**POSTGRADUATE INSTITUTE OF SCIENCE**
UNIVERSITY OF PERADENIYA

Co-funded by the Erasmus+ Programme of the European Union

DS&AI

# M.Sc. Programme in Data Science and Artificial Intelligence

## 1. INTRODUCTION

Data Science is an emerging interdisciplinary field of Statistics and Computer Science for which the foundational topics are Data Manipulation, Data Analysis with Statistics, Machine Learning, Artificial Intelligence, High Performance Computing, Data Communication with Information Visualization, and Data at Scale (Working with Big Data). Data Science is also an integral part of research in many fields, such as machine translation, speech recognition, robotics, search engines, digital economy in addition to business intelligence, operational intelligence, biological sciences, medical informatics, health care, social sciences and the humanities.

With the rapid development in computational capabilities, Machine Learning and Artificial Intelligence has become one of the key areas in computing. Machine Learning and Artificial Intelligence play a prominent role in Data Science and Business Analytics. Furthermore, there is an explosive growth of data due to the rapid development of the Internet and digital technology. Moreover, data storage capacity and processing speed have increased dramatically and cost effectively. Thus, many industries deal with large volumes of data and this has resulted in a great demand and a good opportunity for graduates with statistical skills, comprehensive knowledge in Machine Learning and Artificial Intelligence and competence in related technologies. While data science and artificial intelligence based technologies are gaining traction and becoming strategic priorities many countries in Asian region are still falling behind regional leaders in the technological market, such as Singapore and Malaysia, while the region as a whole is far behind China. There is a shortage of qualified graduates whose skills span these areas, which has been identified in studies done in public and private companies in different Asian countries including Sri Lanka. While it is of utmost importance to train personnel to undertake responsibilities related to the real-world applications, the Sri Lankan education institutions have not paid due attention in the recent past to this rapidly growing area of science.

The proposed curriculum has been developed as a collaboration between European partners (Leiden University, Netherlands; Athens University of Economics and Business, Greece; University of Minho, Portugal; Skybridege Partners, Greece) and Asian partners (Asian Institute of Technology, KhonKaen University and Walailak University, Thailand; InstitutTeknologi Bandung, UniversitasSumatrae Utara and Syiah Kuala University, Indonasia; University of Peradeniya and University of Sri Jayawardenapura, Sri Lanka) and has been co-funded by the Erasmus+ programme of the European Union. Together, the eight partners from Asia, working with the four European partners, have been uniquely positioned to develop a world class Master's program in Data Science and Artificial Intelligence that is truly national in scope.

A systematic requirement analysis was conducted considering existing programmes in Data Science and a stakeholder analysis in each Asian partner country including students, academics and industry in each country. Each individual course curriculum has been developed by academic experts in the relevant field and has been reviewed by at least two more independent teams following a strict quality assurance process.

Thus, the proposed M.Sc. Programme in Data Science and Artificial Intelligence has the assuarance of been designed to cater for the current demands for postgraduate training and research in the fields of Computational Statistics, Artificial Intelligence, Machine Learning and big-data.

## 2. OBJECTIVES OF THE PROGRAMME

The objectives of the programme are to provide graduates an adequate coverage of skills in Data Science and Artificial Intelligence with the focus of applications in the relevant fields and methods of presentation and interpretation of results. At the completion of the course, the candidate will be competent as a data scientist in a research institute, planning institute or a government institute or as an information and data scientist in industries such as Finance, Banking, Retail, Telecommunication, Media and Insurance with complementary skills in statistics andcomputerscience.

## 3. PROGRAMME ELIGIBILITY

Applicants must possess a science-based degree (e.g. Physical Science/Engineering related degree) with Statistics and Computer Science (with at least 9 credits from each subject) or an equivalent qualification acceptable to the Postgraduate Institute of Science. Depending on the courses successfully completed at the degree level and on the recommendation of the programme advisor a candidate may be exempted from some of the preliminary courses. Graduates, who lack knowledge in Mathematics required for successful following of this programme, are expected to follow a course in fundamentals of Mathematics (None credit courses).

## 4. PROGRAMME FEE

Students registered for the M.Sc. degree by course work shall pay the Programme fee in two *(1/2 at the registration and the balance at the end of the second semester)* or three *(1/3rd at the registration, another 1/3rd at the end of the first semester and the balance at the end of the second semester)* instalments. Anadditional payment should be made at the end of the first year to continue for the M.Sc. degree by research.Other payments including registration fee, medical fee, library subscription, examination fee and deposits (science and library) should be paid according to the procedure stipulated by the PGIS.

## 5. THE PROGRAMME STRUCTURE AND DURATION

This programme consists of three options for completion.

### 5.1 Master of Data Science Degree Programme(SLQF Level 9)

The Master of Data Science and Artificial Intelligence degree (M Data Science and AI) can be obtained by completing course work only (without conducting any research project).

Course work, comprising of theory courses, and laboratory and/or fieldwork, shall be conducted over a period of two semesters of 15 weeks each. The total duration of the degree, including examinations, shall be about 12 months. Satisfactory completion of a minimum of 30 credits of course work with a GPA of not less than 3.00 is required for the successful completion of the degree - SLQF Level 9 (Students who do not satisfy the above criteria but obtain a GPA in the range 2.75 to 2.99 for course work of 25 credits are eligible for the Postgraduate Diploma in Data Science and Artificial Intelligence - SLQF Level 8, and those who obtain a GPA in the range 2.75 to 2.99 for course work of 20 credits are eligible for Postgraduate Certificate - SLQF Level 7).

**5.2Master of Science in Data Science and Artificial Intelligence Degree Programme(SLQF Level 10)**

In addition to Master of Data Science and Artificial Intelligence degree (5.1), the Master of Science in Data Science and Artificial Intelligence Degree (M.Sc. Data Science and Artificial Intelligence) requires a research project. The duration of the entire programme shall be 24 months inclusive of 5.1. Completion of all the requirements of 5.1 with a GPA of not less than 3.00 is a prerequisite for the M.Sc. Data Science and Artificial Intelligence. The research project for this degree should be conducted on full-time basis, and completed during the second year. The research component is allocated 30 credits, totalling 60 credits for the entire programme. After successful completion of the research project, the student shall be eligible for the award of the M.Sc. in Data Science and Artificial Intelligence - SLQF Level 10 (Students who do not complete the research project within the stipulated time period shall be awarded the Master of Data Science and Artificial Intelligence - SLQF Level 9).

**5.3 Extension of the programme for M.Phil. (SLQF Level 11) or Ph.D. (SLQF Level 12)**

After conducting research for a period of six months in the M.Sc. Data Science and Artificial Intelligence students who have demonstrated exceptional progress may apply for upgrading the degree status to M.Phil. The student should continue the research project and any additional research work/assignments recommended by the PGIS for a total of two years (60 credits of research) to qualify for the award of the M.Phil. degree (SLQF Level 11).

During the second year of research, students who have demonstrated exceptional and continuous progress may apply for upgrading the degree status from M.Phil. to Ph.D. The student should continue the research project and any additional research work/assignments recommended by the PGIS for another year on full-time basis (additional 30 credits) to qualify for the award of the Ph.D. degree (SLQF Level 12).

