INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

FOUNDATIONS OF DATA SCIENCE								
IV Semester: CE								
Course Code	**Category**	**Hours/Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
ACEE15	Core	3	0	0	3	40	60	100
Contact Classes: 48	**Tutorial Classes: Nil**	**Practical Classes: Nil**				**Total Classes: 48**		
Prerequisite: Probability and Statistics								

I. COURSE OVERVIEW:

Data science is an emerging field that encompasses several STEM domains and offers exciting career prospects in a wide range of engineering applications. Given the rapid adoption of data analytics in engineering and the societal importance of many of its applications, there is an urgent need to train the next generation of civil engineers in data science. The course aims at empowering students with the necessary tools to apply statistics in a civil engineering context, and perform data transformation, data wrangling, visualization, and modeling. Students learn how to gather and analyze data as part of the engineering design process, apply systems thinking to an engineering or societal phenomenon, collaborate with peers to find solutions, and effectively present solutions to an audience.

II. COURSE OBJECTIVES:

The students will try to learn:

I. The data science process, data transformation, visualization, and modelling using appropriate tools.
II. Different types of machine learning approaches and practice visualization techniques using tools.
III. The handle large volumes of data with case studies and effectively present solutions to an audience.

III. COURSE OUTCOMES:

At the end of the course students should be able to:

CO1 Explain the data science process for handling large data.
CO2 Demonstrate the data manipulation functions using appropriate tools like numpy and pandas.
CO3 Outline the data cleaning methods for analyzing the data.
CO4 Examine the data preprocessing for standardizing and normalizing the data.
CO5 Utilize different data visualization techniques using various tools.
CO6 Analyze the large volumes of data in practical scenarios using data visualization techniques.

IV. COURSE CONTENT:

MODULE –I: INTRODUCTION (9)

Data Science: Benefits and uses, facets of data, Data Science Process: Overview, defining research goals, retrieving data, data preparation, exploratory data analysis, build the model, presenting findings and building applications, data mining, data warehousing, basic statistical descriptions of data.

MODULE -II: DATA MANIPULATION (9)

Python Shell, Jupyter Notebook, IPython Magic Commands, NumPy arrays, universal functions, aggregations, computation on arrays, fancy indexing, sorting arrays, structured data, data manipulation with pandas, data Indexing and selection, handling missing data, hierarchical indexing, combining datasets, aggregation and grouping, string operations, working with time series, high performance.

MODULE -III: DATA CLEANING AND PRE-PROCESSING (10)

Data Cleaning: Need for cleaning the Data, treating of missing values, treating duplicate values, treating the bad data, Data Imputation and methods, Fuzzy Matching and its applications.

Data Preprocessing: Standardizing the data, Normalizing the data, Outliers and their detection, treating the outliers.

MODULE -IV: DATA VISUALIZATION (10)

Importing matplotlib, simple line plots, simple scatter plots, visualizing errors, density and contour plots, histograms, legends, colors, subplots, text and annotation, customization, three-dimensional plotting geographic data with base map, visualization with seaborn.

MODULE -V: HANDLING LARGE DATA (10)

Problems, techniques for handling large volumes of data, programming tips for dealing with large data sets, case studies: predicting malicious URLs, building a recommender system, tools and techniques needed, research question, data preparation, model building, presentation and automation.

V. TEXTBOOKS:

1. David Cielen, Arno D. B. Meysman, and Mohamed Ali, "*Introducing Data Science*", Manning Publications, 2016.
2. Jake VanderPlas, "*Python Data Science Handbook*", O'Reilly, 2016.

VI. REFERENCE BOOKS:

1. Robert S. Witte and John S. Witte, "*Statistics*", Eleventh Edition, Wiley Publications, 2017.
2. Allen B. Downey, "*Think Stats: Exploratory Data Analysis in Python*", Green Tea Press, 2014.

VII. ELECTRONICS RESOURCES:

1. https://www.programiz.com/python-programming
2. https://towardsdatascience.com
3. https://www.vedantu.com

VIII. MATERIAL ONLINE:

1. Course Outline Description
2. Lecture notes
3. PowerPoint presentation
4. Definitions and Terminology
5. Tutorial Question Bank
6. Case Studies
7. Real life Examples
8. Complex Engineering Problems
9. Tech Talk Topics
10. Concept Video Topics
11. Open-ended Exercises
12. Assignments
13. Model Question Paper – I
14. Model Question Paper – II

15. GATE Question Bank
16. Previous Question Papers and Solutions