Sales Forecast for BNK Cafe Using Machine Learning Algorithms

Gabriel Natividad Jr., Mark Winfred Turco, Joshua Villanueva, Nhilbert Jay Valente, and Ara Abigail Ambita

Division of Physical Sciences and Mathematics
University of the Philippines Visayas
Miagao, Iloilo, Philippines
{gfnatividad, maturco, nevalente, jcvillanueva5}@up.edu.ph

This study explores the use of machine learning algorithms to forecast daily sales for BNK Cafe, leveraging two years of historical sales data. The research evaluates the performance of four machine learning models—Support Vector Machine (SVM), Linear Regression, Random Forest, and XGBoost—on weekly and monthly resampled datasets. Feature engineering techniques were applied to extract temporal patterns, while model performance was assessed using RMSE, MAPE, and R^2 metrics. XGBoost emerged as the most accurate model, excelling in capturing complex trends and seasonal variations, followed by Random Forest. Linear Regression served as a baseline, while SVM showed moderate success. Findings highlight the potential of advanced machine learning methods in improving inventory management, waste reduction, and operational efficiency in the food and beverage industry. Future work suggests incorporating contextual data and extending the analysis to multi-year trends for enhanced predictive accuracy.

1. Introduction

Running a restaurant deals with a lot of area that needs to be managed, the employees, customer satisfaction, the over-all running cost and especially the preparation of raw ingredients because this is one of the key factors in producing good quality food. Early Morning preparation such as buying and processing of ingredients is a vital part of running a restaurant. A restaurant that manages ingredient supply does not only improve product quality and customer satisfaction but can also reduce food waste and spoilage and in return increase profitability (Foti, 2024). However, over prepping or too much stock may lead to food and money waste while under prepping usually impacts the image of the restaurant. With the use of Machine Learning Algorithms, prediction of sales and trends can be evaluated properly, According to Madein Team (2023), "Forecasting involves the use of past and present data to make an informative prediction" which will be helpful for owners and managers plan and strategize ahead making restaurant operations run more efficiently.

According to Seyedan and Mefekheri in their paper entitled “ Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities “ Businesses that employ forecasting are more open and flexible to changes that may arise to trends. In contrast, without forecasting, Businesses are slower to adapt to the ever changing market demands leading to missed opportunities and profits. Predicting what days or months are peak season for customers of a business that would enable owners and managers to prepare more efficiently.

1.1. Literature Review

In the past, companies mostly relied on intuition and conventional techniques in decision making, which often resulted in uneven and less-than-ideal results and outcomes. Recently, machine learning has emerged as a groundbreaking solution for predicting sales and providing insights that are data-based, which greatly improves accuracy and efficiency. Instead of relying on speculation, machine learning models utilize historical sales data to forecast and predict demand trends which allows businesses to adjust and fine-tune their operations and enhance their decision-making (Smolic, 2024). By effectively anticipating sales peak period, companies can modify and change their inventory, staffing levels, waste mitigation, and customer satisfaction. The transition from instinct-based decision making to data-driven and data informed strategies helps companies to be equipped for better growth and competitiveness.

Nowadays, our time is undergoing a profound and rapid change that is powered by our technology; Digitalization, information, communications technology, and most of all, artificial intelligence. With this in mind, many researchers in the field of business and economics have suggested an approach utilizing artificial intelligence for decision-making. This is evident in retail industries that advance technologies which enable corporations to make accurate sales predictions and make reliable and appropriate decisions in order to optimize their business resources. One of these models is the Support Vector Regression (SVR) Machine Learning model. To enhance inventory management in pharmaceutical chain enterprises, accurate drug sales forecasting is essential. Utilizing the SVR machine learning model, the researchers achieved high-accuracy predictions of drug sales. Considering the impact of promotional strategies, they incorporated promotion factors into the SVR model. Experiments using real sales data from a major chain pharmacy (Company S) revealed that the SVR model with promotion factors achieved a 91% accuracy rate, significantly outperforming traditional time series models in forecasting drug sales (Liu et al., 2021). Another one of these machine learning methods used in practice is XGBoost, that is one technique best suited for forecasting. Its success as an AI includes its scalability and efficiency of running ten times faster than other existing machine learning algorithms. To demonstrate this, the researchers have compared it to the publicly available Walmart retail dataset from a kaggle competition that spans 1,913 days of sales data from multiple states. The goal is to forecast daily sales for the next 28 days based on historical data. Through feature engineering, they have identified the most relevant features for sales and input them into the model for forecasting. The model, a highly efficient and flexible gradient boosting tree model, excels in both classification and regression tasks. The results of the experiment have shown that the XGBoost-based model achieves superior performance, offering faster computation and higher prediction accuracy (Dairu & Zhang, 2021).

2. Data and Methodology

2.1. Data

Dataset

The data used throughout this project is the gathered data from the sales of BNK Cafe starting from 1st of January, 2021 till the 31st of December, 2022.

Data Fields

- **Date**. Indicates the complete detail of the year, month, and day that the sales were documented
- **Sales**. Signifies the amount of profit and revenue gathered on that specific date.

Caveats

The data is wholly dependent on the sales document that was gathered by the BNK Cafe in the span of two years starting from the year 2022 till the end of 2023.

1. **Limited Data Fields**
    - The dataset contains only two columns: Date and Sales. There are no additional features such as:
        - External factors (e.g., weather, holidays, promotions).
        - Economic indicators or demographic data.

- This restricts the model to rely solely on temporal patterns, which may lead to limited predictive power.

2. **Lack of Contextual Data**
    - Without details such as product categories, locations, or customer segments, it is impossible to discern the drivers of sales trends.
    - Changes in sales might be due to external factors (e.g., marketing campaigns, weather changes) that are not captured.
3. **Short Time Frame**
    - With only two years of data, seasonal patterns (e.g., yearly trends) may not be fully captured. A longer time frame would provide better insights into multi-year trends and cyclicality.
4. **Lack of Post-Pandemic Recovery Trends**
    - If this dataset spans 2021–2022, it reflects post-pandemic sales. It may not accurately capture trends before or during the pandemic, leading to biased conclusions.

2.2. Methodology

The methodology employed for this research focuses on sales for BNK Cafe using machine learning algorithms. The dataset spans two years, with daily sales data provided. This study utilizes various regression algorithms, including Support Vector Machine (SVM), Linear Regression, Random Forest, and XGBoost, to predict sales based on engineered time-series features. The implementation and evaluation pipeline includes data preprocessing, feature engineering, model training, hyperparameter tuning, evaluation, and visualization.

2.2.1. Data Description

The dataset comprises 730 daily entries of sales data with the following columns:

- Date: The date of the recorded sales (in DD/MM/YYYY format).
- Sales: The daily sales values in integer format.

The dataset was used exclusively for supervised machine learning, with *Sales* as the target variable. The absence of additional contextual data necessitated reliance on time-based features extracted from the *Date* column.

