# M.Sc. Data Science Part – II Syllabus
# Credit Based System and Grading System
# Academic Year 2024-25

## M.Sc. Part – II Semester – III

| Course Code | Course Type | Course Title | Credits |
|---|---|---|---|
| SIPDSCC611 | Core Subject (Major) | Machine Learning | 4 |
| SIPDSCC612 | Core Subject (Major) | Big Data Analytics | 4 |
| SIPDSEL611 | Core Subject (DSC) | Database Administration and Security | 3 |
| SIPDSCCP611 | Core Subject Practical (Major) | Machine Learning – Practical | 2 |
| SIPDSCCP612 | Core Subject Practical (Major) | Big Data Analytics – Practical | 2 |
| SIPDSELP611 | Core Subject Practical (DSC) | Database Administration and Security – Practical | 1 |
| SIPDSRP611 | Research Project | Project | 6 |
| | | **Total Credits** | **22** |

# SIPDSCC611 : Machine Learning

| M.Sc. (Data Science) | Semester – III |
|---|---|
| **Course Name: Machine Learning** | **Course Code: SIPDSCC611** |
| **Credits** | **4** |

## Course Objective:

- To introduce students to the fundamental concepts and principles of Machine Learning.
- To familiarize students with various types of machine learning algorithms and their applications. To provide hands-on experience in implementing and evaluating machine learning models.
- To develop critical thinking and problem-solving skills in the context of machine learning.
- To enable students to apply machine learning techniques to real-world problems and datasets.
- To understand the limitations and challenges of machine learning and develop strategies to address them.

## Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Understand the foundational concepts and principles of Machine Learning.
- **CO2:** Apply supervised and unsupervised learning techniques, including classification algorithms and clustering algorithms.
- **CO3:** Evaluate the performance of Machine Learning models using classification metrics, ROC/AUC curve analysis, and cross-validation techniques.
- **CO4:** Implement regression models (such as linear regression and logistic regression) and understand their applications in predictive analysis.
- **CO5:** Utilize dimensionality reduction techniques (Like PCA) for feature reduction and selection, and apply association rule mining algorithms (such as the Apriori algorithm) for discovering meaningful patterns in datasets.

| Unit | Contents | No. of Lectures |
|---|---|---|
| **I** | **Introduction to Machine Learning:** Machine Learning(ML), Need for Machine Learning, ML from Knowledge-driven to Data Driven, Applications of Machine Learning, Problems suitable for Machine Learning, Advantages, Disadvantages and Challenges of Machine Learning, Challenges of ML. General architecture of ML systems, **Underlying Concepts in Machine Learning:** Inductive Learning, Generlization, Bias and Variance, Overfitting and Underfitting, Parametic and Non Parametric algorithms. **Types of Machine Learning:** Supervised and Unsupervised Learning, Workflow, Semisupervised Learning, Reinforced Learning | 15 |
| **II** | **Classification and Regression: Classification:** Binary Classification- Assessing Classification performance, Class probability Estimation Assessing class probability Estimates, Multiclass Classification. **Regression:** Assessing performance of Regression- Error measures, | 15 |

| | | |
|---|---|---|
| | Overfitting- Catalysts for Overfitting, Case study of Polynomial Regression. **Theory of Generalization:** Effective number of hypothesis, Bounding the Growth function, VC Dimensions, Regularization theory. | |
| **III** | **Linear Models:** Least Squares method, Multivariate Linear Regression, Regularized Regression, Using Least Square regression for Classification. Perceptron, Support Vector Machines, Soft Margin SVM, Obtaining probabilities from Linear classifiers, Kernel methods for non-Linearity. **Logic Based and Algebraic Model: Distance Based Models:** Neighbours and Examples, Nearest Neighbours Classification, Distance based clustering-K means Algorithm, Hierarchical clustering | 15 |
| **IV** | **Rule Based Models:** Rule learning for subgroup discovery, Association rule mining. **Tree Based Models**: Decision Trees, Ranking and Probability estimation Trees, Regression trees, Clustering Trees. Probabilistic Model: Normal Distribution and Its Geometric Interpretations, Naïve Bayes Classifier, Discriminative learning with Maximum likelihood, Probabilistic **Models with Hidden variables:** Estimation-Maximization Methods, Gaussian Mixtures, and Compression based Models. | 15 |

