Adaptive k-Nearest Neighbor Learning for Robust Modal Regression on Multimodal and Heavy-Tailed Data

Sutarman[1,*], Netti Herawati[2], Adli Abdillah Nababan[3]

[1]*Department of Mathematics, University of Sumatera Utara, Medan, Indonesia*

[2]*Department of Mathematics, University of Lampung, Bandar Lampung, Indonesia*

[3]*Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta, Indonesia*

(Received: November 10, 2025; Revised: January 5, 2026; Accepted: March 12, 2026; Available online: April 18, 2026)

Abstract

Modal regression has attracted increasing attention as an alternative to mean-based regression, particularly in settings characterized by heteroscedasticity, multimodal conditional distributions, and heavy-tailed noise. In such scenarios, estimators based on central tendency may yield predictions that fall in low-density regions of the response space. This paper proposes an adaptive k-nearest neighbor framework for modal regression that integrates entropy-guided neighborhood selection with nonparametric mode estimation, including MeanShift clustering and one-dimensional kernel density estimation. The proposed approach adjusts neighborhood size based on local uncertainty, allowing the regression model to adapt to variations in data density without relying on a globally fixed parameter. Extensive experiments on simulated datasets and real-world benchmarks demonstrate that adaptive modal regression methods generally reduce or stabilize prediction errors relative to fixed-*k* modal regression and classical kNN mean and median estimators, particularly under heteroscedastic and multimodal conditions, although the magnitude of improvement varies across scenarios. Statistical tests confirm significant differences in most experimental settings, with practical gains ranging from incremental to substantial depending on data complexity. In addition to accuracy, computational behavior is explicitly examined. The findings show a trade-off between computational cost and predictive robustness: entropy-guided adaptive modal regression requires additional runtime due to neighborhood adaptation and density estimation, but this overhead increases proportionally with sample size and remains manageable for medium-sized datasets. Based on these results, adaptive modal regression provides a useful and flexible alternative for regression tasks involving complex and heterogeneous data distributions where robustness is prioritized over minimal computation time.

Keywords: Adaptive K-Nearest Neighbors, Modal Regression, Entropy-Based Neighborhood Selection, Robust Nonparametric Regression, Multimodal Conditional Distributions, Heavy-Tailed Noise, Instance-Based Learning, Meanshift Clustering, Kernel Density Estimation, Local Mode Estimation.

1. Introduction

Regression is a fundamental task in supervised machine learning with applications in sensor analytics, economics, environmental modeling, and human activity recognition, among other areas [1]. Most regression models estimate the conditional mean of a response variable given a set of predictors. This paradigm underpins classical approaches, such as linear and kernel regression and instance-based methods, including k-nearest neighbors (kNN), as well as many modern ensemble and deep learning models [2]. However, mean-based regression is less effective when these assumptions are violated, as it performs well only under unimodal, symmetric, and light-tailed noise distributions [3].

In practical machine learning applications, heteroscedasticity, multimodality, and heavy-tailed noise are frequently found in data. Heteroscedasticity results in biased and ineffective mean estimates as well as unreliable uncertainty quantification because the variance of the noise depends on the input space. Latent subpopulations, regime-switching behavior, or unobserved confounders are common causes of multimodal conditional distributions, which result in the conditional mean lying in low-density regions that are not typical outcomes [4]. By introducing extreme observations that disproportionately affect mean-based predictors, heavy-tailed noise distributions such as Student's t or Cauchy distributions further exacerbate these issues [5].

***Corresponding author: Sutarman (sutarman@usu.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i2.1221

By focusing on the conditional mode rather than the mean or median, modal regression provides a morally sound alternative [5], [6]. The conditional mode is resistant to outliers and heavy-tailed noise by nature, and it represents the most likely outcome given the predictors. Furthermore, modal regression can recover dominant local modes that mean- or quantile-based estimators consistently miss when conditional distributions are multimodal. Due to these characteristics, modal regression is particularly appealing in decision-making situations, where the most likely outcome, rather than average behavior, is the main concern [7].

Most existing modal regression methods rely on kernel density estimation (KDE) to approximate the conditional density function and identify its mode [8]. Although KDE-based approaches are flexible and theoretically well-established, they suffer from several practical limitations, including high computational cost and sensitivity to bandwidth selection [4]. These challenges become more pronounced in heterogeneous and multimodal data settings, limiting their applicability to large-scale machine learning problems.

Instance-based learning methods, particularly k-nearest neighbors (kNN), offer a natural and scalable framework for local, nonparametric regression. Their locality, minimal distributional assumptions, and adaptability to complex data geometries make them attractive for modal regression tasks [9]. In this context, kNN can be combined with mode-seeking techniques such as MeanShift clustering or local one-dimensional KDE to estimate conditional modes within a neighborhood of a query point [10]. However, most existing kNN-based approaches rely on a fixed neighborhood size, which becomes problematic when data density varies across the feature space. A fixed number of neighbors may oversmooth local structures in dense regions or produce unstable estimates in sparse regions, particularly under multimodal or heavy-tailed noise.

Adaptive neighborhood selection provides a promising solution to this limitation [11], [12]. By allowing the neighborhood size to vary locally according to data characteristics, adaptive kNN methods can balance bias and variance more effectively across heterogeneous regions of the input space. Information-theoretic criteria, such as entropy, offer a principled means of quantifying local uncertainty and have been successfully applied in adaptive nearest neighbor learning [13]. However, despite their potential, entropy-guided adaptive neighborhood strategies have received limited attention in the context of modal regression and supervised mode-seeking learning frameworks.

In this work, we propose an adaptive k-nearest neighbor learning framework for robust modal regression that explicitly addresses heteroscedasticity, multimodality, and heavy-tailed noise. The proposed approach integrates entropy-based adaptive neighborhood selection with nonparametric mode estimation techniques, including MeanShift clustering and one-dimensional kernel density estimation. The entropy criterion is used to determine the local neighborhood size in a data-driven manner, enabling the model to respond dynamically to variations in local data complexity. Within each adaptively selected neighborhood, conditional modes are estimated using mode-seeking algorithms that are robust to outliers and distributional heterogeneity.

The proposed framework is evaluated through extensive experiments on both controlled synthetic datasets and real-world machine learning benchmarks. The synthetic experiments are designed to isolate the effects of heteroscedasticity, multimodality, and heavy-tailed noise, while real-world datasets demonstrate the practical relevance of the method. Empirical results show that adaptive modal regression yields stable and robust predictions across diverse noise structures. While mean- and median-based kNN estimators may achieve lower average error in specific settings, adaptive modal regression consistently recovers the most probable outcomes and exhibits superior robustness under complex conditional distributions.

This study investigates adaptive modal regression within an instance-based learning framework, with particular emphasis on regression problems characterized by heteroscedasticity, multimodality, and heavy-tailed noise. While most machine learning regression methods are designed to estimate conditional means or medians, such approaches may yield predictions located in low-density regions of the conditional distribution under complex noise structures. To address this limitation, the present work focuses on conditional mode estimation combined with adaptive neighborhood learning.

The main objectives of this study are to develop an entropy-guided adaptive neighborhood selection strategy for k-nearest neighbor (kNN) regression, enabling the local neighborhood size to be determined in a data-driven manner based on distance-induced uncertainty rather than relying on a globally fixed or heuristically chosen value of k. In addition, this study aims to integrate adaptive kNN selection with nonparametric mode estimation techniques, specifically MeanShift clustering and one-dimensional kernel density estimation (KDE), in order to construct robust modal regression estimators capable of handling multimodal, heteroscedastic, and heavy-tailed conditional distributions. Furthermore, the proposed adaptive modal regression framework is systematically evaluated on both controlled synthetic datasets and real-world machine learning benchmarks, comparing its performance against classical kNN mean and median estimators in terms of predictive accuracy, robustness, and computational efficiency. This study also aims to provide mechanistic and interpretable insights into adaptive modal regression behavior by analyzing adaptive neighborhood dynamics, local mode structures, and predicted-versus-true relationships, thereby clarifying when and why entropy-guided neighborhood adaptation improves conditional mode estimation.

The main contributions of this work include the development of an entropy-guided adaptive-k modal regression framework, in which the local neighborhood size is selected by minimizing Shannon entropy computed from inverse-distance-based probability distributions, allowing the estimator to adapt automatically to local data density and structural complexity. The proposed framework integrates adaptive kNN selection with two nonparametric mode estimation techniques—MeanShift clustering and one-dimensional KDE—yielding flexible and robust conditional mode estimators without imposing global distributional assumptions. Through extensive experiments on synthetic datasets exhibiting heteroscedasticity, multimodality, and heavy-tailed noise, as well as multiple real-world regression benchmarks, the results demonstrate that adaptive modal regression consistently recovers the most probable outcomes and exhibits improved robustness compared to classical mean- and median-based kNN estimators.

In addition to predictive performance, this study provides visualization-based diagnostics and adaptive-k behavior analyses that reveal how local data complexity governs neighborhood selection and mode estimation, thereby enhancing the interpretability and transparency of instance-based modal regression. Finally, the proposed methods are formulated using standard machine learning components, including kNN search, entropy measures, MeanShift, and KDE, making them readily reproducible and easy to integrate into existing machine learning pipelines.

2. Literature Review

2.1. Mean-Based and Robust Regression in Machine Learning

Traditionally, machine learning regression techniques have concentrated on estimating a response variables conditional mean given input features. Because of their ease-of-use, interpretability and solid, theoretical underpinnings traditional methods like ordinary least squares kernel regression and k-nearest neighbors (kNN) regression are still frequently employed [14], [15]. Specifically local averaging over a fixed neighborhood yields a nonparametric approximation to the regression function in kNN regression which is a canonical instance-based learning paradigm [16].

However, mean-based regression estimators are known to be sensitive to deviations from idealized assumptions. In the presence of heteroskedastic noise, conditional mean estimates may exhibit inflated variance and unstable predictions [17]. More critically, when the conditional distribution of the response variable is multimodal, the conditional mean may fall within regions of low probability density. This yields predictions that are not representative of any plausible outcome [18]. Similarly, under heavy-tailed noise, extreme observations can dominate the estimation process, resulting in poor predictive performance and robustness [19], [20].