2.2.2. Data Preprocessing

- **Date Parsing**: The Date column was converted into Python's datetime format to enable time-series analysis.
- **Missing Values**: No missing values were observed in the dataset, eliminating the need for imputation.
- **Feature Engineering**: New features were derived from the Date column to capture temporal patterns:
  - **Month**: Numerical representation of the month (1–12).
  - **Year**: Extracted year of the entry.
  - **Week**: ISO calendar week of the year.
  - **Lagged Sales**: Previous day's sales value, used as an additional predictor.
- **Normalization**: The Sales column was normalized using Min-Max scaling to ensure uniform feature ranges, particularly for models sensitive to feature magnitudes (e.g., SVM).

2.2.3. Resampling of Data

The original dataset provided daily sales data for a span of two years. To better capture temporal patterns and reduce noise, the data was resampled into **weekly** and **monthly aggregates**. These resampling steps served the following purposes:

- Weekly Resampling: Highlighted short-term trends and cyclicality, such as weekly peaks and dips.
- Monthly Resampling: Helped capture long-term sales trends, smoothing out day-to-day fluctuations while preserving seasonality.

By resampling the data, the aggregated data allows the models to capture meaningful trends and seasonality without being overwhelmed by day-to-day noise. Additionally, weekly and monthly forecasts align better with business decision-making cycles, making them more practical for the. The resampled datasets were then used for exploratory data analysis and model training, allowing for comparisons between models trained on different granularities.

2.2.4. Model Selection and Training

Four machine learning algorithms were implemented to evaluate their predictive performance:

- Support Vector Machine (SVM):
  - Utilized for its ability to capture non-linear relationships in data.
  - A linear kernel was selected for initial trials, with hyperparameters tuned using GridSearchCV.
- Linear Regression:
  - Chosen as a baseline model due to its simplicity and interpretability.
  - The model was trained on normalized features without additional tuning.
- Random Forest:
  - Employed for its ability to handle non-linearities and high variance in data.
  - GridSearchCV was used to optimize hyperparameters such as the number of trees, maximum depth, and minimum samples per split.
- XGBoost:
  - Selected for its superior performance in gradient-boosted regression tasks.
  - Hyperparameter tuning was conducted using GridSearchCV with an extensive search over learning rate, maximum depth, and number of estimators.

2.2.5. Train-Test Split

The dataset was split into training (80%) and testing (20%) subsets based on chronological order to preserve the temporal structure of the data. This ensured that future dates in the test set were not influenced by their presence in the training set, maintaining the integrity of the forecasting exercise.

2.2.6. Model Evaluation

Models were evaluated using the following metrics:

- **Root Mean Square Error (RMSE)**: Assessed the magnitude of prediction errors.
- **Mean Absolute Percentage Error (MAPE)**: Measured the model's predictive accuracy in percentage terms.
- **Coefficient of Determination or R-Squared (R^2):** Shows how well the data fit with the Regression model. This also quantifies how well the model explains the variability in the target variable (sales).

Separate evaluations were conducted on both the training and testing sets to identify overfitting or underfitting.

2.2.7. Software and Tools

The entire analysis was conducted using Python, with the following libraries:

- **pandas**: Data manipulation and analysis.
- **scikit-learn**: Model training, evaluation, and hyperparameter tuning.
- **xgboost**: Implementation of the XGBoost regressor.
- **matplotlib**: Visualization of results.

2.3. Evaluating Forecast Accuracy

The performance of the forecasting models was evaluated using two widely adopted metrics: Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). These metrics provide insights into both the absolute and relative accuracy of the models.

1. **Root Mean Square Error (RMSE)**: RMSE measures the magnitude of the prediction errors. It is computed as:

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

where y_i represents the actual sales values, $\widehat{y_i}$ represents the predicted sales, and n is the number of observations. RMSE penalizes larger errors more heavily, making it sensitive to outliers. Lower RMSE values indicate better model performance.

2. **Mean Absolute Percentage Error (MAPE)**: MAPE measures the average percentage error between predicted and actual values, making it a scale-independent metric. It is calculated as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} |\frac{y_i - \widehat{y_i}}{y_i}| \times 100$$

MAPE provides a clear interpretation of model accuracy in percentage terms, with lower values indicating better performance.

3. BNK Cafe Sales Forecast Model

3.1. Support Vector Regression(SVM) Model

SVM was chosen for its strength in modeling non-linear relationships, an essential requirement given the temporal nature of sales data. Its versatility, particularly with kernel methods, enabled it to uncover complex interactions between features such as lagged sales, time, and seasonal indicators. By employing a grid search to fine-tune parameters like C, epsilon, and kernel, the model was optimized to balance prediction accuracy and generalizability. This makes SVM a useful tool for forecasting, especially when the data has patterns that aren't purely straightforward or linear.

The Support Vector Machine (SVM) model provided reliable predictions for both the weekly and monthly datasets. Its outputs revealed a strong ability to track overall trends in sales, maintaining close alignment with the actual values. However, the model exhibited some limitations in responding to abrupt spikes or dips in sales, particularly in the weekly dataset. This is likely due to its optimization approach, which prioritizes minimizing errors within a predefined margin set by the epsilon parameter.

