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 09:40:39.509537
- **End Time**: 2024-09-20 09:41:30.846912
- **Time Taken**: 0:00:51.337375
- **Total Input/Output Tokens Used**: 7916/3875
- **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: 3.5/7

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	1
3.5 Check for Duplicate Records in Data	1
4.2 Verify Data Split Proportion	0.5
5.3 Ensure Model Output Shape Aligns with Expectation	0
6.1 Verify Evaluation Metrics Implementation	0
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 includes functions for reading sequence data, combining datasets, and checking for duplicates, which align with the requirement of loading data as expected.
- (test_construct_new_library.py) The code snippet imports data_processing.construct_new_library and data_processing.constants modules. It defines a fixture metadata_fixture that reads a CSV file 'Input_Data.csv' from BACKEND_DATA_DIR and returns the data as a dictionary. The fixture is used in the test_create_and_clean_variant_df_fixture test function.
- (test_model_training.py) The code imports data_processing.model_training as mt, reads data from 'combined_dataset.csv', and prepares encoding data for training models.
- (test_scrape_aaindex_matrix.py) The code snippet imports necessary modules and defines a test function test_aaindex_scrape_or_load. It 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. Finally, it asserts the shape of the property_dataset to be (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 uses pandas to read CSV files and sort values based on 'Test Set R Squared'.

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': ['pd.read_csv', 'to_dict', 'create_variant_dataframe', 'clean_variant_dataframe']}
- {'File Path': '/Users/weilinhan/FixML/repos/proteiML/tests/test_model_training.py', 'Referenced Functions': ['prepare_encoding_data']}
- {'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', 'sort_values']}

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 does not explicitly verify the format of the data as specified in the requirement.
- (test_construct_new_library.py)
- (test_model_training.py) The code does not explicitly verify the format of the data.
- (test_scrape_aaindex_matrix.py) The code snippet does not explicitly verify if the data matches the expected format in terms of shape, data types, or values.
- (test_ensemble_and_cross_validate.py) The code checks the columns of knr_encoding_data_fixture and compares them with the expected columns ['Encoding Dataset', 'Test Set R Squared'].

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': ["[assert list(knr_encoding_data_fixture.columns) == ['Encoding Dataset', 'Test Set R Squared']](#)"]}

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 includes a function 'check_for_duplicates' to verify and remove duplicate records in the loaded data.
- (test_construct_new_library.py)
- (test_model_training.py) The code does not check for duplicate records in the loaded data.
- (test_scrape_aaindex_matrix.py) The code snippet does not include a check for duplicate records in the loaded data.
- (test_ensemble_and_cross_validate.py) No specific check for duplicate records in the provided code.

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 to verify data split proportion as required.
- (test_construct_new_library.py)
- (test_model_training.py) The code splits the data into training and testing sets but does not explicitly verify the split proportion.
- (test_scrape_aaindex_matrix.py) The code snippet does not involve data split proportion verification.
- (test_ensemble_and_cross_validate.py) No data split verification in the provided 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': []}

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 validation as specified in the requirement.
- (test_construct_new_library.py)
- (test_model_training.py) The code trains models but does not verify the output shape alignment.
- (test_scrape_aaindex_matrix.py) The code snippet does not involve model output shape verification.
- (test_ensemble_and_cross_validate.py) No model output shape verification in the provided 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': []}

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) The code does not explicitly verify the evaluation metrics implementation.
- (test_scrape_aaindex_matrix.py) The code snippet does not include verification of evaluation metrics implementation.
- (test_ensemble_and_cross_validate.py) No evaluation metrics verification in the provided 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': []}

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 as required.
- (test_construct_new_library.py)
- (test_model_training.py) The code does not evaluate the model's performance against thresholds.
- (test_scrape_aaindex_matrix.py) The code snippet does not evaluate model performance against thresholds.
- (test_ensemble_and_cross_validate.py) No model performance evaluation against thresholds in the provided 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': []}