School of Engineering and Sciences

M. Tech in Data Science

Academic Batch: 2023-2025

Department of Computer Science Engineering
SRM University - AP, Andhra Pradesh.

Department of Computer Science and Engineering
SRM University-AP, Andhra Pradesh
Curriculum
2023-2025

Semester-I			
CODE	**COURSE NAME**	**CREDIT**	**L - T – P**
MAT 500	Mathematics and Statistical Foundations for Data Science	3	3-0-0
AML500	Advanced Algorithms and Analysis	3	3-0-0
AML501	Machine Learning Techniques	3	3-0-0
AML501L	Machine Learning Techniques Lab	2	0-0-3
DSC 501	Computational Essentials for Data Science	3	3-0-0
DSC 501L	Computational Essentials for Data Science Lab	2	0-0-3
DSC 502	Big Data Analytics-I	3	3-0-0
DSC 502L	Big Data Analytics-I Lab	2	0-0-3
EGL501	English for Research Paper Writing	1	1-0-0
	Total	**22**	**16-0-9**

Semester-II			
CODE	**COURSE NAME**	**CREDIT**	**L - T – P**
MAT561	Optimization Techniques	3	3-0-0
AML504	Data Warehousing and Pattern Mining	3	3-0-0
AML504L	Data Warehousing and Pattern Mining Lab	2	0-0-3
DSC 503L	Big Data Analytics-II Lab	2	0-0-3
AML 505	Deep Learning Techniques	3	3-0-0
AML 505 L	Deep Learning Techniques Lab	2	0-0-3
RM101	Research Methodology & IPR	2	2-0-0
	Elective-I	3	3-0-0
	Elective-II	3	3-0-0
	Total	**23**	**17-0-9**

Semester-III			
CODE	**COURSE NAME**	**CREDIT**	**L - T – P**
DSC 680	Project Work- Phase I	13	0-0-26
	Total	**13**	**0-0-26**

Semester-IV			
CODE	**COURSE NAME**	**CREDIT**	**L – T - P**
DSC 681	Project Work –Phase II	13	0-0-26
	Total	**13**	**0-0-26**

Total: 22+23+13+13 **Total: 71 credits**

List of Electives

List of Electives					
Course Code	**Course Name**	**L**	**T**	**P**	**C**
AML557	Introduction to High Performance Computing	3	0	0	3
AML562	Statistical Modelling for Computer Sciences	3	0	0	3
AML563	Fuzzy Logic and its Applications	3	0	0	3
AML553	Information Retrieval	3	0	0	3
AML554	Pattern Recognition	3	0	0	3
AML552	Knowledge Engineering and Expert Systems	3	0	0	3
DSC 550	Time series analysis and forecasting	3	0	0	3
DSC 551	Complex Networks Analysis	3	0	0	3

DSC 552	Recommender Systems	3	0	0	3
DSC 553	Big Data Security and Privacy	3	0	0	3
DSC 554	Spatial Data Science and Visualization	3	0	0	3
DSC 555	Data Science for Healthcare	3	0	0	3
DSC 556	Business Intelligence and Data Analytics	3	0	0	3

SEMESTER-I

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
MAT500	Mathematics and Statistical Foundations for Data Science	BS	3	0	0	3

UNIT I: PROBABILITY

Classical, relative frequency and axiomatic definitions of probability, addition rule and conditional probability, multiplication rule, total probability, Bayes' Theorem and independence, Bayesian Inference,

UNIT II: STATISTICS

Basic Statistics, Estimate, Learning, Regression, Linear Regression, Multiple Linear Regression, Logistic Regression, Classification, Bayes Theorem for Classification, Sampling Methods, and Resampling Methods

UNIT III: RANDOM VARIABLES

Discrete, continuous and mixed random variables, probability mass, probability density and cumulative distribution functions, mathematical expectation, moments, moment generating function, Chebyshev's inequality, Introduction to Stochastic Processes (SPs), Stationary Processes, Discrete-time Markov Chains (DTMCs), Continuous-time Markov Chains (CTMCs)

UNIT IV: LINEAR ALGEBRA

Finite dimensional vector spaces over a field; linear combination, linear dependence and independence; basis and dimension; inner-product spaces, linear transformations; matrix representation of linear transformations, Eigen values and eigenvectors, rank and nullity, inverse and linear transformation, Cayley-Hamilton Theorem

UNIT V: LINEAR ALGEBRA

Subset Selection, Shrinkage Methods, Dimension Reduction Methods, Support Vector Machine, principal Component analysis,

TEXT BOOKS

1. Sheldon Ross, A First Course in Probability, 7th Edition, Pearson, 2006
2. J. Medhi, Stochastic Processes, 3rd Edition, New Age International, 2009.

REFERENCES BOOKS

1. S.M. Ross, Stochastic Processes, 2nd Edition, Wiley, 1996.
2. Stephen H Friedberg, Arnold J Insel, Lawrence E. Spence, Linear Algebra. 4th Edition, Pearson, 2006.
3. Kenneth M Hoffman, Ray Kunz, Linear Algebra, 2nd Edition, Pearson
4. An Introduction to Statistical Learning - with Applications in R by Gareth James and Springer.

SEMESTER-I

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML500	Advanced Algorithms and Analysis	C	3	0	0	3

UNIT I: INTRODUCTION TO AAA

Defining Key Terms: Algorithm complexity, Greedy method, Dynamic Programming, Backtracking, Branch-and-bound Techniques; Examples for understanding above techniques; Memory model, linked lists and basic programming skills.

UNIT II: NP COMPLETENESS

Overview - Class P - Class NP - NP Hardness - NP Completeness - Cook Levine Theorem - Important NP Complete Problems. Heuristic and Randomized algorithms.

UNIT III

Use of probabilistic inequalities in analysis, Amortized Analysis - Aggregate Method - Accounting Method - Potential Method, competitive analysis, applications using examples.

UNIT IV: GEOMETRIC ALGORITHMS

Point location, Convex hulls and Voronoi diagrams, Arrangements, graph connectivity, Network Flow and Matching: Flow Algorithms - Maximum Flow – Cuts - Maximum Bipartite Matching - Graph partitioning via multi-commodity flow, Karger'r Min Cut Algorithm, String matching and document processing algorithms.

UNIT V: APPROXIMATION ALGORITHMS

Approximation algorithms for known NP hard problems - Analysis of Approximation Algorithms Use of Linear programming and primal dual; local search heuristics; Parallel algorithms: Basic techniques for sorting, searching, merging, list ranking in PRAMs and Interconnection.

TEXT BOOKS

1. Allan Borodin and Ran El-Yaniv: Online Computation and Competitive Analysis, Cambridge University Press, 2005.
2. Michael T Goodric and Roberto Tamassia, “Algorithm Design: Foundations, Analysis and Internet Examples”, John Wiley and Sons, 2002.

REFERENCES

1. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, “Introduction to Algorithms”, Third Edition, The MIT Press, 2009.
2. Sanjoy Dasgupta, Christos Papadimitriou and Umesh Vazirani, “Algorithms”, Tata McGraw-Hill, 2009.
3. RK Ahuja, TL Magnanti and JB Orlin, “Network flows: Theory, Algorithms, and Applications”, Prentice Hall Englewood Cliffs, NJ 1993.
4. Joseph JáJá: Introduction to Parallel Algorithms 1992.
5. Rajeev Motwani, Prabhakar Raghavan: Randomized Algorithms, Cambridge University Press, 1995.
6. Jiri Matousek and Bernd Gärtner: Understanding and Using Linear Programming, 2006.