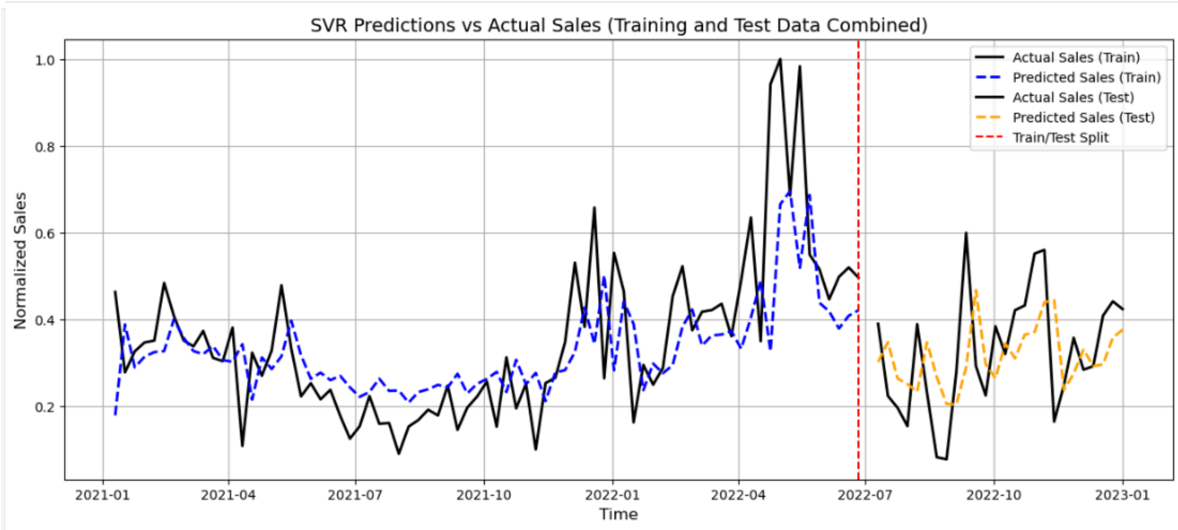

Fig 1. [Weekly] SVR Prediction vs Actual Sales

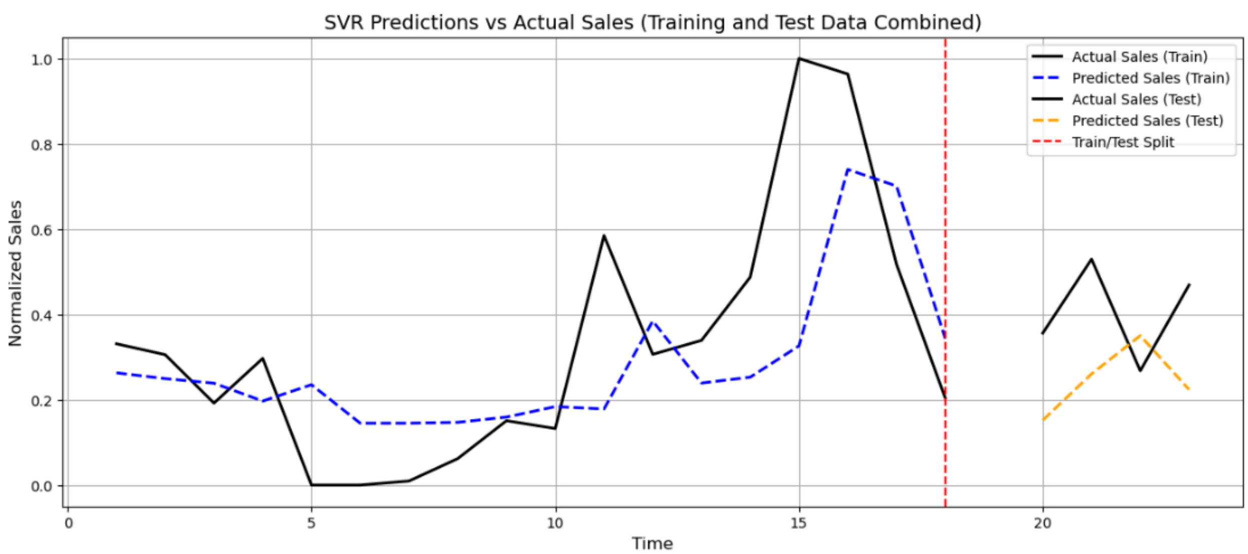

Fig 2. [Monthly] SVR Prediction vs Actual Sales

Moreover, the use of a polynomial kernel, which was identified through hyperparameter tuning, allowed the SVM to effectively capture nonlinear patterns present in the sales data. While this kernel increased the model's flexibility, the inherent trade-off between bias and variance limited its responsiveness to sudden changes which is a common challenge for SVM in time-series tasks.

3.2. Linear Regression

Linear regression was included as a benchmark due to its simplicity and interpretability. By providing a clear view of how basic linear relationships explain sales trends, it offered a foundation for evaluating the incremental benefits of more sophisticated models. Despite its shortcomings with non-linear patterns, its ease of implementation and ability to provide immediate insights justified its role in this study.

As a baseline model, linear regression offered a straightforward yet insightful perspective on the sales data. The model was effective in capturing long-term trends, particularly in the monthly dataset, where sales patterns are smoother and more consistent. However, its simplicity became a limitation when applied to the weekly dataset, where sharp fluctuations and irregular peaks were less accurately predicted.

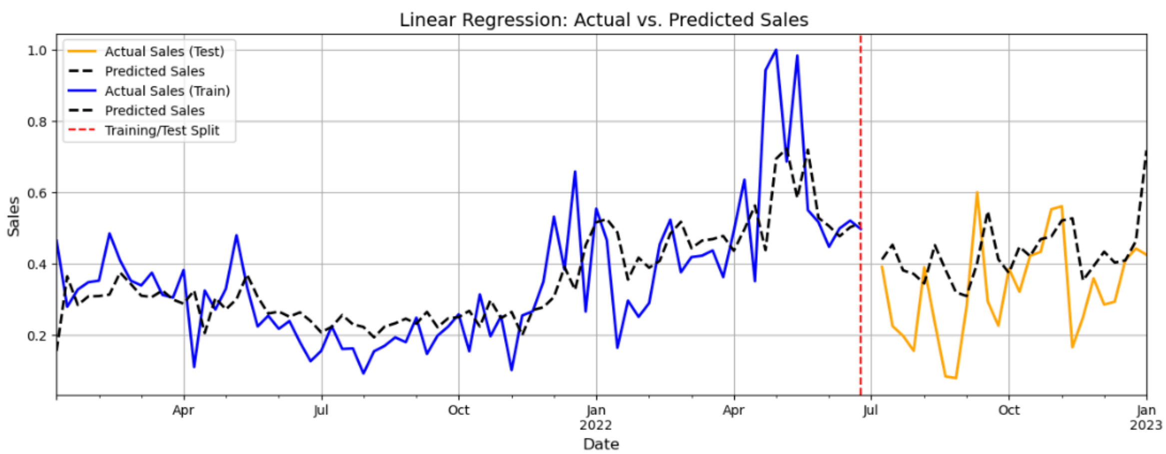

Fig 3. [Weekly] Linear Regression Prediction vs Actual Sales

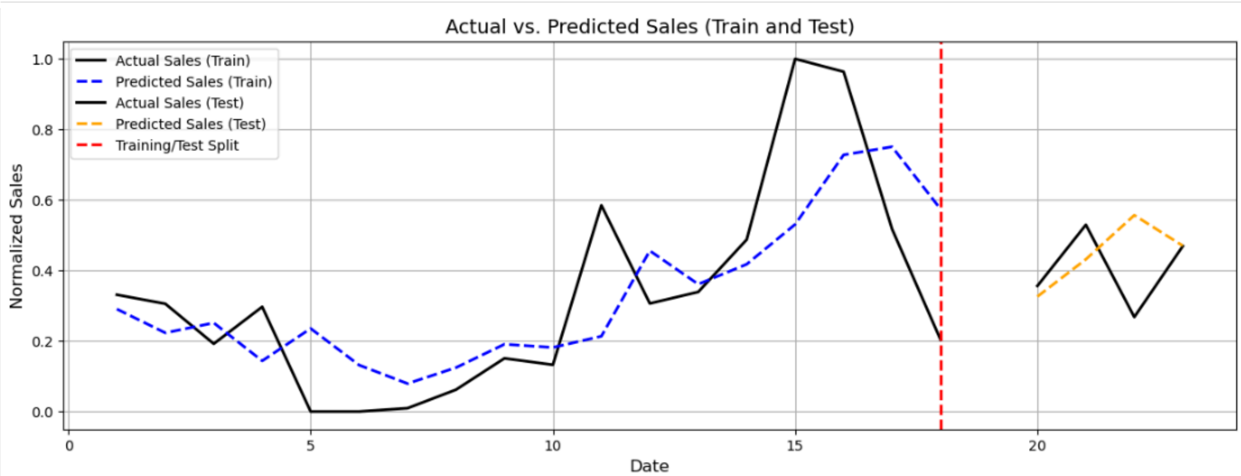

Fig 4. [Monthly] Linear Regression: Prediction vs Actual Sales

3.3. Random Forest (RF) Model

Random Forest was selected for its proven ability to manage complex, non-linear datasets. Its ensemble nature not only enhanced prediction stability but also reduced the risk of overfitting, making it particularly suitable for forecasting tasks. Hyperparameter tuning, including adjustments to the number of trees, maximum depth, and minimum samples per split, further optimized its performance. This balance of flexibility and reliability makes Random Forest a highly effective model for time-series forecasting.

