# **EXPERIMENT REPORT**

| **Student Name** | Nutan Thapa |
| --- | --- |
| **Project Name** | Predicting Sales Revenue |
| **Date** | 8th Oct 2023 |
| **Deliverables** | Prediction\_model.ipynb |

| 1. **EXPERIMENT BACKGROUND** | | |
| --- | --- | --- |
| Provide information about the problem/project such as the scope, the overall objective, expectations. Lay down the goal of this experiment and what are the insights, answers you want to gain or level of performance you are expecting to reach. | | |
| **1.a. Business Objective** | **Explain clearly what is the goal of this project for the business. How will the results be used? What will be the impact of accurate or incorrect results?**  The primary goal of this project for the business is to forecast the total sales revenue across all 10 stores spanning the three states (California, Texas, and Wisconsin) and across all three item categories (hobbies, foods, and household) for the upcoming 7 days.  How will the results be used?  Inventory Management: With accurate forecasting, the retailer can optimize inventory levels by ensuring that there's neither an overstock nor a stockout situation. This can lead to reduced holding costs and increased customer satisfaction.  Resource Allocation: Staffing requirements at each store can be adjusted based on the anticipated sales volumes. For example, if higher sales are expected in the coming week, additional staff can be scheduled in advance.  Promotional Activities: If the forecasted sales are lower than desired, promotional campaigns or discounts can be planned to boost sales. Conversely, if high sales are expected, the retailer might choose to hold off on certain promotions.  Financial Planning: Cash flow management, especially in businesses with thin margins like retail, can benefit from accurate sales forecasts. This aids in budgeting and financial planning.  Supply Chain Optimization: With the forecasted sales data, supply chain processes can be streamlined. It ensures timely ordering from suppliers and minimizes disruptions.  Impact of Accurate Results:  Enhanced Profitability: By preventing overstocking or stockouts, the business can save on costs and enhance profit margins.  Improved Customer Satisfaction: Ensuring that products are available as per demand leads to better customer experience and loyalty.  Strategic Decision Making: Accurate sales data provides a foundation for making informed strategic decisions about expansion, discontinuing certain products, or introducing new ones.  Impact of Incorrect Results:  Financial Losses: Overestimations might lead to overstocking, resulting in higher holding costs, potential spoilage (especially for perishable foods), and possibly more significant discounts to clear out excess inventory.  Missed Opportunities: Underestimations might lead to stockouts, where the business misses out on potential sales and faces customer dissatisfaction.  Operational Disruptions: Incorrect forecasts can disrupt the harmony of the supply chain, leading to either rush orders (with associated increased costs) or excess supply that needs to be stored.  Reputation Damage: Consistently being out of stock or having too much unsold stock might harm the brand's reputation.  In summary, the forecasting model's results have the potential to significantly influence the operational, financial, and strategic aspects of the business. Ensuring the accuracy of the forecast is paramount to optimize benefits and minimize potential pitfalls. | |
| **1.b. Hypothesis** | Present the hypothesis you want to test, the question you want to answer or the insight you are seeking. Explain the reasons why you think it is worthwhile considering it,  Hypothesis:  "Special events, such as holidays and cultural celebrations, have a significant impact on the sales revenue for a specific item in a specific store."  Question to Answer:  "Do special events lead to a statistically significant increase or decrease in sales revenue for specific items in a particular store?"  Reasons for Consideration:  Historical Observations: Preliminary analysis of the data may have shown spikes or drops in sales around specific dates, aligning with particular events.  Business Intuition: Retailers often observe changes in shopping behavior around special events. For instance, certain items might sell more during Christmas, while other items might see a surge during a cultural festival.  Promotional Alignment: If special events are indeed driving sales, businesses can better align their promotional strategies, advertising campaigns, and stock inventory to these events.  Pricing Decisions: Understanding the impact of events on sales can inform dynamic pricing decisions, potentially allowing for premium pricing during high demand driven by an event or discounts to clear stock post-event.  Operational Preparedness: If events significantly influence sales, stores need to be operationally prepared with adequate staffing, security, and logistics.  