SEMESTER-I

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML501	Machine Learning Techniques	C	3	0	0	3

UNIT I: INTRODUCTION
Introduction to Machine Learning: Introduction. Different types of learning, Hypothesis space and inductive bias, Evaluation. Training and test sets, cross validation, Concept of over fitting, under fitting, Bias and Variance.

Linear Regression: Introduction, Linear regression, Simple and Multiple Linear regression, Polynomial regression, evaluating regression fit.

UNIT II: DECISION TREE LEARNING:
Introduction, Decision tree representation, appropriate problems for decision tree learning, the basic decision tree algorithm, hypothesis space search in decision tree learning, inductive bias in decision tree learning, issues in decision tree learning, Python exercise on Decision Tree.
Instance based Learning: K nearest neighbor, the Curse of Dimensionality, Feature Selection: forward search, backward search, univariate , multivariate feature selection approach, Feature reduction (Principal Component Analysis) , Python exercise on kNN and PCA.

Recommender System: Content based system, Collaborative filtering based.

UNIT III: PROBABILITY AND BAYES LEARNING:
Bayesian Learning, Naïve Bayes, Python exercise on Naïve Bayes, Logistic Regression. Support Vector Machine: Introduction, the Dual formulation, Maximum margin with noise, nonlinear SVM and Kernel function, solution to dual problem.

UNIT IV
Artificial Neural Networks: Introduction, Biological motivation, ANN representation, appropriate problem for ANN learning, Perceptron, multilayer networks and the back propagation algorithm,

UNIT V: ENSEMBLES:
Introduction, Bagging and boosting, Random forest, Discussion on some research papers. Clustering: Introduction, K-mean clustering, agglomerative hierarchical clustering, Python exercise on k-mean clustering.

TEXTBOOKS

1. Machine Learning. Tom Mitchell. First Edition, McGraw- Hill, 1997.

2. Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020.

REFERENCES

1. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012.

2. Christopher Bishop, “Pattern Recognition and Machine Learning” Springer, 2007.

SEMESTER-I

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML501L	Machine Learning Techniques Lab	C	0	0	3	2

LIST OF PRACTICAL EXPERIMENTS

1. Basic exercises on Python Machine Learning P
2. ackages such as Numpy, Pandas and matplotlib.
3. Given a dataset. Write a program to compute the Covariance, Correlation between a pair of attributes. Extend the program to compute the Covariance Matrix and Correlation Matrix.
4. Given a set of sample points in N dimensional feature space. Write a program to fit the points with a hyper plane using Linear Regression. Calculate sum of residual error.
5. Write a program that provides option to compute different distance measures between two points in the N dimensional feature space. Consider some sample datasets for computing distances among sample points.
6. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
7. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Python ML library classes can be used for this problem.
8. Write a program to implement feature reduction using Principle Component Analysis
9. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
10. Given a dataset for classification task. Write a program to implement Support Vector Machine and estimate it test performance.
11. Write a program to implement perceptron for different learning task.
12. Write programs to implement ADALINE and MADALINE for given learning task.
13. Build an Artificial Neural Network by implementing the Back propagation algorithm and test the same using appropriate data sets.
14. Write a program to implement K means clustering algorithm. Select your own dataset to test the program. Demonstrate the nature of output with varying value of K

SEMESTER-I

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
DSC 501	Computational Essentials for Data Science	C	3	0	0	3

UNIT I: INTRODUCTION TO DATA SCIENCE

Introduction: Overview of Data Science, The Art of Data Science, The Data Scientist. Data Wrangling: Data Ingestion, Data Review, Data Cleaning, Variable Roles, Feature Selection, Missing Data, Feature Creation, Preparing the Metadata, Preparing for Model Building, Save the Dataset, A Template for Data Preprocessing. Introduction data science and Python: Installing python, A glance at the essential Python packages: Numpy, Scipy, pandas, Scikit-learn, IPyhotn, Matplotlib, NetworkX.

UNIT II: DATA COLLECTION, PRE-PROCESSING AND ANALYTICS

Data Collection Strategies, Data Pre-Processing Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization. Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics

UNIT III: DATA CLASSIFICATION

A Pattern for Predictive Modeling: Loading the Dataset, Building a Decision Tree Model, Python exercise on Decision Tree. Instance based Learning: K nearest neighbor, Model Performance, Evaluating Model Generality, Model Tuning, Comparison of Performance Measures, save the Model to File, A template for Predictive Modeling.

UNIT IV: DATA VISUALIZATION

Data Visualization - Basic principles, ideas and tools for data visualization – Case Study 1 on industry projects – Case Study 2: Create Complex visualization dataset - Data Science and Next-generation data scientists. Visualising Data: Preparing the Dataset, Scatter Plot, Bar Charts, Saving Plots to File, Adding Spice to the Bar Chart, Alternative Bar Charts, Box Plots.

UNIT V: ADVANCES IN DATA SCIENCE

Data loading and preprocessing with pandas, Working with categorical and textual data, Data preprocessing with NumPy, Creating NumPy arrays, NumPy fast operation and computations. Data Analytics using R: Introduction to R, creating a dataset, Getting started with graphs, Basic data management, Advanced data management.

TEXTBOOKS

1. The Essential of Data Science: Knowledge Discovery Using R, Graham J.Williams, The R Series, 2017.
2. Python Data Science Essentials, , Alberto Boschetti, Luca Massaron, 2015.
3. Cathy O'Neil and Rachel Schutt , "Doing Data Science", O'Reilly, 2015

REFERENCES

1. Python Data Science Handbook: Essentials Tools for Working with Data, 1st edition, Jake VanderPlas, O'REILLY.

SEMESTER-I

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
DSC 501L	Computational Essentials for Data Science Lab	C	0	0	3	2

LIST OF PRACTICAL EXPERIMENTS

1. Python basics, creating and manipulating a List, and dictionaries
2. Basic exercises on Python Packages such as Numpy, SciPy, Pandas, Scikit-learn, Matplotlib, NetworkX.
3. Basic exercises on IPython Notebook. Write a program using Datasets and code used in the book.
4. Creating a Data Frame and Matrix-like Operations on a Data Frame from CSV or text files and Applying functions to Data Frames
5. Write a program to implement how to structurally load, manipulate, preprocess, and polish data with pandas and NumPy.
6. Write a program to implement working with categorical and textual data.
7. Write a program to implement creating NumPy arrays, Unidimensional and Multidimensional arrays.
8. Write a program to demonstrate the Introducing EDA (Exploratory Data Analysis), Feature creation.
9. Write a program to demonstrate the different Scoring functions (binary, multilabel classification).
10. Write a program to implement the building custom scoring functions.
11. Write a program to demonstrate the different feature selection methods.
12. Implement different data structures in R.
13. Simple Linear Regression – Fitting, Evaluation and Visualization and Multiple Linear Regression, Lasso and Ridge Regression in Python and R.

SEMESTER-I

Course Code	Course Title	Core/ Elective	Credits			
			L	T	P	C
DSC 502	Big Data Analytics - I		3	0	0	3

.

UNIT I

Understanding Big Data – Concepts and Terminology – Big Data Characteristics – Different types of Data – Big Data Storage concepts – Clusters – File systems and distributed file systems – Sharding – Replication – CAP theorem – BASE - Hadoop Distributed File System (HDFS) Architecture - HDFS commands for loading/getting data - Accessing HDFS through Java program.

