International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Comparative Analysis of ARIMA and VARAlgorithms for Performance Analysis of High-Speed Diesel Pumps

Smita Mahajan[1*], Shivali Amit Wagle[2], Nihar Ranjan[3], Santosh Borde[4]

Submitted: 20/12/2023 **Revised**: 31/01/2024 **Accepted**: 10/02/2024

Abstract: The demand for precise and efficient forecasting of High-Speed Diesel (HSD)pump performance is critical for optimizing fuel distribution, operational planning, and resource allocation in the petroleum industry. This paper presents a comprehensive comparison analysis of implementing two widely used time series forecasting algorithms, Auto regressive Integrated Moving Average (ARIMA) and Vector Auto Regression (VAR), for predicting vibration in electrical systems. The study spans a year-long dataset collected at various intervals, including seconds, minutes, hours, days, weeks, months, and yearly intervals, leveraging data from voltage, current, and temperature sensors. The research analyzes "Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE)" as three critical indicators for evaluating how well ARIMA and VAR perform. The analysis reveals that ARIMA consistently outperforms VAR across all intervals, demonstrating superior accuracy in predicting vibration levels. The data The dataset collected from a range of sensors provides a diverse and rich source of information, effectively capturing the electrical system's dynamic behavior. The results highlight the significance of selecting an appropriate forecasting model for time series data, especially system reliability and maintenance applications. This research contributes to the ongoing discourse on algorithm selection in time series forecasting for electrical systems and provides valuable insights for practitioners and researchers alike. The findings underscore the importance of considering the dataset's specific characteristics and the nature of the target variable when choosing between ARIMA and VAR algorithms for predictive modeling.

Keywords: *ARIMA, VAR, Time series forecasting, prediction*

1. Introduction

High-Speed Diesel (HSD) pumps play a pivotal role in various industrial and operational domains, where their optimal performance is critical for ensuring the efficiency and reliability of machinery. In contemporary industrial setups, the ability to predict and manage the performance of these pumps has become increasingly vital, as unexpected failures can lead to costly downtime and disruptions. This paper presents an in-depth investigation into the predictive modeling of HSD pump performance, utilizing a comparative analysis between Vector Auto regression (VAR) modeling and traditional statistical methods. The study references a real dataset gathered over a year. It contains essential operating data, such as temperature readings, vibration measurements of the HSD pump, and three-phase voltage and current.

The amalgamation of this data from these diverse sensors provides a rich sourceof information crucial for understanding the intricate dynamics influencing the performance of the HSD pump. Through rigorous analysis and modeling, this paper endeavors to forecast the behaviour of the pump, aiming to proactively detect pat- terns, anomalies, and potential performance deviations. Furthermore, utilising two distinct methodologies—VAR modeling and conventional statistical approaches—enables a comparative assessment of their efficacy in predicting the HSD pump performance. The VAR model, known for its ability to handle multivariate time series data, will be juxtaposed against traditional statistical methods to assess their respective strengths and weaknesses in this predictive context.

The multidimensional nature of the dataset, encompassing electrical parameters (3-phase voltage and current), thermal information (temperature readings), and mechanical insights (vibration data), offers a unique opportunity to employ a holistic approach to pump performance prediction. This comprehensive analysis seeks to forecast potential issues or deviations in the pump's behaviour and establish a comparative understanding of the predictive accuracy and efficiency between the VAR model and conventional statistical ARIMA model. The implications of this research are far-reaching, with the potential to significantly enhance predictive maintenance strategiesand operational efficiencies in

1,2Artificial Intelligence and Machine Learning Department, Symbiosis Institute of Technology, (Pune campus), Lavale, Pune, 412115, Maharashtra, India.
3,4JSPM'S Rajarshi Shahu College of Engineering, Pune-411033, Maharashtra, India
rajansmita@gmail.com
kulkarnishivali@yahoo.co.in
nihar.pune@gmail.com
santoshborde@yahoo.com
**Corresponding author: rajansmita@gmail.com*

industrial settings reliant on HSD pump functionality. This comparative analysis contributes valuable insights into predictive modeling for HSD pumps, providing a framework for optimized performance forecasting and proactive maintenance.

2. Literature survey

Various machine learning models have been applied to analyze and predict energy consumption patterns of hourly and daily energy consumption data from Kaggle. These models include Gaussian Processes, Support Vector Regression, K-Nearest Neighbor Regression, Multi-Layer Perceptron, and Linear Regression. Additionally, time series forecasting models, specifically ARIMA and VAR, were employed using Python. Using scenarios with and without meteorological data, these models were used to predictthe energy usage of South Korean households. The results obtained from this research highlight the effectiveness of different methods for energy consumption prediction. Among the techniques investigated, Support Vector Regression emerged as the most accurate method for energy consumption prediction, demonstrating its robustness in capturing consumption trends. Following closely in performance are Multilayer Perceptron and Gaussian Process Regression. These models displayed notable forecasting capabilities and are essential tools for enhancing our understanding of household energy consumption dynamics, particularly in the context of South Korean households [1].

The complexities surrounding the severity of road traffic accidents in the UK. Our approach is a powerful blend of various analytical tools, including machine learning algorithms, econometric methods, and traditional statistical techniques, all applied to examine longitudinal historical data. Our comprehensive analytical framework encompasses descriptive, inferential, bivariate, and multivariate approaches, incorporating correlation analysis through Pearson's and Spearman's Rank Correlation Coefficients, multiple and logistic regression models, Multi-collinearity Assessment, and Model Validation. To mitigate heteroscedasticity and autocorrelation in error terms, we have enhanced the precision and reliability of our regression analyses by leveraging the Generalized Method of Moments (GMM). Furthermore, we have harnessed the potential of the Vector Autoregressive (VAR) model and the Autoregressive Integrated Moving Average (ARIMA) models to enable precise time-series forecasting [2].

