Guidelines for B.Tech Honors in Engineering Regulations-R24

(Effective for the students admitted into I year from the Academic Year 2024-2025 onwards)

1) Introduction

The goal of introducing B.Tech (Honors) is to facilitate the students to choose additionally the specialized courses of their choice and build their competence in a specialized area in the UG level. The programme is a best choice for academically excellent students having good academic record and interest towards higher studies and research.

All the students pursuing regular B.Tech with prerequisite CGPA are eligible to the register Honors degree course. **A student has to acquire 18 more credits, in addition to 160 credits (without back log history and meeting other guidelines) required, for the award of the B.Tech Honors degree. Out of the 18 extra credits required to obtain the Honors degree, at least SIX Credits (i.e., two courses of 3 credits each) must be earned from NPTEL / SWAYAM MOOC Courses.** The additional courses shall be advanced subjects in the concerned department / discipline. The department concerned will determine required courses for award of Honors degree. The subjects in the Honors degree would be a combination of core (Theory and Lab) and some electives.

2) Objectives

The objectives of initiating the B.Tech (Honors) degree certification are:

a) To encourage the undergraduates towards higher studies and research
b) To prepare the students to specialize in core Engineering streams
c) To attain the high-level competence in the specialized area of UG programme
d) To learn the best educational and professional skills in the specialized area after the completion of his undergraduate courses.
e) To provide the opportunity to learn the advanced courses m the specified undergraduate programme

3) Applicability and Enrolment

a) To all B.Tech (Regular and Lateral Entry) students admitted in Engineering & Technology with CGPA of 7.0 up to II Year I Semester (III Semester), without any backlogs and backlog history.

b) It may be noted that both regular degree and Honors degree are to be completed in 4 Years for Regular students and 3 Years for lateral entry admitted students, without any backlog history.

c) For applicability of Honors degree, both regular B.Tech and Honors degree courses shall be successfully completed.

d) Transfer of credits from a particular minor to regular B.Tech or another major degree and vice-versa shall not be permitted

4) Entry level

a) The B.Tech students (both Regular and Lateral Entry) pursing a major degree programme can register for Honors degree at their choice in the same department / allied (as mentioned in AICTE Handbook) offering major degree from IV semester onwards.

b) Students registering for Honors degree shall select the subjects from same branches / department based on the recommendations of BoS committee. For example, if a student pursuing major degree in Electrical & Electronics Engineering, select subjects in Electrical & Electronics Engineering only and he / she will get major and Honors degree in Electrical & Electronics Engineering

c) Students shall be permitted to select a maximum of two subjects per semester from the list of subjects specified for Honors degree other than Lab courses.

d) The students shall complete Honors degree without supplementary appearance within stipulated period as notified by college / JNTU-GV for the completion of regular major B.Tech programme.

e) Honors degree shall not be awarded at any circumstances without completing the regular major B.Tech programme in which a student got admitted

f) If a student is detained due to lack of attendance, he/ she shall not be permitted to register the courses for Honors degree

g) The subjects completed under Honors degree programme shall not be considered as equivalent subjects in case the student fails to complete the major degree programme

h) Students completed their degree shall not be permitted to register for Honors degree

5) Structure of Honors in B.Tech

a) The student shall earn additional 18 credits for award of Honors degree from same branch / department / allied (as mentioned in AICTE Handbook) registered for major degree
b) Students can complete Honors degree courses either in the college or online from platforms like NPTEL/SWAYAM etc…
c) The overall attendance in each semester of regular B.Tech courses and Honors degree courses shall be computed separately
d) Student having less than 65% attendance in Honors courses shall not be permitted for "Honors Course (s) semester end examinations".
e) A student detained due to lack of attendance in regular B.Tech programme shall not be permitted to continue Honors programme
f) The teaching, examinations (internal and external) and evaluation procedure of Honors degree courses offered in offline is similar to regular B.Tech courses
g) Students may choose theory or practical courses to fulfill the minimum credit requirement.
h) Students shall be allowed to take maximum two subjects per semester pertaining to their Honors degree other than lab courses
i) The students registered for minor shall not be permitted to register for B.Tech (Honors)

6) Credits requirement

a) A Student will be eligible to get B.Tech (Honors), if he / she complete an additional 18 credits. These may be acquired either in offline or online like NPTEL / SWAYAM etc by doing 8 / 12 / 16 week courses covering 2 / 3 / 4 credits.
b) The colleges offering Honors degree courses shall be ready to teach the courses in offline at their college in the concerned departments. Curriculum and the syllabus of the courses shall be approved by the Board of Studies.
c) Students shall produce a certificate issued by the NPTEL / SWAYAM etc., conducting agency as a proof of credit attainment.
d) The teaching and evaluation procedure of Honors courses offering in offline mode shall be similar to that of regular B.Tech courses
e) After successful completion of all major and Honors degree courses with specified CGPA the College / University will award B.Tech (Honors)

7) Procedure to Apply for Honors degree

a) The department offering the Honors will announce courses required before the start of the session.
b) The interested students shall apply for the Honors course to the HOD of the concerned department.
c) The whole process should be completed within one week before the start of every session.
d) Selected students shall be permitted to register the courses for Honors degree.

8) To Join in Honors Program

a) Each department offering the Honors degree shall submit the final list of selected students to the principal.
b) The selected students shall submit a joining letter to the Principal through the concerned HoD.
c) The department offering Honors shall maintain the record of student pursing the Honors degree.
d) With the approval of Principal and suggestion of advisor /mentor, students can choose courses from the approved list and shall register the courses within a week as per the conditions laid down in the structure for the Honor degree.
e) Each department shall communicate the Honors courses registered by the students to the time table drafting committee and accordingly time table will be drafting. Time table drafting committee shall see that no clash in time tables.
f) If the student wishes to withdraw / change the registration of subject / course, he/she shall inform the same to advisor/mentor, subject teacher, HoDs of minor department and parent department and Principal within two weeks after registration of the course.

9) Procedure for Monitoring the Progress of the Scheme

The students enrolled in the Honor courses will be monitored continuously at par with the prevailing practices and examination standards. An advisor / mentor from parent department shall be assigned to a group of students to monitor the progress.

10) Allocation of seats for Honors degree

Total number of seats offered for Honors degree shall be a maximum of 60 (based on merit).

11) Examinations

a) The examination for the Honors degree courses offered in offline shall be conducted along

with regular B.Tech programme.

b) The examinations (internal and external) and evaluation procedure of Honors degree courses offered in offline is similar to regular B.Tech courses.

c) It may be noted that both major and Honors courses (from IV Semester to VII Semester) are to be completed in 4 Years for Regular students and 3 Years for lateral entry admitted students.

d) There is no supplementary examination for the failed subjects in an Honors degree programme.

e) Examination Fees: Examination Fees will be as per the College norms

f) For awarding the class, CGPA obtained in Major Degree only will be considered.

g) For awarding the Honor's, obtained credits only will be considered.

h) The student can complete these MOOCs NPTEL courses during III year I semester to IV year II semester course completion and these courses will be included in the IV year II Semester grade memo.

College offering B.Tech Honors Degree in the following domains, and the student can take any one of domain to get B.Tech Honors by satisfying eligibility criteria.

S.No	Specialization	Offered By	Honors (For Students)
1	Artificial Intelligence and Machine Learning	CSE	CSE/CSD
2	Data Science	CSE	CSE/CSM

1. Artificial Intelligence and Machine Learning

S.No	Subject Code	Year of Study	Subject	L	T	P	C
1	R24AMT201	II-II	Computational Statistics and Data Analysis	3	0	0	3
2	R24AMT302	III-I	Machine Learning Using Python	3	0	0	3
3	R24AMT303	III-II	Advance Machine Learning	3	0	0	3
4	R24AMT404	IV-I	Deep Learning	3	0	0	3
5	R24AIM405	---	NPTEL/MOOC Course-I	0	0	0	3
6	R24AIM406	---	NPTEL/MOOC Course-II	0	0	0	3
			Total	**12**	**0**	**0**	**18**

2. Data Science

S.No	Subject Code	Year of Study	Subject	L	T	P	C
1	R24DST201	II-II	Statistical Analysis and Programming for Data Science	3	0	0	3
2	R24DST302	III-I	Data Handling and Visualization	3	0	0	3
3	R24DST303	III-II	Big Data Technologies	3	0	0	3
4	R24DST404	IV-I	Time series analysis and Forecasting	3	0	0	3
5	R24DSM405	---	NPTEL/MOOC Course-I	0	0	0	3
6	R24MDS406	---	NPTEL/MOOC Course-II	0	0	0	3
			Total	**12**	**0**	**6**	**18**

Chairperson
Board of Studies (CSE)

R24AMT201 Computational Statistics and Data Analysis 3 0 0 3

(Honors/Minor Course: Artificial Intelligence and Machine Learning)

Course Objectives:

To introduce several statistical techniques found to be serving as tools even today in the development of machine learning and artificial intelligence based computer algorithms.

