AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

Course Overview

19CSE304 Foundations of Data Science
Lecture 0

Course Outcomes

- **CO1**: Understand the statistical foundations of data science
- **CO2**: Apply pre-processing techniques over raw data so as to enable further analysis.
- **CO3:**Conduct exploratory data analysis and create insightful visualizations to identify
- patterns.
- **CO4**: Identify machine learning algorithms for prediction/classification and to derive insights
- **CO5**: Analyze the degree of certainty of predictions using statistical test and models.

Course Details

Course Code :- **19CSE304**

Course Title :- **Foundations of Data Science**

Course L-T-P :- **2-0-0-3**

Course Credits :- **3**

Course Type :- **Lab Based Course**

Course Objective

- The objective is to teach primary tools for exploration, visualizations and descriptive statistics, for prediction are machine learning and optimization, and for inference are statistical tests and models.

- Through understanding a particular domain, the students learn to ask appropriate questions about their data and correctly interpret the answers provided by inferential and computational tools.

Course Syllabus

- **Unit 1**
- Introduction, Causality and Experiments, Data Preprocessing: Data cleaning, Data reduction, Data transformation, Data discretization. Visualization and Graphing: Visualizing Categorical Distributions, Visualizing Numerical Distributions, Overlaid Graphs, plots, and summary statistics of exploratory data analysis, Randomness, Probability, Introduction to Statistics, Sampling, Sample Means and Sample Sizes.
- **Unit 2**
- Descriptive statistics – Central tendency, dispersion, variance, covariance, kurtosis, five-point summary, Distributions, Bayes Theorem, Error Probabilities; Permutation Testing, Statistical Inference; Hypothesis Testing, Assessing Models, Decisions and Uncertainty, Comparing Samples, A/B Testing, P-Values, Causality.
- **Unit 3**
- Estimation, Prediction, Confidence Intervals, Inference for Regression, Classification, Graphical Models, Updating Predictions.
-

5	dataset generation and EDA
6	Descriptive Statistics Mathematics and Statistics in Python- Sampling
7	Applying ML models and performance analysis
8	Regression Analysis
9	Web scraping
10	Applied Data science: Capstone

Text Books/References

- **Text Book:**
- Ani Adhikari. John DeNero, Computational and Inferential Thinking: The Foundations of Data Science. GitBook, 2019.
- **Reference(s):**
- Shmueli G, Bruce PC, Yahav I, Patel NR, Lichtendahl Jr KC. Data mining for business analytics: concepts, techniques, and applications in R. John Wiley & Sons; 2018..
- Schutt R, O'Neil C. Doing data science: Straight talk from the frontline. First Edition, O'Reilly Media, Inc.; 2013.

Course Evaluation Policy

Assessment	Internal/External	Weightage (%)
Continuous Assessment (50) •Quiz (2): 2 x 5m •Timely lab submission -5 m •Lab evaluation -10 m •Blog writing- EDA on generated dataset (individual) - 10m •Revised Blog Submission-5 m •Quiz on Blog Writing-10 m	Internal	50
Mid-Term Examination (50 m)	Internal	20
End Semester Examination (100 m)	External	30

What is Data Science

- Data Science is all about:

- Asking the correct questions and analyzing the raw data.
- Modeling the data using various complex and efficient algorithms.
- Visualizing the data to get a better perspective.
- Understanding the data to make better decisions and finding the final result.

What is Data Science

- Data science is a deep study of the massive amount of data, which involves extracting meaningful insights from raw, structured, and unstructured data that is

  processed using the

  scientific method, different

  technologies, and algorithms.

- It is a multidisciplinary field that u
  manipulate the

  data so that you can find

  something new and meaningful.

Data Science Applications