**Books and References**

| Sr. No. | Title | Author/s | Publisher | Edition | Year |
|---|---|---|---|---|---|
| 1 | Machine Learning: Concepts, Techniques and Applications | T V Geetha S, Sendhilkumar | CRC Press, Taylor and Francis | First Edition | 2023 |
| 2 | Machine Learning: The Art and Science of Algorithms that Make Sense of Data | Peter Flach | Cambridge University Press | First Edition | 2012 |
| 3 | Introduction to Statistical Machine Learning with Applications in R | Hastie, Tibshirani, Friedman | Springer | Second Edition | 2012 |
| 4 | Introduction to Machine Learning | Ethem Alpaydin | PHI | Second Edition | 2012 |

# SIPDSCCP611 : Machine Learning Practical

| M.Sc. (Data Science) | Semester – III |
|---|---|
| **Course Name: Machine Learning Practical** | **Course Code: SIPDSCCP611** |
| **Credits** | **2** |

**List of Practical: (Make use of python)**

| | |
|---|---|
| **1.** | Implementing a K-Nearest Neighbor (KNN) algorithm (e.g. to classify handwritten digits). |
| **2.** | Building a decision tree model using the ID3 algorithm (e.g. to predict whether a customer will churn or not). |
| **3.** | Developing a Support Vector Machine (SVM) model (e.g. to classify email messages as spam or not spam) |
| **4.** | Building a Naïve Bayes classifier (e.g. to classify movie reviews as positive or negative sentiment). |
| **5.** | Implementing linear regression (e.g. to predict housing prices based on features such as size and location). |
| **6.** | Using logistic regression (e.g. to predict whether a credit card transaction is fraudulent or not). |
| **7.** | Evaluating a classification model using metrics such as accuracy, precision, recall, and F1 score. |
| **8.** | Applying hierarchical clustering (e.g. to group customer segments based on their purchasing behavior). |
| **9.** | Implementing the K-means clustering algorithm (e.g. to identify distinct clusters in a customer demographic dataset). |
| **10.** | Utilizing Principal Component Analysis (PCA) for dimensionality reduction to improve the efficiency and interpretability of a model. |

# SIPDSCC612 : Big Data Analytics

| M.Sc. (Data Science) | Semester – III |
|---|---|
| **Course Name: Big Data Analytics** | **Course Code: SIPDSCC612** |
| **Credits** | **4** |

## Course Objective:

- Understand the Big Data Platform and its Use cases.
- Provide an overview of Apache Hadoop.
- Provide HDFS Concepts and Interfacing with HDFS.
- Understand Map Reduce Jobs Provide hands on Hadoop Eco System.
- Apply analytics on Structured, Unstructured Data.

## Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Identify Big Data and its Business Implications.
- **CO2:** List the components of Hadoop and Hadoop Eco-System.
- **CO3:** Access and Process Data on Distributed File System.
- **CO4:** Manage Job Execution in Hadoop Environment.
- **CO5:** Develop Big Data Solutions using Hadoop Eco System.
- **CO6:** Analyze Infosphere BigInsights Big Data Recommendations.