- To understand the fundamentals of statistics, including types of data, descriptive statistics, and data visualization techniques for univariate and bivariate analysis.
- To gain foundational knowledge of probability theory and various probability distributions, including discrete and continuous types, and apply concepts like Bayes' theorem and Monte Carlo simulations.
- To develop skills in hypothesis testing and statistical inference for analyzing and drawing conclusions from data.
- To acquire the ability to pre-process data effectively and evaluate model performance using various metrics, normalization techniques, and cross-validation strategies.
- To apply statistical methods and machine learning techniques for data analysis, including dimensionality reduction, feature selection, regression, classification, and clustering algorithms. To learn the data processing techniques required to get applied statistical and machine learning algorithms.

Course Outcomes

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

Mapping with POs and PSOs															Dok
Course Code	**Course Outcomes (COs)**	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PSO1**	**PSO2**	
R24AMT201.1	Understand the basic concepts of statistics and perform descriptive data analysis and visualization	3	2	1	2	1	-	-	-	-	1	-	3	1	L1, L2, L3
R24AMT201.2	Apply probability concepts and various probability distributions in solving real-world problems.	3	3	2	3	2	-	-	-	-	1	-	3	2	L1, L2 L3
R24AMT201.3	Analyze data using hypothesis testing and interpret statistical results.	3	3	2	3	2	-	-	-	-	1	-	3	2	L1, L3
R24AMT201.4	Perform data preprocessing and evaluate model performance using appropriate statistical measures.	3	3	3	3	3	-	-	-	-	2	-	3	3	L4, L5

R24AMT201.5	Apply statistical techniques and implement machine learning models for data analysis and decision-making.	3	3	3	3	3	-	-	-	1	2	2	3	3	L4, L5, L6

SYLLABUS

UNIT-I: Introduction to Statistics: 10 Hours

What is statistics, Statistical Data – Categorical, Numerical (Continuous), Univariate and Bivariate Analysis, Mean, Median, Mode, Standard Deviation, Harmonic Mean, Data Visualization – Line, Scatter, Boxp lots, Histogram. Descriptive statistics: qualitative and quantitative Variable, discrete variable, population, sample, random sample. **CO's-CO1**

Self Learning Concepts: Descriptive statistics, and data visualization techniques

UNIT- II: Probability and Distribution: 15 Hour Probability, Random Variable, Joint and Conditional Probability, Baye's Theorem, Monte Carlo Method, Probability Distributions, Characterizing a Distribution, Discrete Distributions, Normal Distributions, Continuous Distributions Derived from the Normal Distribution, Poisson Distribution. Other Continuous Distributions: Lognormal, Weibull, Exponential, Uniform. **CO's-CO2**

Self Learning Concepts: Probability theory and various probability distributions

UNIT – III: Hypothesis and Statistical Tests: 10 Hour

Null hypothesis, Alternative hypothesis, Typical Analysis procedures, Hypothesis Concept, Errors, p-Value, z-value, Crucial value, Test on Numerical Data – Distribution of a Sample Mean, Comparison of Two Groups, Comparison of Multiple Groups, degree of freedom, T-test, Z-test, ANOVA analysis. **CO's-CO3**

Self Learning Concepts: Probability testing and statistical inference

UNIT- IV: Data Pre-processing and Performance Analysis: 10 Hours

Data Pre-processing steps: data cleaning – missing data, noisy data, binning method, regression, clustering, data transformation - attribute selection, data reduction – feature selection, dimensionality reduction. Normalization: Decimal Scaling, Min-Max scaling, Z-score Performance metrics: Confusion matrix, sensitivity, specificity, F1 score, Recall, Precision, ROC-AUC Curve, Cross validation technique – K-fold Model evaluation: Residual error, Bias, Variance, Mean square error, RMSE, Loss. **CO's-CO4**

Self Learning Concepts: Normalization techniques and cross-validation strategies

UNIT-V: Statistical Methods and Machine Learning: 15 Hours

Statistical Methods: Dimensionality Reduction Techniques – Principal Component Analysis, Discriminate Analysis, Feature Selection – Chi2 square method, Variance Threshold, Recursive

Feature Elimination, Outliers detection methods, Resampling – Random, under-sampling and over re-sampling. **Machine Learning**: Introduction to Machine Learning: Supervised and unsupervised ML, Regression (Linear regression, Logistic regression), Classification (Naïve-Bayes classifier), Clustering (K-means, K-mediod). **CO's-CO5**

Self Learning Concepts: Statistical techniques and Machine learning

Board of Studies : Computer Science and Engineering

Approved in BOS No: 02, 9th May, 2025

Approved in ACM No: 02

Text Books:

1. 1.Thomas Haslwanter, "An Introduction to Statistics with Python with Applications in the Life Sciences", Springer International Publishing Switzerland 2016, ISBN: 978-3-319-28315-9, ISBN: 978-3-319-28316-6 (eBook)
2. Allen B. Downey, "Think Stats", Second Edition,O'Reilly Media,ISBN: 978-1-491-90733-
3. Foundations of Machine Learning, Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, MIT Press, Second Edition, 2018.

Reference Books:

1. José Unpingco, "Python for Probability, Statistics, and Machine Learning", Springer International Publishing Switzerland, ISBN: 978-3-319-30715-2, DOI: 10.1007/978-3-319 30717-6, ISBN: 978-3-319-30717-6 (eBook)
2. Claus Weihs, Olaf Mersmann, Uwe Ligges, "Foundations of Statistical Algorithms", CRC Press,ISBN:978-1-4398-7887-3(eBook-PDF).

Web References:

1. http://file.allitebooks.com/20151204/Foundations%Statistical%20Algorithms.pdf
2. http://onlinestatbook.com/Online_Statistics_Education.pdf
3. https://upload.wikimeda.org/wikipedia/commons/8/82/Statistics.pdf
4. http://cnx.org/content/coll0522/1.38/pdf
5. http://www.greenteapress.com/thinkstats/thinkstats.pdf

MOOC / Video Lectures available at:

1. https://www.udemy.com/course/introduction-to-bayesian-statistics/(Free Course)
2. https://www.youtube.com/watch?v=xxpc-HPKN28
3. https://udacity.com/course/intro-to-statistics--st101#(Free Course)
4. https://nptel.ac.in/courses/111/105/111105090/
5. https://ntpel.ac.in/courses/111/105/111105077/

Internal Assessment Pattern

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)
L1	35	--
L2	40	--
L3	25	25
L4	--	35
L5	--	25
L6	--	15
Total (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

Level 1-Remembering

1. What is statistics?
2. Define categorical data.
3. What is a histogram?
4. List different types of variables in statistics.
5. What is meant by standard deviation?
6. What is a random variable?
7. Define Bayes' Theorem.
8. What is the Poisson distribution?
9. What is a null hypothesis?
10. What is data cleaning?
11. Define normalization.
12. What is a confusion matrix?
13. What is supervised learning?

Level 2 – Understanding

1. Explain the difference between univariate and bivariate analysis.
2. Compare mean, median, and mode.
3. Describe a box plot with an example.
4. Explain the concept of continuous vs. discrete data.
5. Explain the difference between joint and conditional probability.
6. Describe the characteristics of the normal distribution.

7. Differentiate between Type I and Type II errors.
8. Explain the concept of a z-test and when it is used.

Level 3 – Applying

1. Calculate the mean, median, and mode of a given dataset.
2. Use a scatter plot to analyze the relationship between two variables.
3. Classify data as categorical or numerical.
4. Apply Bayes' theorem to solve a real-world problem.
5. Use the Monte Carlo method to simulate a probability experiment.
6. Conduct a t-test on a sample dataset.
7. Calculate the degrees of freedom in a two-sample test.
8. Apply z-score normalization to a sample dataset.
9. Perform dimensionality reduction using PCA.
10. Apply k-means clustering to segment a dataset.

Level 4 – Analyzing

1. Analyze a dataset to find outliers using box plots.
2. Distinguish between qualitative and quantitative variables in real-world examples.
3. Analyze the differences between discrete and continuous probability distributions.
4. Compare Poisson and normal distributions with examples.
5. Analyze a scenario to decide whether to use a t-test or z-test.
6. Interpret the result of a hypothesis test using p-value.
7. Analyze the performance of a classification model using precision and recall.
8. Compare RMSE and MSE in terms of model error evaluation.
9. Analyze how bias and variance impact model performance.
10. Compare linear regression and logistic regression.

Level 5 – Evaluating

1. Evaluate which measure of central tendency (mean, median, mode) is most appropriate in skewed data.
2. Justify the use of standard deviation over range in measuring dispersion
3. Evaluate which probability distribution best fits a real-world dataset.
4. Critique the use of Monte Carlo simulation in statistical modeling.
5. Evaluate the validity of conclusions drawn from a statistical test.
6. Judge the appropriateness of ANOVA in comparing more than two groups.
7. Evaluate different data preprocessing techniques for handling noisy data.
8. Critique the effectiveness of k-fold cross-validation in model validation.
9. Justify the selection of a specific machine learning algorithm for a given problem.