The multi-objective optimization method using a SVM-Support Vector Machineand Genetic In light-duty diesel engines operating at high altitudes, the algorithm efficiently lowers NOx emissions and fuel consumption while preserving power output and a minimum smoke limit[3]. As per [4], Support vector regression is an efficient technique for predicting diesel engine performance, providing high accuracy and reducing testing time and cost. According to [5], NPN models using expanded training dataare the best for predicting hydrocarbon emissions from diesel engines operating at constant speeds and loads, with ISO 8178-4 emission tests providing suitable data for predictions. According to [6], the Convo-LSTM model outperforms other deep learning algorithms [29] [30] in wireless mesh networks for traffic prediction and performance predictionof High-Speed Diesel pumps.

Using a genetic algorithm, the adaptive Support Vector Regression model accurately predicts diesel engine system reliability, even with a small dataset and varying system lifetimes as per [7]. Artificial neural networks can effectively predict and describe diesel engine sound quality, reflecting the nonlinear relationship between objective parameters and subjective satisfaction as experimented by [8]. Artificial neural network models provide more accurate predictions of engine performance, torque, and emissions in diesel engines fueled with bio diesel-alcohol mixtures compared to linear regression models [9]. Using support vector regression, the proposed multi- condition performance prediction method by [10]for centrifugal pumps effectively predicts performance under multiple operating conditions, improving pump design. As per [11], Machine learning algorithms [28] accurately predict emission and performance responses in CI engines fuelled with metal-oxide-based nanoparticles, reducing CO and NOx emissions and improving engine performance.

The neural approximation of nonlinear model predictive control (NMPC) improves stability and performance in diesel engine air path control, achieving high-speed computation and satisfying constraints like compressor surge and choke [12]. Younis etal. [13] have stated that Artificial neural networks (ANN) have a greater prediction accuracy than multiple linear regression models in predicting the gross heat value of diesel fuel combustion, with the best models being back-propagation networks (8-8-1) and (8-5-1).

3. Methodologies

3.1 Data acquisition and Processing

The process for gathering and preparing the data is described as follows before going into detail about the methods used to model the data. Several sensors were used to gather the data over the span of a year. Eight sensors in all were positioned to gather information. One output variable and seven input variables make up the gathereddata. Table 1 describes each of these parameters. This mostly consists of vibration, voltage, current, and temperature sensors.

Table 1 Data description

Collection of Data for Parameter	An explanation
Date	The specific data that specifies the interval of time during which the data were gathered
Three-phase currents	IR, IY, and IB
Three-Phase Voltage:	VR, VY, and VB
Temperature	Output of the Temperature sensor
Vibration	The output parameter to be observed

Eight thousand nine hundred and sixty (8960) of these tuples comprise the whole data. Each one contains a single time stamp identifier (Date / Time), seven input variables (three-phase voltage, current, and temperature sensor readings), and one output variable (the reading from the vibration sensor).

Algorithms used for Network Performance Prediction

Resource planning, network optimization, and management all depend on the ability to predict network traffic. To anticipate network traffic, a variety of statistical and non-statistical algorithms can be used. For output prediction, the current study employs both methods.

3.2 Vector Auto Regression

A time series is made up of observations made at predetermined intervals of time. The frequency of observations determines whether a time series is hourly, daily, weekly, monthly, quarterly, or annual, ending a variety of variables related to the series' essential characteristics [14]. When two or more-time series affect one another, a forecastingtechnique called Vector Auto Regression (VAR) can be applied. Put alternatively, there exists a bidirectional relationship between the time series when multiple time series affect one another[15].

A statistical model called VAR is employed to capture the linear inter dependencies between several time series data [16]. Multivariate time series data are incorporated into univariate autoregressive models (AR models). Every variable in the system is represented in a VAR model as a linear combination of its historical values and the historical values of every other variable. VAR models are widely used in economics, finance, and other fields to model the joint behaviour of multiple variables over time [17]. They provide a flexible framework for capturing inter dependencies and dynamic relationships among time series data [18].

- Multivariate Time Series: VAR models are employed when several time series variables interact. All of the system's variables are modeled as functions of their historical values as well as the historical values of each other.
- Order pp: pp is the model order, in the VAR model pp indicates the number of lag observations [19]. For instance, a VAR(2) model has the two most recent time points.
- Stationarity: Prior to developing a VAR model, it is typically essential to ensure

that the time series variables exhibit stationarity, a prerequisite similar to thatof univariate autoregressive models [20]. Various tests, including the "Augmented Dickey-Fuller (ADF)" test, can be employed to assess and confirm the stationarity of the time series variables.

- Estimation: Estimating the coefficients in a VAR model is typically done using

techniques like least squares [21].

- Impulse Response Function (IRF): VAR models frequently analyses the variables' dynamic reactions to inputs [22]. Each variable's response to a single input in any of the variables can be seen by the Impulse Response Function.

AR is known as an auto-regressive part of model because each variable (Time Series) is viewed as a function based on prior values. The predictors of the series are their lags or time-delayed values. This forecasting model differs from others primarily in that both are unidirectional, meaning that predictors affect Y but not the latter.

However, vector auto regression, or VAR, is bi-directional. Stated differently, the variables influence each other. In auto-regression models, the time series is representedas a linear mixture of its lags. Put another way, the values of the series in the past predict their future values.

