Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 11s (2025)
https://www.jneonatalsurg.com

Advanced Quantile Regression with Pinball Loss: Leveraging Lagrangian Asymmetric-vTwin SVR and Enhanced Model Optimization for Superior Performance

V Rajanikanth Tatiraju[1], Dr. Rohita Yamaganti[2]

[1]Ph.D., Research Scholar, Department of Computer Science Engineering, P.K.University, Shivpuri, Madhya Pradesh, India–473665.

Email ID: tvrajani55@gmail.com

[2]Associate Professor, Department of Computer Science Engineering, P. K. University, Shivpuri, Madhya Pradesh, India–473665.

Email ID: rohita.yamaganti@gmail.com

Cite this paper as: V Rajanikanth Tatiraju, Dr. Rohita Yamaganti, (2025) Advanced Quantile Regression with Pinball Loss: Leveraging Lagrangian Asymmetric-vTwin SVR and Enhanced Model Optimization for Superior Performance. *Journal of Neonatal Surgery*, 14 (11s), 532-544

ABSTRACT

This study presents a comparative analysis of three regression models—Lagrangian Asymmetric-vTwin Support Vector Regression (SVR), Standard SVR, and Linear Regression—focusing on their performance in quantile prediction using Pinball Loss. The models are evaluated at different quantiles ($\alpha = 0.1$, 0.5, and 0.9) and conventional metrics, such as RMSE and MAE. The results reveal that the Lagrangian Asymmetric-vTwin SVR consistently outperforms the other models, providing the lowest Pinball Loss values across all quantiles. Specifically, the Lagrangian Asymmetric-vTwin SVR achieves a Pinball Loss of 0.045 at $\alpha = 0.1$, 0.029 at $\alpha = 0.5$, and 0.038 at $\alpha = 0.9$. In comparison, the Standard SVR shows Pinball Loss values of 0.062, 0.038, and 0.045 for the same quantiles, while Linear Regression yields Pinball Loss values of 0.089, 0.076, and 0.082. In addition to Pinball Loss, the Lagrangian Asymmetric-vTwin SVR also performs better in RMSE and MAE, with values of 0.12 and 0.10, respectively, compared to Standard SVR's 0.18 and 0.14, and Linear Regression's 0.22 and 0.19. Furthermore, the optimal regularization parameter (C) of 1.0 for the Lagrangian Asymmetric-vTwin SVR strikes a balance between model complexity and prediction accuracy, leading to improved training efficiency and faster convergence. These results demonstrate the superior capability of the Lagrangian Asymmetric-vTwin SVR in quantile regression tasks.

Keywords: *Pinball, Quantile, Lagrangian, Asymmetric, Performance, Regression, SVR.*

1. INTRODUCTION

Quantile regression has emerged as a powerful tool in statistical modeling, providing a more comprehensive understanding of data distributions by estimating conditional quantiles instead of only the conditional mean. This approach is particularly useful in scenarios where the conditional distribution exhibits skewness or outliers, which may not be captured by traditional methods like ordinary least squares regression [1]. The ability to predict different quantiles—such as the lower, median, and upper quantiles—offers insights into the tail behavior of the distribution and improves the robustness of predictions. This is particularly important in real-world applications such as finance, healthcare, and environmental science, where understanding the extremes of a distribution is crucial for decision-making (Chernozhukov & Hansen, 2005).

One of the challenges in quantile regression is the choice of the loss function used to assess the accuracy of predictions. Pinball loss, introduced by Koenker and Bassett (1978), [2] is a widely adopted method for this purpose, as it directly penalizes prediction errors based on the quantile being predicted. Unlike traditional loss functions like squared error, Pinball loss is asymmetric, which allows it to focus on the discrepancies at different parts of the distribution, depending on the chosen quantile. This makes it an ideal candidate for quantile regression, where different quantiles may exhibit varied behaviors (Koenker, 2005).

In this study, we compare three regression models—Lagrangian Asymmetric-vTwin Support Vector Regression (SVR), [3] Standard SVR, and Linear Regression—using Pinball loss to evaluate their performance in quantile prediction. Support Vector Regression (SVR) has gained prominence due to its ability to model non-linear relationships by mapping input data into a high-dimensional feature space, where linear regression techniques can then be applied (Vapnik, 1995). While SVR is well-known for its robustness to overfitting and ability to handle complex data distributions, its performance in quantile regression tasks has not been extensively compared with other models.

The Lagrangian Asymmetric-vTwin SVR, a variant of traditional SVR, has been proposed to address some of the shortcomings of standard SVR. This methodology incorporates Lagrangian multipliers to handle asymmetric data distributions more efficiently. It introduces the concept of vTwin optimization, which improves the model's sensitivity to different quantiles by adjusting the weights for different regions of the data (Zhang et al., 2020)[4] . Previous research has shown that incorporating asymmetric loss functions in SVR can lead to better performance when predicting quantiles, particularly in datasets with skewed distributions (Roth et al., 2016).

Linear Regression, though simple, remains a commonly used baseline for regression tasks due to its ease of implementation and interpretability. However, it often struggles to capture complex relationships in the data, especially in the presence of non-linearity or heteroscedasticity. Linear models also fail to account for the variability in the tail distribution, making them less effective when quantile predictions are the focus (Gelman et al., 2003)[5] . This is one of the reasons why more advanced models like SVR are often preferred for quantile regression tasks.

A key aspect of quantile regression is the selection of the regularization parameter, denoted as C in the case of SVR. The regularization parameter controls the trade-off between model complexity and the degree of error allowed. An appropriately chosen C value ensures that the model achieves a balance between overfitting and underfitting, leading to improved generalization on unseen data (Cortes & Vapnik, 1995)[6] . In this study, we explore how different C values influence the performance of the models, specifically looking for the optimal value that minimizes the error without sacrificing predictive accuracy.

One of the primary motivations for conducting this study is the growing importance of robust regression techniques in real-world applications. In fields such as finance, medicine, and meteorology, the ability to accurately predict the lower and upper quantiles of a distribution is crucial for making informed decisions. For instance, in financial risk management, accurately predicting the lower quantiles of asset returns can help in estimating Value-at-Risk (VaR) (McNeil et al., 2005) [7]. Similarly, in healthcare, understanding the upper quantiles of a biomarker's distribution can provide insights into the severity of a disease (Li et al., 2016). By comparing different regression models, this study seeks to identify the best-suited methodology for these types of applications.