The Random Forest model delivered robust and consistent predictions, excelling in its ability to capture the complex dynamics of both weekly and monthly sales data. For the weekly dataset, the model effectively accounted for subtle trends and seasonal patterns, providing predictions that closely mirrored the actual sales values. Similarly, in the monthly dataset, Random Forest performed well in modeling broader, long-term trends while preserving predictive accuracy for individual periods.

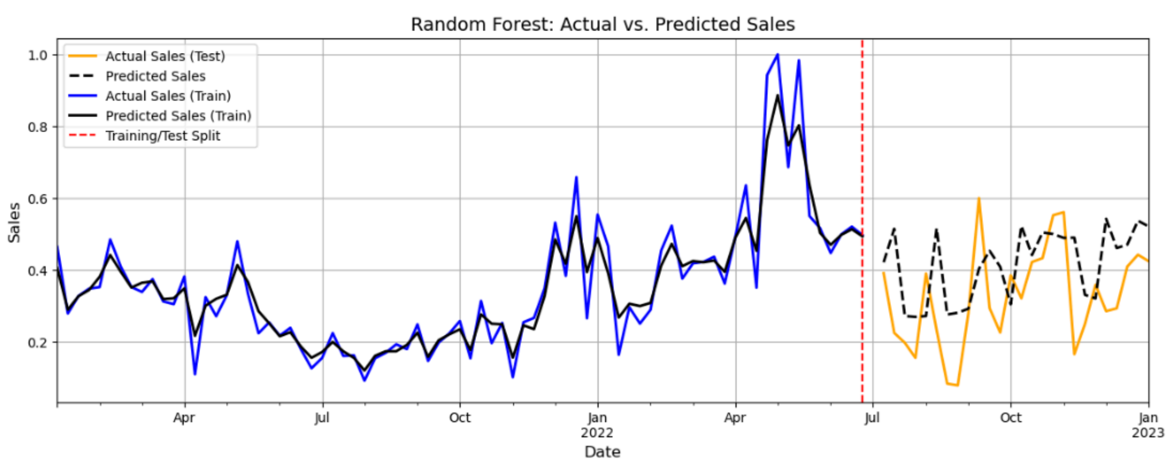

Fig 5. [Weekly] Random Forest Prediction vs Actual Sales

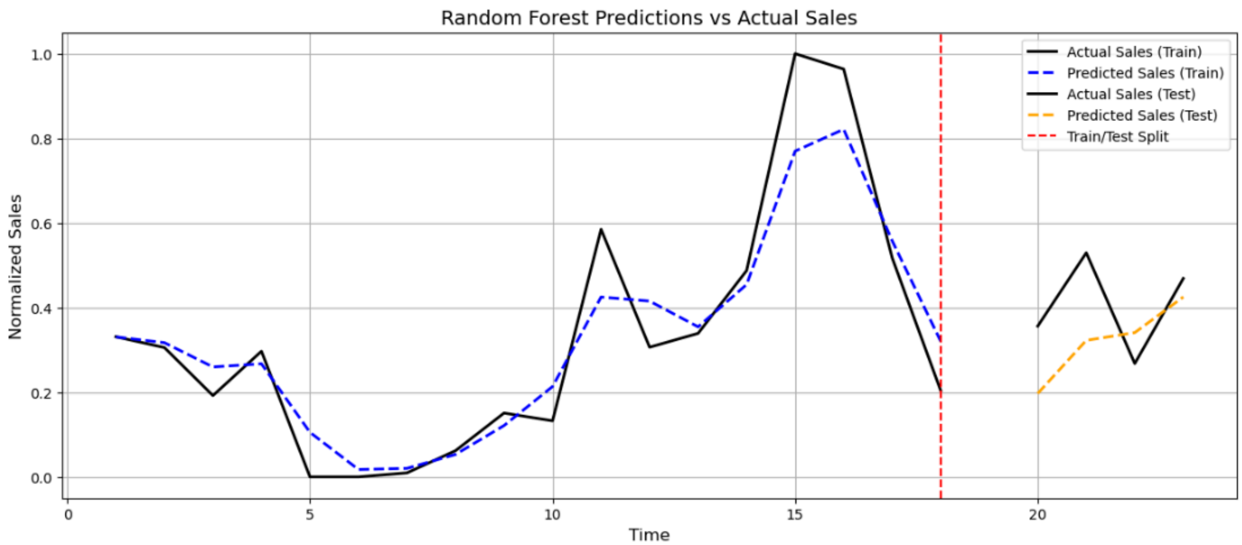

Fig 6. [Monthly] Random Forest Prediction vs Actual Sales

The model's ensemble approach, which aggregates the predictions of multiple decision trees, allowed it to handle the inherent variability and noise in the sales data. By leveraging features such as lagged sales and seasonal indicators, it captured both short-term and long-term patterns with high precision. The comparable RMSE and MAPE values for training and testing datasets also underscored its ability to generalize without overfitting.

3.4. XG Boost Model

XGBoost was chosen for its state-of-the-art capabilities in regression tasks, particularly in datasets with high variability and non-linearity. Its ability to effectively handle noise and interactions between features makes it an excellent choice for time-series forecasting. By fine-tuning its parameters through grid search, XGBoost was optimized to deliver high precision while avoiding overfitting. This model's performance underscores its potential as a leading tool for forecasting applications in dynamic environments.

XGBoost emerged as the most accurate and reliable model in this study, delivering superior performance across both the weekly and monthly datasets. Its predictions closely matched actual sales values, especially during periods of sharp changes in the weekly dataset, where other models struggled. For the monthly dataset, XGBoost effectively captured seasonal trends and broader sales patterns, showcasing its strength in handling complex and high-dimensional data.

