ST. ALOYSIUS COLLEGE(AUTONOMOUS), JABALPUR

Reaccredited 'A+ 'Grade by NAAC(CGPA:3.68/4.00)
College with Potential for Excellence by UGC
DST-FIST Supported & STAR College Scheme by DBT

Faculty of Science

Bachelor of Science (B.Sc.) VI Semester
SUBJECT: COMPUTER SCIENCE
Paper-Core
Data Science using Python

Course Outcomes

CO. No.	Course Outcomes	Cognitive Level
CO 1	Understand database concepts, applications, structure, the need for a database, and ER-Model terminologies.	U, A
CO 2	Be able to understand the fundamentals of Relational Algebra and relational calculus	K,U
CO 3	To gain skills in creating the logical design of databases, including the E R method and normalization approach.	U
CO 4	Know about SQL functions and operators	U, A
CO 5	Understand the knowledge of Database and transaction management.	U

Credit and Marking Scheme

	Credits	Marks		Total Marks
		Internal	External	
Theory	4	40	60	**100**
Practical	2	40	60	**100**
Total	**6**	**200**		

Evaluation Scheme

	Marks	
	Internal	External
Theory	3 Internal Exams of 20 Marks (During the Semester) (Best 2 will be taken)	1 External Exams (At the End of the Semester)
Practical	3 Internal Exams (During the Semester) (Best 2 will be taken)	1 External Exams (At the End of the Semester)

`

ST. ALOYSIUS COLLEGE(AUTONOMOUS), JABALPUR

Reaccredited 'A+ 'Grade by NAAC(CGPA:3.68/4.00)
College with Potential for Excellence by UGC
DST-FIST Supported & STAR College Scheme by DBT

Content of the Course

Theory

No. of Lectures (in hours per week): 2 Hrs. per week

Total No. of Lectures: 60 Hrs.   **Maximum Marks:** 60

Units	Topics	No. of Lectures
I	Introduction to Data Science, Data Science and Artificial Intelligence, Basic concept of data mining: KDD process and its steps, Types of data in data science, Data Science functionalities: Introduction to ML and its types Preparing Data, Data Preprocessing and its Needs, Data Cleaning, Data Integration and Transformation, Data Normalization, Data Reduction, Principal Component Analysis (PCA), Discretization and Concept Hierarchy Generation.	12
II	Association Rule Mining: Basic Concept of Support and Confidence, Apriori Algorithms for mining frequent item-sets. Improving the accuracy of Apriori Algorithm FP-Growth algorithm. Classification and Prediction: Decision Tree Induction, Bayesian Classification, KNN, And Classification by Backpropagation: Neural Network, ANN and SVM, Classifier Accuracy.	12
III	Cluster Analysis: Introduction, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Outlier Analysis. Cluster Model accuracy. Python for Data Science, Data types in Python, Input-output statements in Python, Control statements in Python, arrays and functions in Python. Operators in Python	12
IV	List and its operation, Dictionary, Data analysis using Python- pandas, importing and reading a CSV sheet, basic exploration of data, converting a Python data structure to a data frame, numerical description of a data frame, understanding iloc() and loc(), tackling Null values, data frames(concatenating, merging, join),	12
V	Numpy, Data frames, Data visualization using Matplotlib, Python libraries for Data Science, scilab, scikit, tensor flow, Information extraction using NLP, NLTK library using NLP, Case study Sentimental Analysis, Reinforcement Learning. Case study: Prediction of the disease in health services by building a model.	12

References

Textbooks:

- Data Mining – Concepts and Techniques - Jiawei Han & Micheline Kamber Harcourt, India.
- Eric Matthes, Python Crash Course: A Hands-On, Project-Based Introduction to Programming (2nd Edition)
- Zed A. ShawLearn Python the Hard Way: 3rd Edition
- John M. ZellePython Programming: An Introduction to Computer Science (3rd Edition)

ST. ALOYSIUS COLLEGE(AUTONOMOUS), JABALPUR

Reaccredited 'A+ 'Grade by NAAC(CGPA:3.68/4.00)
College with Potential for Excellence by UGC
DST-FIST Supported & STAR College Scheme by DBT

List of Practical
Max Marks = 100 (60 External+40 Internal)

1. Create a list of random numbers and classify them as discrete or continuous variables.
2. Convert a numerical variable into a categorical variable based on specific criteria.
3. Calculate the mean, mode, median, standard deviation, variance, covariance, and correlation of a given dataset.
4. Perform a regression analysis to determine the relationship between two numerical variables.
5. Use the pandas library to read a CSV file using the read_csv() function.
6. Use functions like head(), tail(), info(), and describe() to get an overview of the data.
7. Convert a Python list, dictionary, or NumPy array to a DataFrame using the pandas library.
8. Calculate statistical measures like mean, median, and standard deviation on DataFrame columns.
9. Use iloc() for integer-based indexing and loc() for label-based indexing to access specific rows or columns in a data frame.
10. Identify and handle missing or Null values using functions like isnull(), fillna(), or dropna().
11. Perform DataFrame operations like concatenating, merging, and joining multiple DataFrames using concat(), merge(), and join() functions.
12. Use NumPy functions for indexing, reshaping arrays, generating random values, and performing mathematical operations on arrays.