| Unit | Contents | No. of Lectures |
|---|---|---|
| **I** | **Introduction:** Introduction to Big Data, Big Data Characteristics, Types of Big Data, Traditional Versus Big Data Approach, Technologies Available for Big Data, Infrastructure for Big Data, Use of Data Analytics, Big Data Challenges, Desired Properties of a Big Data System, Case Study of Big Data Solutions.<br>**Analytical Theory and Methods:** Clustering and Associated Algorithms, Association Rules, Apriori Algorithm, Candidate Rules, Applications of Association Rules, Validation and Testing, Diagnostics, Regression, Linear Regression, Logistic Regression, Additional Regression Models. | 15 |
| **II** | **Hadoop:** Introduction, What is Hadoop?, Core Hadoop Components, Operating System for Big Data, Concepts, Hadoop Architecture, Hadoop Ecosystem, Hive, Hadoop Limitations, Recommendation Systems. | 15 |
| **III** | **Interactive Data Analysis with Spark Shell:** REPL Commands, Using the Spark Shell as a Scala Shell, Number Analysis, Log Analysis.<br>**Writing a Spark Application:** Hello World in Spark, Compiling and Running the Application, Monitoring the Application, Debugging the Application.<br>**Introducing Spark Streaming:** Spark Streaming Is a Spark Add-on, High-Level Architecture, Data Stream Sources, Receiver, Destinations, Application Programming Interface (API), Streaming Context, Basic Structure of a Spark Streaming Application, Discretized Stream (DStream), | 15 |

| | | |
|---|---|---|
| | Creating a Dstream, Processing a Data Stream, Output Operations, Window Operation. | |
| **IV** | **Introduction:** Kafka origin , Benefits, Use Cases, Messaging System. Fundaments of Kafka: Topics and Partitions, Producers and Consumers, Brokers and Clusters.<br>**Kafka CLI :** Creating Kafka Topics, Sending data to Kafka , Kafka Console Consumer, Kafka Consumer Group , Kafka Brokers, Topic Replication, and Controller<br>**Apache Kafka With Zookeeper:** Start ZooKeeper, Single Node-Single Broker Configuration, Start Producer to Send Messages, Start Consumer to Receive Messages, Single Node-Multiple Brokers Configuration<br>**Creating a Topic:** Start Producer to Send Messages, Start Consumer to Receive Messages<br>**Kafka Real Time Example:** Creating Twitter Producer, Application Kafka monitoring, Kafka Connect | 15 |

**Books and References**

| Sr. No. | Title | Author/s | Publisher | Edition | Year |
|---|---|---|---|---|---|
| 1 | Spark : The Definitive Guide | Bill Chambers & Mataei Zaharia | O'Reilly Media, Inc. | 1st | 2018 |
| 2 | Big Data Analytics with Spark : A Practitioner's Guide to Using Spark for Large Scale Data Analysis | Mohammed Guller | Apress | 1nd | 2015 |
| 3 | Practical Big Data Analytics Hands-on Techniques to Implement Enterprise Analytics and Machine Learning Using Hadoop, Spark, NoSQL and R | Nataraj Dasgupta | Packt | Illustrated | 2018 |
| 4 | Big Data Analytics | Radha Shankarmani | Wiley | Second | 2016 |
| 5 | Kafka : The Definitive Guide | Neha Narkhede, Gwen Shapira , Todd Palino | O'Reilly | 1st | 2015 |

# SIPDSCCP612 : Big Data Analytics Practical

| M.Sc. (Data Science) | Semester – III |
|---|---|
| **Course Name: Big Data Analytics Practical** | **Course Code: SIPDSCCP612** |
| **Credits** | **2** |

**List of Practical:**

| | |
|---|---|
| **1.** | Install, configure and run Hadoop and HDFS |
| **2.** | Implement a program in Pig. |
| **3.** | Implement word count/ frequency program using MapReduce. |
| **4.** | Configure the Hive and implement the application in Hive. |
| **5.** | Implement Spark SQL. |
| **6.** | Implement machine learning with Spark or Hadoop. |
| **7.** | Implement Spark Streaming. |
| **8.** | Demonstrate Spark Shell commands. |
| **9.** | Implement Decision tree classification technique |
| **10.** | Implement an application that store big data in Hbase/ Mongodb/ Pig using Hadoop/R |

# SIPDSEL611 : Database Administration and Security

| M.Sc. (Data Science) | Semester – III |
|---|---|
| **Course Name: Database Administration and Security** | **Course Code: SIPDSEL611** |
| **Credits** | **3** |

## Course Objective:

- Installing Oracle Software and creation of an Oracle Database using DBCA.
- Managing Database instances and ASM instances.
- Managing and controlling database network environment.
- Define and devise transaction management, concurrency control, crash recovery components and managing storage structures.
- Controlling user security and designing Database backup and recovery procedures.
- Take Decisions related to Database Maintenance.

## Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Do Database installation and creation using DBCA and design Oracle Database Architecture.
- **CO2:** Manage Database instances and ASM instances.
- **CO3:** Setup Oracle Network environment.
- **CO4:** Manage database storage structures.
- **CO5:** Configure User security, Concurrency control, Database maintenance, Crash Recovery.

| Unit | Contents | No. of Lectures |
|---|---|---|
| **I** | **Oracle Overview and Architecture:** An overview of logical and physical storage structures, Oracle memory structures, Oracle background processes, connecting to oracle instance, processing SQL command. **Managing Oracle:** starting up the oracle instance, managing sessions, shutting down the oracle instance, instances messages, and instance alerts. **Control and Redo Log Files:** Managing the control files, Maintaining and monitoring redo log files **Managing tables, indexes and constraints:** Storing data (create, alter, analyzing, querying table information), Managing indexes, Managing constraints | 15 |
| **II** | **Managing Users and Security:** Profiles, managing users, managing privileges, managing roles, querying role information. **Introduction to Network Administration:** Network design considerations, network responsibilities for the DBA, network configuration, Overview of oracle Net features, Oracle Net Stack Architecture **Managing Data Concurrency:** Locking mechanism, Oracle data concurrency management, Enque mechanism, Monitoring and resolving locking conflicts Managing Undo Data, DML and undo data generation, Monitor and administer undo data, Difference between undo data and redo data, Configuring undo retention, | 15 |

| | | |
|---|---|---|
| | Undo retention guarantee, Undo Advisor | |
| **III** | **Backup and Recovery Overview**: Database backup, restoration and recovery, Types of failure in oracle environment, defining a backup and recovery strategy, Testing the backup and recovery plan. **Database Maintenance:** Managing optimizer statistics, Preferences for Gathering Statistics, Managing the Automatic Workload Repository (AWR), Statistic Levels, Automatic Database Diagnostic Monitor (ADDM), Advisory framework, Automated Maintenance Tasks, Server-generated alerts, Setting alert thresholds, Reacting to alerts, Alert types and clearing Alerts **Information Security Overview:** The Importance of Information Protection, The Evolution of Information Security, Justifying Security Investment, Security Methodology, How to Build a Security Program, The Impossible Job, Strategy and Tactics, Risk Analysis: Threat Definition, Types of Attacks, Risk Analysis. Secure Design Principles: The CIA Triad and Other Models, Defense Models, Zones of Trust, Best Practices for Network Defense. **Web and Database security techniques:** SQL Injection, Cross Site Scripting, Database privileges, Multilevel databases, Query modification, Social engineering and Phishing Attacks. | 15 |

**Books and References**

| Sr. No. | Title | Author/s | Publisher | Edition | Year |
|---|---|---|---|---|---|
| 1. | Student Guides for Oracle Database Administration | Rajesh Bhatiya, Padmaja Potineni | Oracle | 2nd | 2020 |
| 2. | Concurrency Control and Recovery in Database Systems | Deepak Khemani | TMH | 1st | 2018 |
| 3. | Databases and Transaction Processing | Lewis, Bernstein, Kifer, Addison Wesley | | 1st | 2001 |

# SIPDSELP611 : Database Administration and Security Practical

| M.Sc. (Data Science) | Semester – III |
|---|---|
| **Course Name: Database Administration and Security Practical** | **Course Code: SIPDSELP611** |
| **Credits** | **1** |

