

Learning From Differences

Oliver Linger

April 2, 2024

Final-Year Project - B.Sc. in Computer Science

Supervisor: Dr. Derek Bridge

Department of Computer Science

University College Cork

Abstract

This study presents an innovative exploration into the realm of machine learning, specifically focusing on enhancing model interpretability and performance through a novel ensemble of model-based and instance-based learning approaches. Grounded in the principles of Case-Based Reasoning (CBR) and the Learning From Differences methodology, we introduce the LingerRegressor and LingerClassifier models. These models aim to leverage the strengths of both learning approaches to address the limitations present in traditional machine learning algorithms, particularly in the contexts of regression and classification tasks.

Our methodology involves a detailed examination of the algorithms' design and implementation, highlighting the integration of novel variations such as context inclusion, duplication based on distance, and the addition of a distance column to further refine the learning process. Through extensive experiments, we demonstrate the models' effectiveness in achieving higher interpretability and robust performance across diverse datasets. The LingerRegressor and LingerClassifier models not only offer new avenues for academic research but also practical implications for developing more transparent and efficient machine learning systems.

The findings underscore the potential of blending model-based and instance-based learning to enhance machine learning models' adaptability and interpretability. Future work will focus on addressing the identified limitations and exploring the applicability of our approach to broader machine learning challenges. This study contributes to the ongoing dialogue on improving machine learning methodologies and paves the way for further innovation in the field.

Declaration of Originality

In signing this declaration, you are confirming, in writing that the submitted work is entirely your own original work, except where it clearly attributed otherwise, and that it has not been submitted partly or wholly for any other educational award.

1. This body of work is all my own, unless clearly stated otherwise, with the full and proper accreditation;
2. With respect to my work: none of it has been submitted to any educational institution contributing in any way to an educational award.
3. With respect to anyone else's work: all diagrams, text, code, or ideas have been duly attributed to the source in a scholarly manner, whether from books, papers, lecture notes or any other student's work, whether published or unpublished, electronically or print.

Signed

Date: April 2, 2024

Acknowledgements

Contents

1	Introduction	9
1.1	Model-Based Learning	9
1.1.1	Advantages of Model-Based Learning:	9
1.1.2	Disadvantages of Model-Based Learning:	10
1.2	Instance-Based Learning	10
1.2.1	Advantages of Instance-Based Learning:	11
1.2.2	Disadvantages of Instance-Based Learning:	11
1.2.3	Motivation for Blended Learning	11
2	Literature Review	13
2.1	Introduction	13
2.2	KNN + ANN for Regression, Utilizing CBR	14
2.2.1	Learning From Differences Paper	14
2.2.2	Algorithms	15
2.3	Learning Adaptations for Case-based classification	18
2.3.1	An NN-CDH Approach for Classification	19
2.3.2	Variant 0: Non-Network C-CDH	20
2.3.3	Variant 1 - 2: C-NN-CDH	21

2.3.4	Variant 3 - 5: C-NN-CDH	21
2.3.5	Evaluation Summary	25
2.3.6	Conclusions on Learning adaptations for case based classification: A Neural Network approach	26
3	Design and Implementation	27
3.1	LingerRegressor and LingerClassifier Models Roadmap	28
3.2	Learning From Differences	30
3.2.1	Introduction and Motivation	30
3.3	Linger Regression Base Model: LingerRegressor (LR)	31
3.3.1	Objective	31
3.3.2	Architecture	32
3.3.3	Components	32
3.3.4	Parameterization	32
3.3.5	Training and Prediction Design	32
3.3.6	Training and Prediction Implementation	34
3.3.7	Evaluation	38
3.3.8	Optimization	38
3.3.9	Limitations and Future Work	38
3.3.10	Conclusion	38
3.4	Linger Classification Base Model: LingerClassifier (LC)	39
3.4.1	Objective	39
3.4.2	Architecture	40
3.4.3	Components	40
3.4.4	Parameterization	40

3.4.5	Training and Prediction Design	40
3.4.6	Training and Prediction Implementation	41
3.4.7	Evaluation	46
3.4.8	Optimization	46
3.4.9	Limitations and Future Work	46
3.4.10	Conclusion	46
3.5	Variations For <code>LingerRegressor</code> (LR) and <code>LingerClassifier</code> (LC) . . .	47
3.5.1	Variation 1: Addition of Context Design and Implementation (LRV1, LCV1)	47
3.5.2	Variation 2: Duplication Based on Distance Design and Implementation (LRV2, LCV2)	50
3.5.3	Variation 3: Addition of Distance Column Design and Implementation (LRV3, LCV3)	53
3.6	Acronym Guide	56
3.7	Learning From differences Implicitly	56
3.7.1	Introduction and Motivation	56
3.8	Linger Implicit Classification Model: <code>LingerImplicitClassifier</code> (LIC)	57
3.8.1	Objective	57
3.8.2	Architecture	57
3.8.3	Components	58
3.8.4	Parameterization	58
3.8.5	Training and Prediction Design	58
3.8.6	Training and Prediction Implementation	60
3.8.7	Optimization	65
3.8.8	Limitations and Future Work	65

3.8.9	Conclusion	66
3.9	Linger Implicit Regression Model: <code>LingerImplicitRegressor</code>	66
3.9.1	Objective	66
3.9.2	Architecture	67
3.9.3	Components	67
3.9.4	Parameterization	68
3.9.5	Training and Prediction Design	68
3.9.6	Training and Prediction Implementation	70
3.9.7	Regression to Classification Converter Design and Implementation	70
3.9.8	Optimization	73
3.9.9	Limitations and Future Work	73
3.9.10	Conclusion	73
3.10	Learning from Differences For Images	76
3.10.1	Introduction and Motivation	76
3.11	Linger Image Classification Model <code>LingerImageClassifier</code> (LIMC)	76
3.11.1	Objective	76
3.11.2	Architecture	76
3.11.3	Components	76
3.11.4	Parameterization	76
3.11.5	Training and Prediction Design	76
3.11.6	Training and Prediction Implementation	76
3.11.7	Optimization	76
3.11.8	Limitations and Future Work	76
3.12	Linger Image Regressor Model <code>LingerImageRegressor</code> (LIMR)	76

3.12.1	Objective	76
3.12.2	Architecture	76
3.12.3	Components	76
3.12.4	Parameterization	76
3.12.5	Training and Prediction Design	76
3.12.6	Training and Prediction Implementation	76
3.12.7	Optimization	76
3.12.8	Limitations and Future Work	76
4	Experiments	77
4.1	Research Objective	77
4.1.1	Data Collection and Preprocessing	79
4.2	Results and Analysis	80
4.2.1	Experiment 1: Glass Dataset Classification	80
4.2.2	Experiment 6: Raisin Dataset Classification	84
4.2.3	Experiment 7: Car Safety Dataset Classification	84
4.2.4	Experiment 8: Human Wealth Dataset Classification	84
4.3	Discussion	84
5	Conclusions and Future Work	86
6	Results	87

List of Figures

List of Tables

3.1	Regression Models based on the algorithms in [5]	28
3.2	Classifier models based on the algorithms in [5]	29
3.3	LingerImplicitClassifier models based on the algorithms in [16]	29
3.4	LingerImplicitRegressor models based on the algorithms in [16]	30
3.5	Variations for LingerRegressor and LingerClassifier	56
4.1	Tested Hyperparameters for Neural Network and Linger Classifier	81
4.2	Best Parameters for KNN Classifier	81
4.3	Best Parameters for Weighted KNN Classifier	81
4.4	Best Parameters for Neural Net Classifier	82
4.5	Best Parameters for Linger Classifier	82
4.6	Average Training Accuracy Results	83
4.7	Average Test Accuracy Results	83

Chapter 1

Introduction

Machine learning encompasses various paradigms and methodologies aimed at enabling computational systems to learn from data and make predictions or decisions without explicit programming. Among the diverse approaches in machine learning, model-based and instance-based learning represent two fundamental strategies, each with distinct advantages and disadvantages.

1.1 Model-Based Learning

Model-based learning involves the construction of explicit representations of the underlying relationships within the data. These representations, often referred to as models, capture the patterns and structures present in the dataset and can be used to make predictions or infer insights from new, unseen data points. For instance, artificial neural networks (ANNs) exemplify model-based learning by discerning trends and patterns in data.

1.1.1 Advantages of Model-Based Learning:

1. **Generalization:** Well-constructed models have the ability to generalize patterns learned from the training data to unseen instances, thereby making accurate predictions on new data samples. This generalization capability is essential for robust performance across diverse datasets and real-world scenarios. For example, a well-trained regression model can accurately predict housing prices in different regions based on historical data. The book [3] provides insights into the generalization capabilities of model based learning.

2. **Robustness:** Models learn underlying patterns and assumptions within data, making them more robust to noisy outliers. By capturing trends within data, models can effectively mitigate noise, enhancing their reliability in real-world scenarios. The book [3] also shows the ability of models to learn underlying patterns within data, making them far more robust to noisy data.
3. **Efficiency:** Once trained, model-based systems can quickly return predictions, making them suitable for real-time applications where rapid decision-making is essential. The journal [6] offers an insight how the efficiency of model based learning can be implemented to make predictions on big data.

1.1.2 Disadvantages of Model-Based Learning:

1. **Sensitivity to Assumptions:** Models often rely on simplifying assumptions about data distribution and relationships. Deviations from these assumptions can significantly degrade performance, leading to inaccurate predictions or biased outcomes. For example, a model trained to predict house prices based on 2017 data may inaccurately predict house prices for 2023. The book "Machine Learning: A Bayesian and Optimization Perspective" by Sergios Theodoridis and Konstantinos Koutroumbas emphasizes the crucial role assumptions play in model sensitivity, stressing their significance in machine learning applications [14]
2. **Interpretability and Decision-Making:** Certain models, like Neural Networks, lack transparency in decision-making despite their high accuracy. In critical domains like healthcare and finance, interpretability is crucial for validating decisions and ensuring trust in the system. This disadvantage of lack of transparency is depicted well in the journal [12].
3. **Limited Flexibility:** Traditional model-based algorithms struggle to capture complex, non-linear relationships in high-dimensional or unstructured data, limiting their performance in tasks with intricate patterns or diverse feature interactions.

1.2 Instance-Based Learning

Instance-based learning, also known as memory-based learning or lazy learning, eschews explicit model construction in favor of storing and manipulating instances or examples from the training data. Instead of deriving general rules or representations,

instance-based methods make predictions based on the similarity between new instances and those observed in the training set. A classic example of instance-based learning is the k-nearest neighbors (KNN) algorithm. The advantages and disadvantages discussed with instance based learning are seen in the paper [2].

1.2.1 Advantages of Instance-Based Learning:

1. **Flexibility:** Instance-based methods adapt to the characteristics of the training data, making them suitable for tasks with complex, non-parametric relationships or evolving patterns. They can capture nuances and irregularities without imposing rigid assumptions.
2. **Explainability:** The transparency of instance-based learning enables analysts to trace predictions back to specific examples in the training set, facilitating interpretability and trust in the system. The journal [12] highlights this issue for black box model based learning such as neural nets.

1.2.2 Disadvantages of Instance-Based Learning:

The disadvantages discussed with instance based learning are discussed in the paper [2].

1. **Computational Complexity:** Storing and manipulating the entire training dataset can lead to high memory consumption and computational overhead during prediction, limiting scalability.
2. **Susceptibility to Noise and Redundancy:** Instance-based approaches are sensitive to noisy or irrelevant features, which can lead to suboptimal performance or overfitting. Redundant instances or outliers may distort similarity measures, compromising model robustness.

1.2.3 Motivation for Blended Learning

The advantages and disadvantages of both model-based and instance-based learning highlight the need for a versatile and robust approach to machine learning. While model-based learning excels in generalization and robustness, it lacks transparency and struggles with complex relationships. Instance-based learning offers flexibility and explainability but suffers from computational complexity and susceptibility to noise.

The motivation for blended learning arises from leveraging the strengths of both approaches while mitigating their weaknesses. Blended methods aim to capitalize on the advantages of model-based and instance-based learning, circumventing their limitations. By integrating the interpretability of instance-based learning with the accuracy of model-based systems, blended learning offers a promising solution for critical domains like healthcare and finance. The journal [12] provides further evidence that it is important to move towards interpretability for our models going forward, so that critical systems are interpretable and repairable by humans going forward.

In recent years, neural networks have gained significant attention in regression tasks [10]. Inspired by case-based reasoning principles [11], researchers are exploring predictive ensemble models based on these principles [15]. Deviating from conventional approaches, some researchers advocate for training neural networks on differences between sets of features to enhance efficiency and transparency [5].

By adopting a blended approach, machine learning systems can achieve greater interpretability, reliability, and performance across diverse applications.

Chapter 2

Literature Review

2.1 Introduction

The exploration of alternative methodologies for enhancing the performance of neural networks in regression tasks has garnered significant attention within the machine learning community in recent years. One innovative approach, inspired by the principles of case-based reasoning (CBR), involves training neural networks to predict differences between problems based on disparities between features rather than the features themselves. This departure from the conventional supervised learning paradigm aims to leverage the inherent advantages of learning from differences, potentially leading to more interpretable and efficient models.

The study under consideration [5] conducts a systematic factorial study to investigate the efficacy of this approach across various datasets and experimental conditions. The findings suggest that neural networks trained on differences demonstrate comparable or even superior performance to those trained using traditional methods, while requiring significantly fewer epochs for convergence. In this section of the literature review, we consider the utilization of case-based reasoning for regression tasks, with a specific focus on the implementation of KNN (K-Nearest Neighbors) and ANN (Artificial Neural Network) frameworks.

As we delve further into the literature review, we extend our exploration of case-based reasoning beyond regression tasks to include classification challenges. The study by Ye et al. [16], titled "Learning Adaptations for Case-based Classification," offers insights into the application of CBR in classification tasks, shedding light on recent advancements and methodologies. Through a detailed analysis of various adaptation

strategies and neural network-based approaches, we aim to uncover the effectiveness of CBR in solving classification problems, thereby providing a comprehensive understanding of its utility across different machine learning tasks.

2.2 KNN + ANN for Regression, Utilizing CBR

The exploration of alternative methodologies for enhancing the performance of neural networks in regression tasks has attracted considerable attention within the machine learning community in recent years. One innovative approach, inspired by the principles of case-based reasoning (CBR), involves training neural networks to predict differences between problems based on disparities between features rather than the features themselves. This departure from the conventional supervised learning paradigm aims to leverage the inherent advantages of learning from differences, potentially leading to more interpretable and efficient models. The study under consideration [5] conducts a systematic factorial study to investigate the efficacy of this approach across various datasets and experimental conditions. The findings suggest that neural networks trained on differences demonstrate comparable or even superior performance to those trained using traditional methods, while requiring significantly fewer epochs for convergence. In this section of the literature review, we consider the utilization of case-based reasoning for regression tasks, with a specific focus on the implementation of KNN (K-Nearest Neighbors) and ANN (Artificial Neural Network) frameworks.