Customized Offerings: Understanding which items are more popular during which events can allow businesses to offer customized bundles, deals, or new product launches tailored to the event.  Economic Impact: Some events might have broader economic implications, leading to either increased or decreased consumer spending. Testing this hypothesis can provide insights into macroeconomic patterns and their influence on sales.  Worthwhileness:  Testing this hypothesis is worthwhile because the results can offer actionable business insights. If a direct correlation between events and sales is established, it provides a clear direction for strategic planning around marketing, inventory management, staffing, and pricing. Furthermore, it allows businesses to quantify the potential sales uplift (or downturn) due to specific events, enabling better financial forecasting. In essence, this hypothesis aims to bridge the gap between external events and internal sales performance, allowing businesses to be more responsive and proactive in their strategies. | |
| **1.c. Experiment Objective** | Detail what will be the expected outcome of the experiment. If possible, estimate the goal you are expecting. List the possible scenarios resulting from this experiment.  Expected Outcome:  The primary expected outcome of the experiment is to determine the statistical significance of the relationship between special events and sales revenue for specific items in a particular store.  Goal Estimation:  Based on the hypothesis, we aim to determine a quantifiable change in sales (either as a percentage increase or decrease) for specific items during the special events compared to regular days. For instance, an estimated goal might be: "We expect sales to increase by an average of 15% for particular items during special events."  Possible Scenarios Resulting from This Experiment:  Strong Positive Correlation:  The results indicate that special events have a significant positive impact on sales revenue.  Action: Businesses can then amplify marketing efforts, stock up on inventory, and perhaps increase prices slightly due to demand during these periods.  No Correlation:  Special events don't seem to influence the sales revenue noticeably.  Action: Businesses might reconsider investing heavily in event-specific promotions and instead focus on other factors influencing sales.  Negative Correlation:  Sales revenue decreases during special events, which could be surprising.  Action: Investigate further – it might be that specific items aren't popular during certain events, or perhaps there's an external factor like increased competition. Strategies can then be realigned.  Item-specific Variances:  The impact of special events varies based on the items in question. For example, festive decorations might see a surge in sales during Christmas, but regular items might not be as affected.  Action: Tailor the inventory and marketing promotions based on the items that are affected by events.  Store-specific Variances:  Not all stores experience the same effect. Some stores might see a surge in sales, while others remain unaffected or even experience a dip.  Action: Understand regional/cultural variations and tailor marketing and inventory strategies accordingly.  Unanticipated External Influences:  Other factors, such as economic conditions, local disruptions, or concurrent events, might skew the results.  Action: It's essential to account for these in the analysis to ensure that the impact attributed to special events isn't due to external influences.  Ambiguous Results:  The results might not be clear-cut, with some events showing a positive impact while others don't or the magnitude of the effect not being consistent.  Action: This might necessitate a more in-depth analysis or segmentation of events or consideration of other external factors.  In conclusion, the results from this experiment will provide a clearer picture of how events impact sales, enabling businesses to make informed decisions. The myriad of possible outcomes underscores the complexity of retail sales and the importance of using data-driven insights for strategic planning. | |

| 1. **EXPERIMENT DETAILS** | | |
| --- | --- | --- |
| Elaborate on the approach taken for this experiment. List the different steps/techniques used and explain the rationale for choosing them. | | |