UNIT II

Big Data Processing Concepts – Parallel Data Processing – Distributed Data Processing – Hadoop – Processing workloads – Batch processing with MapReduce – Map and Reduce Tasks – MapReduce Example

UNIT III

Big Data Analysis Techniques – Quantitative Analysis – Qualitative Analysis – Data Mining – Statistical Analysis – Machine Learning – Semantic Analysis.

UNIT IV

Hadoop echo system and its components– Flume - Sqoop - Kafka - Pig - Spark - Hive.

UNIT V

NoSQL databases: Introduction, NoSQL vs SQL, Types of NoSQL databases- Key-value data store, Dynamo DB, Document Store, MongoDB, Wide-column store, HBase, Cassandra, Graph Store, Neo4j.

TEXT BOOKS

1. Big Data Fundamentals: concepts, Drivers and Techniques: Pearson Education, 2016
2. Hadoop The Definitive Guide, IV edition, O'Reilly publications

REFERENCES:

1. Programming, Hive, O'Reilly publications
2. Programming Pig, O'Reilly publications
3. Learning Spark, O'Reilly publications
4. NOSQL for Mere Mortals, O'Reilly publications

SEMESTER-I

Course Code	Course Title	Core/ Elective	Credits			
			L	T	P	C
DSC502L	Big Data Analytics -I Lab		0	0	3	2

LIST OF PRACTICAL EXPERIMENTS

1. Hadoop installation.
2. Building Hadoop single node and multi-node clusters.
3. Working with Hadoop Distributed File system commands.
4. MapReduce application for word counting on Hadoop cluster.
5. MapReduce application for matrix multiplication.
5. Working with managed and external tables in HIVE.
6. Write Pig Latin program to sort, group, join, project, and filter the data.
7. Data ingestion using Flume.
8. Perform import and export operations using Sqoop.
9. Data stream processing using Kafka.

SEMESTER-I

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
EGL501	English for Research Paper Writing	HS	1	0	0	1

UNIT I
Planning and Preparation, Word Order, Breaking up long sentences, Structuring Paragraphs and Sentences, Being Concise and Removing Redundancy, Avoiding Ambiguity and Vagueness

UNIT II
Clarifying Who Did What, Highlighting Your Findings, Hedging and Criticising, Paraphrasing and Plagiarism, Sections of a Paper, Abstracts. Introduction

UNIT III
Review of the Literature, Methods, Results, Discussion, Conclusions, The Final Check.

UNIT IV
Key skills are needed when writing a Title, key skills are needed when writing an Abstract, key skills are needed when writing an Introduction, skills needed when writing a Review of the Literature,

UNIT V
Skills are needed when writing the Methods, skills needed when writing the Results, skills are needed when writing the Discussion, skills are needed when writing the Conclusions

UNIT VI
Useful phrases, how to ensure paper is as good as it could possibly be the first- time submission

TEXTBOOKS

1. Gold bort R (2006) Writing for Science, Yale University Press (available on Google Books).
2. Day R (2006) How to Write and Publish a Scientific Paper, Cambridge University Press.

REFERENCES:

1. Highman N (1998), Handbook of Writing for the Mathematical Sciences, SIAM. Highman's book.
2. Adrian Wallwork, English for Writing Research Papers, Springer New York Dordrecht Heidelberg London, 2011.

SEMESTER-II

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
MAT561	Optimisation Techniques	BS	3	0	0	3

UNIT I

Historical Development; Engineering applications of Optimization; Art of Modeling, Objective function; Constraints and Constraint surface; Formulation of design problems as mathematical programming problems. Classification of optimization problems, Optimization techniques – classical and advanced techniques**,**

Introduction to Operation Research: Operation Research approach, scientific methods, introduction to models and modeling techniques, general methods for Operation Research models, methodology and advantages of Operation Research, history of Operation Research.

UNIT II

Linear Programming (LP): Introduction to LP and formulation of Linear Programming problems, Graphical solution method, alternative or multiple optimal solutions, Unbounded solutions, Infeasible solutions, Maximization – Simplex Algorithm, Minimization – Simplex Algorithm using Big-M method, Two phase method, Duality in linear programming, Integer linear programming.

UNIT III

Allocation problems and Game Theory: Introduction to Transportation problems, Transportation problem –Methods of basic feasible solution -Optimal solution–MODI Method. Assignment problem-Hungarian method

Game theory: Two people-zero sum game-mixed stages -Dominance properties

UNIT IV

Sequential optimization; Representation of multi stage decision process Types of multi stage decision problems; Concept of sub optimization and the principle of optimality. Recursive equations –Forward and backward recursions; Computational procedure in dynamic programming (DP), Discrete versus continuous dynamic programming; Multiple state variables; curse of dimensionality in DP; Problem formulation and application in Design of continuous beam and optimal geometric layout of atruss

UNIT V

Network Analysis: Network definition and Network diagram, probability in PERT analysis, project time cost trade off, introduction to resource smoothing and allocation

Sequencing: Introduction, processing N jobs through two machines, processing N jobs through three machines, processing N jobs through m machines.

Inventory Model: Introduction to inventory control, deterministic inventory model, EOQ model with quantity discount.

TEXT BOOKS

1. Hamdy A. Taha, Operations Research, Prentice Hall, Pearso.

2. J. S Arora, Introduction to optimum design, IInd edition, Elsevier India Pvt. Ltd.,

REFERENCES

1. S. S Rao, Optimization: theory and application, Wiley Eastern Ltd., New Delhi.
2. Wayne L. Winston - Operations Research_ Applications and Algorithms-Duxbury Press (2003).
3. Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin, Network Flows: Theory, Algorithms, and Applications, Pearson.
4. J K Sharma, Operations Research Theory and Applications, MacMillan India Ltd.
5. N D Vohra, Quantitative Techniques in management, Tata McGraw Hill.
6. Payne T A, Quantitative Techniques for Management: A Practical Approach, Reston Publishing Co. Inc., Virginia.
7. AchilleMessac, Optimization in practice with MATLAB, Cambridge University Press, 2015.

SEMESTER-II

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML504	Data Warehousing and Pattern Mining	C	3	0	0	3

UNIT – I

Data warehouse concepts, Data warehouse modeling, Data Cube and OLAP, schemas for multidimensional data models, concept hierarchy, measures, and indexing techniques. Data warehouse – design and usage, implementation, architectural components, Role of Metadata, Dimensional Modeling, Data Extraction, Transformation and Loading, Data Quality.

UNIT – II

Classification and prediction; Cluster Analysis – Types of Data in Cluster Analysis, Partitioning methods, Hierarchical Methods; Transactional Patterns and other temporal based frequent patterns.

Mining Time series Data, Periodicity Analysis for time related sequence data, Trend analysis, and Similarity search in Time-series analysis.

UNIT – III

Mining Data Streams, Methodologies for stream data processing and stream data systems, Frequent pattern mining in stream data, Sequential Pattern Mining in Data Streams, Classification of dynamic data streams.

UNIT – IV

Web Mining, Mining the web page layout structure, mining web link structure, mining multimedia data on the web, Automatic classification of web documents and web usage mining; Distributed Data Mining.

UNIT - V

Recent trends in Distributed Warehousing and Pattern Mining, Class Imbalance Problem; Graph Mining; Social Network Analysis.