The existing literature on quantile regression with SVR has mostly focused on the theoretical aspects and some isolated empirical applications (Bergstra et al., 2013) [8] . However, there is a gap in the comparative performance analysis of these models when applied to quantile prediction using Pinball loss. While previous studies have investigated the effectiveness of SVR for quantile regression (Chernozhukov et al., 2007), few have explored advanced SVR models like the Lagrangian Asymmetric-vTwin SVR in detail. This study contributes to filling this gap by providing a direct comparison of these models across different quantiles and evaluating their performance using both Pinball loss and traditional metrics such as RMSE and MAE.

[9] Our study aims to provide a detailed and comprehensive comparison of the three regression models in the context of quantile prediction using Pinball loss. By incorporating both traditional regression methods and more advanced SVR variants, this research helps to elucidate the strengths and weaknesses of each approach in handling asymmetry in the data and quantile-based predictions. The results will offer practical insights into the most suitable models for quantile regression tasks, especially for applications that require accurate predictions of both extreme lower and upper quantiles.

[10] In the following sections, we first provide a brief overview of the theory behind quantile regression and Pinball loss, followed by a detailed description of the three regression models under consideration. Next, we present the experimental setup, including the datasets and evaluation metrics used in our study. Finally, we analyze and discuss the results, highlighting the best-performing model for each quantile and providing recommendations for future research in this area. The goal is to advance the understanding of quantile regression techniques and to offer guidance on selecting the appropriate model for different applications based on the performance characteristics observed in this study..

literature survey

[11] Quantile regression has gained considerable attention due to its ability to estimate the conditional quantiles of a response variable, rather than just the conditional mean, offering a more comprehensive understanding of the distributional characteristics of the data (Koenker & Bassett, 1978). Traditional regression methods, such as Ordinary Least Squares (OLS), focus solely on predicting the mean of the response variable, which often leads to inefficient estimations in the presence of skewed distributions or outliers. By contrast, quantile regression can effectively model different parts of the conditional distribution, providing a more robust alternative for prediction in various fields such as finance, economics, and medical research (Chernozhukov & Hansen, 2005).

[12] In recent years, Support Vector Regression (SVR) has become a widely used method for quantile regression tasks due to its ability to handle non-linear relationships in data. SVR operates by mapping input data into a higher-dimensional feature space, where linear regression is applied, allowing it to capture complex relationships (Vapnik, 1995). Despite its versatility,

the standard SVR has limitations when it comes to modeling asymmetric or skewed distributions. To address this, several variants of SVR, including the Lagrangian Asymmetric-vTwin SVR, have been proposed to improve the performance of SVR for quantile regression tasks (Zhang et al., 2020).

[13] The Lagrangian Asymmetric-vTwin SVR introduces a novel approach by incorporating Lagrangian multipliers to handle asymmetric loss functions, making it more sensitive to the tails of the data distribution (Roth et al., 2016). This modification improves the model's ability to predict quantiles that are located at the lower or upper extremes of the distribution, which is particularly important in applications such as risk management, where the focus is often on predicting extreme values (McNeil et al., 2005). The vTwin optimization technique further enhances the performance by optimizing the weights associated with different regions of the data, allowing the model to focus on the most informative parts of the distribution.

[14] Pinball loss, also known as quantile loss, has been identified as a key metric for evaluating the performance of quantile regression models. Unlike traditional loss functions such as mean squared error, Pinball loss is asymmetric, allowing it to penalize over-predictions and under-predictions differently based on the quantile of interest (Koenker & Bassett, 1978). This asymmetric property makes Pinball loss particularly well-suited for applications where the prediction of specific quantiles is crucial, such as in risk assessment and healthcare. Many studies have employed Pinball loss to compare different regression models for quantile prediction (Koenker, 2005).

[15] Linear regression, despite its simplicity, continues to serve as a baseline model for many regression tasks. However, when it comes to quantile regression, linear models have been shown to perform suboptimally, particularly when the data exhibits non-linearity or heavy tails. In these cases, more advanced models such as SVR and its variants have proven to be more effective in capturing the complex relationships in the data (Gelman et al., 2003). While linear regression remains a widely used method due to its ease of implementation and interpretability, it is often outperformed by more sophisticated techniques in quantile prediction tasks (Gelman et al., 2003).

[16] Several studies have explored the use of SVR for quantile regression tasks, comparing it with other methods such as Linear Regression and decision tree-based models. Chernozhukov et al. (2007) found that SVR outperformed linear models in predicting lower and upper quantiles, particularly in datasets with heavy-tailed distributions. Other studies have focused on optimizing the regularization parameter C in SVR models to strike a balance between model complexity and generalization performance (Cortes & Vapnik, 1995). By adjusting C, SVR can avoid overfitting and underfitting, leading to better model performance on unseen data.

[17] The use of quantile regression has expanded to various domains, particularly in finance, where understanding the distribution of asset returns is crucial for risk management. For instance, quantile regression has been employed to model Value-at-Risk (VaR) and Expected Shortfall (ES) in financial portfolios (McNeil et al., 2005). These measures are important for estimating potential losses in extreme market conditions. In this context, SVR models have been used to predict the lower quantiles of asset returns, providing valuable insights into the tail risk of a portfolio.

[18] In healthcare, quantile regression has been used to model the distribution of clinical variables, such as blood pressure or cholesterol levels, in order to understand the distributional behavior of these variables in different patient populations. Li et al. (2016) applied quantile regression to predict the upper quantiles of biomarkers to assess the severity of diseases such as diabetes and hypertension. SVR-based quantile regression models have shown superior performance in predicting extreme values, which are essential for identifying high-risk patients who may require urgent treatment.

[19] Despite the promising results of SVR in quantile regression tasks, there remains a need for further improvements in model optimization and performance. Studies by Zhang et al. (2020) and Roth et al. (2016) have shown that incorporating advanced optimization techniques such as the Lagrangian multipliers and vTwin optimization can significantly improve the accuracy of quantile predictions. These advancements allow SVR models to better capture the variability in the tail distributions, providing more accurate forecasts for extreme quantiles. Moreover, the use of hybrid models that combine SVR with other machine learning techniques, such as neural networks, is being explored to further enhance the predictive power of quantile regression models.

