**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**WORK INTEGRATED LEARNING PROGRAMMES**

M .Tech (DSE) & M Tech(AIML)

**Course Handout(w.e.f. October 2022)**

Part A: Content Design

|  |  |
| --- | --- |
| **Course Title** | **Introduction to Statistical Methods** |
| **Course No(s)** | AIML ZC418 / DSECT ZC418 |
| **Credit Units** | 4 |
| **Credit Model** | 2 - 0.5 - 1.5.  2units for class room hours, 0.5 unit for Tutorial, 1.5 units for  Student preparation. 1 unit = 32 hours |
| **Content Authors** | Dr YVK Ravi Kumar |
| **Date** | October ,2022 |

**Course Description**

|  |
| --- |
| Basic probability concepts, Conditional probability, Bayes Theorem, Probability distributions, Continuous and discrete distributions, Transformation of random variables, estimating mean, variance, covariance, Hypothesis Testing, Maximum likelihood, ANOVA – single factor, dual factor, time series analysis: AR, MA, ARIMA, SARIMA, sampling based on distribution, statistical significance, Gaussian Mixture Model, Expectation Maximization. |

**Course Objectives**

|  |  |
| --- | --- |
| **CO1** | Introducing basic concepts of probability and statistics to students |
| **CO2** | Students will be able to apply statistical techniques to understand the dats |
| **CO3** | Students will be able to do statistical analysis of the model / algorithm |

**Text Books**

|  |  |
| --- | --- |
| **No** | **Author(s), Title, Edition, Publishing House** |
| T1 | Statistics for Data Scientsists, An introduction to probability ,statistics and Data Analysis,Maurits Kaptein et al,Springer 2022 |
| T2 | Probability and Statistics for Engineering and Sciences,8th Edition, Jay L Devore, Cengage Learning |
| T3 | Introduction to Time Series and Forecasting, Second Edition, Peter J Brockwell, Richard ADavis, Springer. |

**Reference Books**

|  |  |
| --- | --- |
| **No** | **Author(s), Title, Edition, Publishing House** |
| R1 | Miller and Freund’s Probability and statistics for Engineers, 8th Edition, PHI |
|  |  |
| R2 | Statistics for Business and Economics by Anderson, Sweeney and Wiliams, CENAGE learning |

**Modular Content Structure**

**Module 1:Basic Probability & Statisitcs**

1.1.Measures of Central Tendency

1.2.Measures of Variability

1.3. Basic Probability concepts

1.3.1.Axioms of Probability

1.3.2.Definition of Probability

1.3.3.Mutually exclusive and independent events

**Module 2: Conditional Probability & Bayes theorem**

2.1. Conditional Probability

2.2. Conditonal Probability of Independent events

2.3. Bayes Theorem

2.4. Introduction to Naïve Bayes concept

**Module 3: Probability Distributions**

3.1. Random Variables

3.1.1. Discrete random variable – Single and two variables

3.1.2. Discrete random variable – Single and two variables

3.1.3. Mean, Variance, Co – Variance of Random variables

3.1.3. Transformation of random variables

3.2. Probability Distributions

3.2.1. Bernoulli Distribution

3.2.2. Binomial Distribution

3.2.3. Poisson Distribution

3.2.4. Normal(Gaussian) distribution

3.2.5. Introduction of t – distribution , F – distribution , Chi Sqaure distribution.

**Module 4: Hypothesis Testing**

4.1. Sampling – random sampling and Stratified sampling

4.2. Sampling distribution – Cental Limit theorem

4.3. Estimation – Interval Estimation,Confidence level

4.4. Testing of Hypothesis

4.4.1.Mean based

4.4.2. Proportions related

4.4.3. ANOVA – Single and dual factor

4.5. Maximum likelihood

**Module 5: Prediction & Forecasting**

5.1. Correlation

5.2. Regression

5.3. Time Series Analysis

5.3.1. Introduction, Components of time series data

5.3.2.MA model – basic and weighted MA model

5.3.3.Time series models

5.3.3.1.AR Model

5.3.3.2. ARIMA Model

5.3.3.3. SARIMA,SARIMAX,VAR,VARMAX

5.3.3.4. Simple exponential smoothing model

**Module 6: Gaussian Mixture model & Expectation Maximization**

**Learning Outcomes:**

|  |  |
| --- | --- |
| **No** | **Learning Outcomes** |
| **LO1** | Clear understanding of the various statistical models to model the data |
| **LO2** | Drawing conclusions from the models selected to understand the data |

**Part B: Course Handout**

|  |  |
| --- | --- |
| **Academic Term** | I semester ,2022 – 23 |
| **Course Title** | **Introduction to Statistical Methods** |
| **Course No** | AIML ZC418 / DSECT ZC418 |

**Course Contents**

**Contact Session 1: Module 1(Module 1:Basic Probability & Statisitcs)**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 1 | Measures of Central Tendency & Measures of Variability,Data – Symmetric & Asymmetric outlier detection, 5 point summary, Introduction to probability | T1 & T2 |
| HW | Problems on data symmetry & outlier identification | T1 & T2 |
| Lab | Display of statistical Data & Understanding the statistical summary | Lab 1 |