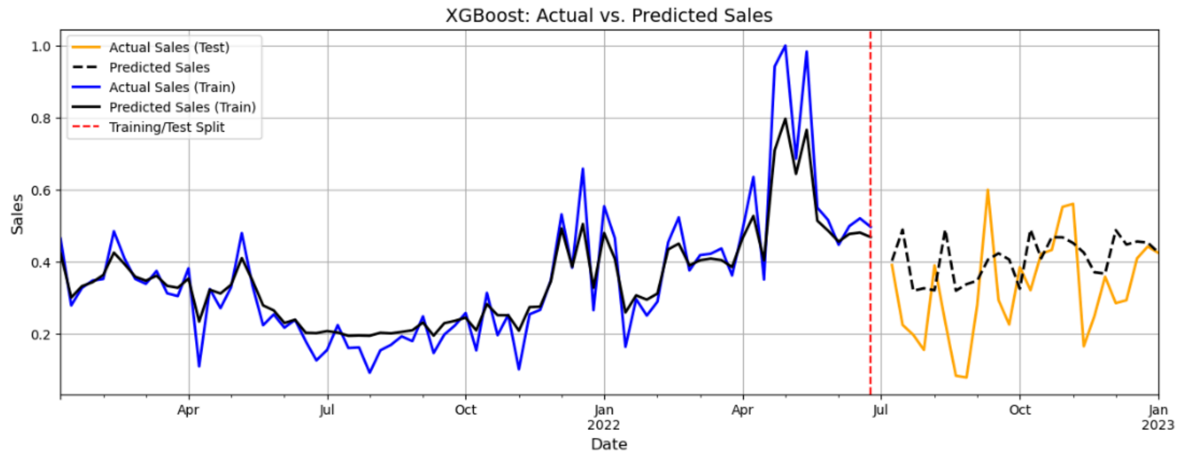

Fig 7. [Weekly] XG Boost Prediction vs Actual Sales

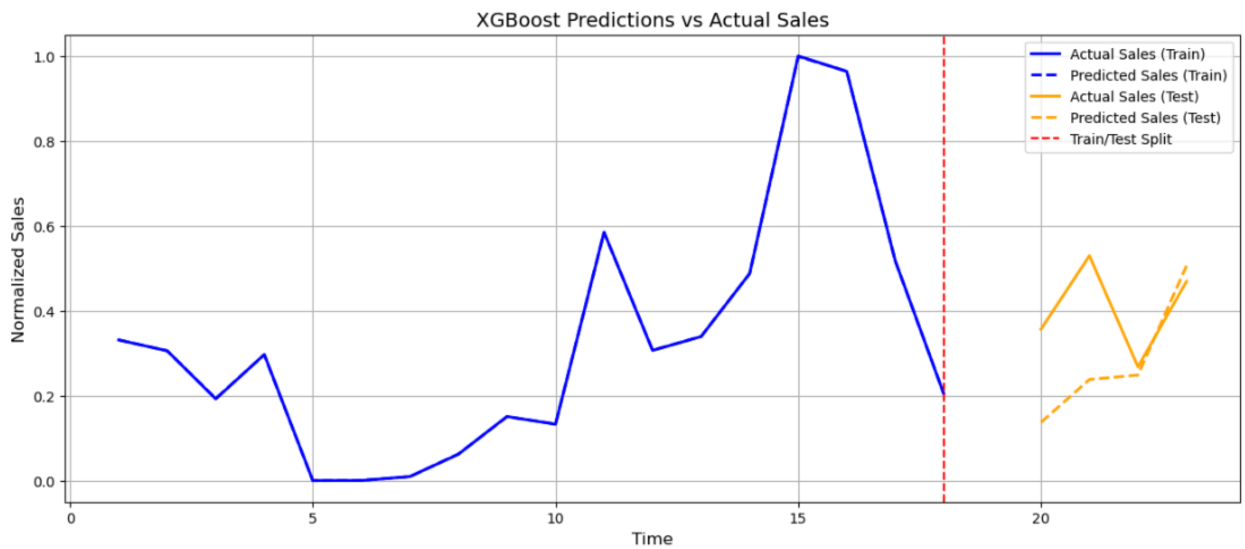

Fig 8. [Monthly] XG Boost Prediction vs Actual Sales

The success of XGBoost can be attributed to its advanced gradient boosting framework, which iteratively minimizes errors by combining predictions from multiple weak learners. This approach allowed it to adaptively correct its shortcomings with each iteration, leading to highly accurate predictions. The extensive hyperparameter tuning conducted during training ensured optimal settings for factors such as learning rate, maximum depth, and the number of estimators.

4. Results and Discussion

The results of the forecasting models are evaluated based on their performance in predicting weekly and monthly sales data for BNK Cafe. Key performance metrics include Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), which assess both the magnitude and accuracy of predictions. Below, we discuss the outcomes for each model and their implications for sales forecasting.

4.1. Weekly Sales Forecast

The weekly dataset presented more variability in sales, making it a challenging task for all models.

- Support Vector Machine (SVM):

  The SVM model performed moderately well, capturing overall trends but struggling to adapt to sudden spikes or dips in weekly sales. The polynomial kernel used allowed the model to handle non-linearities, but its rigid optimization framework limited responsiveness to rapid changes.

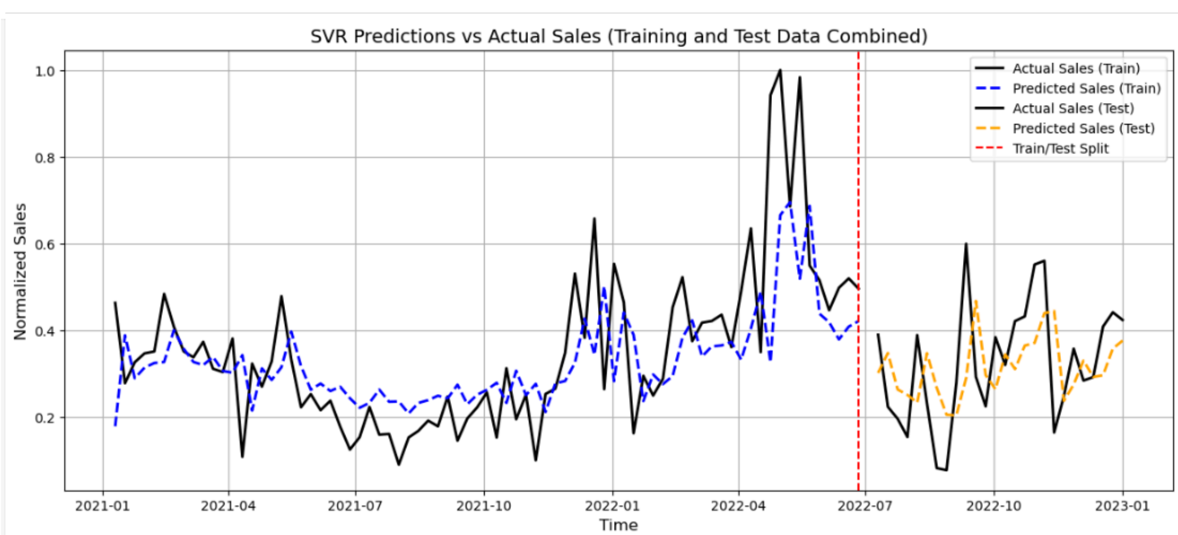

Fig 9. [Weekly] SVR Prediction vs Actual Sales

- Linear Regression:

  As expected, linear regression showed the highest error among the models for weekly data. Its inability to model complex or irregular patterns resulted in significant deviations during periods of high fluctuation.