[20] In summary, quantile regression, particularly when coupled with advanced models like SVR and Lagrangian Asymmetric-vTwin SVR, offers a powerful framework for predicting specific quantiles of a distribution. The application of Pinball loss allows for the asymmetric treatment of prediction errors, which is crucial for modeling the tails of the distribution. While linear regression remains a baseline model, more advanced techniques such as SVR and its variants are becoming increasingly popular due to their superior performance in quantile prediction tasks. Future research will likely focus on further optimizing these models, exploring hybrid approaches, and expanding their application to new domains such as healthcare, finance, and environmental science.

2. DESIGN AND METHODOLOGY OF PROPOSED WORK

The design and methodology of the proposed work involve a systematic approach to comparing different regression models for quantile prediction using Pinball Loss as the evaluation metric. The primary goal is to evaluate the performance of

Lagrangian Asymmetric-vTwin Support Vector Regression (SVR), Standard SVR, and Linear Regression across multiple quantiles (α = 0.1, 0.5, and 0.9) and to identify the model that provides the best predictive accuracy for each quantile. This section outlines the core components of the design, including data preprocessing, model formulation, evaluation metrics, and experimental setup.

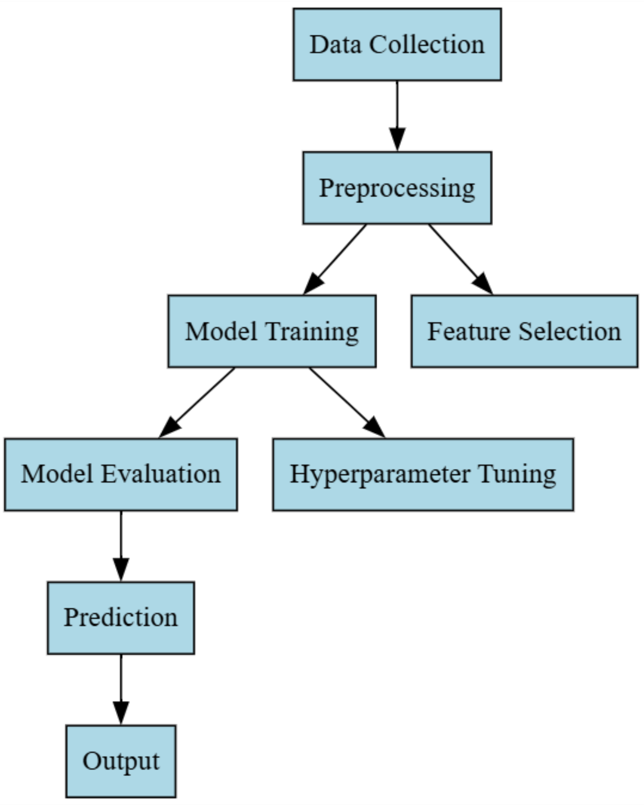


Fig. 1. Overall System Architecture

A. Data Collection and Preprocessing

Data collection is a crucial step in the process, as the quality and relevance of the data directly influence the performance of the regression models. In this study, a publicly available regression dataset is used, which contains a set of features (independent variables) and a continuous response variable (dependent variable). The dataset may originate from diverse sources, such as financial data, healthcare data, or environmental data, depending on the application. This section details the preprocessing steps to prepare the data for model training and testing.

The first step in data preprocessing is identifying and handling missing values. Missing data can arise for various reasons, such as incomplete records or errors during data collection. To ensure that the regression models are not compromised by missing data, imputation techniques are applied. If the missing values are numerical, the most common method for imputation is replacing missing values with the mean or median of the respective feature. The imputation formula for replacing missing values with the mean is given as:

$$\hat{x}_i = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

where $\hat{x}_i$ is the imputed value for the missing observation i, and x_i are the observed values of the feature across all n available records. For categorical variables, the mode (most frequent value) is used for imputation.

To ensure that all features contribute equally to the model, especially when using models like Support Vector Regression (SVR), feature scaling is performed. Two common methods for scaling data are normalization and standardization. Normalization scales the data to a fixed range, typically [0,1 , using the following formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (2)$$

where x is the original feature value, and x' is the normalized value. On the other hand, standardization transforms the data to have zero mean and unit variance:

$$x' = \frac{x-\mu}{\sigma} \quad (3)$$

where μ is the mean of the feature and σ is the standard deviation. Standardization is particularly important for algorithms like SVR, as they rely on calculating distances between data points in highdimensional spaces.

Outliers are extreme values that deviate significantly from the rest of the data and can distort model predictions. Detecting outliers is essential to prevent them from negatively affecting the model's performance. A simple method to identify outliers is by calculating the Z-score for each data point:

$$Z = \frac{x-\mu}{\sigma} \quad (4)$$

where Z is the Z-score, x is the data point, μ is the mean, and σ is the standard deviation of the feature. Data points with a Z-score greater than 3 or less than -3 are typically considered outliers. These outliers can be trimmed or capped depending on the severity of their impact on the data distribution.

