Introduction to sklearn.

R:

R-Basics: Basic Data types in R, R – Operators, Variable assignments, Arithmetic with R.

Vectors: Creation, naming, Selection of Vector, Calculating, comparing Total winnings.

Matrices: Analyze, Naming of Matrix, Selection of Matrix element, Addition of Row, Col Matrix arithmetic.

Factors: Factor, Factor levels, Summarizing factors, Ordered factors, Comparing Ordered factors.

Data Frames: Creating data frame, Selection of data frame elements, Sorting of the data frame elements.

Lists: Creating a list, Creating named list, Selecting elements from a list, Adding information to the list.

Hands-on Development:

Basic Data Structure in Python(3.7) and its applications using sequential conditional and iterative flow of control (String,List,Dictionaries,Tuples,Set)

Indexing & Slicing with shorthand's notation.

Formatted and f'-string formatting in console output

Design and implementation of function in Python (3.7) that takes variable arguments

Design and implementation of method overloading using weakly typed (implicit) design.

Implementation of Iterator in Python(3.7)

Design and Implementation of customize Generator using python(3.7)

Implementation of Map Filter with console input.

Creation of Packages and Sub Packages using Python (3.7)

Implementation of Lambda expression using Python (3.7)

Lambda expression and its association with Map, Filter

Regular Expression in Python (3.7)

IO and Exception Handling using Python (3.7)

Design and implementation of Multithreading application using Python (3.7)

Implementation of Numpy arrays and its functionalities.

Read write update .csv,.txt,.html,.xml using Pandas

Implementation of Pandas –Series, DataFrame and uses of their inbuilt methods

Data visualization using matplotlib libraries

Basic design and architecture of OOP in Python (3.7)

Design and implementation of Decorator Pattern using Python (3.7)

Data Exploration,Feature Engineering ,Predictive modeling using linear regression algorithm (using Scikit-Learn,Numpy,Pandas,Matplot Library) on available EcommerceCustomer dataset of Kaggle repository

Data Exploration,Feature Engineering ,Predictive modeling using KNN algorithm (using Scikit-Learn, Pandas , Numpy, Matplot Library) on available EcommerceCustomer dataset of Kaggle repository

Data Exploration,Feature Engineering ,Predictive modeling using Logistic Regression algorithm (using Scikit-Learn, Pandas , Numpy, Matplot Library) on available Titanic dataset of Kaggle repository

Introduction and basic building block of R

Sequential conditional and iterative flow of control using R

Design and implementation Bayesian algorithm.

Design and Implementation of PDF PMF Estimator in R

Design and implementation of PCA with R.

Data exploration and feature engineering with R using OpenCV

Implementation of SVM for predictive modeling using R.

PGDS192 – DBMS Lab (Python and R)

Hands on Database Schema, Keys, Schema Diagram, Relational Query Languages, SQL, SQL Data Definition, SQL Queries, Set Operations, Null Values, Aggregate Functions, Nested Subqueries, Modification of the Database, Join Expressions, Views, Transactions, Integrity Constraints, Triggers, Relational Algebra, Relational Calculus.

Hands-on Entity-Relationship Model, Constraints, Reduction to Relational Schemas, E-R Design Issues, Extended E-R Features, Functional Dependency, Normalization using functional dependencies, Decomposition, 1NF, 2NF, Boyce-Codd Normal Form, 3NF, Normalization using Multi-valued Dependencies, 4NF, 5NF.

DataBase Connection through Java/Python/PHP.

JIS Institute of Advanced Studies & Research Kolkata
JIS University
Centre for Data Science
3 Year M.Tech in Data Science (Executive)
Detailed Course Structure || Semester 2

PGDS201 – Big Data Analytics

Introduction to Big Data: OLTP System, RDBMS for structured data, LOB, Polyglot Persistence – ACID, ETL, MPP, NAS BOX – SAN.

Apache Hadoop: Hadoop overview and History, Overview of Hadoop, Ecosystem, Hadoop's core – HDFS and MapReduce.

Apache Spark: Spark First Principles, DataFrames, Spark Types and Datasets, Spark SQL, RDDs.

Relational data stores with Hadoop: Hive, MySQL, Sqoop.