TEXT BOOKS

1. Jiawei Han and M Kamber, Data Mining Concepts and Techniques, Second Edition, Elsevier Publication, 2011.

2. Vipin Kumar, Introduction to Data Mining - Pang-Ning Tan, Michael Steinbach, Addison Wesley, 2006

REFERENCES:

1. G Dong and J Pei, Sequence Data Mining, Springer, 2007.
2. Ralph Kimball, Margy Ross, The Data Warehouse Toolkit, 3rd edition, Publisher: Wiley, 2013.

SEMESTER-II

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML504L	Data Warehousing and Pattern Mining Lab	C	0	0	3	2

LIST OF PRACTICAL EXPERIMENTS

1. Basic exercises on Python Packages such as Numpy, Pandas and matplotlib.
2. Given a dataset. Write a program to compute the Mean, Median, Mode, Standard deviation, Covariance, Correlation between a pair of attributes.
3. Write a query to implementation OLAP operations in a data cube.
4. Write a program to implement data pre-processing techniques.
5. Write a program to implement data transformation using different normalization techniques.
6. Write a program that provides option to compute different distance measures between two points in the N dimensional feature space. Consider some sample datasets for computing distances among sample points.
7. Write a program to demonstrate the working of APRIORI algorithm. Use an appropriate data set to generate frequent patterns.
8. Write a program to demonstrate the working of stream mining algorithm. Use an appropriate data set to generate frequent patterns.
9. Write a program to implement K means clustering algorithm. Select your own dataset to test the program. Demonstrate the nature of output with varying value of K.
10. Write a program to demonstrate web page layout structure, web link structure.
11. Write a program to demonstrate graph mining considering a suitable dataset.

SEMESTER-II

Course Code	Course Title	Core/ Elective	Credits			
			L	T	P	C
DSC 503L	Big Data Analytics -II Lab		0	0	3	2

LIST OF PRACTICAL EXPERIMENTS

1. Apache Spark installation
2. Run Spark applications such as word count and matrix multiplication
3. Programs based on DataFrame API
4. Programs based on Spark MLlib - Machine learning library
5. Design an E-commerce product catalog system using MongoDB.
6. CURD operations in Cassandra.
7. Working with HBase commands.
8. Create multiple nodes and build relationships using Neo4j CQL.

SEMESTER-II

Course Code	Course Title	Core/ Elective	Credits			
			L	T	P	C
AML 505	Deep Learning Techniques	C	3	0	0	3

UNIT- I

Introduction: Overview of machine learning, linear classifiers, loss functions

Introduction to Tensor Flow: Computational Graph, Key highlights, Creating a Graph, Regression example, Gradient Descent, Tensor Board, Modularity, Sharing Variables, Keras

UNIT – II

Activation Functions: Sigmoid, ReLU, Hyperbolic Fns, Soft max

Perceptrons: What is a Perceptron, XOR Gate

Artificial Neural Networks: Introduction, Perceptron Training Rule, Gradient Descent Rule, vanishing gradient problem and solution

UNIT – III

Convolutional Neural Networks: Introduction to CNNs, Kernel filter, Principles behind CNNs, Multiple Filters, problem and solution of under fitting and over fitting, Transfer Learning: AlexNet, Google Net, VGG16

UNIT – IV

Recurrent Neural Networks: Introduction to RNNs, Unfolded RNNs, Seq2Seq RNNs, LSTM, GRU, Encoder Decoder architectures, Encoder Decoder architectures, Attention function

UNIT – V

Deep Learning applications: Image segmentation, Object detection, Attention model for computer vision tasks, Text Summarization, Multimodal emotion recognition, Speech Recognition, Video Analytics

TEXT BOOKS:

1. Good fellow, I., Bengio,Y., and Courville, A., Deep Learning, MIT Press, 2016.
2. Josh Patterson, Adam Gibson, Deep Learning: A Practitioner's Approach, OReilly, 2017.

REFERENCES:

1. Bishop, C., M., Pattern Recognition and Machine Learning, Springer, 2006.
2. Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.
3. Golub, G., H., and Van Loan, C.,F., Matrix Computations, JHU Press,2013.
4. Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004.

SEMESTER-II

Course Code	Course Name	Course Category	Credits			
			L	**T**	**P**	**C**
AML 505L	Deep Learning Techniques Lab	C	0	0	3	2

LIST OF PRACTICAL EXPERIMENTS

1. Installation and working on python, Juypter, and its different libraries for deep learning (Tensor Flow, NumPy, Kera, Pandas, Matplotlib, etc.)
2. To implement a Multilayer Perceptron (MLP) using Keras with TensorFlow, and fine-tune neural network hyperparameters for regression problem (house price prediction).
3. To implement a MLP using keras with TensorFlow for classification problem (heart disease predication).
4. To implement a Convolution Neural Network (CNN) for dog/cat classification problem using keras.
5. To Implement a CNN for object detection in the given image.
6. To implement a Recurrent Neural Network (RNN) for predicating time series data.
7. To implement a Long Short-Term Memory (LSTM) for predicating time series data.
8. To implement a Seq2Seq Model for Neural Machine Translation in Keras.
9. To implement an Encoder-Decoder Recurrent neural network model for Neural Machine Translation.
10. To implement a Gated Recurrent Unit (GRU) for time series data predication.
11. To implement use Transfer Learning to explore feature extraction with the VGG16 and ResNet architectures
12. To implement Multimodal emotion recognition using Transformers

SEMESTER-II

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
RM101	Research Methodology and IPR	ES	2	0	0	2

UNIT I

Meaning of research problem, Sources of research problem, Criteria Characteristics of a good research problem, Errors in selecting a research problem, scope, and objectives of research problem.

Approaches of investigation of solutions for research problem, data collection, analysis, interpretation, Necessary instrumentations

UNIT II

Effective literature studies approaches, analysis Plagiarism, Research ethics,

UNIT III

Effective technical writing, how to write report, Paper Developing a Research Proposal, Format of research proposal, a presentation and assessment by a review committee

UNIT IV

Nature of Intellectual Property: Patents, Designs, Trade and Copyright. Process of Patenting and Development: technological research, innovation, patenting, development. International Scenario: International cooperation on Intellectual Property. Procedure for grants of patents, Patenting under PCT.

UNIT V

Patent Rights: Scope of Patent Rights. Licensing and transfer of technology. Patent information and databases. Geographical Indications.

UNIT VI

New Developments in IPR: Administration of Patent System. New developments in IPR; IPR of Biological Systems, Computer Software etc. Traditional knowledge Case Studies, IPR and IITs.

TEXTBOOKS

1. Stuart Melville and Wayne Goddard, "Research methodology: an introduction for science & engineering students'"
2. Wayne Goddard and Stuart Melville, "Research Methodology: An Introduction"

REFERENCES

1. Ranjit Kumar, 2nd Edition, “Research Methodology: A Step by Step Guide for beginners”
2. Halbert, “Resisting Intellectual Property”, Taylor & Francis Ltd ,2007.
3. Mayall, “Industrial Design”, McGraw Hill, 1992.
4. Niebel, “Product Design”, McGraw Hill, 1974.
5. Asimov, “Introduction to Design”, Prentice Hall, 1962.
6. Robert P. Merges, Peter S. Menell, Mark A. Lemley, “ Intellectual Property in New Technological Age”, 2016.
7. T. Ramappa, “Intellectual Property Rights Under WTO”, S. Chand, 2008

SEMESTER-III

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
DSC680	Project Work - Phase I	PR	0	0	26	13

COURSE OBJECTIVES

DSC 501 & DSC 502 is a mandatory two semesters long project work course (spreading over third and fourth semester), culminating to dissertation writing and defending. This course is aimed to prepare our graduate students for a career in the high growth field of DSC 501 & DSC 502. After extensive literature survey, students will learn how to define a real-world problem with its scopes and challenges and then after experimenting with cutting edge artificial intelligence, data science and machine learning techniques learn to propose a solution. Students will learn to compare the performance metrics of their implementation with others' using industry standard software and powerful multicore CPU and GPU. Students will also learn to effectively communicate their research findings by writing and presenting papers and defending their dissertation.