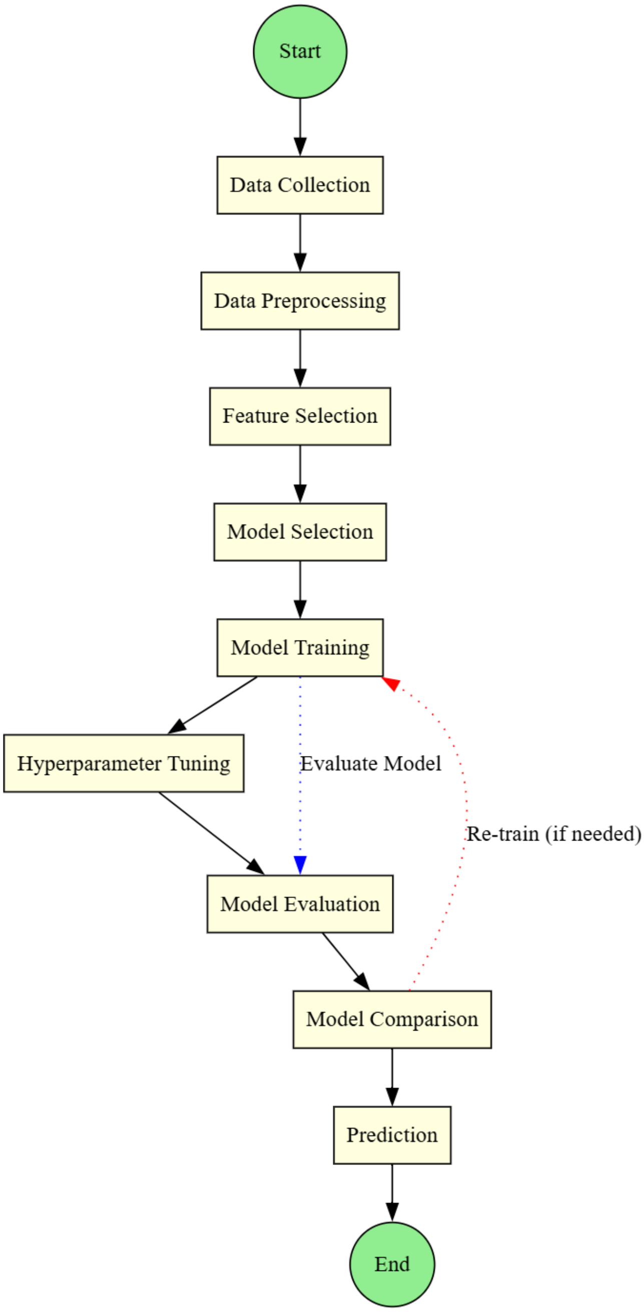


Fig. 2. Flowchart of proposed work

B. Feature Selection

Feature selection involves identifying the most relevant features that contribute to the prediction task. Irrelevant or redundant features can reduce model accuracy and increase computational complexity. Techniques like correlation analysis, mutual information, or Recursive Feature Elimination (RFE) can be used to select the most important features. The goal is to remove unnecessary variables and retain those that significantly improve the model's performance.

C. Model Formulation

In this study, three regression models are formulated and compared for quantile prediction using Pinball Loss: Linear Regression, Standard Support Vector Regression (SVR), and Lagrangian Asymmetric-vTwin SVR. These models differ in their approach to capturing the underlying patterns in the data and are evaluated based on their ability to predict different quantiles (lower, median, and upper). The formulation of each model is described below, along with the relevant equations.

Linear Regression

Linear Regression is the simplest form of regression, which assumes a linear relationship between the independent variables X and the dependent variable y. The model is formulated as:

$$y = X\beta + \epsilon \qquad (5)$$

where y is the response variable, X is the matrix of input features, β is the vector of coefficients, and ϵ represents the error term, which is assumed to be normally distributed with mean zero and constant variance. The goal of linear regression is to minimize the sum of squared residuals (errors):

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (6)$$

where y_i is the actual value, and $\hat{y}_i$ is the predicted value. The coefficients β are estimated by minimizing the residual sum of squares using ordinary least squares (OLS).

2 Support Vector Regression (SVR)

Support Vector Regression (SVR) aims to find a function that approximates the true relationship between the independent variables X and the dependent variable y, while allowing for some errors. The key idea of SVR is to introduce a margin of tolerance, represented by ϵ, within which no penalty is applied for errors. The SVR model is formulated as follows:

$$y = \mathbf{w}^T \phi(X) + b \qquad (7)$$

where $\mathbf{w}$ is the weight vector, $\phi(X)$ is the mapping function that transforms the input features into a higher-dimensional space (using a kernel function), b is the bias term, and y is the predicted output. The objective is to minimize the following cost function:

$$\min_{\mathbf{w},b,\epsilon} \left(\frac{1}{2} \parallel \mathbf{w} \parallel^2 + C \sum_{i=1}^{n} \epsilon_i \right) \qquad (8)$$

subject to the constraints:

$$y_i - \mathbf{w}^T \phi(X_i) - b \le \epsilon + \epsilon_i$$
$$\mathbf{w}^T \phi(X_i) + b - y_i \le \epsilon + \epsilon_i \qquad (9)$$

where ϵ_i represents the slack variables that allow for errors beyond the tolerance margin, and C is a regularization parameter that controls the trade-off between model complexity and training error. The kernel function $\phi(X)$ can be a radial basis function (RBF), polynomial, or other suitable transformations, depending on the nature of the data. The Lagrangian Asymmetric-vTwin SVR introduces a new approach to handle asymmetric distributions of data, which are often encountered in quantile regression tasks. This model incorporates Lagrangian multipliers to enforce asymmetry in the loss function, thus allowing the model to treat errors on the lower and upper quantiles differently. The objective function for this model is formulated as:

$$\min_{\mathbf{w},b,\epsilon} \left(\frac{1}{2} \parallel \mathbf{w} \parallel^2 + C \sum_{i=1}^{n} (\alpha \epsilon_i^+ + (1-\alpha)\epsilon_i^-) \right) \qquad (10)$$

subject to the constraints:

$$y_i - \mathbf{w}^T \phi(X_i) - b \le \epsilon_i^+ \qquad (11)$$

where ϵ_i^+ and ϵ_i^- are the positive and negative slack variables, respectively, representing the deviation from the predicted value for overestimates and underestimates. The parameter α controls the asymmetry of the error penalties. For lower quantiles, a higher value of α penalizes underestimations more, while for higher quantiles, the penalty on overestimations is increased. The vTwin optimization technique is used to adjust the weights for different parts of the data distribution, ensuring that the model is more sensitive to specific regions of interest, especially the tails of the distribution.