2.2.1 Learning From Differences Paper

The paper by [5] introduces the innovative concept of training a neural network based on the disparities between a case and its closest neighbors. Titled "A Factorial Study of Neural Network Learning from Differences for Regression" [5], the study aims to present novel learning methodologies geared towards enhancing performance and fostering a more interpretable learning process. The research evaluates three model variations: a benchmark neural network, a learning from differences model, and an augmented learning from differences model incorporating the original context. The context here refers to the base case, from which differences with its nearest neighbors are calculated.

According to [5], the study's key findings underscore a significant increase in accuracy when incorporating contextual information in the learning process. Moreover, comparable or superior results were achieved with fewer training epochs compared to the basic neural network. This enhancement is attributed partly to the broader training

dataset that encompasses neighbors greater than 1. Notably, the paper advocates for a hybrid approach, combining learning from differences with direct feature-based learning, which yields superior results.

The related work for the paper [5] has addressed the use of neural networks in the case-based reasoning process. These papers have focused on exploiting the base case using the idea that adaptation of information is acquirable from the differences between pairs [7]. Since then, various different approaches have been used. More recently, in relation to using neural networks to predict the differences between problems, a few interesting points have been investigated. These points show the ability of a neural network to correct the solution of the most similar retrieved case. These preliminary works expressed in the paper showed that the use of differences plus context yielded superior results [4]. The previous studies examined in the paper [5] did not examine the impact of different parameters during the CBR (case-based reasoning) experiments. The paper [5] made an effort to explore the effect of epochs in the learning process.

The graphical representations in the paper highlight a distinct trend: rapid attainment of high accuracy levels with the learning from differences method, particularly when contextual information is included. Conversely, the basic neural network requires a substantially larger number of epochs to achieve comparable accuracy, if not worse. This disparity suggests the efficiency of the learning from differences approach.

The paper suggests a marginal performance enhancement and a substantial reduction in training epochs through the adoption of the differences method and inclusion of contextual information. However, it is crucial to acknowledge the time overhead associated with retrieving and training using similar cases.

2.2.2 Algorithms

The training algorithm for learning from differences is described in Algorithm 1, while the prediction algorithm is outlined in Algorithm 2. Additionally, a variant of the training algorithm with context included is presented in Algorithm 3, and the corresponding prediction algorithm is provided in Algorithm 4.

Context which is utilized in 3, and 4. Context is the case used to retrieve (Nearest Neighbors). In the training algorithm it is X_i^{train} , and in prediction it is X_j^{test} . These are included in the training, and prediction in Variant 2.

The paper [5] has an associated library where they have implemented a regression-based machine learning model.

Algorithm 1 Training Algorithm for Learning from Differences

Input: Dataset D , neural network model, number of neighbors n , number of epochs E

- 1: Split D into training set D^{train} and test set D^{test} with a 80%-20% split
 - 2: Initialize ΔD^{train} as an empty list
 - 3: **for** each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 - 4: Retrieve n similar cases from D^{train} using nearest neighbors
 - 5: **for** each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**
 - 6: Compute the differences: $\Delta D_{ij}^{train} = (X_i^{train} - X_j^{train}, y_i^{train} - y_j^{train})$
 - 7: **end for**
 - 8: **end for**
 - 9: Train the MLPRegressor neural network model using ΔD^{train} over a few epochs for later predictions.
 - 10: **Return** Trained neural network model $TrainedNN()$
-

Algorithm 2 Prediction Algorithm for Learning from Differences

Input: Test dataset D^{test} , trained neural network model $TrainedNN()$, number of neighbors n

- for** each case $X_j^{test} \in D^{test}$ **do**
- Retrieve n similar cases from D^{train} using nearest neighbors
- Initialize empty list $y^{pred}predictions$
- for** each similar case $(X_i^{train}, y_i^{train}) \in D^{train}$ **do**
- Compute the differences: $\Delta X_{ij}^{test} = X_j^{test} - X_i^{train}$
- Use the trained neural network to predict: $TrainedNN(\Delta X_{ij}^{test}) = \Delta X_{ij}^{pred}$
- Adapt neighbor's y values: $y^{pred}predictions_j = X_i^{train} + \Delta X_{ij}^{pred}$
- end for**
- Average the adapted y values: $y_j^{pred} = \frac{1}{n} \sum_{i=1}^n y^{pred}predictions$
- Compare y_j^{pred} to y_j^{test} to evaluate accuracy
- end for**
-

Algorithm 3 Training Algorithm for Learning from Differences with context

Input: Dataset D , neural network model, number of neighbors n , number of epochs E

- 1: Split the dataset D into training set D^{train} and test set D^{test} using an 80%-20% split.
 - 2: Initialize ΔD^{train} as an empty list.
 - 3: **for** each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 - 4: Retrieve n similar cases from D^{train} using nearest neighbors.
 - 5: **for** each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**
 - 6: Compute the differences, with concatenated context : $\Delta D_{ij}^{train} = ((X_i^{train} - X_j^{train}) : X_i^{train}, y_i^{train} - y_j^{train})$.
 - 7: Append ΔD_{ij}^{train} to ΔD^{train} .
 - 8: **end for**
 - 9: **end for**
 - 10: Train the MLPRegressor neural network model using ΔD^{train} .
 - 11: **return** Trained neural network model $TrainedNN()$.
-

Algorithm 4 Prediction Algorithm for Learning from Differences with context

Input: Test dataset D^{test} , trained neural network model $TrainedNN()$, number of neighbors n

```
for each case  $X_j^{test} \in D^{test}$  do
  Retrieve  $n$  similar cases from  $D^{train}$  using nearest neighbors
  Initialize  $y_j^{pred}$  as an empty list
  for each similar case  $(X_i^{train}, y_i^{train}) \in D^{train}$  do ‘
    Initialize empty list  $y^{pred}predictions$ 
    Compute the differences:  $\Delta X_{ij}^{test} = ((X_j^{test} - X_i^{train}) : X_j^{test})$ 
    Use the trained neural network to predict:  $\Delta X_{ij}^{pred} = TrainedNN(\Delta X_{ij}^{test})$ 
    Adapt neighbor’s  $y$  values:  $y^{pred}predictions_j = y_i^{train} + \Delta X_{ij}^{pred}$ 
  end for
  Average the adapted  $y$  values:  $y_j^{pred} = \frac{1}{n} \sum_{i=1}^n y_i^{pred}$ 
  Compare  $y_j^{pred}$  to  $y_j^{test}$  to evaluate accuracy
end for
```

The new model proposed in the paper is tested using the factorial study methodology. This process involved finding a network structure through experimental means for each dataset. It entailed training and testing using a wide range of hyperparameters to determine the optimal settings.

Despite its contributions, the paper falls short by confining its attention to regression tasks and not considering classification tasks and exploring additional variations of the learning from differences model beyond context inclusion. Incorporating original values for the base case in the generated differences dataset provides a valuable framework for understanding the observed disparities. While the model with contextual inclusion demonstrates slight improvements over the base model in learning from differences, a more comprehensive examination of the methodology is warranted for further refinement and enhancement.

The paper tested three models using the same factorial methodology: a base version, which is a basic neural network trained empirically; a differences model employing the (CBH) Case-based Heuristic for predictions, (CBH) guides how cases are retrieved with respect to the base case, and how the retrieved cases are adapted to produce a prediction, (CBH) relies on the adaptation of previous problems to solve the current issue; and a differences model with its base context included. The results yielded interesting insights, with similar performance for all three models and a slight improvement in the differences + context model. Notably, the differences and differences + context models required significantly fewer epochs to achieve accuracy compared to the basic neural network, which achieved similar accuracy after more epochs.

Another consideration in the paper [5] was the number of neighbors for both training

and prediction. The study revealed no consistent trend across datasets, suggesting dataset-specific requirements.

In conclusion, the learning from differences method achieved equal or higher accuracy compared to the base neural network model.

2.3 Learning Adaptations for Case-based classification

The paper by Ye et al. [16], titled "Learning Adaptations for Case-based Classification," investigates the application of Case-Based Reasoning (CBR) for classification tasks. It explores recent advancements in CBR, particularly focusing on replacing traditional rule-based learning with Case-Based Heuristic (CBH) network models for adaptation. The study introduces a novel model comprising three variations: segmentation of adaptation knowledge based on the classes of the source cases, training a single neural network on differences between problem solutions, and adapting from an ensemble of source cases followed by a majority vote.

The "NN-CDH" method, as proposed in the paper, represents a novel classification approach that operates on pairs of cases, where one serves as the source and the other as the target. These pairs form the basis of adaptation knowledge, which is learned through neural network models. The methodology adopted in [16] emphasizes learning adaptation knowledge through case pairs processed by neural networks.

The study introduces the first neural network-based approach to classification, termed "C-NN-CDH" or Classification with Neural Network-based Case Difference Heuristic. This approach employs neural networks to learn adaptation knowledge from pairs of cases. According to the findings, this methodology outperforms traditional statistical models, as evidenced in [8].

The first approach involves segmenting the dataset by class and training a neural network for all segmented differences set between base cases. This method offers faster training but slightly less accuracy, according to the paper. The use of these CBR methods provides at least two benefits: inertia-free lazy learning and the ability to assess different cases when running classification. Inertial free means that the model does not make assumptions based on the underlying structure of the data, and lazy learning pertains to an algorithm that will postpone generalization until it is certain a new instance needs to be classified. It will only classify as the data comes; it won't try and learn or adapt preemptively.

In the process of CBR, adaptation is often treated as the most difficult. Various methods have been developed by academics over the years to tackle the issue. For example, Leake et al. [9] extrapolate a method in which the case base adaptation cases are populated from previously successful adaptations. Another approach is to have a stored base case and query and retrieve the case most similar to it, and adapting the retrieved case to the base case [4].

In the paper by Ye et al. [16], several design questions had to be addressed. How to approach case differences and how to select case pairs for training. The paper in question [16] considers standard pair selection methods as well as a new training per selection approach based on class-class classification.

The paper [16] also addresses the application of (CDH) method for classification. It discusses the value distance metric, which is a probabilistic method to measure the similarity that allows the comparison of nominal values in a single-dimensional space.

2.3.1 An NN-CDH Approach for Classification

To achieve adaptation, a neural network is integrated into the process. The source and target cases become case pairs, and their adaptation is learned through a neural network. The neural network computes the difference between the source and target pairs. This difference calculation is a crucial step in the Case-Based Reasoning (CBR) system. Once the difference is computed, the predicted result obtained from the neural network is applied to the source solution. This adapted solution is then considered as the final result of the classification process.

The NN-CDH approach represents a fusion of neural network techniques with the principles of Case-Based Reasoning, offering a dynamic and adaptable method for classification tasks. When calculating the difference between the problem and the solution values, it is important to implement a difference function. This presents an interesting challenge for nominal values. The method presented in [16] uses an implicit calculation through machine learning techniques. This replaces traditional (CDH) case difference heuristic approaches for difference calculations. Through the use of a neural network, the paper's method hopes to include the base context in the differences' calculation. This method for generating differences has been aptly named ("C-CDH") case difference heuristic approach for classification.

When calculating the difference, an issue arrives in that we cannot explicitly calculate differences through subtraction, which was the case in the previous paper by Learning

From Differences [5]. An implicit way of learning from differences is required. In the paper by Ye et al. [16], this is called "C-CDH".

This presents an interesting challenge for nominal values. The method presented in [16] uses an implicit calculation through machine learning techniques. When discussing the Neural Network-based Case-Driven Hierarchical (NN-CDH) Approach described in [16], notation will be homogenized with the algorithms from the previous paper.

For clarification when going through the algorithms, Retrieved cases = Source Cases, Target case = The case which we are attempting to predict its y label values.

Multiple variants were created with the paper by Ye et al. [16].

2.3.2 Variant 0: Non-Network C-CDH

Variant 0, as described in [16], serves as a foundational test bed implementation for classification tasks. The algorithms presented here outline the process of building and utilizing a dictionary of case pairs, where each pair consists of a source case and its corresponding target case. Through the use of nearest neighbors, the algorithm identifies similar cases within the training dataset and organizes them based on their source solution. Subsequently, during prediction, the algorithm retrieves the nearest similar cases for each test case and utilizes the corresponding target solutions for classification. This approach embodies the essence of case-based reasoning, where solutions are adapted from similar past cases to classify new instances effectively. The following algorithms detail the steps involved in training and predicting using Variant 0's case-based heuristic approach.

Algorithm 5 Variant 0, Classification Using Case Based Heuristic, Training

Input: Dataset D , number of neighbors 1

Split D into training set D^{train} and test set D^{test} with an 80%-20% split

Create *CasePairs()* dictionary.

for each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**

 Retrieve 1 similar cases from D^{train} using nearest neighbors.

for each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**

CasePairs(y_j^{train}).append($[X_j^{train} : X_i^{train}, y_i^{train}]$)

end for

end for

Algorithm 6 Variant 0, Classification Using Case Based Heuristic, Prediction

Input: Test dataset D^{test} , $CasePairs()$

```
1: Initialize Empty list  $y^{pred}$ 
2: for each case  $X_j^{test} \in D^{test}$  do
3:   Retrieve 1 similar cases from  $D^{train}$  using nearest neighbors
4:   Initialize  $y_j^{pred}$  as an empty list
5:   for each similar case  $(X_i^{train}, y_i^{train}) \in D^{train}$  do
6:     Retrieve  $r = 1NearestNeighbor(CasePairs(y_i^{train}), [X_i^{train} : X_j^{test}])$ 
7:      $y_j^{pred} = r$ 
8:   end for
9: end for
10: return  $y^{pred}$ 
```

2.3.3 Variant 1 - 2: C-NN-CDH

Variants 1 - 2, known as C-NN-CDH, represent a pivotal advancement in classification methodologies by integrating neural networks into the case-based heuristic framework. In this variant, the traditional case-based heuristic approach is augmented with the power of neural networks to adapt solutions from similar cases. The algorithms presented here delineate the training and prediction processes, where a neural network model learns adaptation knowledge from case pairs derived from the training dataset. Unlike Variant 0, which relies solely on case similarity for adaptation, C-NN-CDH leverages the expressive capacity of neural networks to capture intricate relationships between cases and their solutions. Through the fusion of case-based reasoning principles with neural network techniques, Variant 1 - 2 offers a dynamic and adaptable approach to classification tasks, capable of learning complex adaptation patterns and improving classification accuracy.