## Programme Summary

| Course Code | Course | Cr. Val | Notional Hrs. | L | P | Ind. Learning | Status (Compulsory /Optional) | Existing/ New |
|---|---|---|---|---|---|---|---|---|
| **Preliminary Courses** | | | | | | | | |
| DS 4101 | Mathematics† | 2 | 100 | 30 | | 70 | Compulsory | Existing SC 401 |
| DS 4102 | Statistical methods* | 3 | 150 | 50 | 50 | 50 | Compulsory | Existing SC 403 |
| **Semester I** | | | | | | | | |
| DS 5101 | Computer Programming for Data Science and AI* | 2 | 100 | | 50 | 50 | Compulsory | New |
| DS 5102 | Data Modelling & Management* | 3 | 150 | 50 | 50 | 50 | Compulsory | New |
| DS 5103 | Regression Analysis* | 3 | 150 | 100 | | 50 | Compulsory | Existing SC 504 |
| DS 5104 | Machine Learning* | 3 | 150 | 50 | 50 | 50 | Compulsory | New |
| DS 5105 | Multivariate Methods I* | 2 | 100 | 30 | | 70 | Compulsory | Existing SC 506 |
| DS 5106 | Multicriteria Optimization & Decision Analysis | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5107 | Distributed Systems | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5108 | Software Development & Project Management for DS& AI | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5109 | Nature Inspired Computing | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5110 | Statistical Computing and Simulation Techniques | 3 | 150 | 50 | 50 | 50 | Optional | Existing SC583 |
| **Semester II** | | | | | | | | |
| DS 5215 | Business Intelligence & Analytics* | 3 | 150 | 50 | 50 | 50 | Compulsory | New |
| DS 5216 | Artificial Intelligence* | 3 | 150 | 100 | | 50 | Compulsory | New |
| DS 5217 | Computational Linguistics | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5218 | Computer Vision | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5219 | Multivariate Methods II | 2 | 100 | 30 | | 70 | Optional | Existing SC519 |
| DS 5220 | Spatio Temporal Data Analysis | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5221 | Knowledge Representation | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5222 | Recent Trends in Machine Learning | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5223 | Human Computer Interaction & Information Visualization | 3 | 150 | 50 | 50 | 50 | Optional | New |
| DS 5224 | Special Topics in Data Science** | 2 | 100 | 30 | | 70 | Optional | Existing SC 587 |
| DS 5225 | Industrial Training* | 5 | 500 | | | 500 | Compulsory | New |
| DS 5299 | Independent Study*[1] | 5 | 500 | | | 500 | Compulsory | Existing SC 599 |
| **Semester III and IV** | | | | | | | | |
| DS 6399 | Research Project*[2] | 30 | 3000 | | | 3000 | Compulsory for M.Sc. (Research) | Existing SC 699 |

Preliminary courses are not considered in the computation of the GPA

†Compulsory for those who do not have a Mathematics background

*Compulsory Courses

*[1] Compulsory for M.Sc. (Course work)

*[2] Compulsory for M.Sc. (Research)

**Special Topics will be notified to students in each year.

## 6. PROGRAMME CONTENTS

**Course Code** :DS 4101
**Course Title** : Mathematics
**No. of Credits** :2
**Pre-requisites** :None
**Compulsory/Optional** :Compulsory for students who have not followed Mathematics as a subject for their bachelors degree.

**Aim(s):** To give students an introduction to mathematical concepts essential in data manipulation which includes calculus, matrix algebra and optimization.

**Intended Learning Outcomes:**
At the completion of the course, students will be able to
- evaluate definite and indefinite integrals.
- calculate the area under a curve.
- perform integrations by substitution and partial differentiation.
- evaluatemulti dimensional data using matrix algebra.

**Time Allocation (Hours):** **Lectures** 30 hrs

**Course content/Course description:**
Number Systems, Inequalities, Elements of Set Theory. Coordinate Geometry: Lines, Circles and Parabolas. Calculus: Limits and Derivatives. Maxima, Minima and Inflextion points. Indefinite Integral, Definite Integral, Evaluation of area between curves, Integration by substitution, Partial differentiation. Matrix Algebra : Matrices, determinants and inverse of a matrix. Matrix approach to solution of system of equations. Rank of a matrix, eigen values, Introduction to numerical optimization.

**Recommended Texts:**
1. Hamming R.W. *Methods of Mathematics Applied to Calculus, Probability, and Statistics*, Dover Publications, 2004.
2. Waner S., Costenobel S., Applied Calculus, $6^{th}$ Edition, Cengage Learning, 2013.

| Assessment | Percentage Mark |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS4102
**Course Title** :Statistical Methods
**No. of Credits** :3
**Pre-requisites** :None
**Compulsory/Optional :** Compulsory

**Aim(s):** To introduce basic probability theory so that students can apply the theoretical knowledge in data analysis.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- identify the sample space and events of an experiment.
- calculate probabilities, conditional probabilities, expectation and variance of random variable.
- derive marginal densities and cumulative densities from a given probability distribution.
- calculate probabilities using Bayes' Rule.
- apply Central Limit Theorem to sampling distributions.
- carry out simple and composite hypothesis tests on mean, variance and proportion.
- estimate parameters using maximum Likelihood.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs

**Course content/Course description:**
Probability: Properties, conditional probability, independence. Discrete random variables: Probability mass functions and cumulative distributions. Some common discrete distributions. Continuous random variables: Marginal and conditional distributions, Bayes' Rule. Expectations and Central Limit Theorem. Sampling from the Normal distribution. Point and Interval estimation. Test of Hypotheses: Simple and composite hypothesis. Maximum likelihood estimation. Generalized Likelihood Ratio Tests. Tests on means and variances.

**Recommended Texts:**

1. Hogg R.V., McKean J., Craig A.T., *Introduction to Mathematical Statistics*, 7th Edition, Pearson, 2012.
2. Casella G., Berger R.L., *Statistical Inference*, 2nd Edition, Cengage Learning, 2008.

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** **:**DS 5101
**Course Title** **:**Computer Programming for Data Science and AI*
**No. of Credits** **:**2
**Pre-requisites** **:**None
**Compulsory/Optional :**Compulsory

**Aim(s):** This course is a laboratory course that provides students with the computer programming background required for success in data science and artificial intelligence.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Prepare data for further analysis using data analytic tools
- Manipulate data sets programmatically
- Perform exploratory data analysis programmatically
- Apply basic text processing techniques to unstructured data sets
- Visualize data sets effectively
- Perform basic statistical analyses programmatically
- Build data-driven predictive models

**Time Allocation (Hours):** **Lectures Practical**60 hrs

**Course content/Course description:**
Fundamentals : Python programming, The Python toolset, Working with data;
Numerical computation using numpy, Data manipulation using pandas , Exploratory data analysis,.
Text processing with nltk;
Data visualization:Matplotlib, Pandas, Visdom;
Statistics: Random variables, Probability distributions, Hypothesis testing using scipy and statsmodels;
Machine learning tools: Scikit-learn,Pytorch

**Recommended Texts:**

1. Downey, A. (2014), *Think Stats*, 2nd edition, O'Reilly.
2. McKinney, W. (2013), *Python for Data Analysis*, O'Reilly.
3. Online tutorials

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code :** DS5102
**Course Title :** Data Modelling and Management
**No. of Credits :** 3
**Pre-requisites :** DS 4103
**Compulsory/Optional :** Compulsory

**Aim(s):** To emphasizes on emerging data models and technologies suitable for managing different types and characteristics of data. Student will develop skills for analyzing, evaluating, modeling and developing database applications with concerns on both technical and business requirements.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Apply data modeling and management concepts.
- Design and organize various types of data using a relational and non-relational data models.
- Analyze the characteristics and requirements of data and select an appropriate data model.
- Identify, implement and perform frequent data operations (CRUD: create, read, update and delete) on relational and NoSQL databases.
- Describe the concepts and the importance of big data, data security, privacy and governance.
- Describe the concepts and the importance of data engineering and data visualization.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs

**Course content/Course description:**
Recall: Relational Data Model and Management: Relational Model Concepts, SQL, Relational Database Design and Normalization, Relational Database Management Systems (RDBMSs).
NoSQL Data Modeling and Management: NoSQL Concepts and Characteristics, Major Categories of NoSQL Data Models, NoSQL Database Design, NoSQL Features and Operations.
Data Distribution: Data Sharding and Replication Models, CAP Theorem.
Transaction Processing and Consistency Models: Transaction Processing Concepts, ACID Model, BASE Model.
Large Scale Data Handling: Big Data characteristics, Big data modeling and management, API development/usage.
Applications and Case Studies.
Data Engineering,
Intro to Related Topics :Data Security, Data Privacy, Legal Issues,Data Governance: Social and ethical Issues, Biasness (gender, religions, etc.)