A typical AR(p) model equation is as shown in equation 1

$$Y_t = \alpha + \beta Y_{t}1 + \beta Y_{t}2 + ... + \beta Y_p + \epsilon_t \quad (1)$$

Where the intercept, α, is a constant, and the coefficients of the lags of Y up to order p are β_1, β_2, and β_p.

Order 'p' denotes using the predictors in the equation up to p-lags of Y.

The error, or ϵ_t, is known as white noise. Each variable's historical values and the other variables' historical values in the system are combined linearly to represent eachvariable in the VAR model.

One set of equations, one for each variable (time series), is used to represent a time series composed of several time series that influence each other. The number of equations is equal to the number of time series interacting.

3.3 "Auto Regressive Integration and Moving Average"

"ARIMA - Auto Regressive Integrated Moving Average uses a moving average and auto-regression algorithm combination to estimate future outcomes based on historical time series data" [23]. Mathematically, it is represented as follows:

$$h_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} \epsilon_t \quad (2)$$

$$y_t = h_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + ... + \phi_q \epsilon_{t-q} \quad (3)$$

Specific terms are derived via auto-regression, while others are derived from movingaverages. When there is an excess difference in the time series, add more MA terms; when there is an under-differentiated term, add more AR terms.

In the presence of non-stationarity within the dataset, the ARIMA (p, d, q) approach incorporates lags at either the first or second differencing levels. Conversely, for stationary data without lag, an alternative technique involves using ARMA (p, q). Here, 'p' signifies the "Moving Average (MA)" order, and 'q' denotes the Autoregressive (AR) order, representing the number of lagged errors considered in forecasting within the ARIMA model.

The widely employed method for achieving stationarity in a time series involvessubtracting the initial value from the current value. The necessity for one or more lags depends on the nature of the time series—whether it is univariate or multivariate. Accordingly, when the data series remains stationary without differentiation, d = 0 signifies the minimal differentiation required to maintain stationarity. In the identification process, the correlogram is plotted as the initial step to assess the presence of autocorrelation (ACF) and partial autocorrelation (PACF) [24].

Subsequently, analyze the auto-regressive and moving averages to identify suitablemodels by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series. The determination of the model order parameters (p, q) is guided by the observation of patterns in the ACF and PACF graphs,and the most optimal model is then selected. Once the best-fitting model is identified, forecasting leverages its parameters (p, d, and q). To assess the efficacy of the chosen model, diagnostic forecasting involves evaluating it using statistically significantmetrics such as mean square error, Bayesian criteria (BIC), and Akaike information criterion (AIC) [24].

ARIMA stands as a widely acknowledged and utilized model for time series fore- casting, amalgamating three crucial components. Moving Average (MA), Integrated(I), and Auto Regressive (AR) as its three main components.

- Auto Regressive (AR): The AR facet embodies the autoregressive aspect of the model, signifying that the value of the time series variable at a specific time is directly tied to its prior values. The parameter "p" denotes the order of the autoregressive component, representing the number of lagged observations incorporated into the model [25].

- Integrated (I): Represented by the "d" parameter, the integrated component corresponds to differencing. "Differencing involves subtracting the previous observation from the current one, a process employed to induce stationarity in the time series. Stationarity streamlines the modeling process" [26].

- Moving Average (MA): The MA component characterizes the moving average element of the model, indicating that the current value of the time series variable is influenced by past white noise (random) errors. The "q" parameter signifies the order of the moving average component [27].

Steps for fitting an ARIMA model:

1. Stationarity: Check and ensure that the time series is stationary. If not, apply differencing until stationarity is achieved.
2. Identify Parameters: Based on the plots of the partial autocorrelation function (PACF) and autocorrelation function (ACF), ascertain the values of pp, dd, and qq.
3. Fit the Model: Using of the established parameters to fit the ARIMA model to thedata.
4. Evaluate Model: Evaluate the model's performance using various metrics anddiagnostic checks.

ARIMA models are famously used in many different fields for predicting the time series and have been successful in doing so.

4. Results

There are 8960 data points in the time series data shown in Table 1. This data issplit 4:1 across the training and test sets, resulting in an observed 80% -20 % split. So, before the data is changed into a more algorithm-friendly format, it is pre-processed.

3.4 Preprocessing of Data

As part of the preparation, the data underwent the following evaluates:

Verify the stationarity of the time series to determine if it is stationary.

1. Verify the coherence of the mean and standard deviation within the dataset by plotting their values across the entire data range using a rolling window. This method provides a robust approach to confirming the stability and consistency of these statistical measures throughout the dataset.

2. "The ADFuller test for stationarity assessment: The ADFuller test is used to findout how well a trend keeps up over time. This is possible by keeping a null hypothesisand an alternate hypothesis". Table 2 displays the test results.

The time series is considered nonstationary, adhering to the null hypothesis suggesting a common root. However, an alternative hypothesis proposes stationarity. The assessment involves determining the *p*-value associated with the null hypothesis, with a predetermined threshold set at 0.05. If the calculated *p*-value falls below this threshold, one may infer that the "null hypothesis holds true, indicating the series is stationary."

3. Employment of seasonal decomposition involves establishing a stationary time series for forecasting. The triad values required for this process are derived through the application of the seasonal decomposition approach. The outcomes of this method encompass three components: residuals, seasonality, and trends, as illustrated in Fig 1.

