

Customer Acquisition Cost Optimisation

Introduction:

OLIST, a leading Brazilian e-commerce platform, is seeking to optimize its marketing efforts by leveraging analytics to measure the effectiveness of customer acquisition campaigns and improve its product offering. In this document, we propose both machine learning and non-machine learning solutions to address the CFO's concern about inflated customer acquisition costs and to improve the efficiency of marketing campaigns.

1. Machine Learning Solutions:

a. Predictive Customer Lifetime Value (CLV) Modeling:

Implementation Strategy:

1. Data collection and preparation: Gather historical customer transaction and interaction data. Clean and preprocess the data.
2. Feature engineering: Create relevant features impacting CLV, such as recency, frequency, monetary value, and customer demographics.
3. Model selection: Choose an appropriate machine learning model for CLV prediction, such as Linear Regression, Random Forest, or Gradient Boosting Machines.
4. Model training and validation: Split the dataset into training and validation sets. Train the model using the training set and evaluate its performance on the validation set.
5. Hyperparameter tuning: Optimize the model's hyperparameters to improve its performance using techniques like Grid Search or Bayesian Optimization.
6. Model evaluation: Assess the final model's performance on a separate test dataset for an unbiased evaluation.
7. Model deployment and monitoring: Deploy the trained model for real-time CLV predictions and continuously monitor its performance.
8. Actionable insights and strategies: Use CLV predictions to develop data-driven marketing strategies and customer retention initiatives to maximize ROI.

Advantages:

- Monetary: Accurate CLV predictions enable targeted marketing spend, potentially reducing acquisition costs and increasing return on investment (ROI).
- Process: Machine learning models can be continuously updated and refined as new data becomes available, ensuring consistent improvements in marketing effectiveness.

Disadvantages:

- Monetary: Developing, deploying, and maintaining machine learning models can be resource-intensive.
- Process: Data quality and availability might be a limiting factor for accurate predictions.

b. Customer Segmentation and Targeted Marketing:

Implementation Strategy:

1. Data gathering: Collect demographics, buying history, preferences, and business encounters. Preprocess data to remove missing numbers, outliers, and inconsistencies.
2. Feature engineering: Use age, gender, geography, recency, frequency, monetary value, product categories, and interaction data to capture customer behavior, preferences, and attributes.
3. Model selection and clustering: Select a clustering approach to group consumers by similarity. K-means, DBSCAN, and hierarchical clustering are prominent customer segmentation methods. t-SNE and UMAP can reduce dimensionality and visualize consumer groupings.
4. Model validation: Train the clustering algorithm on the prepared dataset and find the best number of clusters (if applicable). Assess cluster quality using silhouette score, Davies-Bouldin index, or Calinski-Harabasz index.
5. Analyze and interpret clusters: Use central tendencies (e.g., mean, median) and feature distributions to characterize each consumer segment. Label segments by attributes.
6. Customize marketing campaigns for each customer segment based on their interests, needs, and behaviors. Customized promotions, content, and communication methods are examples.
7. Monitor and assess targeted marketing efforts using KPIs like conversion rate, click-through rate, and ROI. Analyze findings to improve marketing strategy.
8. Maintain the segmentation model with fresh consumer data to keep it accurate and relevant. Review customer categories and alter marketing techniques.

Advantages:

- Monetary: Targeting specific customer segments can yield higher conversion rates, leading to lower acquisition costs.
- Process: Machine learning algorithms can identify and exploit complex patterns in customer behavior, leading to more effective marketing campaigns.

Disadvantages:

- Monetary: Implementing targeted marketing campaigns can be more expensive than broad-based efforts.
- Process: Segmentation might lead to neglecting some potential customers.

2. Non-Machine Learning Solutions:

a. Marketing Mix Modeling (MMM):

Implementation Strategy:

1. Goals and scope: Set MMM project goals like optimizing marketing budget allocation or increasing ROI. Choose the time period, area, and product lines.
2. Get data: Collect historical data on marketing, sales, economic conditions, seasonality, and competition. Internal marketing and sales databases, industry studies, and market research.
3. Preprocessing: Clean, organize, and aggregate data for analysis. Manage missing values, outliers, and marketing activity categorization.
4. Feature engineering: Create relevant variables (features) for modeling, such as marketing expenditure per channel, promotional activities, and external factors like GDP growth or competitor moves.
5. Select a statistical or machine learning model for your analysis. Linear regression, time-series models, and Bayesian methods are common. Choose based on data complexity, variable relationships, and model interpretability.
6. Model development: Input features and fit the selected model to historical data. For model validation, data is separated into training and testing sets.
7. Model evaluation: Use R-squared, MAE, and MSE to evaluate the model's accuracy and validity. Optimize the model.
8. Analyze results: Analyze the model's results to determine how each marketing channel affects sales or the performance metric. Find channels with increasing ROI, diminishing returns, or synergy.
9. Optimization and recommendations: Use the model's insights to guide marketing budget allocation and optimization. Use scenario analysis to evaluate marketing strategy.
10. Monitor and update: Add fresh data to the model to keep it relevant and accurate. Market information and changes should inform marketing tactics.

Advantages:

- Monetary: MMM can provide insights into the effectiveness of different marketing channels, allowing for more efficient allocation of resources.
- Process: MMM can be applied using historical data, which might be more readily available than the data needed for machine learning models.

Disadvantages:

- Monetary: MMM requires data on marketing spend and ROI, which can be challenging to collect and maintain.
- Process: MMM might not capture complex customer behaviors or account for emerging trends.

b. A/B Testing:

Implementation Strategy:

1. Goal setting: Set a test goal like increasing conversion rates, user engagement, or bounce rates.
2. Select variables: Test which parts of your website or app. Headlines, calls-to-action, pictures, and layouts. Limit variables to simplify analysis.
3. Variations: Design A and B with the variables. Make sure the changes matter.
4. Split the traffic: Randomly assign users to version A or B so they only see one variation. For an even distribution, split 50/50 or change the ratio.

Advantages:

- Monetary: A/B testing allows for data-driven decisions and reduces the risk of wasting resources on ineffective campaigns.
- Process: A/B testing is relatively easy to implement and does not require advanced analytical skills.

Disadvantages:

- Monetary: Running multiple tests might be resource-intensive and time-consuming.
- Process: A/B testing might not capture complex customer behaviors or interactions between marketing variables.

Chosen Solution:

The optimal solution for OLIST is to implement a Predictive CLV modeling solution to optimize marketing spend. This approach balances the advantages and disadvantages of all other methods while ensuring the most efficient use of marketing resources.

Known Limitations:

- Data quality and availability
- Costs and resources required for model development, deployment, and maintenance
- Potential neglect of some customer segments

Success Metrics:

- Reduced customer acquisition cost
- Increased return on marketing investment
- Improved customer lifetime value
- Higher customer retention rates
- Greater overall revenue and profitability

By implementing the suggested solutions, OLIST can optimize its customer acquisition costs and improve the effectiveness of its marketing campaigns, leading to increased revenue and profitability.

Customer Churn

Introduction:

OLIST, a leading Brazilian e-commerce platform, aims to minimize customer churn and reinforce retention efforts by leveraging analytics. The company has already built a churn model to identify 'at-risk' customers. In this document, we propose both machine learning and non-machine learning solutions for creating an effective retention strategy to prevent churn.

1. Machine Learning Solutions:

a. Personalized Marketing and Recommendations:

Implementation Strategy:

1. Collect data from purchase history, browsing behavior, demographics, and customer feedback. Clean, relevant, and privacy-compliant data.
2. Feature engineering: Determine customer preferences based on product attributes, demographics, and past interactions. Combining or transforming features can create new ones.

3. Split data: Use the data to train and test the model.
4. Model selection: Choose a machine learning algorithm like collaborative filtering, content-based filtering, matrix factorization, or neural networks.
5. Train the model using the training data. Adjusting model parameters reduces errors and improves prediction.
6. Model evaluation: Test the trained model on the testing set for accuracy and generalization. Precision, recall, and MAP are good evaluation metrics.
7. To improve performance, fine-tune the model's hyperparameters, features, or algorithm based on evaluation results.
8. Model deployment: Integrate the trained model into your marketing systems or recommendation platforms to generate customer-specific content, product recommendations, and promotions.
9. Maintain accuracy and relevance by monitoring and retraining the model. Personalize your marketing strategy with customer feedback and engagement metrics.

Advantages:

- Monetary: Personalized marketing and recommendations can increase customer engagement, leading to higher lifetime value and revenue.
- Process: Machine learning algorithms can analyze vast amounts of data to create highly relevant and targeted recommendations, improving overall customer experience.