2.3.4 Variant 3 - 5: C-NN-CDH

Variants 3 - 5 of the classification neural network using case-based heuristics introduce a novel approach to classification by grouping case pairs based on their source solutions. In this framework, multiple specialized adaptation neural networks are employed, each trained to adapt cases of specific solutions to all solutions. Inspired by the ensemble of adaptations for classification approach, these variants share a common grouping and training method but differ in their testing strategies. The algorithms delineate the training process, where case pairs are organized and used to train specialized neural networks for adaptation. The variation arises in the prediction algorithm, where different strategies are employed to classify test instances based on the learned adaptation knowledge. The following section details the nuances of each variant's

Algorithm 7 Variant 1 - 2: Classification Using Neural Net Case Based Heuristic (Training)

Input: Dataset D , neural network model, number of neighbors n , number of epochs E
 Split D into training set D^{train} and test set D^{test} with an 80%-20% split
 Create *CasePairs* list.
for each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 Retrieve 1 similar case from D^{train} using nearest neighbors.
 for each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**
 if Variant 1 **then**
 $CasePairs.append([X_j^{train} : X_i^{train}, y_i^{train}])$
 else
 if Variant 2 **then**
 $CasePairs.append([X_j^{train} : X_i^{train} : y_j^{train}, y_i^{train}])$
 end if
 end if
end for
end for
 Train the neural network model using *CasePairs*()
return Trained neural network model *TrainedNN*()

Algorithm 8 Variant 1 - 2: Classification Using Neural Net Case Based Heuristic (Prediction)

Input: Test dataset D^{test} , trained neural network model *TrainedNN*()
 1: Initialize Empty list y^{pred}
 2: **for** each case $X_j^{test} \in D^{test}$ **do**
 3: Retrieve 1 similar case from D^{train} using nearest neighbors
 4: **for** each similar case $(X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 5: Use the trained neural network to predict: $y_j^{pred} = TrainedNN([X_i^{train} : X_j^{test} : y_i^{train}])$
 6: **end for**
 7: **end for**
 8: **return** y^{pred}

prediction methodology and discusses their implications for classification accuracy and adaptability.

Algorithm 9 Variant 3 - 5: Classification Neural Network Using Case Based Heuristic (Training)

Input: Dataset D , number of neighbors 1

Split D into training set D^{train} and test set D^{test} with an 80%-20% split

Create $CasePairs()$ dictionary.

Create $AdaptNN()$, a dictionary to store each Neural Network

List of unique Classes $UniqueClasses$

for each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**

Retrieve 1 similar case from D^{train} using nearest neighbors.

for each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**

$CasePairs(y_j^{train}).append([X_j^{train} : X_i^{train}, y_i^{train}])$

end for

end for

for each unique $Class$ in $UniqueClasses$ **do**

Create a new Classification Neural Network, trained on Case Pairs

$AdaptNN(Class).fit(CasePairs(Class))$

end for

return Trained neural network models dictionary $AdaptNN()$.

In variant (5), the prediction method utilizes a Class-to-Class approach for retrieving neighbors. This method retrieves the most similar neighbor for each class from the training dataset. Once the nearest neighbors for each class are identified, the corresponding specialized neural networks associated with those classes are employed to generate predictions for the target instance. Subsequently, a majority vote is conducted among the predictions from the specialized neural networks to determine the final classification outcome.

It is worth considering re-implementing the algorithm to allow for the retrieval of multiple neighbors (N) from each class during the prediction stage. This adjustment could potentially lead to improved results, particularly when dealing with large and varied datasets.

Variant (5) stands out as the most successful approach in the study, primarily due to its Class-to-Class retrieval of neighbors during the prediction stage. By employing multiple specialized neural networks, the variant reduces training time while effectively leveraging implicit learning through differences for classification tasks. This method demonstrates its efficacy in adapting to diverse datasets and achieving robust classification performance.

Algorithm 10 Variant 3 - 5: Classification Using Neural Net Case-Based Heuristic (Prediction)

Input: D^{test} , $AdaptNN()$, $UniqueClasses$, and $CasePairs()$

Initialize an empty list y^{pred}

for each case $X_j^{test} \in D^{test}$ **do**

if Variant 3 **then**

 Retrieve 1 similar case from D^{train} using nearest neighbors

for each similar case $(X_i^{train}, y_i^{train}) \in D^{train}$ **do**

$y_j^{pred} = AdaptNN(y_i^{train}).predict([X_i^{train} : X_j^{test}])$

end for

else

if Variant 4 **then**

 Create an empty list $predictions$

 Retrieve N similar cases from D^{train} using nearest neighbors

for each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**

$predictions.append(Adapt(y_i^{train}).predict([X_i^{train} : X_j^{test}]))$

end for

 Perform a majority vote on $predictions$

$y_j^{pred} = MajorityVote(predictions)$

else

if Variant 5 **then**

 Create an empty list $preds$

for each unique class $Class$ in $UniqueClasses$ **do**

 Retrieve 1 similar case from $CasePairs(Class)$

for each similar case $(X_i^{CasePair}, y_i^{CasePair}) \in CasePairs(Class)$ **do**

for each $(X_z^{train}, X_q^{train}) \in X_i^{CasePair}$ **do**

$preds.append(AdaptNN(Class).predict(X_z^{train} : X_j^{test}))$

end for

end for

 Perform a majority vote on $preds$

$y_j^{pred} = MajorityVote(preds)$

end for

end if

end if

end if

end for

return y^{pred}

2.3.5 Evaluation Summary

Within the paper [16] the experiments conducted on two different datasets provided answers to several key questions.

The effectiveness of neural network learning adaptation knowledge was demonstrated through consistent performance of one or more C-NN-CDHs, which often outperformed the neural network classifier. Despite the possibility of neural networks learning to discard the source problem and rely solely on the target problem, the non-zero weights associated with the source problem and significant performance differences between C-NN-CDHs and the neural network indicate that C-NN-CDHs effectively learn adaptation knowledge. This suggests that if a neural network can handle the classification task directly, it can also learn adaptation knowledge or the relation between pairs of cases in the task domain.

Surprisingly, variant (1) performed almost identically to variant (2), which also considers the source solution in adaptation. This suggests that the source solution, heavily coupled with the source problem, may not provide additional useful information for adaptation.

In terms of segmenting case pairs by source case solution, variants (1, 2) generally outperformed variants (3, 4) in accuracy on most datasets. This could be because a single adaptation neural network in (1, 2) is well-trained with all pairs of cases, while a specialized adaptation neural network in (3, 4) is trained with segmented examples. However, variants (3, 4) showed faster convergence during training due to training on segmented examples.

The impact of the ensemble of adaptations was investigated, with EAC-NN-CDH (variant (4)) performing similarly to its counterpart variant (3) without ensemble. This suggests that the generalization power of C-NN-CDH produces stable predictions that an ensemble version does not significantly alter.

The usefulness of the class-to-class approach for adaptation was demonstrated by C2C-NN-CDH (variant (5)), which performed differently from and often better than the other C-NN-CDHs. C2C-NN-CDH reaches its prediction by collecting evidence from diverse source cases from all classes, which can provide more global support, especially when there are multiple classes. Moreover, the C2C approach offers the possibility of explanation with contrastive evidence.

2.3.6 Conclusions on Learning adaptations for case based classification: A Neural Network approach

The conclusion of the paper [16] being that (variant (5)) is worth investigating using its C2C-NN-CDH approach to classification. As it yielded the most promising results when compared to all the other variants tested in the paper. The grouping of the dataset based on class outcome and training a specialized neural network, has its benefits in terms of speedup. However, the retrieval of N cases across all classes during testing, has shown to yield better results, in a way that allows an interpretation of the results.

Chapter 3

Design and Implementation

Objectives and Contributions

The design chapter of this paper serves as a comprehensive guide, presenting various algorithms and their implementations in the context of our novel framework. Our primary objectives are to elucidate the intricacies of these algorithms, introduce innovative variations, and showcase the integration of our framework within the broader machine learning landscape.

The tables provided below offer a structured roadmap through our design chapter, delineating where and how different algorithms are described. Notably, the algorithms discussed originate from seminal works such as [5] and [16]. However, in addition to presenting established methodologies, we introduce novel variations that represent fresh approaches to classification and regression tasks.

Key innovations include Variant 2, which introduces duplication based on distances, and Variant 3, incorporating a column indicating each neighbor's distance to the base case. Notably, we employ a new type of classifier, distinct from those mentioned in [5]. Variants 1, 2, and 3 exemplify our endeavor to explore uncharted territory within the learning from differences methodology, offering novel perspectives on classification.

Furthermore, our introduction of the **LingerImplicitRegressor** presents a groundbreaking approach to regression tasks. Leveraging methodologies outlined in [16], this novel regressor extends the learning from differences framework, offering a fresh perspective on regression problems.

By presenting a cohesive framework that integrates established algorithms, innovative

variations, and our novel contributions, these tables provide a clear overview of the design landscape. Our aim is to facilitate understanding and navigation through the complexities of machine learning methodologies, empowering researchers and practitioners to leverage our framework effectively in their endeavors.

3.1 LingerRegressor and LingerClassifier Models Roadmap

The following models are based on the algorithms outlined in [5]. Variations such as Variant 2 and Variant 3 introduce innovative adaptations to the base models, providing enhanced functionality and performance.

LingerRegressor		
Full Name	Acronym	Description
LingerRegressor Section: 3.3	LR	Base Linger regressor, no variations
LingerRegressor Variant 1 Section: 3.5.1	LRV1	Linger regressor with the addition of context, where context is the original base case.
LingerRegressor Variant 2 Section: 3.5.2	LRV2	Linger regressor integrated with duplication based on distance, where differences are duplicated based on the neighbors' proximity to the base case.
LingerRegressor Variant 3 Section: 3.5.3	LRV3	Linger regressor integrated with a distance column indicating the distance of each neighbor from the base case.

Table 3.1: Regression Models based on the algorithms in [5]

LingerClassifier			
Full Name		Acronym	Description
LingerClassifier	Section: 3.4	LC	Base Linger classifier, no variations
LingerClassifier	Variant 1 Section: 3.5.1	LCV1	Linger classifier with the addition of context, where context is the original base case.
LingerClassifier	Variant 2 Section: 3.5.2	LCV2	Linger classifier integrated with duplication based on distance, where differences are duplicated based on the neighbors' proximity to the base case.
LingerClassifier	Variant 3 Section: 3.5.3	LCV3	Linger classifier integrated with a distance column indicating the distance of each neighbor from the base case.

Table 3.2: Classifier models based on the algorithms in [5]

LingerImplicitClassifier and LingerImplicitRegressor Models

The classifier models presented here are based on algorithms from [16], with Variations 1 and 2 also originating from the same source.

The regressor models in this section extend the learning from differences framework outlined in [16] to regression tasks. This innovative approach maintains the core principles of the methodology while introducing adaptations tailored to address the unique challenges of regression problems.

LingerImplicitClassifier		
Full Name	Acronym	Description
LingerImplicitClassifier Section: 3.8	LIC	Base Linger Implicit Classifier selects the N nearest neighbors from each class
LingerImplicitClassifier Variant 1 Section: 3.8.5	LICV1	Selects N random items from each class
LingerImplicitClassifier Variant 2 Section: 3.8.5	LICV2	Selects the single nearest neighbor from each class

Table 3.3: LingerImplicitClassifier models based on the algorithms in [16]

LingerImplicitRegressor		
Full Name	Acronym	Description
LingerImplicitRegressor Section: 3.3	LIR	Base Linger Implicit Regressor selects the N nearest neighbors from each class during prediction
LingerImplicitRegressor Variant 1 Section: 3.9.5	LIRV1	Selects N random items from each class
LingerImplicitRegressor Variant 2 Section: 3.9.5	LIRV2	Selects the single nearest neighbor from each class
Regression to Classification Converter		
Full Name	Acronym	Description
Regression to Classification Converter Section: 3.9.7	RTCC	Converts regression problems into classification problems

Table 3.4: LingerImplicitRegressor models based on the algorithms in [16]

3.2 Learning From Differences

3.2.1 Introduction and Motivation

Our design process is anchored in the seminal work of [5], which laid the groundwork for the learning from differences method. Building upon this foundation, our primary objective is to enhance and extend these algorithms, creating a more dynamic and adaptable framework.

Inspired by Scikit-learn’s modular design philosophy, we aim to develop models that not only match but surpass its rigorous hyperparameter testing capabilities. Our goal is to introduce multiple new hyperparameters accessible through familiar pipelines, thereby enhancing flexibility and functionality across diverse machine learning tasks and datasets.

As we progress from implementing base models to introducing variations, our focus shifts to optimizing the performance of both the **LingerRegressor** and **LingerClassifier**. These variations are meticulously crafted to improve training efficiency and predictive accuracy, ultimately enhancing overall model performance.

Named **LingerRegressor** and **LingerClassifier**, our custom models are designed to leverage the training from differences approach, distinguishing them within the machine learning landscape.

Central to our efforts is ensuring the seamless testability of introduced variations, akin to Scikit-learn’s comprehensive hyperparameter testing framework. This rigorous analysis empowers stakeholders to make informed decisions about the efficacy and applicability of our models to specific tasks.

However, we encountered the critical challenge of compatibility during model development. Our codebase is structured as a downloadable package, seamlessly integrating with Scikit-learn’s ecosystem to ensure interoperability across various datasets and environments. This compatibility enhances accessibility and usability for practitioners in diverse machine learning endeavors.

3.3 Linger Regression Base Model: LingerRegressor (LR)

3.3.1 Objective

The Linger Regression Base Model integrates principles of Case-Based Reasoning (CBR) into regression analysis to enhance interpretability and create a performative model. The objectives include:

- Developing a dynamic regression model using differences between a case and its nearest neighbors, adapting retrieved solutions to the target solution.
- Ensuring compatibility with Scikit-learn for accessibility, ease of implementation, and extensive hyperparameter testing.
- Enabling effective interpretation of predictions for users.
- Implementing the learning from differences algorithm with enhancements for improved performance.
- Creating an easily extendable code base to allow for variations.

Its intended use case is for regression problems that could benefit from a more transparent learning process.

3.3.2 Architecture

The `LingerRegressor` combines K-nearest neighbors (KNN) and `MLPRegressor` from Scikit-learn for flexibility and accuracy. A K-nearest neighbors (KNN) is used in conjunction with a `MLPRegressor` in the fit phase of the machine. During the predict phase, a K-nearest neighbors (KNN) is used once again with the now trained `MLPRegressor` to get the differences' prediction or K neighbors.

3.3.3 Components

- KNN module: Computes nearest neighbors and extracts differences during both training and prediction phases.
- `MLPRegressor`: Offers flexibility in activation functions, solvers, and hyperparameters for regression.
- Supplementary Features: Weighted KNN, Context Addition, Additional Results Column, and Duplication Based on Distance. Each supplementary feature can be activated through a hyperparameter.

3.3.4 Parameterization

Extensive parameterization allows fine-tuning of KNN, `MLPRegressor`, and additional features. All the same hyperparameters are available as in a conventional (KNN) and neural network, with additional hyperparameters for `LingerRegressor`.

3.3.5 Training and Prediction Design

Training Design

Given dataset D , split it into D^{train} and D^{test} with an 80%-20% split through random selection. For each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$, retrieve n similar cases from D^{train} , $(X_j^{train}, y_j^{train}) \in D^{train}$ using Scikit-learn's neighbors module. Compute the differences between C_i and its neighbors and append them to ΔD^{train} as $(X_i^{train} - X_j^{train}, y_i^{train} - y_j^{train})$.