**List of Practical:**

| | |
|---|---|
| **1.** | **Creating Oracle Database** |
| a. | Create a Database with the Database Configuration Assistant (DBCA) |
| b. | Create a Database design template with the DBCA |
| c. | Generate database creation scripts with the DBCA |
| **2.** | **Managing the Oracle Instance** |
| a. | Start and Stop the Oracle database and components. |
| b. | Use Enterprise Manager (EM) |
| **3.** | **Administering Use Security** |
| a. | Create and manage database user account |
| b. | Authentication users |
| c. | Grant and revoke privileges |
| d. | Create and manage roles |
| e. | Implement standard password security features |
| **4.** | **Managing Schema Objects** |
| a. | Create and modify tables |
| b. | Define constraint |
| c. | View the columns and contents of a table |
| **5.** | **Managing undo Data** |
| a. | Monitor and administer undo |
| b. | Configure undo retention |
| **6.** | **Configure the Oracle Network Environment** |
| a. | Create additional listener |
| b. | Create Net Service aliases |
| c. | Configure connect-time failover |
| **7.** | Performing Databases backups |
| **8.** | Performing database Recovery |

# SIPDSRP611 : Project

| M.Sc. (Data Science) | Semester – III |
|---|---|
| **Course Name: Project** | **Course Code: SIPDSRP611** |
| **Credits** | **6** |

## M.Sc. Part – II Semester – IV

| Course Code | Course Type | Course Title | Credits |
|---|---|---|---|
| SIPDSCC621 | Core Subject (Major) | Deep Learning | 4 |
| SIPDSCC622 | Core Subject (Major) | Next Generation Technologies | 4 |
| SIPDSEL621 | Core Subject (DSC) | Time Series and Forecasting | 3 |
| SIPDSCCP621 | Core Subject Practical (Major) | Deep Learning – Practical | 2 |
| SIPDSCCP622 | Core Subject Practical (Major) | Next Generation Technologies – Practical | 2 |
| SIPDSELP621 | Core Subject Practical (DSC) | Time Series and Forecasting – Practical | 1 |
| SIPDSRP621 | Research Project | Project / | 6 |
| | | **Total Credits** | **22** |

# SIPDSCC621 : Deep Learning

| M.Sc. (Data Science) | Semester – IV |
|---|---|
| **Course Name: Deep Learning** | **Course Code: SIPDSCC621** |
| **Credits** | **4** |

## Course Objective:

- To know the importance of deep learning.
- To acquire knowledge of the basics of neural networks.
- To implement neural networks using computational tools for a variety of problems.
- To explore various deep learning algorithms.

## Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Develop algorithms simulating human brain.
- **CO2:** Analyze ANN learning and memory-based learning.
- **CO3:** Explore the essentials of Deep Learning and Deep Network architectures.
- **CO4:** Define, train and use a Deep Neural Network for solving real world problems that require artificial Intelligence based solutions.
- **CO5:** Use deep learning methodology in real world application.

| Unit | Contents | No. of Lectures |
|---|---|---|
| **I** | **Neural Networks:** The Neuron –Expressing Linear Perceptrons as Neurons – Feed-Forward Neural Networks – Linear Neurons and their Limitations – Sigmoid, Tanh and Relu Functions – Softmax Output Layers. **Neural Learning:** Measuring Errors - Gradient Descent – Delta Rule and Learning Rate – Backpropagation – Stochastic and Minibatch Gradient – Test Sets, Validation Sets and Overfitting – Preventing Overfitting in Deep Neural Networks – **Other Optimization Algorithms:** Adagrad, RMSProp, Adadelta, Adam | 15 |
| **II** | **Convolution Neural Networks:** Neurons in Human Vision – Shortcomings of Feature Selection –Scaling Problem in Vanilla Deep Neural Networks – Filters and Feature Maps – Description of Convolutional Layer – Maxpooling – Convolution Network Architecture – Image Classification **Pre-Trained Models**: Self-Supervised Pretraining, AlexNet, VGG, NiN, GoogleNet, Residual Network (ResNet), DenseNet, Region-Based CNNs (R-CNNs) – Transfer Learning - FSL | 15 |
| **III** | **Recurrent Neural Networks:** Sequence-to-Sequence Modeling – Embedding - Recurrent Neural Networks - Bidirectional RNNs, Analyzing Variable Length Inputs – Tackling seq2seq Problem – Beam Search and Global Normalization – Recurrent Neural Networks (RNN)– Hidden States – Perplexity – Character-level Language Models –**Modern RNNs:** Gated Recurrent Units (GRU), Long Short Term Memory (LSTM), Bidirectional Long Short Term Memory (BLSTM), Deep Recurrent Neural Network, Bidirectional RNN. **Attention Models and Transformers:** | 15 |

