

PROJECT # 02 – ***AUTO ML AGENT***

|  |
| --- |
| TOPIC : PROJECT # 02 – *AUTO ML AGENT* |
| COURSE CODE : CIC - 201 |
| SUBMITTED TO : Maham Ashraf |
| SUBMITTED BY : Umar Ayaz | 22FA – 008 – CS |

**SRS DOCUMENTATION (AUTO ML AGENT)**

**1. *Introduction***

**1.1 Purpose**

AutoMLAgent is a tool designed to make machine learning easier and faster for everyone. It takes care of tedious tasks like cleaning and preparing data, so users can focus on analyzing results and building better models. Whether you're a beginner or an experienced data analyst, AutoMLAgent simplifies the process of creating classification and regression models, saving you time and effort.

**1.2 Scope**

AutoMLAgent handles everything from data cleaning to running machine learning models. Here's what it does:

* Converts text data into numbers so it can be processed.
* Deals with missing values in the dataset by filling them automatically.
* Scales numeric data using smart techniques to improve accuracy.
* Finds the most important feature in your dataset for predictions.
* Lets you choose between classification and regression models.
* Shows useful visualizations and performance metrics to understand how well the model works.

This tool is perfect for students, data analysts, and developers who want a quick and easy way to explore data and build machine learning models.

**1.3 Key Terms**

* **Naive Bayes**: A simple but powerful method for classification tasks.
* **Linear Regression**: A method to predict numbers (like house prices) based on existing data.
* **StandardScaler**: A tool that adjusts numbers to make them easier for models to process.
* **MinMaxScaler**: A tool that squeezes numbers into a specific range, usually between 0 and 1.

**1.4 References**

* Pandas Documentation: <https://pandas.pydata.org/>
* Scikit-learn Documentation: <https://scikit-learn.org/>

**2. *Overview***

**2.1 What It Does**

AutoMLAgent makes building machine learning models easy by combining the most common steps into one tool. Instead of writing code for each step, you can use this tool to handle everything with just a few inputs.

**2.2 Features**

1. **Data Cleaning and Preparation**:
   * Converts text data into numbers.
   * Fills missing values with smart guesses (like averages or most common values).
   * Scales numbers to make the data consistent.
2. **Feature Selection**:
   * Automatically finds the feature in your dataset that has the most impact on predictions.
3. **Model Selection**:
   * **Classification**: Builds models to classify data into categories (like spam vs. non-spam).
   * **Regression**: Builds models to predict numbers (like stock prices).
4. **Visualizations**:
   * Creates scatter plots and box plots to show relationships between data features.
5. **Performance Metrics**:
   * Provides accuracy scores, confusion matrices, and error rates to evaluate how well the model works.

**2.3 Who Can Use It**

* **Students**: Great for learning how data preparation and machine learning work.
* **Data Analysts**: Helps analyze datasets and build quick models for insights.
* **Developers**: A fast way to prototype machine learning workflows without writing repetitive code.

**2.4 System Requirements**

* Works on Windows, macOS, or Linux.
* Requires Python 3.7 or higher and libraries like Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn.

**2.5 Constraints**

* The input file must be a CSV.
* The user needs to specify which column is the target.

**2.6 Assumptions**

* Users have a basic understanding of data and machine learning.
* The dataset doesn’t have duplicate column names.

**3. *What AutoMLAgent Can Do***

**3.1 Core Features**

1. Load a CSV file and understand its structure.
2. Handle missing or text-based data automatically.
3. Scale numbers in the dataset if needed.
4. Identify the most important feature in the data.
5. Build a model (classification or regression) based on user input.
6. Automatically pick the best classification method based on the type of data (Gaussian for continuous data, Categorical for discrete data).
7. Visualize data relationships and show performance metrics.

**3.2 Nice-to-Have Features**

* It’s fast! AutoMLAgent can handle datasets of up to 10,000 rows in just a few seconds.
* It provides clear instructions at every step to guide users.
* It gracefully handles errors, like invalid inputs or missing files.

