

Machine Learning ICP 6

Load a dataset in CSV format from Kaggle or any public API like UCI.

Note: You can work on different datasets to perform ICP tasks.

1. Load a CSV dataset using Pandas.
2. Display the first and last 5 rows.
3. Get dataset information (.info()) and summary statistics (.describe()).
4. Select specific columns from a DataFrame.
5. Filter rows based on conditions (e.g., all rows where age > 25).
6. Use .loc[] and .iloc[] for selection.
7. Identify missing values in a dataset.
8. Fill missing values using mean/median/mode.
9. Drop rows or columns with too many missing values.
10. Remove duplicate values.
11. Convert data types (e.g., string to datetime).
12. Normalize column names (remove spaces, change to lowercase).
13. Group data by a categorical column and calculate summary statistics.
14. Use .groupby() with functions like sum(), mean(), count().
15. Sort the dataset by multiple columns.
16. Set and reset indexes for better data handling.
17. Merge two DataFrames using different join types (inner, outer, left, right).
18. Concatenate DataFrames vertically and horizontally.
19. Generate basic plots (histogram, boxplot, line plot) using .plot().
20. Use Pandas with Matplotlib for advanced visualizations.