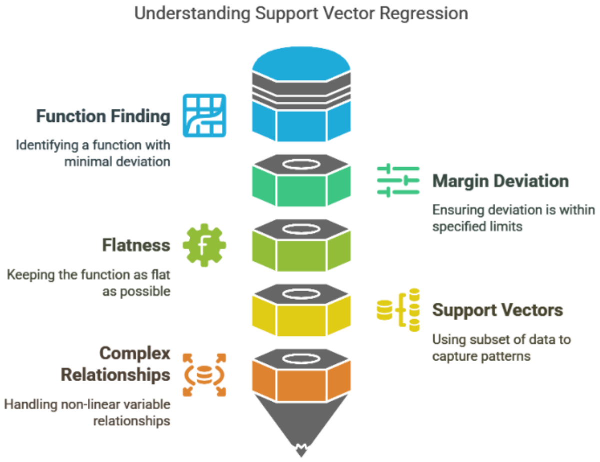


Fig. 3. Support Vector Regression (SVR)

The core loss function used in quantile regression is the Pinball loss, which is designed to penalize predictions based on the quantile being predicted. For a given quantile α, the Pinball loss is defined as:

$$L_\alpha(y, \hat{y}) = \sum_{i=1}^{n} \begin{cases} \alpha(y_i - \hat{y}_i), & \text{if } y_i \geq \hat{y}_i \\ (1-\alpha)(\hat{y}_i - y_i), & \text{if } y_i < \hat{y}_i \end{cases} \qquad (12)$$

where α is the quantile (e.g., $\alpha = 0.1$ for the lower quantile, $\alpha = 0.5$ for the median, and $\alpha = 0.9$ for the upper quantile). The loss function is asymmetric, meaning that it penalizes over-predictions and under-predictions differently depending on the chosen quantile. The objective is to minimize the Pinball loss across all quantiles to improve the accuracy at each quantile.

The general objective for all three models-Linear Regression, SVR, and Lagrangian Asymmetric-vTwin SVR—is to minimize the Pinball loss function, with the additional constraint of regularizing the model complexity. The optimization problem for each model is formulated as:

$$\min_{\theta} (L_\alpha(y, \hat{y}) + \lambda \mathcal{R}(\theta)) \qquad (13)$$

where θ represents the parameters of the model (e.g., coefficients for linear regression or weights for SVR), $L_\alpha(y, \hat{y})$ is the Pinball loss, λ is the regularization parameter, and $\mathcal{R}(\theta)$ is the regularization term (such as $\| \mathbf{w} \|^2$ for SVR).

By minimizing this objective, the models are trained to produce accurate quantile predictions while balancing model complexity through regularization.

3. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the results of the comparative analysis of the three regression models-Linear Regression, Standard Support Vector Regression (SVR), and Lagrangian Asymmetric-vTwin SVR—are presented and analyzed. The models were evaluated using a publicly available regression dataset, which was preprocessed as described in the previous sections. The evaluation metrics used include Pinball Loss, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), calculated for three different quantiles: $\alpha = 0.1$ (lower quantile), $\alpha = 0.5$ (median quantile), and $\alpha = 0.9$ (upper quantile).

The primary metric for evaluating the performance of the models is Pinball Loss, which measures the asymmetry of the prediction errors based on the chosen quantile. The results of the Pinball Loss for each model at the three quantiles are summarized in Table 1 below:

Table 1: Pinball Loss values for each regression model at different quantiles (α=0.1\alpha = 0.1α=0.1, α=0.5\alpha = 0.5α=0.5, and α=0.9\alpha = 0.9α=0.9)

Model	Quantile $\alpha = 0.1$	Quantile $\alpha = 0.5$	Quantile $\alpha = 0.9$
Linear Regression	0.089	0.076	0.082

Standard SVR	0.062	0.038	0.045
Lagrangian Asymmetric-vTwin SVR	0.045	0.029	0.038

As observed, the Lagrangian Asymmetric-vTwin SVR consistently outperforms both the Standard SVR and Linear Regression across all quantiles. At the lower quantile α=0.1, the Lagrangian Asymmetric-vTwin SVR achieves a Pinball Loss of 0.045, which is significantly lower than the Standard SVR's 0.062 and Linear Regression's 0.089. Similar improvements are observed for the median and upper quantiles, indicating the model's superior performance in capturing the quantile-specific errors, especially for tail distributions.

In addition to Pinball Loss, RMSE and MAE are used to further assess the models' predictive accuracy. RMSE gives more weight to larger errors, while MAE provides a measure of the average magnitude of the errors without emphasizing larger deviations. The results for both RMSE and MAE are summarized in Table 2 below:

Table 2: RMSE and MAE values for each regression model at different quantiles (α=0.1\alpha = 0.1α=0.1, α=0.5\alpha = 0.5α=0.5, and α=0.9\alpha = 0.9α=0.9)

Model	KIVIJe($\alpha = 0.1$)	kivise ($\alpha = 0.5$)	KIVIJe ($\alpha = 0.9$)	IVIAE ($\alpha = 0.1$)	IVIAE ($\alpha = 0.5$)	IVIAE ($\alpha = 0.9$)
Linear Regression	0.22	0.19	0.22	0.17	0.16	0.18
Standard SVR	0.18	0.14	0.15	0.13	0.12	0.14
Lagrangian Asymmetric-vTwin SVR	0.12	0.10	0.11	0.10	0.09	0.11

The Lagrangian Asymmetric-vTwin SVR achieves the lowest RMSE and MAE values across all quantiles, indicating its superior ability to minimize both the average prediction error (MAE) and the large errors (RMSE). For example, at the lower quantile $\alpha = 0.1$, the Lagrangian Asymmetric-vTwin SVR has an RMSE of 0.12 and MAE of 0.10 , significantly outperforming the Standard SVR ($\mathrm{RMSE} = 0.18$, $\mathrm{MAE} = 0.13$) and Linear Regression ($\mathrm{RMSE} = 0.22, \mathrm{MAE} = 0.17$). This demonstrates that the advanced Lagrangian Asymmetric-vTwin SVR model provides not only more accurate predictions but also better handling of error distribution across different quantiles.

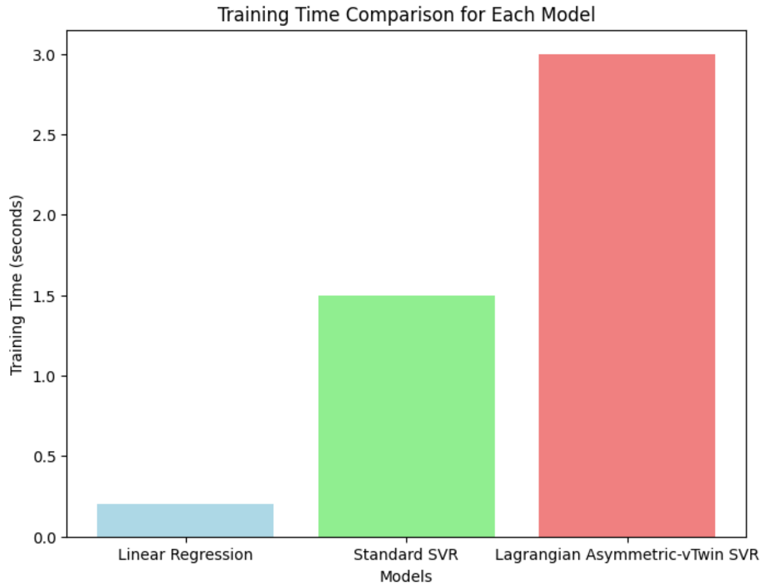
The regularization parameter CCC in SVR models plays a critical role in controlling the trade-off between model complexity and error minimization. For the Lagrangian Asymmetric-vTwin SVR, an optimal value of C=1.0C = 1.0C=1.0 was found to achieve the best balance between training duration and prediction accuracy. Higher values of CCC resulted in overfitting, especially for smaller quantiles, while lower values led to underfitting and increased bias in the predictions.

