# **EXPERIMENT REPORT**

| **Student Name** | Nutan Thapa |
| --- | --- |
| **Project Name** | Forecasting Sales Revenue |
| **Date** | 7th Oct 2023 |
| **Deliverables** | Forecasting\_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,  Question:  Can our chosen time-series algorithm effectively project a week's total sales across all 10 stores and three item categories?  Insight:  We aim to discern patterns in sales data to forecast upcoming sales, identifying potential weekly trends or standout sales days.  Why This Matters:  Historical Data: We've leveraged past sales data, a crucial element for time-series forecasting, to predict future outcomes.  Business Decisions: Accurate forecasts can guide inventory, resource allocation, and financial planning, having a direct impact on profitability.  Algorithmic Strength: We've employed modern forecasting techniques, potentially yielding higher accuracy.  Feedback Loop: Implementing this model allows for continuous refinements, improving our forecasts over time.  In essence, testing this hypothesis is vital as it intersects with core business goals, utilizes available data, and sets a standard for future forecasting endeavors. | |
| **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.  To accurately forecast the next 7 days' total sales revenue across all stores and item categories.  Estimated Goal:  Maintain prediction accuracy with an RMSE below a certain level, based on the best results from our initial models.  Possible Scenarios:  Successful Forecasting: Accurate predictions leading to better business decisions and profit optimization.  Suboptimal Forecasting: Gives a general sales trend but requires further model refinement.  Inaccurate Forecasting: Need to review data, model, or consider external influencing factors.  Overfitting: Model performs well on training data but not on real-world data.  Computational Issues: Some models might be resource-intensive, affecting real-time forecasting efficiency.  Regular model evaluations and updates are essential regardless of the scenario. | |

| 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**  Steps Taken for Data Preparation:  Loading Data: Imported three datasets: 'calendar.csv', 'items\_weekly\_sell\_prices.csv', and 'sales\_train.csv'.  Pivoting: Converted sales data from wide to long format for easier merging.  Merging: Integrated sales data with calendar info and item prices.  Revenue Calculation: Derived daily revenue for each item by multiplying sales by item price.  Grouping & Aggregation: Summarized daily revenues for each store and then aggregated across all stores.  Data Splitting: Segregated data into training (all but the last 7 days) and testing sets (last 7 days).  Rationale:  Merging and pivoting were essential to unify scattered data.  Focusing on total revenue aligned with the project's forecasting goal.  Skipped Steps:  Outlier Handling: Did not explicitly address outliers, preserving potential genuine sales spikes.  Seasonality Decomposition: Skipped this, relying on models to handle seasonality.  Note for Future:  External Factors & Item-level Forecasting: Incorporating external indicators or diving into item-level data might enhance forecasting. | |
| **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. Revenue Calculation: Derived 'item\_daily\_revenue' from 'Sales' and 'sell\_price'. 2. Aggregate Sales: Formed 'Total\_sales\_revenue' by summing daily revenues across stores.   Rationale:  Streamlined sales data into one metric for easier forecasting.  Removed Features:   1. Item-specific Sales: Dropped after aggregating for total revenue. 2. Redundant Columns: Removed unnecessary columns post-merging.   Potentially Important Features for Future:   1. Store-wise Sales: Useful for analyzing regional trends. 2. Item Price History: Can provide insights into market dynamics. | |
| **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:**   1. **ARIMA (Autoregressive Integrated Moving Average)**  * Rationale: ARIMA is a standard method for time series forecasting. It combines autoregressive, differencing, and moving average components to model a time series. * Hyperparameters Tuned: * ARIMA order: The parameters (p,d,q) where:     p is the number of lag observations included in the model (lag order). d is the number of times that the raw observations are differenced (degree of differencing).  q is the size of the moving average window (order of moving average).  **Values tested:** (5,1,0), (0,1,1), (1,1,0), and several others based on ACF/PACF plots.  **Reason for Choice:** It's often a first choice due to its effectiveness and simplicity.   1. SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors)  * Rationale: While ARIMA works well with non-seasonal data, SARIMAX extends ARIMA to effectively handle seasonal components and external regressors. * Hyperparameters Tuned:   SARIMA order: (p,d,q)(P,D,Q,s) where the capital letters represent seasonal components.  Values tested were chosen based on seasonality patterns observed in the data.  Reason for Choice: Useful when dealing with data that exhibits seasonal patterns.   1. Holt-Winters (Exponential Smoothing)  * Rationale: This model captures level, trend, and seasonality in time series data, making it effective for many real-world cases. * Hyperparameters: The model has built-in smoothing parameters for level, trend, and seasonality.   **Reason for Choice:** Given its built-in handling of seasonality and trend without needing differencing, it's a strong choice for data with clear patterns.  **Models Not Trained:**   1. Prophet: We considered using Facebook's Prophet, but given its requirement for additional installations and setup, we decided against it for this experiment. Prophet is an additive model that works well with daily data exhibiting multiple seasonalities. 2. LSTM: While LSTMs can be effective for time series forecasting, they often require larger datasets and more compute power. Given our dataset size and the overhead in tuning an LSTM, we decided against it for this experiment.   Hyperparameters Potentially Important for Future Experiments:   * Seasonality in SARIMAX: The s parameter in SARIMAX represents the seasonality. Given the retail nature of the dataset, it's crucial to identify the correct seasonality (e.g., 7 for weekly patterns). * Smoothing Parameters in Holt-Winters: Adjusting these can improve the responsiveness of the model to changes in trend or seasonality.   **Final Choice:**  We chose the Holt-Winters model for our primary forecasting tool because it gave the lowest RMSE among the tested models, making it the most accurate model for our dataset. | |
|  |  | |