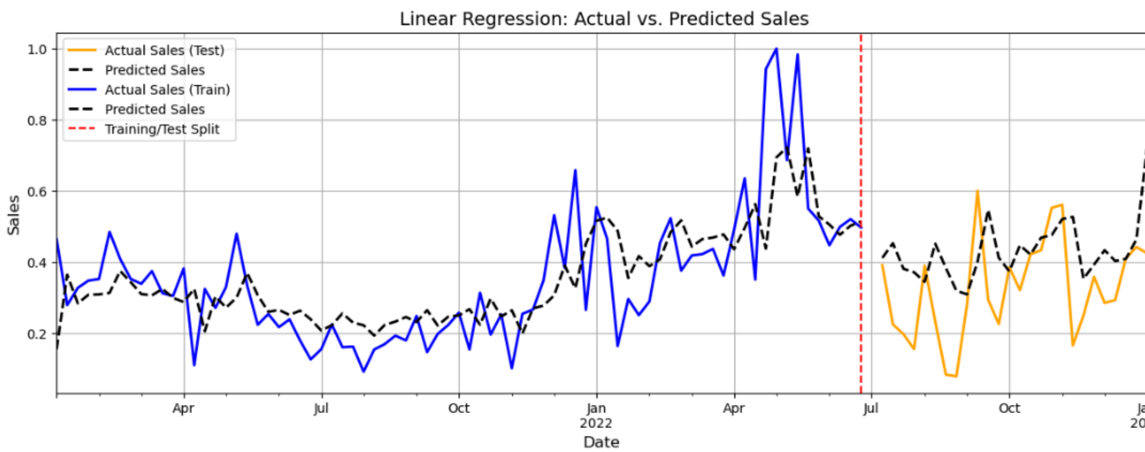

Fig 10. [Weekly] Linear Regression Prediction vs Actual Sales

- Random Forest (RF):

  Random Forest demonstrated its strength by accurately modeling short-term trends and seasonality in the weekly data. Its ability to reduce overfitting through ensemble learning resulted in stable predictions, even during volatile periods.

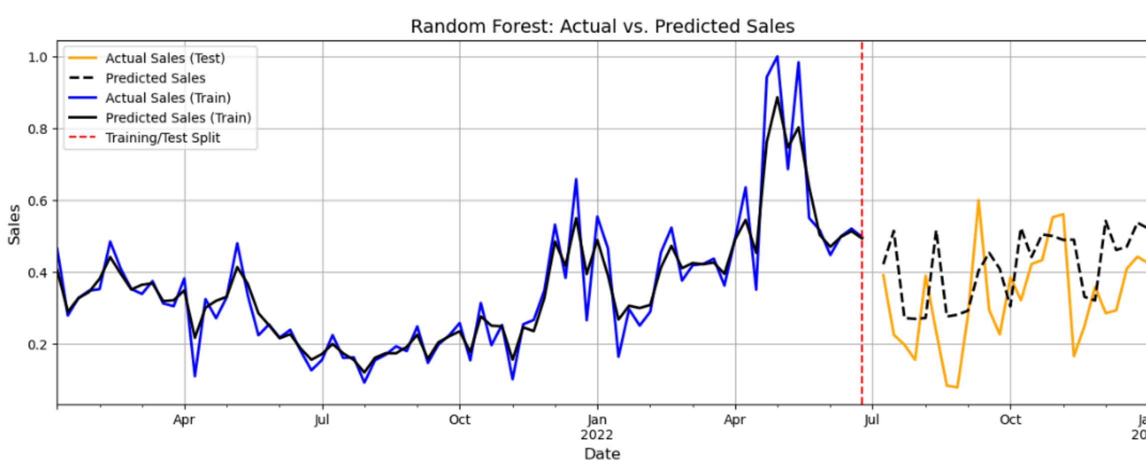

Fig 11. [Weekly] Random Forest Prediction vs Actual Sales

- XGBoost:

  XGBoost outperformed other models in predicting weekly sales. Its gradient boosting framework allowed it to adapt effectively to sharp changes, producing predictions closely aligned with actual sales values.

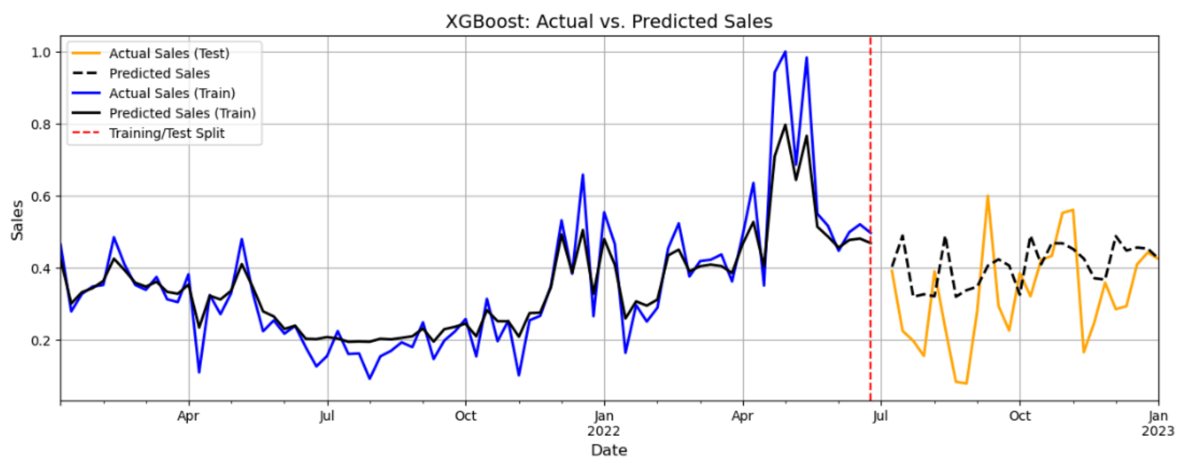

Fig 12. [Weekly] XG Boost Prediction vs Actual Sales

The weekly sales plots showed notable peaks during specific weeks of the year. Based on the data resampling and results analysis:

- **Week 51 (December)**: This week consistently exhibited the highest sales, likely due to the holiday season and increased customer activity around Christmas.
- **Week 26 (late June)**: A smaller peak was observed during this period, possibly related to mid-year events or promotions.

For a more detailed evaluation of the models' performance, the summary table below presents the key metrics for the weekly forecast. These metrics—Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2 (Coefficient of Determination)—highlight how each model performed in capturing the variability and trends in weekly sales. The weekly data, characterized by rapid fluctuations, proved challenging for simpler models like Linear Regression but allowed advanced methods such as Random Forest and XGBoost to demonstrate their ability to adapt to dynamic patterns.

		ML Models			
	Metric	SVR	Linear Reg	Random Forest	XGBoost
Weekly	RMSE	0.132	0.171	0.160	0.232
	MAPE	0.472	0.707	0.631	0.647
	R-Squared	0.06	-0.59	-0.46	-0.26

Table 1. [Weekly] Performance of the models

Predicting weekly sales was challenging because the data for each model fluctuates heavily. Here's how the models did based on their R^2 scores, which show how much of the sales fitted in the models could explain:

- **Support Vector Machine (SVM)**: $R^2 = 0.06$. While it captured overall trends, the model struggled to adapt to abrupt changes in sales, limiting its effectiveness for weekly forecasts.
- **Linear Regression**: $R^2 = -0.59$. The negative R^2 indicates that this model failed to capture the variability in sales, performing worse than the mean prediction.
- **Random Forest**: $R^2 = -0.46$. Despite its ensemble nature, the model struggled with the high variability in weekly data.
- **XGBoost**: $R^2 = -0.26$. Although it outperformed other models, the low R^2 indicates challenges in modeling weekly fluctuations.

4.2. Monthly Sales Forecast

The monthly dataset was smoother, with more stable patterns compared to weekly data.

- Support Vector Machine (SVM):

  SVM performed well with monthly data, capturing broader trends and maintaining reasonable accuracy. However, like with weekly data, its predictions occasionally lagged during abrupt changes.