Algorithm 11 Training Algorithm for Learning from Differences

- 1: Split D into training set D^{train} and test set D^{test} with an 80%-20% split
 - 2: Initialize ΔD^{train} as an empty list
 - 3: **for** each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 - 4: Retrieve n similar cases from D^{train} using nearest neighbors
 - 5: **for** each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**
 - 6: Compute the differences: $\Delta D_{ij}^{train} = (X_i^{train} - X_j^{train}, y_i^{train} - y_j^{train})$
 - 7: Append ΔD_{ij}^{train} to ΔD^{train}
 - 8: **end for**
 - 9: **end for**
 - 10: Train the MLPRegressor neural network model using ΔD^{train} over a few epochs for later predictions.
 - 11: **Return** Trained neural network model $TrainedNN()$
-

Prediction Design

During prediction, for each case X_j^{test} , retrieve n similar cases from D^{train} . Compute the differences $\Delta X_{ij}^{test} = X_j^{test} - X_i^{train}$ for each neighbor. Use the trained neural network to predict ΔX_{ij}^{test} , resulting in ΔX_{ij}^{pred} . For each set of retrieved neighbors from D^{train} , calculate $y_i^{train} + \Delta X_{ij}^{pred}$ and add each n neighbors adapted y values. Once summed, divide by n to obtain y_j^{pred} . Compare y_j^{pred} to y_j^{test} to evaluate accuracy.

Algorithm 12 Prediction Algorithm for Learning from Differences

Input: Test dataset D^{test} , trained neural network model $TrainedNN()$, number of neighbors n

- Initialize empty list y_j^{pred}
- for** each case $X_j^{test} \in D^{test}$ **do**
- Initialize empty list y_j^{pred}
- Retrieve n similar cases from D^{train} using nearest neighbors
- Initialize empty list $y_j^{pred_predictions}$
- for** each similar case $(X_i^{train}, y_i^{train}) \in D^{train}$ **do**
- Compute the differences: $\Delta X_{ji}^{test} = X_j^{test} - X_i^{train}$
- Use the trained neural network to predict: $TrainedNN(\Delta X_{ji}^{test}) = \Delta X_{ji}^{pred}$
- Adapt neighbor's y values: $y_{ji}^{pred} = y_i^{train} + \Delta X_{ji}^{pred}$
- end for**
- Average the adapted y values: $y_j^{pred} = \frac{1}{n} \sum_{i=1}^n y_{ji}^{pred}$
- end for**
- Compare y_j^{pred} to y_j^{test} to evaluate accuracy
-

3.3.6 Training and Prediction Implementation

Training Implementation

During training, the algorithm fits the base Linger Regressor to the input dataset X and the corresponding label vector y . It computes the nearest neighbors for each sample in X and constructs the input features and target labels for the regressor by calculating the differences between each sample and its neighbors. These differences are then used to train the regressor, and the trained model along with the training data are stored for later use. It makes use of scikit-learns [13] `NearestNeighbors`, `numpy` and `MLPRegressor`.

Algorithm 13 Training Implementation Algorithm for Base Linger Regressor

Input: Dataset, X , label vector y , number of neighbors $n_neighbors_1$

```
1: function FIT( $X, y$ )
2:   Increment  $n\_neighbours\_1$  ▷ Fit scikit-learns NearestNeighbors function to  $X$ 
3:    $neighbors = NearestNeighbors(n\_neighbors = n\_neighbors\_1).fit(X)$ 
4:   Compute nearest neighbors for each sample in  $X$ 
5:   Compute the indexes and distances, of nearest neighbors in  $X$ 
6:    $distances, indices = neighbors.kneighbors(X)$ 
7:    $differences\_X = []$ 
8:    $differences\_y = []$ 
9:   if  $variant == 1$  (LRV1) then
10:    Call  $addition\_of\_context\_func\_fit(X, y, indices, differences\_X, differences\_y)$ 
11:     $differences\_X, differences\_y = addition\_of\_context\_func\_fit(X, y, indices, differences\_X, differences\_y)$ 
12:  else if  $variant == 2$  (LRV2) then
13:    Call  $duplicate\_difference\_based\_on\_distance(differences\_X, differences\_y, distances)$ 
14:     $differences\_X, differences\_y = duplicate\_difference\_based\_on\_distance(differences\_X, differences\_y, distances)$ 
15:  else if  $variant == 3$  (LRV3) then
16:    Call  $add\_additional\_distance\_column(differences\_X, distances)$ 
17:     $differences\_X = add\_additional\_distance\_column(differences\_X, distances)$ 
18:  else
19:    for each sample  $i$  in nearest neighbors indices do
20:       $base \leftarrow$  Index of base case
21:      for each neighbor  $n$  of  $i$  do
22:        Append  $(X[base] - X[n])$  to  $differences\_X$ 
23:        Append  $(y[base] - y[n])$  to  $differences\_y$ 
24:      end for
25:    end for
26:  end if
27:  Fit MLPRegressor with  $differences\_X$  and  $differences\_y$ 
28:  Set  $train\_X = X$ 
29:  Set  $train\_y = y$ 
30:  return Fitted LingerRegressor
31: end function
```

Prediction Implementation

During prediction, the algorithm utilizes the fitted Nearest Neighbors model to identify the nearest neighbors for each sample in the input data X . It then computes the

differences between each sample in X and its corresponding neighbors, preparing the input features for the regressor. After predicting the differences using the trained regressor, the algorithm aggregates the predictions to obtain the final predicted target values for each sample in X . If weighted KNN is enabled, the predictions are weighted based on the distances to the neighbors, providing more influence to closer neighbors in the prediction process. This offers a different form of nearest neighbors. Where values are weighted based on their proximity to the base case. It makes use of scikit-learns [13] `NearestNeighbors`, and `numpy`.

Algorithm 14 Prediction Implementation Algorithm for Base Linger Regressor

Input: Input data, X , number of neighbors $n_neighbours_2$

```
1: function PREDICT( $X$ )
2:   Fit Nearest Neighbors model with  $n\_neighbours\_2$  using  $train\_X$ 
3:   Compute nearest neighbors for each sample in  $X$ ,
4:   Compute the indexes and distances, of nearest neighbors in  $X$ 
5:   Initialize empty list  $differences\_test\_X$ 
6:    $is\_sparse\_X \leftarrow$  Check if  $X$  is sparse
7:   if  $is\_sparse\_X$  then
8:     Convert  $X$  to dense array
9:   end if
10:  for each sample  $test$  in  $X$  do
11:    for each nearest neighbor  $nn\_in\_training$  in  $indices[test]$  do
12:      Append  $(X[test, :] - train\_X[indices[test][nn\_in\_training], :])$  to
         $differences\_test\_X$ 
13:    end for
14:  end for
15:  if  $variant == 1$  (LRV1) then
16:    Call  $addition\_of\_context\_func\_pred(differences\_test\_X, X)$ 
17:     $differences\_test\_X = addition\_of\_context\_func\_pred(differences\_test\_X, X)$ 
18:  else if  $variant == 3$  (LRV3) then
19:    Call  $add\_additional\_distance\_column(differences\_test\_X, distances)$ 
20:     $differences\_X = add\_additional\_distance\_column(differences\_test\_X, distances)$ 
21:  end if
22:   $predictions \leftarrow$  Predict using LingerRegressor with  $differences\_test\_X$ 
23:  Split  $predictions$  into batches of size  $n\_neighbours\_2$ 
24:  Initialise  $y\_pred$  as an empty list
25:   $y\_pred = []$ 
26:  if Weighted KNN is enabled then
27:    for each  $indexes, differences, dist$  in  $zip(indices, predictions, distances)$  do
28:      Compute weights for each neighbor
29:       $weights = 1 \div (dist + 1e - 8)$ 
30:      Compute weighted sum of target values
31:       $weighted\_sum = \sum(w \times (train\_y[i] + d) \text{ for } i, d, w \text{ in } zip(indexes, differences, weights))$ 
32:      Compute weighted average
33:       $weighted\_avg = weighted\_sum \div \sum(weights)$ 
34:      Append  $weighted\_avg$  to  $y\_pred$ 
35:    end for
36:  else
37:    for each  $indexes, differences$  in  $zip(indices, predictions)$  do
38:      Compute sum of target values
39:       $sum\_result = \sum((train\_y[i] + d) \text{ for } i, d \text{ in } zip(indexes, differences))$ 
```


3.3.7 Evaluation

Evaluation metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) to assess model accuracy and performance.

3.3.8 Optimization

To optimize the model, we consider various strategies:

- Hyperparameter Tuning: Grid search or randomized search cross-validation for optimal hyperparameters.
- Feature Engineering: Experimentation with different feature combinations to improve model performance.
- Regularization Techniques: L1 and L2 regularization to prevent overfitting.

3.3.9 Limitations and Future Work

The Linger Regression Base Model has several limitations:

- Sensitivity to Noise: The model may produce inaccurate predictions when faced with noisy or irrelevant data.
- Interpretability: While the model enhances interpretability compared to black-box models, it may still be challenging to interpret in complex scenarios.

Future work includes:

- Enhancing Noise Robustness: Investigating methods to improve model performance in the presence of noisy data.
- Interpretability Improvements: Exploring techniques to enhance model interpretability for complex datasets.

3.3.10 Conclusion

In conclusion, the `LingerRegressor` model presents a novel approach to regression analysis by integrating principles of Case-Based Reasoning (CBR) and the learning

from differences methodology. Through the combination of K-nearest neighbors (KNN) and `MLPRegressor`, the model offers flexibility, accuracy, and interpretability in regression tasks. The implementation of supplementary features further enhances its adaptability and performance. Despite its limitations, the model shows promise in addressing the challenges of transparent and dynamic regression modeling. Future research can focus on refining the model's robustness to noise and improving its interpretability for complex datasets. Overall, the `LingerRegressor` model contributes to the advancement of machine learning methodologies, providing a valuable tool for regression analysis in diverse domains where a more interpretable model is needed.

3.4 Linger Classification Base Model:

`LingerClassifier` (LC)

3.4.1 Objective

The objective of the Linger Classification Base Model (`LingerClassifier`) is to extend the learning from differences methodology to classification tasks. The primary goals include:

- Developing a versatile classification model that utilizes differences between cases for prediction.
- Ensuring compatibility with Scikit-learn for seamless integration and extensive hyperparameter testing.
- Providing interpretable predictions to facilitate user understanding.
- Implementing enhancements to the learning from differences algorithm for improved classification performance.
- Designing a modular and extensible codebase to support future variations and enhancements.

The `LingerClassifier` is tailored for classification problems where interpretability and adaptability are critical.

3.4.2 Architecture

Similar to the regression counterpart, the `LingerClassifier` integrates K-nearest neighbors (KNN) and Scikit-learn's `MLPClassifier` for classification tasks. During training, KNN is used to compute nearest neighbors, and during prediction, the differences are used to adapt the class labels.

3.4.3 Components

The key components of the `LingerClassifier` include:

- **KNN Module:** Computes nearest neighbors and extracts differences during training and prediction phases.
- **MLPClassifier:** Offers flexibility in activation functions, solvers, and hyperparameters for classification.
- **Additional Features:** Similar to the regression model, features like weighted KNN, context addition, additional results' column, and duplication based on distance can be activated through hyperparameters.

3.4.4 Parameterization

The model provides extensive parameterization options for fine-tuning KNN, `MLPClassifier`, and additional features. Hyperparameters for the classification model mirror those available in standard KNN and neural networks, with additional parameters specific to `LingerClassifier`.

3.4.5 Training and Prediction Design

The training and prediction algorithms for the `LingerClassifier` are similar to those of the `LingerRegressor`, with adjustments made to accommodate classification tasks.

Training Algorithm Design

During training, the algorithm computes differences between cases and their nearest neighbors, similar to the regression model. However, the adaptation of class labels is

based on the majority class among the nearest neighbors.

Algorithm 15 Training Algorithm for LingerClassifier

Split the dataset D into training set D^{train} and test set D^{test} using an 80%-20% split.
Initialize ΔD^{train} as an empty list.
for each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 Retrieve n similar cases from D^{train} using nearest neighbors.
 for each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**
 Compute the differences: $\Delta D_{ij}^{train} = (X_i^{train} - X_j^{train}, y_i^{train} - y_j^{train})$.
 Append ΔD_{ij}^{train} to ΔD^{train} .
 end for
end for
Train the MLPClassifier neural network model using ΔD^{train} .
return Trained neural network model $TrainedNN()$.

Prediction Algorithm Design

During prediction, the algorithm adapts class labels based on the majority class among the nearest neighbors' labels.

Algorithm 16 Prediction Algorithm for Learning from Differences for Classification

Input: Test dataset D^{test} , trained neural network model $TrainedNN()$, number of neighbors n

- 1: Initialize empty list y^{pred}
- 2: **for** each case $X_j^{test} \in D^{test}$ **do**
- 3: Retrieve n similar cases from D^{train} using nearest neighbors
- 4: Initialize empty list y_j^{pred}
- 5: **for** each similar case $(X_i^{train}, y_i^{train}) \in D^{train}$ **do**
- 6: Compute the differences: $\Delta X_{ji}^{test} = X_j^{test} - X_i^{train}$
- 7: Use the trained neural network to predict: $TrainedNN(\Delta X_{ji}^{test}) = \Delta X_{ji}^{pred}$
- 8: Adapt neighbor's y value: $yy_{ji}^{pred} = y_i^{train} + \Delta X_{ji}^{pred}$
- 9: **end for**
- 10: Perform majority voting on y_j^{pred} predictions to obtain y_j^{pred}
- 11: **end for**
- 12: Compare y^{pred} to y^{test} to evaluate accuracy

3.4.6 Training and Prediction Implementation

Training Algorithm Implementation

The provided algorithm outlines the prediction process for the Linger Classifier implemented in Python. It begins by fitting a Nearest Neighbors model with a specified

number of neighbors using the training data. Next, it computes the nearest neighbors for each sample in the input data and initializes an empty list to store the differences between the test samples and their corresponding neighbors. If the input data is sparse, it is converted to a dense array for compatibility. The algorithm then iterates through each test sample and its nearest neighbors to calculate the feature differences. After computing the differences, it proceeds to make predictions using the trained classifier. If weighted KNN is enabled, it calculates the weighted results based on the differences and distances to determine the item with the highest weighted sum as the prediction. Otherwise, it finds the most common item among the predictions. Finally, the predicted target values are aggregated and returned as the output of the algorithm.

