Data Analytics Lesson Plan
Spring 6th Semester, 2022-23

School of Computer Engineering,
KIIT Deemed to be University
(Institute of Eminence)

Course Details

Program(s)	Academic Session, Semester	Subject Name	Subject Code	Credit
B.Tech [CSE, IT, CSCE, CSSE]	Spring 2022-23 (December-May) 6th Sem	Data Analytics	IT-3006	3

Course Committee Faculties

Course Code : IT 3032 Course Name : Data Analytics (Cr : 3)		
Dr. Sarita Tripathy (C)	sarita.tripathyfcs@kiit.ac.in	9437696333
Dr. Mainak Bandyopadhyay	mainak.bandyopadhyayfcs@kiit.ac.in	8009387402
Dr.Siddharth Swarup Rautaray	siddharthfcs@kiit.ac.in	8763999449
Dr. Manjusha Pandey	manjushafcs@kiit.ac.in	8763999448
Dr. Rajat Kumar Behera	rajatkumar.beherafcs@kiit.ac.in	9886072882
Dr. Satarupa Mohanty	satarupafcs@kiit.ac.in	9124941515
Dr. Hrudaya Kumar Tripathy	hktripathyfcs@kiit.ac.in	9437432185
Dr. Minakhi Rout	minakhi.routfcs@kiit.ac.in	9861108580
Mr. Ajay Anand	ajay.anandfcs@kiit.ac.in	9178181981
Dr. Ambika Prasad Mishra	ambikaprasad.mishra@gmail.com	9692144071
Dr. Soumya Ranjan Mishra	soumyaranjan.mishrafcs@kiit.ac.in	8096177170
Mrs. Jayanti Dansana	jayantifcs@kiit.ac.in	9438509667
Dr. Arup Abhinna Acharya	aacharyafcs@kiit.ac.in	9861058079
Dr. Jagannath Singh	jagannath.singhfcs@kiit.ac.in	9861085883

Course Outcomes

Prerequisite: DBMS

	At the end of the course, the students will be able to:
CO1	Make use of data science concepts to handle the big data. [L3]
CO2	Examine the statistical concepts for finding relationships among variables and estimate the data samplings [L4, L5]
CO3	Select the data analytics techniques & models for both data prediction and performance analysis. [L5, L6]
CO4	Develop rules using frequent itemsets and association mining. [L6]
CO5	Solve real-time problems using classification and clustering techniques. [L6]
CO6	Apply the mining techniques for data streams. [L3]

Syllabus

Module	Title	Contents	Hours
1.	**Introduction to Data Science**	Introduction to Data, Data Science, Challenges of Traditional Systems, Evolution of Analytic Scalability, Types of Computing (Distributed, Parallel, Grid), Data Analytics Lifecycle, Introduction to Big Data (Characteristics) and Hadoop (Hadoop Ecosystem, MapReduce, Hbase, Pig, Hive, Sqoop, NOSQL), Visualizations.	8
2.	**Statistical Concepts**	Data Exploration: Distribution of a single variable, Basic Concepts (populations and samples, data sets, variables, and observations, types of data), Descriptive measures for categorical variables, descriptive measures for numerical variables, Outliers and Missing values, Finding relationships among variables: Categorical variables, Numerical variables, Categorical variables and a Numerical variable. Sampling and distributions: Terminology, Estimation, Confidence Interval estimation, Sampling distributions, Confidence interval, Hypothesis testing, Chi-square test for independence.	9
3.	**Data Analytics**	IIntroduction, Types of Data Analytics, Importance of Data Analytics, Data Analytics Applications, Regression Modelling Techniques: Linear Regression, Multiple Linear Regression, Non Linear Regression, Logistic Regression, Time Series Analysis, Performance analysis (RMSE, MAPE).	8

Syllabus(contd.)

Module	Title	Contents	Hours
4.	**Frequent Itemsets and Association**	Introduction to Frequent Itemsets, Market-Basket Model, Algorithm for Finding Frequent, Itemsets, Association Rule Mining, Apriori Algorithm and Correlations.	5
5.	**Classification & Clustering**	Introduction to classification and clustering, Distance-Based Algorithms: K Nearest Neighbour (KNN), Decision Tree-Based Algorithms: Decision Tree (ID3 Algorithm), Support Vector Machines (Linear), Naves Bayes. Overview of Clustering Techniques, Hierarchical Clustering, Partitioning Methods, K- Means Algorithm.	6
6.	**Data Streams**	Introduction to Mining Data Streams, Data Stream Management Systems, Data Stream Mining, Examples of Data Stream Applications, Stream Queries, Issues in Data Stream Query, Processing, Sampling in Data Streams, Filtering Streams, Counting Distinct Elements in a Stream, Estimating Moments.	6

Books

- **Text Book:**
  - Data Analytics, Radha Shankarmani,M. Vijayalaxmi, Wiley India Private Limited, ISBN: 9788126560639.