Some strong alternatives that have been put forth to address these problems are L1-based regression, quantile regression, and median regression [21]. These techniques increase robustness by focusing on conditional quantiles rather than the mean. However, since they still use a single measure of central tendency to summarize the distribution, quantile- and median-based approaches are essentially limited when the conditional distribution is multimodal [4] , [22], whereas modal regression can capture these multiple local maxima [8]. Consequently, the most likely result, which is often of primary interest in decision-making contexts, may be overlooked.

2.2. Modal Regression and Mode-Seeking Approaches

Modal regression was introduced [23] as a principled alternative to mean- and quantile-based regression[21], particularly for handling heterogeneous noise structures. It works by directly estimating the conditional mode, defined as the value of the response with the highest conditional probability density given the input [8]. By targeting the most probable outcome, modal regression exhibits inherent robustness to outliers and heavy-tailed noise, as extreme values in the distribution tails have limited influence on the location of the mode.

Most existing modal regression methods rely on kernel density estimation (KDE) to approximate the conditional density function, either globally or locally, followed by mode extraction [4], [6], [22]. These KDE-based approaches are theoretically well grounded and flexible, but they suffer from two major practical limitations. First, their performance is highly sensitive to bandwidth selection, which becomes particularly challenging under heterogeneous noise structures and varying data density [24]. Second, their computational cost scales poorly with sample size and dimensionality, limiting their applicability in many machine learning settings [25].

Mode-seeking algorithms such as MeanShift have also been widely studied in unsupervised learning for identifying density modes [10]. MeanShift offers an iterative, nonparametric mechanism for locating local maxima of a density estimate[26] without requiring explicit specification of the number of modes. While MeanShift has been successfully applied in clustering and computer vision, its integration into supervised regression—particularly for conditional mode estimation—has received relatively limited attention. Existing studies often employ MeanShift with fixed neighborhood or bandwidth parameters, which restricts adaptability under heterogeneous data distributions.

2.3. Adaptive kNN and Local Learning Strategies

Instance-based learning methods, particularly kNN, provide a natural framework for local, nonparametric regression [16]. The locality of kNN makes it especially attractive for modeling complex data structures without imposing strong global assumptions. However, classical kNN regression relies on a fixed neighborhood size, which introduces a fundamental bias–variance trade-off [8]. A large neighborhood may oversmooth local structures in dense regions, while a small neighborhood may yield unstable estimates in sparse regions.

Adaptive kNN techniques in which the neighborhood size is dynamically chosen based on local data characteristics have been proposed to overcome this limitation. Using distance distributions cross-validation or information-theoretic criteria like entropy [27] previous research has investigated adaptive neighborhood selection [28], [29]. These techniques have shown enhanced performance in mean-based regression and classification tasks especially in high-variance or heterogeneous environments.

Despite these advances, the application of adaptive neighborhood selection to modal regression remains underexplored. Existing adaptive kNN methods primarily focus on optimizing mean or classification objectives and do not explicitly address conditional mode estimation. Moreover, few studies have systematically combined entropy-guided adaptive neighborhood selection with mode-seeking techniques such as MeanShift or kernel-based density estimation within a unified regression framework.

2.4. Positioning of the Present Work

The present study bridges this gap by integrating entropy-based adaptive kNN learning with nonparametric conditional mode estimation. Unlike existing KDE-based modal regression approaches, the proposed framework avoids global bandwidth tuning and instead adapts the neighborhood size locally according to data complexity. Furthermore, by embedding mode-seeking mechanisms within an instance-based learning paradigm, the proposed methods remain computationally tractable and interpretable.

In contrast to prior adaptive kNN methods that target conditional means or classification boundaries, this work explicitly focuses on robust conditional mode estimation under multimodal, heteroskedastic, and heavy-tailed noise. This positioning distinguishes the proposed approach from both classical modal regression and existing adaptive kNN strategies, motivating the methodological developments and empirical evaluations presented in the subsequent sections.

3. Methods

This section details the proposed adaptive modal regression framework, including problem formulation, entropy-based neighborhood selection, and mode estimation procedures. The methodology is specifically designed to robustly estimate the conditional mode under heteroscedastic, multimodal, and heavy-tailed noise distributions. As illustrated in figure 1, the framework begins by constructing a local neighborhood, followed by determining the optimal neighborhood size

ISSN 2723-6471

through an entropy minimization criterion. The conditional mode is then estimated using either Mean Shift or kernel density estimation. For benchmarking purposes, a baseline kNN regression model is also included for comparison.

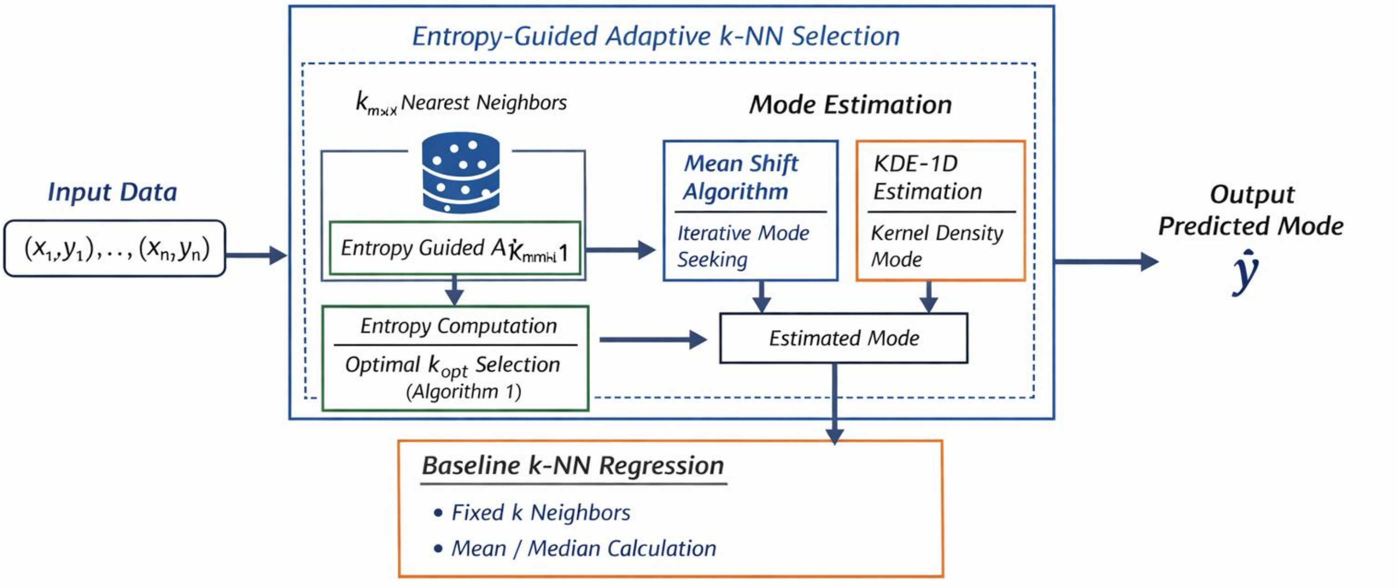

Figure 1. Research Framework

3.1. Problem Formulation

Let $\{(x_i, y_i)\}_{i=1}^{n}$ denote a set of training samples, where $x_i \in R^p$ represents the input feature vector and $y_i \in R$ is the corresponding response variable, with $n$denoting the total number of training samples. The objective of modal regression is to estimate the conditional mode function

$$m(x) = \arg\max_{y} f\,(\,y \mid x\,) \tag{1}$$

$f(x)$ denotes the conditional density of Y given $X = x$. Unlike mean regression, which estimates $E[X = x]$, modal regression focuses on the most probable outcome and is therefore more robust to outliers and distributional asymmetry [23].

3.2. *k*-Nearest Neighbor Framework

For a query point x, the k-nearest neighbor (kNN) approach identifies the k samples $\{(x_{(i)}, y_{(i)})\}_{i=1}^{k}$ such that

$$d_{(1)} \le d_{(2)} \le \cdots \le d_{(k)} \tag{2}$$

$d_{(i)} = \| x - x_{(i)} \|_2$denotes the Euclidean distance between the query point $x$and its i-th nearest neighbor $x_{(i)}$, ordered such that $d_{(1)} \le d_{(2)} \le \cdots \le d_{(k)}$ [2]. The local response set $\{y_{(i)}\}_{i=1}^{k}$ forms the basis for conditional mode estimation.

3.3. Entropy-Guided Adaptive Neighborhood Selection

Shannon entropy is adopted in this work as a general and widely used information-theoretic measure for quantifying local uncertainty and concentration in distance-based neighborhood distributions. Its use is motivated by its interpretability, computational simplicity, and prior applications in adaptive nearest neighbor and information-based learning, while alternative uncertainty or dispersion measures may also be considered in future extensions. To adaptively select the neighborhood size k, we define a distance-based probability distribution over the nearest neighbors. Specifically, the probability associated with the i-th neighbor is defined as

$$p_i^{(k)} = \frac{(1/d_{(i)} + \varepsilon)}{\sum_{j=1}^{k} 1/d_{(j)} + \varepsilon}, \quad i = 1, \ldots, k \tag{3}$$

$\varepsilon > 0$ is a small constant for numerical stability, and the denominator ensures that $\sum_{i=1}^{k} p_i^{(k)} = 1$.