**Recommended Texts:**

1. A. Meier and M. Kaufmann, SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management, Springer, 2019, ISBN 978-3658245481

2. M. Kleppmann, Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems, O'Reilly, 2017, ISBN 978-1449373320
3. D. Sullivan, NoSQL for Mere Mortals, Addison-Wesley, 2015, ISBN 978-0-1340-2321-2
4. P. Sadalage and M. Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Addison-Wesley Professional, 2013, ISBN 978-0-3218-2662-6

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code :**DS5103
**Course Title :**Regression Analysis
**No. of Credits :**3
**Pre-requisites :**DS 4102
**Compulsory/Optional :**Compulsory

**Aim(s):** To acquaint students with Least Square methods and concept of linear and multiple regression, correlation, and its applications, To approach the material with matrices algebra, To develop the ability to build linear and nonlinear regression models.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- write regression model in matrix form.
- calculate least square estimates and Maximum Likelihood estimates of regression parameters.
- identify remedies for hetroscedasticity and multicollinearity.
- analyse data using multiple regression techniques.
- perform kernel smoothing.
- use unsupervised learning method to fit regression models

**Time Allocation (Hours):** **Lectures** 45 hrs

**Course content/Course description:**
Simple linear regression and correlation, lack of fit, residual plots, Extension to multiple linear regression, Matrix approach to linear regression, Linear models. Multiple linear regression, Interpretation of coefficients. Inferences in regression analysis. Sequential and partial regression sums of squares. Analysis of aptness of the model. Model selection procedures. Introduction to non-linear regression.
Introduction to regression models, matrix formulation, Gram-Schmidt theory, Regression LSEs, Regression MLEs under normality, confidence sets and LR tests, Cross sectional modeling and Hetroscedasticity, Multicollinearity, Ridge regression, weighted and generalized least squares, Multiple Regression, Logistic Regression, Non-linear and maximum Likelihood Modeling, Splines and other bases, Kernel smoothing.

**Recommended Texts:**

1. Kutner M.H., *Applied Statistical Models*, McGraw-Hill education, 2013.
2. Christensen R. ,*Analysis of Variance, Design and Regression: Linear modelling for unbalanced data*, Chapman & Hall/CRC 2, 2015

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** **:**DS 5104
**Course Title** **:**Machine Learning
**No. of Credits** **:**3
**Pre-requisites** **:**None
**Compulsory/Optional :**Compulsory

**Aim(s):** To introduce students from a variety of science and engineering backgrounds to the fundamentals of machine learning and prepares them to perform R&D involving machine learning techniques and applications. Students learn to design, implement, and evaluate intelligent systems incorporating models learned from data.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Formulate a practical data analysis and prediction problem as a machine learning problem.
- Identify the characteristics of the data set required for a particular machine learning problem.
- Train and test supervised regression and classification models, unsupervised learning and density estimation models, and reinforcement learning models.
- Integrate a trained machine learning model into an online software system.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical**30 hrs

**Course content/Course description:**
Introduction to Machine Learning. Supervised Learning: Linear regression, logistic regression and generalized linear models; Generative probabilistic models; Convex optimization and quadratic programming; Support vector machines; Decision trees and ensemble models (should not be in AI); Non-parametric methods .
Deep Learning: Perceptrons and inspiration from neuroscience; Multilayer neural networks and backpropagation; Optimization techniques; best practices; loss curve analysis.
Learning Theory: Bias-variance tradeoff; Regularization; model selection, and feature selection; Generalization bounds and VC dimension.
Unsupervised Learning: Clustering: k-means; Gaussian mixture models; Principal components analysis; Independent components analysis; Autoencoders.
Reinforcement Learning: Markov decision processes and the Bellman equations; Value iteration, policy iteration; Q-learning

**Laboratory Session(s):**
1. Linear regression models 2. Logistic regression 3. Support vector classification 4. Decision trees 5. Single-layer and multi-layer neural networks 6. Multi-layer back-propagation, regularization, hyperparameter search 7. Model selection, feature selection 8. Clustering with k-means and GMMs 9. Principal components analysis and autoencoders 10. Value iteration and policy iteration 11. Q-learning 12. Deploying a machine learning model

**Recommended Texts:**

1. Mitchell, T. (1997), *Machine Learning*, McGraw-Hill.
2. Bishop, C. (2006), *Pattern Recognition and Machine Learning*, Springer.
3. Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*, MIT Press.
4. Hastie, T., Tibshirani, R., and Friedman, J. (2016), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition, Springer.
5. Sutton, R.S. and Barto, A.G. (2018), *Reinforcement Learning: An Introduction*, 2nd

| edition, MIT Press. | |
|---|---|
| **Assessment** | **Percentage Mark** |
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5105
**Course Title** :Multivariate Methods I
**No. of Credits** :2
**Pre-requisites** :DS 4102
**Compulsory/Optional** :Compulsory

**Aim(s):** To teach the student about multivariate visualization, multivariate normal distribution, hypothesis testing for multivariate data and statistical methods that uncover surprising but valid linkages between variables and explain and predict their measured values.

**Intended Learning Outcomes:**
At the end of the course students will be able to:
- list methods of visualizing multivariate data sets
- identify the usage of multivariate normal distribution
- perform statistical tests of the mean value vector of a multivariate normal distribution
- perform statistical tests of two or several populations of a multivariate normal distribution
- conduct methods and techniques for validation of multivariate normal distribution.
- use principal component and factor analysis for multivariate data sets

**Time Allocation (Hours):** **Lectures** 30 hrs

**Course content/Course description:**
Introduction to multivariate analysis. Multivariate normal distribution. Expected values. Variance - Covariance matrix. Principal component analysis (PCA). Interpretation using illustrative examples. Factor analysis. Comparison with PCA, factor loadings, rotations, Interpretation. Cluster Analysis.

**Recommended Texts:**

1. Johnson R. A. and Wichern D.W., *Applied Multivariate Statistical Analysis*, $6^{th}$ Edition, Pearson publications, 2007.
2. RencherA.C. ,*Multivariate Statistical inference & Applications*, Wiley Interscience, 1997.


| **Assessment** | **Percentage Mark** |
|---|---|

| | |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5106
**Course Title** :Multicriteria Optimization & Decision Analysis
**No. of Credits** :3
**Pre-requisites** :None
**Compulsory/Optional** :Optional

**Aim(s):** To give students an understanding of the decision making process and multicriteria decision analysis methods and optimization processes for finding optimal solutions to problems with multiple decision alternatives and conflicting objectives.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Describe the decision making processes typically used by organizations.
- Formulate a decision making scenario as a multicriteria decision analysis problem.
- Identify and formulate different types of mathematical programming problems including formulations with constraints and multiple objectives.
- Analytically solve simple Pareto optimization problems that are special cases for the application of Karush-Kuhn-Tucker conditions and the Lagrange multiplier theorem.
- Apply methods of multicriteria optimization and decision analysis to real world problem domains.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs

**Course content/Course description:**
Decision making: Decision making processes in organizations, The problem of multiple criteria.
Multicriteria decision analysis methods and approaches: Analytic hierarchy process (AHP), Analytic network process (ANP), Multiattribute utility theory.
Multicriteria optimization: Multi-objective optimization problems, Pareto optimal solutions, Scalarization methods, Multicriteria linear programming, Multicriteria integer programming.
Applications (examples): Health care, Project management, Water management.
Tools and software: Expert Choice, Decisionarium.