Algorithm 17 Training Implementation Algorithm for Base Linger Classifier

Input: Dataset, X , label vector y , number of neighbors $n_neighbors_1$

```
1: function FIT( $X, y$ )
2:   Increment  $n\_neighbors\_1$ 
3:    $neighbors \leftarrow$  Fit scikit-learn's NearestNeighbors function to  $X$ 
4:   Compute nearest neighbors for each sample in  $X$ 
5:    $distances, indices \leftarrow neighbors.kneighbors(X)$ 
6:    $differences\_X \leftarrow []$ 
7:    $differences\_y \leftarrow []$ 
8:   if  $variant == 1$  (LCV1) then
9:     Call  $addition\_of\_context\_func\_fit(X, y, indices, differences\_X, differences\_y)$ 
10:     $differences\_X, differences\_y = addition\_of\_context\_func\_fit(X, y, indices, differences\_X, differences\_y)$ 
11:  else if  $variant == 2$  (LCV2) then
12:    Call  $duplicate\_difference\_based\_on\_distance(differences\_X, differences\_y, distances)$ 
13:     $differences\_X, differences\_y = duplicate\_difference\_based\_on\_distance(differences\_X, differences\_y, distances)$ 
14:  else if  $variant == 3$  (LCV3) then
15:    Call  $add\_additional\_distance\_column(differences\_X, distances)$ 
16:     $differences\_X = add\_additional\_distance\_column(differences\_X, distances)$ 
17:  else
18:    for each sample  $i$  in nearest neighbors indices do
19:       $base \leftarrow$  Index of base case
20:      for each neighbor  $n$  of  $i$  do
21:        Append  $(X[base] - X[n])$  to  $differences\_X$ 
22:        Append  $(y[base] - y[n])$  to  $differences\_y$ 
23:      end for
24:    end for
25:  end if
26:  Fit MLPClassifier with  $differences\_X$  and  $differences\_y$ 
27:  Set  $train\_X = X$ 
28:  Set  $train\_y = y$ 
29:  return Fitted LingerClassifier
30: end function
```

Prediction Algorithm Implementation

The prediction algorithm for the Linger Classifier, implemented in Python, outlines the process of predicting target values for the input data. Initially, it fits a Nearest Neighbors model with a specified number of neighbors using the training data. Then, it

computes the nearest neighbors for each sample in the input data and initializes an empty list to store the feature differences between the test samples and their corresponding neighbors. If the input data is sparse, it is converted to a dense array for compatibility. The algorithm iterates through each test sample and its `NearestNeighbors` to calculate the feature differences. After computing the differences, it proceeds to make predictions using the trained `LingerClassifier`. If weighted KNN is enabled, it calculates the weighted sum of target values for each neighbor based on the differences and distances, and selects the item with the highest weighted sum as the prediction. Otherwise, it finds the most common item among the predictions. Finally, the predicted target values are aggregated and returned as the output of the algorithm.

Algorithm 18 Prediction Algorithm for Linger Classifier

Input: Input data, X , number of neighbors $n_neighbours_2$

```
1: function PREDICT( $X$ )
2:   Fit Nearest Neighbors model with  $n\_neighbours\_2$  using  $train\_X$ 
3:   Compute nearest neighbors for each sample in  $X$ 
4:   Initialize empty list  $differences\_test\_X$ 
5:    $is\_sparse\_X \leftarrow$  Check if  $X$  is sparse
6:   if  $is\_sparse\_X$  then
7:     Convert  $X$  to dense array
8:   end if
9:   for each sample  $test$  in  $X$  do
10:    for each nearest neighbor  $nn\_in\_training$  in  $indices[test]$  do
11:      Append  $(X[test, :] - train\_X[indices[test][nn\_in\_training], :])$  to
       $differences\_test\_X$ 
12:    end for
13:  end for
14:  if  $variant == 1$  (LCV1) then
15:    Call  $addition\_of\_context\_func\_pred(differences\_test\_X, X)$ 
16:     $differences\_test\_X = addition\_of\_context\_func\_pred(differences\_test\_X, X)$ 
17:  else if  $variant == 3$  (LCV3) then
18:    Call  $add\_additional\_distance\_column(differences\_test\_X, distances)$ 
19:     $differences\_X = add\_additional\_distance\_column(differences\_test\_X, distances)$ 
20:  end if
21:   $predictions \leftarrow$  Predict using LingerClassifier with  $differences\_test\_X$ 
22:  Split  $predictions$  into batches of size  $n\_neighbours\_2$ 
23:  Initialise  $y\_pred$  as an empty list
24:   $y\_pred = []$ 
25:  if Weighted KNN is enabled then
26:    for each  $indexes, differences, distance$  in
       $zip(indices, differences\_test\_X, distances)$  do
27:      Compute weights for each neighbor
28:       $weights = 1 \div (dist + 1e - 8)$ 
29:      Compute weighted sum of target values,
30:       $weighted\_sum = []$ 
31:       $weighted\_sum = [w \times (train\_y[i] + d)]$  for  $i, d, w$  in  $zip(indexes, differences, weights)$ 
32:      Find the item with the highest weighted sum
33:       $max\_sum = max(weighted\_sum)$ 
34:      Append  $max\_sum$  to  $y\_pred$ 
35:    end for
36:  else
37:    for each  $indexes, differences$  in  $zip(indices, predictions)$  do
38:       $results = [train\_y[i] + d \text{ for } i, d \text{ in } zip(indexes, differences)]$ 
39:      Find the most common item in  $results$ 
```


3.4.7 Evaluation

Evaluation metrics for the `LingerClassifier` include accuracy, precision, recall, F1-score, and confusion matrix analysis. These metrics provide insights into the model's performance and its ability to classify instances accurately.

3.4.8 Optimization

To enhance the performance of the `LingerClassifier`, several optimization strategies are considered:

- **Hyperparameter Tuning:** Utilize grid search or randomized search cross-validation to fine-tune the hyperparameters for optimal classifier performance.
- **Feature Engineering:** Experiment with different feature combinations and transformations to uncover informative patterns in the data and improve classifier accuracy.
- **Regularization Techniques:** Apply L1 and L2 regularization methods to the classifier to mitigate overfitting and enhance generalization ability.

These optimization strategies aim to maximize the predictive capability of the `LingerClassifier` model.

3.4.9 Limitations and Future Work

The `LingerClassifier` shares similar limitations and avenues for future work as the regression model, including sensitivity to noise and interpretability challenges. Future work may focus on improving noise robustness and enhancing interpretability for complex classification tasks.

3.4.10 Conclusion

The `LingerClassifier` offers a novel approach to classification tasks by integrating the learning from differences methodology. Its modular design and extensible architecture

make it well-suited for a wide range of classification problems, particularly those where interpretability and adaptability are paramount.

For more details on the `LingerClassifier`, and `LingerRegressor` variations, refer to Table 3.5.

3.5 Variations For `LingerRegressor` (LR) and `LingerClassifier` (LC)

3.5.1 Variation 1: Addition of Context Design and Implementation (LRV1, LCV1)

Objective

The objective of adding context is to incorporate the original attribute values of the source cases into the generated differences array. This ensures that the new datasets used for fitting and prediction contain both the original attribute values of the source cases and the differences between them and their nearest neighbors. This may aid in the learning process and improve the model. Its design is the same for both the `LingerRegressor` and `LingerClassifier`.

Training Design

Algorithm 19 Training Algorithm for Learning from Differences, with context included in `LingerClassifier` and `LingerRegressor`

- 1: Split the dataset D into training set D^{train} and test set D^{test} using an 80%-20% split.
 - 2: Initialize ΔD^{train} as an empty list.
 - 3: **for** each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 - 4: Retrieve n similar cases from D^{train} using nearest neighbors.
 - 5: **for** each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**
 - 6: Compute the differences, with concatenated context: $\Delta D_{ij}^{train} = ((X_i^{train} - X_j^{train}) : X_i^{train}, y_i^{train} - y_j^{train})$.
 - 7: Append ΔD_{ij}^{train} to ΔD^{train} .
 - 8: **end for**
 - 9: **end for**
 - 10: Train the `MLPClassifier` or `MLPRegressor` neural network model using ΔD^{train} .
 - 11: **return** Trained neural network model $TrainedNN()$.
-

Prediction Design

Algorithm 20 Prediction Algorithm for Variant 1, context included.

Input: Test dataset D^{test} , trained neural network model $TrainedNN()$, number of neighbors n

- 1: Initialize y^{pred} as an empty list.
 - 2: **for** each case X_j^{test} in D^{test} **do**
 - 3: Retrieve n similar cases from D^{train} using nearest neighbors.
 - 4: Initialize y_j^{pred} as an empty list.
 - 5: **for** each similar case $(X_i^{train}, y_i^{train})$ in D^{train} **do**
 - 6: Compute the differences: $\Delta X_{ji}^{test} = ((X_j^{test} - X_i^{train}) : X_j^{test})$.
 - 7: Use the trained neural network to predict: $Trained\ NN(\Delta X_{ji}^{test}) = \Delta X_{ji}^{pred}$.
 - 8: Adapt neighbor's y values: $y_{ji}^{pred} = y_i^{train} + \Delta X_{ji}^{pred}$.
 - 9: **end for**
 - 10: **if** Regression **then**
 - 11: Average the adapted y values: $y_j^{pred} = \frac{1}{n} \sum_{i=1}^n y_{ji}^{pred}$.
 - 12: **else if** Classification **then**
 - 13: Perform majority voting on y_j^{pred} to obtain y_j^{pred} .
 - 14: **end if**
 - 15: **end for**
 - 16: Compare y_j^{pred} to y_j^{test} to evaluate accuracy.
-

Training Implementation

This algorithm is called during the Fit stage of both the `LingerClassifier`, and `LingerRegressor`.

Algorithm 21 Context Addition Function for Base Linger Classifier

Input: Dataset X , label vector y , nearest neighbor indices $indices$, differences in X $differences_X$, differences in y $differences_y$ **Output:** Updated differences in X $differences_X$, updated differences in y $differences_y$

```
1: function addition_of_context_func_fit( $X$ ,  $y$ ,  $indices$ ,  $differences\_X$ ,  
    $differences\_y$ )  
2:   for  $i$  in  $indices$  do  
3:      $base \leftarrow i[0]$   
4:      $neighbors \leftarrow i[1:]$   
5:     for  $n$  in  $neighbors$  do  
6:        $base\_case \leftarrow X[base]$   
7:        $diff \leftarrow X[base] - X[n]$   
8:        $combined\_list \leftarrow []$   
9:       for  $pair$  in  $zip(diff, base\_case)$  do  
10:        for  $item$  in  $pair$  do  
11:          Append  $item$  to  $combined\_list$   
12:        end for  
13:      end for  
14:      Append  $np.array(combined\_list)$  to  $differences\_X$   
15:      Append  $y[base] - y[n]$  to  $differences\_y$   
16:    end for  
17:  end for  
18:  return  $differences\_X$ ,  $differences\_y$   
19: end function
```

Prediction Implementation

This algorithm is called during the predict algorithm for both the `LingerClassifier`, and `LingerRegressor`.

Algorithm 22 Context Addition Function for Base Linger Classifier during Prediction

Input: Differences in test data *differences_test_X*, input data *X* **Output:** Combined differences in test data *combined_differences_test_X*

```
1: function addition_of_context_func_pred(differences_test_X, X)
2:   combined_differences_test_X  $\leftarrow$  []
3:   duplicated_test_data  $\leftarrow$  [item for item in X for _ in range(n_neighbours_2)]
4:   num_items  $\leftarrow$  length of duplicated_test_data
5:   for item_pos in range(num_items) do
6:     combined_list  $\leftarrow$  []
7:     for pair in zip(differences_test_X[item_pos], duplicated_test_data[item_pos])
8:       do
9:         for item in pair do
10:          Append item to combined_list
11:        end for
12:      Append np.array(combined_list) to combined_differences_test_X
13:    end for
14:  return combined_differences_test_X
15: end function
```

3.5.2 Variation 2: Duplication Based on Distance Design and Implementation (LRV2, LCV2)

Objective

A novel variation introduced in this study involves duplicating differences based on distance, aiming to amplify the influence of neighbors closely resembling the base case. This innovative approach represents a departure from conventional methods in the learning from differences paradigm. By magnifying the impact of differences that exhibit stronger correlation with the base case, this technique effectively mitigates the influence of outliers, leading to more robust model performance.

This novel approach is specifically integrated into the fitting process of both the `LingerClassifier` and `LingerRegressor` models. By emphasizing the significance of closely resembling neighbors, this technique enhances the model's ability to discern meaningful patterns in the data, thereby improving its predictive accuracy and generalization capabilities.

1. **Normalize Distances:** Normalize the distances from the base case to its nearest neighbors represented by $Distances_{ij}$. We apply a logarithmic transformation to the distances: $Distances_{ij} = \log_e(Distances_{ij} + 1)$, with the addition of 1 to prevent a logarithm of 0.
2. **Determine Maximum and Minimum Distances:** Determine the maximum distance $MaxDistance$ and minimum distance $MinDistance$ in $Distances_{ij}$.
3. **Compress Distances:** Compress the distances within $Distances_{ij}$ to a range between 0 and 10 using the formula:

$$Distances_{ij} = \text{round} \left(10 - \frac{(Distances_{ij} - MinDistance)}{(MaxDistance - MinDistance) \times 10} \right)$$

4. **Duplicate Items:** Duplicate each item ΔD_{ij}^{train} by the corresponding distance in $Distances_{ij}$. It's important to note that this duplication process can significantly increase the size of the dataset and, consequently, the training time overhead. The process of duplication is only used in the training process of the algorithm.

