**AI4ALL Personal Project Guide**

Answer the following questions to guide you in developing your personal project. You can also consider making slides to showcase everything you have done.

Use screenshots or images wherever possible to help you illustrate your point in this document.

You should use the code provided in the personal project notebook, but adapt it and build on it using what you learn by reading articles.

**Part 1: Exploring the Data**

1. Which dataset are you using?
2. Describe your dataset. How many rows are there? What are the names of some interesting columns? Which data type are the columns?
3. How did you explore your dataset? What did you learn about the data and what patterns did you find? Describe the methods you used.
4. Did you encounter any difficulties with your dataset? How did you work around them? (examples: missing data, messy data, mixed data types, etc.)

**Part 2: Feature Engineering**

1. What new features did you create based on your existing dataset? For each new feature, describe how you created it and why you think it will be more useful later.

|  |  |  |
| --- | --- | --- |
| **New Feature Name** | **How it was created** | **Why it will be useful** |
|  |  |  |
|  |  |  |
|  |  |  |

**Part 3: Unsupervised Machine Learning (Clustering)**

1. What question do you want to answer using clustering and your data?
2. What type of clustering algorithms did you try? How well did each work?
3. If you had to pick a final algorithm to answer your question, which would it be? Why?

**Part 4: Supervised Machine Learning (Classification)**

1. What question do you want to answer using classification and your data?
2. What are you trying to classify (or predict)? How many options are there for that variable (e.g., predicting high/medium/low is 3 options)? How does the accuracy of your classification algorithm change if you increase or decrease the number of options being predicted?

1. What types of classification algorithms did you try? How well did each work?
2. If you had to pick a final algorithm to answer your question, which would it be? Why?

**Optional: Web Scraping**

1. Where did you get the data you decided to web scrape?
2. Would this data be good for machine learning? Why or why not?