**Recommended Texts:**


1. Ishizaka, A., and Nemery, P. (2013). *Multi-Criteria Decision Analysis:Methods and Software,* John Wiley & Sons.
2. Ehrgott, M. (2000), *Multicriteria Optimization*, Springer.
3. Kaliszewski, I., Miroforidis, J., and Podkopaev, D. (2016), *Multiple Criteria Decision Making by Multiobjective Optimization: A Toolbox*, Springer.
4. *Journal of Multi-Criteria Decision Making*, Wiley.

| | |
|---|---|
| **Assessment** | **Percentage Mark** |
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5107
**Course Title** :Distributed Systems
**No. of Credits** :3
**Pre-requisites** :None
**Compulsory/Optional** :Optional

**Aim(s):** To introduces the concepts of distributed systems, cloud computing, and blockchain. Students learn to create required distributed infrastructure and ecosystems for DS&AI applications. Students earn skills on deployment, monitoring, and management of distributed systems.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Explain the main concepts of distributed systems, cloud computing, and blockchain,
- Setup distributed environment for DS&AI applications,
- Utilize distributed file systems,
- Create the process pipeline and deploy cloud services,
- Monitor and management the usage of network resources,
- Implement applications with blockchain/smart contracts.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical**30 hrs

**Course content/Course description:**
<u>Distributed systems</u> :Network topologies, protocols, and synchronization; Content-delivery network (CDN) and Software-defined network (SDN); Types of distributed file systems (DFS) and applications; Distributed caching and scheduling.
<u>Cloud Computing:</u> Characteristics of cloud computing; Virtualization and containers; Parallel computing and GPU/multi-processor programming; Service models (SaaS, IaaS, and PaaS); Ecosystems and container-orchestration systems; Recent trends in cloud computing.
<u>Blockchain</u>: Distributed ledgers/blockchain; Cryptography and hash algorithms; Proof-of-Work (PoW) and consensus algorithms; Cryptocurrency and mining; Smart contracts; Related tools and applications

<u>Laboratory Session(s)</u>:
1. Distributed environment setup
2. Distribute file systems
3. Virtualization and containers
4. Container-orchestration systems
5. GPU/multi-processor programming

6. Dimension reduction
7. API developments
8. Blockchain/smart contracts

**Recommended Texts:**

1. Kshemkalyani, A. D. (2011). Distributed Computing: Principles, Algorithms, and Systems, Cambridge University Press.
2. Kirk, D. B. and Hwu, W. W. (2016). Programming Massively Parallel Processors: A Hands-on Approach, 3rd edition, Morgan Kaufmann.
3. Rafaels, R. (2018). Cloud Computing: 2018, 2nd edition, CreateSpace Independent Publishing Platform.

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5108
**Course Title** :Software Development & Project Management for DS & AI
**No. of Credits** :3
**Pre-requisites** :None
**Compulsory/Optional** :Optional

**Aim(s):** To emphasizes on modern and important software development, software process, and project management. Student will tailor the software development process and project management for DS&AI projects, including planning, iterative development, test driven development, continuous integration/continuous delivery, versioning, and deliverables. Students learn to apply knowledge to the problems in DS&AI domains.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Explain the importance of software development and project management,
- Explain how model-driven development works in a DevOps and agile environments,
- Create model and data versioning,
- Apply the principles of project management to DS & AI project.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical**30 hrs

**Course content/Course description:**
<u>Software Development and Software Process</u>: Introduction to Modern Software Process (Agile practices and frameworks, Model-driven development), Test-Driven Development (TDD), Test Automation, Continuous Integration/Continuous Delivery (CI/CD), Configuration Management, DevOps.
<u>Project Management</u>: Project Integration Management, Project Scope Management, Project Time Management, Project Cost Management, Project Quality Management, Project

Human Resources Management, Project Communications Management, Project Risk Management, Project Procurement Management, Project Stakeholder Management.
<u>DS & AI project management</u>.

**Recommended Texts:**

1. Project Management Institute. (2017*). A Guide to the Project Management Body of Knowledge (Pmbok Guide)*, 6th edition, The Stationery Office Ltd.
2. Beck, K. and Andres, C. (2004). *Extreme Programming Explained: Embrace Change: Embracing Change*, 2nd Edition, Addison-Wesley Professional.
3. Forsgren, N., Humble, J., and Kim, G. (2018). *Accelerate: The Science of Lean Software and Devops: Building and Scaling High Performing Technology Organizations*, 1st Edition, IT Revolution Press.
4. Rubin, K.S. (2012). *Essential Scrum: A Practical Guide to the Most Popular Agile Process (Addison-Wesley Signature): A Practical Guide To The Most Popular Agile Process (Addison-Wesley Signature Series (Cohn))*, 1st edition, Addison-Wesley Professional.
5. Humble, J. and Farley, D. (2010), *Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation (Addison-Wesley Signature Series (Fowler))*, 1st edition, Addison-Wesley Professional.

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** **:**DS 5109
**Course Title** **:**Nature Inspired Computing
**No. of Credits** **:**3
**Pre-requisites** **:**None
**Compulsory/Optional :**Optional

**Aim(s):** To Introduce the students to the field of nature-inspired metaheuristic methods for search and optimization, including the latest trends in nature-inspired algorithms and other forms of natural computing. The students will be exposed not only to paradigms of nature-inspired metaheuristic methods (originating from, for example, biology, living thing behavior and natural phenomena), but also to their applications.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- demonstrate fundamental insights of nature-inspired computation;
- implement nature-inspired methods into concrete algorithms;
- apply nature-inspired algorithms to some search and optimization applications.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical**30 hrs

| |
|---|
| **Course content/Course description:**<br>Bio-Inspired Algorithms: Genetic Algorithm, Clonal Selection Algorithm<br>Living ThingBehaviours : Particle Swarm Optimization, Jaya algorithm, Grey Wolf Optimizer, Biogeography-Based Optimization.<br>Other Paradigms: Differential Evolution, Cuckoo Search, Simulated Annealing, Evolutionary Strategies.<br>Strategies for Empowering the Search: Chaotic-Based Approach, Oppositional-Based Approach, Elitism-Based Approach, Multi-species Approach, Cooperative Approach, Hybridizing Approach.<br>Nature-Inspir  nature Inspired Computing in Applications: Handle the Constraints of Problem, Machine Learning Construction, Cluster Data, Optimize Engineering Problem, Handle Multi-objective.<br>Practical Sessions: The lectures are accompanied by practical sessions where students work in small groups to learn algorithms or solve problems. |
| **Recommended Texts:**<br>1. Xin-She Yang (2014), Nature-Inspired Optimization Algorithms (Elsevier Insights) 1st Edition, Elsevier Insights.<br>2. David E. Goldberg (1989), Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Professional; 1 edition.<br>3. Dan Simon (2013), Evolutionary Optimization Algorithms: Biologically-Inspired and Population-Based Approaches to Computer Intelligence, John Wiley & Sons.<br>4. RavipudiVenkata Rao (2019), Jaya: An advanced Optimization Algorithm and its Engineering Applications, Springer International Publishing AG, part of Springer Nature 2019. |

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

| |
|---|
| **Course Code** :DS 5110<br>**Course Title** : Statistical computing and Simulation Techniques<br>**No. of Credits** :3<br>**Pre-requisites** :DS 4102<br>**Compulsory/Optional :** Optional |
| **Aim(s):** To introduces Monte Carlo methods, collectively one of the most important toolkits for modern statistical inference. |
| **Intended Learning Outcomes:**<br>At the end of the course students will be able to:<br>• generate random numbers using different algorithms.<br>• generate random variates of a given probability distribution.<br>• approximate functions and probabilities using different computational methods. |

- simulate random events in R using MCMC, Metropolis-Hastings Algorithm etc.
- conduct bootstrap sampling and Gibbs sampling.
- apply different variance reduction techniques.
- useMonte Carlo methods in statistical programming and simulations.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hours

**Course content/Course description:**
Introduction to R, generating random variables, Methods of approximating functions and probabilities, computational methods in linear algebra, bootstrap method, Markov Chain Monte Carlo method (MCMC), Gibbs sampling, Metropolis-Hastings Algorithm, variance reduction methods, EM and related algorithms. Practical simulation models using case studies.

