ML template

28 June 2024 11:55

1. - Problem Definition
     - Define the problem you're trying to solve
     - Identify the type of machine learning task (classification, regr
     - Set clear objectives and success criteria

2. - Data Collection
     - Gather relevant data from available sources
     - Ensure data quality and relevance to the problem

3. - Data Preprocessing
     - Load and inspect the data
     - Handle missing values
     - Convert data types if necessary
     - Handle outliers
     - Encode categorical variables

4. - Exploratory Data Analysis (EDA)
     - Analyze data distributions
     - Identify patterns and relationships
     - Visualize key insights
     - Generate hypotheses

5. - Feature Engineering
     - Create new features based on domain knowledge and insigh
     - Select relevant features
     - Perform feature scaling or normalization if required

6. - Data Splitting
     - Split data into training, validation, and test sets

ression, clustering, etc.)

ts from EDA

7. Model Selection
    - Choose appropriate algorithms based on the problem type a
    - Consider model interpretability and complexity

8. Model Training
    - Train selected models on the training data
    - Use cross-validation to assess model performance

9. Model Evaluation
    - Evaluate models on the validation set
    - Use appropriate metrics (accuracy, F1-score, RMSE, etc.)
    - Analyze model errors and limitations

10. Hyperparameter Tuning
    - Optimize model hyperparameters using techniques like grid s

11. Final Model Selection and Testing
    - Select the best-performing model
    - Evaluate the final model on the test set

12. Model Interpretation
    - Analyze feature importance, shaps
    - Interpret model coefficients or decision rules and final outpu

13. Deployment Planning
    - Prepare the model for deployment (if applicable)
    - Consider scalability and maintenance requirements

14. Documentation and Reporting
    - Document the entire process, including decisions and ration
    - Prepare a final report or presentation of findings

15. Continuous Monitoring and Updating
    - Plan for model monitoring in production
    - Set up processes for regular model updates and retraining

nd data characteristics

search or random search

t

ale