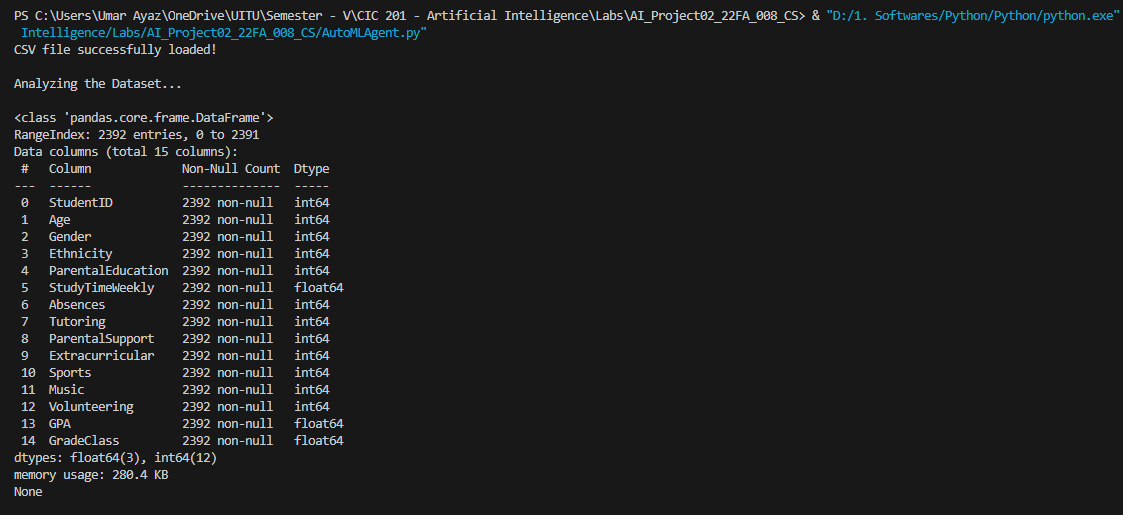
**4. *How It Works***

**4.1 Example Workflow**

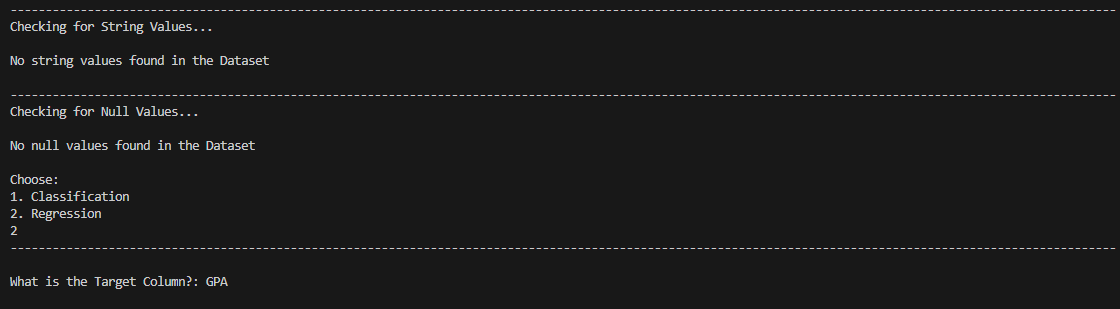
1. **Load the Data**: Upload a CSV file (e.g., Student\_Performance.csv).
2. **Prepare the Data**: Handle missing values, convert text to numbers, and scale numeric data.
3. **Select the Model**: Choose between classification and regression.
4. **Visualize Results**: View performance metrics and plots showing data relationships.

**USING REGRESSION**

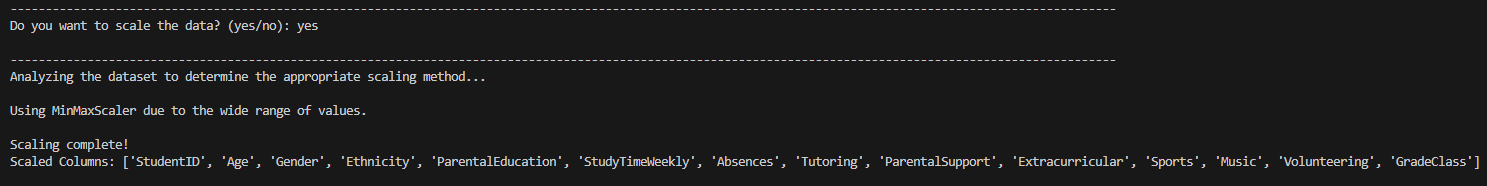
*Information about the Dataset*

**

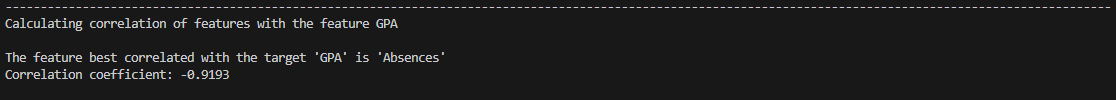
*Checking for String & Null Values | Choosing Regression and the Target Column*