**Recommended Texts:**

1. Ross S.M., *Simulation (Statistical Modeling and Decision Sciences)*, 4thEdition.APress, 2006.
2. Law A.M., Kelton W.D., *Simulation Model and Analysis*, 3rd Edition, McGraw-Hill, 1999.
3. Borshchev A., *The Big Book of Simulation Modeling, Multimethod Modeling with AnyLogic*, Anylogic North America, 2013.

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** **:**DS 5215
**Course Title** **:**Business Intelligence & Analytics
**No. of Credits** **:**3
**Pre-requisites** **:**DS 5102
**Compulsory/Optional :**Compulsory

**Aim(s):** Business intelligence (BI) is a process of analyzing business data to obtain business insights and actionable intelligence and knowledge, in order to support better business decision making and capture new business opportunities. This course will give students an understanding of the principles and practices of business intelligence and data analytics to support organizations in conducting their business in a competitive environment.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Explain the concepts characteristics of BI and data analytics
- Describe multiple business problem/decision making domains requiring BI and data
- Analytics.

- Apply BI and data analytic tools and technologies to develop BI applications.
- Integrate BI applications with other information systems as part of a business process.
- Define a BI strategy for an organization 6. Manage a BI project for an organization.
- Describe big data analytics and applications

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs

**Course content/Course description:**

Introduction to Business Intelligence: BI Definition; BI Concepts; Business Intelligence, Analytics, and Data Science; Business Intelligence to Support Decisions.
Data Warehousing for BI: DWdesign; Multidimensional data modelling and analysis; ETL process.
Categories of Data analytics: Descriptive Analytics; Predictive Analytics; Prescriptive Analytics.
Descriptive Analytics: Descriptive Statistics; Business Performance Management; Data Visualization and Dashboard Design.
Predictive Analytics: Data Mining (Text Analytics and Text Mining, Web Analytics, Web Mining, and Social Analytics); Predictive Modeling.
Overview of Prescriptive Analytics: Optimization; Multi-Criteria Systems.
Technical Aspects: BI Architecture; BI Tools and Technologies.
BI Applications: BI Maturity; BI Strategies; BI Project (case study).
Overview of Big Data: Big Data Analytics; Example of Big Data Applications.

**Recommended Texts:**

1. Business intelligence, analytics, and data science, by Ramesh Sharda; DursunDelen; Efraim Turban, Pearson Publisher, 2018.
2. Business Analytics (2nd Ed.) by James Evans, Pearson, 2017. 2. Business Analysis for Business Intelligence (1st Ed) by Bert Brijs, Auerbach Publications, 2013. 3. Business Intelligence Guidebook (1st Ed) by Rick Sherman, Morgan Kaufmann, 2014. 4. Fundamentals of Business Intelligence by Wilfried Grossmann and Stefanie Rinderle-Ma, Springer, 2015.

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5216
**Course Title** :Artificial Intelligence
**No. of Credits** :3
**Pre-requisites** :None

**Compulsory/Optional :**Compulsory

**Aim(s):** To introducing students to fundamentals of Artificial Intelligence. Students will be exposed to several techniques on planning and decision procedures ranging from precise to uncertain and temporal reasoning with applications to intelligent agents.

**Intended Learning Outcomes:**

At the end of the course students will be able to:

- Demonstrate fundamental insights into practical planning and decision procedures;
- Reason under uncertainty;
- Apply planning techniques into intelligent agents.

**Time Allocation (Hours):** **Lectures** 45 hrs

**Course content/Course description:**

Introduction to AI :What is (artificial) intelligence, History of artificial intelligence.

Intelligent agents: Planning and decision, Decision trees and searching techniques, Heuristic algorithms.

Constrained planning Recall on propositional and predicate logic, Unification and resolution, Prolog and/or constraint solvers.

Planning under uncertainty: Bayesian networks, (Partially observable) Markov decision networks.

Temporal planning: Temporal reasoning, Scheduling.

**Recommended Texts:**

1. Russel, S. and Norvig, P. (2013), *Artificial Intelligence: A Modern Approach*, 3rd edition,Pearson.
2. Ghallab, M., Nau, D., and Traverso, P. (2004) *Automated Planning: Theory & Practice*, 1st edition, Morgan Kaufmann Publishers and Elsevier.
3. Bratko, I. (2011), *Prolog Programming for Artificial Intellige*nce, 4th edition, Pearson.

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code :**DS 5217
**Course Title :**Computational Linguistics
**No. of Credits :**3
**Pre-requisites :**DS 5101, DS 5104, DS 5216
**Compulsory/Optional :**Optional

**Aim(s):** To provide students with the fundamental knowledge on computational linguistics, especially text processing. Students would learn the fundamentals of language, text mining, natural language processing and its applications. Students should be able to apply the pre-processing and parsing methods for natural languages. Students could employ techniques and models for NLP problems, design and implement solutions required in NLP applications.

**Intended Learning Outcomes:**

At the end of the course students will be able to:

- Explain the fundamentals of language, text mining, natural language processing and its applications
- Apply the pre-processing and parsing methods for natural languages
- Describe and employ suitable techniques and models for NLP problems
- Design and implement NLP applications

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs

**Course content/Course description:**

Introduction of language, text mining, and natural language processing: Lexical processing, Syntactic processing, Semantic processing, Reference resolution, text classification, Information extraction, Sequence to sequence transformation.

Pre-processing: Tokenization, Normalization, Stemming, Lemmatization.

Word Representation: Vector space model, Word embedding, Application in Text Classification & Clustering.

Syntactic processing: Constituent grammar, Constituent parsing techniques (Rule-based, Machine learning-based), Dependency grammar, Dependency parsing techniques, Semantic Analysis: Word sense similarity, Semantic analyser, Semantic role labelling

Applications on Computational Linguistics: Information extraction (NER and Relation Extraction), Sentiment analysis (Document level sentiment analysis and aspect-based sentiment analysis), Dialogue system, Machine translation

**Laboratory Session(s):** A list of specific lab sessions

1. Exploring available NLP Toolkit: NLTK, spacy, etc.
2. Pre-processing
3. Word representation
4. Syntactic processing
5. Semantic analysis

**Recommended Texts:**

1. Jurafsky, D., Martin, J. H., “Speech and Language Processing”, Prentice Hall, 2nd Edition, 2008 (https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf).
2. Bird, S. , Klein, E., et al. (2009),*Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, 1st Edition, O'Reilly Media.

| Assessment | Percentage Mark |
|---|---|

| | |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5218
**Course Title** :Computer Vision
**No. of Credits** :3
**Pre-requisites** :DS 5104, DS 5216
**Compulsory/Optional** :Optional

**Aim(s):** To introduce the concepts of computer vision with emphasis on state-of-the-art methods used in vision applications.

