|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SCHOOL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE** | | | | | **DEPARTMENT OF COMPUTER SCIENCE ENGINEERING** | | | | |
| **Program Name:** B. Tech | | | | **Assignment Type: Lab** | | | **Academic Year:**2025-2026 | | |
| **Course Coordinator Name** | | | | Venkataramana Veeramsetty | | | | | |
| **Instructor(s) Name** | | | | |  | | --- | | Dr. V. Venkataramana (Co-ordinator) | | Dr. T. Sampath Kumar | | Dr. Pramoda Patro | | Dr. Brij Kishor Tiwari | | Dr.J.Ravichander | | Dr. Mohammand Ali Shaik | | Dr. Anirodh Kumar | | Mr. S.Naresh Kumar | | Dr. RAJESH VELPULA | | Mr. Kundhan Kumar | | Ms. Ch.Rajitha | | Mr. M Prakash | | Mr. B.Raju | | Intern 1 (Dharma teja) | | Intern 2 (Sai Prasad) | | Intern 3 (Sowmya) | | NS\_2 ( Mounika) | | | | | | |
| **Course Code** | | | 24CS002PC215 | **Course Title** | | AI Assisted Coding | | | |
| **Year/Sem** | | | II/I | **Regulation** | | R24 | | | |
| **Date and Day**  **of Assignment** | | | Week5 - Monday | **Time(s)** | |  | | | |
| **Duration** | | | 2 Hours | **Applicable to**  **Batches** | |  | | | |
| **AssignmentNumber: 9.1**(Present assignment number)/**24**(Total number of assignments) | | | | | | | | | |
|  | **Q.No.** | **Question** | | | | | | ***Expected Time***  ***to complete*** |  |
|  | 1 | **Lab 17– AI for Data Processing: Data cleaning and preprocessing scripts**  The objective of this lab is to enable students to understand and apply **AI-assisted coding tools** for automating and enhancing data preprocessing tasks. Students will:   1. Gain practical experience in **cleaning, transforming, and standardizing real-world datasets** with issues such as missing values, duplicates, outliers, inconsistent formats, and noisy text. 2. Learn to **leverage AI coding assistants** to generate preprocessing scripts, while critically evaluating and refining the AI-generated code for accuracy, efficiency, and best practices. 3. Develop the ability to design **end-to-end preprocessing pipelines** that prepare raw data for downstream machine learning and analytics applications. 4. Build confidence in **combining human expertise with AI assistance**, ensuring data quality and integrity in diverse domains such as customer feedback, healthcare, and finance.   **Lab Question 1: Customer Feedback Dataset**  You are given a CSV file containing customer feedback collected from an e-commerce website. The dataset includes columns: customer\_id, feedback\_text, rating, and date. However, the file has many missing values, typos, and inconsistent date formats.   * **Task 1:** Use an AI-assisted coding tool to generate a script that detects and fills missing rating values with the column’s median and standardizes the date column into YYYY-MM-DD format. * **Task 2:** Clean the feedback\_text column by removing stopwords, correcting common spelling mistakes, and converting text to lowercase using AI suggestions. Compare the AI-generated preprocessing code with your manually written version.   **Prompt:**  Write a Python script that fills missing rating values with the column’s median, converts dates into YYYY-MM-DD format, and cleans the feedback text by lowercasing, removing stopwords, and fixing common spelling mistakes.  **Code:**  **Screenshot 2025-10-25 184539**  Screenshot 2025-10-25 184555  **Comparison:**  **AI version**: Fast to generate, suitable for simple data.  **Manual refined version**: More **robust**, **readable**, and **error-tolerant**, ideal for production or evaluation purposes.  **Lab Question 2: Medical Records Dataset**  A hospital provides you with a dataset of anonymized medical records containing attributes like patient\_id, age, gender, blood\_pressure, and cholesterol. Some columns include outliers and inconsistent categorical labels (e.g., Male, M, male).   * **Task 1:** Write a script (with AI assistance) to detect and handle outliers in the blood\_pressure column using statistical methods (e.g., IQR or z-score). * **Task 2:** Standardize categorical values in the gender column and encode them into numeric form. Let the AI-assisted coding tool propose the preprocessing pipeline, then refine the pipeline manually based on your understanding.   **Prompt:**  Write a Python script that detects and handles outliers in a medical dataset using the IQR method and standardizes inconsistent gender labels (like ‘Male’, ‘M’, ‘male’) before encoding them numerically.  **Code:**  **Screenshot 2025-10-25 185419**  Screenshot 2025-10-25 185433  **Lab Question 3: Financial Transactions Dataset**  A bank gives you transaction data with columns: transaction\_id, amount, currency, timestamp, and merchant. The dataset contains multiple issues: different currency units (USD, INR, EUR), timestamps in various time zones, and duplicated rows.   * **Task 1:** Use AI-assisted coding to write a script that removes duplicate transactions and converts all amount values into a single currency (e.g., USD) using a provided conversion dictionary. * **Task 2:** Normalize the timestamp column into UTC format and create a new column transaction\_hour for downstream time-series analysis. Compare the AI’s preprocessing code against your own optimized version.   **Prompt:**  Write a Python script that removes duplicate transactions from a financial dataset, converts all amounts into USD using given exchange rates, and standardizes timestamps into UTC while creating a transaction\_hour column  **Code:**  **Screenshot 2025-10-25 185909Screenshot 2025-10-25 185928** | | | | | | Week 9- Monday |  |