As we know the liquor sales is one of the worlds biggest business market all around the world where it can generate a large impact on the revenue .

as per the studies and data people all around the world use different types of liquor and each country is getting taxes and other benefits through the sales.

Lowa Liquor is a retail store that specializes in selling various types of alcoholic beverages.

The goal of the business is to identify the factors causing the decline in sales and develop strategies to increase sales revenue.

The store may be implementing general strategies such as marketing and promotion campaigns, but there is no evidence that these strategies are effective.

The team is concerned about the reasons behind this decline and wants to identify the factors contributing to it.

Considering all these factors we can analyse how to increase the sales of Lowa Liquor by proper marketing and personalised advertisements.

The key is to use data analysis and machine learning techniques to identify the factors contributing to the decline in sales and develop strategies to improve sales revenue.

This may involve analyzing sales data and customer behavior, identifying patterns and trends, and using this information to make datadriven decisions about pricing, product mix, promotions, and inventory management.

When fitting multiple regression models, there are several important points to note:

Variable selection: Carefully choose the independent variables (predictors) to include in the model. Consider their relevance and potential impact on the dependent variable.

Multicollinearity: Check for multicollinearity, which occurs when independent variables are highly correlated with each other. Multicollinearity can affect the stability and interpretability of the regression coefficients.

Model evaluation: Assess the goodness of fit of the model using appropriate metrics such as R-squared, adjusted R-squared, and p-values of the coefficients. These metrics help evaluate how well the model explains the variation in the dependent variable and the statistical significance of the independent variables.

Cross-validation and overfitting: Guard against overfitting, which occurs when the model is too complex and performs well on the training data but fails to generalize to new data. Use techniques like cross-validation to estimate the model's performance on unseen data and ensure its reliability.

Model validation: Validate the final model using independent data or by applying it to new observations. This helps assess the model's performance and its ability to generalize beyond the original dataset.

By considering these points, we can effectively fit and evaluate multiple regression models, providing valuable insights into the relationships between variables and predicting the value of the dependent variable.

So we have built almost 15 different models using hyper parameter and gridsearchcv .

Like model names .

Based on the above information, **the random forest regression model** seems to have achieved excellent performance based on several evaluation metrics:

**1. R-squared**: The R-squared value of **0.99916** indicates that the model explains approximately 99.916% of the variance in the target variable. A high R-squared value suggests that the model fits the data very well, and the majority of the variability in the target variable is captured by the model.

**2. Adjusted R-squared**: The adjusted R-squared value of **0.99916** is identical to the R-squared value in this case. This suggests that the model contains no unnecessary variables or overfitting issues, as the adjusted R-squared is usually lower than the R-squared when there are excessive variables in the model.

**3. Test RMSE**: The test root mean squared error (RMSE) of **0.0332** indicates that, on average, the model's predictions have an error of approximately 0.0332 units when applied to unseen test data. A lower RMSE value suggests better predictive performance, so the provided RMSE value is relatively low.

**4. Train RMSE**: The train RMSE of **0.0289** represents the average error of the model's predictions on the training data. A lower train RMSE suggests that the model is fitting the training data well, with small discrepancies between the predicted values and the actual values.

**5. MAPE**: The Mean Absolute Percentage Error (MAPE) of **0.163** indicates the average percentage difference between the predicted and actual values. A lower MAPE indicates better accuracy, and the provided MAPE value is relatively low.

Based on these metrics, the random forest regression model appears to be performing exceptionally well, demonstrating a high level of accuracy and predictive power.

Hence **RANDOM FOREST REGRESSOR** is predicting **‘Volume sold litre’** well compared to other Regression model.