**Intended Learning Outcomes:**

At the end of the course students will be able to:

- Explain key concepts of computer vision.
- Extract discriminative features from image/video data and use them for pattern classification.
- Analyse, examine, and evaluate existing practical computer vision systems.
- Apply computer vision algorithms from standard libraries and tools to build prototype computer vision systems for real scenarios.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs

**Course content/Course description:**

<u>Introduction to Computer Vision:</u> Mathematical foundation for Computer Vision, Geometry of image formation, Image filtering and Edge detection, Image segmentation.
<u>Feature detection and matching:</u> Feature detection methods, Feature description methods, Feature matching.
<u>Object Recognition:</u> Image classification, Bag of words, Convolutional neural networks, Generative adversarial networks and recurrent neural networks, Performance evaluation
<u>Motion Analysis and Tracking:</u> Camera models, Two-view geometry, Stereo, Optical flow, Structure from motion.
<u>Case Studies and Applications of Computer Vision:</u> Large scale image search and feature indexing, Trajectory analysis, Image caption generation, Action recognition, Other state of the art applications of computer vision. .

**Recommended Texts:**


1. Szeliski, R. (2010). Computer Vision: Algorithms and Applications. Springer.
2. Forsyth, D.A. and Ponce, J. (2011). Computer Vision: A Modern Approach. 2nd Ed. Prentice Hall.
3. Goodfellow, I., Bengio, Y. and Courville, A. (2016). Deep Learning. MIT Press
4. Hartley, R and Zisserman, A. (2004), Multiple View Geometry in Computer Vision. $2^{nd}$ Ed. Cambridge University Press.


| **Assessment** | **Percentage Mark** |
|---|---|

| | |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5219
**Course Title** :Multivariate Methods II
**No. of Credits** :2
**Pre-requisites** :DS 5105
**Compulsory/Optional** :Optional

**Aim(s)**: To introduce further topics in Multivariate Analysis as an extension to SC506.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- identify the relevant method to reduce high dimensional data
- use methods for multiple inference
- evaluate covariance structure models

**Time Allocation (Hours):** **Lectures** 30 hrs

**Course content/Course description:**
Two-groups Discriminant analysis. Multiple-group Discriminant analysis. Multivariate analysis of variance. Canonical correlation. Covariance structure models. Multivariate Data Visulization; Multidimensional scaling, correspondence analysis, Biplots.

**Recommended Texts:**

1. Johnson R.A. and Wichern D.W., *Applied Multivariate Statistical Analysis*, 6$^{th}$ Edition, Pearson publications, 2007
2. RencherA.C.,*Multivariate Statistical inference & Applications*, Wiley Interscience, 1997.

| Assessment | Percentage Mark |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5220
**Course Title** :Spatio Temporal Data Analysis
**No. of Credits** :3

| |
|---|
| **Pre-requisites :**DS 5101, DS 5102, DS 5104<br>**Compulsory/Optional :**Optional |
| **Aim(s): To provide** Students should understand with problems, methods, algorithms, and novel computational techniques in the analysis of spatio-temporal databases. Students will apply these understanding in spatio-temporal data projects. |
| **Intended Learning Outcomes:**<br>At the end of the course students will be able to:<br>• Explain the problems and methods (minimum methods are clustering and predictive learning) on the spatio-temporal data mining<br>• Apply modeling skill for realizing spatio-temporal data projects<br>• Apply integration skill for realizing spatio-temporal data projects<br>• Apply visualization skill for realizing spatio-temporal data projects |
| **Time Allocation (Hours):** **Lectures** 30hrs **Practical** 30 hrs |
| **Course content/Course description:**<br><u>Description on Spatio-Temporal Data:</u> Examples on real case applications on Spatio-Temporal Data, Definition and Properties of Spatio-Temporal Data, Defining Features for Spatio-Temporal Data, including the data instance, Data Similarity.<br><u>Modeling on Spatio-Temporal Data:</u> Clustering on Spatio-Temporal Data, Classification on Spatio-Temporal Data, Predictive learning on Spatio-Temporal Data.<br><u>Integration on Spatio-Temporal Data Projects:</u> Capturing Time Series Data, Capturing Spatial Data, Pre-processing on Time Series Data (time and frequency domain), Pre-processing on Spatial Data, Data Management for the Spatio-Temporal Data.<br><u>Visualization on Spatio-Temporal Data:</u> Visual mapping, Visualizing geo-spatial data, Visualizing spatio-temporal data,<br>Applications of Spatio-Temporal Data Analysis. |
| **Recommended Texts:**<br>1. Cressie, N., Wikle, C. K., "Statistics for Spatio-Temporal Data", John Wiley & Sons, 1st edition, 2015<br>2. Zheng, Y., "Urban Computing", MIT, 2019<br>3. Hsu, W., Lee, M. L., Wang, J., "Temporal and Spatio-Temporal Data Mining", IGI Global, 2007 |

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

| |
|---|
| **Course Code :**DS 5221<br>**Course Title :**Knowledge Representation<br>**No. of Credits :**3 |

**Pre-requisites :**DS 5216
**Compulsory/Optional :**Optional

**Aim(s):** To Introduce the students to the field of knowledge representation, with the goal of reasoning about knowledge. The students will be exposed to specialized knowledge representations stemming from applications in different domains, such as, semantic web and cognitive robotics.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- Use logical formalisms to effectively describe knowledge, belief, events, and situations
- Identify the components of nonmonotonic reasoning and its usefulness as representation mechanism for knowledge systems
- Design real world knowledge-based systems

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical**30 hrs

**Course content/Course description:**
Knowledge Representation and Reasoning: Description Logics and ontologies,Conceptual Graphs and Linked Data, Nonmonotonic Reasoning, Answer Sets, Belief Revision, Model-Based and Case-Based Reasoning.
Classes of Knowledge and Specialized Representations:Reasoning about Knowledge and Belief, Situation Calculus, Event Calculus.
Knowledge Representation in Applications: Semantic Web, Cognitive Robotics, Knowledge Engineering.

**Practical Sessions:** The lectures are accompanied by practical session where the students work in small groups to solve real problems based on the knowledge acquired during the lectures.

**Recommended Texts:**

1. vanHarmelen, F., Lifschitz, V., and Porter, B., editors (2007), *Handbook of Knowledge Representation,* 1st edition, Elsevier.
2. Heath, T. and Bizer, C. (2011), *Linked Data: Evolving the Web into a Global Data Space*, 1st edition. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool.

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

| |
|---|
| **Course Code :**DS 5222<br>**Course Title :**Recent Tends in Machine Learning<br>**No. of Credits :**3<br>**Pre-requisites :**DS 5104<br>**Compulsory/Optional :**Optional |
| **Aim(s):** To providing students with a deeper understanding of machine learning techniques and a wider variety of extant learning models. Students will be prepared to develop advanced machine learning applications and perform research at a state-of-the-art level. |
| **Intended Learning Outcomes:**<br>At the end of the course students will be able to:<br>• Design, train, test, and deploy modern convolutional neural networks (CNNs).<br>• Utilize the principles of adversarial learning to increase the robustness of a machine learning model.<br>• Design, train, test, and deploy generative adversarial networks (GANs).<br>• Utilize recurrent neural networks (RNNs) to model and predict time series.<br>• Utilize deep neural networks to solve difficult tabula rasa reinforcement learning problems.<br>• Apply state-of-the-art machine learning methods to solve problems in speech processing, speech synthesis, natural language understanding, natural language synthesis, computer vision, and intelligent agent design. |
| **Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs |
| **Course content/Course description:**<br>Overview of modern machine learning methods.<br>Convolutional neural networks: Fundamentals, Inception modules, Residual layers, Squeeze and excitation, Detection models, Semantic segmentation models, Instance-aware segmentation models.<br>Deep belief networks<br>Transfer learning<br>Automatic learning<br>Deep unsupervised learning: Generative adversarial networks (GANs), Variationalautoencoders.<br>Practical techniques for deep learning models: Weight initialization, Dropout, Adam optimization, Batch normalization.<br>Time series processing: Hidden Markov models (HMMs), Recurrent neural networks (RNNs) and backpropagation through time, Word embedding for natural language processing, Long short term memory (LSTM) units, Gated recurrent units (GRUs), Attention mechanisms for RNNs.<br>Deep Reinforcement learning: Policy gradients, Actor/critic methods, Imitation learning, Exploration/exploitation, Meta learning, Monte Carlo methods,<br>Applications: Speech recognition, Speech synthesis, Conversational agents, Recommendation systems, Anomaly detection, Computer vision systems |

