



NEW HORIZON COLLEGE OF ENGINEERING

New Horizon Knowledge Park, Ring Road, Marathalli
Autonomous College Permanently Affiliated to VTU, Approved by AICTE & UGC
Accredited by NAAC with 'A' Grade, Accredited by NBA

A Hands-on workshop on “Data Viz: NLP Sentiment Analytics”

On June 7, 2024, the Alumni Association and the Department of Computer Science and Engineering (Data Science) at New Horizon College of Engineering organized a workshop titled "Data Viz: NLP Sentiment Analytics." This workshop was held from 09:30 AM to 11:00 AM at the Data Science Computer Lab-1. The session was conducted by Mr. Krishnav Dave, Founder & CEO of PreProd Corp. This report summarizes the activities and learnings from the workshop, focusing on the processes of procuring data, storing it in a local environment, performing web scraping using BeautifulSoup (bs4), and understanding the concepts of monolithic and microservices architectures.

The workshop aimed to provide participants with practical knowledge on how to perform sentiment analysis using natural language processing (NLP) techniques and visualize the results effectively. The specific focus was on working with data related to U.S. presidents.

Key Topics Covered

1. Procuring Data

- **Data Sources:** Identification of reliable sources for gathering data on U.S. presidents, such as official websites, news articles, and historical databases.
- **Data Collection:** Methods for collecting textual data, speeches, and public opinions.

2. Storing Data in Local Environment

- **Environment Setup:** Steps to set up the local environment for data storage and analysis. This includes installing necessary Python libraries and tools.
- **Data Storage:** Techniques for storing the collected data in a structured format, such as CSV or JSON files, to facilitate easy retrieval and processing.

3. Web Scraping Using BeautifulSoup (bs4)

- **Introduction to BeautifulSoup:** Overview of BeautifulSoup, a Python library used for web scraping purposes.
- **Web Scraping Process:** Detailed steps on how to scrape data from websites. This involves:
 - **Sending HTTP Requests:** Using libraries such as `requests` to fetch the content of web pages.
 - **Parsing HTML Content:** Utilizing BeautifulSoup to parse the HTML content of the web pages.
 - **Extracting Data:** Techniques for extracting specific pieces of information (e.g., text, links, and images) from the parsed HTML content.
- **Storing Scraped Data:** Methods for storing the scraped data in a structured format for further analysis.

4. Monolithic and Microservices Architectures

- **Monolithic Architecture:** Understanding the traditional approach where a single, unified application is responsible for multiple tasks. Monolithic systems are simpler to develop initially but can become difficult to maintain and scale as they grow.

- **Microservices Architecture:** Learning about the modern approach where an application is divided into small, independent services that communicate over well-defined APIs. This architecture improves scalability, maintainability, and flexibility.

Practical Session: Working with Presidential Data


During the practical session, participants engaged in a hands-on project involving the analysis of U.S. presidential data. The following steps were undertaken:

1. **Procurement of Data**
 - Participants identified key websites and sources containing speeches, public statements, and biographical information of U.S. presidents.
 - Data was collected using manual downloads and automated scripts.
2. **Local Storage Setup**
 - The local environment was set up with necessary tools, including Python, Jupyter Notebook, and libraries such as pandas, BeautifulSoup, and requests.
 - Data was stored in CSV files for easy access and manipulation.
3. **Web Scraping Exercise**
 - **Step-by-Step Guide:**
 - **Import Libraries:** Imported necessary libraries (`requests`, `BeautifulSoup`, `pandas`).
 - **Fetch Web Page:** Used the `requests` library to fetch the HTML content of a web page containing presidential data.
 - **Parse HTML:** Utilized `BeautifulSoup` to parse the HTML content.
 - **Extract Information:** Extracted relevant information, such as the text of speeches and public statements.
 - **Store Data:** Saved the extracted data into a CSV file for subsequent analysis.

Understanding Monolithic and Microservices Architectures

- **Monolithic Architecture:** Participants learned that a monolithic architecture involves a single application handling all aspects of a task. While simpler to develop initially, monolithic systems can become complex and unwieldy as they scale.
- **Microservices Architecture:** Participants explored how microservices architecture divides an application into smaller, independent services that can be developed, deployed, and maintained separately. This approach enhances scalability and flexibility, making it easier to manage large applications.

The "Data Viz: NLP Sentiment Analytics" workshop provided participants with valuable insights into the processes of procuring and storing data, as well as the practical application of web scraping techniques using BeautifulSoup. The inclusion of monolithic and microservices architectures helped participants understand different approaches to building and maintaining applications. The hands-on experience with U.S. presidential data allowed participants to apply theoretical knowledge in a practical setting, enhancing their understanding of data collection, storage, and analysis.