Training Design

Algorithm 23 Duplication Based on Distance in LingerClassifier and LingerRegressor

Input: Distances $Distances_{ij}$, Base dataset ΔD^{train}

- 1: Normalize Distances:
 - 2: **for** each distance $Distances_{ij}$ **do**
 - 3: $Distances_{ij} \leftarrow \log_e(Distances_{ij} + 1)$ ▷ Apply logarithmic transformation
 - 4: **end for**
 - 5: Determine Maximum and Minimum Distances:
 - 6: $MaxDistance \leftarrow$ maximum value of $Distances_{ij}$
 - 7: $MinDistance \leftarrow$ minimum value of $Distances_{ij}$
 - 8: Compress Distances:
 - 9: **for** each distance $Distances_{ij}$ **do**
 - 10: $Distances_{ij} \leftarrow \text{round} \left(10 - \frac{(Distances_{ij} - MinDistance)}{(MaxDistance - MinDistance) \times 10} \right)$
 - 11: **end for**
 - 12: Duplicate Items:
 - 13: **for** each item ΔD_{ij}^{train} in ΔD^{train} **do**
 - 14: Duplicate ΔD_{ij}^{train} by the corresponding distance in $Distances_{ij}$
 - 15: **end for**
-

Training Implementation

Algorithm 24 Difference Duplication Based on Distance

```
1: function duplicate_difference_based_on_distance(differences_X,  
   differences_y, distances)  
2:   distances_X  $\leftarrow$  []  
3:   for d in distances do  
4:     neighbors  $\leftarrow$  d[1 :]  
5:     for n in neighbors do  
6:       append n to distances_X  
7:     end for  
8:   end for  
9:   distances_array  $\leftarrow$  np.array(distances_X)  
10:  log_values  $\leftarrow$  np.log(distances_array + 1)  
11:  min_val  $\leftarrow$  np.min(log_values)  
12:  max_val  $\leftarrow$  np.max(log_values)  
13:  if min_val == max_val then  
14:    scaled_distances_rounded  $\leftarrow$  np.ones_like(log_values).astype(int)  
15:  else  
16:    scaled_distances  $\leftarrow$   $10 - ((\log\_values - \min\_val) / (\max\_val - \min\_val) \times$   
17:    10)  
18:    scaled_distances_rounded  $\leftarrow$  np.round(scaled_distances).astype(int)  
19:  end if  
20:  filtered_data_X  $\leftarrow$  []  
21:  filtered_data_y  $\leftarrow$  []  
22:  for each item, count in zip(differences_X, scaled_distances_rounded) do  
23:    if count  $\neq$  0 then  
24:      append item to filtered_data_X  
25:    end if  
26:  end for  
27:  for each item, count in zip(differences_y, scaled_distances_rounded) do  
28:    if count  $\neq$  0 then  
29:      append item to filtered_data_y  
30:    end if  
31:  end for  
32:  duplicated_list_X  $\leftarrow$  []  
33:  duplicated_list_y  $\leftarrow$  []  
34:  for each item, count in zip(filtered_data_X, scaled_distances_rounded) do  
35:    for  $\_$   $\leftarrow$  1 to count do  
36:      append item to duplicated_list_X  
37:    end for  
38:  end for  
39:  for each item, count in zip(filtered_data_y, scaled_distances_rounded) do
```

3.5.3 Variation 3: Addition of Distance Column Design and Implementation (LRV3, LCV3)

Objective

Variation 3 introduces a novel approach by incorporating the distance between each base case and its nearest neighbor into both the training and prediction phases of the neural network. This innovative method aims to leverage the proximity information to enhance the learning process and predictive performance of the model.

By integrating distance information, the neural network gains insight into the spatial relationships between data points. This allows the model to potentially learn associations between differences and their corresponding distances, leading to improved accuracy and effectiveness in both training and prediction tasks.

The motivation behind this variation lies in its potential to provide the model with a deeper understanding of the underlying data structure. By considering distance as a factor in the learning process, the model can better capture nuanced patterns and relationships, resulting in more robust and reliable predictions.

Design for Training

In the fit phase, the distances between each base case D_i^{train} and its nearest neighbor in D^{train} are computed and stored in $Distances_{ij}$. To incorporate these distances into the training data, each distance in $Distances_{ij}$ is concatenated with the corresponding difference ΔD^{train} . This results in a new dataset where each difference is paired with its corresponding distance, ensuring a one-to-one ratio between distances and differences.

Algorithm 25 Training Algorithm for Learning from Differences, with Distance Column

Input: Dataset D , Neural network model, Number of neighbors n , Number of epochs E

- 1: Split D into training set D^{train} and test set D^{test} with an 80%-20% split
 - 2: Initialize ΔD^{train} and $Distances$ as empty lists
 - 3: **for** each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**
 - 4: Retrieve n similar cases from D^{train} using nearest neighbors
 - 5: **for** each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**
 - 6: Compute the differences: $\Delta D_{ij}^{train} = (X_i^{train} - X_j^{train}, y_i^{train} - y_j^{train})$
 - 7: Compute the distance: $Distances_{ij} = ||X_i^{train}, X_j^{train}||$ (Euclidean distance)
 - 8: Concatenate $Distances_{ij}$ as a new attribute to ΔD_{ij}^{train}
 - 9: **end for**
 - 10: **end for**
 - 11: Train the MLPRegressor neural network model using ΔD^{train} over E epochs
 - 12: **return** Trained neural network model $TrainedNN()$.
-

Design for Predict

In the prediction phase, distances between each test case X_j^{test} and its nearest neighbors in D^{train} are stored in $Distances_{ji}$. These distances are concatenated with the corresponding differences ΔX^{test} , ensuring a one-to-one ratio between distances and differences in the prediction dataset.

Algorithm 26 Prediction Algorithm for Learning from Differences with Additional Distance Column

Input: Test dataset D^{test} , Trained neural network model, Number of neighbors n , $TrainedNN()$

```

1: Initialize  $\Delta D^{train}$  and  $Distances$  as empty lists
2: for each case  $X_j^{test} \in D^{test}$  do
3:   Retrieve  $n$  similar cases from  $D^{train}$  using nearest neighbors
4:   for each similar case  $(X_i^{train}, y_i^{train}) \in D^{train}$  do
5:     Compute the differences:  $\Delta X_{ij}^{test} = (X_j^{test} - X_i^{train})$ 
6:     Compute the distance:  $Distances_{ji} = ||X_j^{test}, X_i^{train}||$  (Euclidean distance)
7:     Add  $Distances_{ji}$  as a new attribute to  $\Delta X_{ji}^{test}$ 
8:     Use the trained neural network to predict:  $TrainedNN(\Delta X_{ji}^{test}) = \Delta X_{ji}^{pred}$ 
9:     Adapt neighbor's  $y$  values:  $y_j^{pred} = y_i^{train} + \Delta X_{ji}^{pred}$ 
10:  end for
11:  if Regression then
12:    Average the adapted  $y$  values:  $y_j^{pred} = \frac{1}{n} \sum_{i=1}^n y_j^{pred}$ 
13:  else if Classification then
14:    Perform majority voting on  $y_j^{pred}$  to obtain  $y_j^{pred}$ 
15:  end if
16:  Compare  $y_j^{pred}$  to  $y_j^{test}$  to evaluate accuracy
17: end for

```

Implementation for Training, and Prediction

The same function is called during both predict and training.

Algorithm 27 Add Additional Distance Column

Input: Differences dataset $differences_X$, Distances of the nearest neighbors $distances$

```

1: function  $add\_additional\_distance\_column(differences\_X, distances)$ 
2:   Adds an additional column with distance values associated with each
    $differences\_X$  during fitting.
3:   Output:  $differences\_X$  with an additional column.
4:   for  $i$  in  $range(len(distances))$  do
5:      $diff\_res \leftarrow$  Convert  $distances[i]$  to numpy array
6:     Append  $diff\_res$  to  $differences\_X[i]$ 
7:   end for
8:   return  $differences\_X$ 
9: end function

```

3.6 Acronym Guide

Table 3.5: Variations for `LingerRegressor` and `LingerClassifier`

Variation	Acronym
Addition of Context	LRV1, LCV1
Duplication Based on Distance	LRV2, LCV2
Addition of Distance Column	LRV3, LCV3

3.7 Learning From differences Implicitly

3.7.1 Introduction and Motivation

This design segment aims to implement and extend the algorithms proposed by Ye et al. [16] for both classification and regression tasks. The primary objective is to develop a unified framework where a single set of algorithms can be employed for both classification and regression tasks, leveraging implicit learning from differences within case pairs.

As outlined in the literature review (see Section 2), the algorithms (Algorithm 9, Algorithm 10) proposed by Ye et al. were chosen as the basis for our model implementation. Variant 5, as demonstrated in the paper, exhibited superior performance among the variants explored, prompting its adoption in our implementation.

Moreover, while the paper focuses on classification tasks, it overlooks the potential application of implicit difference learning within case pairs for regression tasks. Thus, our secondary motivation is to explore the adaptability of Variant 5 to regression models. We seek to investigate whether the framework outlined in Algorithm 9 and Algorithm 10 can be extended and optimized for regression tasks effectively.

By extending the application of Variant 5 to regression tasks, we aim to provide a comprehensive and versatile framework for both classification and regression tasks, harnessing the power of implicit learning from differences within case pairs.

3.8 Linger Implicit Classification Model:

LingerImplicitClassifier (LIC)

3.8.1 Objective

The objective of the Linger Implicit Classification Base Model (`LingerImplicitClassifier`) is to implement the learning from differences methodology seen in [16] for classification tasks. The primary goals include:

- Developing a versatile classification model that learns through implicit differences in case pairs.
- Ensuring compatibility with Scikit-learn for seamless integration and extensive hyperparameter testing.
- Providing interpretable predictions to facilitate user understanding, for a variety of tasks.
- Implementing enhancements to the learning from differences algorithm for improved classification performance through hyperparameter testing.
- Designing a modular and extensible codebase to support future variations and enhancements.

The `LingerImplicitClassifier` is tailored for classification problems where interpretability and adaptability are critical, Where explicit differences may not be easily calculable. It adds another layer of interpretability, as both the target case and retrieved case are visible within the case pair which the Neural Nets are trained on.

3.8.2 Architecture

The `LingerImplicitClassifier` leverages a hybrid architecture incorporating K-nearest neighbors (KNN) and Scikit-learn's `MLPClassifier` to tackle classification tasks. In the training phase, KNN plays a pivotal role in computing nearest neighbors, thereby facilitating the creation of case pairs. During the prediction phase, KNN is again utilized to retrieve the most similar cases from each class.

The process of generating case pairs for predictions involves KNN's functionality to identify the nearest neighbors. These neighbors serve as crucial elements in forming

case pairs, enabling the classifier to make informed predictions based on similarities observed within the dataset.

The integration of KNN with the `MLPClassifier` offers a comprehensive approach to classification tasks, combining the strengths of both algorithms. While KNN excels in identifying similarities and forming case pairs, the `MLPClassifier` contributes by providing a powerful neural network-based classification model capable of learning intricate patterns and relationships within the data.

By harnessing the complementary capabilities of KNN and the `MLPClassifier`, the `LingerImplicitClassifier` aims to achieve robust and accurate classification results across diverse datasets, making it a versatile tool for various machine learning applications.

3.8.3 Components

The key components of the `LingerClassifier` include:

- **KNN Module:** Computes nearest neighbors and extracts differences during training and prediction phases.
- **`MLPClassifier`:** Offers flexibility in activation functions, solvers, and hyperparameters for classification.
- **Additional Features:** Class to class pairs, random N selected neighbors, and N nearest neighbors.

3.8.4 Parameterization

The model provides extensive parameterization options for fine-tuning KNN, `MLPClassifier`, and additional features. Hyperparameters for the classification model mirror those available in standard KNN and neural networks, with additional parameters specific to `LingerClassifier`.

3.8.5 Training and Prediction Design

The training and prediction algorithms for the `LingerImplicitClassifier`.

Training Algorithm Design

Algorithm 28 Training Algorithm for LingerImplicitClassifier

Input: Dataset D , number of neighbors 1

Split D into training set D^{train} and test set D^{test} with an 80%-20% split

Create *CasePairs()* dictionary.

Create *AdaptNN()*, a dictionary to store each Neural Network

List of unique Classes *UniqueClasses*

for each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**

 Retrieve 1 similar case from D^{train} using nearest neighbors.

for each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**

CasePairs(y_j^{train}).append($[X_j^{train} : X_i^{train}, y_i^{train}]$)

end for

end for

for each unique *Class* in *UniqueClasses* **do**

 Create a new Classification Neural Network, trained on Case Pairs

AdaptNN(*Class*).fit(*CasePairs*(*Class*))

end for

return Trained neural network models dictionary *AdaptNN*()

Prediction Algorithm Design

During the prediction phase, various methods exist for retrieving neighbors within each class. These methods can be categorized into three variants:

1. **Variant 1:** (LICV1) Selecting N random items from each class.
2. **Variant 2:** (LICV2) Selecting the single nearest neighbor from each class.

Each variant may be suitable depending on the specific characteristics of the dataset and the objectives of the prediction task.

Algorithm 29 Prediction Algorithm for `LingerImplicitClassifier`

Input: Dataset D^{test} , Dictionary $AdaptNN()$, Set $UniqueClasses$, and Dictionary $CasePairs()$

```
Initialize an empty list  $y^{pred}$ 
for each case  $X_j^{test} \in D^{test}$  do
    Create an empty list  $preds$ 
    for each unique class  $Class$  in  $UniqueClasses$  do
        if Base version: (LIC) then
            Retrieve  $N$  nearest neighbors from  $CasePairs(Class)$ 
        else if variant 1: (LICV1) then
            Select  $N$  random items from  $CasePairs(Class)$ 
        else if variant 2: (LICV2) then
            Select the single nearest neighbor from  $CasePairs(Class)$ 
        end if
        for each selected case  $(X_i^{CasePair}, y_i^{CasePair})$  do
            for each  $(X_z^{train}, X_q^{train}) \in X_i^{CasePair}$  do
                 $preds.append(AdaptNN(Class).predict(X_z^{train} : X_j^{test}))$ 
            end for
        end for
        Perform a majority vote on  $preds$ 
         $y_j^{pred} = MajorityVote(preds)$ 
    end for
end for
return  $y^{pred}$ 
```

3.8.6 Training and Prediction Implementation

Training Algorithm Implementation

The `fit` function of the `LingerImplicitClassifier` model trains the classifier using the provided training data X and target labels y . It begins by handling input data, converting sparse matrices to dense arrays if necessary. Next, it initializes structures to store trained adaptation neural networks and unique classes. The algorithm then computes the nearest neighbors for each sample in X , generates pairs of cases (consisting of base and retrieved cases), and organizes the data based on class labels. Finally, it trains adaptation neural networks for each class label using the generated case pairs and target labels. Overall, this process enables the

`LingerImplicitClassifier` model to adapt to specific instances retrieved during inference, enhancing its performance on the task.

Algorithm 30 Fit Function for LingerImplicitClassifier

Input: Training data X , target labels y

```
1: function FIT( $X, y$ )
2:    $is\_sparse\_X \leftarrow$  Check if  $X$  is sparse
3:    $is\_sparse\_y \leftarrow$  Check if  $y$  is sparse
4:   if  $is\_sparse\_X$  then
5:     Convert  $X$  to dense array
6:   end if
7:   if  $is\_sparse\_y$  then
8:     Convert  $y$  to dense array
9:   end if
10:   $classes\_ \leftarrow$  Get unique classes from  $y$ 
11:   $adaptation\_networks\_ \leftarrow \{\}$ 
12:   $neighbours \leftarrow$  Fit Nearest Neighbors with  $X$ 
13:   $distances, indices \leftarrow$  Compute nearest neighbors for each sample in  $X$ 
14:  Initialize  $case\_pairs\_X$  and  $case\_pairs\_y$ 
15:  for each  $i$  in  $indices$  do
16:     $base \leftarrow i[0]$ 
17:     $neighbors \leftarrow i[1:]$ 
18:    for each  $n$  in  $neighbors$  do
19:       $case\_x \leftarrow [X[base], X[n]]$ 
20:       $case\_y \leftarrow [y[base], y[n]]$ 
21:      Append  $case\_x$  to  $case\_pairs\_X$ 
22:      Append  $case\_y$  to  $case\_pairs\_y$ 
23:    end for
24:  end for
25:  Initialize dictionaries:  $case\_pairs\_by\_class, cases\_without\_concatenation,$   

    $case\_pairs\_y$ 
26:  for each ( $source, target$ ) in  $zip(case\_pairs\_X, case\_pairs\_y)$  do
27:    Append  $source[0]$  to  $cases\_without\_concatenation[target[1]]$ 
28:     $problem\_query \leftarrow$  Concatenate  $source[0]$  and  $source[1]$ 
29:    Append  $problem\_query$  to  $case\_pairs\_by\_class[target[1]]$ 
30:    Append  $target[0]$  to  $case\_pairs\_y[target[1]]$ 
31:  end for
32:  for each  $class\_label, cases$  in  $case\_pairs\_by\_class$  do
33:    Initialize MLPClassifier with parameters
34:    Fit MLPClassifier with  $cases$  and  $case\_pairs\_y[class\_label]$ 
35:    Store the fitted MLPClassifier in  $adaptation\_networks\_ [class\_label]$ 
36:  end for
37: end function
```

Prediction Algorithm Implementation

The `predict` function of the `LingerImplicitClassifier` model predicts class labels for the input samples. It creates nearest neighbor finders for each class, then iterates through the test samples to predict class labels. Depending on the chosen variant, it either selects nearest neighbors using (LICV1) or a single nearest neighbor using (LICV2). Finally, it performs majority voting to decide the final classification.