**Laboratory Session(s):**

1. Preparation environment of recent trends machine learning tools/library such as pip install etc.
2. CNN: Residual layers
3. Deep belief networks
4. Generative adversarial networks (GANs)
5. Weight initialization with GANs /Adam optimization
6. Weight initialization/Batch normalization
7. Time series processing
8. Deep Reinforcement learning
9. Speech recognition application
10. Recommendation systems application
11. Anomaly detection application
12. Computer vision systems application

**Recommended Texts:**

1. Goodfellow, I., Bengio, Y., and Courville, A. (2016), *Deep Learning*, MIT Press.
2. Sutton, R.S. and Barto, A.G. (2018), *Reinforcement Learning: An Introduction*, 2nd edition, MIT Press.

Journals and Magazines:
*IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*
*Journal of Machine Learning Research (JMLR)*. Microtome

| Assessment | Percentage Mark |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5223
**Course Title** :Human Computer Interaction & Information Visualization
**No. of Credits** :3
**Pre-requisites** :DS 5102
**Compulsory/Optional** :Optional

**Aim(s):** To providestudentsthe knowledge in understanding the principles, processes and techniques for design, implementation and evaluation of interactive systems to maximize usability and to enhance user experience of data-driven systems. Students would learn the methods and techniques to present information to enhance the understanding of data.

**Intended Learning Outcomes:**

At the end of the course students will be able to:

- Explain capabilities of both humans and computers and the theoretical foundation of human computer interaction (HCI)
- Adopt the process of design thinking for development of interactive systems

- Employ tools in HCI for implementation of systems with maximized usability and enhanced user experience
- Explain the fundamentals of information visualization
- Summarize dynamic, real-time and spatial datasets across categories, space, and time through visualization tools.

**Time Allocation (Hours):** **Lectures** 30 hrs **Practical** 30 hrs

**Course content/Course description:**
Introduction to HCI: Humans and computers, Interaction..
Design process: Interaction and design basics, User centred design, Design rules, Implementation, Evaluation techniques, Design thinking.
Models and theories of HCI: Cognitive models, Communication and collaboration models, Task analysis, Models of the system, Modelling rich interaction .
Introduction information visualization: Theories of data graphics, Static and moving patterns, Visual objects and data objects.
Data and visualization design, implementation and evaluation: Data types and visualizations, Space Perception and the Display of Data in Space, Interacting with Visualizations, Visualization design, implementation and evaluation.

**Recommended Texts:**

1. Alan Dix et. al. (2003), *Human Computer Interaction*, 3rd edition, Pearson.
2. Colin Ware (2012), *Information Visualization: Perception for Design*, 3rd edition, Morga Kaufmann.
3. Cole NussbaumerKnaflic (2015), *Storytelling with Data: A Data Visualization Guide for Business Professionals*, 1st edition, Wiley.
4. Steve Wexler, Jeffrey Shaffer and Andy Cotgreave (2017), *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios,* 1st edition, Wiley.

| Assessment | Percentage Mark |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 5224
**Course Title** :Special Topics in Data Science
**No. of Credits** :2
**Pre-requisites** :None
**Compulsory/Optional :** Optional

| |
|---|
| **Aim(s):** To introduce emerging concepts, tools and technologies to students. |
| **Intended Learning Outcomes:**<br>At the end of the course students will be able to:<br>• describe specific concept/tool/technology.<br>• describe the principles of the concept/tool/technology discussed.<br>• apply the specific concept/too/technology in practical applications. |
| **Time Allocation (Hours):** **Lectures** 30 hrs |
| **Course content/Course description:**<br>The special topics will be different in different years and will be based on the latest developments in Data Science field. |

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

| |
|---|
| **Course Code** :DS 5225<br>**Course Title** :Industrial Training<br>**No. of Credits** :5 (Non-GPA)<br>**Pre-requisites** :<br>**Compulsory/Optional** :Compulsory |
| **Aim(s):**The primary aim of this course is to embed strong synergies with the regional industries in so as to improve the employability of graduates through viable links with the labour market. |
| **Intended Learning Outcomes:**<br>At the end of the course students will be able to:<br>• Acquire techniques and skills and adapt to emerging technologies<br>• Apply theoretical knowledge in real conditions<br>• Analyze a business environment and propose appropriate business strategies<br>• Communicate effectively and become team members and leaders. |
| **Time Allocation (Hours):** 500 notional hours |
| **Course content/Course description:** |
| **Recommended Texts:** |

| **Assessment** | **Percentage Mark** |
|---|---|

| | |
|---|---|
| **External Supervisor's Evaluation** | 20% |
| **Training Report** | 30% |
| **Oral presentation** | 50% |

**Course Code** :DS 5299
**Course Title** :Independent Study
**No. of Credits** :5
**Pre-requisites** :
**Compulsory/Optional** :Compulsory

**Aim(s):**To prepare students through major steps of a case study including topic selection related to big data and carryout literature survey, model building, analysis and presentation of findings.

**Intended Learning Outcomes:**
At the end of the course students will be able to:

- conduct a literature review for a big data related problem
- clearly define the objective and methodology of a research problem
- conduct a mini research project or a case study in statistics related problem
- prepare presentations on the analysis outcomes

**Time Allocation (Hours):** 500 notional hours

**Course content/Course description:**
Students will study the information on selected research papers and present them in the form of seminars. By involving in an industry related study students will write research proposals and present it.

**Recommended Texts:**

1. Backwell, J. and Martin, J., *A Scientific approach to Scientific writing*, Springer, 2011.
2. Postgraduate Institute of Science (2016) Guidelines for writing M.Sc. Project Report/M.Phil. Thesis/ PhD Thesis.

| Assessment | Percentage Mark |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

**Course Code** :DS 6399
**Course Title** : Research Project
**No. of Credits** :30
**Pre-requisites** :GPA of 3.00 at SLQF 9

| |
|---|
| **Compulsory/Optional :**Optional |
| **Aim(s):**To prepare the student to conduct a research independently by training them to plan, design and conduct a scientific research, to gather reliable scientific data, analyse, and interpret, and to develop skills in scientific writing. |
| **Intended Learning Outcomes:**<br>At the end of the course students will be able to:<br>• apply the scientific method.<br>• design a research project.<br>• complete a research project.<br>• follow ethical issues in scientific research<br>• identify the patenting process in research<br>• prepare presentations at national/international conferences.<br>• produce a thesis conforming to the requirements of the PGIS.<br>• write manuscripts for publication in refereed journals. |
| **Time Allocation (Hours):**3000 notional hrs. (one year duration, full-time) |
| **Course content/Course description:**<br>The students will conduct sufficient amount of work on a chosen research topic under the guidance provided by an assigned supervisor/s, make a presentation of research findings at a national/international conference, and produce a thesis. |

| **Assessment** | **Percentage Mark** |
|---|---|
| **In-course** | 50% |
| **End-semester** | 50% |

## 7. PROGRAMME EVALUATION

### Evaluation of Course work

Based on the scheme given below, the overall performance of a student in a given course shall be evaluated by the respective instructor(s) and a grade shall be assigned.