COURSE CONTENT

In project phase I students will choose from a wide range of real-world problems that needs knowledge and expertise of data science and machine learning techniques for solving them. Problems and concepts may be defined based on extensive literature survey of research articles published in highly reputed journals. Significance of proposed problem and the state-of-the art of the problem domain to be explored first. Then students will propose their innovative ideas that mitigate the challenges of the problem. Industry relevant tools may be used for solving the problem and demonstrating the results. Students are required to publish their research findings in reputed journals and conferences. The progress of their projects will be regularly assessed by the designated project guides.

In the second phase of the project work, students will start writing their dissertation. Simultaneously they will work in their project for better solutions and more publications. Students will submit their dissertation at least two weeks in advance to the internal and external examiners before the date of final viva-voce. Successful Défense of the dissertation will be considered as partial requirement for awarding M. Tech degree in Data Science.

COURSE LEARNING OUTCOME

1. Conduct state-of-the-art literature review in identified problem domain that requires Data Analytics.
2. Develop an in-depth understanding in the concept of uncovering business intelligence from large amount of web data mining.
3. Design innovative products and software services by harnessing the power of Data Science in broad application fields ranging from computer vision, internet of things to advanced autonomous systems.
4. Evaluate the proposed solution through extensive performance experiments.

5. Effectively communicate research findings in terms of reports and presentations.
6. Inculcate independent research ability that addresses fundamental problems

SEMESTER-IV

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
DSC681	Project Work - Phase II	PR	0	0	26	13

COURSE OBJECTIVES

DSC 501 & DSC 502 is a mandatory two semesters long project work course (spreading over third and fourth semester), culminating to dissertation writing and defending. This course is aimed to prepare our graduate students for a career in the high growth field of DSC 501 & DSC 502. After extensive literature survey, students will learn how to define a real-world problem with its scopes and challenges and then after experimenting with cutting edge artificial intelligence, data science and machine learning techniques learn to propose a solution. Students will learn to compare the performance metrics of their implementation with others' using industry standard software and powerful multicore CPU and GPU. Students will also learn to effectively communicate their research findings by writing and presenting papers and defending their dissertation.

COURSE CONTENT

In project phase I students will choose from a wide range of real-world problems that needs knowledge and expertise of data science and machine learning techniques for solving them. Problems and concepts may be defined based on extensive literature survey of research articles published in highly reputed journals. Significance of proposed problem and the state-of-the art of the problem domain to be explored first. Then students will propose their innovative ideas that mitigate the challenges of the problem. Industry relevant tools may be used for solving the problem and demonstrating the results. Students are required to publish their research findings in reputed journals and conferences. The progress of their projects will be regularly assessed by the designated project guides.

In the second phase of the project work, students will start writing their dissertation. Simultaneously they will work in their project for better solutions and more publications. Students will submit their dissertation at least two weeks in advance to the internal and external examiners before the date of final viva-voce. Successful Défense of the dissertation will be considered as partial requirement for awarding M. Tech degree in Artificial Intelligence and Machine Learning.

COURSE LEARNING OUTCOME

1. Conduct state-of-the-art literature review in identified problem domain that requires Data Analytics.
2. Develop an in-depth understanding in the concept of uncovering business intelligence from large amount of web data mining.
3. Design innovative products and software services by harnessing the power of Data Science in broad application fields ranging from computer vision, internet of things to advanced autonomous systems.
4. Evaluate the proposed solution through extensive performance experiments.
5. Effectively communicate research findings in terms of reports and presentations.
6. Inculcate independent research ability that addresses fundamental problems.

Electives

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML557	Introduction to High Performance Computing	E	3	0	0	3

UNIT I: INTRODUCTION TO PIPELINING AND INSTRUCTION LEVEL PARALLELISM

Introduction to pipelining – Types of pipelining – Hazards in pipelining - Introduction to instruction level parallelism (ILP) – Challenges in ILP - Basic Compiler Techniques for exposing ILP - Reducing Branch costs with prediction - Overcoming Data hazards with Dynamic scheduling - Hardware-based speculation - Exploiting ILP using multiple issue and static scheduling - Exploiting ILP using dynamic scheduling, multiple issue and speculation - Tomasulo's approach, VLIW approach for multi-issue.

UNIT II: MULTI PROCESSORS AND THREAD – LEVEL PARALLELISM

Introduction to multi processors and thread level parallelism - Characteristics of application domain - Systematic shared memory architecture - Distributed shared – memory architecture – Synchronization – Multithreading - Multithreading-fined grained and coarse grained, superscalar and super pipelining, hyper threading. Vector architectures; organizations and performance tuning; GPU architecture and internal organization, Elementary concepts in CUDA programming

UNIT III: MEMORY HIERARCHY

Introduction to cache performance - Cache Optimizations - Virtual memory - Advanced optimizations of Cache performance - Memory technology and optimizations - Protection: Virtual memory and virtual machines - multi-banked caches, critical word first, early restart approaches, hardware pre-fetching, write buffer merging.

UNIT IV: PARALLEL PROGRAMMING

Introduction to parallel computing platforms; (Open MP, MPI, Open CL, Open ACC) with performance improvement analysis done using real-life AI and ML applications

UNIT V: INTER CONNECTION AND NETWORKS

Introduction to inter connection networks and clusters - interconnection network media - practical issues in interconnecting networks- examples - clusters - designing a cluster – System on Chip (SoC) Interconnects – Network on Chip (NOC).

TEXTBOOKS

1. Sterling, Thomas, Maciej Brodowicz, and Matthew Anderson. "High performance computing: modern systems and practices", Morgan Kaufmann, 2017.
2. Hennessy, John L., and David A. Patterson. "Computer architecture: a quantitative approach", Elsevier, 2011.

REFERENCES

1. Wang, Endong, Qing Zhang, Bo Shen, Guangyong Zhang, Xiaowei Lu, Qing Wu, and Yajuan Wang. "High-performance computing on the Intel Xeon Phi." Springer 5, 2014.
2. Sanders, Jason, and Edward Kandrot. “CUDA by example: an introduction to general-purpose GPU programming”, Addison-Wesley Professional, 2010.
3. Chandra, Rohit, Leo Dagum, David Kohr, Ramesh Menon, Dror Maydan, and Jeff McDonald. Parallel programming in Open MP. Morgan kaufmann, 2001.
4. Kaeli, David R., Perhaad Mistry, Dana Schaa, and Dong Ping Zhang. Heterogeneous computing with Open CL 2.0. Morgan Kaufmann, 2015.
5. Farber, Rob. Parallel programming with Open ACC. Newnes, 2016.

