

Test Evaluation Report

Metadata

Template used: /Users/weilinhan/FixML/fixml/src/fixml/data/templates/eval_report.md.jinja

Run details

- **Checklist Path:** /Users/weilinhan/FixML/fixml/src/fixml/data/checklist/checklist.csv
- **Repository Path:** /Users/weilinhan/FixML/repos/proteiML
- **HEAD Commit:** eef65f8d65240e785ca57060d2a0b05d23635aa6
- **Start Time:** 2024-09-20 09:45:14.429417
- **End Time:** 2024-09-20 09:46:03.126450
- **Time Taken:** 0:00:48.697033
- **Total Input/Output Tokens Used:** 7916/3819
- **Successful/Failed Calls:** 5/0 (100% Success Rate)
- **Files Evaluated:**
 - /Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py
 - /Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py
 - /Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py
 - /Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py
 - /Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py
- **LLM Model Used:** gpt-3.5-turbo

Evaluation Summary

Completeness Score: 4.5/7

Completeness Score per Checklist Item:

ID Title	is_Satisfied
2.1 Ensure Data File Loads as Expected	1
3.2 Data in the Expected Format	0.5
3.5 Check for Duplicate Records in Data	1
4.2 Verify Data Split Proportion	1
5.3 Ensure Model Output Shape Aligns with Expectation	0.5
6.1 Verify Evaluation Metrics Implementation	0.5
6.2 Evaluate Model's Performance Against Thresholds	0

Evaluation Details

2.1 Ensure Data File Loads as Expected

Requirement: Verify if data-loading functions correctly fetch datasets from predefined sources, such as a data file in repository, or online repositories, such as a url. Verify if the data-loading functions gracefully handle errors when unable to load data or correctly handle edge cases.

Observations:

- (test_format_mutation_dataset.py) The code imports functions for reading sequence data, combining datasets, and checking for duplicates.
- (test_construct_new_library.py) The code snippet imports functions from 'data_processing.construct_new_library' and 'data_processing.constants' modules. It defines a fixture 'metadata_fixture' that reads a CSV file and returns the data as a dictionary. The fixture is used in the test function 'test_create_and_clean_variant_df_fixture' to create and clean a variant dataframe.
- (test_model_training.py) The code includes fixtures for loading data and preparing encoding data.
- (test_scrape_aaindex_matrix.py) The code snippet imports necessary modules and defines a test function test_aaindex_scrape_or_load. The function checks if SCRAPE_AAINDEX is True, then it scrapes aaindex indices, finds accession numbers, and saves the dataset to a CSV file. If SCRAPE_AAINDEX is False, it reads the dataset from a CSV file. The assertion checks if the shape of the property_dataset is (20, 6).
- (test_ensemble_and_cross_validate.py) The code imports data_processing.ensemble_and_cross_validate module and uses data from BASE_DIR to load datasets. The code defines fixtures for loading data files.

Function References:

- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py', 'Referenced Functions': [[read_seq_data](#)], [combine_seq_datasets](#)], [check_for_duplicates](#)]}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py', 'Referenced Functions': [[create_variant_dataframe](#)], [clean_variant_dataframe](#)]}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': [[pd.read_csv](#)]}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py', 'Referenced Functions': [[scrape_aaindex_indicies](#)], [find_accession_numbers](#)]}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py', 'Referenced Functions': [[pd.read_csv](#)]}}

3.2 Data in the Expected Format

Requirement: Verify if the data matches the expected format that specified by the function. This step must involves checking the shape of data, data types, values, and any other properties that specified by the function.

Observations:

- (test_format_mutation_dataset.py) The code processes sequence data and checks for duplicates.
- (test_construct_new_library.py)
- (test_model_training.py) The code includes functions to prepare encoding data and train models.
- (test_scrape_aaindex_matrix.py) Not applicable
- (test_ensemble_and_cross_validate.py) The code checks the columns of the loaded data and compares them with expected values.

