

NOTIFICATION

Sub: Revised syllabus of M.Sc. Statistics IV Sem programme.
Ref: Academic Council approval vide agenda

No.: ಎಸಿಸಿ:ಶೈ.ಮ.ಸಾ.ಸ.3:11 (2024-25) dtd 08.01.2025.

The Revised syllabus of M.Sc. Statistics programme IV Semester STS 553 Data Science and STS 557 Data Mining courses which are approved by the Academic Council at its meeting held on 08.01.2025 is hereby notified for implementation with effect from the academic year 2024-25 IV semester batch and onwards.

Copy of the Syllabus shall be downloaded from the University Website
(www.mangaloreuniversity.ac.in)


REGISTRAR

To,

1. The Registrar (Evaluation), Mangalore University.
2. The Chairman, Dept. of Studies in Statistics, Mangalore University.
3. The Chairman, PG BOS in Statistics, Dept. of Studies in Statistics, Mangalore University.
4. The Superintendent (ACC), O/o the Registrar, Mangalore University.
5. The Asst. Registrar (ACC), O/o the Registrar, Mangalore University.
6. The Director, DUIMS, Mangalore University – with a request to publish in the Website.
7. Guard File.

Soft Core	STS 553 : DATA SCIENCE	No. of credits: 3
-----------	------------------------	-------------------

Course Outcomes:

- CO1: Develop in depth understanding of the key technologies in data science.
- CO2: Understand the theory of certain Statistical method of data Science.
- CO3: Gain practical, hands-on experience with statistics programming languages and big data tools through coursework.
- CO4: Demonstrate knowledge of statistical data analysis techniques utilized in company decision making.
- CO4: Apply principles of Data Science to the analysis of real-life data. .
- CO5: Employ cutting edge tools and technologies to analyze data of Data Science Apply algorithms to build machine intelligence.

Unit 1:

Introduction to data science, and its product, data science environment/ecosystem.

Data visualization - data graphics, elements of visual perception, data manipulation – basics of SQL- merging, aggregating and iterating. (12 hrs)

Unit 2:

Data preparation - data cleaning, methodologies and its validation. Python/R application to Data preparation (8 hrs)

Unit 3:

Machine learning: Learning, Supervised learning and Unsupervised learning, Semisupervised learning, and Reinforcement learning; Machine learning (ML). Evaluation metrics of ML models. Classification and regression learning model, naïve Bayes model, knearestneighbour methods, model selection, ROC curve, cross validation. (9 hrs)

Unit 4:

Artificial Neural network – Perceptron, feed-forward neural network, back propagation, gradient descent optimization rule. Support vector machines – maximal margin, primal and dual formulation. Tree classifiers-decision trees, ID3 algorithm CART. Random Forests, multi-modal data processing- components and applications. (13 hrs)

A mini project on data science by self-study (for internal evaluation only).

Reference:

1. Richard O Duda, Peter E Hart, and David D Stork (2012): Pattern Classification, John Wiley and Sons
 2. Oswald Campesato (2020): Artificial Intelligence Machine Learning and Deep Learning, Mercury Learning And Information, New Delhi
 3. Charles Elkan (2013) Predictive analytics and data mining -e-book
 4. Trevor Hastie, Robert Tibshirani & Jerome Friedman (2001): The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, New York.
 5. James D. Miller (2017): Statistics for Data Science, Packt Publishing Ltd.
 6. Wes McKinney(2018): Python for Data Analysis, O'Reilly Media Inc.
 7. Gopal M (2018) Applied Machine Learning, McGraw Hill Education (India) Private Limited, Chennai.
-


Chairman
P.G. BOS in Statistics
Mangalore University
Mangalagangothri - 574 199

Soft Core	STS 557 : DATA MINING TECHNIQUES	No. of credits :3
------------------	---	--------------------------

Course Outcomes:

- CO1: Able to handle big data and able to work for companies.
- CO2: Gain knowledge about basic concepts of Machine Learning and Identify machine learning techniques suitable for a given problem
- CO3: Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining
- CO4: Apply Dimensionality reduction techniques.
- CO5: To assess the strengths and weaknesses of the different algorithms, identify the application area of algorithms and apply them.
- CO6: Apply data mining techniques as well as methods in integrating and interpreting the data sets and improving effectiveness, efficiency and quality for data analysis.

Unit I

Data Mining – motivations and importance, Knowledge Discovery in Databases (KDD) process - search, induction, querying, approximation and compression. Kinds of data considered models – predictive, descriptive, and prescriptive analysis. inter-connections between Statistics, Data Mining,

Applications of data mining.

(10 hrs)

Unit II

Data marts, databases and data warehouses - OLTP systems, multidimensional models

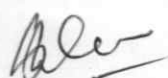
– data cubes, OLAP operations on data cubes, multidimensional schemas.

Supervised learning – classification and prediction, statistical classification

Linear Discriminants-Mahalanobis' linear discriminant, Fisher's linear discriminant;

General Bayesian classifier, Regression based classification,

(10 hrs)



Unit III

Unsupervised learning – Clustering problem, similarity and distance measures, Partitioning algorithms-k-means & k-medoids (PAM) algorithms. Density based clustering algorithms (DBSCAN). (10 Hrs)

Unit IV

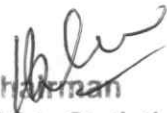
Computational methods useful in data mining: Expectation-Maximisation (EM) algorithm, Genetic algorithm, Resampling Techniques - Gibbs sampler, Markov Chain Monte Carlo (MCMC) method., Bootstrap sampling,

(10 Hrs)

References:

1. Jiawei Han, Micheline Kamber: (2002):Data Mining-Concepts and Techniques, Morgan Kaufman Publishers, U.S.A
2. Margaret.H.Dunham (2005): Data Mining-Introductory and Advanced Topics, Pearson Education.
3. Trevor Hastie, Robert Tibshirani & Jerome Friedman (2001):The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, New York,
4. Michael Berthold, David J. H and (Eds): (2003) Intelligent Data Analysis - An Introduction (2nd Ed), Springer.
5. J.P. Marques de Sa: (2001):Pattern Recognition - Concepts, Methods and Applications, Springer
6. Rajan Chattamvelli: (2009): Data Mining Methods, Narosa Publishing House.

~~~~~

  
Chairman

P.G. BOS in Statistics  
Mangalore University  
Mangalagangothri - 574 199