Course Structure for T.Y.B.Sc. (CS) Statistics As per NEP 2.0 (2024 Pattern)

Sem.	Course Type	Course Code	Course Name	Theory / Practical	Credits
V	Minor	COS-310-MN(A)	Predictive Analytics	Theory	02

CBCS Syllabus as per NEP 2020 for T.Y.B.Sc. Computer Science (2024 Pattern)

Name of the Programme	**:** B.Sc. Computer Science
Program Code	**:** COS
Class	**:** T.Y.B.Sc. (Computer Science)
Semester	**:** V
Course Type	**:** Minor (Theory)
Course Name	**:** Predictive Analytics
Course Code	**:** COS-311-MN (A)
No. of Credits	**:** 2
No. of Teaching Hours	**:** 30

Course Objectives:

1. To understand the concepts of correlation, regression, and time series analysis and their applications.
2. To analyze relationships between variables using correlation techniques.
3. To learn and apply the least squares method for fitting linear, quadratic, and exponential models for predictive analytics.
4. To develop a strong foundation in regression modeling, including parameter estimation, residual analysis, and model evaluation.
5. To explore different time series components and apply appropriate smoothing and forecasting techniques.
6. To implement autoregressive models for real-world forecasting problems in finance, business, and environmental sciences.
7. To apply correlation, regression, and time series methods in Data Science applications like stock market prediction, sales forecasting, and anomaly detection.

Course Outcomes:

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

CO1. Students will be able to compute and interpret correlation coefficients for understanding relationships in bivariate data.

CO2. Students will be able to apply correlation techniques for feature selection, dimensionality reduction, and exploratory data analysis in machine learning.

CO3. Students will be able to fit linear, quadratic, and exponential models using the least squares method for predictive analytics.

CO4. Students will be able to estimate regression parameters, interpret coefficients, and validate models using residual analysis and R^2.

CO5. Students will be able to analyze time series data, identify trends and seasonality, and apply appropriate smoothing techniques.

CO6. Students will be able to implement autoregressive models and forecasting techniques for real-world time series applications.

CO7. Students will be able to utilize statistical modeling techniques for data-driven decision-making across various domains in Data Science.

Topics and Learning Points

UNIT– 1 Correlation and Its Applications (05L)

Bivariate data**,** bivariate frequency distribution**,** Concept of correlation between two variables: positive correlation, negative correlation, no correlation**,** Importance of correlation in data science and machine learning**,** Interpretation of correlation in the context of feature selection and exploratory data analysis (EDA)**,** Scatter diagram**,** interpretation of the type of correlation from a scatter diagram, Real-world applications of correlation in data-driven decision-making.**,** Covariance between two variables: Definition, computation, the effect of change of origin and scale, Karl Pearson's coefficient of correlation (r)**:** Definition, computation for ungrouped and grouped data**,** and interpretation.

Properties

(i) $-1 \leq r \leq 1$ (without proof).

(ii) Effect of change of origin and scale (without proof).

Spearman's rank correlation coefficient**:** Definition, derivation of formula, computation, and interpretation (without ties). In case of ties, compute Karl Pearson's correlation coefficient between ranks**,** applications of correlation in Data Science

UNIT– 2 Regression Analysis **(12L)**

Understanding the concept of regression. Difference between regression and correlation. Concept of dependent (response) and independent (predictor) variables. Regression equation: $Y=\beta_0 + \beta_1 X+ \varepsilon$. Assumptions of simple linear regression. Estimation of Model Parameters. Concept of residuals, coefficient of determination R^2, Adjusted R^2. Hypothesis Testing in Simple Linear Regression: Testing for significance of regression coefficients, ANOVA for regression. Checking assumptions: Linearity, Homoscedasticity (constant variance of errors), and Normality of residuals. Introduction to multiple regression, Regression equation: $Y=\beta_0 + \beta_1 X_1+ \beta_2 X_2+ \beta_3 X_3+\ldots. \beta k X_k +\varepsilon$. Assumptions of simple linear regression. real-world applications. Non-linear regression: Second-degree curve $Y=a + bX + cX^2$, and exponential models $Y= ab^X$, $Y= aX^b$, Applications in predictive modeling

UNIT– 3 Logistic Regression **(05L)**

Understanding classification problems in data science Difference between linear and logistic regression. Concept of odds and log-odds, Logistic function and its interpretation. Applications of logistic regression in real-world scenarios. Extending logistic regression to multiple predictors. Evaluating model performance with classification metrics.

UNIT– 4 Time Series Analysis **(08L)**

Introduction, meaning, and importance of time series in Data Science, Definition, Components of Time Series: i) The Trend ii) Seasonal variation iii) Cyclical variation iv) Irregular variation, Additive and Multiplicative Model, Methods of trend estimation and smoothing: (i) moving average, (ii) curve fitting by least square principle, (iii) exponential smoothing.