| 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:   1. ARIMA Model:   RMSE: 11,922.50   1. SARIMAX Model:   RMSE: 12,739.34   1. Holt-Winters Model:   RMSE: 8,173.58  **Analysis:**  The Holt-Winters model outperformed the ARIMA and SARIMAX models by a significant margin. This suggests that the Holt-Winters model was better able to capture the underlying patterns in the data.  The ARIMA model's performance, while not as good as Holt-Winters, was still decent. The choice of parameters (5,1,0) might have been suitable for the dataset, but ARIMA might have struggled if significant seasonal patterns weren't addressed with differencing.  The SARIMAX model performed slightly worse than the ARIMA model. One potential reason might be that the seasonal components or external regressors used in the SARIMAX model might not have been optimally chosen or tuned.  **Main Underperforming Cases/Observations:**  Seasonality Handling: If strong weekly or monthly seasonal patterns existed, ARIMA wouldn't have been the ideal choice without explicitly modeling that seasonality. SARIMAX aims to address this, but its performance can suffer if seasonality is incorrectly identified.  Irregularities or Outliers: If there were special events (like promotions, holidays, etc.) that caused sales spikes, models like ARIMA or SARIMAX might not handle them well unless explicitly modeled.  External Factors: Factors not considered in the model, like marketing campaigns, stock availability, or economic conditions, could impact sales and, therefore, the forecast's accuracy.  Potential Root Causes:   1. Data Preprocessing: Any preprocessing steps that might have changed the natural patterns of the data could impact the model's performance. 2. Model Assumptions: Each model has underlying assumptions (e.g., ARIMA assumes stationarity). If these assumptions are not met, performance might be affected. 3. Hyperparameter Tuning: While we tested several orders for ARIMA and SARIMAX, exhaustive tuning might yield better results.   For future experiments, it might be worth diving deeper into these underperforming observations to identify patterns or factors that weren't initially considered. Incorporating these insights could lead to better forecasting accuracy. | |
| **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)**  **Results Analysis:**  For the task of projecting a week's sales revenue across all store locations and product categories for the American retail chain, various models were rigorously tested. Among these, the Holt-Winters method, distinguished by its least RMSE, stood out as the superior forecasting tool for our scenario.  **Business Implications:**   1. **Precision in Forecasting:** With Holt-Winters at our disposal, the retailer gains an edge in forecasting sales, paving the way for informed decisions related to inventory, manpower planning, and promotional campaigns. Such accurate predictions mean the business can tailor its strategies to real demand, ensuring efficient use of resources. 2. **Potential Pitfalls of Inaccurate Projections:**  * Over-Prediction:  1. Capital Lockdown: Predicting sales beyond actual demand might lead to excess inventory, tying up capital, and increasing storage costs. For goods with a limited shelf life, this can also equate to waste. 2. Misallocated Resources: Channeling resources based on inflated forecasts can mean missing out on employing them where they are genuinely needed.  * Under-Prediction:  1. Missed Revenue Opportunities: Lowballing forecasts can result in stock shortages, disappointing customers, and lost sales. In the long run, this could erode brand loyalty in a market where competition is fierce. 2. Operational Hurdles: Scrambling to meet an unanticipated demand surge can strain operations, be it sourcing additional stock or stretching the workforce thin.   3. **Relative Impact Assessment:** Evaluating the models on their RMSE, had the retailer gravitated towards ARIMA or SARIMAX instead of Holt-Winters, they'd have compromised on the accuracy. This could translate to higher susceptibility to the challenges outlined above, pushing up operational costs and missing revenue windows.  To Summarize, the selection of a forecasting tool is pivotal for a retail business, as it ripples through the operational, fiscal, and strategic facets of operations. The inherent risks of over or under-predicting sales make it imperative to lean on the most reliable model. For our retailer, Holt-Winters clearly fits the bill, optimizing their approach to upcoming market demands. | |
| **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.  **Outcome Reflection:**  The Holt-Winters model emerged as the superior forecasting tool in our experiment, highlighting the importance of model selection tailored to specific business contexts.  Insights Gained:   * Model efficacy can vary greatly even with similar forecasting objectives. * Even small improvements in forecasting accuracy can translate to substantial operational and financial benefits.   **Future Direction:**  Given the success of Holt-Winters, it's worth delving deeper into its variants or parameter tuning to potentially further enhance accuracy. However, it's also essential to explore hybrid models or new methods as data grows and market dynamics change. The current approach is promising, but resting on laurels in the ever-evolving retail landscape isn't advisable. | |
| **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.**  **Potential Next Steps and Experiments:**   1. Parameter Tuning for Holt-Winters: Dive deeper into Holt-Winters parameters to further optimize the model.   Expected Uplift: Moderate. Fine-tuning may provide incremental improvements in accuracy.   1. Hybrid Model Exploration: Combine Holt-Winters with other algorithms to capture different data patterns.   Expected Uplift: High. Hybrid models can capitalize on the strengths of individual models, potentially yielding more accurate results.   1. Incorporate External Data: Include data like holiday promotions or economic indicators to better capture sales influences.   Expected Uplift: Moderate to High. External factors can significantly influence sales.   1. Model Interpretability: Explore models that offer more interpretability to provide business insights along with forecasts.   Expected Uplift: Low in terms of accuracy, but high in terms of business insights.  Ranking:   1. Hybrid Model Exploration 2. Incorporate External Data 3. Parameter Tuning for Holt-Winters 4. Model Interpretability   Deployment Recommendation:  As the Holt-Winters model achieved the desired business outcome, it's recommended to:  Develop an API for the model to integrate with the existing business system.  Set up a continuous training pipeline to regularly retrain the model with new data.  Monitor the model's performance in real-time to catch any drift or changes in patterns. | |