**

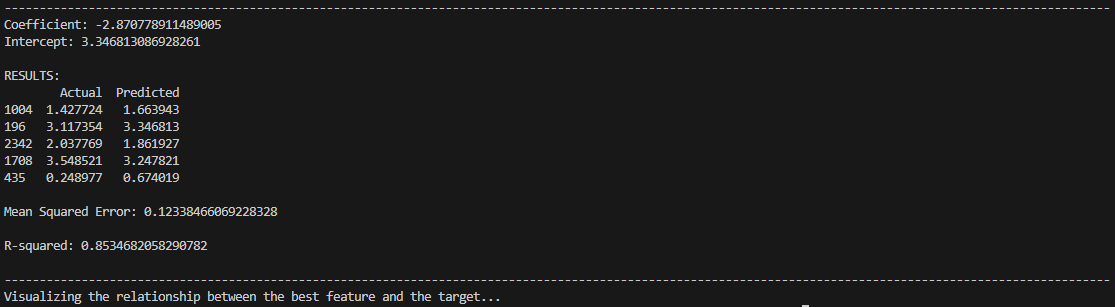
*Scaling the Dataset*



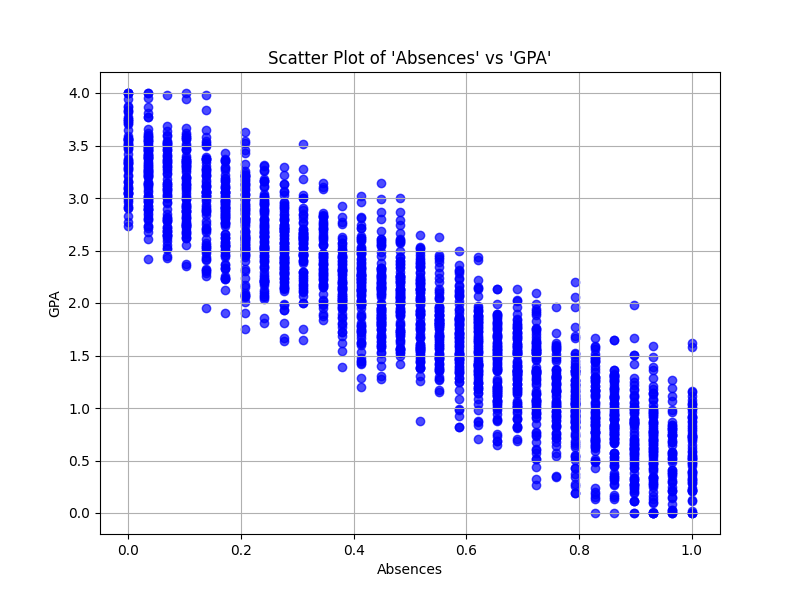
*Finding out the Correlation with the Target Feature*