| | | |
|---|---|---|
| | **Attention Mechanism:** Attention Cues, Attention Pooling, Scoring Functions, Self Attention and Positional Encoding; –Bidirectional Encoder Representations from Transformers (BERT) – Generative Pre-trained Transformers | |
| **IV** | **Advanced Neural Networks:** Generative Adversarial Networks – Generator, Discriminator, Training, GAN variants; Autoencoder: Architecture, Denoising and Sparcity; DALL-E, DALL-E 2 and IMAGEN | 15 |

**Books and References**

| Sr. No. | Title | Author/s | Publisher | Edition | Year |
|---|---|---|---|---|---|
| 1 | Deep Learning | Ian Goodfellow, Yoshua Bengio, Aaron Courville | MIT Press | | 2016 |
| 2 | Neural Networks and Deep Learning | Michael Nielsen | Determination Press | | 2015 |
| 3 | Deep Learning with Python | Francois Chollet | Manning Publications | First | 2017 |
| 4 | Dive into Deep Learning | Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola | Cambridge University Press | | 2024 |

# SIPDSCCP621 : Deep Learning Practical

| M.Sc. (Data Science) | Semester – IV |
|---|---|
| **Course Name: Deep Learning Practical** | **Course Code: SIPDSCCP621** |
| **Credits** | **2** |

**List of Practical:**

| | |
|---|---|
| **1.** | Python Primer<br>Revisiting Data Preprocessing<br>Setting up Deep-Learning workstations<br>Working with different data types and file formats |
| **2.** | Simple Classification Tasks<br>Working with MNIST – IMDB Datasets |
| **3.** | Training a CNN from Scratch<br>Using pre-trained CNNs |
| **4.** | Visualizing what CNNs are learning – Intermediate Activations, Convnet Filters, Heatmaps |
| **5.** | Exploring Multi-Input, Multi-output Models<br>Hyper-parameter Tuning |
| **6.** | Language Modeling using RNN<br>Practicing of Stacking Layers in Bidirectional RNNs |
| **7.** | Transfer Learning models for classification problems<br>Exploring Hugging-face API |
| **8.** | Text Generation Using LSTM |
| **9.** | Image generation from Text using GAN |
| **10.** | Demonstrate video to text with LSTM model. |

# SIPDSCC622 : Next Generation Technologies

| M.Sc. (Data Science) | Semester – IV |
|---|---|
| **Course Name: Next Generation Technologies** | **Course Code: SIPDSCC622** |
| **Credits** | **4** |

## Course Objective:

- To acquire knowledge of the basics of Microsoft Azure.
- To implement virtual machines in Azure.
- To acquire knowledge about web3 Technologies Blockchain.
- To acquire knowledge about the basics of Flask.
- To acquire knowledge about Docker and Kubernetes.

## Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Configure and manage virtual machines in Azure.
- **CO2:** Understand blockchain technology and develop blockchain based solutions.
- **CO3:** Build application using Flask
- **CO4:** Understand the basic concepts of containerization and how Docker operates on various systems.
- **CO5:** Understand the basics and architecture of Kubernetes.