Level 6-Create

1. Design an end-to-end feature selection pipeline using Chi-Square method, Variance Threshold, and Recursive Feature Elimination for a given high-dimensional dataset.
2. Develop a dimensionality reduction solution combining PCA and Discriminate Analysis for a dataset suffering from multi collinearity. Explain your approach.
3. Construct an outlier detection strategy integrating multiple methods (like Z-score, IQR, and DBSCAN-based detection) for a highly imbalanced dataset.
4. Create a balanced dataset by applying appropriate random sampling, under-sampling, and over-sampling techniques. Justify your choice of resampling methods.
5. Propose a new feature selection framework that improves upon traditional Chi-Square and Recursive Feature Elimination for datasets with non-linear feature interactions.
6. Design a machine learning model that first uses unsupervised learning (clustering) to segment data, followed by supervised learning (classification) to predict labels.
7. Formulate a novel regression model by hybridizing concepts from linear and logistic regression for datasets with mixed continuous and categorical targets.
8. Develop a comparison framework to evaluate the performance of K-Means and K-Medoids on a noisy dataset. Include performance metrics and visualization techniques.
9. Create a complete machine learning workflow (data preprocessing, feature selection, model building, and evaluation) for a real-world classification task using Naïve Bayes.

Chairperson
Board of Studies (CSE)

R24AMT302 Machine Learning Using Python 3 0 0 3

(Honors/Minor Course: Artificial Intelligence & Machine Learning)

Course Objectives:

1. Students understand issues and challenges of Machine Learning.
2. Should be able to select data, model, model complexity etc.
3. Understand the strengths and weaknesses of many popular machine learning approaches.

Course Outcomes

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

Course Code	Course Outcomes	Mapping with POs and PSOs							Dok
		PO1	PO2	PO3	PO4	PS01	PS02	PS03	
R24AMT302.1	Understand the basic concepts of various types of machine learning	3	2	3	2	3	3	3	L1, L2
R24AMT302.2	Assess the suitability of clustering algorithms in solving a particular problem	3	3	3	3	3	2	3	L2, L3
R24AMT302.3	Apply various dimensionality reduction techniques for the extraction of features from high dimensionality data	3	3	3	1	3	2	2	L3, L4
R24AMT302.4	Develop scaling up machine learning techniques	3	3	3	3	2	2	-	L4, L5
R24AMT302.5	Choose a suitable regression technique that is appropriate for a particular dataset by analyzing the trade-off of computational complexity versus convergence speed.	3	2	2	2	3	3	-	L4, L5

SYLLABUS

UNIT-I: Brief Introduction to Machine Learning: 15 Hours

Supervised Learning, Unsupervised Learning, Ensemble Learning, Reinforcement Learning. **Supervised Learning:** Decision Tree Induction, Naïve Bayes Classification, Rule based Classification, K-Nearest Neighbor. **CO's–CO1**

Self-Learning Topics: Performance evaluation metrics of Classifiers

UNIT-2: Unsupervised Learning: 15 Hours

Clustering, Partitioned Clustering (K-Means), Hierarchical Clustering, BIRCH, CURE, Density based Clustering (DBSCAN).

Self-Learning Topics: Performance evaluation metrics of Clustering. **CO's–CO2**

UNIT-3: Feature Analysis: **15 Hours**

Dimensionality Reduction, Feature Selection, Feature Projection, Filter Method, Wrapper Method, Embedded Method, Feature Projection, Principal Component Analysis (PCA). **CO's–CO3**

Self-Learning Topics: Linear Discriminant Analysis (LDA).

UNIT-4: Regression: **15 Hours**

Linear and Logistic regression, Regularization - L1 & L2 regularization, dropout methods, Lasso Regression, Ridge regression, Hypothesis testing of Regression Model, R-square and goodness of fit, Influential Observations. **CO's–CO4**

Self-Learning Topics: Multiple Linear Regressions

UNIT-5: Regression: **15 Hours**

Linear and Logistic regression, Regularization - L1 & L2 regularization, dropout methods, Lasso Regression, Ridge regression, Hypothesis testing of Regression Model, R-square and goodness of fit, Influential Observations.

Models of ANNs: Feed-forward & feedback, Multi-layer Feed forward Networks, Delta learning rule for Multi-Perceptron layer, Generalized delta learning rule, Error back-propagation training networks, Recurrent NN. **CO's–CO5**

Self-Learning Topics: Multiple Linear Regression

Board of Studies : Computer Science and Engineering

Approved in BOS No: 02,9th May, 2025

Approved in ACM No: 02

Expert Talk (To be Delivered by SMEs from Industries) CO's/ PO's / PSO's

1. Multiple Linear Regression CO5 PO1,PO2,PO3,PO4,PSO1,PSO2

Text Books:

1. Shalev-Shwartz, S., Ben-David, S., (2014), *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press
2. R. O. Duda, P. E. Hart, D. G. Stork (2000), *Pattern Classification*, Wiley-Blackwell, 2nd Edition

Reference Books:

1. Mitchell Tom (1997), *Machine Learning*, Tata McGraw-Hill
2. C. M. Bishop (2006), *Pattern Recognition and Machine Learning*, Springer-Verlag New York, 1st Edition

Web References:

1. MathWorks Machine Learning Workflow eBook (PDF)
2. Coursera – Machine Learning with Python

3. NPTEL – Machine Learning Course
4. Geeks for Geeks – Machine Learning

Internal Assessment Pattern:

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)
L1	35	10
L2	30	10
L3	25	20
L4	10	35
L5	--	25
TOTAL (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

L1: Remember

1. Define Supervised Learning and give two examples.
2. What is Ensemble Learning in machine learning?
3. List the types of Supervised Learning algorithms.
4. Write the full form of KNN and mention its use.
5. What are the basic components of a Decision Tree?

L2: Understand

1. Differentiate between Supervised and Unsupervised Learning.
2. Explain the working of the Naïve Bayes classifier with an example.
3. Describe the steps involved in building a Decision Tree.
4. How does K-Nearest Neighbor classify a new data point?
5. Explain the concept of Reinforcement Learning with a real-life analogy.

L3: Apply

1. Apply Decision Tree classification to a dataset of students' exam results to predict pass/fail.
2. Given a dataset of email spam detection, choose an appropriate classifier and justify.
3. Train a simple KNN classifier using 5 sample data points and classify a new entry.
4. Use Naïve Bayes classification for sentiment analysis of movie reviews.
5. Construct a scenario where reinforcement learning could be applied in robotics.

L4: Analyze

1. Compare Decision Tree and Rule-based Classification on the basis of interpretability.
2. Analyze the role of feature selection in improving KNN performance.

3. What are the limitations of Naïve Bayes in high-dimensional data?
4. Compare Supervised, Unsupervised, and Reinforcement Learning with examples.
5. Analyze how Ensemble Learning improves model accuracy over single learners.

L5: Evaluating

1. Evaluate the effectiveness of Decision Tree over KNN for a multiclass dataset.
2. Which classifier would you prefer for medical diagnosis: Naïve Bayes or Rule-based? Justify.
3. Assess the strengths and weaknesses of Supervised Learning in stock market prediction.
4. Critically evaluate the use of Ensemble Methods like Random Forests over individual classifiers.
5. Justify the choice of classifier for real-time face recognition in surveillance systems.

Chairperson
Board of Studies (CSE)

R24AMT303 Advanced Machine Learning 3 0 0 3

(Honors/Minor Course: Artificial Intelligence and Machine Learning)

Course Objectives:

1. To provide in-depth knowledge of ensemble learning techniques such as bagging, boosting, and stacking for improved model performance.
2. To explore structured and probabilistic models like RBMs, HMMs, and Markov Random Fields for complex pattern recognition.
3. To introduce advanced regularization techniques and optimization strategies for enhancing model generalization and dealing with uncertainty.
4. To equip students with the ability to design and implement time series forecasting models using real-world datasets.
5. To develop the skills to apply machine learning methods in Natural Language Processing (NLP) tasks such as classification, information retrieval, and summarization.