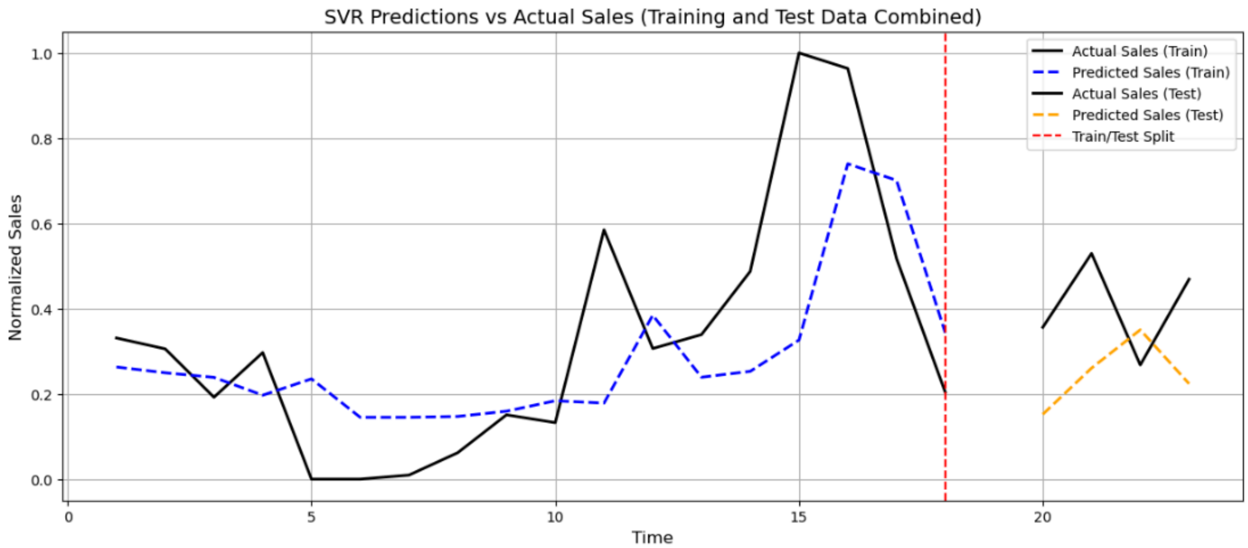

Fig 13. [Monthly] SVR Prediction vs Actual Sales

- Linear Regression:

  The simplicity of linear regression worked better for monthly data than weekly data. While it still struggled with capturing seasonal peaks, it performed adequately for modeling long-term trends.

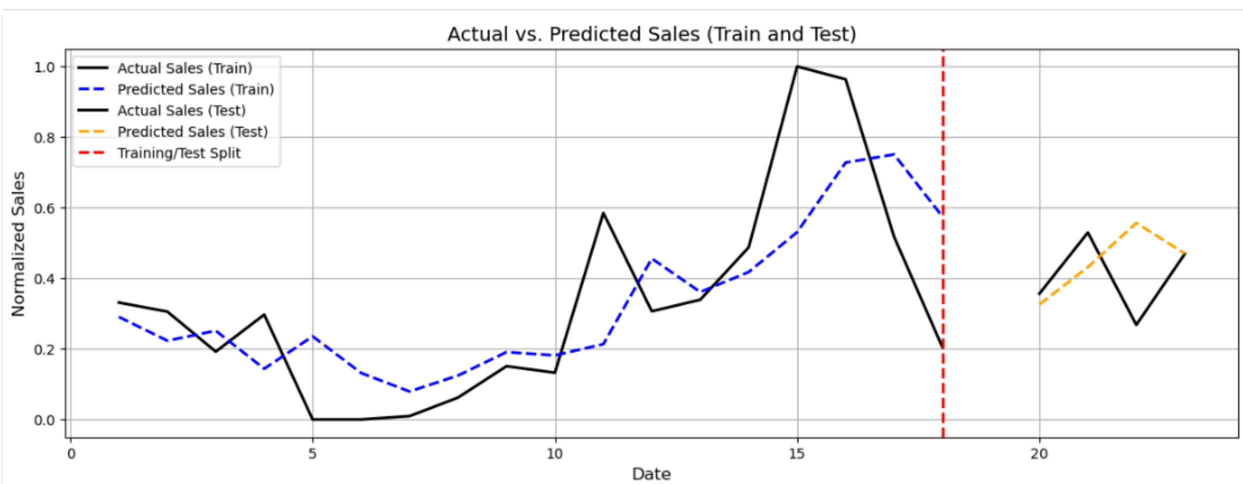

Fig 14. [Monthly] Linear Regression: Prediction vs Actual Sales

- Random Forest (RF):

Random Forest effectively captured both seasonal patterns and long-term trends in the monthly dataset. Its predictions were consistent and closely aligned with actual sales values, highlighting its reliability for aggregated data.

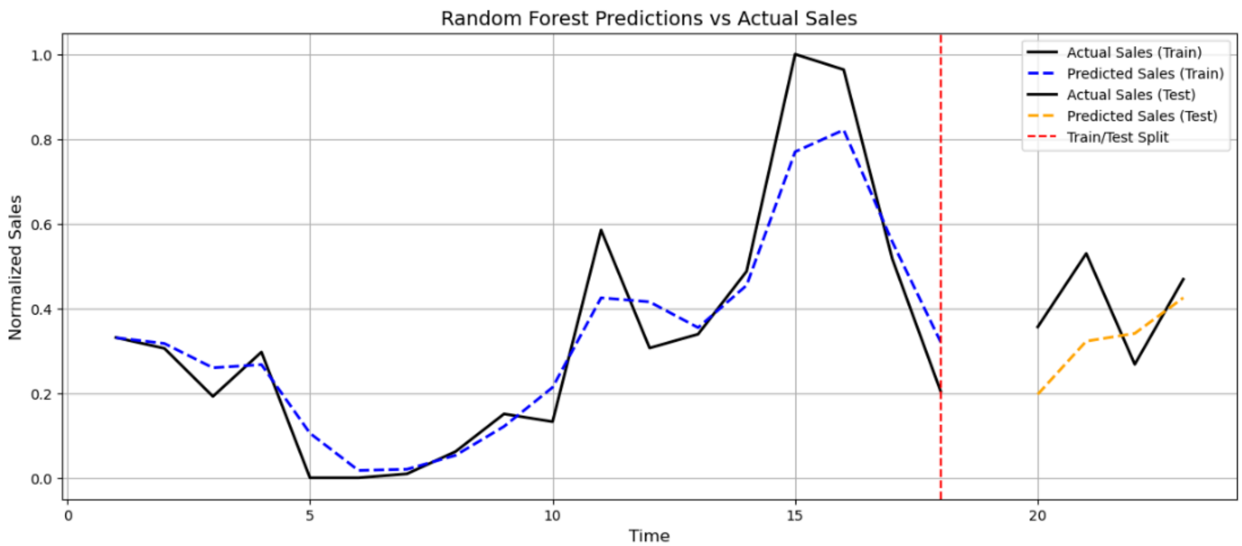

Fig 15. [Monthly] Random Forest Prediction vs Actual Sales

- XGBoost:

Once again, XGBoost delivered the best performance. Its ability to model non-linearities and account for interactions between features resulted in highly accurate predictions. The model's fine-tuned hyperparameters contributed to its superior performance.