**Contact Session 2: Module 1(Module 1:Basic Probability & Statisitcs)**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 2 | Axioms of Probability,Mutually exclusive and independent events,Problem solving to understand basic probability concepts | T1 & T2 |
| HW | Problems on probability | T1 & T2 |
| Lab |  |  |

**Contact Session 3: Module 2(Conditional Probability & Bayes theorem)**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 3 | Introduction to conditional probability,indepents events, Total probability | T1 & T2 |
| HW | Problems on conditional probability | T1 & T2 |
| Lab |  |  |

**Contact Session 4: Module 2(Conditional Probability & Bayes theorem)**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 4 | Bayes theorem(with proof),Introduction to Naïve Bayes concept. | T1 & T2 |
| HW | Problems on Bayes theorem | T1 & T2 |
| Lab | Bayes theorem & Naïve Bayes Concept | Lab 2 |

**Contact Session 5: Module 3: Probability Distributions**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 5 | Random variables - Discrete & continuous Expectaion of a random variable,mean and variance of a random variable – Sinlge random random variable & Joint distributions | T1 & T2 |
| HW | Problems on random variables | T1 & T2 |
| Lab | Probability Distributions & Sampling | Lab 3 |

**Contact Session 6: Module 3: Probability Distributions**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 6 | Bernoulli,Binomial, Poisson and Norma distributions. Inroduction to t – distribution , F – Distribution and Chi Sqaure distributions | T1 & T2 |
| HW | Problems on Probability distributions | T1 & T2 |
| Lab | Probability Distributions & Sampling | Lab 3 |

**Contact Session 7: Module 4: Hypothesis Testing**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 7 | Sampling – random sampling and Stratified sampling,Sampling distribution – Cental Limit theorem,Estimation– Interval Estimation,Confidence level | T1 & T2 |
| HW | Problems on Interval estimation | T1 & T2 |
| Lab |  |  |

**Contact Session 8:**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 8 | REVISION OF THE TOPICS COVERED |  |
| HW |  |  |
| Lab |  |  |

**MID SEMESTER EXAMINATION**

**Contact Session 9 : Module 4: Hypothesis Testing**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS – 9 | Testing of Hypothesis - mean and proportions related models (one mean, two mean, one proportion and Several proportions with small and big samples wherever applicable) | T1 & T2 |
| HW | Problems on Testing of Hypothesis | T1 & T2 |
| Lab |  |  |

**Contact Session 10: Module 4: Hypothesis Testing**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS – 10 | Maximum likelihood ,ANOVA – Single and dual factor | T1 & T2 |
| HW | Problems on ANOVA | T1 & T2 |
| Lab | ANOVA | Lab 4 |

**Contact Session 11: Module 5: Prediction & Forecasting**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 11 | Correlation & regression | T1 & T2 |
| HW | Problems on correlation and regression | T1 & T2 |
| Lab | Regression – Analysis of model summary | Lab 5 |

**Contact Session 12: : Module 5: Prediction & Forecasting**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 12 | Time Series Analysis - Introduction, Components of time series data, Moving Averages and weighted moving averges model | T3 |
| HW | Problems on Time series | T3 |
| Lab | Time series | Lab 6 |

**Contact Session 13: Module 5: Prediction & Forecasting**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 13 | Time series models - AR Model,ARMA Model,ARIMA | T3 |
| HW | Problems on Time Series | T3 |
| Lab | Time series | Lab 6 |

**Contact Session 14: Module 5: Prediction & Forecasting**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 14 | Time Series Models - SARIMA,SARIMAX,VAR,VARMAX and Simple exponential smoothing model | T3 |
| HW | Problems on Time Series |  |
| Lab | Time series | Lab 6 |

**Contact Session 15: Module 6: Gaussian Mixture model & Expectation Maximization**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 15 | Gaussian Mixture model & Expectation Maximization | Class Notes |
| HW | Problems on GMM & EM |  |
| Lab |  |  |

**Contact Session 16:**

|  |  |  |
| --- | --- | --- |
| **Contact Session** | **List of Topic Title** | **Reference** |
| CS - 16 | **REVISION OF THE SYLLABUS** |  |
| HW |  |  |
| Lab |  |  |

**Detailed Plan for Lab work**

|  |  |  |  |
| --- | --- | --- | --- |
| **Lab No.** | **Lab Objective** | **Lab Sheet Access URL** | **Session Reference** |
| 1 | Display of statistical Data & Understanding the statistical summary |  | 1 |
| 2 | Bayes theorem & Naïve Bayes Concept |  | 4 |
| 3 | Probability Distributions & Sampling |  | 5 & 6 |
| 4 | ANOVA |  | 10 |
| 5 | Regression – Analysis of model summary |  | 11 |
| 6 | Time series |  | 12 - 14 |
|  |  |  |  |

**Evaluation Scheme**:

Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Name | Type | Duration | Weight | Day, Date, Session, Time |
| EC-1(a) | Quizzes – 1 & 2 | Online |  | 10% |  |
| EC-1(b) | Assignments - 2 | Online |  | 20% |  |
| EC-2 | Mid-Semester Test | Closed Book |  | 30% |  |
| EC-3 | Comprehensive Exam | Open Book |  | 40% |  |

**Note:**

Syllabus for Mid-Semester Test (Closed Book): Topics in Session Nos. 1 to 8

Syllabus for Comprehensive Exam (Open Book): All topics (Session Nos. 1 to 16)