Disadvantages:

- Monetary: Developing, deploying, and maintaining recommendation systems can be resource-intensive.
- Process: Data quality, privacy concerns, and algorithm biases might impact recommendation effectiveness.

b. Predictive Analytics for Customer Satisfaction:

Implementation Strategy:

1. Goal setting: Identify your project's goal, such as predicting customer satisfaction, identifying satisfaction factors such as reviews made, repeat orders, social sharing etc, or preventing customer churn.
2. Collect data from surveys, feedback forms, purchase history, and customer support interactions. Validate data.

3. Feature engineering: Determine customer-satisfying features like response time, product quality, and pricing. Combining or altering existing features can create new ones.
4. Split data: Use the data to train and test the model.
5. Model selection: Logistic regression, decision trees, random forests, or neural networks may work for your problem. Data type, complexity, and accuracy determine this choice.
6. Train the model using the training data. Adjusting model parameters reduces errors and improves prediction.
7. Model evaluation: Test the trained model on the testing set for accuracy and generalization. Use precision, recall, F1-score, or AUC-ROC to evaluate.
8. To improve performance, fine-tune the model's hyperparameters, features, or algorithm based on evaluation results.

Advantages:

- Monetary: Early identification of dissatisfied customers can lead to proactive interventions, reducing churn and maintaining revenue.
- Process: Machine learning models can identify patterns in customer behavior that indicate dissatisfaction, enabling targeted interventions.

Disadvantages:

- Monetary: Implementing predictive analytics systems might require significant investment in infrastructure and resources.
- Process: The accuracy and effectiveness of predictive models are dependent on data quality and availability.

2. Non-Machine Learning Solutions:

a. Customer Surveys and Feedback:

Implementation Strategy:

- 1) Identify the target audience and problem that you wish to focus on, such as purchase experience, delivery time, pricing, support received.
- 2) Create questions for the surveys based upon the target and issue identified in step 1.
- 3) Use a platform such as Google Forms or Typeform to create your form and then embed it into the OLIST site to show at the required timing or action by user.
- 4) After a period of time, review the information gathered and act upon it.

Advantages:

- Monetary: Regularly collecting customer feedback can provide insights into potential issues, enabling timely interventions to reduce churn.

- Process: Surveys and feedback collection can be relatively simple to implement and manage.

Disadvantages:

- Monetary: Survey incentives and administration might have associated costs.
- Process: Low response rates and potential biases in survey results can impact decision-making.

b. Loyalty Programs and Incentives:

Implementation Strategy:

- 1) Analyze the sort of benefits that customers would appreciate from a loyalty program whilst also considering the associated cost to said rewards.
- 2) Dry run the algorithm used to calculating revenue/reward to ensure there are no loopholes which can be exploited thus possibly causing financial damage to the business.
- 3) Implement the system to a small user set to double check as stated in step 2.
- 4) As time goes by review the revenue vs cost aspect to the loyalty program to ensure it is competitive whilst also making business sense.

Advantages:

- Monetary: Loyalty programs can increase customer retention, leading to higher lifetime value and revenue.
- Process: Implementing loyalty programs can be relatively straightforward and can be customized to suit different customer segments.

Disadvantages:

- Monetary: Costs associated with loyalty programs and incentives can be significant.
- Process: Determining the right balance of rewards and incentives to optimize customer retention can be challenging.

Summary:

The optimal solution for OLIST is to implement personalized marketing and recommendations as this way the trained model can work out the products and items that interests a particular customer more than just the standard marketing email typically sent out to their complete customer base – this will also help bring back old customers as it can possibly help get customers interested in products that they may not of known about or realized that they need such a product beforehand.

Known Limitations:

- Data quality and availability
- Costs and resources required for model development, deployment, and maintenance
- Potential biases in survey results and algorithm outputs

Success Metrics:

- Reduced customer churn rate
- Increased customer retention rate
- Improved customer satisfaction and engagement
- Higher customer lifetime value
- Greater overall revenue and profitability

By implementing the suggested solutions, OLIST can effectively reduce customer churn and enhance retention efforts, leading to increased customer satisfaction, loyalty, and profitability.

