**Project Summary: Retail Price Optimization Using Machine Learning**

**Introduction**

In the competitive landscape of retail, setting the optimal price is critical for maximizing profitability and market share. This project focuses on developing predictive models to optimize retail prices based on historical sales data, various product features, and competitive pricing. By leveraging machine learning algorithms, the goal is to predict the total price a consumer is willing to pay, thus enabling data-driven pricing strategies.

**Objective**

The primary objectives of this project include:

* To analyze historical sales data and identify key factors influencing retail pricing.
* To develop predictive models that estimate total retail prices based on various input features.
* To compare different machine learning algorithms to determine the most effective model for price optimization.

**Data Collection**

The dataset used in this project comprises historical sales data from the retail sector, including features such as:

* **Quantity Sold (qty)**: The number of units sold.
* **Unit Price (unit\_price)**: The price per unit of the product.
* **Competitor Prices (comp\_1)**: Prices set by competitors for similar products.
* **Product Score (product\_score)**: A rating indicating product quality or customer satisfaction.
* **Competitive Price Difference (comp\_price\_diff)**: The price difference between the product and its competitors.
* **Total Price (total\_price)**: The target variable representing the final retail price.

**Data Preprocessing**

Data preprocessing involved several steps:

* **Data Cleaning**: Removal of any missing or irrelevant data points to ensure accuracy in the analysis.
* **Feature Engineering**: Creation of new features such as comp\_price\_diff to enhance model performance.
* **Data Transformation**: Normalization and encoding of categorical variables to prepare the dataset for machine learning algorithms.

**Model Development**

Multiple regression models were developed and evaluated, including:

1. **Linear Regression**
2. **Decision Tree Regressor**
3. **Random Forest Regressor**
4. **Gradient Boosting Regressor**
5. **Support Vector Regressor**

Each model was trained on a training dataset and validated using a test dataset.

**Model Evaluation**

The models were evaluated using the following metrics:

* **Mean Squared Error (MSE)**: Measures the average squared difference between predicted and actual values.
* **Mean Absolute Error (MAE)**: Measures the average absolute difference between predicted and actual values.
* **R-squared (R²)**: Indicates the proportion of variance explained by the model.

**Results**

The performance metrics for each model are summarized as follows:

| **Model** | **MSE** | **MAE** | **R²** |
| --- | --- | --- | --- |
| **Linear Regression** | 539,263.6 | 430.43 | 0.8293 |
| **Decision Tree** | 150,606.1 | 153.93 | 0.9523 |
| **Random Forest** | 120,655.7 | 91.71 | 0.9618 |
| **Gradient Boosting** | 94,873.2 | 99.83 | 0.9700 |
| **Support Vector Regressor** | 3,323,279.0 | 1,038.42 | -0.0520 |

* **Best Model Based on MSE**: Gradient Boosting achieved the lowest MSE of 94,873.2, indicating strong predictive accuracy.
* **Best Model Based on MAE**: Random Forest exhibited the smallest MAE of 91.71, suggesting low average prediction error.
* **Best Model Based on R²**: Gradient Boosting also led with an R² of 0.9700, showing it explains approximately 97% of the variance in total price.

**Conclusion**

The project successfully demonstrated the effectiveness of machine learning models in predicting retail prices. The Gradient Boosting model emerged as the best-performing model overall, while the Random Forest model excelled in terms of average error. These insights provide a robust foundation for implementing data-driven pricing strategies in the retail sector.

**Future Work**

Future work may include:

* **Hyperparameter Tuning**: Further optimizing model parameters to enhance performance.
* **Integration with Real-Time Data**: Developing a dynamic pricing tool that updates prices based on real-time competitor data and sales performance.
* **Exploration of Additional Features**: Investigating other potential features, such as seasonality and promotional activities, to improve model accuracy.