Course Outcomes

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

Course Code	Course Outcomes (COs)	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	Dok
R24AMT 303.1	Analyse and apply ensemble techniques such as bagging, boosting, and stacking to improve prediction accuracy.	3	3	2	2	3	1	1	1	2	2	1	3	2	L1,L2
R24AMT 303.2	Design and implement structured models including Restricted Boltzmann Machines and Hidden Markov Models for data representation and sequence modeling	3	3	3	2	3	1	1	1	2	2	1	3	1	L2,L3
R24AMT 303.3	Apply regularization and optimization methods to reduce overfitting and improve the robustness of machine learning models.	3	3	3	2	3	1	1	1	2	2	1	3	1	L3,L4
R24AMT 303.4	Construct and evaluate forecasting models such as ARIMA and Holt- Winters for time series data analysis in financial and other domains.	3	3	3	3	3	1	1	1	2	2	1	2	1	L3,L4

R24AMT 303.5	Utilize machine learning techniques to solve NLP problems such as text classification, summarization and question answering using both supervised and unsupervised methods	3	3	3	2	3	1	1	1	2	2	1	3	2	L5,L6

SYLLABUS

Unit I: **12 Hours**

Ensemble Methods and Boosting Techniques: Bagging and Bootstrapping: Weak Learners, Bias-Variance Trade off, Random Forests, Voting Ensemble Methods, Applications of Ensemble Models in Real-World Problems Boosting Algorithms: Adaptive Boosting (AdaBoost), XG-Boost, Light-GBM, Cat-Boost, Stacking Techniques. **CO's-CO1**

Self Learning Concepts: Visualize the Bias-Variance Trade-off with models trained on under fitting vs over fitting data. Analyze use-cases in credit scoring, healthcare diagnostics, and stock price prediction

Unit II: **15 Hours**

Structured and Probabilistic Models: Structured Models: Restricted Boltzmann Machines (RBMs): Contrastive Divergence, Size of Mini-Batch, Monitoring Learning Progress and Over fitting, Markov Random Fields, Hidden Markov Models (HMMs).

Bayesian Modeling: Dealing with Uncertainty in ML, Bayesian Inference and Gaussian Processes, Bayesian Neural Networks, Randomized Methods in ML. **CO's-CO2**

Self Learning Concepts: Explore Contrastive Divergence (CD) as a training algorithm

Implement small-scale BNNs using libraries like Tensor Flow Probability, Pyro, or Edwar.

Unit III: **10 Hours**

Regularization and Optimization in ML: Regularization Techniques: Bias and Variance, Early Stopping, L1 and L2 Regularization, Sparse Coding, Dropout Layers, Optimization Techniques, Avoiding Over-fitting in Deep Learning Models. **CO's-CO3**

Self Learning Concepts: Visualize regularization effects on model weights**.** Implement dropout in Keras or PyTorch and visualize its impact on model generalization.

Unit IV: **15 Hours**

Time Series and Forecasting Models: Trend Analysis, Cyclical and Seasonal Components, Smoothing and Moving Averages, Box-Jenkins Methodology, Holt-Winters Method, Auto-Correlation, ARIMA Models, Real-Time Applications of Time Series Analysis in Financial Markets. **CO's-CO4**

Self Learning Concepts: Apply to real datasets using the ARIMA framework**.** Visualize trends using line plots**,** rolling means**,** and polynomial regression

Unit V: **15 Hours**

Natural Language Processing (NLP): Introduction to NLP, NLP Pipelines, Tokens, Lemmatization, POS Tagging, Text Classification Named Entity Recognition (NER), Question Answering and Information Retrieval, Text Simplification and Summarization, Supervised and Unsupervised Learning Methods in NLP. **CO's-CO5**

Self Learning Concepts: Tokenize and lemmatize a paragraph from a news article. Use `spaCy` to perform NER on articles, resumes, or tweets.

Board of Studies : Computer Science and Engineering

Approved in BoS No: 02,9th May, 2025

Approved in ACM No: 02

Text Books:

1. **Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani** An Introduction to Statistical Learning: with Applications in R Springer, 2nd Edition, 2021. ISBN: 978-1-0716-1418-1
2. **Aurelien Gerona** Hands-On Machine Learning with Scikit-Learn, Kera's, and Tensor Flow O'Reilly Media, 2nd Edition, 2019. ISBN: 978-1-492-03264-9 (Great practical guide with hands-on code for ensemble models, NLP, and time series)

Reference Books:

1. **Trevor Hastie, Robert Tibshirani, Jerome Friedman** The Elements of Statistical Learning: Data Mining, Inference, and Prediction Springer, 2nd Edition, 2009. ISBN: 978-0-387-84858-7
2. **Ian Goodfellow, Yoshua Bengio, Aaron Courville** Deep Learning MIT Press, 2016. ISBN: 978-0-262-03561-3 (Excellent for understanding RBMs, CNNs, RNNs, and Bayesian techniques)
3. **Kevin P. Murphy** Machine Learning: A Probabilistic Perspective MIT Press, 2012.ISBN: 978-0-262-01802-9 (Covers probabilistic models and Bayesian learning in depth)
4. **MichelTom(1997)**MachineLesning.TanMcGraw-Hill
5. **C.M. BISHOP (2006),** Patem Recognitions and Machat Looming Springer-VerlagNewYork,1stEden
6. **Steven Bird Ewan Klen, Edward Loper** (2009) Natural Language Processing with Python: Analysing Text with the Natural, Language Toolkit, O'RELLY publications.

Internal Assessment Pattern

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)

L1	35	--
L2	40	--
L3	25	40
L4	--	35
L5	--	15
L6	--	10
Total (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

L1: Remember

1. Define bagging and boosting.
2. What is a weak learner?
3. What is a Restricted Boltzmann Machine (RBM)?
4. List any two applications of Markov Random Fields.
5. DefineL1andL2regularization.
6. What is early stopping?
7. What is the Box-Jenkins methodology?
8. Define ARIMA.
9. What is tokenization?
10. List applications of NLP.

L2: Understand

1. Explain the bias-variance tradeoff in ensemble learning.
2. Describe how Random Forests work.
3. Explainthe working of Hidden Markov Models (HMM).
4. Describe contrastive divergence used in RBMs.
5. Explain how regularization helps in reducing over fitting.
6. Differentiate between sparse coding and dropout.
7. Explain trend, cyclical, and seasonal components in time series.
8. Describe the concept of auto-correlation.
9. Explain text summarization and text simplification.
10. Describe the steps involved in an NLP pipeline.

L3: Apply

1. Implement a voting ensemble classifier on a given data set using scikit-learn.

2. Apply bagging and boosting techniques on a classification problem and compare results.
3. Apply HMM to model part-of-speech tagging for a sentence.
4. Implement a simple RBM for binary image reconstruction.
5. Apply L1 and L2 regularization to a linear regression model using a data set.
6. Implement drop out in a deep neural network using Tensor Flow/Keras.
7. Apply Holt-Winters method for sales fore casting.
8. Fit an ARIMA model to predict stock market data.
9. Implement a text classification model using supervised learning.
10. Apply Named Entity Recognition (NER) using spa Cyor NLTK.

L4: Analysing

1. Analyze the impact of hyper parameters on the performance of XG Boost.
2. Compare and contrast the performance of AdaBoost, Gradient Boosting, and Random Forest on noisy data.
3. Analyze the role of mini-batch size in RBM training.
4. Compare the characteristics of MRF sand HMMs.
5. Analyze the effect so fover-regularization on model performance.
6. Examine the trade-offs between bias and variance in regularized models.
7. Analyze the significance of smoothing in time series fore casting.
8. Compare moving averages and exponential smoothing techniques.
9. Analyze the performance of supervised vs unsupervised approaches in text classification.
10. Compare traditional NLP techniques with deep learning-based methods.

L5: Evaluating

1. Critically evaluate the use of ensemble methods in healthcare predictive modeling.
2. Justify the choice of Cat Boost over other ensemble methods in handling categorical features.
3. Evaluate the effectiveness of Bayesian Neural Networks in uncertain environments.
4. Assess the advantages and limitations of probabilistic graphical models.
5. Justify the use of Bayesian modeling over traditional ML in high uncertainty scenarios.
6. Evaluate the benefits of Gaussian Processes in regression tasks.
7. Evaluate the effectiveness of ARIMA versus machine learning models for time series forecasting.
8. Justify the selection of forecasting methods for financial data sets.
9. Evaluate the effectiveness of information retrieval models in question answering systems.
10. Assess the challenges in applying NLP techniques to multilingual datasets.

L6: Create and Design:

1. Design an end-to-end NLP pipeline for a customer support Chabot. Include stages like tokenization, POS tagging, NER, text classification, and QA. Justify your architecture choices.
2. Create a Named Entity Recognition (NER) model for legal or medical documents. Define the entity classes, explain your training strategy, and propose evaluation metrics.
3. Develop a question answering system for a university FAQ page. Incorporate retrieval-based and generative techniques, and justify which method is more effective for your use case.
4. Propose an NLP-based text simplification tool for language learners. Explain the algorithm and how it balances grammatical integrity and vocabulary simplicity.
5. Design a news summarization system using both extractive and abstractive techniques. Compare their results and evaluate them using ROUGE or BERT Score.
6. Create a pipeline for information retrieval from a large corpus (e.g., Wikipedia articles). Integrate indexing, tokenization, victimization, and retrieval models like BM25 or transformer-based retrievers.

