INSTITUTE	**FACULTY OF TECHNOLOGY**
PROGRAM	**BACHELOR OF TECHNOLOGY (COMPUTER ENGINEERING)**
SEMESTER	**5**
COURSE TITLE	**DATA SCIENCE ESSENTIALS**
COURSE CODE	**01CE0517**
COURSE CREDITS	**3**

Objective:

1 To provide a strong foundation for data science and application areas related to it and understand the underlying core concepts and emerging technologies in data science.

Course Outcomes: After completion of this course, student will be able to:

1 Explain the fundamental concepts, tools, and workflow of data science, and demonstrate their significance in solving real-world problems.

2 Apply data acquisition, cleaning, and preprocessing techniques to prepare structured and unstructured datasets for analysis.

3 Perform exploratory data analysis and interpret statistical measures such as central tendency, variability, and distributions.

4 Implement basic machine learning algorithms (e.g., Linear Regression, Logistic Regression, Naive Bayes, Decision Tree) and evaluate their performance using appropriate metrics.

5 Analyze ethical issues related to data privacy, fairness, and bias, and apply ethical frameworks to decision-making in data science projects.

Pre-requisite of course:NA

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
2	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Introduction to data science** Overview of data science and its applications, Emergence of data science, Outlining the core competencies of a data scientist and data science, Linking data science with big data and AI, data science workflow and process, Role of python in data science, Tool for Data Science.	4

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
2	**Data Acquisition and Management** Introduction to different data formats (structured, unstructured, semi- structured), overview of data acquisition techniques (surveys, web scraping, APIs), Data cleaning techniques to Handle missing data and outliers, Data preprocessing- Issues in high dimensional data, Dimensionality reduction and feature subset selection.	8
3	**Data analysis** Exploratory Data Analysis: Introduction, Exploring relationships and patterns in data, Feature engineering and selection, Predictive vs Descriptive analytics., Statistics for Data Analysis: Descriptive statistics (Measures of central tendency and variability) and data summarization, Central Limit Theorem, Sampling Distribution	10
4	**Machine Learning for data science** Definition, Types of learning, Evaluation and performance measures, overfitting and underfitting, Linear Regression: Model, Cost Function gradient descent, Simplifying Models through Regularization, Logistic Regression, Naive Bayes, Decision Tree.	10
5	**Ethical Issues in Data Science** Privacy and Data Protection: Overview of privacy concerns in data science, Ethical considerations in data collection and usage, Bias and Fairness in Data Science: Fairness considerations in machine learning models, Techniques for measuring and mitigating bias in data science, Ethical Decision-making in Data Science: frameworks and principles, Ethical dilemmas, Ethical guidelines for next-generation data scientists	4
Total Hours		**36**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Practical-1** a. Hands on practical on Jupyter notebook and google colab., b. Explore and import the features of various packages.	2
2	**Practical-2** Working with Numpy	2
3	**Practical-3** Working with Pandas	2
4	**Practical-4** Hands on practical to clean noisy data by following techniques:i. Dropping ii. Mean iii. Median iv. Mode	2
5	**Practical-5** Hands on practical with data preprocessing techniques: a. Handling categorical data i. Label Encoding ii. Dummy Encoding iii. One-hot encoding	2

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
6	**Practical-6** Hands on practical for features scaling on a real-world dataset: a. Normalization b. Standardization	2
7	**Practical-7** Implement measures of central tendency and variability on diabetes dataset to learn and apply statistical analysis.	2
8	**Practical-8** Perform Exploratory Data Analysis (EDA) on student dataset to analyse performance of student.	2
9	**Practical-9** Hands on practical with sklearn package to build linear regression model on estate dataset and its evaluation.	2
10	**Practical-10** Apply Logistic Regression algorithm on Cancer Dataset and perform diagnostic classification operation.	2
11	**Practical-11** Apply Decision Tree algorithm on a weather forecasting dataset to predict humidity and evaluate model performance using accuracy score and mean square error.	2
12	**Practical-12** Write a python script: a. Implement Naïve Bayes classification Model on a real-world dataset. b. Evaluate model performance using RMSE.	2
13	**Practical-13** Implement Support Vector Machine (SVM) algorithm on an insurance dataset for classification tasks.	2
14	**Practical-14** Conduct a case study to analyse and explore ethical issues in the field of data science.	2
Total Hours		**28**

Textbook :

1 Data Science from Scratch: First Principles with Python, Joel Grus,, O'Reilly Publication, 2019

References:

1 Data Science for Dummies, Data Science for Dummies, Lillian Pierson, Wiley Publication, 2021

2 Practical statistics for data scientists, Practical statistics for data scientists, Peter Bruce, Andrew Bruce and Peter Gedeck, O'Reilly Publication, 2017

3 Headfirst Statistics, Headfirst Statistics, Dawn Griffiths, O'Reilly Publication, 2008

4 Machine Learning for Absolute Beginners, Machine Learning for Absolute Beginners, Oliver Theobald, Scatterplot Press, 2017

References:

5 Python for data analysis, Python for data analysis, Wes McKinney, O'Reilly Publication, 2017

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
10.00	20.00	40.00	30.00	0.00	0.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

4 Students will use supplementary resources such as online videos, NPTEL videos, e-courses, Virtual Laboratory.

Supplementary Resources:

1 https://www.coursera.org/programs/milap-faculty-program- mm3kt/browse?collectionId=&productId=_Fk2Gi3cEeiHghIydZ_0lA&productType=s12n&query=data+science&showMiniModal=true&source=search

DR. KRUNAL NARUMAL VAGHELA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)