*Evaluation Scheme*

- For all courses a minimum of 80% attendance is expected.
- The evaluation of each course shall be based on within course and end of course examinations, and assignments. The weightage of marks given below can generally be used as a guideline in the computation of the final grade.
  End of course examination 50 - 60%
  Continuous assessments (mid-semester examination, assignments, etc.) 40 - 50%
- Courses with laboratory and/or fieldwork shall be evaluated, where applicable, on a continuous assessment basis.
- The minimum grade a student should achieve to pass a course is C.
- Students will be informed of the evaluation scheme by the instructor at the beginning of a given course.

*Grade Points and Grade Point Average (GPA)*

The Grade Point Average (GPA) will be computed using the grades earned for core courses and optional courses, taken for credit. Preliminary courses, industrial training, research project and seminar will be evaluated on a pass/fail basis.

On completion of the end of course examination, the instructor(s) is/are required to hand over the grades of a given course to the programme coordinator who will assign the Grade Points using the following table:

| Grade | Grade Point |
|---|---|
| A+ | 4.0 |
| A | 4.0 |
| $A^-$ | 3.7 |
| $B^+$ | 3.3 |
| B | 3.0 |
| $B^-$ | 2.7 |
| $C^+$ | 2.3 |
| C | 2.0 |
| F | 0.0 |

The Grade Point Average (GPA) will be computed using the formula:

GPA = $\frac{\sum c_i g_i}{\sum c_i}$ where $c_i$ is the number for the $i^{th}$ course, and

$g_i$ is the grade point for the $i^{th}$ course

*Make-up Examinations*
'Make-up' examinations may be given only to students who fail to sit a particular examination due to medical or other valid reasons acceptable to the PGIS.

*Repeat Courses*
If a student fails a course or wishes to improve his/her previous grade in a course, he/she shall repeat the course and course examinations at the next available opportunity. However, he/she may be exempted from repeating the course, and repeat only the course examinations if recommended by the teacher-in-charge or M.Sc. Programme Coordinator. The student may repeat the same course or a substituted (new) optional course in place of the original course.A student is allowed to repeat five credits of coursework free-of-charge. The maximum number of credits a candidate is allowed to repeat is fifteen. The maximum grade, a candidate could obtain at a repeat attempt is a B and he/she is allowed to repeat a given course only on two subsequent occasions.

**Evaluation of Industrial Training**

Industrial training will be evaluated on the basis of a report by the external supervisor (at the placement organization), a written report and oral presentation by the student.

**Evaluation of Research Project**

Research project will be evaluated on the basis of a written report (M.Sc. project report) and oral presentation (see Section 6.0 of the PGIS Handbook for the format of the project report).

## 8. TEACHING PANEL

| | Name and Affiliation | Field of specialization |
|---|---|---|
| 1. | Dr.H.T.K. Abeysundara, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, *B.Sc. (Perad.), M.Sc., Ph.D. (Texas Tech)* | *Statistics (Asymptotic Theory and Functional data analysis)* |
| 2. | Dr. S.P. Abeysundara, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, *B.Sc. (Perad.), M.Sc., Ph.D. (Texas Tech)* | *Statistics(Nonlinear modelling & Optimization)* |
| 3. | Dr.H.R.O.E. Dayaratna, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, *B.Sc. (Perad.), Ph.D. (Tokiyo)* | *Computer Networking* |
| 4. | Prof. W. B. Daundasekara, Dept. of Mathematics, Faculty of Science, Univ. of Peradeniya, *B.Sc. (Perad.), M.A., Ph.D. (Alabama)* | *Mathematics* |
| 5. | Dr.P.L. Gamage, Department of Statistics, Faculty of Science, Univ. of Colombo, *B.Sc. (Cmb.), M.Sc., Ph.D. (Texas Tech)* | *Asymptotic Theory* |
| 6. | Mr. P.M.P.C. Gunathilake, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, *B.Sc. (Perad.), MPhil. (Perad)* | *Data Structures, Programming, Computer Architecture* |
| 7. | Dr.D.S.K. Karunasinghe, Dept. of Eng. Mathematics, Faculty of Engineering, Univ. of Peradeniya,*B.Sc.Eng (Perad.), Ph.D. (NUS)* | *Mathematics & Statistics* |
| 8. | Prof. S.R. Kodituwakku, Dept. of Statistics & Computer Science, Faculty of Science, Univ. of *Peradeniya, B.Sc. (Perad.), M.Sc. (AIT), Ph.D.* | *Database Systems and Distributed Systems* |

| | | |
|---|---|---|
| | *(RMIT)* | |
| 9. | Dr.L.S. Nawarathna, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya *B.Sc. (Perad.), Ph.D. (University of Texas)* | *Method comparison studies, Biostatistics* |
| 10. | Dr.U.H.G.R.D. Nawarathna, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, B.Sc. (*Perad.), Ph.D. (North Texas)* | *Medical Image processing, Machine Learning* |
| 11. | Dr. R. Palamakumbura, Dept. of Eng. Mathematics, Faculty of Engineering, Univ. of Peradeniya, *B.Sc. (Perad.), M.Sc., Ph.D. (Texas Tech)* | *Pattern generation in coupled mechanical systems, Statistics* |
| 12. | Prof. A.A.I. Perera, Dept. of Mathematics, Faculty of Science, Univ. of Peradeniya, *B.Sc. (Perad.), M.Sc. (Oslo), Ph.D. (RMIT)* | *Mathematics* |
| 13. | Dr.A.A.S. Perera, Dept. of Mathematics, Faculty of Science, Univ. of Peradeniya, *B.Sc. (Perad.), Ph.D. (SUNY/Albany)* | *Mathematics* |
| 14. | Dr. K. Perera, Dept. of Eng. Mathematics, Faculty of Engineering, Univ. of Peradeniya, *B.Sc. (J'pura), Ph.D. (SUNY/Albany)* | *Statistics* |
| 15. | Dr.U.A.J. Pinidiyaarachchci, Dept. Statistics and Computer Science, Univ. of Peradeniya, B.Sc.(*Perad.), Ph.D.(Uppsala)* | *Computer Vision* |
| 16. | Dr. R. Siyambalapitiya, Dept. Statistics and Computer Science, Faculty of Science, Univ. Peradeniya, B.Sc. (*Perad.*), Ph.D. (Perad.) | *Operating Systems* |
| 17. | Dr. T. M. H. A. Usoof, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, *B.Sc.(Perad.), Ph.D.(Umeå)* | *Human Computer Interaction, Technology in Education* |
| 18. | Dr. C. Walgampaya, Dept. of Eng. Mathematics, Faculty of Engineering, Univ. of Peradeniya, *B.Sc.Eng (Perad.), Ph.D. (Louisville)* | *Click fraud detection, Automatic web robots and Agents* |
| 19. | Prof. P. Wijekoon, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, *B.Sc. (Kel.), Ph.D. (Dortmund)* | *Statistics* (*Linear Models and Multivariate Statistics)* |
| 20. | Dr. R.D. Yapa, Dept. Statistics and Computer Sc., Faculty of Science, Univ. of Peradeniya, *B.Sc.(J'pura), M.Sc. (Cmb.), Ph.D.(Hiroshima)* | *Image processing, Data Mining and Bioinformatics* |

## PROGRAMME COORDINATORS

Mr. Prabhath Gunathilake, M.Phil
Department of Statistics & Computer Science
Faculty of Science
University of Peradeniya
Peradeniya

Dr.Hemalika T.K. Abeysundara
Department of Statistics & Computer Science
Faculty of Science
University of Peradeniya
Peradeniya