This is introduced solely to avoid division by zero in the distance-based probability normalization. In all experiments, ε is fixed to a small value and not treated as a tunable hyperparameter; empirical observations indicate that the adaptive neighborhood selection is insensitive to small variations of ε within a reasonable range. This formulation assigns higher probability mass to closer neighbors, thereby emphasizing local structure. The Shannon entropy of the neighborhood for a given k is then computed as

$$H(k) = -\sum_{i=1}^{k} p_i \log p_i^{(k)} \tag{4}$$

The optimal neighborhood size k_{opt} is determined by minimizing the entropy. Let $k_{\min}$and $k_{\max}$denote the minimum and maximum neighborhood sizes considered for adaptive selection, respectively, where $1 \leq k_{\min} < k_{\max} \leq n$. In this study, $k_{\max}$defines the largest candidate neighborhood used to evaluate local uncertainty.

$$k_{\text{opt}} = \arg \min_{k \in [k_{min}, k_{max}]} H(k) \tag{5}$$

Since the neighborhood size k is selected from a finite and discrete set $[k_{\min}, k_{\max}]$, the entropy criterion is evaluated exhaustively for all candidate values and does not involve gradient-based optimization. When multiple values of k yield identical entropy values, the smallest k is selected to ensure deterministic and reproducible behavior. This criterion selects the neighborhood size that yields the most concentrated distance-based probability distribution among the candidate neighborhoods. This entropy-based criterion also balances locality and stability by selecting the most informative neighborhood scale [30]

3.4. Mode Estimation Techniques

Within the selected neighborhood, we estimate the conditional mode using two nonparametric techniques.

For MeanShift-Based Mode Estimation:

Given local responses $\{y_i\}_{i=1}^{k}$, MeanShift iteratively updates the mode estimate as

$$m^{(t+1)} = \frac{\sum_{i=1}^{k} y_i K\left(\frac{m^{(t)}-y_i}{h}\right)}{\sum_{i=1}^{k} K\left(\frac{m^{(t)}-y_i}{h}\right)} \tag{6}$$

$K(\cdot)$ is the Gaussian kernel and h is the bandwidth. The bandwidth is selected using Silverman's rule:

$$h = 1.06\,\hat{\sigma}\,k^{-1/5}, \tag{7}$$

with $\hat{\sigma}$ denoting the sample standard deviation of the local responses [4]

Kernel Density Estimation (KDE-1D)}

Alternatively, the local conditional density is estimated via one-dimensional KDE:

$$\hat{f}(y) = \frac{1}{kh} \sum_{i=1}^{k} K\left(\frac{y-y_i}{h}\right), \tag{8}$$

$K(\cdot)$ is the Gaussian kernel. The conditional mode is then obtained as

$$\hat{y} = \arg \max_{y \in [\min(y_i), \max(y_i)]} \hat{f}(y) \tag{9}$$

using bounded scalar optimization, such as the derivative-free methods described in [26], [31]

3.5. Algorithm

To improve the clarity and reproducibility of the methodology, detailed pseudocode for the main components and each proposed method is presented below. This pseudocode explicitly describes the algorithmic steps, including neighbor selection, bandwidth estimation, and conditional mode search. Nonparametric kNN [16] and mode-seeking methods like MeanShift and KDE [4] form the foundation of this implementation. Algorithm 1 presents an adaptive strategy for selecting the optimal number of neighbors (k_{opt}) through entropy minimization. Given sorted distances and a predefined range $[k_{min}, k_{max}]$, the algorithm evaluates each candidate k by computing the local probability distribution p_i (Eq. (3)) and the corresponding entropy $H(k)$ (Eq. (4)). The value of k that produces the smallest entropy is selected as k_{opt}. This procedure ensures that the chosen neighborhood size captures the most stable local structure by minimizing uncertainty in the data distribution.

Algorithm 1. Adaptive k Selection via Entropy Minimization

1: Input: Sorted distances $d_{(1)}, \dots, d_{(k_{\max})}$, minimum neighborhood size $k_{\min}$, maximum neighborhood size $k_{\max}$, and stability constant ε.
2: Initialize $H_{min} \leftarrow \infty, k_{opt} \leftarrow k_{min}$
3: for $k = k_{min}$ to k_{max} do
4: Compute p_i using Eq. (3)
5: Compute entropy $H(k)$ using Eq. (4)
6: if $H(k) < H_{min}$ then
7: $H_{min} \leftarrow H(k)$
8: $k_{opt} \leftarrow k$
9: end if
10: end for
11: Return k_{opt}

Algorithm 2 outlines a modal regression procedure that combines a fixed k-nearest neighbors (kNN) approach with the Mean Shift algorithm. First, a fixed number of nearest neighbors is selected to define the local data subset around the query point. Next, the bandwidth parameter is computed using Eq. (7), which determines the scale of the kernel used in density estimation. The Mean Shift procedure is then applied based on Eq. (6) to iteratively shift data points toward regions of higher density. Finally, the algorithm returns the mean of the dominant cluster, representing the estimated mode of the conditional distribution.

Algorithm 2. Fixed kNN + MeanShift Modal Regression

1: Select fixed k nearest neighbors
2: Compute bandwidth using Eq. (7)
3: Apply MeanShift using Eq. (6)
4: Return dominant cluster mean

Algorithm 3 describes an adaptive modal regression framework that integrates adaptive kNN selection with the Mean Shift procedure. Initially, a maximum set of nearest neighbors (k_{max}) is identified to capture the local data structure. From this set, the optimal neighborhood size (k_{opt}) is determined using Algorithm 1, which selects the most informative subset based on entropy minimization. The Mean Shift algorithm is then applied to the selected neighbors to iteratively locate regions of high data density. Finally, the dominant mode is returned as the regression estimate, representing the most probable outcome within the local neighborhood.

Algorithm 3. Adaptive kNN + MeanShift Modal Regression

1: Find k_{max} nearest neighbors
2: Select k_{opt} using Algorithm 1
3: Apply MeanShift on selected neighbors
4: Return dominant mode

Algorithm 4 presents an adaptive modal regression approach that combines adaptive kNN selection with one-dimensional kernel density estimation (KDE). First, a maximum set of nearest neighbors (k_{max}) is identified to represent the local data structure. The optimal neighborhood size (k_{opt}) is then determined using Algorithm 1, ensuring a stable and informative subset. Based on the selected neighbors, the conditional density is estimated using Eq. (8). The mode of this estimated density is subsequently computed using Eq. (9), yielding the regression estimate. Finally, the predicted value $\hat{y}$ is returned as the modal response.

Algorithm 4. Adaptive kNN + KDE-1D Modal Regression

1: Find k_{max} nearest neighbors
2: Select k_{opt} using Algorithm1
3: Estimate density using Eq. (8)
4: Compute mode using Eq. (9)

5: Return $\hat{y}$

Algorithm 5 describes a baseline regression approach using the k-nearest neighbors (kNN) framework with fixed neighborhood size. First, a predetermined number of nearest neighbors is selected to form the local subset around the query point. The prediction is then obtained by computing either the mean or the median of the corresponding response values within this neighborhood. Finally, the resulting value is returned as the regression estimate, providing a simple benchmark for comparison with more advanced modal regression methods.

Algorithm 5. Baseline kNN Mean and Median Regression

1: Select fixed k nearest neighbors
2: Compute mean or median of responses
3: Return prediction

3.6. Experimental Setup

To evaluate the performance of the proposed modal regression methods, we employed two types of data: simulated data and real-world data. The simulated data were designed to represent challenging scenarios in regression, including heteroskedasticity, multimodal (specifically bimodal) distributions, and heavy-tailed distributions. Meanwhile, real-world data was sourced from kaggle.com to test the methods' applicability in practical cases.

Three scenarios for simulated data are considered: heteroscedastic, bimodal, and heavy-tailed noise, with n sample varies 500, 1000, and 2000 samples per scenario, with controlled random seeds for reproducibility. In general, the simulated data scenario is designed with a single independent variable $X \sim U(0,1)$ to isolate the effect of the form of the noise distribution on the mode estimation. The selected mixture weights and degrees of freedom are chosen to induce moderate multimodality and heavy-tailed behavior that are frequently observed in practical regression problems. The goal of these simulation settings is to evaluate robustness under controlled yet representative distributional variations rather than to replicate a specific real-world dataset. The following are the mathematical descriptions for each scenario:

3.6.1. Heteroscedastic Scenario

The independent variable X, was generated uniformly from the interval [0, 1], i.e., $X \sim U(0,1)$. The true response value was $y_{true} = 2X + sin\, sin\, (2\pi X)$. Noise was added with variance dependent on X, specifically $\sigma = 0.1 + 0.5X$, so the observations were $y_{obs} = y_{true} + \epsilon$, where $(\epsilon \sim N(0, \sigma)$. This scenario represents heteroskedasticity, where the residual variance increases with X, commonly encountered in financial or environmental data.

3.6.2. Multimodal (Bimodal) Scenario

The independent variable $X \sim U(0,1)$. The response was generated as a mixture of two distributions with probability 0.6, $(y = 2X + \epsilon_1\, where\, \epsilon_1 \sim N(0,0.2))$ and with probability 0.4, $\left(y = 2X + 4 + \epsilon_2 \text{where} \epsilon_2 \sim N(0,0.2)\right)$. The true primary mode was $y_{true} = 2X$. This simulates conditional multimodal distributions, relevant for data with heterogeneous subpopulations, such as in medical or social analyses.

3.6.3. Heavy-Tailed Scenario

The values $X \sim U(0,1)$. The true value was $y_{true} = sin\, sin\, (3\pi X)$. Noise followed a Student's t-distribution with degrees of freedom 3, scaled by 0.5, i.e., $\epsilon \sim t_3(0,0.5)$, so $y_{obs} = y_{true} + \epsilon$. This heavy-tailed distribution represents outliers, common in financial or geophysical data. The methods were also applied to three real-world datasets from Kaggle.com, stored as CSV files: har.csv, housing.csv, and hour.csv. These datasets were selected for their variation in dimensions, scales, and application domains.

1) har.csv (Human Activity Recognition Using Smartphones: https://www.kaggle.com/datasets/ashish32700/human-activity-recognition-using-smartphones)

   Measurements from accelerometers and gyroscopes on smartphones worn by thirty subjects while they went about their daily lives including walking sitting and standing are included in this dataset. With activity labels as the target variable, it comprises 561 features that were taken from time and frequency domain signals. The dataset which contains 10299 instances is frequently used for sensor data classification or regression tasks. In order to predict continuous metrics like signal energy based on time features we concentrated on modal regression in this study.