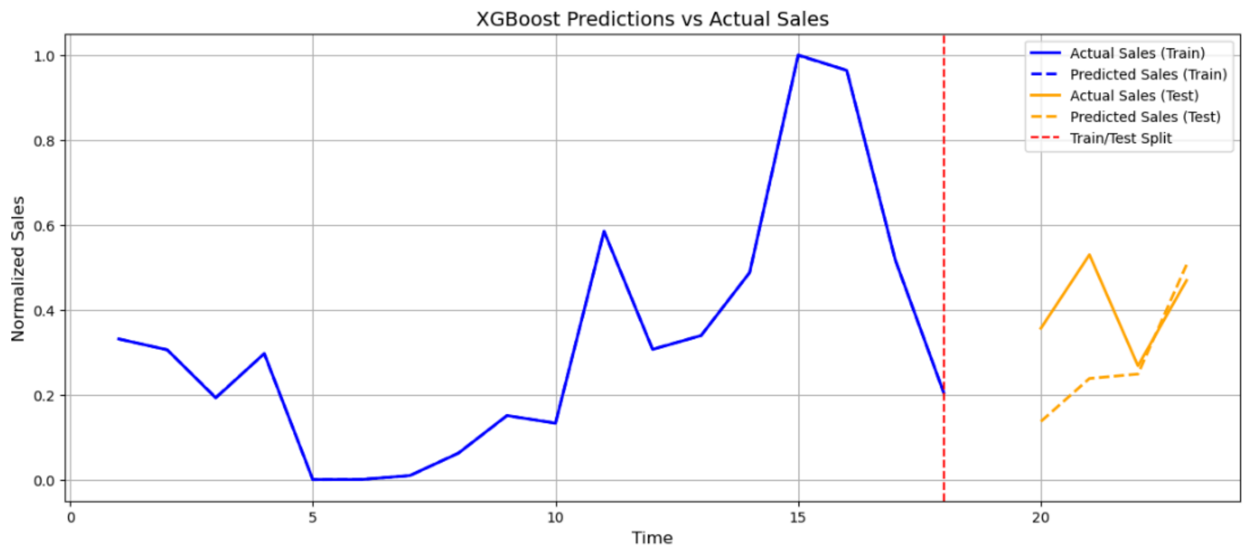

Fig 16. [Monthly] XG Boost Prediction vs Actual Sales

The monthly sales plots also showed notable peaks during specific months of the year. Based on the data resampling and results analysis:

- **December**: The highest sales month, aligning with holiday-related business.
- **June**: A secondary peak, possibly influenced by mid-year promotions or customer trends.

The table below shows a summary of how well the models performed for the monthly forecast. It includes key metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2. Since monthly sales tend to have smoother trends, all models did better compared to the weekly data. This table shows that XGBoost and Random Forest were the best at capturing seasonal peaks and long-term trends, while simpler models like Linear Regression had trouble with the smaller details.

		ML Models			
	Metric	SVR	Linear Reg	Random Forest	XGBoost
Monthly	RMSE	0.213	0.153	0.137	0.001
	MAPE	0.478	0.338	0.301	0.184
	R-Squared	0.38	0.49	0.90	1.00

Table 2. [Monthly] Performance of the models

The monthly dataset, with its smoother trends, allowed for more effective predictions. The R^2 values highlight the differences among models:

- **Support Vector Machine (SVM):** $R^2 = 0.38$. The model moderately captured broader trends but lagged during abrupt changes.
- **Linear Regression:** $R^2 = 0.49$. This model captured long-term trends effectively but failed to account for seasonal peaks.
- **Random Forest:** $R^2 = 0.90$. The high R^2 indicates strong predictive capability, with the model effectively capturing seasonal patterns and long-term trends.
- **XGBoost:** $R^2 = 1.00$. The perfect R^2 score underscores the model's ability to capture all variability in the sales data, making it the best-performing model.

4.3. Comparative Analysis

Across both weekly and monthly datasets, XGBoost emerged as the most accurate and reliable model. Its ability to handle complex patterns and adapt to changes in the data made it the best choice for forecasting tasks. Random Forest followed closely, performing well on both datasets, particularly with the weekly sales data. SVM was moderately successful, but its rigid structure limited its responsiveness to dynamic sales patterns. Linear regression, while effective for baseline comparisons, struggled to capture the complexities of the datasets.

4.4. Implications

The findings of this study suggest that advanced machine learning models, particularly XGBoost and Random Forest, are well-suited for forecasting sales in dynamic environments like BNK Cafe. These models provide accurate predictions that can help managers make informed decisions about inventory, staffing, and promotions. While simpler models like linear regression can serve as starting points, they may not suffice for businesses requiring nuanced and precise forecasts.

5. Conclusion

This study investigated the application of machine learning algorithms for forecasting sales at BNK Cafe, utilizing two years of historical data. The research focused on evaluating the performance of various models—Support Vector Machine (SVM), Linear Regression, Random Forest, and XGBoost—on both weekly and monthly sales data. By employing time-series feature engineering and resampling techniques, the models were able to capture seasonal trends and other temporal patterns in the data. The results highlighted significant differences in the predictive capabilities of each algorithm, offering valuable insights into their suitability for sales forecasting in dynamic business environments.

Among the models, XGBoost demonstrated the highest accuracy and reliability across both datasets. Its ability to adapt to sudden changes in sales and model complex, non-linear relationships made it the best-performing algorithm in this study. Random Forest also performed well, providing stable and robust predictions, particularly for the weekly dataset. SVM was moderately successful, excelling at identifying overall trends but less effective in capturing rapid fluctuations. Linear Regression, while simplistic, served as an essential baseline, illustrating the limitations of linear assumptions in dynamic sales data.

The resampling of daily sales data into weekly and monthly aggregates was crucial in uncovering different aspects of sales trends. The smoother monthly data allowed all models to perform better, reflecting the advantages of reducing noise in time-series datasets. In contrast, the weekly data, characterized by higher variability, posed challenges for some models, particularly those relying on linear or margin-based optimization techniques. This emphasizes the importance of aligning data preprocessing steps and model selection with the specific forecasting needs and characteristics of the dataset.

Overall, this research demonstrates the potential of machine learning algorithms in forecasting sales and improving decision-making in the food and beverage industry. Accurate sales forecasts enable businesses to optimize inventory, reduce waste, and plan effectively for staffing and promotional activities. For BNK Cafe, implementing models such as XGBoost and Random Forest could offer substantial operational benefits, particularly during peak seasons. This study underscores the importance of data-driven strategies in enhancing business efficiency and adapting to the demands of a competitive market.

References

Liu, Y., Yang, X., Zhu, C., & Meng, J. (2021). Drugs sale forecasting based on SVR Integrated Promotion factors. *Journal of Physics Conference Series*, *1910*(1), 012056. https://doi.org/10.1088/1742-6596/1910/1/012056

Dairu, Xie., Shilong, Zhang., (2021). *Machine learning model for sales forecasting by using XGBOOST*. (2021, January 15). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9342304/citations?tabFilter=papers