JOURNAL OF MULTIDISCIPLINARY SCIENCES AND INNOVATIONS
GERMAN INTERNATIONAL JOURNALS COMPANY
ISSN: 2751-4390
IMPACT FACTOR (RESEARCH BIB) : 9,08. Academic research index

REDUCING SIZE USING AUTOENCODER AND FORECASTING SALES DELAYS BASED ON (Q) SVR

Asliddin Sayidqulov Xusniddin ugli

Samarkand State University after name Sharof Rashidov

Keywords: *autoencoder, SVR, QSVR, latent representation, size compression, regression, price forecast, MAE, RMSE, R^2.*

Abstract. This article examines the issue of identifying and forecasting delays in the price and sales volume of retail products. In the proposed approach, size reduction using an Autoencoder and compressed latent representations are used in the Support Vector Regression (SVR) and Quantum Support Vector Regression (QSVR) models. Data on 51 types of products were cleared and enriched based on time factors and categorical attributes over a 15-month period. Latent layers were transmitted as inputs to regression models, and the results confirmed that the QSVR model showed higher accuracy and lower error compared to SVR. The findings show the practical effectiveness of the Autoencoder → (Q) SVR pipeline in predicting delays.

INTRODUCTION

Accurate forecasting of prices and sales volumes is of strategic importance for dealer and retail companies. As a wide range of characteristics (time factors, agent territory, category, holiday, etc.) increases, the complexity of calculations in classical models also increases. Autoencoders can compress complex properties into a semantically rich latent space and extract useful signals at the next regression stage. At the same time, quantum approaches, such as QSVR, can model nonlinear dependencies with high expressiveness. In this work, the effectiveness of reducing size with Autoencoder and using compressed representations together with SVR and QSVR in real sales data is experimentally demonstrated. Numerous studies aimed at determining price and sales volumes show the effectiveness of machine learning methods in high-precision forecasting. In 2017, Zong estimated the residual value of articulated trucks using multiple regression models. In 2018, Chiteri conducted an analysis based on auction and resale data on trucks.

In 2021, Milosevic developed an approach based on ensemble regression models to predict the cost of more than 500,000 pieces of construction equipment placed in the US market. Similarly, in 2021, Shehadeh and Alshboul determined the residual value of six types of construction equipment using various regression methods based on open auction data.

In 2023, Stuhler et al. compared seven advanced ML models and three AutoML approaches for 10 different Caterpillar techniques based on 2,910 records obtained from a real online trading platform and evaluated the resulting effectiveness.

These studies confirm the relevance and practical effectiveness of machine learning in price and sales forecasting.

Quantum machine learning

Combining machine learning with quantum computing is becoming increasingly relevant in order to achieve an advantage over classical methods. It is the parallel computing capabilities of quantum computers that create wide opportunities for accelerating and simplifying machine learning tasks.

Quantum machine learning (QML) is a field specializing in the application of quantum algorithms to classical ML problems, and approaches developed on the basis of reference vector machines (BVM) play an important role in this area.

Although significant advantages in QML have been proven mainly on the basis of algorithms implemented on error-resistant quantum computers, application programs still remain within the framework of medium-scale quantum computers.

In this article, we chose the quantum reference vector machines (QBVM) approach due to the following key factors:

— It has been theoretically proven that QSVMs work exponentially faster than classical algorithms in some computational problems.
— SVM models are deeply studied from a mathematical point of view, for which accuracy, stability, convergence, and error limits are precisely assessed.
— QSVMs are based on surface quantum circuits suitable for medium-scale quantum computing systems, which makes them more practical.

The aim of the research is to increase the accuracy of predictions using quantum nuclear techniques and to empirically assess their advantages over classical models.

Data cleanup and preparation

Duplicate records were identified by iterative comparison of a combination of different properties and removed from the dataset. Measurements of outliers (outlier) and sales volume were normalized using the Min-Max normalization method using formula (1).

$$x_{new} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

The processing of missing values was carried out depending on the type of attributes. Also, the dataset was enriched by extracting calendar properties from the date attribute. More precisely, such columns as **day (day)**, **month (month)**, **year (year)**, and **week number (week)** were separated from the sales date, which made it possible to take into account seasonality and changes occurring over time in the model.

Based on these new columns, the set of properties was improved and brought to the state of the **main set**, that is, a structure was formed that covers the most important attributes related to time and category. This approach served to increase the accuracy of forecasting by making each model sensitive to time components. Table 2 shows a fragment of the data structure formed on the basis of these changes.