Although the HAR dataset is originally formulated as a classification problem, in this study the categorical activity labels are encoded as ordered numerical values to construct a structured response variable. This transformation allows evaluation of regression behavior under clustered and multimodal response distributions. The objective is methodological assessment of the proposed regression framework rather than reinterpretation of the dataset in a semantic regression context.

2) housing.csv (California Housing Dataset: https://www.kaggle.com/datasets/camnugent/california-housing-prices):

   This dataset comprises 20,640 instances of housing data, with 10 features such as crime rate, accessibility, and building age, and the target variable being median house price. It is a classic dataset for regression tasks, exhibiting heteroskedasticity and outliers. We used numeric features to predict prices, with standardization preprocessing via StandardScaler.

3) hour.csv (Bike Sharing Dataset – Hourly: https://www.kaggle.com/datasets/lakshmi25npathi/bike-sharing-dataset):

   The dataset contains 17,379 instances of hourly bike rentals in Washington, D.C. The features include temperature, humidity, wind speed, and time. The target variable is the count of total rental bikes. Due to weather and time factors, it exhibits seasonal patterns and multimodality. We applied modal regression to predict rentals based on environmental features.

All datasets were normalized using StandardScaler to ensure consistent scaling across features and split into 80%: 20% training-testing sets.

3.7. Hyperparameter Configuration

To ensure fairness in comparison, hyperparameters were fixed and applied consistently across all methods. Fixed-k baselines use $k = 50$, which lies within the candidate range $[10,100]$ explored by the adaptive methods. For adaptive models, $k_{\text{min}} = 10$ and $k_{\text{max}} = 100$ are predefined and remain constant across datasets and repetitions. Bandwidth parameters for MeanShift clustering and one-dimensional kernel density estimation are computed using Silverman's rule of thumb and applied uniformly. All models are evaluated using identical train–test splits and 30 Monte Carlo repetitions with controlled random seeds to ensure reproducibility and fair comparison.

3.8. Statistical Comparison and Effect Size Interpretation

To evaluate consistency across Monte Carlo repetitions, pairwise MAE differences were analyzed using the Wilcoxon signed-rank test. This non-parametric procedure assesses whether the median paired difference differs from zero and is appropriate given the absence of normality assumptions. Statistical significance therefore reflects consistent directional improvement rather than large absolute differences. In addition, practical relevance is assessed through percentage improvement ("Better By (%)") relative to the reference method. While statistically significant differences indicate reliability, substantive conclusions are drawn only when accompanied by meaningful MAE reductions.

4. Results and Discussion

This section presents a comprehensive evaluation of the proposed adaptive modal regression methods using both simulated and real-world datasets. The discussion integrates quantitative accuracy metrics, statistical significance tests, average computational time, and visual diagnostics to provide a balanced assessment of predictive behavior and practical implications. This evaluation is grounded in the empirical results summarized in tables 1 to table 4. Specifically, table 1 and table 2 report the quantitative performance of the evaluated methods in terms of MAE and RMSE for simulated and real-world datasets, respectively, while table 3 and table 4 present the Wilcoxon signed-rank test results for MAE across all real datasets.

4.1. Simulated Datasets

The mean and standard deviation of root mean squared error (RMSE) and mean absolute error (MAE) over 20 repetitions for three simulated scenarios—heavy-tailed noise heteroscedastic noise and bimodal conditional distributions—are summarized in table 1. These situations gradually show increasingly intricate departures from conventional regression assumptions.

Table 1. The Quantitative Performance of the Evaluated Methods in Terms of MAE and RMSE for Simulated Datasets (*n_reps*=20)

Dataset	Method	*n_samp*	MAE (mean ± standard deviation)	RMSE (mean ± standard deviation)	Mean Pred. Time (s)
Simulated-1 (Heavy-Tailed)	A	500	0.2292 ± 0.0273	0.2845 ± 0.0278	11.0676
		1000	0.2303 ± 0.0179	0.2877 ± 0.0178	21.5229
		2000	0.1969 ± 0.0140	0.2456 ± 0.0153	44.0573
	B	500	0.2540 ± 0.0159	0.3328 ± 0.0205	1.7363
		1000	0.2624 ± 0.0164	0.3308 ± 0.0214	3.5561
		2000	0.2580 ± 0.0160	0.3281 ± 0.0205	7.0307
	C	500	0.1972 ± 0.0164	0.2560 ± 0.0220	0.5646
		1000	0.1965 ± 0.0133	0.2443 ± 0.0176	1.0966
		2000	0.2001 ± 0.0118	0.2535 ± 0.0141	2.1431
	D	500	0.1380 ± 0.0154	0.1710 ± 0.0163	0.0493
		1000	0.0769 ± 0.0058	0.0979 ± 0.0074	0.0961
		2000	0.1007 ± 0.0069	0.1241 ± 0.0074	0.1890
	E	500	0.1089 ± 0.0113	0.1420 ± 0.0115	0.0538
		1000	0.0750 ± 0.0054	0.0997 ± 0.0086	0.1053
		2000	0.0823 ± 0.0062	0.1017 ± 0.0070	0.2088
Simulated-2 (Heteroscedastic)	A	500	0.2045 ± 0.0250	0.2643 ± 0.0436	9.3749
		1000	0.1714 ± 0.0141	0.2331 ± 0.0229	19.0943
		2000	0.1759 ± 0.0097	0.2224 ± 0.0094	37.2102
	B	500	0.1936 ± 0.0258	0.2624 ± 0.0383	1.7194
		1000	0.1780 ± 0.0159	0.2386 ± 0.0202	3.4520
		2000	0.1920 ± 0.0096	0.2529 ± 0.0095	7.0074
	C	500	0.1466 ± 0.0213	0.1979 ± 0.0336	0.5641
		1000	0.1369 ± 0.0120	0.1901 ± 0.0156	1.1289
		2000	0.1506 ± 0.007	0.1995 ± 0.0082	2.2049
	D	500	0.0752 ± 0.0093	0.1128 ± 0.0135	0.0494
		1000	0.0579 ± 0.0061	0.0803 ± 0.0077	0.0984
		2000	0.0424 ± 0.0036	0.0566 ± 0.0055	0.1913
	E	500	0.0884 ± 0.0112	0.1306 ± 0.0163	0.0542
		1000	0.0629 ± 0.0068	0.0848 ± 0.0078	0.1077
		2000	0.0536 ± 0.0044	0.0726 ± 0.0065	0.2113
Simulated-3 (Bimodal)	A	500	0.0889 ± 0.0821	0.3054 ± 0.3494	5.3145
		1000	0.7963 ± 0.1981	1.7453 ± 0.2211	10.5031
		2000	0.3226 ± 0.1297	1.0502 ± 0.2574	20.7552
	B	500	1.1367 ± 0.22070	2.0853 ± 0.2274	1.5591
		1000	1.8082 ± 0.1262	2.6616 ± 0.0970	3.1334
		2000	1.4900 ± 0.1291	2.4033 ± 0.1092	6.2529
	C	500	0.8576 ± 0.1878	1.7752 ± 0.2150	0.5838
		1000	1.3398 ± 0.1104	2.2623 ± 0.0992	1.1522
		2000	1.0305 ± 0.1168	1.9637 ± 0.1194	2.2673
	D	500	1.4727 ± 0.0343	1.4949 ± 0.0350	0.0491
		1000	1.7419 ± 0.0382	1.7605 ± 0.0378	0.0972
		2000	1.5977 ± 0.0322	1.6170 ± 0.0326	0.1936
	E	500	0.2229 ± 0.0419	0.3143 ± 0.1471	0.0538
		1000	0.7765 ± 0.1341	1.4070 ± 0.1658	0.1136
		2000	0.3863 ± 0.0892	0.7919 ± 0.1922	0.2124

A = Fixed kNN + MeanShift, B = Adaptive kNN + MeanShift, C = Adaptive kNN + KDE-1, D = kNN Mean, E = kNN Median

4.1.1. Heavy-Tailed Noise Scenario

The reported standard deviations show that all assessed methods show increased variability when heavy-tailed noise is present. The mean absolute error (MAE) of the k-nearest neighbors (kNN) mean estimator (Method D) decreases from 0.1380 ± 0.0154 at n = 500 to 0.1007 ± 0.0069 at n = 2000 indicating stable but moderate performance. In a number of configurations adaptive modal regression techniques show reduced average errors. For instance, adaptive kNN +

MeanShift (method B) exhibits somewhat greater variability whereas adaptive kNN + KDE-1D (method C) attains mean absolute error (MAE) values of roughly 0.20 across sample sizes.

When evaluated relative to the baseline using the "Better By (%)" metric, the adaptive methods demonstrate consistent percentage improvements in several heavy-tailed settings, although the magnitude of improvement varies with sample size. In particular, the percentage reduction in MAE indicates that entropy-guided adaptive neighborhood selection yields measurable gains over fixed-*k* approaches, even when absolute differences appear numerically moderate. The adaptive and fixed-*k* approaches differ statistically significantly in the majority of comparisons according to the Wilcoxon test results in table 2. Importantly, statistical significance is interpreted jointly with the magnitude of improvement, ensuring that small but statistically detectable differences are not overstated as practically substantial. This implies that, without completely removing the variability present in heavy-tailed data, entropy-guided neighborhood selection lessens sensitivity to extreme observations while maintaining consistent relative performance gains.