Delivery Date Prediction

Introduction:

OLIST, a Brazilian e-commerce giant, is looking to leverage the power of analytics to improve its processes and product offerings. One of the areas that needs improvement is the estimated delivery date provided to customers. Currently, the estimated delivery time is twice the actual delivery time, which is causing a negative impact on the company's customer satisfaction and retention. This document proposes machine learning (ML) and non-ML solutions to address this problem, highlighting their advantages and disadvantages, and summarizing the best solution along with its known limitations, success metrics, and implementation considerations.

I. Machine Learning Solutions

Time Series Forecasting

Implementation Strategy:

1. **Preprocessing:** Remove inconsistencies, outliers, and missing values. Also, use a structured format like CSV or DataFrame. Creating relevant features to help the model identify patterns may require feature engineering.
2. Select the most important delivery date-affecting features, such as shipping method, distance, and destination. Correlation analysis or recursive feature elimination can help you choose the best model features.

3. Split the dataset into training and testing sets. The training set will train the model, while the testing set will test it. Training is usually 80% and testing 20%.
4. Select a time series forecasting machine learning model. ARIMA, SARIMA, Exponential Smoothing, LSTM are popular options.
5. Train the chosen model on the training dataset. The model learns patterns and relationships between features and the target variable (delivery date) using historical data.
6. Improve model performance by tuning hyperparameters. This step is crucial to finding the best parameters for your problem. Find optimal hyperparameter values using Grid Search or Randomized Search.
7. Evaluate the model's performance on the testing dataset. Evaluate the model's predictions using MAE, MSE, and RMSE.
8. Deploy the model to production after testing it. The model may be integrated into an order management system to automatically predict delivery dates for new orders.

Advantages:

- Process: Able to predict future delivery dates by analyzing historical delivery data
- Process: Can account for trends, seasonal patterns, and other factors affecting delivery times
- Monetary: Can be continuously updated with new data to improve accuracy

Disadvantages:

- Process: Requires a large amount of historical data for accurate predictions
- Process: Can be sensitive to outliers and noise in the data
- Process: May not account for sudden changes in delivery conditions (e.g., weather, natural disasters, etc.)

Regression Models

Implementation Strategy:

1. Collect data: Track package deliveries by origin, destination, weight, shipping method, and delivery date. Data improves model accuracy.
2. Data preprocessing: Remove outliers, handle missing values, and convert categorical variables like shipping method into numerical values using one-hot encoding.
3. Engineering new features to predict delivery dates. For instance, you can calculate the package's distance, shipping day, and holiday season.

4. Split the data into training and testing sets. The training set will train the model, while the testing set will test it. 70-30 or 80-20 is typical.
5. Regression model: Select a regression model to predict delivery dates. Linear regression, decision trees, gradient boosting machines, and random forests are regression models. If your dataset is large, consider deep learning methods like neural networks.
6. Model training: Train your regression model with training data. Cross-validation, grid search, or random search the model's hyperparameters.
7. Model evaluation: Compare the test set delivery dates to your model's predictions. Use MAE, MSE, or R-squared to evaluate.
8. To improve model performance, repeat steps 5-7. Test regression models, feature engineering, and hyperparameter tuning.
9. Deploy the model: After testing, integrate it into your business operations to predict delivery dates in real time. To maintain accuracy, monitor the model's performance and add new data.
10. Share results: To improve transparency and satisfaction, share predicted delivery dates with customers, suppliers, and internal teams.

Advantages:

- Process: Can incorporate multiple features (e.g., distance, order size, delivery location, etc.) to predict delivery time
- Process: Can be trained on historical data and continuously updated
- Process: Can be easily interpreted and explained to stakeholders

Disadvantages:

- Process: May require a large amount of data for accurate predictions
- Process: Can be sensitive to feature selection and data quality

II. Non-Machine Learning Solutions

Rule-Based Systems

Implementation Strategy:

1. Domain knowledge: Ask delivery companies about delivery times and cutoffs.
2. Create a list of rules for shipping pricing and delivery times as per the knowledge gathered in step 1.
3. Refine and update rules: Pricing and delivery times change over time therefore will require regular review.

Advantages:

- Process: Can be easily implemented without the need for extensive data
- Process: Transparent and easy to understand
- Monetary: Can be quickly modified based on expert knowledge

Disadvantages:

- Process: May not account for complex patterns and relationships in delivery times
- Monetary & Process: Can become difficult to maintain as the number of rules increases.