3.5 "Augmented Dickey-Fuller test"

The ADF test plays a crucial role in evaluating the stationarity of time series dataand finds widespread application in time series analysis, econometrics, and financial modeling. Its primary purpose is to ascertain the presence of a unit root in a giventime series dataset. A unit root is a characteristic of a time series that reflects a random walk behavior—characterized by a consistent mean and variance, with inherently unpredictable values[27]. In econometrics and finance, the ADF test serves as a valuable tool for assessing the stationarity of a time series. By determining the existenceof a unit root, the ADF test provides insights into whether the data exhibits a stationary or non-stationary behavior, crucial for making informed decisions in various analytical and modeling contexts [15].

Stationarity plays a crucial role in the field of time series analysis [16]. A time series is considered stationary when its statistical characteristics, such as mean and variance, remain consistent over time. In contrast, non-stationary time series exhibit patterns like trends or seasonality, introducing complexities in the analysis and modeling process.

The ADF test involves estimating the presence of a unit root in a time seriesand determining whether it can be removed through differencing (making the series stationary). The test provides a p-value, and based on this p-value, one can decide whether to reject the null hypothesis that a unit root is present. The ADF test helpsyou assess whether a time series is stationary. A low p-value (typically less than 0.05) leads to rejecting the null hypothesis and suggests that the time series is likely stationary. Conversely, a high p-value implies that the null hypothesis cannot be rejected, and the time series may be non-stationary [20]. The Augmented Dickey-Fuller (ADF) test interpretation involves assessing the p-value obtained from the test statistic. The ADF test is commonly used to determine whether a time series is stationary. Here's astep-by-step guide on how to interpret the ADF test results:

- Null Hypothesis (H0): "The null hypothesis of the ADF test is that the time serieshas a unit root, which implies that it is non-stationary".
- Alternative Hypothesis (H1): "The alternative hypothesis is that the time series does not have a unit root, indicating stationarity".
- Test Statistic and Critical Values: "The ADF test produces a test statistic and critical values. The test statistic is compared with critical values to make a decision.

If the test statistic is less than the critical value, the null hypothesis is rejected".

- P-value: "The primary measure for decision-making is the p-value. If the p-value is less than a chosen significance level (commonly 0.05), the null hypothesis is rejected.

This suggests evidence against the presence of a unit root and in favour of stationarity". If p-value < 0.05: Reject the null hypothesis. If p-value >= 0.05: Fail to rejectthe null hypothesis.

Interpretation:

- Rejecting the Null Hypothesis: If you reject the null hypothesis, it suggests that the time series is likely stationary. This is a positive outcome for many time series analyses, as stationarity simplifies modeling.
- Failing to Reject the Null Hypothesis: If you fail to reject the null hypothesis,

it implies that there is not enough evidence to conclude that the time series is stationary. In this case, the data may exhibit a unit root, indicating non-stationarity.

- Consideration of Lag Order: In some cases, the ADF test involves choosing the lag

order for the test. Different lag orders might produce different results. It's importantto consider the context of the data and potentially experiment with different lag orders to find the most appropriate one.

The time series data is visualised as shown in Figure 1. This exploration shows the relations, dependability and sheds light on what variables are changing with time and how they are changing (trends) with time. Histogram, Pair Plot. A Stationary series is one whose mean and variance do not change with time. The ADF generates a tuple consisting of 6 parameters; the ADF test statistic (TS), p-value, number of lags used, number of observations used, critical values (CV) at 1%,5%, 10%levels and themaximized information criterion (icbest).

3.6 Results of ADF

ADF Test is executed to check for stationarity. Each input value is a time series data.So when all the series are checked for stationarity, the output achieved is shown in Table 2 and Table 3.

3.7 Exploratory Data Analysis

The Data used for this work is as shown in figure 2 Before heading into the data pre- processing part, it is important to visualize what variables are changing with time and

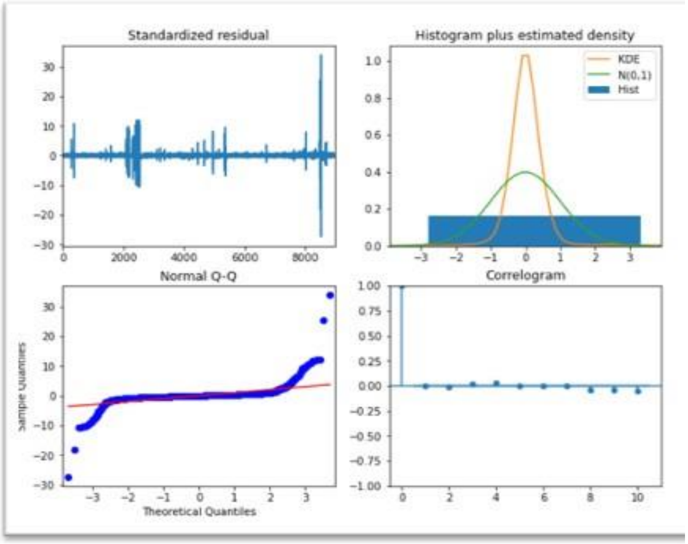

Fig. 1 The seasonal de-compose of the observations.

Table 2 Adifuller Test Statistics with lag leap 1

Multivariate Input parameters							Output
Test Parameters	IR	IY	IB	VR	VY	VB	Vibration
TS	-22.624	-22.608	-22.603	-17.441	-17.432	-17.44	-20.793
p-value	0	0	0	4.71E-30	4.76E-30	4.71E-30	0
no of Lags	33	33	33	37	37	37	37
no of Observations	8839	8839	8839	8834	8835	8835	8835
CV (1%)	-3.431	-3.431	-3.431	-3.431	-3.431	-3.431	-3.431
CV (5%)	-2.862	-2.862	-2.862	-2.862	-2.862	-2.862	-2.862
CV (10%)	-2.567	-2.567	-2.567	-2.567	-2.567	-2.567	-2.567

how they are changing (trends) with time. The Histogram which shows the distribution is as shown in Figure 3 The relation between the data with the output variable and within the input features can be well visualised using the pair plot as shown in Figure 4.