Non – Relational data stores with Hadoop: Why NoSQL, Hbase, Cassandra, Mongo DB.

Querying Data Interactively: Drill, Phoenix, Presto.

Cluster Management: YARN, (Tez, Mesos, ZooKeeper, Oozie, Zeppelin, Hue), Explanation.

Feeding Data to Cluster: Apache Kafka- Setting up and publishing web logs, Apache Flume – Setting up and publishing logs.

Analyzing Data Streams: Apache Spark Streaming.

PGDS202 – Machine Learning

Introduction to Pattern Recognition: Different Paradigms of Pattern Recognition, Representations of Patterns and Classes, Metric and non-metric proximity measures, Feature Extraction, Different approaches to Feature Selection. Component Analysis and Dimension Reduction - The Curse of Dimensionality, Principal Component Analysis.

Tree Classifiers: (a) Decision Trees: CART, C4.5, ID3. (b) Random Forests.

Supervised Methods: Bayesian Decision Theory, Discriminative Classifiers: Decision Boundary, Discriminant analysis, Support Vector Machines and Introduction to Neural Network.

Unsupervised Methods: (a) Fisher Linear Discriminant and Locally Linear Embedding, (b) Clustering - K-Means, Expectation Maximization and Mean Shift.

Parametric Techniques: (a) Maximum Likelihood Estimation (b) Bayesian Parameter Estimation (c) Sufficient Statistics.

Non-Parametric Techniques: (a) Kernel Density Estimators (b) Nearest Neighbour Methods.

Classifier Ensembles: (a) Bagging (b) Boosting / AdaBoost

Graphical Models: (a) Introductory ideas and relation back to earlier topics (b) Bayesian Networks (c) Sequential Models - State-Space Models, Hidden Markov Models and Dynamic Bayesian Networks.

Introduction to Machine Intelligence, Problem Space Representation, Heuristic Search Techniques, Knowledge Representation, Predicate Logic, Reasoning under uncertainty, Statistical Reasoning, Planning, Learning, Expert System Design, Expert System Shell, Case Studies.

PGDS291 – Big Data Analytics Lab

Hadoop Configuration and architectural framework of Apache Spark setup with virtualization.

Design and deployment of map with key-value tuple to multiple cluster and reduce for multi row dependent functional operations.

Statistical analysis of numeric with pivot table using excel sheet.

Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations (Design/Development of Solutions).

Experiment and implementation with PIG LATIN LANGUAGE – PIG

Design implementation of HIVE OPERATIONS

Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.

Implementation of Spark Streaming; Optimization with Partitioners; Deployment of Wide and Narrow Dependencies; Capstone Project.

PGDS292 – Machine Learning Lab

Implementation of Polynomial Curve Fitting

Implementation Bayesian curve fitting

The Curse of Dimensionality using Python

Implementation of Nearest-neighbour methods

Bayesian Linear Regression implementation

Implementation of the perceptron algorithm

Gradient descent optimization

Implementation of Diagonal approximation

Implementation of Bayesian Neural Network

Implementation of Radial Basis Function in ANN.

Implementation of Logistic Regression.

SVMs for regression.

JIS Institute of Advanced Studies & Research Kolkata
JIS University
Centre for Data Science
3 Year M.Tech in Data Science (Executive)
Detailed Course Structure || Semester 3

PGDS301 – Advanced Machine Learning with Python

Preliminaries of Artificial Neural networks; Deep Networks: Deep Feed Forward Network, Regularization for Deep Learning, Optimization for Training Deep Models.

Convolutional Networks (CNN), Recurrent (RNN) and Recursive Net; Autoencoders; Representation Learning; Monte Carlo Methods; Deep generative Models.

Reinforcement Learning in the Context of Deep Learning: Multi-armed Bandits, Finite Markov Decision Processes, Q-Learning, Deep Q-Learning, Policy Optimization.

Introduction to Parallel and Distributed Deep Learning; Application of Deep Learning Techniques in Computer Vision, Natural Language Processing.