| Unit | Contents | No. of Lectures |
|---|---|---|
| I | **Getting started with Microsoft Azure:** What is Azure? Overview of cloud computing, Azure services. Azure Resource Manager: Why use Resource Manager? The classic deployment model, Role based Access Control, The Azure portal: Dashboard and hub. **Creating and viewing resources: Azure Virtual Machines:** What is Azure Virtual Machines? Billing, Service level agreement, Virtual machine models, Azure Resource Manager model, Classic/Azure Service Management model **Virtual machine components:** Virtual machine, Disks, Virtual Network, Availability set. **Create virtual machines:** Create a virtual machine with the Azure portal, Create a virtual machine with a template | 15 |
| II | **Connecting to a virtual machine:** Remotely access a virtual machine, Network connectivity **Configuring and managing a virtual machine:** Disks, Fault domains and update domains, Image capture **Scaling Azure Virtual Machines:** Resource Manager virtual machines, Classic virtual machines **Azure Storage:** Storage accounts: General-purpose storage accounts, Blob storage accounts. Storage services: Blob storage, File storage, Table storage, Queue Storage, Redundancy | 15 |
| III | **Introduction to Web3 Technologies Blockchain:** Growth of blockchain technology, Distributed systems, the history of blockchain and Bitcoin, Blockchain, Consensus, CAP theorem and blockchain, Decentralization using blockchain, Methods of decentralization, Routes to decentralization, Blockchain and full ecosystem decentralization, The consensus problem, | 15 |

| | | |
|---|---|---|
| | Analysis and design, Classification, Algorithms, Bitcoin: Overview, Cryptographic keys, Transactions, Blockchain Mining, Bitcoin network, Wallets, Bitcoin payments, Innovation in Bitcoin, Advanced protocols, Bitcoin investment, and buying and selling Bitcoin | |
| **IV** | **Introduction to Flask:** Installation, Basic application structure, routing, variables, redirect and errors, Templates, cookies, session, webforms, Databases.<br>**Introduction to Docker:** Introduction to Docker Architecture and Container Life Cycle, Understanding images and containers, Create and Implement docker images using Dockerfile, Container Lifecycle and working with containers. To Build, deploy and manage web or software application on Docker Engine, Publishing image on Docker Hub.<br>**Introduction to Kubernetes:** Install and configure Kubernetes, Spin Up a Kubernetes Cluster, Check the Nodes of Your Kubernetes Cluster, Installing kubectl to manage cluster and deploy Your First Kubernetes Application | 15 |

**Books and References**

| Sr. No. | Title | Author/s | Publisher | Edition | Year |
|---|---|---|---|---|---|
| 1 | Fundamentals of Azure | Michael Collier, Robin Shahan | Microsoft Press | 2nd | 2016 |
| 2 | Mastering Flask Web Development | Daniel Gaspar, Jack Stouffer | Packt | 2nd | |
| 3 | Mastering Bitcoin: Programming the Open Blockchain | Andreas M. Antonopoulos | O'Reilly | 3rd | 2017 |
| 4 | Kubernetes and Docker - An Enterprise Guide: Effectively containerize applications, integrate enterprise systems, and scale applications in your enterprise | Scott Surovich, Marc Boorshtein | Packt | | 2020 |

# SIPDSCCP622 : Next Generation Technologies Practical

| M.Sc. (Data Science) | Semester – IV |
|---|---|
| **Course Name: Next Generation Technologies Practical** | **Course Code: SIPDSCCP622** |
| **Credits** | **2** |

**List of Practical:**

| | |
|---|---|
| **1.** | **Introduction to Microsoft Azure:** |
| **a.** | Creating Virtual Machines with Azure Portal |
| **b.** | Creating Virtual Machines with template |
| **2.** | **Configuring and managing Virtual machine in Azure** |
| **3.** | **Scaling Azure virtual machine** |
| **4.** | **Introduction to Flask:** |
| **a.** | Create Flask application |
| **b.** | Show the use of cookies and sessions |
| **5.** | **Implement Docker commands** |
| **6.** | **Implement Kubernetes commands** |

# SIPDSEL621 : Time Series and Forecasting

| M.Sc. (Data Science) | Semester – IV |
|---|---|
| **Course Name: Time Series and Forecasting** | **Course Code: SIPDSEL621** |
| **Credits** | **3** |

## Course Objective:

- Present time series in an informative way, both graphically and with summary statistics
- Model time series to analyses the underlying structure(s) in both the time and frequency domains.

## Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Forecast the trend pattern exhibited by the given data by using various methods
- **CO2:** Run and interpret time series models and regression models for time series
- **CO3:** Use the Box-Jenkins approach to model and forecast time series data empirically.
- **CO4:** Analyze and estimate the cyclic components using special processes.

| Unit | Contents | No. of Lectures |
|---|---|---|
| I | **Introduction To Trend :** Introduction to times series data, application of time series from various fields, Components of a time series, Decomposition of time series. Trend: Estimation of trend by free hand curve method, method of semi averages, fitting a various mathematical curve, and growth curves. **Trend And Seasonal Component:** Method of moving averages. Detrending. Effect of elimination of trend on other components of the time series. Seasonal Component: Estimation of seasonal component by Method of simple averages, Ratio to Trend, Ratio to moving average and Link relatives. | 15 |
| II | **Forecasting:** Variate component method: Stationary Time series: Weak stationary, auto correlation function and correlogram of moving average. Forecasting: Exponential smoothing methods, Short-term forecasting methods: Brown's discounted regression, Box-Jenkins Method. **Cyclic Component:** Deseasonalization , Cyclic Component: Harmonic Analysis. Some Special Processes: Moving-average (MA) process and Autoregressive (AR) process of orders one and two, Estimation of the parameters of AR (1) and AR (2) – Yule-Walker equations. | 15 |
| III | **Non-stationary time series:** Non-stationary time series. Time series with non-stationary variance. Non-stationary mean. ARIMA (p,d,q) models. The use of Box-Jenkins methodology to determination of order of integration. **The unit root problem:** The unit root problem. Spurious trends and regressions. Unit root tests (Dickey-Fuller). ADF test and the choice of the number of lags. Other unit root tests. **Unit root and structure changes:** Non-stationary time series, TSP or DSP: methodology of research. Segmented trends and structure changes. | 15 |

| | | |
|---|---|---|
| | **Regressive dynamic models:** Regressive dynamic models. Autoregressive models with distributed lags (ADL). **Vector autoregression model and co-integration:** Time series co-integration. Co-integration regression. Testing of co-integration. Vector autoregression and co-integration. Co-integration and error correction model. | |

**Books and References**

| Sr. No. | Title | Author/s | Publisher | Edition | Year |
|---|---|---|---|---|---|
| 1 | Practical Time Series Analysis Prediction with Statistics and Machine Learning | Aileen Nielsen | O'Reilly Media | 1st | 2019 |
| 2 | The Analysis of Time Series: An Introduction with R | Chris Chatfield , Haipeng Xing | Chapman and Hall/CRC | 7th | 2019 |
| 3 | Time series analysis and its applications | Robert H. Shumway, David S. Stoffer | Springer New York | | 2000 |

# SIPDSELP621 : Time Series and Forecasting Practical

| M.Sc. (Data Science) | Semester – IV |
|---|---|
| **Course Name: Time Series and Forecasting Practical** | **Course Code: SIPDSELP621** |
| **Credits** | **1** |

**List of Practical:**

| | |
|---|---|
| **1.** | Fitting and plotting of modified exponential curve |
| **2.** | Fitting and plotting of Gompertz curve |
| **3.** | Fitting and plotting of logistic curve |
| **4.** | Fitting of trend by Moving Average Method |
| **5.** | Measurement of Seasonal indices Ratio-to-Trend method |
| **6.** | Measurement of Seasonal indices Ratio-to-Moving Average method |
| **7.** | Measurement of seasonal indices Link Relative method |
| **8.** | Calculation of variance of random component by variate difference method |
| **9.** | Forecasting by exponential smoothing |
| **10.** | Forecasting by short term forecasting methods |

# SIPDSRP621 : Project

| M.Sc. (Data Science) | Semester – IV |
|---|---|
| **Course Name: Project** | **Course Code: SIPDSRP621** |
| **Credits** | **6** |