Algorithm 31 Predict Function for `LingerImplicitClassifier`

Input: Test data X

```
1: function PREDICT( $X$ )
2:    $predictions \leftarrow []$ 
3:    $is\_sparse\_X \leftarrow$  Check if  $X$  is sparse
4:   if  $is\_sparse\_X$  then
5:     Convert  $X$  to dense array
6:   end if
7:    $nbrs\_by\_class \leftarrow$  Create nearest neighbor finders for each class
8:   for each  $sample$  in  $X$  do
9:     if  $variant == 2$  (LICV2) then
10:       $class\_predictions \leftarrow$  Predict class labels for  $sample$  using 1
      NearestNeighbor
11:    else
12:       $class\_predictions \leftarrow$  Predict class labels for  $sample$  using  $N$ 
      NearestNeighbors
13:    end if
14:     $majority\_prediction \leftarrow$  Perform majority voting on  $class\_predictions$ 
15:    Append  $majority\_prediction$  to  $predictions$ 
16:  end for
17:  return  $predictions$ 
18: end function
```

Algorithm 32 Helper: Create Nearest Neighbor Finders

```
1: function _CREATE_NBRs_BY_CLASS
2:   nbrs_by_class  $\leftarrow \{\}$ 
3:   for each class_label in classes_ do
4:     n_neighbors  $\leftarrow 1$  if variant == 2 (LICV2) else n_neighbours_2
5:     nbrs  $\leftarrow$  NearestNeighbors(n_neighbors).fit(cases_without_concatenation[class_label])
6:     nbrs_by_class[class_label]  $\leftarrow$  nbrs
7:   end for
8:   return nbrs_by_class
9: end function
```

Algorithm 33 Helper: Predict Class Labels

```
1: function _predict_class(sample, nbrs_by_class)
2:   class_predictions  $\leftarrow \{\text{class\_label} : [] \text{ for class\_label in } \textit{classes\_}\}$ 
3:   for each class_label, nbrs in nbrs_by_class.items() do
4:     if variant == 1 (LICV1) then
5:       indices  $\leftarrow$  Randomly select n_neighbours_2 indices from
        range(len(cases_without_concatenation[class_label]))
6:     else
7:       distances, indices  $\leftarrow$  nbrs.kneighbors([sample])
8:     end if
9:     nearest_neighbors  $\leftarrow$  indices[0]
10:    for each neighbor_index in nearest_neighbors do
11:      neighbor_sample  $\leftarrow$  cases_without_concatenation[class_label][neighbor_index]
12:      merged_case_pair  $\leftarrow$  Concatenate neighbor_sample and sample
13:      adapted_prediction  $\leftarrow$  Predict with adaptation_networks_[class_label]
        using merged_case_pair
14:      Append adapted_prediction[0] to class_predictions[class_label]
15:    end for
16:  end for
17:  return class_predictions
18: end function
```

Algorithm 34 Helper: Perform Majority Voting

```
1: function _majority_vote(class_predictions)
2:   all_predictions  $\leftarrow$  [prediction for predictions in class_predictions.values() for prediction in p]
3:   prediction_counter  $\leftarrow$  Counter(all_predictions)
4:   most_common_predictions  $\leftarrow$  prediction_counter.most_common(1)
5:   If the case occurs where there is no distinct most common.
6:   if len(most_common_predictions) > 1 then
7:     majority_prediction  $\leftarrow$  Randomly select from most_common_predictions
8:   else
9:     majority_prediction  $\leftarrow$  most_common_predictions
10:  end if
11:  return majority_prediction
12: end function
```

3.8.7 Optimization

To enhance the performance of the `LingerImplicitClassifier`, several optimization strategies are considered:

- **Hyperparameter Tuning:** Utilize grid search or randomized search cross-validation to fine-tune the hyperparameters for optimal classifier performance.
- **Feature Engineering:** Experiment with different feature combinations and transformations to uncover informative patterns in the data and improve classifier accuracy.
- **Regularization Techniques:** Apply Regularization techniques to improve accuracy, and reduce noise.

3.8.8 Limitations and Future Work

The `LingerImplicitClassifier` including sensitivity to noise. Future work may focus on improving noise robustness and enhancing interpretability for complex classification tasks.

3.8.9 Conclusion

The `LingerImplicitClassifier` offers a novel approach to classification tasks by integrating the learning from differences implicitly through case pairs. It performs well for a variety of classification tasks. That could benefit from a more interpretable learning methodology

3.9 Linger Implicit Regression Model: `LingerImplicitRegressor`

3.9.1 Objective

The objective of the `LingerImplicitRegressor` is to pioneer a learning methodology seen in [16] tailored specifically for regression tasks, known as Learning from Differences. The primary objectives encompass:

- Creation of a flexible regression model capable of learning implicitly from differences observed in pairs of cases.
- Adaptation of regression tasks into the Learning from Differences framework to unlock its potential.
- Seamless integration with Scikit-learn to ensure compatibility and enable extensive hyperparameter optimization.
- Provision of interpretable predictions crucial for user comprehension across a wide spectrum of regression scenarios.
- Continuous enhancement of the Learning from Differences algorithm through rigorous hyperparameter testing to achieve superior regression performance.
- Designing a modular and extensible codebase to accommodate future variations and improvements.

The `LingerImplicitRegressor` is designed to address regression challenges where explicit differences may be elusive, prioritizing interpretability and adaptability. It implements the classification methodology described in [16] but facilitates its use for regression tasks. This is a novel approach. It provides added transparency by exposing

both the target case and the retrieved case within the case pair, facilitating a deeper understanding during model training.

3.9.2 Architecture

The `LingerImplicitRegressor` utilizes a hybrid architecture integrating K-nearest neighbors (KNN) with Scikit-learn's `MLPClassifier`. In the training phase, KNN plays a pivotal role in computing nearest neighbors, thereby facilitating the creation of case pairs. During the prediction phase, KNN is again utilized to retrieve the most similar cases from each class.

The process of generating case pairs for predictions involves KNN's functionality to identify the nearest neighbors. These neighbors serve as crucial elements in forming case pairs, enabling the regressor to make informed predictions based on similarities observed within the dataset.

The integration of KNN with the `MLPClassifier` offers an interesting approach to regression tasks, combining the strengths of both algorithms. It places regression values within specified ranges, offering a new type of solution to classic regression issues. While KNN excels in identifying similarities and forming case pairs, the `MLPClassifier` contributes by providing a powerful neural network-based classification model capable of learning intricate patterns and relationships within the data. Classifying regression values.

By harnessing the complementary capabilities of KNN and the `MLPClassifier`, the `LingerImplicitRegressor` aims to achieve robust and accurate regression results across diverse datasets while using a classifier, making it a versatile tool for various machine learning applications. This new approach to regression tasks that allows for categorization of regression results. It processes a regression task and converts it to a classification task, this could be useful for problems where ranges of values are highly important. This would make certain regression tasks more interpretable.

3.9.3 Components

The key components of the `LingerRegressor` include:

- KNN Module: Computes nearest neighbors and extracts differences during training and prediction phases.

- **MLPClassifier**: Offers flexibility in activation functions, solvers, and hyperparameters for classification.
- **Additional Features**: Class to class pairs, random N selected neighbors, and N nearest neighbors.

3.9.4 Parameterization

The model provides extensive parameterization options for fine-tuning KNN, **MLPClassifier**, and additional features. Hyperparameters for the classifier model mirror those available in standard KNN and neural networks, with additional parameters specific to **LingerRegressor**.

3.9.5 Training and Prediction Design

The training and prediction algorithms for the **LingerImplicitRegressor**.

Training Algorithm Design

Algorithm 35 Training Algorithm for LingerImplicitRegressor

Input: Dataset D , number of neighbors 1, N Classes

Split D into training set D^{train} and test set D^{test} with an 80%-20% split

Create *CasePairs()* dictionary.

Create *AdaptNN()*, a dictionary to store each Neural Network

Convert training data y values into classes, using the algorithm 36

$D^{train}y, unique_ranges, y_{train}^{max}, y_{train}^{min} = RegToClassConverterFunction(D^{train}, N)$

Convert test y values based on the train y value range. So that a new range is the same for both.

$D^{test}y, unique_ranges, y_{test}^{max}, y_{test}^{min} = RegToClassConverterFunction(D^{train}, N, y_{train}^{max}, y_{train}^{min})$

List of unique Classes *UniqueClasses*

for each case $C_i = (X_i^{train}, y_i^{train}) \in D^{train}$ **do**

 Retrieve 1 similar case from D^{train} using nearest neighbors.

for each similar case $(X_j^{train}, y_j^{train}) \in D^{train}$ **do**

$CasePairs(y_j^{train}).append([X_j^{train} : X_i^{train}, y_i^{train}])$

end for

end for

for each unique *Class* in *UniqueClasses* **do**

 Create a new Classification Neural Network, trained on Case Pairs

$AdaptNN(Class).fit(CasePairs(Class))$

end for

return Trained neural network models dictionary *AdaptNN()*.

Prediction Algorithm Design

In the prediction phase for regression, several methods exist for retrieving neighbors within each class. These methods can be categorized into three variants:

1. **Variant 1:** (LIRV1) Selecting N random items from each class.
2. **Variant 2:** (LIRV2) Selecting the single nearest neighbor from each class.

The prediction algorithm mirrors Algorithm 29. The key distinction lies in the preprocessing step, where continuous y values are converted into classes to enable accuracy testing.

3.9.6 Training and Prediction Implementation

Both the `LingerImplicitClassifier` and `LingerImplicitRegressor` share the same training and prediction implementations. This shared functionality is encapsulated in a base class, allowing for code reuse and maintenance. The algorithm 36 found in the 3.9.7 section is used during the data preprocessing stage in the same way as `sklearn LabelEncoder`. To convert the regression problem into a classification problem.

Training Implementation

The training algorithm for `LingerImplicitRegressor` is outlined in 3.8.6. The function 3.9.7 is called before training to convert y values into class labels. For both training and test datasets

Prediction Implementation

The prediction algorithm for `LingerImplicitRegressor` is outlined in 3.8.6.

3.9.7 Regression to Classification Converter Design and Implementation

(RTCC) This function serves to categorize continuous y target values for training, validation, and testing. It returns a human-readable range with their corresponding numeric prediction. This ensures that the results are interpretable to humans at the end of the process. Central to its operation is the parameter equal division, which indicates whether samples should be equally distributed into N classes. This feature enables uniform representation across different segments of the target variable, promoting fairness and consistency in the categorization process.

Regression to Classification Converter Design

The "Regression to Classification Converter" algorithm (Algorithm 36) is designed to transform continuous regression target values into categorical classes. It takes as input a dataset D , the desired number of segments N , and optional parameters such as y_{\max} , y_{\min} , and Equal division.

If Equal division is set, the algorithm divides the data into N segments based on nearly equal counts. Otherwise, it calculates the range for each segment based on the provided or inferred maximum and minimum values of y . It then iterates through each value in y , determining the corresponding category based on the segment ranges. The algorithm returns the transformed y values, along with the unique ranges, minimum and maximum values of y , providing a categorical representation suitable for classification tasks.

Algorithm 36 Converter algorithm, regression to classification

Input: Dataset D , number of segments N , optional: y_{\max} , y_{\min} , Equal division

```
1:  $(X, y) \in D$ 
2: Get the minimum and maximum values of the labels
3: Ensure that the max and min are used from the test set so that  $y$  values are classified
   the same in both fit and predict.
4: if Equal division then
5:   Divide the data into  $N$  segments based on nearly equal counts
6:    $quantiles \leftarrow \text{np.linspace}(0, 1, N + 1)$ 
7:   if  $unique\_ranges$  is None then
8:      $unique\_ranges \leftarrow \text{np.quantile}(y, quantiles)$ 
9:   end if
10:   $categories \leftarrow \text{np.digitize}(y, unique\_ranges[: -1])$ 
11:   $category\_counts \leftarrow \text{Counter}(categories)$ 
12:  Return  $categories, unique\_ranges, y_{\min}, y_{\max}$ 
13: else
14:   if  $y_{\max}$  and  $y_{\min}$  are provided then
15:     Leave  $y_{\max}$  and  $y_{\min}$  as is
16:   else
17:      $y_{\max} \leftarrow \text{MAX}(y)$ 
18:      $y_{\min} \leftarrow \text{MIN}(y)$ 
19:   end if
20:   Calculate the range for each segment for  $y$  values
21:    $segmentRange \leftarrow \frac{y_{\max} - y_{\min}}{N}$ 
22:   Initialize an empty list:  $categories$ 
23:   Initialize an empty dictionary:  $unique\_ranges$ 
24:    $current\_index \leftarrow 0$ 
25:   for each value  $val$  in  $y$  do
26:     for  $i$  in  $\text{range}(N)$  do
27:       if  $i < N - 1$  then
28:         if  $y_{\min} + i \times segmentRange \leq val < y_{\min} + (i + 1) \times segmentRange$  then
29:            $category \leftarrow (y_{\min} + i \times segmentRange) - (y_{\min} + (i + 1) \times$ 
30:              $segmentRange)$ 
31:           if category not in  $unique\_ranges$  then
32:              $unique\_ranges[category] \leftarrow current\_index$ 
33:              $current\_index \leftarrow current\_index + 1$ 
34:           end if
35:           Append  $unique\_ranges[category]$  to  $categories$ 
36:           break
37:         end if
38:       else
39:         if  $y_{\min} + i \times segmentRange \leq val \leq y_{\max}$  then
            $category \leftarrow (y_{\min} + i \times segmentRange) - (y_{\max})$ 
```

Regression to Classification Converter Implementation

This function is encapsulated within a class called `RegressionToClassificationConverter`. The `transform` function carries out the conversion of regression values to classification values.