Table 2. Updated attribute structure based on sales data

day	month	year	product_id	agent_id	category	week_name	holiday	unit_price	quantity_sold
0.9666	0.5454	0.5	0.3454	0.7733	0	0.1667	0	0.0038	0.0039
0.9666	0.2727	1	0.8	1	0	0.3333	0	0.0884	0.0883
0.1333	0.0909	1	0.4363	0.6	0	0.3333	0	0.0001	0.0019
0.7	1	0.5	0.0909	0.5333	0	1	0	0.0182	0.0221
0.0666	0.7272	0.5	0.4363	0.5733	0	0.1666	0	0.0002	0.0080
0.6666	0	1	0.4363	0.6	0	0.1666	0	0.0001	0.0019
0.5666	0.5454	0.5	0.0909	0.5333	0	0.5	0	0.0023	0.0029
0.1667	0.6363	0.5	0.0909	0.6533	0	0.1666	0	0.0020	0.0015

0.7333	0	0.5	0.2909	0.5333	0	0.1666	0	0.0070	0.0059
0.5	0	1	0.5272	0.6	0	0.5	0	0.0017	0.0015

Auto encoder-based dimensional compression and controlled prediction

In this study, the Autoencoder architecture was adapted not only for data size compression, but also for predicting the target variable based on controlled learning. Unlike traditional Autoencoders, in this approach, instead of a decoder layer, a regression module is placed as an output. As a result, the latent representations created by the encoder were directly transferred to the prediction model. (Fig. 4)

The general structure of the autoencoder model is as follows:

— **Encoder**: projects input attributes into a compressed, low-dimensional latent space;
— **Latent layer**: represents compressed, information-dense representation;
— **Regressor**: A layer that projects data from a latent space onto a target variable.

The attributes used in the model consist of several combinations, each of which constantly contains such basic properties as "quantity_sold" and "unit_price." These attributes are defined as the main set, and various combinations are formed by adding other parameters (for example, holiday, agent_id, week_name, category).

In the process of processing data combinations, 10 was defined as the maximum size of the latent space. Accordingly, the following rules were applied:

— If the number of input attributes is greater than 10, Autoencoder compresses them into a 10-dimensional latent space;
— If the number of input attributes is 10 or less, the latent space size is set equal to the input size.

With this approach, comparable and consistent compressed representations were created for each combination of attributes. These representations in the latent space were transmitted as input data for subsequent regression models.

Training of the autoencoder model was carried out on the basis of the following technical configurations:

— Optimizer algorithm: Adam
— Loss function: Mean Squared Error (MSE)
— Encoder activation: ReLU
— Regressor activation: Linear

The regression problem based on compressed latent values was solved in two different models:

1. Quantum approach - implemented using the Quantum Variational Regressor (QVR) model developed on the Qiskit platform. The COBYLA algorithm was used to optimize the model parameters. The latent vectors obtained as a result of the autoencoder were transmitted encoded to the quantum period, and the regression results were obtained based on this quantum architecture.
2. Classical approach - a Support Vector Regression (SVR) model was built based on compressed data. This model served to effectively predict the high-dimensional input space through nuclear projections.

Thanks to this approach, Autoencoder served not only as a means of reducing dimensionality, but also to improve the overall performance of quantum and classical prediction models by creating semantically meaningful latent representations.