Function References:

- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py', 'Referenced Functions': []}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py', 'Referenced Functions': []}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': [[mt.prepare_encoding_data](#)], [mt.train_knr_model](#)], [mt.train_mpnr_model](#)], [mt.train_rfr_model](#)]}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py', 'Referenced Functions': []}}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py', 'Referenced Functions': [[list](#)], [mean](#)], [type](#)]}}

3.5 Check for Duplicate Records in Data

Requirement: Verify if there is no duplicate record in the loaded data.

Observations:

- (test_format_mutation_dataset.py) The code calls the check_for_duplicates function to identify and handle duplicate records.
- (test_construct_new_library.py)
- (test_model_training.py) No specific check for duplicate records in the code.
- (test_scrape_aaindex_matrix.py) Not applicable
- (test_ensemble_and_cross_validate.py) No specific check for duplicate records in the loaded data.

Function References:

- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py', 'Referenced Functions': [[check for duplicates](#)]}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py', 'Referenced Functions': []}

4.2 Verify Data Split Proportion

Requirement: Verify if the data is split into training and testing sets by the proportion specified by the function. Verify the split by checking the actual fraction of data points in the training and test sets.

Observations:

- (test_format_mutation_dataset.py) The code does not include functionality related to data split proportion verification.
- (test_construct_new_library.py)
- (test_model_training.py) The code includes functions to split data for training and testing.
- (test_scrape_aaindex_matrix.py) Not applicable
- (test_ensemble_and_cross_validate.py) No specific verification of data split proportion in the code.

Function References:

- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': [[mt.prepare encoding data](#)]}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py', 'Referenced Functions': []}

5.3 Ensure Model Output Shape Aligns with Expectation

Requirement: Verify if the structure of the model's output matches the expected format specified by the function, such as checking the dimensions of the output versus the number of labels in classification task.

Observations:

- (test_format_mutation_dataset.py) The code does not involve model output shape verification.
- (test_construct_new_library.py)
- (test_model_training.py) The code includes functions to train different models.
- (test_scrape_aaindex_matrix.py) Not applicable
- (test_ensemble_and_cross_validate.py) The code checks the columns and dimensions of the model's output.

Function References:

- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': ['[mt.train_knr_model](#)', '[mt.train_mpnr_model](#)', '[mt.train_rfr_model](#)']}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py', 'Referenced Functions': ['[columns](#)', '[mean](#)', '[type](#)']}

6.1 Verify Evaluation Metrics Implementation

Requirement: Verify if the evaluation metrics are correctly implemented and appropriate for the model's task. Verify if the metric computations with expected values to validate correctness.

Observations:

- (test_format_mutation_dataset.py) The code does not include evaluation metrics implementation verification.
- (test_construct_new_library.py)
- (test_model_training.py) No specific check for evaluation metrics implementation in the code.
- (test_scrape_aaindex_matrix.py) Not applicable
- (test_ensemble_and_cross_validate.py) The code creates evaluation metrics dataframes and asserts their types and values.

Function References:

- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py', 'Referenced Functions': ['[type](#)', '[np.float64](#)']}

6.2 Evaluate Model's Performance Against Thresholds

Requirement: First compute evaluation metrics for both the training and testing datasets, then verify if these metrics exceed threshold values, indicating acceptable model performance.

Observations:

- (test_format_mutation_dataset.py) The code does not evaluate model performance against thresholds.
- (test_construct_new_library.py)
- (test_model_training.py) No specific check for evaluating model performance against thresholds in the code.
- (test_scrape_aaindex_matrix.py) Not applicable
- (test_ensemble_and_cross_validate.py) No specific evaluation of model performance against thresholds in the code.

Function References:

- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_format_mutation_dataset.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_construct_new_library.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_scrape_aaindex_matrix.py', 'Referenced Functions': []}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_ensemble_and_cross_validate.py', 'Referenced Functions': []}