| **2.a. Data Preparation** | **Describe the steps taken for preparing the data (if any). Explain the rationale why you had to perform these steps. List also the steps you decided to not execute and the reasoning behind it. Highlight any step that may potentially be important for future experiments**  Data Preparation Steps:  Creating a Continuous Date Range:  Rationale: Ensuring that the dataset contains an entry for every date within the desired range. This helps in maintaining consistency and prevents any disruptions when working with time series data.  Handling Missing Event Data:  Rationale: Missing event data might lead to gaps or inconsistencies in the dataset. By filling these missing events with a placeholder like 'no specific event' and 'none', we are ensuring that the dataset remains continuous and we can account for days without any specific events.  Grouping by Date:  Rationale: To handle any potential duplicates in the event data. By grouping and joining events, it helps to consolidate data and ensure a unique entry per date.  Merging Calendar Dataframes:  Rationale: Combining data from different sources (calendar data and event data) to create a comprehensive dataset with all the necessary information.  Filtering Training Data:  Rationale: To focus on a specific subset of the data based on state, store, category, department, and item ID. This step is crucial to narrow down the dataset and work with targeted information.  Melting Sales Data:  Rationale: To reshape the data in a long format where each row represents sales for a specific day. This makes it easier to join with other datasets and perform time series analysis.  Integrating Sell Prices:  Rationale: It's essential to combine the sales data with the selling price to calculate revenue.  Calculating Daily Selling Price and Revenue:  Rationale: Provides a clear picture of the daily sales revenue which is crucial for any sales forecasting or time series modeling.  Steps Not Executed:  Outlier Detection:  Not explicitly executed in the provided functions.  Reasoning: Depending on the nature of the data, removing or adjusting outliers could be essential. However, it wasn't explicitly done in these steps. Outliers in sales data can sometimes represent valid sales spikes.  Normalization or Standardization:  Not executed in the provided functions.  Reasoning: For certain models, especially those sensitive to the scale of features, normalization or standardization can be vital. It wasn't a part of this processing pipeline.  Handling Additional Missing Values:  Although we addressed missing events, other potential missing values (e.g., in sales data) weren't explicitly handled.  Reasoning: There might not have been other missing values, or the models in use might handle them inherently.  For Future Experiments:  Feature Engineering: Given that sales data can have temporal patterns, features like lag values, moving averages, or seasonality components can be beneficial.  Handling Categorical Variables: While events were dealt with, if other categorical variables get introduced, proper encoding strategies would be essential.  Advanced Imputation Techniques: Instead of using placeholders, advanced imputation techniques like KNN imputation or model-based imputation might provide more accurate data.  The provided steps lay a robust foundation for data processing tailored for time series analysis. However, as highlighted, certain steps can potentially enhance the quality of predictions in future experiments. | |
| **2.b. Feature Engineering** | **Describe the steps taken for generating features (if any). Explain the rationale why you had to perform these steps. List also the feature you decided to remove and the reasoning behind it. Highlight any feature that may potentially be important for future experiments**  Feature Generation Steps:   1. Daily Volume of Sales: Direct indicator of demand; essential for forecasting. 2. Weekly Volume of Sales: Captures weekly sales trends; useful for spotting patterns like weekend spikes. 3. Daily Selling Price: Indicates daily price trends; key for revenue calculation. 4. Sales Revenue: Metric for profitability. 5. Event-related Features: Impact of holidays/events on sales.   Features Removed:   1. Whole Week Selling Price: Removed because the more granular daily selling price offers sufficient information.   Potential Features for Future:   1. Lag Features: For capturing short-term trends. 2. Moving Averages: Smoothes out fluctuations, highlights trends. 3. Seasonality Components: Captures inherent seasonal sales patterns. 4. Promotional Flags: Indicates sales spikes due to offers. 5. External Factors: Like weather or economic indicators.   In essence, the current features focus on daily sales dynamics, with potential future features aiming to capture temporal patterns and external factors. | |
| **2.c. Modelling** | **Describe the model(s) trained for this experiment and why you choose them. List the hyperparameter tuned and the values tested and also the rationale why you choose them. List also the models you decided to not train and the reasoning behind it. Highlight any model or hyperparameter that may potentially be important for future experiments**  **Models Trained for the Experiment:**  **Linear Regression:**  **Rationale: Linear Regression serves as a baseline model. It's straightforward, doesn't require tuning, and offers a point of comparison to assess the relative performance of more complex models.**  **Hyperparameters: Linear regression doesn't involve hyperparameters in its basic form.