Lab Part:

Hands-on Deep Networks: Deep Feed Forward Network, Regularization for Deep Learning, Optimization for Training Deep Models, Convolutional Networks (CNN), Recurrent (RNN) and Recursive Net; Autoencoders; Representation Learning; Deep generative Models; Reinforcement Learning in the Context of Deep Learning: Multi-armed Bandits, Q-Learning, Deep Q-Learning, Policy Optimization; Application of Deep Learning Techniques in Computer Vision, Natural Language Processing.

PGDS302 – Cloud Computing

Concept of VM: what is virtual machine, What is a hyper visor, Common hyper visor like virtual box & VMware, Elastic nature of machine resources, How to create your own VM image in virtual box, How to manage multiple virtual machine in one machine, Kubernetes, CLI & GUI.

Docker container basics: Basics of docker containers, Difference between docker container & VM, Command to start & stop a docker container, Connecting to the external network, Sharing with the whole file system, Creating your own docker container.

Proprietary cloud vendor examples: AWS free tier, Google free tier, Azure free tier, Some common service using the above 3, Some common pipelines in the above 3.

IOT Cloud interaction: Web services, data ingestion and analytics in the cloud, Common readymade cloud analytics , system, Write your own cloud data ingestion service using MQTT, Write your own cloud data ingestion service using HTTP, Basic analysis and visualization after data ingestion, Reverse communication from cloud to IOT device using notification, Project on IOT to cloud communication and vice-versa.

JIS Institute of Advanced Studies & Research Kolkata
JIS University
Centre for Data Science
3 Year M.Tech in Data Science (Executive)
Detailed Course Structure || Semester 4

PGDS401 – Experimental IoT

Sensors and Actuator: Analog Input, Digital Input, Camera Input, Touch Screen Input, Network Input, Bluetooth Input, SPI Input, I2C Input, RS 232 Input, Audio Port Input, Analog Output, Digital Output, Bluetooth Output, SPI Output, I2C Output, RS232 Output.

Sensors and Actuator Internals: Temperature, Pressure and Relative Humidity Sensor, Soil Humidity Sensor, Gass Sensor, PIR Sensors, Proximity Sensors and Limit Switch, V,I,COS PHI, Optocoupler.

RPI Hardware Details: PIN OUT Diagram, GPIO Intrnal Circuitry, Alternate function pin circuitry, Detailed Hardware Specs of RPI

Arduino UNO Hardware Details: ATmega 328P, Digital I/O Pins, PWM Digital I/O Pins, Analog Input Pin, DC, Current Input Pin, SRAM, EEPROM.

Node MCU firmware with with ESP8266 Development Kit: ESP 8266 wifi Module with TCP/IP Protocol, Auduino IDE for ESP8266. Interfacing with ESP8266.

Raspberry Pi Setup and Administration: OS LOADING, Post boot configuration, SSH Configuration, Serial Console on Rpi, Wiring Pi, I2c, SPI setup, DHCP server and DHCP client configuration, Wi- Fi Configuration, IP Configuration, Port Forwarding On rpi, Maximum Driving Capacity Of Individual and All Pins, Optocoupler based relay driving circuitry for different voltage relays, Driving High Current using Rpi, Rpi as router, (already covered) Camera Connection and troubleshooting 8MP.

Rpi Interfacing and Programming: GPIO IN/OUT, GPIO PULL HIGH/LOW, HTTP SERVER in python, Cronjob, MQTT Publisher Program, MQTT Broker, MQTT Subscriber, HTTP server, HTTP Client, RPi ADC Programming, Actuator On/Off and Sensor Reading Monitoring in HTTP server (already covered), Analog output control with Rpi, GSM Modem, SPI interfacing, I2C interfacing, Touch screen Interfacing, Bluetooth Communication RPI, Raspberry Pi Interupts, Rpi to Rpi communication BT/Ethernet/Wi-Fi/RS232, Camera handling using Rpi, FR using Rpi.

IoT cloud Interfacing, communication and Simulations: Blynk (IoT platform with free Cloud, Android mobile apps) Rapberry Pi – Azure IoT Simulations (using ECS-6), IoT solutions with Microsoft Azure, Arduino-MQTT - Microsoft Azure, DHT11 sensor data to IoT Hub using Node MCU.

