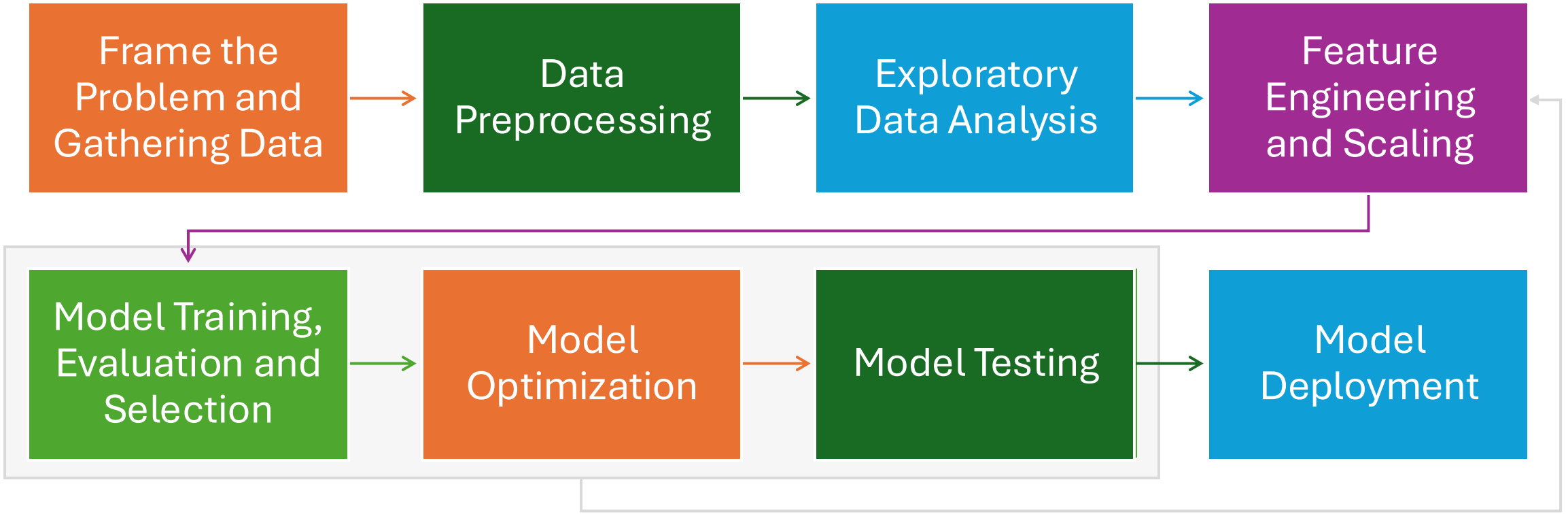
Complete Workflow of a Machine Learning Project

IRON KAGGLE

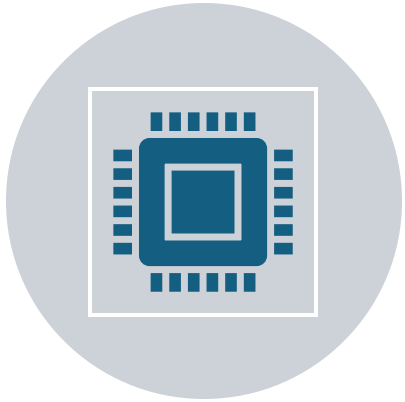
ML Workflow


Frame the Problem and Gathering Data
Data Preprocessing
Exploratory Data Analysis
Feature Engineering and Scaling
Model Training, Evaluation and Selection
Model Optimization
Model Testing
Model Deployment

Data Preprocessing and Exploration

EXPLORATORY DATA ANALYSIS (EDA): **DISTRIBUTION**, **OUTLIERS**.

APPLIED TRANSFORMATIONS: SCALING, ENCODING, HANDLING MISSING VALUES.



DATA SPLIT: TRAINING SET (80%) AND TEST SET (20%).

Building and Evaluating the Model

- Model Selection
  - Use **pipeline** process with a list of Regression models.
  - Evaluated with metrics: **MAE**, MSE, RMSE, **R^2**.
  - Selection: **RandomForestRegresor**
    - **R2: 0.9486, MAE: 537.10**
- Train the RandomForestRegresor model
  - Data split: training set (80%) and test set (20%).
- Optimization and Hyperparameter tuning:
  - **GridSearchCV**.
- Evaluate the Results
- Save the model

Results and Insights of the Project

Current Model Evaluation Metrics

- MAE: 449.10
- MSE: 462461.12
- RMSE: 680.04
- **R2: 0.94**

Prediction for a Real Data Set

- A slight improvement (0.95 or 0.96), if the new data are consistent with previous patterns.
- **If there is additional noise as the Original Training Data, might decrease slightly (0.92 or 0.93).**

Insights

- Don't skip workflow steps
- The model gave you what you gave it
- All results are Key results

Next steps

- Test with real DataSet
- Continuous refinement