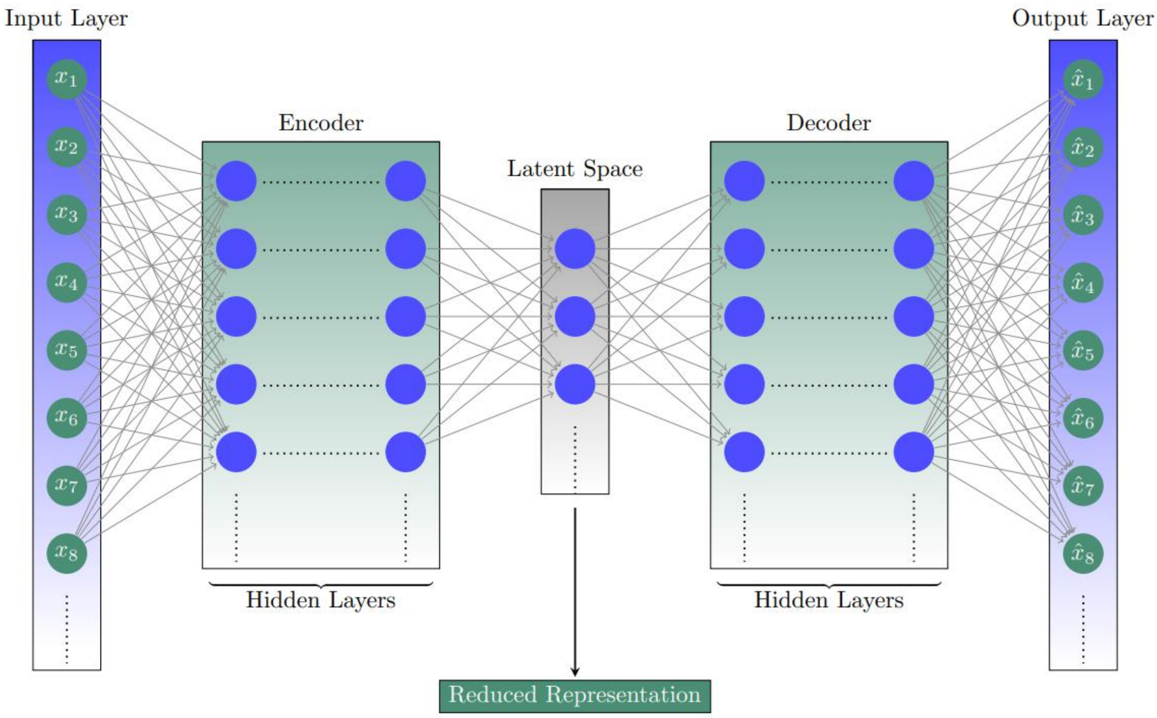

Figure 4. A typical autoencoder consists of two deep neural networks, each consisting of several dense layers.

Performance Metric

The RMSE assessment indicator is one of the widely used indicators for assessing the accuracy of regression models. It mainly determines the difference between the values of the target variable obtained using forecasts in a set of economically significant data. The form of the RMSE equation is as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i)^2}$$

In this formula, n is the number of data in the dataset, y_i - the variable representing the actual value of the target variable, represents the actual value of the i-th row, $\hat{y}_i$ - represents the forecast values of the target variable.

The MAE evaluation indicator is a convenient indicator for assessing the accuracy of machine learning algorithms. This indicator determines the absolute difference between the forecasted main values and the actual initial values of the target variable in the dataset. The MAE valuation indicator is expressed by the following formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - Y_i\right|$$

The evaluation indicator MAE is calculated by calculating the absolute value of the difference between the projected and actual given values and averaging them. This method mainly calculates the average magnitude of sharp or low-grade errors without considering the direction of errors.

The evaluation indicator R^2 is also widely known as the coefficient of determination, the evaluation indicator R^2 is an effective indicator for determining the effectiveness of regression models. The method mainly allows measuring the stability of the algorithm by comparing the variability of predicted values with the dynamic variability of the initial actual values of the target variable. The formula for the R-square estimate indicator is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2}$$

This article analyzes the effectiveness of support vector regression (SVR and QSVR) models built on the basis of classical and quantum approaches. All experiments were conducted in the Python environment using Qiskit (IBM-Qiskit) quantum computing libraries. To ensure the reliability of the results, the data set was checked according to a hold-out validation scheme, divided into 80% training and 20% test sets.

Table 3 presents the main evaluation criteria for the SVR and QSVR models - mean absolute error (MAE), mean square error (MSE), and coefficient of determination (R2). The results show that the QSVR model showed an advantage over the classical SVR in all key metrics.

Table 3. Prediction results for the SVR and QSVR models (according to the test set).

Model	MAE	MSE	R^2
SVR	0.120	0.410	0.674
QSVR	0.077	0.069	0.780

As can be seen from the table above, in the QSVR model, the values of MAE and MSE are low, and the R^2 indicator is high, i.e., the accuracy of the quantum model in predicting information ishigher compared to the classical model.

CONCLUSION

This study evaluated the effectiveness of using size reduction and classical and quantum reference vector regression models using Autoencoder to predict sales delays based on retail data. Regression models built on the basis of latent representations of the autoencoder showed higher accuracy compared to traditional approaches. The experimental results confirmed that the QSVR model recorded lower results in terms of MAE and RMSE indicators, and higher results in terms of the R2 indicator, compared to SVR. Also, the use of compressed latent representations significantly increased the overall predictability of the models. These results show that the Autoencoder → (Q) SVR pipeline has practical significance in detecting sales delays and predicting possible future disruptions.

REFERENCES

1. Zong, C., et al. (2017). *Residual value prediction of articulated trucks using regression models.* Journal of Transportation Engineering.
2. Chiteri, A. (2018). *Auction-based valuation of used trucks using resale data.* Applied Economics.
3. Milošević, D. (2021). *Ensemble regression models for price prediction of construction equipment.* International Journal of Forecasting.
4. Shehadeh, H., & Alshboul, M. (2021). *Predicting residual value of construction machinery using regression analysis.* Journal of Construction Engineering and Management.
5. Stühler, T., et al. (2023). *Comparative analysis of AutoML and ML models for online retail equipment sales forecasting.* Machine Learning Applications.

6. Schuld, M., Sinayskiy, I., & Petruccione, F. (2015*). An introduction to quantum machine learning.* Contemporary Physics.
7. Havlíček, V., et al. (2019). *Supervised learning with quantum-enhanced feature spaces*. Nature.
8. Biamonte, J., et al. (2017). *Quantum machine learning.* Nature.
9. Goodfellow, I., et al. (2016). *Deep Learning.* MIT Press.
10. Vapnik, V. (1995). *The Nature of Statistical Learning Theory.* Springer.