The Pierson's correlation, which plays an important role in understanding the dependency and impact of one variable on another, can be seen with the help of Figure 5.

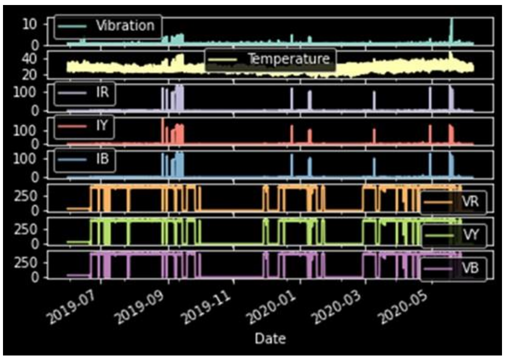

Fig. 2 Time series Data Collected over the year

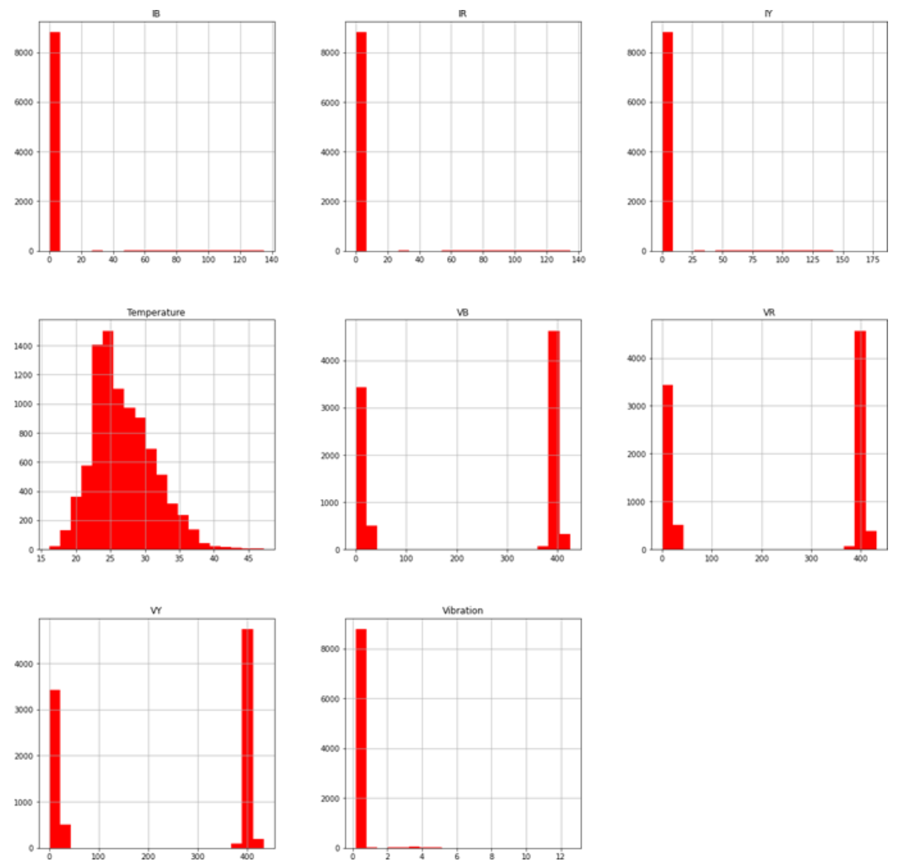

Fig. 3 Data Distribution for data visualization

Table 3 Adifuller Test Statistics with leap 2

	Multivariate Input parameters						Output
Test Parameters	IR	IY	IB	VR	VY	VB	Vibration
TS	-34.9219	-34.741	-34.833	-27.94	-27.94	-27.944	-35.923
p-value	0	0	0	0	0	0	0
no of Lags	36	36	36	37	37	37	37
no of Observations	8835	8835	8835	8834	8834	8834	8834
CV (1%)	-3.431	-3.431	-3.431	-3.431	-3.431	-3.431	-3.431
CV (5%)	-2.862	-2.862	-2.862	-2.862	-2.862	-2.862	-2.862
CV (10%)	-2.567	-2.567	-2.567	-2.567	-2.567	-2.567	-2.567