Chairperson
Board of Studies (CSE)

R24DST404 TIME SERIES ANALYSIS AND FORECASTING 3 0 0 3

(Honors/Minor Course: Data Science)

Course Objectives:

The main objectives of the course is to

1. Understand the Basics of Time Series Analysis.
2. Learn the Foundations of Time Series Model Building.
3. Gain Proficiency in Time Series Forecasting Models.
4. Explore Advanced Time Series Models and Techniques.
5. Apply Time Series Models Using Software Tools.

Course Outcomes

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

Course Code	Course Outcomes (COs)	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	Do k
R24DST404.1	Understand Time Series Data	3	2	-	-	2	1	-	-	-	-	2	3	2	L1, L2
R24DST404.2	Build Time Series Models and Test for Stationary	3	2	3	-	3	-	-	-	-	-	-	2	3	L2, L3
R24DST404.3	Apply Forecast with ARIMA, SARIMA, and Exponential Smoothing Models	3	-	3	-	3	1	-	-	-	-	2	3	2	L3, L4
R24MB301.4	Evaluate Advanced Time Series Models and Techniques	2	2	1	3	3	-	-	2	2	3	2	2	3	L4 L5
R24DST404.5	Implement Real-World Time Series Forecasting Applications	2	2	3	-	3	2	-	-	-	-	2	3	2	L5, L6

SYLLABUS

UNIT I **14 Hours**

Introduction to Time Series Analysis: Definition and Importance of Time Series Data, Applications in Business and Economics, Components of Time Series: Trend, Seasonal, Cyclical, Irregular, Time Series Data Visualization and Smoothing Techniques, Time Series vs. Cross-Sectional Data.

CO's-CO1

Self -Learning Concepts: Explore Time Series in Business Contexts.

UNIT II **12 Hours**

Stationary and Model Building: Concept of Stationary and Non-Stationary, Differencing and Transformation, Autocorrelation Function (ACF) and Partial ACF (PACF), Identification of AR,MA, ARMA, and ARIMA models, Model Selection Criteria (AIC, BIC) **CO's-CO2**

Self- Learning Topic: Autocorrelation Function (ACF) and Partial ACF (PACF)

UNIT III **10 Hours**

Time Series Forecasting Models: Autoregressive Integrated Moving Average (ARIMA) Models, Seasonal ARIMA (SARIMA) Models, Exponential Smoothing Methods: Simple Exponential Smoothing, Holt's Linear Trend Model, Holt-Winters Seasonal Model, Model Diagnostics and Forecast Accuracy Measures (MAE, RMSE, MAPE) **CO's-CO3**

Self- Learning Topic: Model Diagnostics and Forecast Accuracy Measures

UNIT IV **13 Hours**

Advanced Time Series Models and Techniques: ARCH and GARCH Models for Volatility Forecasting, Vector Auto regression (VAR) Models, Intervention Analysis and Transfer Function Models, State-Space Models and Kalman Filter (Introductory), Handling Multivariate Time Series **CO's-CO4**

Self- Learning Topic: Vector Auto regression (VAR) Models

UNIT V **14 Hours**

Applications and Software Implementation: Time Series Applications in Finance, Marketing, Operations, and HR, Case Studies: Retail Sales Forecasting, Stock Market Prediction, Demand, Forecasting, Time Series in Python (pandas, stats models, Prophet) and R (forecast, ts, tseries), Project Work: Real-Time Forecasting and Analysis using Software Tools, Ethical Considerations in Forecasting and Data Interpretation **CO's-CO5**

Self- Learning Topic: Time Series Forecasting in Python and R

Board of Studies: Computer Science and Engineering

Approved in BOS No: 02, 9th May, 2025

Approved in ACM No: 02

Expert Talk (To be Delivered by SMEs from Industries) **CO's/ PO's / PSO's**

1. Applications of Time Series Analysis in Business and Economics CO1-PO1,PO2,PO5,PSO1,PSO2
2. Building Predictive Models: Stationary and Model Identification in Time Series CO2-PO2, PO6, and PSO2.
3. Advances in Time Series Forecasting: ARIMA, SARIMA, and Exponential Smoothing Techniques CO3-PO3,PO4,PO5,PSO3
4. Real-World Applications and Software Implementation in Time Series Forecasting CO5-PO3,PO5,PO9,PSO1,PSO3.

Text books:

1. Time Series Analysis: Forecasting and Control, George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, 5th Edition, Wiley.
2. "Forecasting: Principles and Practice", Rob J Hyndman, George Athanasopoulos, 3rd

Edition (Free online version), OTexts.

3. "Introductory Time Series with R", Paul S.P. Cowpertwait, Andrew V. Metcalfe, 2nd Edition, Springer
4. "Time Series Analysis and Its Applications: With R Examples", Shumway, R.H., Stoffer, D.S. 4th Edition, Springer
5. "Applied Econometric Time Series", Walter Enders, 5th Edition, Wiley.

References:

1. "Time Series: A Data Analysis Approach Using R", Jonathan D. Cryer, Kung-Sik Chan, 2nd Edition.
2. "The Analysis of Time Series: An Introduction", Chris Chatfield, 6th Edition
3. "Hands-On Time Series Analysis with R: Build effective models for forecasting, anomaly detection, and classification using R", Rami Krispin, Packt Publishing.

Web references:

1. **https://robjhyndman.com/.**
2. https://stackoverflow.com/questions/tagged/time-series.

Internal Assessment Pattern

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)
L1	17	--
L2	33	--
L3	33	17
L4	17	33
L5	--	33
L6	--	17
Total (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

L1 Remembering

1. What are the key components of time series data?
2. Define time series and give an example of time series data in business.
3. What is the difference between stationary and non-stationary time series data?
4. Define differencing in the context of time series analysis.
5. What is the difference between ARIMA and SARIMA models?

6. Define the term "Exponential Smoothing."
7. Define ARCH and GARCH models.
8. What is a Vector Autoregression (VAR) model, and in what scenarios is it used?
9. What are the common software tools used for time series analysis?
10. What are the ethical considerations when interpreting time series data?

L2 Understanding

1. Explain the differences between trend, seasonal, cyclical, and irregular components of time series data.
2. How does time series data differ from cross-sectional data?
3. Explain how the Autocorrelation Function (ACF) and Partial ACF (PACF) help in identifying the appropriate model for time series data.
4. Explain the principle behind Exponential Smoothing and how it differs from the ARIMA model.
5. Explain the working mechanism of GARCH models for volatility forecasting.
6. Explain how time series forecasting models are implemented in Python and R.
7. What are the major applications of time series analysis in business?
8. Explain the concept of differencing in time series analysis.
9. How does the ACF and PACF help in identifying the appropriate time series model?
10. Describe the key concepts behind Holt-Winters Exponential Smoothing

L3 Applying

1. Illustrate how smoothing techniques can be applied to time series data to identify the trend component.
2. Apply the concept of differencing to make a non-stationary time series stationary.
3. Use ACF and PACF plots to identify potential AR and MA models.
4. Apply the Holt-Winters Seasonal model to forecast future values for a given time series dataset.
5. Demonstrate the steps to forecast using ARIMA and SARIMA models
6. Apply an ARCH/GARCH model to forecast volatility in financial time series data.
7. Use VAR models to analyze the relationship between multiple time series variables.
8. Use Python (pandas, statsmodels, Prophet) or R (forecast, ts) to implement a forecasting model for a business dataset.
9. Use ACF and PACF plots to identify whether an AR, MA, or ARMA model is suitable for a given time series.
10. Given a time series, apply differencing to make the series stationary.

L4 Analyzing (Analysis Level)

1. Analyze a given time series dataset to separate the trend, seasonal, and irregular components using graphical methods
2. Analyze a given time series data and determine whether it is stationary or non-stationary.
3. Identify the most appropriate model (AR, MA, ARMA, ARIMA) based on the ACF and PACF plots.
4. Analyze a given time series dataset using ARIMA and SARIMA models. Compare their performance based on forecasting accuracy.
5. Analyze the volatility trends in a given financial time series dataset using GARCH models.
6. Analyze the outcomes of a real-time forecasting project and discuss how well the model performed in predicting future trends.
7. Analyze a non-stationary time series to identify appropriate transformations and test for stationary.
8. Analyze the ACF and PACF of a given series and suggest the model to use (AR, MA, ARMA, or ARIMA).
9. Identify when to use a simple Exponential Smoothing model versus a Holt-Winters model.
10. Assess the impact of ethical issues on the interpretation and application of forecasting results.

