SAL Institute of Technology and Engineering Research

Subject Name : Python for Data Science

Subject Code: 3150713

Semester–V Mid Exam Syllabus

	Topics		
1.	**Overview of Python and Data Structures:** Basics of Python including data types, variables, expressions, objects and functions. Python data structures including String, Array, List, Tuple, Set, Dictionary and operations them.	**06**	**20**
2.	**Data Science and Python:** Discovering the match between data science and python: Considering the emergence of data science, Outlining the core competencies of a data scientist, Linking data science, big data, and AI , Understanding the role of programming, Creating the Data Science Pipeline, Preparing the data, Performing exploratory data analysis, Learning from data, Visualizing, Obtaining insights and data products, Understanding Python's Role in Data Science, Considering the shifting profile of data scientists, Working with a multipurpose, simple, and efficient language, Learning to Use Python Fast ,Loading data, Training a model, Viewing a result. Introducing Python's Capabilities and Wonders: Why Python?, Grasping Python's Core Philosophy, Contributing to data science, Discovering present and future development goals, Working with Python, Getting a taste of the language, Understanding the need for indentation, Working at the command line or in the IDE, Performing Rapid Prototyping and Experimentation, Considering Speed of Execution, Visualizing Power, Using the Python Ecosystem for Data Science, Accessing scientific tools using SciPy, Performing fundamental scientific computing using NumPy, Performing data analysis using pandas, Implementing machine learning using Scikit-learn, Going for deep learning with Keras and TensorFlow, Plotting the data usingmatplotlib, Creating graphs with NetworkX, Parsing HTML documents using Beautiful Soup.	**04**	**15**

3.	Getting Your Hands Dirty With Data: Understanding the tools:Using the Jupyter Console, Interacting with screen text, Changing the window appearance, Getting Python help, Getting IPython help, Using magic functions, Discovering objects, Using Jupyter Notebook, Working with styles, Restarting the kernel, Restoring a checkpoint, Performing Multimedia and Graphic Integration, Embedding plots and other images, Loading examples from online sites, Obtaining online graphics and multimedia. Working with Real Data: Uploading, Streaming, and Sampling Data, Uploading small amounts of data into memory, Streaming large amounts of data into memory, Generating variations on image data, Sampling data in different ways, Accessing Data in Structured Flat-File Form ,Reading from a text file Reading CSV delimited format, Reading Excel and other Microsoft Office files, Sending Data in Unstructured File Form, Managing Data from Relational Databases, Interacting with Data from NoSQL Databases, Accessing Data from the Web. Conditioning Your Data: Juggling between NumPy and pandas, Knowing when to use NumPy, Knowing when to use pandas, Validating Your Data, Figuring out what's in your data, Removing duplicates, Creating a data map and data plan, Manipulating Categorical Variables, Creating categorical variables, Renaming levels, Combining levels, Dealing with Dates in Your Data, Formatting date and time values, Using the right time transformation, Dealing with Missing Data, Finding the missing data, Encoding missingness, Imputing missing data, Slicing and Dicing: Filtering and Selecting Data, Slicing rows, Slicing columns, Dicing, Concatenating and Transforming, Adding new cases and variables, Removing data, Sorting and shuffling, Aggregating Data at Any Level. Shaping Data: Working with HTML Pages, Parsing XML and HTML, Using XPath for data extraction, Working with Raw Text, Dealing with Unicode, Stemming and removing stop words, Introducing regular expressions, Using the Bag of Words Model and Beyond, Understanding the bag of words model, Working with n grams, Implementing TF-IDF transformations, Working with Graph Data, Understanding the adjacency matrix, Using NetworkX basics.	**10**	**30**