**  **Results:**  **Mean Squared Error: 0.2716**  **Mean Absolute Error: 0.4280**  **R-squared: 0.7235**  **Random Forest Regressor:**  **Rationale: Random Forests are ensemble methods that are less prone to overfitting, can capture non-linear relationships, and provide feature importances.**  **Hyperparameters Used:**  **n\_estimators: 100, which denotes the number of trees in the forest.**  **random\_state: 42, for reproducibility of results.**  **Results:**  **Mean Squared Error: 0.0042**  **Mean Absolute Error: 0.0090**  **R-squared (10-fold CV average): 0.9480 with a standard deviation of 0.0486.**  **LightGBM:**  **Rationale: LightGBM is a gradient boosting framework that is efficient for large datasets and can outperform other algorithms, especially when appropriately tuned.**  **Hyperparameters Tuned:**  **boosting\_type: 'gbdt', which stands for gradient boosting decision tree.**  **objective: 'regression' since it's a regression task.**  **metric: 'l2', which stands for the squared loss.**  **num\_leaves: 31, controls tree complexity.**  **learning\_rate: 0.05, regulates the step size in the boosting process.**  **feature\_fraction: 0.9, fraction of features used for each iteration.**  **early\_stopping\_rounds: 10, which means training will stop if the validation metric doesn’t improve for 10 rounds.**  **Results:**  **Mean Squared Error: 0.0201**  **Mean Absolute Error: 0.0745**  **R-squared: 0.9796**  **Models Not Trained and Reasoning:**  **Support Vector Regression (SVR):**  **Reasoning: Given the dataset's nature and given that tree-based methods and linear regression were already chosen, SVR might not have added significant value in this context. Moreover, SVR can be more computationally intensive on larger datasets.**  **Deep Neural Networks:**  **Reasoning: Given the performance of the Random Forest and LightGBM models, neural networks might add computational overhead without necessarily guaranteeing improved performance. Furthermore, tree-based models offer better interpretability, which might be essential for business applications.**  **Important for Future Experiments:**  **Gradient Boosting Machines (like XGBoost):**  **XGBoost is a gradient boosting library similar to LightGBM. Given the promising results from LightGBM, XGBoost might also provide good results and could be worth testing in future experiments.**  **Hyperparameter Optimization Algorithms:**  **Bayesian Optimization or Genetic Algorithms could be explored to fine-tune hyperparameters more efficiently, especially for LightGBM and Random Forest.**  **Feature Engineering:**  **Given the strong performance of the models, especially LightGBM, it might be beneficial to investigate feature engineering to further enhance the models' predictive power.**  **In summary, the experiment utilized a variety of models spanning from simple linear regression to more complex ensemble models. The chosen models and their hyperparameters were informed by the dataset's nature and the regression problem at hand. Future experiments can further explore optimization techniques and other models based on evolving requirements and dataset properties.** | |
|  |  | |

| 1. **EXPERIMENT RESULTS** | | |
| --- | --- | --- |
| Analyse in detail the results achieved from this experiment from a technical and business perspective. Not only report performance metrics results but also any interpretation on model features, incorrect results, risks identified. | | |
| **3.a. Technical Performance** | Score of the relevant performance metric(s). Provide analysis on the main underperforming cases/observations and potential root causes.  Performance Metrics:  Linear Regression:  Mean Squared Error: 0.2716  Mean Absolute Error: 0.4280  R-squared: 0.7235  Random Forest Regressor:  Mean Squared Error: 0.0042  Mean Absolute Error: 0.0090  R-squared (10-fold CV average): 0.9480  LightGBM:  Mean Squared Error: 0.0201  Mean Absolute Error: 0.0745  R-squared: 0.9796  Analysis:  Underperformance:  Linear Regression: This model has the highest errors among the three. The MSE and MAE values are notably larger, and the R-squared value, while decent, is significantly lower than those of the Random Forest and LightGBM models.  Potential Root Causes:  Linearity Assumption: Linear regression assumes that the relationship between the independent and dependent variables is linear. If the data has non-linear patterns, then the linear model might not capture these, leading to reduced performance.  Feature Interactions: Both Random Forest and LightGBM can handle interactions between features naturally. In contrast, a basic linear model won't account for these unless manually added, which might be a cause for its reduced performance.  Model Complexity: Linear regression is a simpler model, especially compared to ensemble methods like Random Forest and gradient boosting methods like LightGBM. The latter can model complex relationships and patterns in the data, while the former might struggle if such complexities exist.  