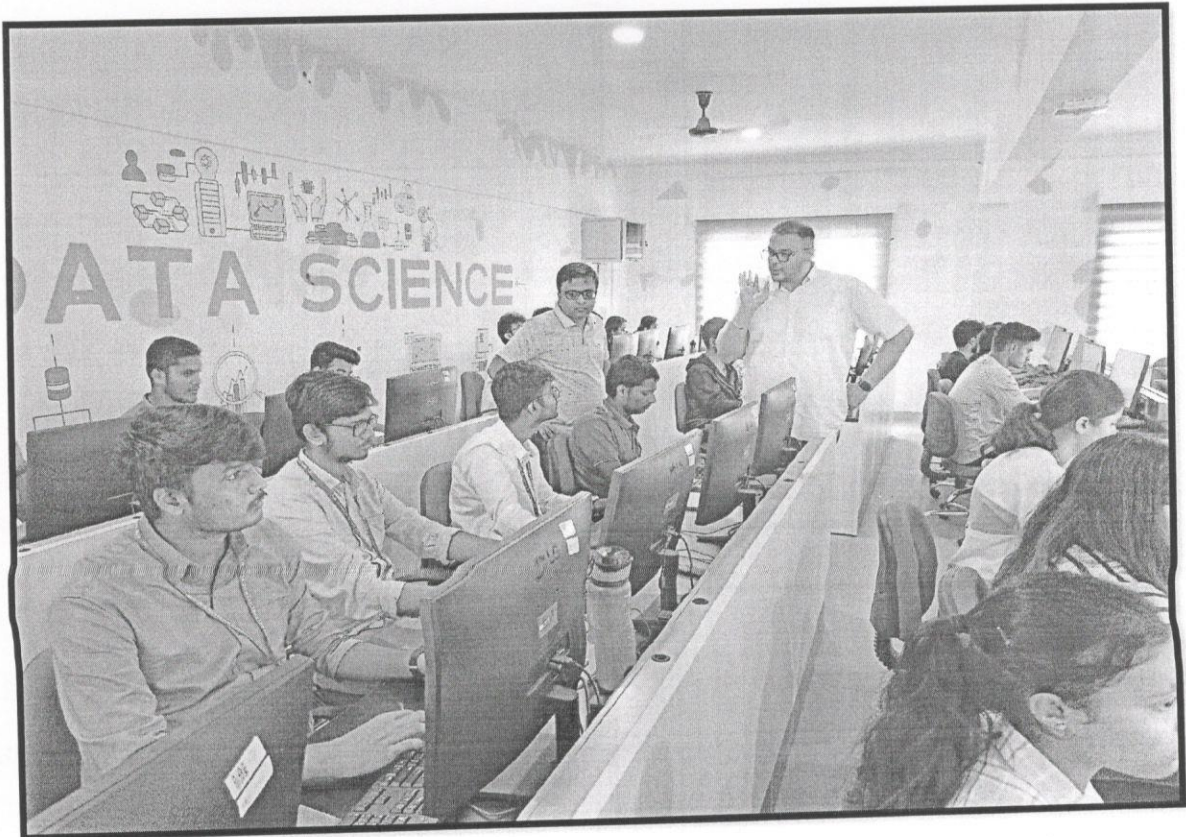
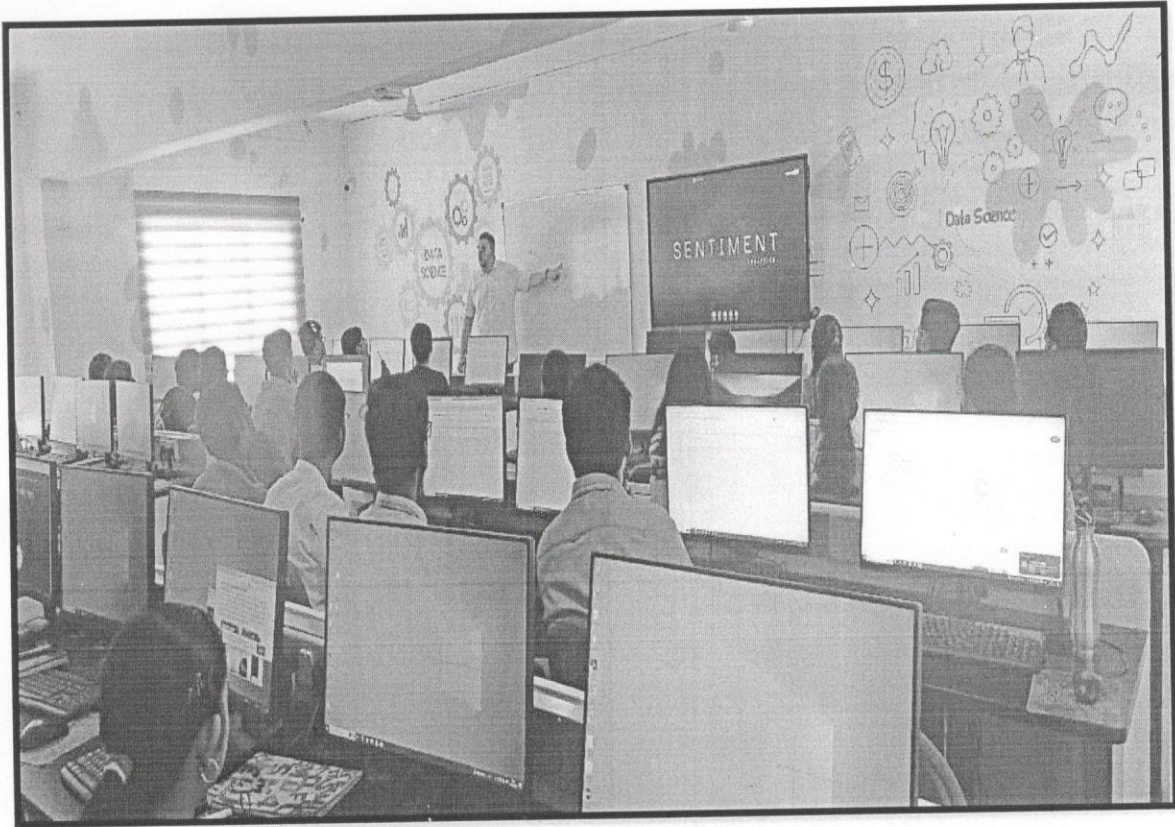
Faculty Coordinator