Process Improvement

Implementation Strategy:

1. Review current delivery times and bottlenecks internally such as stock issues and manpower shortages.
2. Find more effective solutions to the issues found in step 1 such as robotic stock movement, more sophisticated stock inventory systems.
3. Identify any issues with the current couriers used when delivering products on the last mile.

Advantages:

- Process: Can identify inefficiencies in the delivery process and help streamline operations
- Process: May lead to more consistent delivery times and better customer satisfaction

Disadvantages:

- Process: Requires a detailed understanding of the delivery process and potential bottlenecks
- Process: May not address the issue of conservative delivery time estimates

Best Solution

The best solution of OLIST to create a regression model of estimated vs actual delivery times whilst also considering for various factors such as delivery distance, order time, stock levels. As this will allow for better delivery accuracy over time as and when more orders are made to better fit the regression model.

Known Limitations

- Data quality and availability may impact the accuracy of ML models
- ML models may not account for sudden changes in delivery conditions (e.g., weather, natural disasters, etc.)
- Process improvements may require significant organizational changes and resources

Success Metrics

- Reduction in the difference between estimated and actual delivery times
- Increase in customer satisfaction and retention rates
- Improved accuracy of delivery date predictions
- Reduction in last-mile delivery issues and inventory management challenges

Implementation Considerations

- Ensure data quality and availability for ML models
- Allocate resources for process improvement initiatives and staff training
- Continuously monitor and update ML models based on new data and feedback
- Communicate changes in delivery date estimates to customers transparently

Fraud Detection

Introduction:

OLIST, a leading Brazilian e-commerce platform, is seeking to protect its business and reputation from fraudulent activities that can result in significant financial losses and damage to its brand image. In this document, we propose both machine learning and non-machine learning solutions to identify and prevent various types of fraud, such as merchant identity, advanced fee, wire transfer scams, and chargeback transactions.

1. Machine Learning Solutions:

a. Supervised Learning for Fraud Detection:

Implementation Strategy:

1. Collect data: Label both fraudulent and non-fraudulent transactions. The dataset should include transaction amount, time, location, user behavior, and other relevant attributes.
2. Preprocess data: Remove missing values, outliers, and categorical variables. To improve model performance, normalize or scale data.

3. Split data: Create training, validation, and testing sets. 70% training, 15% validation, 15% testing is a common split ratio.
4. Handle imbalanced data: Fraudulent transactions are rare, creating imbalanced data. Balance the dataset using oversampling, undersampling, or synthetic data (SMOTE).
5. Model selection: Use logistic regression, support vector machines, decision trees, random forests, or neural networks for binary classification. Find the best model for your dataset by experimenting.
6. Model training: Train the algorithm with the training set. Avoid overfitting and fine-tune hyperparameters by monitoring model performance on the validation set.
7. Model evaluation: Use accuracy, precision, recall, F1-score, and ROC curve area to evaluate the model's performance on the testing set. Choose the best-performing model.
8. Model deployment: Install the trained model to detect fraudulent transactions in real time or batch processing.

Advantages:

- Monetary: Accurate fraud detection can prevent significant financial losses and protect the company's revenue.
- Process: Machine learning models can learn from historical data to identify patterns and predict fraudulent transactions.

Disadvantages:

- Monetary: Developing, deploying, and maintaining machine learning models can be resource-intensive.
- Process: Models might require continuous updates as fraudsters adapt their techniques and create new patterns.

b. Anomaly Detection:

Implementation Strategy:

1. Data collection: Collect transaction logs, user profiles, and device data.
2. Preprocessing: Remove inconsistencies, missing values, and outliers. To scale features, normalize them.
3. Preprocessed data feature engineering: Extract relevant features. These features should detect fraud patterns. Transaction amount, frequency, and time are common.
4. Select a suitable anomaly detection algorithm. Z-score, IQR, K-means, DBSCAN, and autoencoders are some popular methods.

Advantages:

- Monetary: Identifying and preventing atypical transactions can reduce the occurrence of fraud and protect company revenue.
- Process: Anomaly detection algorithms can identify new and emerging fraud patterns that might not be detected through supervised learning.

Disadvantages:

- Monetary: Implementing anomaly detection algorithms can be costly, as they often require large amounts of data for training.
- Process: A high number of false positives might occur, potentially affecting customer experience.