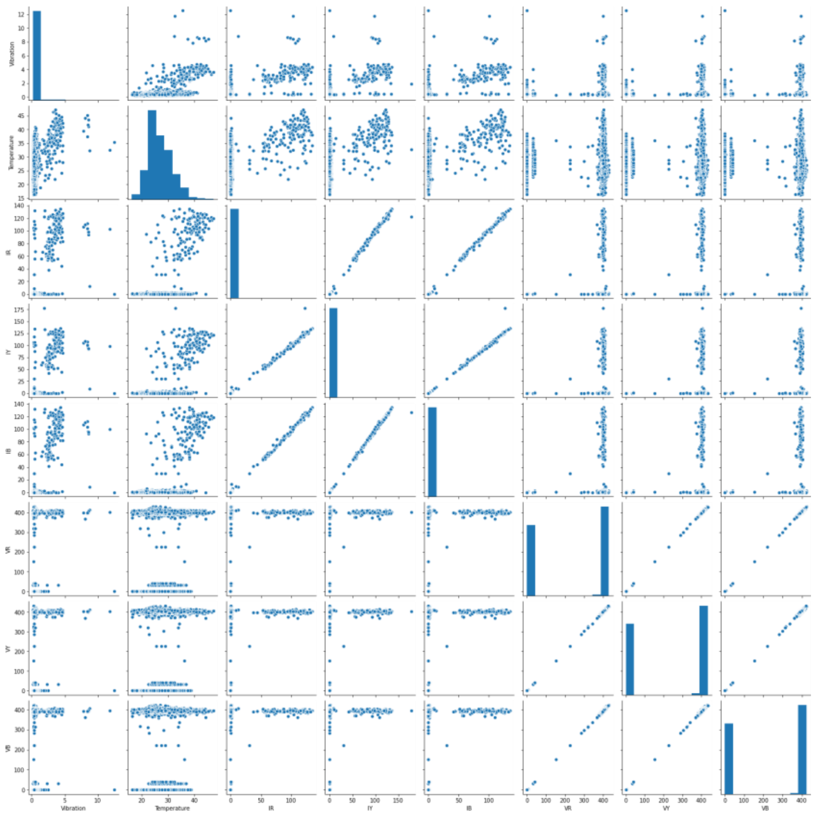

Fig. 4 Pair Plot for data visualization

3.8 Results of ARIMA

The models undergo training by utilizing data derived from the preprocessed training set, with ARIMA and VAR being trained separately. The fully trained model undertakes the task of predicting outcomes based on the test set data. The model's predictions are then evaluated by comparing them to the actual output values within the dataset. Among the two forecasting methods, ARIMA and VAR, the former demonstrates a notably higher degree of line fitting, aligning closely with others, while the latter exhibits the least degree of line fitting.

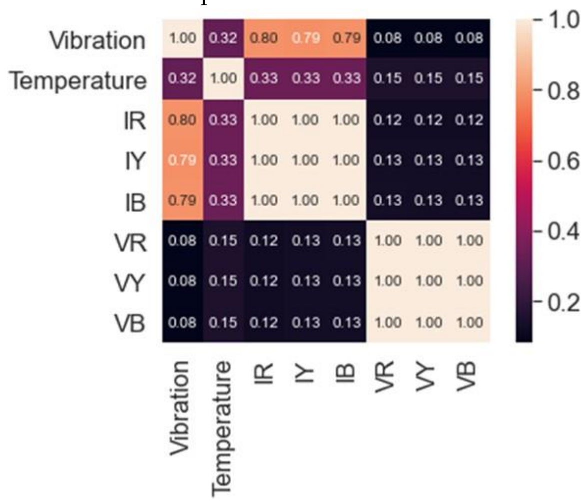

Fig. 5 Correlation Matrix for data visualization

The aforementioned model is evaluated across multiple time intervals as Hourly, Daily, Weekly, Monthly, and Yearly.

The results are evaluated across the following three metrics:

- "Mean Absolute Errors (MAE): Absolute difference between actual and predicted traffic".
- "Mean Square Error (MSE): Square of the difference between actual and predicted traffic".

• “Root Mean Squared Error (RMSE): Square root of the MSE value”.

The results are mentioned in the tables below:

To enhance result interpretability, visual representations in the form of graphs illustrating the comparison between predicted and actual outputs for the implementedalgorithms are presented in Fig 6 to Fig.7

3.9 Results of VAR

Fitting the VAR Model: The VAR model is fitted to the second differenced data usinga maximum lag of 15 and AIC as the information criterion.

• Forecasting: Predictions are made for 100 steps ahead using the fitted VAR model.

• Inverting the Differencing Transformation: The function invert transformation is defined to revert the differencing transformation applied to the forecasted data.

• Visualization: The actual vs forecasted values are plotted for each sensor data(columns in the dataset). The color for the forecasted and actual plots is shown in Figure.

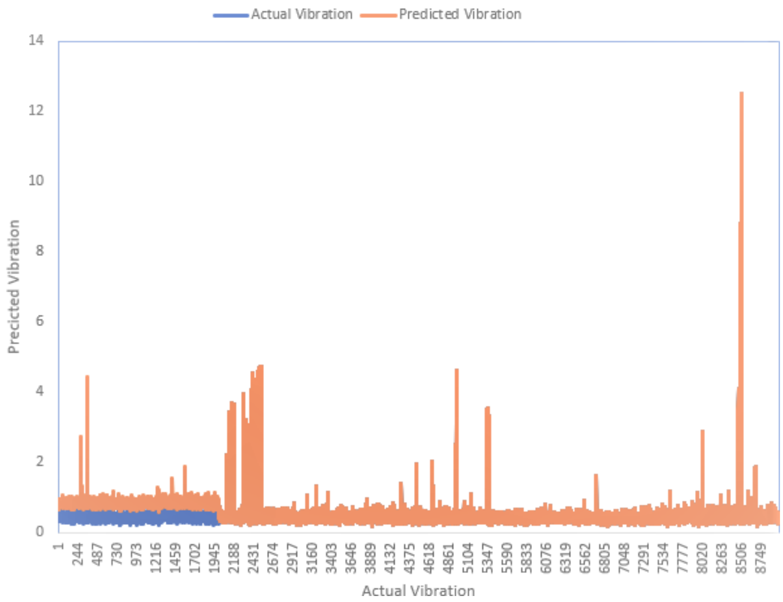

Fig. 6 Actual vs Predicted vibrations using ARIMA algorithm

• Mean Squared Error (MSE): The MSE between the test data and the forecastedvalues is calculated using mean squared error from the sklearn.metrics module.

