

ML Presubmission Guidelines

1. Title of the Project

2. Problem Statement

- Clearly define the problem you're solving.
- Mention whether it's a classification, regression, clustering, or recommendation task.
- Include real-world application or motivation behind the problem.

3. Data Collection

A. Web Scraping (if applicable)

- Mention the source URL(s).
- Tools/Libraries Used: `BeautifulSoup`, `Selenium`, `Scrapy`, etc.
- Ethical scraping (respect `robots.txt`, delay requests).
- Include a brief explanation or sample code snippet for scraping.

B. Dataset Info

- Dataset size (rows × columns)
- Data format (CSV, JSON, SQL, etc.)
- Number of features
- Target variable (if supervised learning)

4. Data Preprocessing

- **Handling Missing Values:** Fill, drop, interpolate?
- **Encoding:** One-hot, Label encoding?
- **Normalization/Scaling:** MinMaxScaler, StandardScaler?
- **Data Splitting:** Train-Test (e.g., 80-20 split)
- **Outlier Detection and Treatment**
- **Feature Engineering:** Derived features, polynomial features, etc.

5. Exploratory Data Analysis (EDA)

- Use visualizations: `matplotlib`, `seaborn`, `plotly`
- Correlation heatmap
- Distribution plots (histograms, boxplots)
- Target vs Feature plots

- Class imbalance check (for classification)
- Key insights derived

6. Modeling

- Algorithms used (explain choice): e.g., Logistic Regression, Random Forest, SVM, XGBoost, etc.
- Hyperparameter tuning: GridSearchCV / RandomizedSearchCV
- Cross-validation strategy used (if any)

7. Evaluation Metrics

- **Classification:** Accuracy, Precision, Recall, F1, ROC-AUC
- **Regression:** RMSE, MAE, R^2
- **Clustering:** Silhouette Score, Davies-Bouldin Index
- Include confusion matrix/ROC curve if applicable

8. Results & Interpretation

- Summarize performance of different models
- Compare training vs validation performance
- Final model selection & justification
- Limitations or assumptions

9. Conclusion

- Brief summary of findings
- Future work (e.g., better data, advanced models, deployment)