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML562	Statistical Modelling for Computer Science	E	3	0	0	3

UNIT I

Introduction to Data- Definition of data; Different kinds of variables; Sampling principles and strategies; Difference between observation and experiment; Examining numerical data; Considering categorical data; Case studies and examples; Analysis and Representation of data, different kind of existing software tools Example: Python, Pandas, scipy. stats, numpy, matplotlib etc.Line Plot, Bar chart, Histogram plot, Box and Whisker Plot, Scatter Plot etc.

UNIT II

Probability, Distributions of random variables, Foundations of random variables- Defining probability, Conditional probability, Sampling from small population; Random variables, Continuous distributions. Normal distribution; Binomial distribution; Negative binomial distribution; Poisson distribution; Central tendencies, Law of large numbers, Central limit theorem.

UNIT III

Foundations of Inference, Inference for categorical data, Inference for numerical data-Point estimates and sampling variability, Confidence intervals for a proportion, Hypothesis testing, Critical values, Covariance and correlation, Significance tests, Effect size. Inference for a single proportion, Difference of two proportions; Testing for goodness of fit using chi-square. One-sample means with the t-distribution, Paired data, Difference of two means, Power calculations for a difference of means, comparing many means with Analysis of variance (ANOVA).

UNIT IV

Introduction to linear regression-Fitting a line, residuals, and correlation; Least square regression, Types of outliers in linear regression; Inference for linear regression.

UNIT V

Multiple and logistic regression- Introduction to multiple regression, Model selection, Checking model conditions using graphs, Multiple regression case studies, Introduction to logistic regression.

TEXT BOOKS

1. Open Intro Statistics - David Diez, Christopher Barr, and Mine Çetinkaya-Rundel.
2. Introduction to Probability - by Dimitri Bertsekas.

REFERENCES

1. Statistical Methods and Machine learning - Jason Brownlee.

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML563	Fuzzy Logic and its Applications	E	3	0	0	3

UNIT I

Introduction and Motivation: History of fuzzy theory; Limitations of classical logic; Introduction to fuzzy set theory in contrast with classical set theory; Introduction to fuzzy logic.

UNIT II

Fuzzy Logical Operators -Fundamental concepts of fuzzy theory: sets, relations, and logic operators Conjunction, Disjunction, Negation.

UNIT III

Fuzzy Inference Systems - Approximate reasoning, fuzzy inference, possibility theory. Separation from probability, Generalized Modus Ponens, Generalized Modus Tollens, Approximate Reasoning.

UNIT IV

Fuzzy Control Systems- The Mamdani Model, The Sugeno Model, Defuzzification methods, Families of implication operators, Hierarchy of implication operators.

UNIT V

Applications - Fuzzy Classification Algorithms, Fuzzy Logic and Neural Networks, Fuzzy Graph Theory, Fuzzy Character Recognition, Fuzzy Expert Systems, Fuzzy Markov Chains, Fuzzy Ranking Algorithms, Fuzzy Facial Recognition, Fuzzy Image Stabilization, Fuzzy Logic in Computer Games**.**

TEXT BOOKS

1. Bede - Mathematics of Fuzzy Sets and Fuzzy Logic
2. Fuzzy Logic: Intelligence, Control, and Information, J. Yen, R. Langari, Prentice Hall, 1999

REFERENCES

1. Chen and Pham - Introduction to Fuzzy Sets, Fuzzy Logic and Fuzzy Control Systems
2. Fuzzy Systems Toolbox-student edition for use with MATLAB, by Mark Beale and Howard Demuth, PWS Publishing Company, 1996 2)
3. Fuzzy Set Theory –Foundations and Applications, George J. Klir , Ute St. Clair, and Bo Yuan, Prentice Hall PTR, 1997 3)
4. Fuzzy Engineering, Bart Kosko, Prentice Hall, 1997 4) Fuzzy Logic with Engineering Applications, by Timothy J. Ross, McGraw Hill, 1995

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML553	Information Retrieval	E	3	0	0	3

UNIT I

Introduction: Overview of Information Retrieval, Architecture of a Search Engine, Acquiring Data : Crawling the Web, Document Conversion, Storing the Documents, Detecting Duplicates, Noise Detection and Removal.Processing Text: Text Statistics, Document Parsing, Tokenizing, Stopping, Stemming, Phrases, Document Structure, Link Extraction, More detail on Page Rank, Feature Extraction and Named Entity Recognition, Internationalization.

UNIT II

Ranking with Indexes Abstract Model of Ranking, Inverted indexes, Map Reduce, Query Processing: Document-at-a-time evaluation, Term-at-a-time evaluation, Optimization techniques, Structured queries, Distributed evaluation, Caching**.**

Queries and Interfaces: Information Needs and Queries, Query Transformation and Refinement: Stopping and Stemming Revisited, Spell Checking and Query Suggestions, Query Expansion, Relevance Feedback, Context and Personalization. Displaying the Results: Result Pages and Snippets, Advertising and Search, Clustering the Results; Translation; User Behavior Analysis.

UNIT III

Retrieval Models: Overview of Retrieval Models; Boolean Retrieval, The Vector Space Model. Probabilistic Models: Information Retrieval as Classification, The BM25 Ranking Algorithm. Ranking based on Language Models: Query Likelihood Ranking, Relevance Models and Pseudo-Relevance Feedback. Complex Queries and Combining Evidence: The Inference Network Model, The Galago Query Language. Models for Web search, Machine Learning and Information Retrieval: Learning to Rank (Le ToR), Topic Models

UNIT IV

Evaluating Search Engines: Test collections, Query logs, Effectiveness Metrics: Recall and Precision, Averaging and interpolation, focusing on the top documents. Training, Testing, and Statistics: Significance tests, setting parameter values

Classification and Clustering

UNIT V

Social Search: Networks of People and Search Engines: User tagging, searching within Communities, Filtering and recommending, Meta search. Beyond Bag of Words: Feature-Based Retrieval Models, Term Dependence Models, Question Answering, Pictures, Pictures of Words, etc., XML Retrieval, Dimensionality Reduction and LSI

TEXTBOOKS

1. Introduction to Information Retrieval. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schuetze, Cambridge University Press, 2007.

REFERENCES

1. Search Engines: Information Retrieval in Practice. Bruce Croft, Donald Metzler, and Trevor Strohman, Pearson Education, 2009.
2. Modern Information Retrieval. Baeza-Yates Ricardo and BerthierRibeiro-Neto. 2nd edition, Addison-Wesley, 2011.

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
AML554	Pattern Recognition	E	3	0	0	3

UNIT I: PATTERN CLASSIFIER

Overview of Pattern recognition – Basics of Probability and Statistics, Linear Algebra, Linear Transformations, Components of Pattern Recognition System, Learning and adaptation Discriminant functions – Supervised learning – Parametric estimation – Maximum Likelihood Estimation – Bayesian parameter Estimation – Problems with Bayes approach– Pattern classification by distance functions – Minimum distance pattern classifier.

UNIT II: CLUSTERING

Clustering for unsupervised learning and classification–Clustering concept – C Means algorithm – Hierarchical clustering – Graph theoretic approach to pattern Clustering – Validity of Clusters.

UNIT III: FEATURE EXTRACTION AND STRUCTURAL PATTERN RECOGNITION

Feature Extraction and Feature Selection: Feature extraction – discrete cosine and sine transform, Discrete Fourier transform, Principal Component analysis, Kernel Principal Component Analysis. Feature selection – class separability measures, Feature Selection Algorithms - Branch and bound algorithm, sequential forward / backward selection algorithms. Principle component analysis, Independent component analysis, Linear discriminant analysis, Feature selection through functional approximation – Elements of formal grammars, Syntactic description – Stochastic grammars – Structural Representation.