Outliers: Linear regression can be sensitive to outliers. If there are any outliers in the data, they can disproportionately affect the regression line, leading to inaccurate predictions. Random Forest and LightGBM are generally more robust to outliers.  Main Underperforming Observations:  To identify specific underperforming observations, we would need to examine the residuals (the difference between the predicted and actual values) from the models. Cases with high residuals are where the model predictions were most off. A detailed analysis would involve:  Plotting the residuals against the predicted values to check for patterns.  Checking if there are any specific groups or categories of data where the model consistently underperforms.  Analyzing the characteristics of these underperforming observations to understand why the model might be getting them wrong.  Recommendations:  Feature Engineering: For linear regression, considering polynomial features or interaction terms might improve performance if non-linearity is suspected.  Robust Regression: If outliers are a concern for the linear model, considering robust regression techniques could be beneficial.  Deep Dive into Data: For observations where predictions were notably off, a detailed data analysis can be conducted to identify any data quality issues or external factors not captured in the dataset that might be influencing the target variable.  In summary, while the Random Forest and LightGBM models perform excellently, the linear regression model lags. A deeper dive into the residuals and the specific characteristics of the underperforming observations can provide more insights into potential areas of improvement. | |
| **3.b. Business Impact** | **Interpret the results of the experiments related to the business objective set earlier. Estimate the impacts of the incorrect results for the business (some results may have more impact compared to others)**  Given the business objective of accurately predicting sales revenue for a specific item in a store on a given date, the models show varying degrees of success:  Linear Regression: It explains 72.35% of the variance in sales revenue. This means that while it has a fairly good grasp on the factors affecting sales, it misses out on almost 28% of the data's complexity, likely non-linear patterns.  Random Forest Regressor: It performs much better, capturing 94.80% of the data's variance. This higher accuracy indicates a better understanding of the factors driving sales, making it a more reliable choice for predicting sales.  LightGBM: The best performer with 97.96% variance explained. For businesses looking to forecast sales, this model would likely provide the most reliable and actionable insights.  Impact of Incorrect Results:  1. Stock Management and Inventory Costs:  If predictions overestimate sales, the store might end up with excess inventory, leading to increased holding costs and potential wastage, especially for perishable items. On the other hand, underestimating sales could lead to stockouts, which would frustrate customers and result in lost sales opportunities.  2. Staffing and Operations:  If the store expects higher sales based on predictions, they might allocate more staff or extended hours, leading to increased operational costs. Conversely, underestimating might lead to understaffing, causing longer customer wait times and a potential decline in service quality.  3. Marketing and Promotions:  Incorrect forecasts might lead to misguided marketing efforts. Overestimating sales might result in less aggressive marketing, potentially missing out on customers, whereas underestimating might lead to unnecessary discounts and promotions.  4. Revenue Projections and Business Strategy:  Inaccurate predictions can skew revenue projections. This might misinform stakeholders and could lead to incorrect strategic decisions about store expansions, renovations, or other capital expenditures.  5. Supplier Relationships:  Regular stockouts or excess inventory can strain relationships with suppliers. It might lead to renegotiated contract terms or missed bulk discount opportunities.  Potential Impact Analysis:  For Linear Regression with 72.35% accuracy, the 27.65% error might have significant impacts in all areas mentioned above. Depending on the magnitude of sales, even a small percentage error can translate into substantial monetary values.  For Random Forest, the 5.2% unexplained variance is far better but still poses risks. However, its overall more accurate predictions might result in fewer operational hiccups.  LightGBM, with only a 2.04% margin of error, offers the best scenario. Still, businesses need to be wary and have contingencies in place for the small percentage of inaccuracies.  Conclusion:  While the LightGBM model offers the best predictive capability, no model is perfect. The business must always consider model predictions as one of many factors in decision-making and maintain flexibility in operations to adjust for any unforeseen sales trends. | |