Step : Fitting the VAR model to the 2nd Differenced Data and Forecasting for 100 steps ahead the MSE value is obtained as Mean Square Error:3.655

5. Conclusion

In conclusion, our comparative analysis of ARIMA and VAR algorithms for predicting vibration in electrical systems reveals valuable insights into their respective performances. Over the course of a year-long dataset collected at various intervals, ARIMA consistently demonstrates superior forecasting accuracy compared to VAR. This superiority is evident across multiple evaluation metrics, including Mean Absolute Error(MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The results emphasize the importance of algorithm selection in time series forecasting, particularly in applications where precision in predicting vibration levels is critical for maintaining the reliability and efficiency of electrical systems. ARIMA’s effectiveness is attributed to its ability to capture and model the underlying temporal patterns present in the dataset, showcasing its suitability for such applications. While VAR is a powerful tool for capturing inter dependencies among multiple time series variables, our findings suggest that for predicting vibration in electrical systems, the univariate approach of ARIMA proves more advantageous.

Our research contributes to the existing body of knowledge by providing empirical evidence supporting the preferential use of ARIMA in scenarios similar to our study.

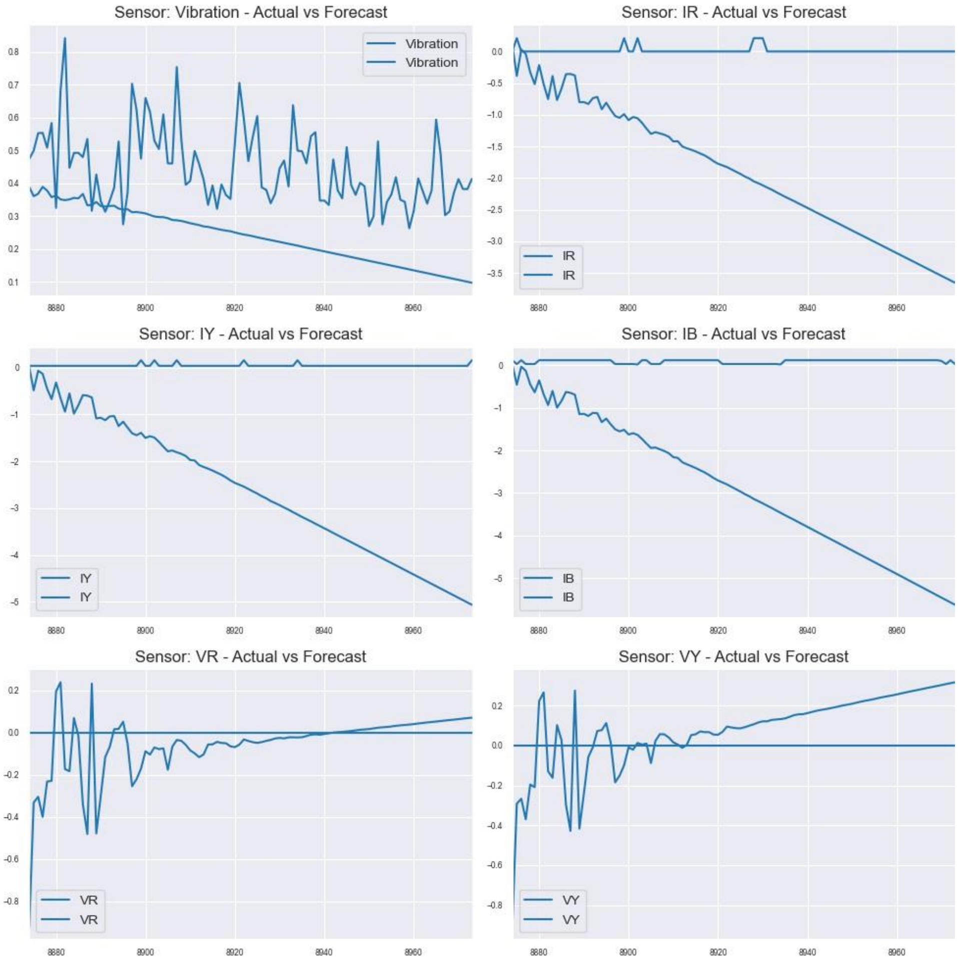

Fig. 7 Actual vs Predicted vibrations using VAR algorithm

However, it is essential to note that the effectiveness of forecasting algorithms is context-dependent, and results may vary based on the nature and characteristics of the dataset.

As the field of time series forecasting continues to evolve, further research can explore hybrid models or advanced machine learning techniques to enhance predictive capabilities. Nonetheless, the insights gained from this study serve as a valuable guide for practitioners and researchers in making informed decisions when selecting forecasting algorithms for similar applications in HSD pumps.

References

[1] Bilal, M., Kim, H., Fayaz, M., Pawar, P.: Comparative analysis of time series forecasting approaches for household electricity consumption prediction. arXiv preprint arXiv:2207.01019 (2022)

[2] Sufian, M.A., Varadarajan, J.: Enhancing prediction and analysis of uk road traffic accident severity using ai: Integration of machine learning, econometric techniques, and time series forecasting in public health research. arXiv preprint arXiv:2309.13483 (2023)

[3] Wang, J., Shen, L.-z., Bi, Y., Lei, J.: Modeling and optimization of a light-duty diesel engine at high altitude with a support vector machine and a genetic algo- rithm. Fuel **285**, 119137 (2021) https://doi.org/10.1016/J.FUEL.2020.119137

[4] Prediction of diesel engine performance using support vector regression technique. Regular (2020) https://doi.org/10.35940/ijitee.j7494.0891020