UNIT – IV: HIDDEN MARKOV MODELS AND SUPPORT VECTOR MACHINE

State Machines – Hidden Markov Models – Training – Classification – Support vector Machine – Feature Selection.

UNIT V: RECENT ADVANCES

Fuzzy logic – Fuzzy Pattern Classifiers – Pattern Classification using Genetic Algorithms – Case Study Using Fuzzy Pattern Classifiers and Perception

TEXTBOOKS AND REFERENCE

1. Andrew Webb, “Stastical Pattern Recognition”, Arnold publishers, London,1999
2. C.M.Bishop, “Pattern Recognition and Machine Learning”, Springer, 2006.
3. M. Narasimha Murthy and V. Susheela Devi, “Pattern Recognition”, Springer 2011.
4. Menahem Friedman, Abraham Kandel, “Introduction to Pattern Recognition Statistical, Structural, Neural and Fuzzy Logic Approaches”, World Scientific publishing Co. Ltd, 2000.
5. Robert J.Schalkoff, “Pattern Recognition Statistical, Structural and Neural Approaches”, John Wiley & Sons Inc., New York, 1992.
6. R.O.Duda, P.E.Hart and D.G.Stork, “Pattern Classification”, John Wiley, 2001
7. S.Theodoridis and K.Koutroumbas, “Pattern Recognition”, 4th Ed., Academic Press. 2009.

Course Code	Course Name	Course Category	CREDITS			
			L	T	P	C
AML552	Knowledge Engineering and Expert Systems	E	3	0	0	3

UNIT I

The nature of Expert Systems Types of applications of Expert Systems relationship of Expert Systems to Artificial Intelligence and to Knowledge-Based Systems. The nature of expertise Distinguishing features of Expert Systems. Benefits of using an Expert System Choosing an application.

UNIT II

Theoretical Foundations What an expert system is; how it works and how it is built. Basic forms of inference: abduction; deduction; induction.

UNIT III

The representation and manipulation of knowledge in a computer; Rule-based representations (with backward and forward reasoning); logic-based representations (with resolution refutation); taxonomies; meronomies; frames (with inheritance and exceptions); semantic and partitioned nets (query handling).

UNIT IV

Basic components of an expert system; Generation of explanations; Handling of uncertainties; Truth Maintenance Systems; Expert System Architectures; An analysis of some classic expert systems; Limitations of first generation expert systems; Deep expert systems; Co-operating expert systems and the blackboard model.

UNIT V

Building Expert Systems Methodologies for building expert systems: knowledge acquisition and elicitation; formalisation; representation and evaluation. Knowledge Engineering tools, Case Study.

TEXTBOOKS

1. P Jackson, Introduction to Expert Systems, Addison Wesley, 1990 (2nd Edition).

REFERENCES

1. Elaine Rich, Kevin Knight, Artificial Intelligence, McGraw-Hill, Inc, 1991 (2nd Edition).
2. Jackson. Jean-Louis Lauriere, Problem Solving and Artificial Intelligence, Prentice Hall, 1990.

SUBJECT CODE	SUBJECT TITLE	CORE/ ELECTIVE	CREDITS			
			L	T	P	C
DSC 550	Time Series Analysis and Forecasting	E	3	0	0	3

UNIT 1

Stochastic processes, Autocovariance and autocorrelation functions, Partial autocorrelation function, White noise process, Estimation of mean autocovariances and autocorrelations, Moving average and autoregressive representations of time series processes, Difference equations, Log operators

UNIT 2

Stationary time series models: Autoregressive processes, Moving average processes, Autoregressive moving average processes

Nonstationary time series models: Autoregressive integrated moving average models, Forecasting

UNIT 3

Model identification, Parameter estimation, Diagnostic checking, Model selection, Seasonal time series models, Testing for unit root

UNIT 4

Vector time series models: Vector autoregressive moving average process, Nonstationary vector autoregressive moving average process, Identification of the models, Model fitting and forecasting

UNIT 5

The Kalman Filter, Generalized method and moments, Modeling time series with changes in regime

TEXT BOOKS

1. Hamilton, James Douglas. *Time series analysis*. Princeton university press, 2020.
2. Wei, William W.S. *Time Series Analysis: Univariate and Multivariate Methods*, Pearson education, 2006.

REFERENCES:

1. Hyndman, R.J., & Athanasopoulos, G. *Forecasting: principles and practice*, OTexts: Melbourne, Australia, 2018.

SUBJECT CODE	SUBJECT TITLE	CORE/ ELECTIVE	CREDITS			
			L	T	P	C
DSC551	Complex Network Analysis	E	3	0	0	3

UNIT I: **INTRODUCTION**
Definition and representation of complex networks; Examples of complex networks; Directed and undirected networks; Weighted networks; Signed networks; Heterogeneous networks; Dynamic networks; Measures and metrics: Centrality, Geodesic distance, Transitivity, Reciprocity, Similarity, Homophily; Large scale structure of complex networks: Components, Shortest paths and small world effect, Degree distribution, Power laws and scale free networks, Clustering coefficient, Assortative mixing; Important problems in complex network analysis

UNIT II:
Important problems in complex network analysis (1) - Link prediction. Definition of link prediction; Similarity based methods; Probabilistic and maximum likelihood models; Link prediction using dimensionality reduction; Deep learning based methods; Temporal link prediction; Link prediction in bipartite networks, Link prediction in heterogeneous networks; Evaluation; Application of link prediction

UNIT III:
Important problems in complex network analysis (2) - Community detection
Definition of communities; Traditional algorithms: Hierarchical clustering, Girvan-Newman algorithms, Spectral clustering, Graph partitioning; Modularity based algorithms; Dynamic algorithms; Algorithms for overlapping communities; Detection of dynamic communities; Deep learning methods for community detection; Evaluation; Application of community detection

UNIT IV:
Important problems in complex network analysis (3) - Node classification and Motif analysis.
Node classification: Problem definition and applications; Iterative classification methods; Label propagation method; Graph regularization method; Evaluation
Motif analysis: Definition of network motifs; Triangle counting and enumeration algorithms; Motifs in temporal networks; Applications of network motifs

UNIT V:
Network representation learning Graph kernels; Spectral methods; Node embedding: Random walk embedding; Graph neural networks

TEXT BOOKS

1. Easley, David, and Jon Kleinberg. *Networks, crowds, and markets*. Cambridge: Cambridge university press, 2010.
2. Newman, M.E.J, *Networks: An Introduction.* Oxford, UK: Oxford University Press, 2013.

REFERENCES:

1. Hamilton, William L. "Graph representation learning." Synthesis Lectures on Artifical Intelligence and Machine Learning 14.3 (2020): 1-159.

2. Kumar, Ajay, et al. "Link prediction techniques, applications, and performance: A survey." *Physica A: Statistical Mechanics and its Applications* 553 (2020): 124289.
3. Fortunato, Santo. "Community detection in graphs." *Physics reports* 486.3-5 (2010): 75-174.
4. Liu, Fanzhen, et al. "Deep learning for community detection: progress, challenges and opportunities." *arXiv preprint arXiv:2005.08225* (2020).
5. Bhagat, Smriti, Graham Cormode, and S. Muthukrishnan. "Node classification in social networks." *Social network data analytics*. Springer, Boston, MA, 2011. 115-148.
6. Milo, Ron, et al. "Network motifs: simple building blocks of complex networks." *Science* 298.5594 (2002): 824-827.
7. Paranjape, Ashwin, Austin R. Benson, and Jure Leskovec. "Motifs in temporal networks." *Proceedings of the tenth ACM international conference on web search and data mining*. 2017.