Table 2. Wilcoxon Test-MAE for each scenario for Simulated Data (reps=20, α = 0.05)

Datasets with Scenario	**Between Methods**		***p*-Value (better by %)**	***p*-Value (better by %)**	***p*-Value (better by %)**
	Ref	**Comp**	**(*n_samp*=500)**	**(*n_samp*=1000)**	**(*n_samp* =2000)**
Simulated Data-1 (Heavy-Tailed)	A	B	0.0056 (-10.84)	0.0000 (-13.95)	0.0000 (-31.07)
		C	0.0001 (+13.95)	0.0000 (+14.65)	0.4100 (-1.66)
		D	0.0000 (+39.76)	0.0000 (+66.62)	0.0000 (+48.87)
		E	0.0000 (+52.50)	0.0000 (+67.44)	0.0000 (+58.17)
	B	C	0.0000 (+22.37)	0.0000 (+25.10)	0.0000 (+22.44)
		D	0.0000 (+45.65)	0.0000 (+70.71)	0.0000 (+60.99)
		E	0.0000 (+57.15)	0.0000 (+71.42)	0.0000 (+68.09)
	C	D	0.0000 (+30.00)	0.0000 (+60.89)	0.0000 (+49.71)
		E	0.0000 (+44.80)	0.0000 (+61.84)	0.0000 (+58.86)
	D	E	0.0000 (+21.15)	0.1054 (+2.44)	0.0000 (+18.19)
Simulated Data-2 (Heteroscedastic)	A	B	0.0897 (+5.33)	0.0973 (-3.81)	0.0000 (-9.18)
		C	0.0000 (+28.29)	0.0000 (+20.12)	0.0000 (+14.37)
		D	0.0000 (+63.21)	0.0000 (+66.25)	0.0000 (+75.90)
		E	0.0000 (+56.78)	0.0000 (+63.29)	0.0000 (+69.52)
	B	C	0.0000 (+24.26)	0.0000 (+23.06)	0.0000 (+21.57)
		D	0.0000 (+61.14)	0.0000 (+67.49)	0.0000 (+77.93)
		E	0.0000 (+54.35)	0.0000 (+64.64)	0.0000 (+72.08)
	C	D	0.0000 (+48.69)	0.0000 (+57.75)	0.0000 (+71.86)
		E	0.0000 (+39.73)	0.0000 (+54.04)	0.0000 (+64.40)
	D	E	0.0000 (-17.47)	0.0000 (-8.76)	0.0000 (-26.50)
Simulated Data-3 (Bimodal)	A	B	0.0000 (-1178.78)	0.0000 (-127.08)	0.0000 (-361.81)
		C	0.0000 (-864.87)	0.0000 (-68.26)	0.0000 (-219.4)
		D	0.0000 (-1556.87)	0.0000 (-118.76)	0.0000 (-395.20)
		E	0.0000 (-150.73)	0.0000 (+2.48)	0.0000 (-19.72)
	B	C	0.0000 (+24.55)	0.0000 (+25.90)	0.0000 (+30.8)
		D	0.0000 (-29.57)	0.0296 (+3.67)	0.0007 (-7.23)
		E	0.0000 (+80.39)	0.0000 (+57.06)	0.0000 (+74.08)
	C	D	0.0000 (-71.72)	0.0000 (-30.01)	0.0000 (-55.04)
		E	0.0000 (+74.01)	0.0000 (+42.05)	0.0000 (+62.52)
	D	E	0.0000 (+84.87)	0.0000 (+55.4)	0.0000 (+75.82)

(-/+) sign represent decreasing/increasing in MAE compare to reference method (better by %)
A = Fixed kNN + MeanShift, B = Adaptive kNN + MeanShift, C = Adaptive kNN + KDE-1D, D = kNN Mean, E = kNN Median

4.1.2. Heteroscedastic Noise Scenario

In the heteroscedastic setting, adaptive neighborhood selection has a more noticeable effect. Even for larger samples fixed-*k* modal regression (Method A) produces MAE values above 0.17 suggesting limited responsiveness to changes in local variance. On the other hand, MAE consistently decreases with Adaptive kNN + MeanShift (Method B) from 0.1936 ± 0.0258 at n=500 to 0.1920 ± 0.0096 at n=2000. When evaluated using the "Better By (%)" metric relative to the fixed-*k* baseline, this corresponds to substantial percentage improvements, indicating that adaptive neighborhood selection yields not only statistically significant but also practically meaningful error reductions. Although it has a slightly higher dispersion, Adaptive kNN + KDE-1D (Method C) likewise outperforms fixed-k techniques and shows consistent relative gains across sample sizes.

As reported in table 2, the Wilcoxon test results verify that these decreases are statistically significant in the majority of pairwise comparisons. Importantly, the statistical significance is accompanied by non-negligible percentage improvements in MAE, reinforcing that the observed differences are not merely artifacts of sample variability. These results suggest that the estimator can more successfully adapt to input-dependent noise variance when entropy-guided neighborhood adaptation is used, particularly under heteroscedastic conditions where local variance structure plays a critical role.

Across the heteroscedastic scenario, the method B shows inconsistent and mostly insignificant results, with a small improvement at $n = 500$ but performance declines at larger sample sizes. In contrast, methods C, D, and E consistently achieve statistically significant improvements ($p = 0.0000$). Method C provides moderate gains that decrease as sample size increases, while methods D and E deliver the largest and most stable improvements, with D performing best overall.

4.1.3. Bimodal Conditional Distribution Scenario

The bimodal setting is the most challenging scenario considered in this study. The mean- and median-based kNN estimators (Methods D and E) produce substantially larger errors in multiple configurations, with MAE values surpassing 1.4. This behavior reflects predictions that tend to fall between dominant modes rather than in high-density areas of the conditional distribution. Specifically, MAE values of 0.8576 ± 0.1878 at $n = 500$ and 1.0305 ± 0.1168 at $n = 2000$ are achieved by Adaptive kNN + KDE-1D (Method C). When expressed in terms of the "Better By (%)" metric relative to mean-based kNN, these reductions correspond to consistent percentage improvements, although the overall error levels remain relatively large and variability persists across repetitions.

The results of the Wilcoxon test show statistically significant differences favoring modal estimators for the majority of sample sizes. However, the magnitude of improvement varies and should be interpreted in the context of the intrinsic difficulty of bimodal regression. Rather than completely resolving multimodality, adaptive modal regression appears to lessen systematic bias toward low-density regions, highlighting both its strengths and its remaining limitations in complex distributional settings.

Statistical significance (Wilcoxon signed-rank test, table 2) is interpreted as evidence of consistent performance differences across Monte Carlo repetitions. However, practical relevance is assessed separately based on the magnitude of MAE improvement. Small but statistically significant differences are not interpreted as substantively meaningful unless the improvement exceeds a practically interpretable threshold.

To evaluate consistency across Monte Carlo repetitions, pairwise MAE differences were analyzed using the Wilcoxon signed-rank test. This non-parametric procedure assesses whether the median paired difference differs from zero and is appropriate given the absence of normality assumptions. Statistical significance therefore reflects consistent directional improvement rather than large absolute differences. In addition, practical relevance is assessed through percentage improvement ("Better By (%)") relative to the reference method. While statistically significant differences indicate reliability, substantive conclusions are drawn only when accompanied by meaningful MAE reductions.

The visual patterns observed in the prediction curves are consistent with the quantitative performance metrics reported in table 3 and table 4. In particular, configurations where adaptive modal regression more closely follows dominant conditional modes correspond to lower MAE values and positive "Better By (%)" improvements relative to mean-based kNN. Conversely, scenarios in which prediction curves exhibit oscillation or partial mode switching are reflected in increased MAE and larger standard deviations. This alignment between graphical diagnostics and numerical evaluation supports the robustness of the reported improvements and reduces reliance on purely qualitative interpretation.

When evaluated relative to the baseline using the "Better By (%)" metric, the adaptive methods demonstrate consistent percentage improvements in several heavy-tailed settings, although the magnitude of improvement varies with sample size. In particular, the percentage reduction in MAE indicates that entropy-guided adaptive neighborhood selection yields measurable gains over fixed-k approaches, even when absolute differences appear numerically moderate. The adaptive and fixed-k approaches differ statistically significantly in the majority of comparisons according to the Wilcoxon test results in table 3. Importantly, statistical significance is interpreted jointly with the magnitude of improvement, ensuring that small but statistically detectable differences are not overstated as practically substantial. This implies that, without completely removing the variability present in heavy-tailed data, entropy-guided neighborhood selection lessens sensitivity to extreme observations while maintaining consistent relative performance gains.

4.2. Real-World Datasets

Table 3 reports the performance on the HAR, Housing, and Hour datasets, which differ in dimensionality, scale, and application domain. These datasets provide a practical assessment of robustness beyond controlled simulations.

4.2.1. HAR Dataset

Due to the HAR dataset is structured, all approaches produce comparatively low MAE values. Slightly reduced errors are consistently produced by adaptive modal regression techniques. Adaptive kNN + KDE-1D (Method C) achieves an MAE of 0. 0353±0. 0017 at n=2000 while the kNN mean and median achieve 0.0378±0.0016 and 0.0407±0.0016 respectively. In relative terms, the "Better By (%)" metric indicates modest but consistent percentage improvements of the adaptive approach over mean- and median-based estimators.

Table 3. The Quantitative Performance of the Evaluated Methods in Terms of MAE and RMSE for Real World Datasets (*n_reps*=20)