*The Final Performance Metrics*

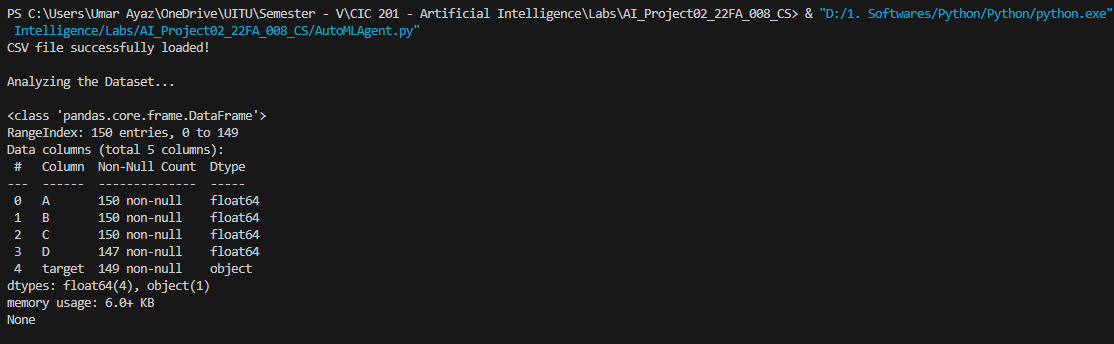


*The Relationship between the Best Feature & the Target*

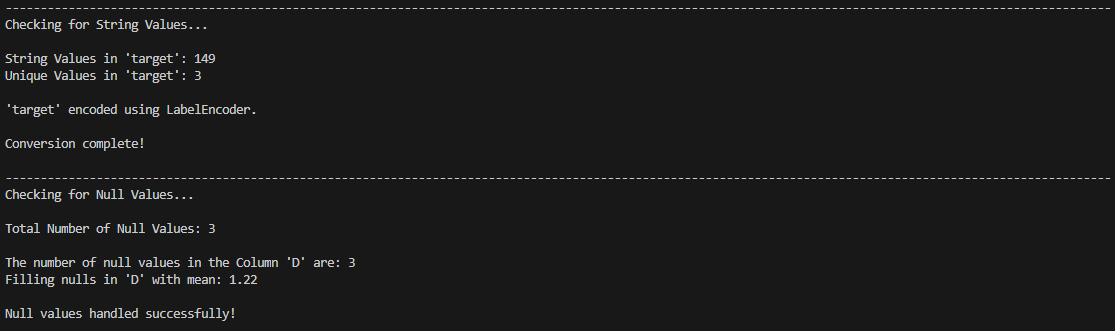
**

**USING CLASSIFICATION**

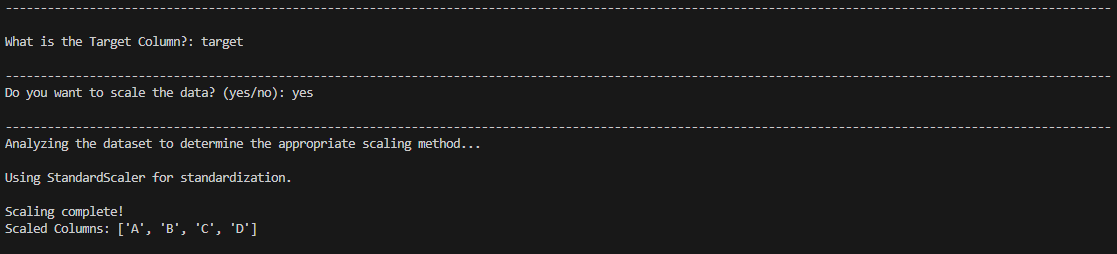
*Information about the Dataset*

**

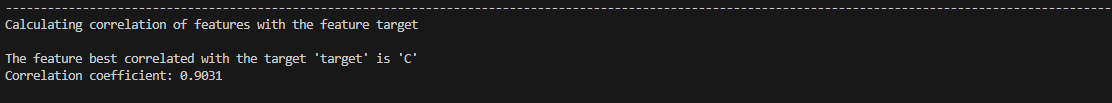
*Converting String Values to Numeric Ones & Filling in the Null Values*



*Scaling the Dataset*

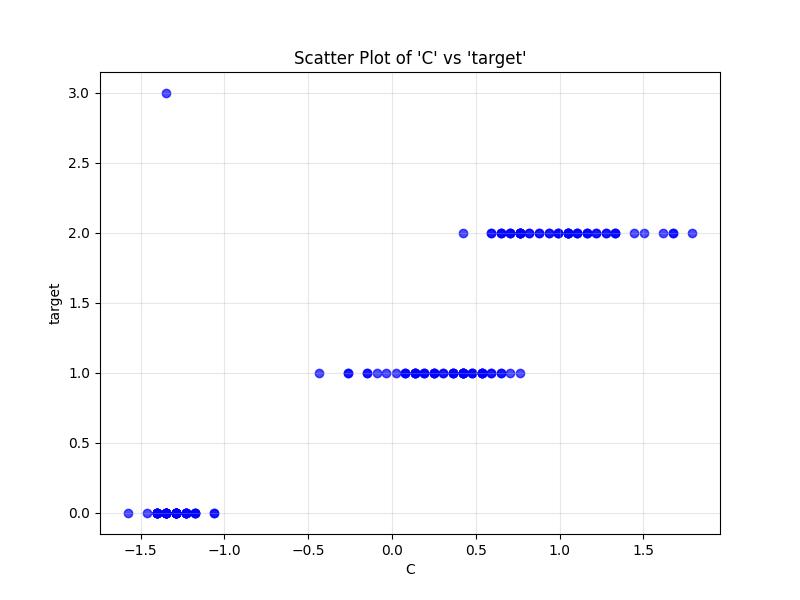


*Finding out the Correlation with the Target Feature*



*The Relationship between the Best Feature and the Target*





*The Final Performance Metrics*