Lab Part:

GPIO IN/OUT. GPIO PULL HIGH/LOW. HTTP SERVER in python. Cronjob. MQTT Publisher Program. MQTT Broker. MQTT Subscriber. Netcat Client. Netcat Server. HTTP server. HTTP Client. RPi ADC Programming. RPi ADC Programming with own designed ADC board via serial port (16F877A IC). RPi sensor Data Aggregating and packet formation. Actuator On/Off and Sensor Reading Monitoring in HTTP

server. Camera handling. Face recognition (Rpi 4) with OpenCV. Object detection & recognition. Speech recognition. Analog output control with Rpi. Platform movement by stepper motor using RPi. RS232 interfacing & communication. GSM Modem. SPI interfacing. I2C interfacing. Remote Camera fid. GPS Module Interfacing. Touch screen Interfacing. Basics of UI of KIOSK. Rpi Controlled Drone for Agriculture. Rpi controlled robotics. Bluetooth Communication RPI. Ble Beacon Tracking with RPI. Audio Interfacing with RPI with Sound Port. Audio Interfacing with Rpi Zero w without sound port. Raspberry Pi Interrupts. Socket Communication. Rpi to Rpi communication BT/Ethernet/Wi-Fi/RS232. Creating VM image in virtual box and to manage multiple virtual machines in one machine. Design and implementation of Kubernetes, CLI & GUI. Creating and analyzing own Docker container. Web services, data ingestion and analytics in the cloud. Implementation of Common readymade cloud analytics system. Implementation of cloud data ingestion service using MQTT. Implementation and customization of cloud data ingestion service using HTTP, basic analysis and visualization after data ingestion. Reverse communication from cloud to IOT device using notification. Capstone Project on IoT to cloud communication and vice-versa.

PGDS402 – Cybersecurity

Introduction to Cyber Security: Overview of Cyber Security, Internet Governance – Challenges and Constraints, Cyber Threats: Cyber Warfare-Cyber Crime-Cyber terrorism-Cyber Espionage, Need for a Comprehensive Cyber Security Policy, Need for a Nodal Authority, Need for an International convention on Cyberspace.

Cyber Security Vulnerabilities and Cyber Security Safeguards: Cyber Security Vulnerabilities-Overview, vulnerabilities in software, System administration, Complex Network Architectures, Open Access to Organizational Data, Weak Authentication, Unprotected Broadband communications, Poor Cyber Security Awareness. Cyber Security Safeguards- Overview, Access control, Audit, Authentication, Biometrics, Cryptography, Deception, Denial of Service Filters, Ethical Hacking, Firewalls, Intrusion Detection Systems, Response, Scanning, Security policy, Threat Management.

Securing Web Application, Services and Servers: Introduction, Basic security for HTTP Applications and Services, Basic Security for SOAP Services, Identity Management and Web Services, Authorization Patterns, Security Considerations, Challenges.

Intrusion Detection and Prevention: Intrusion, Physical Theft, Abuse of Privileges, Unauthorized Access by Outsider, Malware infection, Intrusion detection and Prevention Techniques, Anti-Malware software, Network based Intrusion detection Systems, Network based Intrusion Prevention Systems, Host based Intrusion prevention Systems, Security Information Management, Network Session Analysis, System Integrity Validation.

Cryptography and Network Security: Introduction to Cryptography, Symmetric key Cryptography, Asymmetric key Cryptography, Message Authentication, Digital Signatures, Applications of Cryptography. Overview of Firewalls- Types of Firewalls, User Management, VPN Security Security Protocols: - security at the Application Layer- PGP and S/MIME, Security at Transport Layer- SSL and TLS, Security at Network Layer-IPSec.

Cyberspace and the Law: Introduction, Cyber Security Regulations, Roles of International Law, the state and Private Sector in Cyberspace, Cyber Security Standards. The INDIAN Cyberspace, National Cyber Security Policy 2013.

Cyber Forensics: Introduction to Cyber Forensics, Handling Preliminary Investigations, Controlling an Investigation, Conducting disk-based analysis, Investigating Information-hiding, Scrutinizing E-mail, Validating E-mail header information, Tracing Internet access, Tracing memory in real-time.