Module 1 Cheatsheet: Data Science and Generative AI

Popular GenAI tools

Name of mode	el Usage	Link
Data Robot	A simple tool useful for data analysis and model building operations	https://www.datarobot.com/
Mostly.AI	Synthetic data generation	https://mostly.ai/
ChatGPT	GPT based model used for text and code generation based on natural language querie	s https://openai.com/chatgpt
DB Sensei	Generate SQL queries for databases using natural language queries	https://dbsensei.com/

Important prompts for data preparation

Task

Read a CSV data file and load it to a data frame.

Data cleaning: Identify and replace missing values per the following guidelines.

- 1. You replace the missing entries in columns containing categorical values with the 1. Identify the attributes with missing values. most frequent entries
- 2. You replace the missing entries in columns with continuous data with the mean value of the column.
- 3. If a value is missing in the target column, you may need to drop that row

Data Normalization: Normalize an attribute to its maximum value.

Converting categorical variable into indicator variables

Prompt Write a Python code that can perform the following tasks:

Read the CSV file, located on a given file path, into a Pandas data frame, assuming that the first rows of the file are the headers for the data. Write a Python to perform the following tasks:

- 2. Segregate these attributes into categorical and continuous valued attributes.
- 3. Drop the entire row if the value is missing in the target variable.
- 4. If the value is missing in a categorical attribute, replace the missing values with the most frequent value in the column.
- 5. If the value is missing in a continuous value attribute, replace the missing values with the mean value of the entries in the column.

Write a Python code to normalize the content under a given attribute in a data frame df to its maximum value. Make changes to the original data, and do not create a new attribute.

Write a Python code to perform the following tasks.

- 1. Convert a data frame df attribute into indicator variables, saved as df1, with the naming convention "Name <unique value of the attribute>".
- 2. Append df1 into the original data frame df.
- 3. Drop the original attribute from the data frame df.

Author(s)

Abhishek Gagneja