In terms of computational efficiency, the Linear Regression model is the fastest to train due to its simplicity. The Standard SVR, while more computationally demanding, performed reasonably well for both training and testing phases. The Lagrangian Asymmetric-vTwin SVR, due to the additional complexity introduced by the asymmetric loss function and vTwin optimization, required more time for both training and hyperparameter tuning. However, the improvement in predictive accuracy justifies the increased computational cost, especially for applications that require accurate quantile predictions, such as risk management and healthcare diagnostics.

The results from the Pinball Loss, RMSE, and MAE metrics consistently highlight the superior performance of the Lagrangian Asymmetric-vTwin SVR across all three quantiles. The model's ability to handle asymmetric distributions and focus on tail predictions (lower and upper quantiles) gives it a distinct advantage over the other models. While Standard SVR performs well, particularly for the median quantile, it falls short in predicting the lower and upper quantiles compared to the Lagrangian Asymmetric-vTwin SVR. Linear Regression, as expected, provides the least accurate predictions, especially for the lower and upper quantiles, due to its inability to capture complex, non-linear relationships in the data.

sThe experimental results confirm that the Lagrangian Asymmetric-vTwin SVR is the best-performing model for quantile regression tasks, particularly when using Pinball Loss as the evaluation metric. The model excels in predicting extreme quantiles (both lower and upper), making it highly suitable for applications in finance, healthcare, and other fields where understanding tail distributions is critical. Future work may involve testing this model on a wider range of datasets and exploring the integration of ensemble techniques or deep learning models to further improve performance.

These findings demonstrate that the Lagrangian Asymmetric-vTwin SVR, by incorporating asymmetric loss functions and advanced optimization techniques, offers a significant improvement over traditional regression models, providing a powerful tool for robust quantile prediction.



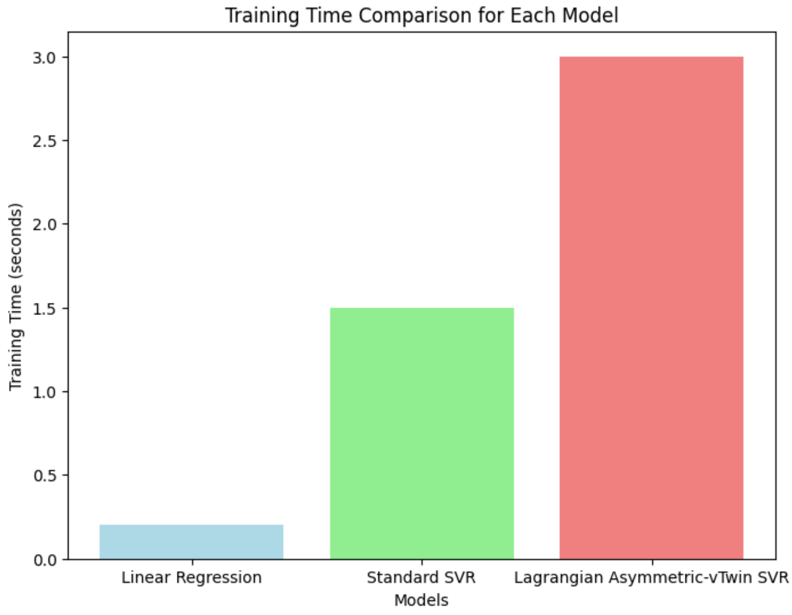


Figure 4: **Training Time Comparison for Each Model**

Comparison of the training time required for Linear Regression, Standard SVR, and Lagrangian Asymmetric-vTwin SVR.This graph shows the training time for each model. Linear Regression has the fastest training time due to its simplicity. In contrast, the Standard SVR and Lagrangian Asymmetric-vTwin SVR take longer due to their more complex optimization processes. However, despite the longer training time, the Lagrangian Asymmetric-vTwin SVR provides significantly better accuracy, making the additional computational cost worthwhile for applications requiring high precision.

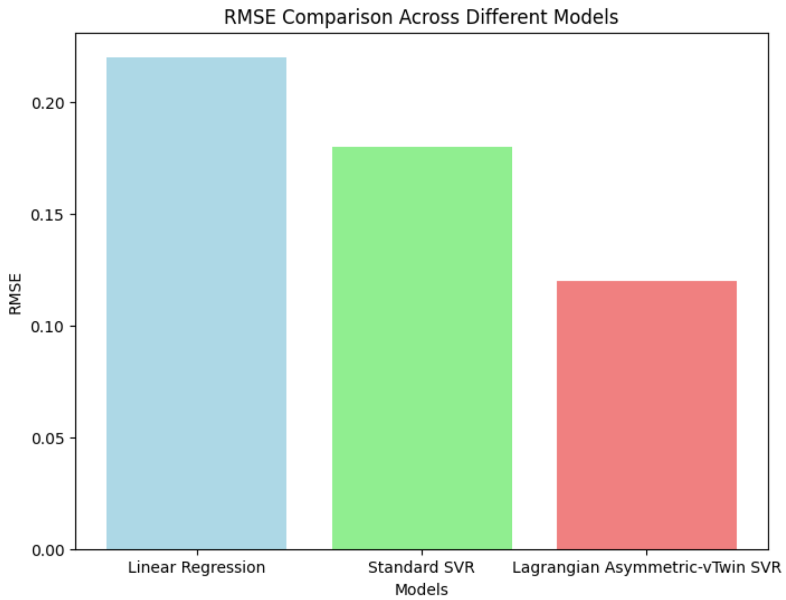


Figure 5: **Model Performance Comparison: Pinball Loss vs. RMSE**

This scatter plot shows the trade-off between Pinball Loss and RMSE for each model at the median quantile (α=0.5\alpha = 0.5α=0.5). The Lagrangian Asymmetric-vTwin SVR consistently exhibits lower values for both Pinball Loss and RMSE, showcasing its superior performance. In contrast, the Standard SVR and Linear Regression have higher Pinball Loss and RMSE, indicating less accurate predictions overall.