L5 Evaluating

1. Evaluate the effectiveness of different time series visualization techniques in understanding the underlying patterns in data.
2. Evaluate the performance of an ARIMA model against other time series models based on error metrics such as AIC and BIC
3. Evaluate the forecast accuracy of different models (ARIMA, SARIMA, Holt-Winters) using error measures such as MAE, RMSE, and MAPE.
4. Evaluate the performance of a GARCH model versus an ARIMA model for forecasting volatility in financial data.
5. Critically evaluate the use of time series forecasting in business contexts and suggest improvements or alternatives based on real-world case studies.
6. Evaluate the performance of different ARIMA models based on selection criteria such as AIC and BIC.
7. Critically assess the forecasting model's performance in a business context using real-time data.
8. Evaluate the effectiveness of different forecasting software tools (Python vs. R) for implementing time series models.

9. Evaluate the effectiveness of GARCH models compared to simpler time series models like ARIMA for volatility forecasting.
10. Critically assess the applicability of VAR models in different business scenarios.

L6 Creating

1. Create visualization for a time series dataset, showing trend, seasonal, and irregular components.
2. Design an ARIMA model for a given time series dataset. Justify your choices of parameters for the model.
3. Construct a forecasting model using Exponential Smoothing for a time series dataset and generate predictions.
4. Develop a multivariate time series model using VAR to understand the interaction between economic indicators.
5. Create a comprehensive time series forecasting model using software tools (Python/R) for a given business case study (e.g., retail sales or stock market prediction).

Chairperson
Board of Studies (CSE)

R24DST201 Statistical Analysis and Programming for Data Science 3 0 0 3

(Honors/Minor Course: Data Science)

Course Objectives:

1. Be prepared with a varied range of expertise in different aspects of data science such as data collection, visualization, processing, and modeling of data sets. Should be able to select data, model, model complexity etc.

Course Outcomes

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

Course Code	Course Outcomes	Mapping with POs and PSOs							Dok
		PO1	PO2	PO3	PO4	PS01	PS02	PS03	
R24DST201.1	Identify the need for data science and solve basic problems using Python built-in data types and their methods.	3	2	3	2	3	3	3	L1,L2
R24DST201.2	Design an application with user-defined modules and packages using OOP concept.	2	3	2	3	3	2	3	L2,L3,L4
R24DST201.3	Employ efficient storage and data operations using NumPy arrays.	3	3	3	1	3	-	2	L3,L4
R24DST201.4	Apply powerful data manipulations using Pandas.	2	3	3	3	2	2	-	L4, L5
R24DST201.5	Develop data pre-processing and visualization using Pandas.	3	2	2	2	3	3	-	L4,L5

SYLLABUS

UNIT-I: Descriptive Statistics and Probability Distributions: **15 Hours**

Statistical Data – Categorical, Numerical (Continuous), Univariate and Bivariate Analysis, Mean, Median, Mode, Standard Deviation.

Descriptive statistics: Qualitative and quantitative Variable, discrete variable, population, sample, random sample.

Probability Distributions: Binomial Distribution, Poisson distribution, Uniform Distribution, Normal Distribution. **CO's–CO1**

Self-Learning Topics: Exponential Distribution

UNIT-2: Introduction to Data Science and Python Programming: **15 Hours**

Introduction to Data Science – Why Python? – Essential Python libraries Python Introduction – Features, Identifiers, Reserved words, Indentation, Comments Built-in Data Types and their Methods: Strings, List, Tuples, Dictionary, Set Type Conversion, Operators, Decision Making,

Looping: Loop control statements Math and Random number functions **CO's–CO2**

Self-Learning Topics: User-defined functions – function arguments & its types.

UNIT-3: Introduction to Numpy: **15 Hours**

NumPy Basics: Arrays and Vectorized Computation – The NumPy ndarray – Creating ndarrays – Data Types for ndarrays – Arithmetic with NumPy Arrays – Basic Indexing and Slicing – Boolean Indexing – Transposing Arrays and Swapping Axes. **CO's–CO3**
Self-Learning Topics: Universal Functions

UNIT-4: Introduction to pandas Data Structures: **15 Hours**
Pandas Data Structures Series, Data Frame
Essential Functionality: Dropping Entries, Indexing, Selection, and Filtering – Function Application and Mapping – Sorting and Ranking. **CO's–CO4**

Self-Learning Topics: Summarizing and Computing Descriptive Statistics

UNIT-5: Data Cleaning and Preparation: **15 Hours**
Data Cleaning and Preparation: Handling Missing Data – Data Transformation: Removing Duplicates, Transforming Data Using a Function or Mapping, Replacing Values, Detecting and Filtering Outliers.

String Manipulation: Vectorized String Functions in pandas. **CO's–CO5**
Self-Learning Topics: Plotting with pandas

Board of Studies : Computer Science and Engineering

Approved in BOS No: 02, 9th May, 2025
Approved in ACM No: 02

Expert Talk (To be Delivered by SMEs from Industries) **CO's/ PO's / PSO's**

1. Vectorized String Functions in pandas CO5,PO1,PO2,PO3,PO4,PSO1,PSO2

Text Books:

1. Y. Daniel Liang, Introduction to Programming using Python, Pearson, 2012.
2. Wes McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly, 2nd Edition, 2018.
3. Jake VanderPlas, Python Data Science Handbook: Essential Tools for Working with...

Reference Books:

1. Wesley J. Chun, *"Core Python Programming"*, Prentice Hall, 2006.
2. Mark Lutz, *"Learning Python"*, O'Reilly, 4th Edition, 2009.

Web References:

1. https://www.edx.org/course/python-basics-for-data-science
2. https://www.edx.org/course/analyzing-data-with-python
3. https://www.coursera.org/learn/python-plotting?specialization=data-science-python

Internal Assessment Pattern:

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)
L1	35	10

L2	30	10
L3	25	20
L4	10	35
L5	--	25
TOTAL (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

L1: Remember

1. Define mean, median, and mode.
2. What is a categorical variable?
3. List types of probability distributions.
4. Define standard deviation.
5. What is the difference between population and sample?

L2: Understand

1. Explain the significance of standard deviation in data analysis.
2. Differentiate between qualitative and quantitative variables.
3. Describe the characteristics of a normal distribution.
4. Explain the concept of a random sample with an example.
5. Compare binomial and Poisson distributions.

L3: Apply

1. Calculate the mean and mode of a given dataset using Python.
2. Use Python to simulate a binomial distribution.
3. Write a program to distinguish between discrete and continuous variables.
4. Use numpy or scipy to compute standard deviation of a list.
5. Apply probability distribution functions to predict outcomes.

L4: Analyze

1. Analyze the shape of a normal vs uniform distribution.
2. Examine the impact of outliers on mean and standard deviation.
3. Identify when to use Poisson vs Binomial distribution.
4. Investigate a dataset to classify variables as qualitative or quantitative.
5. Compare univariate and bivariate analysis with examples.

L5: Evaluating

1. Justify the use of standard deviation over variance in real-time analysis.
2. Evaluate the suitability of the normal distribution in stock price prediction.

3. Critique the limitations of mean in skewed distributions.
4. Assess the effectiveness of different sampling methods.
5. Argue which central tendency measure is best for income data.

Chairperson
Board of Studies (CSE)

R24DST302 **Data Handling and Visualization** **3 0 0 3**

(Honors/Minor Course: Data Science)

Course Objectives:

1. Be prepared with a varied range of expertise in different aspects of data science such as data collection, visualization, processing, and modeling of data sets.

Course Outcomes

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

Course Code	Course Outcomes	Mapping with POs and PSOs							Dok
		PO1	PO2	PO3	PO4	PS01	PS02	PS03	
R24DST302.1	Understand basics of Data Visualization	2	2	3	2	3	3	3	L1, L2
R24DST302.2	Implement visualization of distributions	3	3	2	3	3	2	3	L2, L3, L4
R24DST302.3	Develop programs on visualization of time series, proportions & associations	2	3	3	1	3	-	2	L3, L4
R24DST302.4	Develop programs on visualization of Scatter plots, Mosaic plots	2	3	2	2	3	2	-	L4, L5
R24DST302.5	Apply visualization on Trends and uncertainty	3	2	2	2	3	3	-	L4, L5

SYLLABUS

UNIT-I: Introduction to Visualization: **15 Hours**

Visualizing Data – Mapping Data onto Aesthetics, Aesthetics and Types of Data, Scales Map Data Values onto Aesthetics, Coordinate Systems and Axes – Cartesian Coordinates, Nonlinear Axes, Coordinate Systems with Curved Axes, Color Scales – Color as a Tool to Distinguish, Color to Represent Data Values, Color as a Tool to Highlight **CO's–CO1**

Self-Learning Topics: Diversity of Visualizations – Amounts, Distributions, Proportions, x–y relationships, Geospatial Data

UNIT-2: Visualizing Distributions: **15 Hours**

Visualizing Amounts – Bar Plots, Grouped and Stacked Bars, Dot Plots and Heat maps, Visualizing Distributions – Histograms and Density Plots, Visualizing a Single Distribution, Visualizing Multiple Distributions at the Same Time, Visualizing Distributions: Empirical Cumulative Distribution Functions and Q-Q Plots, Empirical Cumulative Distribution Functions **CO's–CO2**

Self-Learning Topics: Highly Skewed Distributions.