3. Non-Machine Learning Solutions:

a. Rule-Based Systems:

Implementation Strategy:

1. Domain knowledge: Ask fraud analysts or investigators about fraud patterns, behaviors, and indicators.
2. Create a list of fraud-detecting rules based on domain knowledge. These rules may include thresholds, patterns, or combinations of fraud-related factors.
3. Refine and update rules: Fraud methods change. Update the rules periodically to detect new fraud trends.

Credit card fraud detection example rules:

1. Over-the-limit transactions.
2. Short-term transactions from various locations.
3. Off-hours transactions.

Advantages:

- Monetary: Rule-based systems can be cost-effective, as they do not require complex algorithms or extensive data sets.
- Process: Implementing and updating rules is relatively straightforward and can be done in-house.

Disadvantages:

- Monetary: Rules might need frequent updating to stay current with evolving fraud techniques, which can be time-consuming and costly.
- Process: Rule-based systems can be inflexible and might not detect new or unknown fraud patterns.

b. Know Your Customer (KYC) and Identity Verification:

Implementation Strategy:

1. Define the scope: Your KYC system may include customer onboarding, periodic reviews, or transaction monitoring. KYC regulations and best practices vary by industry and jurisdiction.
2. Risk-based strategy: Classify customers by risk with a risk assessment framework. This framework should consider customer type, geography, transaction volume, and business relationship. High-risk customers may need stricter KYC.
3. KYC guidelines: Establish customer identification, verification, and monitoring policies. Define the minimum customer type and risk level information needed and the process for obtaining, verifying, and updating it.
4. Customer identification and data collection: Get customer names, addresses, and birthdays. Customers or trusted third-parties can provide this information.
5. Identity verification: Use trusted, independent sources like passports and driver's licenses to verify the customer's identity. Biometric authentication, video identification, or face-to-face interviews may be needed in some cases.

Advantages:

- Monetary: Implementing KYC and identity verification can prevent fraudulent actors from using the platform, reducing financial losses.
- Process: These measures can enhance customer trust and protect the company's reputation.

Disadvantages:

- Monetary: KYC and identity verification can be expensive and resource-intensive to implement and maintain.
- Process: Some legitimate customers might find the verification process intrusive or time-consuming, potentially affecting user experience and conversion rates.

Summary:

The optimal solution for OLIST would to implement anomaly detection modelling – this is due to the nature of which OLIST conducts its business such as an ecommerce business where customers may make last minute purchases of which they do not wish to have to go through the slow process of KYC – Anomaly detection is perfect for this as it can analyse the customers previous purchase history, refund rate, device location and payment method to calculate the risk of fraud before allowing a transaction.

Known Limitations:

- Data quality and availability
- Costs and resources required for model development, deployment, and maintenance
- Potential impact on customer experience

Success Metrics:

- Reduced incidence of fraud
- Lower financial losses due to fraud
- Improved customer trust and satisfaction
- Enhanced company reputation and brand image

By implementing the suggested solutions, OLIST can effectively combat fraud in the e-commerce space, protect its revenue, and maintain a strong reputation, ultimately leading to increased customer trust and satisfaction.

Price Optimisation

Introduction:

OLIST, a leading Brazilian e-commerce platform, is seeking to optimize its pricing strategy to maximize sales and revenue. In this document, we propose both machine learning and non-machine learning solutions to address the sales team's need for a price optimization algorithm that accounts for various factors such as customer segmentation, location, competitor pricing, and special occasions.

1. Machine Learning Solutions:**a. Dynamic Pricing Algorithm:****Implementation Strategy:**

1. Machine learning model: Choose a model based on data and goals. Linear regression, decision trees, and neural networks are popular.
2. Data prep: Clean and preprocess data to fit your machine learning model. Training and testing sets.
3. A library: Choose a machine learning library to simplify model training. Scikit-learn, TensorFlow, and PyTorch are popular Python libraries.
4. Model evaluation: Assess the model's accuracy using MSE or R-squared on the test dataset.

5. Integrate the model into eCommerce: Implement the trained machine learning model into OLIST to dynamically update prices based on the defined factors.

Advantages:

- Monetary: Maximizes revenue by setting optimal prices based on real-time demand and customer preferences.
- Process: Machine learning models can be continuously updated and refined as new data becomes available, ensuring consistent improvements in pricing effectiveness.