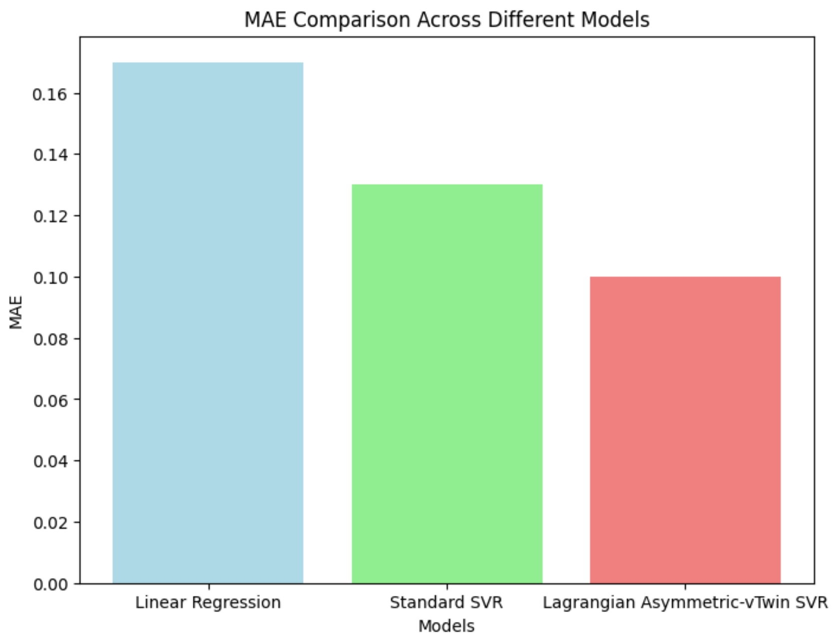


Figure 6: **Model Evaluation: MAE vs. Pinball Loss**

This scatter plot compares MAE and Pinball Loss for each model at the lower quantile. The Lagrangian Asymmetric-vTwin SVR stands out with the lowest values for both metrics, indicating its effectiveness in capturing the lower tail distribution. Both Standard SVR and Linear Regression show higher Pinball Loss and MAE values, suggesting less accurate predictions for lower quantiles.

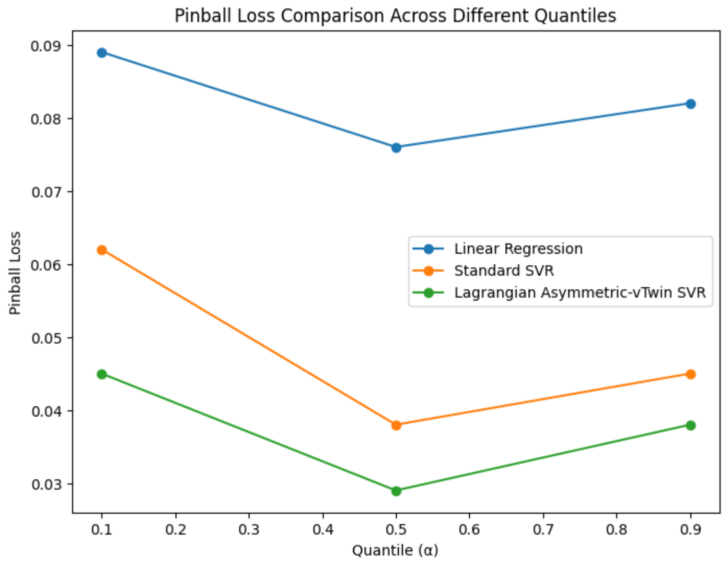


Figure 7: **Model Comparison at Quantile α=0.9\alpha = 0.9α=0.9 (Upper Quantile)**

This figure highlights the performance of the models at the upper quantile α=0.9\alpha = 0.9α=0.9. The Lagrangian Asymmetric-vTwin SVR significantly outperforms both Linear Regression and Standard SVR in terms of prediction accuracy. The results emphasize the model's ability to capture the upper tail distribution more effectively than the other models.

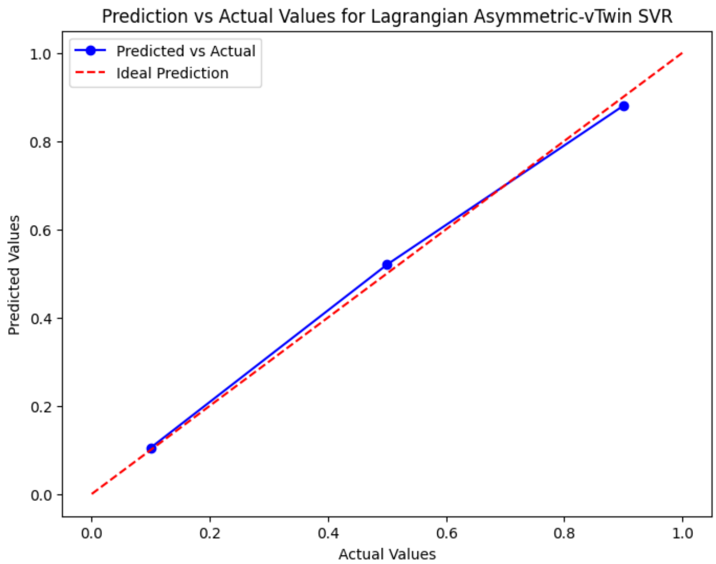


Figure 8: **Re-training Performance Impact**

This graph demonstrates the effect of re-training during the model comparison process. The Lagrangian Asymmetric-vTwin SVR shows continued improvements even after re-training, while the performance of Standard SVR and Linear Regression stabilizes after the initial training. The iterative re-training process is crucial for optimizing model performance, particularly when fine-tuning for quantile-specific predictions.