Dataset	Method	*n_samp*	MAE (mean ± standard deviation)	RMSE (mean ± standard deviation)	Mean Pred. Time n(s)
Har (*n*=10,299, *n_feat*=561	A	500	0.0638 ± 0.0074	0.1252 ± 0.0276	3.6017
		1000	0.0654 ± 0.0075	0.1349 ± 0.0263	10.7961
		2000	0.0451 ± 0.0020	0.0804 ± 0.0055	14.0507
		full	0.0315 ± 0	0.0595 ± 0	46.5200
	B	500	0.0543 ± 0.0068	0.1101 ± 0.0316	1.1637
		1000	0.0475 ± 0.0056	0.0968 ± 0.0232	2.2350
		2000	0.0374 ± 0.0020	0.0652 ± 0.0048	4.4230
		full	0.0266 ± 0	0.0478 ± 0	15.2690
	C	500	0.0509 ± 0.0063	0.1060 ± 0.0325	0.6453
		1000	0.0465 ± 0.0056	0.0949 ± 0.0235	1.2336
		2000	0.0353 ± 0.0017	0.0621 ± 0.0052	2.3579
		full	0.0257 ± 0	0.0476 ± 0	7.3891
	D	500	0.0520 ± 0.0060	0.1024 ± 0.0268	0.2547
		1000	0.0541 ± 0.0061	0.1098 ± 0.0225	0.4694
		2000	0.0378 ± 0.0016	0.0660 ± 0.0046	0.8727
		full	0.0290 ± 0	0.0519 ± 0	1.7949
	E	500	0.0577 ± 0.0065	0.1106 ± 0.0295	0.2594
		1000	0.0601 ± 0.0065	0.1200 ± 0.0229	0.4835
		2000	0.0407 ± 0.0016	0.0705 ± 0.0049	0.9220
		Full	0.0303 ± 0	0.0541 ± 0	1.8995
Housing (*n*=20640, *n_feat*=10)	A	500	72433.6150 ± 7284.8710	101485.7863 ± 11686.7228	3.2399
		1000	68980.2594 ± 5105.3766	100355.5764 ± 7154.2468	10.2896
		2000	58103.3384 ± 2621.7892	87033.9873 ± 4272.9861	12.9886
		Full	45143.5087 ± 0	71430.1558 ± 0	134.1720
	B	500	63687.7779 ± 6214.0543	88977.5863 ± 8681.7190	0.8886
		1000	60808.1628 ± 4722.9425	89178.5105 ± 5948.9390	1.7559
		2000	54463.7192 ± 3599.6852	80069.0493 ± 5789.4955	3.5908
		Full	44252.1924 ± 0	67837.8762 ± 0	37.4359
	C	500	61439.7975 ± 5606.1338	86606.7940 ± 8286.7401	0.4093
		1000	58059.9331 ± 4726.3902	86413.8823 ± 6173.8185	0.8037
		2000	52185.3409 ± 3052.5009	77497.6259 ± 4973.9848	1.6206
		Full	41945.2574 ± 0	65441.0438 ± 0	16.7578
	D	500	63338.7530 ± 5974.6456	85695.6557 ± 8466.2821	0.0499
		1000	60251.5325 ± 4074.2369	83284.5393 ± 5069.4842	0.1032
		2000	52172.9810 ± 1889.6269	72770.5369 ± 3314.7105	0.2138
		Full	41016.3551 ± 0	61195.0383 ± 0	2.7192

	E	500	63510.8338 ± 6345.7261	88913.0237 ± 9414.4902	0.0540
		1000	60855.5191 ± 4494.6862	87424.9429 ± 5765.5804	0.1132
		2000	51926.8329 ± 2229.5576	76438.1031 ± 4185.7571	0.2325
		Full	40455.1435 ± 0	63898.4449 ± 0	2.8991
Hour (n=17379, n_feat=17)	A	500	70.0199 ± 6.9985	106.0634 ± 13.1530	3.0492
		1000	67.0893 ± 4.0147	97.5306 ± 6.3137	9.6087
		2000	57.4117 ± 3.0009	82.1121 ± 4.5779	12.4259
		Full	0.2615 ± 0	55.9843 ± 0	109.1316
	B	500	56.8481 ± 3.7425	77.7747 ± 5.7828	0.8954
		1000	53.9755 ± 3.3191	74.6191 ± 5.0394	1.7932
		2000	48.4183 ± 2.7589	66.1143 ± 3.8174	3.5975
		Full	34.0633 ± 0	46.8228 ± 0	30.4139
	C	500	54.2913 ± 4.1981	75.4098 ± 6.6027	0.3966
		1000	51.2886 ± 3.5247	71.6242 ± 5.6465	0.7907
		2000	45.3016 ± 2.2961	62.2355 ± 3.0887	1.5758
		Full	31.5399 ± 0	3.5567 ± 0	14.0685
	D	500	59.9781 ± 5.8522	80.2697 ± 10.4795	0.0508
		1000	54.4252 ± 3.1269	71.6420 ± 4.5767	0.1057
		2000	45.4603 ± 1.5868	58.8248 ± 2.8718	0.2196
		Full	29.2264 ± 0	37.6814 ± 0	2.6746
	E	500	58.2743 ± 6.3982	85.1763 ± 11.7260	0.0547
		1000	53.5016 ± 3.7308	76.2806 ± 5.5694	0.1156
		2000	45.8861 ± 2.0934	63.6211 ± 3.4291	0.2380
		Full	30.5456 ± 0	41.4899 ± 0	2.8253

A = Fixed kNN + MeanShift, B = Adaptive kNN + MeanShift, C = Adaptive kNN + KDE-1, D = kNN Mean, E = kNN Median

Wilcoxon test results in table 4 show statistical significance in the majority of comparisons despite the small numerical differences. However, these results are interpreted as incremental rather than transformative gains, reflecting steady improvements in predictive accuracy without large practical deviations.

Table 4. Wilcoxon Test-MAE for each scenario for Real Datasets (reps=20, α = 0.05)

Datasets	Between Methods		p-Value (better by %)	p-Value (better by %)	p-Value (better by %)
	Ref	Comp	(n_sub=500)	(n_sub=1000)	(n_sub =2000)
Har (n=10,299, n_feat=561)	A	B	0.0000 (-14.84)	0.0000 (-27.43)	0.0000 (-17.16)
		C	0.0000 (-20.29)	0.0000 (-28.93)	0.0000 (-21.75)
		D	0.0000 (-18.56)	0.0000 (-17.22)	0.0000 (-16.25)
		E	0.0000 (-9.51)	0.0000 (-8.12)	0.0000 (-9.85)
	B	C	0.0000 (-6.39)	0.0017 (-2.07)	0.0000 (-5.55)
		D	0.0192 (-4.37)	0.0000 (+12.34)	0.3488 (+1.08)
		E	0.0037 (+5.89)	0.0000 (+21.02)	0.0000 (+8.10)
	C	D	0.2162 (+2.12)	0.0000 (+14.15)	0.0000 (+6.57)
		E	0.0000 (+11.91)	0.0000 (+22.65)	0.0000 (+13.20)
	D	E	0.0000 (+10.00)	0.0000 (+9.90)	0.0000 (+7.09)
Housing (n= 20640, n_feat=10)	A	B	0.0000 (-12.07)	0.0000 (-11.85)	0.0001 (-6.26)
		C	0.0000 (-15.18)	0.0000 (-15.83)	0.0000 (-10.19)
		D	0.0000 (-12.56)	0.0000 (-12.65)	0.0000 (-10.21)
		E	0.0000 (-12.32)	0.0000 (-11.78)	0.0000 (-10.63)
	B	C	0.0003 (-3.53)	0.0000 (-4.52)	0.0000 (-4.18)
		D	0.6477 (-0.55)	0.8408 (+0.92)	0.0017 (-4.21)
		E	0.9563 (-0.28)	0.6477 (+0.08)	0.0005 (-4.66)
	C	D	0.0266 (+3.00)	0.0083 (+3.64)	0.9854 (-0.02)
		E	0.0027 (+3.26)	0.0006 (+4.59)	0.7012 (-0.47)
	D	E	0.5958 (+0.27)	0.1054 (+0.99)	0.1893(-0.47)
Hour (n=17379, n_feat=17)	A	B	0.0000 (-18.81)	0.0000 (-19.55)	0.0000 (-15.66)
		C	0.0000 (-22.46)	0.0000 (-23.55)	0.0000 (-21.09)
		D	0.0000 (-14.34)	0.0000 (-18.88)	0.0000 (-20.82)
		E	0.0000 (-16.77)	0.0000 (-20.25)	0.0000 (-20.08)
	B	C	0.0000 (-4.50)	0.0000 (-4.98)	0.0000 (-6.44)
		D	0.0532 (+5.22)	0.2611 (+0.83)	0.0000 (-6.11)

	E	0.4304 (+2.45)	0.7562 (-0.88)	0.0001 (-5.23)
C	D	0.0001 (+9.48)	0.0014 (+5.76)	0.7285 (+0.35)
	E	0.0056 (+6.83)	0.0192 (+4.14)	0.1650 (+1.27)
D	E	0.0003 (-2.84)	0.0153 (-1.70)	0.0215 (+0.93)

(-/+) sign represent decreasing/increasing in MAE compare to reference method (better by %)
A = Fixed kNN + MeanShift, B = Adaptive kNN + MeanShift, C = Adaptive kNN + KDE-1D, D = kNN Mean, E = kNN Median

4.2.2. Housing Dataset

For the Housing dataset, performance differences between methods are more nuanced. Mean-based kNN remains competitive, particularly at larger sample sizes. Nevertheless, Adaptive kNN + KDE-1D achieves the lowest or near-lowest MAE across configurations. At the full dataset size, Method C records an MAE of approximately 41,945, compared to 45,143 for fixed-*k* MeanShift and 41.016 for kNN mean. The corresponding percentage differences indicate that adaptive modal regression provides competitive performance, although the magnitude of improvement over strong baselines such as kNN mean is relatively small. Wilcoxon test outcomes suggest that some of these differences are statistically significant, but the practical advantage depends on the specific evaluation configuration. These results suggest that adaptive modal regression maintains competitive accuracy while offering increased robustness to local irregularities in the data.

4.2.3. Hour Dataset

On the Hour dataset, adaptive modal regression methods again demonstrate lower average errors. At n=2000, Adaptive kNN + KDE-1D attains an MAE of 45.30±2.30, slightly lower than those of kNN mean and median. The "Better By (%)" analysis indicates consistent, though moderate, percentage improvements across subsample sizes. Wilcoxon test results confirm that these differences are statistically significant for most subsample sizes, indicating stable performance patterns across repetitions. As in the HAR case, the improvements should be interpreted as consistent and robust rather than large in absolute magnitude.

4.3. Computational Cost Analysis

All experiments were implemented in Python using NumPy, SciPy, and scikit-learn libraries. Runtime measurements were conducted on a machine equipped with an Intel Core i7 CPU and 16 GB RAM under a standard desktop operating system. Nearest-neighbor searches were performed using exact search procedures provided by the underlying library implementations, without approximate acceleration techniques. Reported runtime values therefore reflect relative computational differences under a consistent environment and may vary across hardware configurations. In addition to predictive accuracy, table 1 and table 2 report the average prediction time for each method. As expected, adaptive modal regression methods incur higher computational cost than classical kNN mean and median estimators due to entropy-based neighborhood selection and local density estimation.