UNIT-3: Visualizing Associations & Time Series 1: 15 Hours

Visualizing Proportions – A Case for Pie Charts, A Case for Side-by-Side Bars, A Case for Stacked Bars and Stacked Densities, Visualizing Proportions Separately as Parts of the Total, Visualizing Nested Proportions – Nested Proportions Gone Wrong, Mosaic Plots and Tree Maps, Nested Pies **CO's–CO3**

Self-Learning Topics: Parallel Sets

UNIT-4: Visualizing Associations & Time Series 2: 15Hours

Visualizing Associations among Two or More Quantitative Variables – Scatter plots, Correlograms, Dimension Reduction, Paired Data Visualizing Time Series and Other Functions of an Independent , Variable – Individual Time Series, Multiple Time Series and Dose–Response Curves, Time Series of Two or More Response Variables **CO's–CO4**

Self-Learning Topics: Build a Scatter plot and suggest dimension reduction

UNIT-5: Visualizing Uncertainty: **15 Hours**

Visualizing Trends – Smoothing, Showing Trends with a Defined Functional Form, Detrending and Time-Series Decomposition, and Visualizing Geospatial Data – Projections, Layers, Choropleth Mapping, Cartograms, Visualizing Uncertainty – Framing Probabilities as Frequencies, Visualizing the Uncertainty of Point Estimates, Curve Fits. **CO's–CO5**

Self-Learning Topics: Hypothetical Outcome Plots

Board of Studies : Computer Science and Engineering

Approved in BOS No: 02, 9th May, 2025

Approved in ACM No: 02

Expert Talk (To be Delivered by SMEs from Industries) CO's/ PO's / PSO's

1. Vectorized String Functions in pandas CO5, PO1, PO2, PO3, PO4, PSO1, PSO2

Text Books:

1. *Claus Wilke*, "Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures"*, 1st edition, O'Reilly Media Inc, 2019.*

Reference Books:

1. Tony Fischetti, Brett Lantz, R: Data Analysis and Visualization, O'Reilly, 2016
2. Ossama Embarak, Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems, Apress, 2018

Web References:

1. https://www.coursera.org/learn/data-visualization
2. https://www.coursera.org/learn/python-for-data-visualization#syllabus

Internal Assessment Pattern:

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)
L1	35	-
L2	30	-
L3	25	35
L4	10	35
L5	--	30
TOTAL (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

L1: Remember

1. What is the purpose of using aesthetics in data visualization?
2. Define Cartesian coordinates and their use in visualizing data.
3. List three types of color scales used in data visualization.
4. What is meant by "mapping data onto aesthetics"?
5. Identify different types of coordinate systems used in data visualization.

L2: Understand

1. Explain how nonlinear axes affect data visualization.
2. Describe the role of color in highlighting data points.
3. Compare Cartesian and curved axes with examples.
4. Discuss how aesthetics influence the readability of a chart.
5. Explain the importance of using appropriate scales in mapping data.

L3: Apply

1. Use a color scale to distinguish between categories in a bar chart.
2. Apply mapping techniques to show data on both x and y axes using a scatter plot.
3. Convert a raw dataset into a visualization using coordinate axes.
4. Implement a visualization showing different data types using appropriate aesthetics.
5. Apply appropriate axis scaling for a dataset containing exponential values.

L4: Analyze

1. Analyze how changing the coordinate system affects the interpretation of a dataset.
2. Distinguish between effective and ineffective use of color in data visualization.
3. Compare multiple types of aesthetic mappings and assess their impact on user perception.
4. Examine the effect of curved vs. linear axes on a time series plot.

5. Analyze visualization and identify flaws in the aesthetic mapping.

L5: Evaluating

1. Evaluate visualization for its effectiveness in conveying trends using color.
2. Critique the choice of coordinate system used in a real-world data visualization example.
3. Justify the use of a particular color scale in a demographic heat map.
4. Assess the quality of visual encoding in dashboard visualization.
5. Evaluate the clarity of a complex plot using multiple aesthetics.

Chairperson
Board of Studies (CSE)

R24DST303 **BIG DATA TECHNOLOGIES** **3 0 0 3**

((Honors/Minor Course: Data Science))

Course Objectives:

1. Understand what big data is and how Big Data Technologies can help organizations achieve a competitive advantage
2. Provide an overview of Apache Hadoop and its ecosystem components.
3. Understand Map Reduce Jobs
4. Processing of Big Data with advanced architectures like Spark.
5. To understand practical machine learning scalable and easy.

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

Course Code	Course Outcomes	Mapping with POs and PSOs							Dok
		PO1	PO2	PO3	PO4	PS01	PS02	PS03	
R24DST303.1	Understand fundamentals of Big Data Technologies	3	2	3	2	3	3	3	L1, L2
R24DST303.2	Investigate Hadoop framework and Hadoop Distributed File system.	3	3	3	3	3	2	3	L2, L3
R24DST303.3	Demonstrate the Map Reduce programming model to process the big data along with Hadoop tools.	3	3	3	1	3	2	2	L3, L4
R24DST303.4	Implement Big Data code in Apache Spark (in PySpark).	3	3	3	3	2	2	-	L4, L5
R24DST303.5	Run Supervised and Unsupervised machine learning on Large-Scale Data.	3	2	2	2	3	3	-	L4, L5

SYLLABUS

UNIT-I: Introduction to Big Data Analytics **15 Hours**

Introduction to Big Data Analytics: Big Data, Scalability and Parallel Processing, Designing Data Architecture, Data Sources, Quality, Pre-Processing and Storing, Data Storage and Analysis, Big Data Analytics Applications and Case Studies

Self-Learning Topics: Data Analytics Applications and Case Studies **CO's-CO1**

UNIT–II: Introduction to Hadoop **13 Hours**

Introduction to Hadoop: Introduction, Hadoop and its Ecosystem, Hadoop Distributed File System, Map Reduce Framework and Programming Model, Hadoop Yarn, Hadoop Ecosystem Tools.

Hadoop Distributed File System Basics: HDFS Design Features, Components, HDFS User Commands. Hadoop Ecosystem Components: Using Apache Pig, Hive, Sqoop, Flume, Oozie.

Self-Learning Topics: HBase. **CO's–CO3**

UNIT-III: Map Reduce, Hive and Pig **14 Hours**

Map Reduce, Hive and Pig: Introduction, Map Reduce Map Tasks, Reduce Tasks and Map Reduce Execution, Composing Map Reduce for Calculations and Algorithms, Hive,

Self-Learning Topics: HiveQL, Pig. **COs–CO3**

UNIT- IV: Large-Scale Data Processing with PySpark **13 Hours**

Apache Spark, Spark programming. (Python and PySpark), RDDs, Data Frames, Spark SQL, PySpark, NumPy, SciPy, Code Optimization, Cluster Configurations, Linear Algebra Computation in Large Scale,

Self-Learning Topics: Distributed File Storage Systems. **COs–CO4**

UNIT-V: Large Scale Machine Learning with Spark **15 Hours**

Basic statistics, Data sources, Pipelines, Extracting, transforming and selecting features, Classification and Regression, Clustering, Collaborative filtering, Frequent Pattern Mining.

Self-Learning Topics: Model selection and tuning. **COs–CO5**

Board of Studies: Computer Science and Engineering
Approved in BOS No: 02, 9th May, 2025
Approved in ACM No: 02

Expert Talk (To be delivered by SMEs from Industries)

	COs	**POs / PSOs**
1. CRISP-DM Methodology	CO5	PO1,PO2,PO3,PO4,PSO1,PSO2

Text Books:

1. Raj Kama! and Preeti Saxena, "Big Data Analytics Introduction to Hadoop, Spark, and Machine-Learning", McGraw Hill Education, 2018, ISBN: 9789353164966, 9353164966
2. Douglas Eadline, "Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem", 1st Edition, Pearson Education, 2016. ISBN13: 978 9332570351

Reference Books:

1. Tom White, "Hadoop: The Definitive Guide", 4th Edition, O"Reilly Media, 2015.ISBN-13: 978- 9352130672
2. Perrin, J. (2020). Spark in action (2nd ed.). (Covers Apache Spark 3 with examples in Java, Python, and Scala) O'Reilly Media Inc.
3. Arshdeep Bahga, Vijay Madisetti, "Big Data Analytics: A Hands-On Approach", 1st Edition, VPT Publications, 2018. ISBN-13: 978-0996025577
4. Damji, J., Wenig, B., Das, T., Lee, D. (2020). Learning spark (2nd ed.) O'Reilly Media Inc.
5. Nudurupati, S. (2021). Essential PySpark for scalable data analytics: A beginner's guide to harnessing the power and ease of PySpark 3 Packt Publishing

Web References:

1. https://www.coursera.org/learn/introduction-to-big-data-with-spark-hadoop
2. https://www.udemy.com/course/the-ultimate-hands-on-hadoop-tame-your-big-data/?couponCode=IND21PM

Internal Assessment Pattern:

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)
L1	35	10
L2	30	10
L3	25	20
L4	10	35
L5	--	25
TOTAL (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

L1: Remember

1. What is Big Data and how is it different from traditional data?
2. List the basic components of Hadoop Distributed File System (HDFS).
3. Define Map and Reduce tasks in the Map Reduce programming model.
4. What is an RDD in PySpark?
5. Name the different types of machine learning algorithms supported in Spark MLlib.

