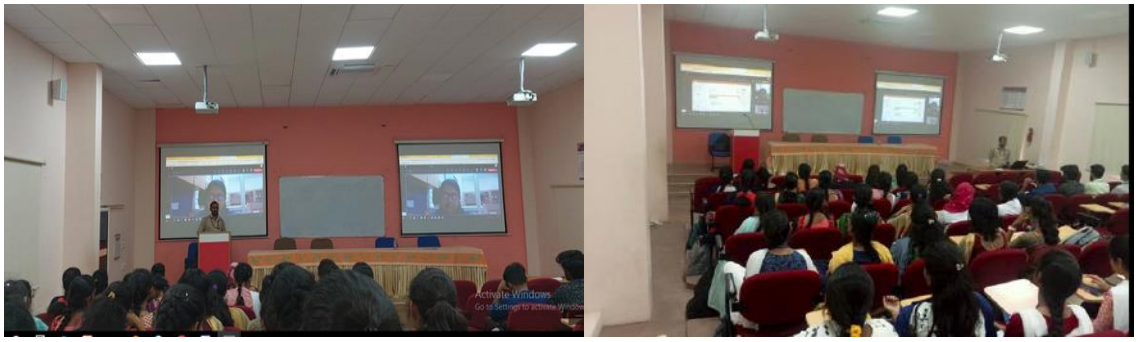
A Report on Guest Lecture on

Exploratory data analysis in machine learning using python

Organized by Department of Computer Science & Technology

27.09.2023



Submitted by : Mr Srikantarao.S, Assistant Professor, Department of Computer Science and Technology.

Resource Person: Dr Anooja Ali, Associate Professor, Department of CSE, REVA University, Bengaluru.

Report Received on 03.09.2023

Convened by: Dr M Sreedevi, Professor & Head, Department of Computer Science & Technology.

Participants: III-year CST Students

Department of Computer Science & Technology, has organized one day Guest lecture on **" Exploratory data analysis in machine learning using python "** on 27-09-2023 (Wednesday), 2:00 PM to 4.15 PM.

Objective:

The aim of this Guest Lecture is that students have to know about Data visualization and their conversion from Data visualization to BI, Types of variables, Tools for Exploratory data analysis, Demonstration of packages in python for data analysis, Jupyter notebook explanation on NumPy, Pandas packages with practical implementation.

The Programme started at 2.00 PM

Welcome Address:

Mr. V. Naveen, Assistant Professor, Event Coordinator, Department of CST, MITS gave few words about the event with a grand welcome to the Resource Person, HOD, Faculty members and participants and thanked organizers and participants for organizing this event.

Resource Person Introduction:

Mr. Srikantarao S, Assistant Professor, Department of CST, MITS, Madanapalle conveyed her greetings to Resource person, Management, Principal, Vice Principals, Deans, all the HODs, Faculty members for giving this opportunity for conducting Online guest lecture. Then he introduced the resource person **Dr.ANOOJA ALI** to the gathering and read about her achievements.

Dr. ANOOJA ALI started the session by extending her heartily thanks to the participants organizing members, HOD, Principal and Management of MITS, Madanapalle for giving her opportunity to share her knowledge and experience in **" Exploratory data analysis in machine learning using python"**

Importance of EDA in Data Science and Machine leaning

The Data Science and machine leaning field are now very important in the business world as they provide many opportunities to make vital business decisions by analyzing hugely gathered data. Understanding the data thoroughly needs its exploration from every aspect. The impactful features enable making meaningful and beneficial decisions; therefore, EDA occupies an invaluable place in Data science and machine learning projects.

Advantages of Using EDA:

Here are a few advantages of using Exploratory Data Analysis -

1. Gain Insights Into Underlying Trends and Patterns

EDA assists data analysts in identifying crucial trends quickly through data visualizations using various graphs, such as box plots and histograms. Businesses also expect to make some unexpected discoveries in the data while performing EDA, which can help improve certain existing business strategies.

2. Improved Understanding of Variables

Data analysts can significantly improve their comprehension of many factors related to the dataset. Using EDA, they can extract various information such as averages, means, minimum and maximum, and more such information is required for preprocessing the data appropriately.

3. Better Preprocess Data to Save Time

EDA can assist data analysts in identifying significant mistakes, abnormalities, or missing values in the existing dataset. Handling the above entities is critical for any organization before beginning a full study as it ensures correct preprocessing of data and may help save a significant amount of time by avoiding mistakes later when applying machine learning models.

4. Make Data-driven Decisions

The most significant advantage of employing EDA in an organization is that it helps businesses to improve their understanding of data. With EDA, they can use the available tools to extract critical insights and make conclusions, which assist in making decisions based on the insights from the EDA.

The resource person delivered her guest lecture on Exploratory Data analysis (EDA) with following topics:

1. Data Cleaning: EDA involves examining the information for errors, lacking values, and inconsistencies. It includes techniques including records imputation, managing missing statistics, and figuring out and getting rid of outliers.
2. Descriptive Statistics: EDA utilizes precise records to recognize the important tendency, variability, and distribution of variables. Measures like suggest, median, mode, preferred deviation, range, and percentiles are usually used.
3. Data Visualization: EDA employs visual techniques to represent the statistics graphically. Visualizations consisting of histograms, box plots, scatter plots, line plots, heatmaps, and bar charts assist in identifying styles, trends, and relationships within the facts.
4. Feature Engineering: EDA allows for the exploration of various variables and their adjustments to create new functions or derive meaningful insights. Feature engineering can contain scaling, normalization, binning, encoding express variables, and creating interplay or derived variables.
5. Correlation and Relationships: EDA allows discover relationships and dependencies between variables. Techniques such as correlation analysis, scatter plots, and pass-tabulations offer insights into the power and direction of relationships between variables.
6. Data Segmentation: EDA can contain dividing the information into significant segments based totally on sure standards or traits. This segmentation allows advantage insights into unique subgroups inside the information and might cause extra focused analysis.
7. Hypothesis Generation: EDA aids in generating hypotheses or studies questions based totally on the preliminary exploration of the data. It facilitates form the inspiration for in addition evaluation and model building.
8. Data Quality Assessment: EDA permits for assessing the nice and reliability of the information. It involves checking for records integrity, consistency, and accuracy to make certain the information is suitable for analysis.

Queries and Feedback session:

- Students asked few queries on Data analysis, and its related areas, Resource person answered for all those queries.
- Students given feedback on the Online guest lecture and said they felt happy for being involved in this type of session as they got clear picture about Data analysis and visualization tools.

Some key takeaways from this Guest lecture are:

- EDA is subjective as it summarizes the features and characteristics of a dataset. So, depending on the project, data scientists can choose from the various plots discussed in this guest lecture to explore the data before applying machine learning algorithms.
- Since the nature of EDA depends on the data, we can say that it is an approach instead of a defined process.
- EDA presents hidden insights from data through visualizations such as graphs and plots.
- Graphical and non-graphical statistical methods can be used to perform EDA.
- Univariate analysis is simpler than multivariate analysis.
- The success of any EDA will depend on the quality and quantity of data, the choice of tools and visualization, and its proper interpretation by a data scientist.
- EDA is crucial in AI-driven businesses such as retail, e-commerce, banking and finance, agriculture, healthcare, and so on.

The session ended by **Dr. M. Sreedevi, Professor, Head of the Department/CST, MITS,** thanking the resource person, faculty members and students and expressed her gratitude to the Management and Principal for giving permission and financial support to organize this programme.