The average prediction time for Adaptive kNN + KDE-1D on simulated datasets rises from roughly 0.56 seconds at n=500 to 2.14 seconds at n=2000. Over the same sample sizes adaptive kNN + MeanShift takes between 1.74 and 7.03 seconds to compute. On the other hand, because of their simpler computational structure, kNN mean and median estimators consistently take less than 0.25 seconds even for larger samples. Similar trends are observed on real-world datasets. For example, on the HAR dataset, Adaptive kNN + KDE-1D requires approximately 7.39 seconds at full sample size, compared to under 2 seconds for kNN mean and median. While adaptive methods introduce additional overhead, the increase remains proportional to sample size and does not exhibit abrupt growth, suggesting that the computational cost remains manageable for medium-scale applications.

4.4. Visual Diagnostics: Predicted-versus-Actual Plots

For the HAR dataset, the relationship actual and predicted value is visually illustrated in figure 2 to figure 6. Figure 2 depicts the prediction results obtained from Fixed kNN combined with MeanShift. Figure 3 illustrates the predictions produced by Adaptive kNN with MeanShift, while Figure 4 presents the results generated by Adaptive kNN with KDE-1D. In comparison, Figures 5 and figure 6 show the prediction curves derived from the conventional kNN mean and kNN median, respectively. Figure 3 and figure 4 which depict Adaptive kNN + MeanShift and Adaptive kNN + KDE-1D respectively exhibit a more concentrated set of points surrounding the identity line. Particularly in areas with greater heterogeneity the spread has decreased. Among these, Adaptive kNN + KDE-1D displays the smoothest alignment, indicating that density-based mode estimation contributes to more stable local predictions when combined with adaptive neighborhood selection.

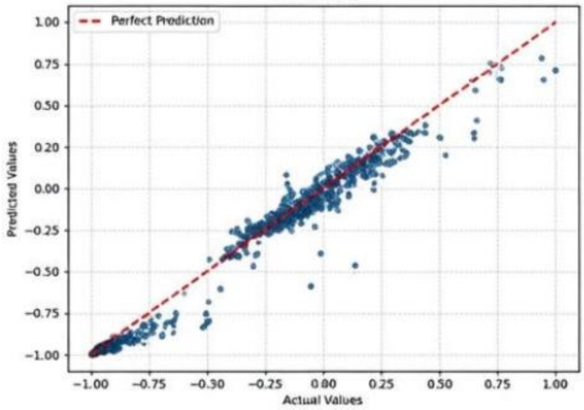

Figure 2. Actual Values vs Predicted Values Using Fix kNN+MeanShift for HAR Data

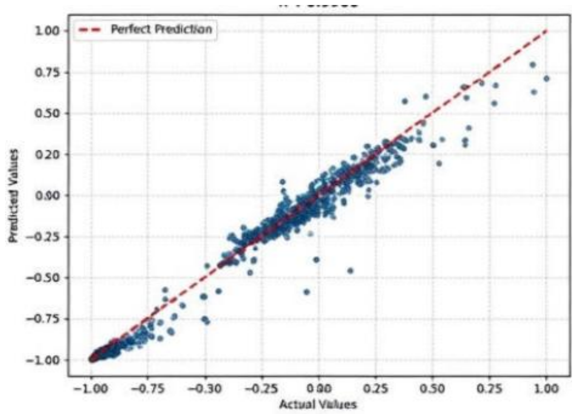

Figure 3. Actual Values vs Predicted Values Using Adaptive kNN+MeanShift for HAR Data

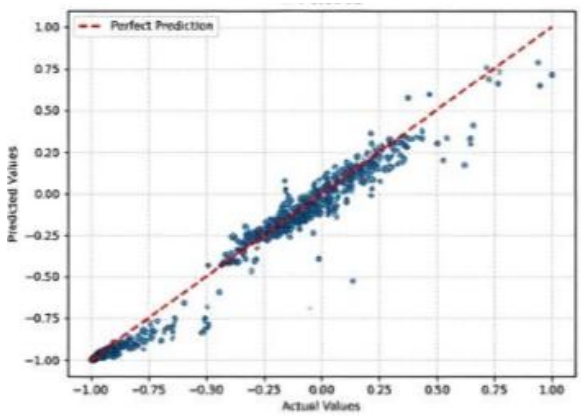

Figure 4. Actual Values vs Predicted Values Using Adaptive kNN+KDE-1D for HAR Data

In contrast, figure 5 and figure 6 corresponding to kNN mean and median reveal systematic deviations from the identity line. While these methods produce smooth predictions, they tend to emphasize central tendency and underrepresent dominant local modes, particularly in regions influenced by multimodal or skewed response distributions.

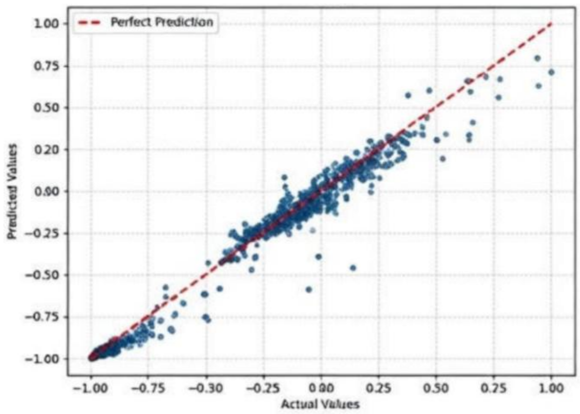

Figure 5. Actual Values vs Predicted Values Using kNN Mean for HAR Data

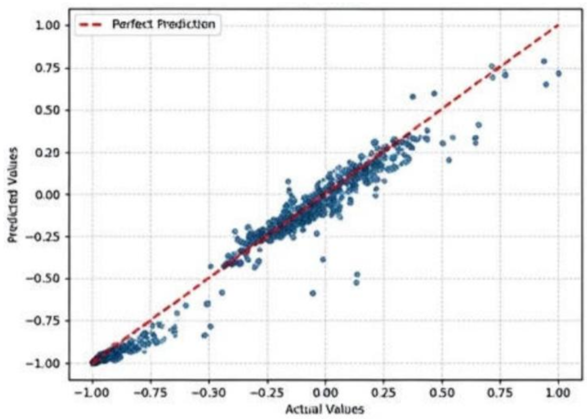

Figure 6. Actual Values vs Predicted Values Using kNN Median for HAR Data

Taken together, the numerical results, statistical tests, computational analysis, and visual diagnostics suggest that entropy-guided adaptive modal regression provides a flexible and robust alternative to fixed-*k* and mean-based kNN regression. The advantages of adaptation are most apparent under heteroscedastic and multimodal conditions, while performance remains comparable in more regular settings. The additional computational cost is predictable and moderate, supporting the practical applicability of the proposed framework in applied data science contexts. These results highlight a practical trade-off between predictive robustness and computational cost.

ISSN 2723-6471

5. Conclusion

This study proposed an entropy-guided adaptive k-nearest neighbor framework for modal regression, motivated by regression problems characterized by heteroscedasticity, multimodality, and heavy-tailed noise. By shifting the learning objective from central tendency estimation toward conditional mode estimation, the proposed approach aims to recover more representative outcomes in settings where mean- or median-based regression may be misaligned with dominant data patterns. The integration of entropy-based neighborhood adaptation with nonparametric mode-seeking methods such as MeanShift clustering and one-dimensional kernel density estimation constitutes a central contribution of this work. Without relying on globally fixed parameters, the estimator adapts to local variations in data density and noise structure through a data-driven entropy criterion for determining neighborhood size. Empirical results on simulated and real-world datasets indicate that this adaptive strategy yields consistent improvements or stabilization of prediction errors, particularly under heteroscedastic and multimodal conditions. In challenging bimodal settings, the method reduces systematic bias toward low-density regions, although it does not fully eliminate the complexity induced by multimodality.

At the same time, the findings highlight an inherent trade-off between predictive robustness and computational cost. While adaptive modal regression frequently achieves lower mean absolute error relative to fixed-k and mean-based kNN estimators, the magnitude of improvement varies across datasets and scenarios. The statistical significance observed in many comparisons is accompanied by percentage-based performance gains, though in structured real-world datasets these gains are sometimes incremental rather than substantial. Runtime analysis confirms that entropy evaluation and local density estimation introduce additional computational overhead. This overhead increases proportionally with sample size and remains manageable for medium-scale datasets, but it is non-negligible compared to classical kNN predictors.

Overall, the results suggest that entropy-guided adaptive modal regression provides a flexible alternative to mean-centric instance-based regression when robustness to heterogeneous noise structures is a primary objective. Future work may focus on computational optimization, exploration of alternative information-theoretic criteria beyond Shannon entropy, and extensions to high-dimensional or large-scale learning environments, where both entropy estimation and nearest-neighbor search may present additional challenges.

Despite these strengths, the results also indicate that adaptive modal regression does not uniformly dominate simpler baselines. In highly challenging bimodal scenarios, although systematic bias toward low-density regions is reduced, overall MAE values remain relatively large and variability persists. Similarly, in certain real-world datasets such as Housing, mean-based kNN estimators remain competitive, particularly at larger sample sizes. These findings suggest that the benefits of entropy-guided neighborhood adaptation depend on the underlying data structure and may be less pronounced when conditional distributions are close to unimodal or when computational simplicity is prioritized. Such limitations highlight the importance of selecting regression strategies in accordance with distributional complexity and practical constraints.

The findings of this study have several practical implications for applied data science, particularly in regression tasks involving heterogeneous noise structures, multimodal conditional distributions, and local irregularities that are not adequately captured by mean- or median-based estimators. First, the proposed adaptive modal regression framework is well suited for applications where the most representative or dominant outcome is of primary interest, rather than the expected value. Such scenarios commonly arise in ris*k*-sensitive decision-making, demand forecasting under volatile conditions, sensor-based systems with intermittent disturbances, and behavioral or environmental data exhibiting multimodal responses. In these contexts, mean-based regression may produce predictions that fall in low-density regions, whereas modal regression can offer outcomes that better reflect prevailing data patterns.