- **Reference Books:**
  - Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services (Editor), Wiley, 2014
  - Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with advanced analystics, John Wiley & sons, 2012.
  - Glenn J. Myatt, Making Sense of Data, John Wiley & Sons, 2007 Pete Warden, Big
  - Data Glossary,O'Reilly, 2011.
  - Jiawei Han, MichelineKamber "Data Mining Concepts and Techniques", Second Edition, Elsevier, Reprinted 2008.
  - Stephan Kudyba, Thomas H. Davenport, Big Data, Mining, and Analytics, Components of Strategic Decision Making, CRC Press, Taylor & Francis Group. 2014
  - Big Data, Black Book, DT Editorial Services, Dreamtech Press, 2015

Lesson Plan

Modules	Lecture Days	Topics/Coverage
1. Introduction to Data Science	1st to 3rd	Introduction to Data, Data Science, Challenges of Traditional Systems, Evolution of Analytic Scalability, Types of Computing (Distributed, Parallel, Grid),
	4th to 7th	Data Analytics Lifecycle, Introduction to Big Data (Characteristics) and Hadoop (Hadoop Ecosystem, MapReduce, Hbase, Pig, Hive, Sqoop, NOSQL), Visualizations.
	8th	***Module 1 Activities***
2.Statistical Concepts	10th - 12th	Data Exploration: Distribution of a single variable, Basic Concepts (populations and samples, data sets, variables, and observations, types of data), Descriptive measures for categorical variables, descriptive measures for numerical variables, Outliers and Missing values
	13th to 15th	Finding relationships among variables: Categorical variables, Numerical variables, Categorical variables and a Numerical variable.
	16th to 18th	Sampling and distributions: Terminology, Estimation, Confidence Interval estimation, Sampling distributions, Confidence interval, Hypothesis testing, Chi-square test for independence.
	19th	***Module 2 Activities***

Lesson Plan

Modules	Lecture Days	Topics/Coverage
3. Data Analytics	20th to 21th	Introduction, Types of Data Analytics, Importance of Data Analytics, Data Analytics Applications
	22th 24th	Regression Modelling Techniques: Linear Regression, Multiple Linear Regression, Non Linear Regression.
	MID SEMESTER	
	25th-27th	Logistic Regression, Time Series Analysis, Performance analysis (RMSE, MAPE).
	28th	***Module 3 Activities***
4. Frequent Itemsets and Association	29th to 31st	Introduction to Frequent Itemsets, Market-Basket Model, Algorithm for Finding Frequent Itemsets,
	32nd to 33rd	Association Rule Mining, Apriori Algorithm and Correlations.
	34th	***Module 4 Activities***

Lesson Plan(contd.)

5. Classification & Clustering	35th to 37th	Introduction to classification and clustering, Distance-Based Algorithms: K Nearest Neighbour (KNN), Decision Tree-Based Algorithms: Decision Tree (ID3 Algorithm),
	38th to 40th	Support Vector Machines (Linear), Naves Bayes. Overview of Clustering Techniques, Hierarchical Clustering, Partitioning Methods, K-Means Algorithm.
	41st	***Module 5 Activities***
6.Data Streams	42nd to 44th	Introduction to Mining Data Streams, Data Stream Management Systems, Data Stream Mining, Examples of Data Stream Applications, Stream Queries, Issues in Data Stream Query, Processing
	45th to 47th	Sampling in Data Streams, Filtering Streams, Counting Distinct Elements in a Stream, Estimating Moments.
	48th	***Module 6 Activities***

Activities

Group formation (For this Semester)

- In the beginning of session, students may be divided into groups (8 to 10 members)

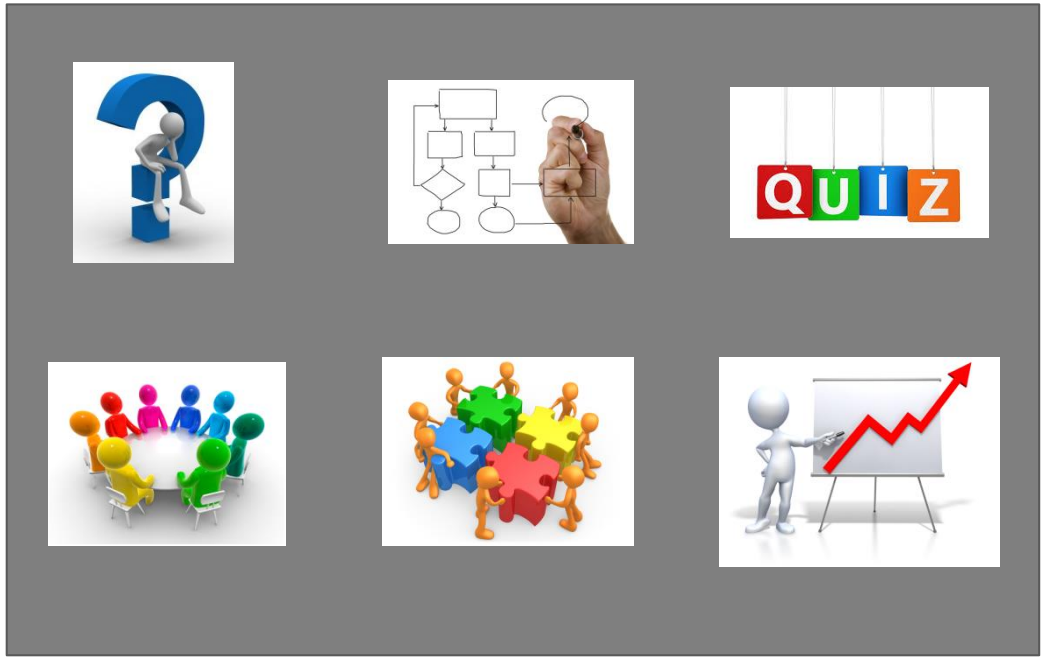
Real world project prototypes

- Right after the group formation, Students will be asked to participate in prototype projects in different domains in real world.
- After every module, they will be preparing the relevant and necessary diagrams/documents/ code/algorithm/test plan as required as part of activity deliverables.

Focus areas have been identified as

- **Real world problem identification & solution approach.**
- **Analyzing the probable solution.**
- **Critical thinking**
- **Creation of design**
- **Interactivity Focus**
- **Reflection**

- **Activities have been identified in every module.**
- **Different focus areas have been identified as applicable across all modules.**



Evaluation Scheme

EXAM		MARKS
End Semester		50
Internal	Mid Semester	20
	Activities	30
Total		**100**