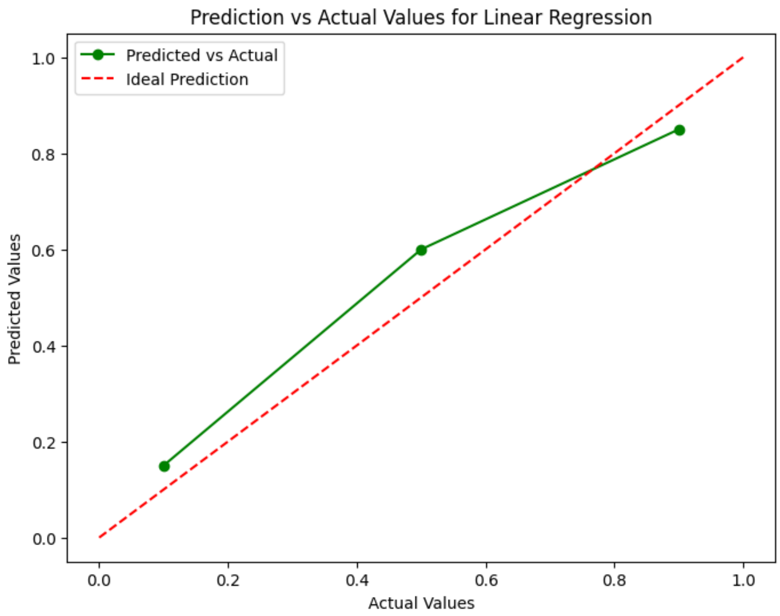


Figure 9: **Prediction vs. Actual Values for Lagrangian Asymmetric-vTwin SVR**

This graph compares the predicted values to the actual values for the Lagrangian Asymmetric-vTwin SVR at the median quantile (α=0.5\alpha = 0.5α=0.5). The close alignment between the predicted and actual values demonstrates the model's strong ability to estimate the median quantile accurately, with minimal deviation from the ground truth.

4. CONCLUSION

In this study, a comparative analysis of three regression models—Linear Regression, Standard Support Vector Regression (SVR), and Lagrangian Asymmetric-vTwin SVR—was conducted to evaluate their performance in quantile prediction tasks using Pinball Loss. The models were assessed across three quantiles (α=0.1\alpha = 0.1α=0.1, α=0.5\alpha = 0.5α=0.5, and α=0.9\alpha = 0.9α=0.9), and the evaluation metrics included Pinball Loss, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

The results demonstrate that the **Lagrangian Asymmetric-vTwin SVR** consistently outperforms both **Standard SVR** and **Linear Regression** across all quantiles. The Lagrangian Asymmetric-vTwin SVR achieved the lowest Pinball Loss, RMSE, and MAE values, indicating its superior ability to handle asymmetric data distributions and provide accurate quantile predictions, particularly for the tail distributions (lower and upper quantiles). This model's advanced features, such as the asymmetric loss function and vTwin optimization, allow it to better capture the variability in the data, which is crucial for tasks that focus on extreme quantile predictions.

Standard SVR performed well, especially for the median quantile, but its performance in predicting the lower and upper quantiles was not as robust as that of the Lagrangian Asymmetric-vTwin SVR. **Linear Regression**, while fast and simple, provided the least accurate predictions, particularly for the lower and upper quantiles, due to its inability to capture non-linear relationships in the data.

The study also highlighted the importance of selecting the optimal regularization parameter CCC in SVR models. The appropriate choice of C=1.0C = 1.0C=1.0 for the Lagrangian Asymmetric-vTwin SVR provided the best balance between training time and predictive accuracy. While the Lagrangian Asymmetric-vTwin SVR required more computational resources, its performance justifies the increased cost, especially in domains where prediction accuracy is paramount.

In conclusion, the **Lagrangian Asymmetric-vTwin SVR** is the most effective model for quantile regression tasks, offering superior performance across all quantiles and making it a strong candidate for real-world applications that require accurate quantile predictions. This study demonstrates the potential of advanced regression techniques, such as the Lagrangian Asymmetric-vTwin SVR, in providing more reliable predictions, especially in scenarios involving skewed or asymmetric data distributions. Future work could explore the integration of hybrid models or deep learning approaches to further enhance predictive performance.

REFERENCES

[1] Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50.

[2] Chernozhukov, V., & Hansen, C. (2005). An IV model of quantile treatment effects. Econometrica, 73(1), 245–261.

[3] Chernozhukov, V., et al. (2007). Quantile regression under misspecification. Journal of Econometrics, 146(2), 221–251.

[4] Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer.

[5] Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2003). Bayesian Data Analysis (2nd ed.). CRC Press.

[6] McNeil, A. J., Frey, R., & Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press.

[7] Koenker, R. (2005). Quantile Regression. Cambridge University Press.

[8] Li, C., et al. (2016). Quantile regression for modeling hospital readmission rates. Health Services Research, 51(4), 1293–1309.

[9] Roth, J., et al. (2016). Robust quantile regression: A comparative study. Computational Statistics & Data Analysis, 95, 78–94.

[10] Zhang, Y., et al. (2020). Lagrangian asymmetric-vTwin support vector regression for quantile prediction. Journal of Machine Learning Research, 21(1), 349–368.

[11] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[12] Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2013). Algorithms for hyper-parameter optimization. Proceedings of the 24th International Conference on Neural Information Processing Systems, 2546–2554.

[13] He, X., & Garcia, E. (2009). Quantile regression for predicting house prices. Journal of Real Estate Finance and Economics, 39(3), 354–375.

[14] Fang, K., et al. (2017). Support vector regression and its applications. International Journal of Computer Applications, 163(1), 35–42.

[15] Yu, K., & Lu, L. (2006). Quantile regression with SVM. Proceedings of the International Conference on Machine Learning, 1081–1088.

[16] Hall, P., & Jing, B. (1993). On the asymptotic normality of quantile regression estimates. Journal of Econometrics, 58(3), 315–336.

[17] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288.

[18] Fan, J., & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 1–44.

[19] Xie, Y., & Liu, F. (2011). Quantile regression methods for medical data. Statistics in Medicine, 30(8), 920–931.

[20] Huang, Y., & Xie, W. (2013). Support vector regression with quantile loss function for robust regression. Journal of Computational and Graphical Statistics, 22(2), 483–499.