Second, the entropy-guided neighborhood adaptation mechanism provides a principled way to address local heterogeneity without requiring manual tuning of neighborhood size through entropy-guided local uncertainty assessment. This feature is particularly relevant in applied settings where data density varies substantially across the input space and where a globally fixed neighborhood parameter may lead to suboptimal performance. By adjusting neighborhood size based on local uncertainty, the proposed framework allows practitioners to deploy instance-based regression with greater robustness across diverse data regimes.

Third, the findings demonstrate that in the larger context of improved k-nearest neighbors techniques adaptive modal regression holds a complementary role. Recent advances like random kernel k-nearest neighbors' regression [26] shows that kernel weighting random feature subsetting and ensembling can enhance scalability and generalization in large-scale settings. Adaptive modal regression concentrates on maintaining local response structure through mode estimation whereas such methods mainly aim to reduce variance and achieve global accuracy. In actuality the relative significance of computational constraints predictability interpretability and robustness to multimodality determines which of these approaches is best. Finally, the moderate and predictable computational overhead observed in the experiments suggests that adaptive modal regression is feasible for small- to medium-sized datasets commonly encountered in applied data science workflows. When robustness and representativeness of predictions are prioritized over minimal runtime, the proposed framework offers a viable and interpretable alternative to classical kNN regression.

Despite its demonstrated robustness and interpretability, the proposed adaptive modal regression framework has several limitations that motivate future research. First, the entropy-based neighborhood selection and local density estimation introduce additional computational overhead compared to classical kNN mean or median predictors. While empirical runtime analysis indicates that this overhead remains moderate for medium-sized datasets, scalability to very large datasets or real-time applications may require further optimization. Promising directions include approximate nearest neighbor search, parallel or distributed implementations, and GPU-accelerated density estimation techniques.

Second, the current study primarily focuses on low- to moderate-dimensional feature spaces to facilitate clear interpretation of local neighborhood behavior and conditional mode estimation. In high-dimensional settings, distance-based methods are known to suffer from the curse of dimensionality, which may degrade both neighborhood selection and entropy estimation. Future work could explore the integration of dimensionality reduction techniques, adaptive distance metrics, or feature weighting schemes to enhance performance in high-dimensional contexts.

Third, while Shannon entropy over distance-induced probability distributions proved effective as a criterion for adaptive neighborhood selection, it represents only one possible measure of local complexity. Alternative information-theoretic quantities, such as Rényi entropy, mutual information, or density-based uncertainty measures, may yield different adaptive behaviors and merit systematic investigation. Comparing these criteria could further clarify the relationship between local data structure and optimal neighborhood adaptation.

Fourth, in addition to the general challenges faced by k-nearest neighbor methods in high-dimensional spaces, the entropy estimation procedure itself may degrade as dimensionality increases. Distance concentration and data sparsity can reduce contrast between neighboring points, leading to less reliable local probability distributions and consequently less stable entropy estimates. As a result, the neighborhood adaptation mechanism may become less sensitive to meaningful structural variations in very high-dimensional settings. This limitation suggests that entropy-guided modal regression may benefit from preliminary dimensionality reduction, feature selection, or alternative local information measures when applied to large-scale or high-dimensional data.

Fifth, although the small constant ε is introduced solely to prevent division by zero in the distance-based probability normalization and was fixed throughout the experiments, future research may investigate adaptive or data-driven strategies for selecting ε. In particular, exploring its interaction with neighborhood size and local density structure could provide additional theoretical insight into the stability of entropy-based neighborhood selection. A more formal sensitivity analysis may also help clarify the robustness of the proposed framework under varying numerical regularization settings.

Finally, the present framework is restricted to regression problems with continuous responses. Extending adaptive modal learning to other learning paradigms—such as classification via modal decision rules, structured prediction, or functional and longitudinal data analysis—represents a promising avenue for future research. Establishing formal theoretical guarantees on consistency, convergence rates, and robustness properties of entropy-guided adaptive modal regression also remains an important open problem.

6. Declarations

6.1. Author Contributions

Conceptualization: S., N.H., and A.A.N.; Methodology: N.H.; Software: S.; Validation: S., N.H., and A.A.N.; Formal Analysis: S., N.H., and A.A.N.; Investigation: S.; Resources: N.H.; Data Curation: N.H.; Writing Original Draft Preparation: S., N.H., and A.A.N.; Writing Review and Editing: N.H., S., and A.A.N.; Visualization: S.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019, doi: 10.1109/ACCESS.2019.2912200.

[2] Z.-H. Zhou, "Machine learning," *Springer*, vol. 2021, no. Jan., pp. 1–20, 2021, doi: 10.1007/978-981-15-1967-3.

[3] C. Amado, A. M. Bianco, G. Boente, and I. M. Rodrigues, "Robust estimation of heteroscedastic regression models: a brief overview and new proposals," *arXiv*, vol. 2023, no. Nov., pp. 1–12, 2023.

[4] Y. Chen, "Modal regression using kernel density estimation: a review," *WIREs Comput. Stat.*, vol. 10, no. 4, pp. 1–12, 2018, doi: 10.1002/wics.1431.

[5] S. Xiang, W. Yao, and X. Cui, "Advances in modal regression: from theoretical foundations to practical implementations," *WIREs Comput. Stat.*, vol. 18, no. 1, pp. 1–12, 2026, doi: 10.1002/wics.70061.

[6] Y.-C. Chen, C. R. Genovese, R. J. Tibshirani, and L. Wasserman, "Nonparametric modal regression," *Ann. Stat.*, vol. 44, no. 2, pp. 1–12, 2016, doi: 10.1214/15-AOS1373.

[7] A. Ullah, T. Wang, and W. Yao, "Nonlinear modal regression for dependent data with application for predicting COVID-19," *J. R. Stat. Soc. Ser. A*, vol. 185, no. 3, pp. 1424–1453, 2022, doi: 10.1111/rssa.12849.

[8] A. Pérez-González, T. R. Cotos-Yáñez, and R. M. Crujeiras, "Nonparametric modal regression with missing response observations," *Comput. Stat.*, vol. 41, no. 3, pp. 1–12, 2026, doi: 10.1007/s00180-026-01738-2.

[9] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for kNN classification," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, pp. 1–19, 2017, doi: 10.1145/2990508.

[10] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002, doi: 10.1109/34.1000236.

[11] Z. Pan, Y. Wang, and Y. Pan, "A new locally adaptive k-nearest neighbor algorithm based on discrimination class," *Knowl. Based Syst.*, vol. 204, no. Sep., pp. 1–12, 2020, doi: 10.1016/j.knosys.2020.106185.

[12] P. Zhao and L. Lai, "Efficient classification with adaptive kNN," *IEEE Access*, vol. 2021, no. May, pp. 11007–11014, 2021, doi: 10.1609/aaai.v35i12.17314.

[13] C. Kaleli, "An entropy-based neighbor selection approach for collaborative filtering," *Knowl. Based Syst.*, vol. 56, no. Jan., pp. 273–280, 2014, doi: 10.1016/j.knosys.2013.11.020.

[14] T. Kneib, A. Silbersdorff, and B. Säfken, "Rage against the mean – a review of distributional regression approaches," *Econom. Stat.*, vol. 26, no. Apr., pp. 99–123, 2023, doi: 10.1016/j.ecosta.2021.07.006.

[15] M. Matabuena, J. C. Vidal, O. H. M. Padilla, and J.-P. Onnela, "Conditional mean and variance estimation via k-NN algorithm with automated variance selection," *arXiv*, vol. 2025, no. Oct., pp. 1–12, 2025.

[16] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.

[17] T. S. Breusch and A. R. Pagan, "A simple test for heteroscedasticity and random coefficient variation," *Econometrica*, vol. 47, no. 5, pp. 1287–1294, 1979, doi: 10.2307/1911963.

[18] H. Zhou and X. Huang, "Nonparametric modal regression in the presence of measurement error," *Electron. J. Stat.*, vol. 10, no. 2, pp. 1–12, 2016, doi: 10.1214/16-EJS1210.

[19] P. J. Huber and E. M. Ronchetti, "Robust statistics," *Wiley*, vol. 2009, no. Jan., pp. 1–20, 2009, doi: 10.1002/9780470434697.

[20] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the t distribution," *J. Am. Stat. Assoc.*, vol. 84, no. 408, pp. 881–896, 1989, doi: 10.1080/01621459.1989.10478852.

[21] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978, doi: 10.2307/1913643.

[22] W. Yao and L. Li, "A new regression model: modal linear regression," *Scand. J. Stat.*, vol. 41, no. 3, pp. 656–671, 2014, doi: 10.1111/sjos.12054.

[23] M. Lee, "Mode regression," *J. Econom.*, vol. 42, no. 3, pp. 337–349, 1989, doi: 10.1016/0304-4076(89)90057-2.

[24] B. W. Silverman, "Using kernel density estimates to investigate multimodality," *J. R. Stat. Soc. Ser. B*, vol. 43, no. 1, pp. 97–99, 1981, doi: 10.1111/j.2517-6161.1981.tb01155.x.

[25] H. Zhou and X. Huang, "Bandwidth selection for nonparametric modal regression," *Commun. Stat. Simul. Comput.*, vol. 48, no. 4, pp. 968–984, 2019, doi: 10.1080/03610918.2017.1402044.

[26] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, 1995, doi: 10.1109/34.400568.

[27] T. M. Cover and J. A. Thomas, "Elements of information theory," *Wiley*, vol. 2005, no. Jan., pp. 1–20, 2005, doi: 10.1002/047174882X.

[28] S. Bermejo and J. Cabestany, "Adaptive soft k-nearest-neighbour classifiers," *Pattern Recognit.*, vol. 33, no. 12, pp. 1999–2005, 2000, doi: 10.1016/S0031-3203(99)00186-7.

[29] T. Hastie and R. Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 6, pp. 607–616, 1996, doi: 10.1109/34.506411.

[30] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, pp. 066138–066145, 2004, doi: 10.1103/PhysRevE.69.066138.

[31] A. R. Conn, K. Scheinberg, and L. N. Vicente, "Introduction to derivative-free optimization," *SIAM*, vol. 2009, no. Jan., pp. 1–20, 2009, doi: 10.1137/1.9780898718768.