

Course Code	ECAP792	Course Title	DATA SCIENCE TOOL BOX		
			WEIGHTAGE		
			CA	ETE(Th.)	ETE (Pr.)
			30	40	30

Course Outcomes:

- C01:** Understand the concept and need for data science.
- C02:** Discuss the various phases in the data analytics lifecycle.
- C03:** Analyze the fundamental areas of study in data science
- C04:** Understand the concept of data preprocessing.
- C05:** Understand the importance of data visualization.
- C06:** Learn the various data visualization software and libraries.
- C07:** Understand different types of machine learning techniques
- C08:** Use Python for developing machine learning algorithms
- C09:** Use various data science tools for developing algorithms

Unit No.	Content
Unit-1	Introduction to Data Science: Why learn data science?, Life cycle of data analytics: Data discovery, Data preparation, Model planning, Model building, Communicate results, Operationalization. Type of data analysis: Descriptive analysis, Diagnostic analysis, Predictive analysis, Prescriptive analysis, types of data analytics.
Unit-2	Data pre-processing: Introduction to data preprocessing, Data preprocessing, Data wrangling, Data types and forms, Possible data error types.
Unit-3	Various data preprocessing operations: Data cleaning, Data integration, Data transformation, Data reduction, Data discretization.
Unit-4	Data Plotting and Visualization: Introduction to data visualization, Visual encoding, Data visualization software, Data visualization libraries, Basic data visualization tools, Advanced data visualization tools, Data visualization types.
Unit-5	Role of statistics in data science: Hypothesis testing, null hypothesis, alternative hypothesis, Statistical significance: Type 1 and type 2 errors, Data science, p-value, ANOVA, Chi-square test.
Unit-6	Machine learning: Introduction, types of machine learning techniques, learning problems and system, designing a learning system, concept of learning task.
Unit-7	Unsupervised learning: Introduction to Clustering algorithms, K Means, K mode, K median, Performance measures of clustering.
Unit-8	Supervised learning: Introduction to Classification algorithms, KNN (k-nearest neighbors) algorithm, Naïve Bayes algorithm, cross-validation and metrics.
Unit-9	Regression models: Introduction to regression, types of regression, Machine linear regression, machine logistic regression, regularization, performance metrics.
Unit-10	Weka: Introduction to weka tool, Data import, Choose model (algorithm), Hands on analysis of clustering and classification algorithms.
Unit-11	Excel data analysis: Introduction to excel data analysis, Data analysis tool pack, Descriptive statistics, Analysis of variance (ANOVA), Regression, Histogram.
Unit-12	R tool: Introduction R, RStudio, Some important R data structures: Vectors, character strings, Matrices, Lists, Dataframe, R programming structure.
Unit-13	NumPy and Pandas: Introduction to python, NumPy, understanding data types in python, Pandas for data analysis, data indexing and selection: Data selection in series, Data selection in DataFrame, Missing data in Pandas, Handling missing data
Unit-14	Machine learning packages in python: Data import, Visualization with Matplotlib, simple line and scatter plots, Seaborn, heatmap, Introducing Sci-kit learn package.

List of Practical/ Experiments

- Build a Binary Classifier for Tweets using a Logistic Regression.
- Build a Naive Bayes Tweet Classifier.
- Create Word Vectors that Capture Dependencies between Words, then Visualize their relationships in Two Dimensions using PCA.
- Transform Word Vectors and Assign them to Subsets using Locality Sensitive Hashing.
- Build your own Spellchecker to Correct Misspelled Words.
- Create Part-of-Speech Tags for a Wall Street Journal Text corpus using Markov models.

- Build your own Autocomplete Language model using a Text corpus from Twitter using N-gram Language models.
- Build your own Continuous Bag-of-Words model to Create Word Embeddings from Shakespeare text.
- Build a Sophisticated Tweet Classifier that places Tweets into Positive or Negative Sentiment categories, using a Deep Neural Network.
- Build your own Next-Word Generator using a simple RNN on Shakespeare Text data.
- Build your own Named Entity Recognition system using an LSTM and Data from Kaggle.
- Build your own Siamese Network that Identifies Question Duplicates in a Dataset from Quora.
- Build a Neural Machine Translation model with Attention that Translates English Sentences into German.
- Create a Tool that Generates Text Summaries.

Readings:

1. Norman matloff, **“The Art of R Programming”**, No starch press, 2011.
2. Jason Bell, **“Machine Learning: Hands-On for Developers and Technical Professionals”**, Wiley Publication, 2015
3. Jake VanderPlas, **“Python Data Science Handbook”**, O’reilly, 2017.
4. Alexander Loth, Nate Vogel and Sophie Sparkes, **“Visual Analytics with Tableau”**, Wiley, 2019.
5. Gypsy nandi and Rupam kumar sharma, **“Data science Fundamentals and Practical approach”**, bpb publisher, 2020.
6. Kenneth Cukier and Viktor Mayer-Schönberger, Big Data: A Revolution That Will Transform How We Live, Work, and Think, Hodder And Stoughton, 2013.
7. Daniel T. Larose , Chantal D. Larose, Data mining and Predictive analytics, Second Ed., Wiley Publication, 2015
8. Bertt Lantz, Machine Learning with R: Expert techniques for predictive modeling, 3rd Edition, April 15, 2019.

Course Code	ECAP794	Course Title	ADVANCE DATA VISUALISATION		
			WEIGHTAGES		
			CA	ETE(Th.)	ETE (Pr.)
			30	40	30

Course Outcomes:

- C01:** discuss the terminology used in Tableau Prep.
- C02:** identify how Tableau Prep approaches data sampling.
- C03:** construct and understand data prep flows that address common scenarios encountered in data preparation, as applied to common data use cases.
- C04:** review the quality of the data and perform exploratory analysis.
- C05:** manage and Connect Data Source.

Unit No.	Content
Unit-1	Introduction to Data Visualization: Acquiring and Visualizing Data, Simultaneous acquisition and visualization, Applications of Data Visualization, Keys factors of Data Visualization. Reading Data from Standard text files (.txt, .csv, XML), Displaying JSON content.
Unit-2	Making charts interactive and animated: Data joins, updates and exits, interactive buttons, Updating charts, Adding transactions, using keys , wrapping the update phase in a function, Adding a Play button to the page, Making the Play button go, Allow the user to interrupt the play, sequence.