L2: Understand

1. Explain the importance of data quality and pre-processing in big data analytics.
2. Describe how Apache Hive simplifies querying data in Hadoop.
3. Explain how MapReduce processes large datasets using the Map and Reduce phases.
4. Illustrate the difference between RDDs and Data Frames in PySpark.
5. Explain the concept of collaborative filtering in recommendation systems.

L3: Apply

1. Given a real-world problem, outline basic big data architecture to solve it.
2. Use Sqoop to import data from a MySQL database into HDFS. Write the command you would use.
3. Write a simple MapReduce job to count the frequency of words in a text file.
4. Write a PySpark program to load a CSV file and calculate the average of a numerical column.
5. Create a pipeline in Spark MLlib to perform classification on a given dataset.

L4: Analyze

1. Analyze how different data sources impact the design of a big data analytics solution.
2. Compare and contrast Apache Pig and Hive in terms of use cases and architecture.
3. Analyze a HiveQL query and explain how it is internally converted to MapReduce tasks.
4. Analyze the impact of partitioning and caching on the performance of a PySpark application.
5. Analyze the strengths and weaknesses of clustering vs. classification in large datasets.

L5: Evaluating

1. Evaluate the role of big data analytics in decision-making with the help of a case study.

2. Critically evaluate the suitability of Hadoop for real-time data processing applications.
3. Evaluate when to use Pig over Hive for big data processing.
4. Evaluate the performance of NumPy vs. PySpark for large-scale linear algebra computations.
5. Evaluate different model selection techniques in Spark and determine the best approach for a given problem.

Chairperson
Board of Studies (CSE)

R24AMT404 **Deep Learning** **3 0 0 3**

(Honors/Minor Course: Artificial Intelligence & Machine Learning)

Course Objectives:

1. To introduce the fundamental concepts of neural networks and their role in building intelligent systems.
2. To develop an understanding of deep learning and how it differs from traditional machine learning and artificial intelligence.
3. To explore the structure and working of Convolutional Neural Networks (CNNs) and how they are used for image-related tasks.
4. To examine advanced deep learning architectures including Recurrent Neural Networks (RNNs), LSTM, GRU, and Generative Adversarial Networks (GANs).
5. To familiarize students with modern CNN architectures such as Alex Net, Dense Net, and the use of transfer learning.

Course Outcomes

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

Course Code	Course Outcomes (COs)	PO1	PO2	PO3	PO4	PO5	PO11	PSO1	PSO2	Dok
R24AMT404.1	Understand the fundamentals of neural networks and their training techniques.	3	2	-	2	3	-	3	2	L1, L2
R24AMT404.2	Analyze the concepts of deeplearning and differentiate between AI, ML, and DL approaches.	3	3	2	-	3	-	3	2	L1, L2 L3
R24AMT404.3	App CNN chitectures and explore modern CNN models such as Alex Net, Dense Net, and PixelNet.	3	3	3	2	3	-	3	3	L2, L3
R24AMT404.4	Examine various deep learning architectures like RNNs, LSTMs, GRUs, and GANs.	3	3	3	-	3	-	3	3	L3, L4
R24AMT404.5	Implement deep learning techniques for real-world applications like object detection, image captioning, and video-text conversion.	3	3	3	3	3	2	3	3	L5

SYLLABUS

UNIT I: The Neural Network **15 Hours**

Introduction, Building Intelligent Machines, the Neuron, Expressing Linear Perceptions as Neurons, Linear Neurons and Their Limitations, Sigmoid. Feed-Forward Neural Networks, Training Feed-Forward Neural Networks, Back propagation Algorithm. **CO's –CO1**

Self-Learning Topics: Compare sigmoid with ReLU, tanh, etc., using plots

UNIT II: Deep Learning **15 Hours**

Introduction, History, Compare artificial intelligence, machine learning with deep learning, Kinds of Machine Learning problems, deep learning applications. **CO's-CO2**

Self-Learning Topics: how self-driving cars use Deep Learning.

UNIT III: Convolution Neural Networks: **15 Hours**

Introduction to CNNs, Convolution, Correlation, Filtering, Convolution Layer, and CNN architectures, Detection and Segmentation, Convolutions for image. Modern Convolution Neural Networks: Deep Convolution Neural Networks (Alex Net), Alex Net Architecture, Variants of CNN: Dense Net, PixelNet **CO's–CO3**

Self-Learning Topics: Compare architectures visually using diagrams.

UNIT IV: **Deep Learning Architectures** **13 Hours**

Recurrent Neural Networks (RNNs), Advanced RNN: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), Generative Adversarial Networks (GANs), Advanced GANs. **CO's–CO4**

Self-Learning Topics: Use a pre-trained GAN to generate faces or digits.

UINT V: **Applications of Deep Learning**: **13 Hours**

Image segmentation, Object Detection, Automatic Image Captioning, Image generation with Generative Adversarial Networks, Video to Text with LSTM models, Computer Vision Basics.

CO's-CO5

Self-Learning Topics: Explore sequence modeling for videos

Board of Studies: Computer Science and Engineering

Approved in BoS No: 02, 9th May, 2025

Approved in ACM No:

Text Books:

1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, "Deep Learning", MIT Press, 2017.
2. Josh Patterson, Adam Gibson, "Deep Learning: A Practitioner's Approach", O'Reilly Media, 2017.
3. François Chollet, "Deep Learning with Python", Manning Publications, 2018.

Reference Books:

1. Bengio, Yoshua. *"Learning Deep Architectures for AI."* Foundations and Trends in Machine Learning, 2.1 (2009): 1127.

Web References:

1. https://cs229.stanford.edu/notes2020spring/cs229-notes-deep_learning.pdf
2. http://perso.ens-lyon.fr/jacques.jayez/Cours/Implicite/Fundamentals_of_Deep_Learning.pdf

Internal Assessment Pattern

Cognitive Level	Internal Assessment #1(%)	Internal Assessment #2(%)
L1	35	--
L2	40	--
L3	25	40
L4	--	35
L5	--	25
Total (%)	**100**	**100**

Sample Short and Long Answers questions of Various Cognitive Levels

L1: Remember

1. Define a neuron in the context of artificial neural networks.
2. What is the function of an activation function in a neural network?
3. Name the main parts of a neuron model.
4. What is the sigmoid function used for in neural networks?
5. Mention the main steps involved in training a feed-forward neural network?
6. Who introduced the concept of deep learning?
7. What is the difference between AI, ML, and DL?
8. When was deep learning first introduced?
9. Mention one real-world use of deep learning.
10. What does "depth" refer to in deep learning?

L2: Understand

1. Explain how neural networks help in building intelligent machines?
2. Explain the flow of data in a feed-forward neural network.
3. Describe the process of training a feed-forward neural network.
4. Explain the concept of deep learning in simple terms.
5. Describe the types of machine learning problems with examples.
6. Describe the relationship between data availability and deep learning performance?
7. Compare artificial intelligence, machine learning, and deep learning?
8. Explain the role of filtering in convolution neural networks.
9. Explain how convolutions are applied in image processing tasks.
10. Interpret how feature maps are generated in convolution layers.

L3: Apply

1. Apply the concept of deep learning to design a model for voice recognition.

2. Demonstrate how you would classify spam emails using a deep learning model.
3. Implement a deep learning framework to identify handwritten digits?
4. Apply the concept of convolution to enhance image edges in a CNN.
5. Use Alex Net to classify a dataset of animal images.
6. Demonstrate how CNNs can be applied for real-time object detection.
7. Implement an LSTM model to analyze and generate text.
8. Use GANs to generate synthetic faces from random noise input.

L4: Analysing

1. Analyze the difference between LSTM and GRU in terms of memory management.
2. Break down the working of a GAN into generator and discriminator roles.
3. Analyze the differences between image segmentation and object detection in terms of feature extraction and use-cases.
4. Analyze how computer vision techniques are used to enhance object detection accuracy in autonomous vehicles.

L5: Evaluating

1. Evaluate the performance of different deep learning models for object detection (e.g., YOLO vs. Faster R-CNN).
2. Justify the use of GANs for image generation over traditional augmentation techniques.
3. Critically assess the accuracy and relevance of captions generated by an image captioning model.
4. Assess the limitations of current deep learning models in understanding and describing video content**.**

Chairperson
Board of Studies (CSE)