3.9.8 Optimization

To enhance the performance of the `LingerImplicitRegressor`, several optimization strategies are considered:

- **Hyperparameter Tuning:** Utilize grid search or randomized search cross-validation to fine-tune the hyperparameters for optimal model performance.
- **Feature Engineering:** Experiment with different feature combinations and transformations to uncover informative patterns in the data and improve regressor accuracy.
- **Regularization Techniques:** Apply regularization techniques to improve accuracy and reduce noise.

3.9.9 Limitations and Future Work

One of the limitations of the `LingerImplicitClassifier` is its sensitivity to noise, particularly when labeling values based on the maximum and minimum. However, opting to evenly distribute samples into classes can mitigate this sensitivity. For future research, exploring alternative methods to bucket regression values based on diverse distributions could be an intriguing avenue.

3.9.10 Conclusion

The `LingerImplicitClassifier` offers a novel approach to regression tasks by integrating the learning from differences implicitly through case pairs. It performs well for a variety of regression tasks, while utilizing a learning from differences classification model. The number of segments greatly affects the outcome of this model.

Algorithm 37 Transform function

```
1: function TRANSFORM( $y$ ,  $min\_val = None$ ,  $max\_val = None$ ,  $unique\_ranges =$   
    $None$ )  
2:   Input:  
3:    $y$  (list or array-like): The continuous target variable.  
4:    $min\_val$  (float): Minimum value of the target variable.  
5:    $max\_val$  (float): Maximum value of the target variable.  
6:    $unique\_ranges$  (array-like): Predefined unique ranges for segments.  
7:   Output:  
8:    $categories$  (list): A list of categories (segment indices) corresponding to each value  
   in  $y$ .  
9:    $unique\_ranges$  (dict or array-like): Unique ranges for segments.  
10:   $min\_val$  (float): Minimum value of the target variable.  
11:   $max\_val$  (float): Maximum value of the target variable.  
12:  if  $min\_val$  is  $None$  and  $max\_val$  is  $None$  then  
13:     $min\_val \leftarrow \min(y)$   
14:     $max\_val \leftarrow \max(y)$   
15:  end if  
16:  if  $equal\_division$  is  $None$  then  
17:     $segment\_range \leftarrow (max\_val - min\_val) / n\_segments$   
18:     $categories \leftarrow []$   
19:     $unique\_ranges \leftarrow \{\}$   
20:     $current\_index \leftarrow 0$   
21:    for  $val$  in  $y$  do  
22:      for  $i$  in  $\text{range}(n\_segments)$  do  
23:        if  $i < n\_segments - 1$  then  
24:          if  $min\_val + i \times segment\_range \leq val < min\_val + (i + 1) \times$   
    $segment\_range$  then  
25:             $category \leftarrow \text{int}(min\_val + i \times segment\_range)$   
26:             $\text{int}((min\_val + (i + 1)) \times segment\_range)$   
27:            if  $category$  not in  $unique\_ranges$  then  
28:               $unique\_ranges[category] \leftarrow current\_index$   
29:               $current\_index \leftarrow current\_index + 1$   
30:            end if  
31:             $categories.append(unique\_ranges[category])$   
32:            break  
33:          end if  
34:        else  
35:          if  $min\_val + i \times segment\_range \leq val \leq max\_val$  then  
36:             $category \leftarrow \text{int}(min\_val + i \times segment\_range)$   
37:             $\text{int}(max\_val)$   
38:            if  $category$  not in  $unique\_ranges$  then  
39:               $unique\_ranges[category] \leftarrow current\_index$   
40:               $current\_index \leftarrow current\_index + 1$   
41:            end if  
42:             $categories.append(unique\_ranges[category])$   
43:            break  
44:          end if  
45:        end if  
46:      end for  
47:    end for  
48:    return  $categories, unique\_ranges, min\_val, max\_val$   
49:  else
```


3.10 Learning from Differences For Images

3.10.1 Introduction and Motivation

3.11 Linger Image Classification Model

LingerImageClassifier (LIMC)

3.11.1 Objective

3.11.2 Architecture

3.11.3 Components

3.11.4 Parameterization

3.11.5 Training and Prediction Design

3.11.6 Training and Prediction Implementation

3.11.7 Optimization

3.11.8 Limitations and Future Work

3.12 Linger Image Regressor Model

LingerImageRegressor (LIMR)

3.12.1 Objective

3.12.2 Architecture

3.12.3 Components

3.12.4 Parameterization

3.12.5 Training and Prediction Design

Chapter 4

Experiments

4.1 Research Objective

The overarching objective of this study is to investigate the efficacy of blended learning approaches in machine learning tasks, particularly focusing on regression and classification problems. Blended learning, which integrates the strengths of model-based and instance-based learning while mitigating their respective weaknesses, offers a promising avenue for enhancing the interpretability, reliability, and performance of machine learning systems.

Specifically, this research aims to achieve the following objectives:

1. **Evaluate Blended Learning Methods:** The first objective is to evaluate the performance of blended learning methodologies in contrast to traditional model-based and instance-based approaches across various regression and classification tasks. This assessment will encompass key metrics including accuracy, interpretability, computational efficiency, and robustness. By comparing the effectiveness of blended learning methods, we aim to determine their potential to enhance the overall performance of machine learning systems.
2. **Investigate Model Interpretability:** The second objective is to explore the interpretability of blended learning models compared to traditional black-box models, such as neural networks. Understanding the interpretability of these models is crucial for enhancing trust and usability in critical domains. By investigating how blended learning methods provide insights into model decisions, we aim to demonstrate their potential to address the interpretability challenges associated with traditional machine learning approaches.

3. **Examine Efficiency and Scalability:** The third objective is to investigate the computational efficiency and scalability of blended learning methods, particularly in scenarios with large datasets. This examination will involve analyzing the memory and computational overhead of blended learning models compared to traditional approaches. By assessing the efficiency and scalability of blended learning methods, we aim to identify their suitability for practical deployment in real-world applications.
4. **Explore Novel Architectures:** The fourth objective is to explore novel architectures and techniques for implementing blended learning, including ensemble models, case-based reasoning principles, and approaches that leverage differences between sets of features. By exploring innovative approaches, we aim to further enhance the performance and interpretability of blended learning methods, thereby advancing the state-of-the-art in machine learning.
5. **Provide Insights for Practical Applications:** The fifth objective is to provide insights and recommendations for the practical application of blended learning methods in real-world scenarios, with a particular focus on critical domains such as healthcare, finance, and decision support systems. By discussing the implications of the findings for stakeholders and suggesting best practices for deploying blended learning models effectively, we aim to facilitate the adoption of these methods in critical domains.
6. **Develop Python Package for Seamless Integration:** The final objective is to develop a Python package that seamlessly integrates with scikit-learn, allowing users to incorporate blended learning methods into their machine learning pipelines and workflows effortlessly. This package should provide compatibility with scikit-learn's APIs, enabling users to import it like any other scikit-learn package and use it seamlessly within their existing workflows. By developing such a package, we aim to enhance the accessibility and usability of blended learning methods for the broader machine learning community.

By addressing these objectives, this research seeks to advance our understanding of blended learning approaches and their potential to address the challenges faced by traditional model-based and instance-based methods. Ultimately, the goal is to contribute to the development of more interpretable, reliable, and efficient machine learning systems with broader applicability across various domains.

4.1.1 Data Collection and Preprocessing

All datasets used in this study were collected from the UCI Machine Learning Repository [1], a widely recognized and reliable source of machine learning datasets. These datasets cover various domains and have been extensively studied in the machine learning community.

Preprocessing For Labelled Datasets

The preprocessing procedure involves several steps to ensure the quality and suitability of the data for machine learning tasks. These steps are applied uniformly across all labeled datasets.

1. **Column Naming:** The columns of the dataset are assigned appropriate names for clarity and consistency throughout the analysis.
2. **Handling Missing Values:** Rows containing missing values in specified features are removed from the dataset. This ensures that the dataset used for training and testing is complete and does not contain any incomplete or unreliable data points.
3. **Index Resetting:** After removing rows with missing values, the index of the dataset is reset to maintain a continuous and consistent index structure.
4. **Data Splitting:** The dataset is split into two subsets: a development set and a test set. This splitting is performed using a train-test split ratio of 80 : 20 to ensure a sufficient amount of data for training while also reserving a portion for evaluating the model's performance.
5. **Development Set Copying:** A copy of the development set is created to preserve the original dataset for future reference or comparison.
6. **Feature Transformation:** Numeric features undergo feature scaling using the `StandardScaler` to ensure that they are on a similar scale, which aids in model convergence and performance. Nominal features are transformed using one-hot encoding to convert categorical variables into a numerical format suitable for machine learning algorithms.

These preprocessing steps collectively prepare the data for subsequent modeling and analysis, ensuring that it is clean, complete, and appropriately formatted for use in machine learning tasks.

4.2 Results and Analysis

4.2.1 Experiment 1: Glass Dataset Classification

Methodology

For Experiment 1, we performed classification tasks on the Glass dataset. Our approach involved a two-step process. Firstly, we trained a neural network (NN) classifier using grid search with cross-validation to explore various hyperparameter combinations. Subsequently, we utilized the best hyperparameters from the NN training to train a Linger Classifier (LC) and its variants, aiming for enhanced robustness. The LC models were then tested for robustness by conducting five runs.

For comparison a KNN, weighted KNN and Neural Netowrk were run on the same data. To guage the relative performance of THE `LingerClassifier`.

The following sections summarize the results obtained from Experiment 1, presenting the hyperparameters used and the corresponding test accuracy achieved.

Table 4.1: Tested Hyperparameters for Neural Network and Linger Classifier

Hyperparameter	Values
Neural Network	
Hidden Layer Sizes	(256, 128), (128, 64), (100,), (200, 100), (300, 200, 100)
Activation	identity, logistic, tanh, relu
Alpha	0.0001, 0.001, 0.01, 0.1
Max Iterations	1000, 1500
Early Stopping	True
Validation Fraction	0.1, 0.2, 0.3
Learning Rate Init	0.001, 0.01, 0.1
Solver	adam, sgd
Beta 1	0.9, 0.95, 0.99
Beta 2	0.999, 0.995, 0.9
Linger Classifier	
Neighbors 1	2, 5, 7, 10, 13, 15, 17, 21
Neighbors 2	2, 5, 7, 10, 13, 15, 17, 21
Weighted KNN	False
Additional Results Column	False
Duplicated on Distance	False
Addition of Context	False

(LC) Results

Table 4.2: Best Parameters for KNN Classifier

Hyperparameter	Value
n_neighbors	2

Table 4.3: Best Parameters for Weighted KNN Classifier

Hyperparameter	Value
n_neighbors	7

Table 4.4: Best Parameters for Neural Net Classifier

Hyperparameter	Value
predictor__activation	relu
alpha	0.0001
beta_1	0.9
beta_2	0.995
early_stopping	True
hidden_layer_sizes	(300, 200, 100)
learning_rate_init	0.01
max_iter	1000
solver	adam
validation_fraction	0.3

Table 4.5: Best Parameters for Linger Classifier

Hyperparameter	Value
activation	relu
addition_of_context	False
additional_results_column	False
alpha	0.0001
beta_1	0.9
beta_2	0.995
duplicated_on_distance	False
early_stopping	True
hidden_layer_sizes	(300, 200, 100)
learning_rate_init	0.01
max_iter	1000
n_neighbors_1	2
n_neighbors_2	5
solver	adam
validation_fraction	0.3
weighted_knn	False

Table 4.6: Average Training Accuracy Results

Classifier	Average Training Accuracy (%)
KNN Classifier	68.2%
Weighted KNN Classifier	70.0%
Neural Net Classifier	73.1%
Linger Classifier	70.04%

Table 4.7: Average Test Accuracy Results

Classifier	Average Test Accuracy (%)
KNN Classifier	60.46%
Weighted KNN Classifier	62.8%
Neural Net Classifier	61.4%
Linger Classifier	60.46%

(LCV1)

(LCV2)

(LCV3)

(LIC)

(LICV1)

(LICV2)

4.2.2 Experiment 6: Raisin Dataset Classification

(LC)

(LCV1)

(LCV2)

(LCV3)

(LIC)

(LICV1)

(LICV2)

4.2.3 Experiment 7: Car Safety Dataset Classification

(LC)

(LCV1)

(LCV2)

(LCV3)

(LIC)

(LICV1)

gained from the analysis.

Chapter 5

Conclusions and Future Work

Chapter 6

Results

Bibliography

- [1] David Aha. Machine learning repository, 1987.
- [2] David W Aha, Dennis Kibler, and Marc K Albert. Instance-based learning algorithms. *Machine learning*, 6:37–66, 1991.
- [3] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [4] Susan Craw, Nirmalie Wiratunga, and Ray C Rowe. Learning adaptation knowledge to improve case-based reasoning. *Artificial intelligence*, 170(16-17):1175–1192, 2006.
- [5] Mathieu D’Aquin, Emmanuel Nauer, and Justine Lieber. A factorial study of neural network learning from differences for regression. In Mark T. Keane and Nirmalie Wiratunga, editors, *Case-Based Reasoning Research and Development. ICCBR 2022. Lecture Notes in Computer Science*, volume 13405, page PageNumbersHere. Springer, Cham, 2022.
- [6] Preeti Gupta, Arun Sharma, and Rajni Jindal. Scalable machine-learning algorithms for big data analytics: a comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(6):194–214, 2016.
- [7] Kathleen Hanney and Mark T Keane. Learning adaptation rules from a case-base. In *European Workshop on Advances in Case-Based Reasoning*, pages 179–192. Springer, 1996.
- [8] Vahid Jalali, David Leake, and Najmeh Forouzandehmehr. Learning and applying case adaptation rules for classification: An ensemble approach. In *IJCAI*, pages 4874–4878, 2017.
- [9] David B Leake and Andrew Kinley. Acquiring case adaptation knowledge: A hybrid approach. 1996.

- [10] Leorey Marquez, Tim Hill, Reginald Worthley, and William Remus. Neural network models as an alternative to regression. In *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences*, volume 4, pages 129–135. IEEE, 1991.
- [11] Christopher K Riesbeck and Roger C Schank. *Inside case-based reasoning*. Psychology Press, 2013.
- [12] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.
- [13] scikit-learn contributors. scikit-learn.
<https://github.com/scikit-learn/scikit-learn>, Accessed: Month Day, Year.
- [14] Sergios Theodoridis. *Machine learning: a Bayesian and optimization perspective*. Academic press, 2015.
- [15] Ian Watson and Farhi Marir. Case-based reasoning: A review. *The knowledge engineering review*, 9(4):327–354, 1994.
- [16] Xiaomeng Ye, David Leake, Vahid Jalali, and David J Crandall. Learning adaptations for case-based classification: A neural network approach. In *Case-Based Reasoning Research and Development: 29th International Conference, ICCBR 2021, Salamanca, Spain, September 13–16, 2021, Proceedings 29*, pages 279–293. Springer, 2021.