References:

1. Gupta and Kapoor : Fundamentals of Mathematical Statistics, Sultan Chandand Sons, New Delhi.
2. Sharma K. V. S. (2001) Statistics made it simple: Do it yourself on PC. Prentce Hall of India, New Delhi.
3. Gupta S. C. and Kapoor V. K. (1987): Fundamentals of Applied Statistics, Sultan Chand and Sons, New Delhi.
4. B. L. Agarwal : Programmed Statistics, New Age International Publishers, NewDelhi.
5. Gupta. S.P.: Statistical Methods,SultanChandandSons,23,Daryaganj,New Delhi
6. Mukhopadhyay Parimal(1999):Applied Statistics, New Central Book Agency, Pvt. Ltd. Calcutta.
7. Moore D. S., Norz W. I, Flinger M. A., (2013), The Basic Practice of Statistics, Sith Edition, Freeman and Company New York
8. Brase C.H. and Brase C.P,(Brockwell, P.J.,& Davis, R.A.(2016).Introduction to Time Series and Forecasting. Springer

Programme Outcomes and Course Outcomes Mapping:

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PO13
CO1	3	2	1	2	2	1	2	1	2	1	1	1	1
CO2	3	3	1	1	2	1	2	1	2	1	1	1	1
CO3	2	3	2	3	3	2	3	2	3	1	1	2	1
CO4	2	3	1	3	3	1	2	1	3	1	1	1	1
CO5	2	3	1	3	3	1	2	1	3	1	1	1	1
CO6	2	3	1	3	3	1	2	1	3	1	1	2	1
CO7	2	3	2	3	3	2	3	2	3	1	1	2	1
CO8	2	3	3	3	3	2	3	2	3	1	1	3	1

Justification

PO1: Comprehensive Knowledge and Understanding

The course outcomes (CO1–CO3) strongly support PO1 by developing a solid conceptual foundation in data types, data sources, big data characteristics, and data science workflows. CO4–CO8 further strengthen theoretical understanding through multivariate analysis, distance measures, statistical modeling, and feature engineering within a multidisciplinary data science context.

PO2: Practical, Professional, and Procedural Knowledge

CO2–CO8 strongly align with PO2 as they emphasize hands-on data handling, industry-relevant analytical techniques, data preprocessing, modeling, and evaluation practices. Students gain practical exposure to professional standards, tools, and procedures required for real-world data science applications.

PO3: Entrepreneurial Mindset and Knowledge

CO3 and CO8 moderately support PO3 by enabling students to design end-to-end data science solutions, identify data-driven opportunities, optimize model performance, and support innovation and decision-making relevant to entrepreneurial and business contexts.

PO4: Specialized Skills and Competencies

CO3–CO8 strongly contribute to PO4 by developing advanced technical and analytical competencies, including workflow design, matrix analysis, proximity measures, statistical modeling, and dimensionality reduction. These outcomes enhance problem-solving ability, adaptability, and technical proficiency.

PO5: Capacity for Application, Problem-Solving, and Analytical Reasoning

CO1–CO8 collectively support PO5 by enabling students to apply learned concepts to practical problems, analyze complex datasets, compare analytical techniques, and make informed decisions using logical reasoning and critical thinking.

PO6: Communication Skills and Collaboration

CO3, CO7, and CO8 moderately support PO6 by requiring interpretation, explanation, and presentation of analytical results and modeling outcomes. These activities promote effective communication and collaborative problem-solving in data-driven team environments.

PO7: Research-related Skills

CO3–CO8 strongly align with PO7 by engaging students in formulating analytical approaches, applying appropriate methodologies, performing data analysis, validating results, and adhering to ethical research practices while reporting findings systematically.

PO8: Learning How to Learn Skills

All COs partially to moderately support PO8 by encouraging independent exploration of datasets, analytical methods, and evolving tools. Students develop self-directed learning habits and adaptability essential for continuous professional growth.

PO9: Digital and Technological Skills

CO1–CO8 strongly support PO9 through extensive use of statistical software, programming tools, and computational techniques for data processing, modeling, and performance improvement in digital environments.

PO10: Multicultural Competence, Inclusive Spirit, and Empathy

The COs partially support PO10 by exposing students to diverse real-world datasets and application domains, fostering sensitivity toward inclusive interpretation and responsible data-driven decision-making across varied populations.

PO11: Value Inculcation and Environmental Awareness

CO4–CO7 partially contribute to PO11 by emphasizing ethical data handling, responsible interpretation of analytical results, and awareness of the societal and environmental implications of data-driven decisions.

PO12: Autonomy, Responsibility, and Accountability

CO3, CO6, CO7, and CO8 moderately support PO12 by enabling students to independently design analytical workflows, justify modeling choices, manage data projects, and take responsibility for outcomes.

PO13: Community Engagement and Service

The COs partially align with PO13 by enabling the application of data science techniques to community-oriented problems and evidence-based decision-making that can support societal development.