| **1** | List all the issues you faced during the experiments (solved and unsolved). Present solutions or workarounds for overcoming them. Highlight also the issues that may have to be dealt with in future experiments. | |

| 1. **FUTURE EXPERIMENT** | | |
| --- | --- | --- |
| Reflect on the experiment and highlight the key information/insights you gained from it that are valuable for the overall project objectives from a technical and business perspective. | | |
| **4.a. Key Learning** | Reflect on the outcome of the experiment and list the new insights you gained from it. Provide rationale for pursuing more experimentation with the current approach or call out if you think it is a dead end.  Model Variation: There's a notable difference in performance between models. Linear Regression couldn't capture complexities as well as Random Forest or LightGBM. This highlights the value of trying multiple models for complex data.  Non-linear Data: The superior performance of Random Forest and LightGBM suggests that our data isn't purely linear. Real-world data often has complex patterns that simpler models might miss.  Hyperparameter Importance: LightGBM's performance underscores the need for proper hyperparameter tuning. Parameters like early stopping can help prevent overfitting.  Interpretability vs. Accuracy: While methods like LightGBM might be more accurate, they're harder to interpret. Depending on the business context, interpretability might be crucial.  Next Steps:  Deepen Hyperparameter Tuning: There's potential for further performance gains with more extensive hyperparameter searches for models like LightGBM.  Feature Engineering: Collaborating with domain experts might yield better features and improve model accuracy.  Consider Neural Networks: Given the data's complexity, exploring deep learning might be beneficial.  Prioritize Model Interpretation: Using tools like SHAP or LIME can help explain complex models to stakeholders.  Conclusion:  The current approach has shown promise, especially with LightGBM. Further refinement and exploration are warranted. | |
| **4.b. Suggestions / Recommendations** | **Given the results achieved and the overall objective of the project, list the potential next steps and experiments. For each of them assess the expected uplift or gains and rank them accordingly. If the experiment achieved the required outcome for the business, recommend the steps to deploy this solution into production.**  Advanced Feature Engineering:  Uplift: High  Diving deeper into feature creation might unearth patterns that models could better exploit. Collaborating with domain experts could yield significant dividends in terms of performance.  Neural Networks:  Uplift: Moderate to High  Given the data's complexity, using neural networks like feed-forward or recurrent architectures might capture intricate patterns, especially if the dataset is large.  Ensemble Models:  Uplift: Moderate  Combining predictions from LightGBM, Random Forest, and other algorithms might yield a more robust and accurate model.  Extended Hyperparameter Tuning:  Uplift: Moderate  While we have tuned some hyperparameters, using tools like GridSearch or RandomizedSearch can explore a broader parameter space and potentially increase model accuracy.  Model Interpretability Tools:  Uplift: Low (in terms of performance) but High (in terms of stakeholder acceptance)  Employing tools like SHAP, LIME, or eli5 can help in breaking down predictions, leading to greater trust in the model's decisions.  Data Augmentation:  Uplift: Moderate  Introducing slight variations in the training data or synthesizing new data might enhance the model's generalization capabilities.  Rank:  Advanced Feature Engineering  Neural Networks  Ensemble Models  Extended Hyperparameter Tuning  Data Augmentation  Model Interpretability Tools  Deployment to Production:  If the experiment has met the business objectives:  Model Finalization:  Retrain the chosen model (e.g., LightGBM) on the entire dataset to leverage all available data.  Integration:  Integrate the model into the business's data pipeline. Ensure real-time or batch prediction capabilities align with business needs.  API Creation:  Develop a RESTful API if the business needs predictions on-the-go or for integration into other systems/apps.  Monitoring & Maintenance:  Once deployed, set up monitoring tools to track model performance over time. This ensures the model remains accurate as new data flows in.  Feedback Loop:  Create a mechanism to capture feedback on predictions. This allows for continuous improvement of the model.  Documentation & Training:  Document the model's workings, assumptions, and potential pitfalls. Also, train the relevant stakeholders on its usage, benefits, and limitations.  Scale & Optimize:  As usage increases, ensure the infrastructure can handle the load. Also, look into model optimization techniques for faster predictions if required.  Periodic Re-training:  Schedule regular intervals to retrain the model with new data to ensure it remains up-to-date with the latest trends. | |