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
DSC 552	Recommender Systems	E	3	0	0	3

UNIT 1: INTRODUCTION

Introduction and basic taxonomy of recommender systems (RS). Traditional and non-personalized RS. Overview of data mining methods for recommender systems (similarity measures, classification, ensembles of classifiers, clustering, SVMs, dimensionality reduction).

UNIT 2: CONTENT-BASED RECOMMENDER SYSTEMS

Domain-specific challenges in recommender systems, Content representation and content similarity, similarity-based retrieval, Feature selection, Item representation Methods for learning user profiles, , text classification methods, advantages and limitations of content-based filtering.

UNIT 3: COLLABORATIVE FILTERING (CF)

Mathematical optimization, Baseline predictor through least squares, Regularization and overfitting, Matrix factorization models and dimensionality reduction, Matrix Decomposition, Latent factor models, The Netflix data challenge, Tensors and their applications

UNIT 4: NEIGHBORHOOD-BASED RS

Formal definition, components of neighborhood methods: Rating normalization, Similarity weight computation, neighborhood selection, dimensionality reduction, graph-based methods, Hybrid recommender systems.

UNIT 5: PERFORMANCE EVALUATION OF RS

General goals of evaluation design: accuracy, coverage, confidence and trust, novelty, serendipity, diversity and scalability, design issues in offline recommendation evaluation, accuracy metrics in offline evaluation: RMSE, MAE, Limitations of evaluation metrics.

TEXT BOOKS

1. C.C. Aggarwal, Recommender Systems: The Textbook, Springer, 2016.
2. F. Ricci, L Rokach, B. Shapira and P.B. Kantor, Recommender systems handbook, Springer 2010.

REFERENCES:

1. Falk, Kim. Practical recommender systems. Simon and Schuster, 2019.

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
DSC 553	Big Data Security and Privacy	E	3	0	0	3

UNIT 1: ENCRYPTION TECHNIQUES

Cipher models: stream cipher and block cipher design principles - Block cipher modes of operation - DES - DDES and TDES - strength of DES - AES – strength of AES – RC4 – Symmetric key distribution. Asymmetric key cipher: RSA cryptosystems & cryptanalysis – ElGamal cryptosystem – Elliptic curve arithmetic and cryptography – Asymmetric key distribution and management.

UNIT 2: MESSAGE AUTHENTICATION

Authentication requirements – Authentication functions – Message authentication codes (MAC) – Hash functions – Security of hash functions and MACS – Digital Signature – certificate authority – Access Control Mechanism – ABAC - CapBAC.

UNIT 3: INTRODUCTION TO SECURITY ANALYTICS

Introduction to Security Analytics – Techniques in Analytics – Analysis in everyday life – Challenges in Intrusion and Incident Identification – Simulation and Security Process, Analytical Softwares and tools, Malware Analysis – static and dynamic analysis - Security Intelligence – Security Breaches.

UNIT 4: APPLICATIONS OF SECURITY ANALYTICS

Access Analytics – Analysis of Log file -Security analysis with text mining –Machine Learning and data mining applications for security: Intrusion detection and network anomaly detection. Big data analytics for security: Analyzing DDOS – Distributed Denial of Service attack: counter based method, and access pattern-based method – Machine learning for Ransom ware detection and prevention.

UNIT 5: SECURITY AND PRIVACY FOR BIG DATA

Anomaly detection in cloud big database metric - Anonymizing and pseudonymizing data - Differential Privacy - Differential Privacy method - Homomorphic encryption - Secure Multiparty Computation - Data Protection Law for Big Data - General Data Protection Regulation (GDPR) - Personal data and Big Data applications - Digital identity protection using blockchain for academic qualification certificates - Model Poisoning Attack - Defense Against Model Poisoning Attack.

TEXT BOOKS

1. William Stallings, “Crpyptography and Network security: Principles and Practices”, Pearson/PHI, 5th Edition, 2010.
2. Douglas R. Stinson,“Cryptography Theory and Practice ”, Chapman & Hall/CRC, 3rd Edition, 2006.

REFERENCES:

1. Mark Talabis, Robert McPherson, I Miyamoto and Jason Martin, “Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data”, Syngress Media, U.S., 2014.
2. Siddhartha Bhattacharyya, “Frontiers in Computational Intelligence” Volume 3, De Gruyter, 2017.

Course Code	Course Name	Course Category	Credits			
			L	T	P	C
DSC 554	Spatial Data Science and Visualization	E	3	0	0	3

UNIT 1: UNDERSTANDING SPATIAL DATA SCIENCE

Spatial Data Science and Applications, Understanding of Spatial Data Science, Introduction to spatial data science, what is spatial special? spatial autocorrelation, map projection, uncertainty, and modifiable areal unit problem. Solution Structures of Spatial Data Science Problems, Four Disciplines for Spatial Data Science and Applications, Open-Source Software for data science and visualization, Python for Data Analytics, Spatial Data Science Problems, Spatial Data vs. Spatial Big Data.

UNIT 2: GEOGRAPHIC INFORMATION SYSTEM (GIS)

What is GIS?, Five Layers of GIS, Spatial Reference Framework, Spatial Data Models, Spatial Data Acquisition Systems, Spatial Data Analysis, Geo-visualization and Information Delivery, Spatial DBMS, and Big Data Systems, Database Management System (DBMS), Spatial DBMS, Spatial Big Data System, Choropleth Mapping- Principles, User-defined choropleths.

UNIT 3: SPATIAL DATA ANALYTICS

Spatial Data Analytics, Proximity and Accessibility, Spatial Autocorrelation, Spatial Interpolation, Spatial Categorization, Minimum Distance to Mean (MDM) and Decision Tree (DT), clustering algorithms such as K-Means and DBSCAN with real-world examples, Network Analysis, Spatial Regression

UNIT 4: PRACTICAL APPLICATIONS OF SPATIAL DATA SCIENCE

Practical Applications of Spatial Data Science, storing data in MongoDB, fetching data from MongoDb, Data-science life-cycle, Data-preparation, pipelines, model-building, classification and regression use cases, clustering use cases.

UNIT 5: DATA VISUALIZATION

What is data visualization, use of data visualization, types of data visualization, exploratory data analysis, numerical data analysis, categorical data analysis, and visualization chats using Python, Spatial Feature Engineering, Feature Engineering Using Map Matching, Feature Engineering using Map Synthesis

TEXT BOOKS

1. Crickard, P., “Mastering geospatial analysis with Python: explore GIS processing and learn to work with GeoDjango, CARTOframes and MapboxGL-Jupyter” Packt Publishing, 2018.
2. Comber, Lex, and Chris Brunsdon, “Geographical data science and spatial data analysis: An introduction in R,” Sage Publication, 2020.

REFERENCES:

1. Lawhead, J., “Learning geospatial analysis with Python,” Packt Publishing Ltd, 2015.
2. Hassan, A. and Vijayaraghavan, J., “Geospatial Data Science Quick Start Guide: Effective techniques for performing smarter geospatial analysis using location intelligence,” Packt Publishing Ltd, 2019.