HOD-CSE(DS)

Head of Department - CSE (Data Science)
New Horizon College of Engineering
Ring Road, Bellandur Post, Bengaluru - 560 103

Glimpses of the event:





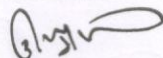
NEW HORIZON COLLEGE OF ENGINEERING

Autonomous College Permanently Affiliated to VTU, Approved by AICTE & UGC
Accredited by NAAC with 'A' Grade. Accredited by NBA

Department of Computer Science & Engineering (Data Science)

Participants list for "Data Viz: NLP Sentiment Analytics" Workshop (07-06-24)

S No	USN	Name	Signature
1	1NH22CD001	A LALITH	A. Lalith
2	1NH22CD002	ABHI S	Abhi.S
3	1NH22CD014	ANKITHA SURAKSHA V	Ankitha
4	1NH22CD015	ARCHANA K	Archana
5	1NH22CD024	BHOOMIKA AMBATI	Bhoomika
6	1NH22CD027	DARSHAN DIDAGUR	Darshan
7	1NH22CD028	DHEERAJ ARVIND	Dheeraj
8	1NH22CD029	DIVYESH RAJ PILLAI	Divyesh
9	1NH22CD034	HARSH KUMAR	Harsh
10	1NH22CD035	ISHANVI REDDY	Ishanvi
11	1NH22CD037	JANIYA MARIA TOMY	Janiya
12	1NH22CD038	JEEVITHA S	Jeevitha.S
13	1NH22CD042	K NIRANJAN	Nirajan
14	1NH22CD043	KANCHUKOMMALA NAVEEN	K. Naveen
15	1NH22CD049	KHUSHI NAGARAJ	Khushi
16	1NH22CD050	KHUSHI SHRIMALI	Khushi
17	1NH22CD061	MANISH KUMAR MALAKAR	Manish
18	1NH22CD064	MANOJ P	Manoj
19	1NH22CD073	NARAYAN PRASHANT NAIK	Narayan
20	1NH22CD075	NIHAL MANOHAR	Nihal
21	1NH22CD077	P HARSHAD ALI KHAN	P. Harshad
22	1NH22CD079	PENMATSA VENKATA SATYA SAHASRA	Penmatsa
23	1NH22CD081	POORNIMA	Poornima
24	1NH22CD083	R DEEPIYA SREE	R. Deepya
25	1NH22CD085	RAKSHITHA G	Rakshitha.G
26	1NH22CD092	ROHAN M	Rohan
27	1NH22CD094	RUCHITHA S K	Ruchitha
28	1NH22CD104	SKANDA PRIYA V	Skandapriya
29	1NH22CD105	SNEHA	Sneha
30	1NH22CD106	SRI LIKITHA NANNURI	Sri Likitha
31	1NH22CD108	SRINIVAS K S	Srinivas
32	1NH22CD111	SURYA YASHASWINI P V N	Surya
33	1NH22CD115	TANISHA SWAMI	Tanisha
34	1NH22CD116	TANUSHREE D KOTIAN	Tanushree
35	1NH22CD122	VENKAT LAKSHMI SHRUTHI A	Venkat Lakshmi
36	1NH22CD127	SREEJITH S	Sreejith
37	1NH22CD128	HRITHIK KUMAR	HRITHIK
38	1NH22CD129	CHINTHAGINJALA TEENU DEEKSHITH	Teenu deekshith
39	1NH22CD139	VIPUL RAJ	Vipul


Coordinator