[5] Watson, D., Bostic, D., Harrison, T.V.: Comparative models of hydrocarbon emis- sions for a diesel engine operating at constant loads and speeds. Transactions of the ASABE **52**, 1079–1087 (2009) https://doi.org/10.13031/2013.27778

[6] Mahajan, S., Harikrishnan, R., Kotecha, K.: Prediction of network traffic in wire-less mesh networks using hybrid deep learning model. IEEE Access **PP**, 1–1 (2022) https://doi.org/10.1109/ACCESS.2022.3140646

[7] Chato, L., Tayeb, S., Latifi, S.: A genetic algorithm to optimize the adaptive support vector regression model for forecasting the

reliability of diesel engine systems. 2017 IEEE 7th Annual Computing and Communication Workshop andConference (CCWC), 1–7 (2017) https://doi.org/10.1109/CCWC.2017.7868462

[8] Fan, W.: Prediction and evaluation of sound quality of diesel engines based on neural network. Noise and Vibration Control (2014)

[9] Tosun, E., Aydın, K., Bilgili, M.: Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. alexandria engineering journal **55**, 3081–3089 (2016) https://doi.org/10.1016/J.AEJ.2016.08.011

[10] Qing-hua, F.: Regression analysis of the curves of the flow and lift when regulating the speed of centrifugal pumps. Chemical Engineering Machinery (2010)

[11] Ağbulut, , Gürel, A., Sarıdemir, S.: Experimental investigation and prediction of performance and emission responses of a ci engine fuelled with different metal-oxide based nanoparticles–diesel blends using different machine learning algorithms. Energy **215**, 119076 (2021) https://doi.org/10.1016/j.energy.2020. 119076

[12] Moriyasu, R., Nojiri, S., Matsunaga, A., Nakamura, T., Jimbo, T.: Diesel engine air path control based on neural approximation of nonlinear mpc. Control Engi- neering Practice (2019) https://doi.org/10.1016/J.CONENGPRAC.201 9.104114

[13] Younis, Y., Abbas, S., Najim, F.T., Kamar, F.H., Nechifor, G.: Comparison of an artificial neural network and a multiple linear regression in predicting the heat of combustion of diesel fuel based on hydrocarbon groups. Revista De Chimie **71**, 66–74 (2020) https://doi.org/10.37358/rc.20.6.8171

[14] Zhang, R., Hu, J.: Production performance forecasting method based on mul- tivariate time series and vector autoregressive machine learning model for waterflooding reservoirs. Petroleum Exploration and Development **48**(1), 201–211 (2021)

[15] Katris, C.: Unemployment and covid-19 impact in greece: A vector autoregression (var) data analysis. Engineering Proceedings **5**(1), 41 (2021)

[16] Rajab, K., Kamalov, F., Cherukuri, A.K.: Forecasting covid-19: vector autoregression-based model. Arabian journal for science and engineering, 1–10 (2022)

[17] Sethi, J.K., Mittal, M.: Analysis of air quality using univariate and multivariate time series models. In: 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 823–827 (2020). IEEE

[18] Montiel Olea, J.L., Stock, J.H., Watson, M.W., *et al.*: Inference in struc- tural vector autoregressions identified with an external instrument. Journal of Econometrics **225**(1), 74–87 (2021)

[19] Adolf, J.K., Loossens, T., Tuerlinckx, F., Ceulemans, E.: Optimal sampling rates for reliable continuous-time first-order autoregressive and vector autoregressive modeling. Psychological Methods **26**(6), 701 (2021)

[20] Heaps, S.E.: Enforcing stationarity through the prior in vector autoregressions. Journal of Computational and Graphical Statistics **32**(1), 74–83 (2023)

[21] Ding, P., Jia, M., Yan, X.: Stationary subspaces-vector autoregressive with exoge- nous terms methodology for degradation trend estimation of rolling and slewing bearings. Mechanical Systems and Signal Processing **150**, 107293 (2021)

[22] Winarno, S., Usman, M., Kurniasari, D., *et al.*: Application of vector error cor- rection model (vecm) and impulse response function for daily stock prices. In: Journal of Physics: Conference Series, vol. 1751, p. 012016 (2021). IOP Publishing

[23] Al-Qazzaz, R.A., Yousif, S.A.: High performance time series models using auto autoregressive integrated moving average. Indones. J. Electr. Eng. Comput. Sci **27**, 422–430 (2022)

[24] Elakkiya, E., Radha, M., Sathy, R.: Application of arima model for predicting cashew nut production in india–an analysis. Int. J. Res. Bus. Manage. **5**, 45–52 (2017)

[25] Mandrikova, O., Fetisova, N., Polozov, Y.: Hybrid model for time series of complex structure with arima components. Mathematics **9**(10), 1122 (2021)

[26] Unnikrishnan, P., Jothiprakash, V.: Hybrid ssa-arima-ann model for forecasting daily rainfall. Water Resources Management **34**, 3609–3623

(2020)

[27] Espasa, A., Peńa, D.: The decomposition of forecast in seasonal arima models. Journal of Forecasting **14**(7), 565–583 (1995)

[28] Khetani, Vinit, et al. "Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains." International Journal of Intelligent Systems and Applications in Engineering 11.7s (2023): 253-262.

[29] Mane, Deepak, et al. "Traffic Density Classification for Multiclass Vehicles Using Customized Convolutional Neural Network for Smart City." Communication and Intelligent Systems: Proceedings of ICCIS 2021. Singapore: Springer Nature Singapore, 2022. 1015-1030.

[30] A. Bhujbal and D. Mane, "A survey on deep learning approaches for vehicle and number plate detection," Int. J. Sci. Technol. Res., vol. 8, no. 12, pp. 1378–1383, 2019.