Disadvantages:

- Monetary: Developing, deploying, and maintaining machine learning models can be resource-intensive.
- Process: Data quality and availability might be a limiting factor for accurate predictions.

b. Market Basket Analysis:

Implementation Strategy:

1. Data gathering: Customer IDs, product IDs, purchase dates, and quantities from OLISTs store. Preprocess and clean data for analysis.
2. Pick one: Market Basket Analysis machine learning algorithm. Apriori and Eclat are popular association rule mining algorithms.
3. Set minimum support, confidence, and lift thresholds to filter relevant association rules. Support shows the frequency of items bought together confidence the likelihood of buying one item when another is bought and lift the strength of the association.
4. The algorithm: Implement the algorithm with a machine learning library. Python's 'mlxtend' library implements the Apriori algorithm.

Advantages:

- Monetary: Identifying and exploiting cross-selling opportunities can increase average order value and revenue.
- Process: Machine learning algorithms can uncover complex patterns in customer purchase behavior, leading to more effective pricing strategies.

Disadvantages:

- Monetary: Implementing market basket analysis requires investment in data infrastructure and analytics capabilities.
- Process: Real-time analysis might be challenging due to high data volumes.

2. Non-Machine Learning Solutions:

a. Competitive Price Monitoring:

Implementation Strategy:

1. Collect data on competitor stores that are in the same markets as yourself and ensure the products are the same SKU, Color, Model, Brand, etc
2. Once confirmed develop a way to pull the data on a regular basis to ensure the data received is always “fresh”
3. After collection, compare prices of OLIST and competitors’ and adjust accordingly when taking into consideration overheads and other related costs.

Advantages:

- Monetary: Tracking competitor pricing can help OLIST stay competitive and capture market share.
- Process: Regular price comparisons can inform pricing decisions without advanced analytical capabilities.

Disadvantages:

- Monetary: Requires continuous monitoring and data collection, potentially leading to increased costs.
- Process: Competitor pricing might not capture unique value propositions of OLIST's products.

b. Cost-Plus Pricing:

Implementation Strategy:

1. Cost production: Calculate the product's total cost, including direct and indirect costs. Calculate cost per unit by dividing total cost by number of units produced.
2. Markup percentage: Set each product's profit margin or markup. This percentage depends on industry standards, competition, and business goals.
3. Calculate markup: Each product's markup is the cost per unit times the markup percentage.
4. Price it: The selling price for each product is the cost per unit plus the markup.
5. Examine pricing: Is your price competitive? Consider product differentiation, target audience, and value perception when comparing prices. If needed, adjust markup percentage.

Advantages:

- Monetary: Ensures a profit margin on each product sold.
- Process: Straightforward and easy to implement pricing strategy.

Disadvantages:

- Monetary: Might not be optimal for maximizing revenue or aligning with customer willingness to pay.
- Process: Ignores market dynamics and customer preferences.

Summary:

The optimal solution for OLIST is to implement dynamic pricing algorithm as this will allow OLIST to leave the day-to-day price optimization to the produced model which will enable price competitiveness when compared to like-for-like stores online whilst also allowing for override pricing for special dates such as Christmas, new year, etc

Known Limitations:

- Data quality and availability
- Costs and resources required for model development, deployment, and maintenance
- Potential difficulties in real-time analysis due to high data volumes

Success Metrics:

- Increased sales volume
- Higher average order value
- Improved revenue and profitability
- Greater market share
- Enhanced customer satisfaction and loyalty

By implementing the suggested solutions, OLIST can optimize its pricing strategy to maximize sales and revenue, leading to increased market share and profitability.

Sentiment Analysis

Introduction:

This document aims to suggest both machine learning (ML) and non-machine learning (non-ML) solutions to analyze customer sentiment from OLIST's customer reviews, which are predominantly in Portuguese. By identifying customer experiences and areas of improvement, OLIST can further enhance its product offerings and processes. We present the advantages and disadvantages of each approach based on monetary and process benefits, summarize the best solution, discuss known limitations, and propose success metrics.

1. Machine Learning Solutions:

a. Sentiment Analysis using Pre-trained NLP Models

Implementation Strategy:

1. Select models for both tasks. Sequence-to-sequence models like LSTM, GRU, or Transformer-based architectures like OpenAI's GPT or Google's BERT can translate. Use SVM, Naive Bayes, RNNs, CNNs, or pre-trained models like BERT or RoBERTa for sentiment analysis.
2. Separate training and validation datasets. Train the translation and sentiment analysis models on source-target language pairs and labeled reviews. Your dataset may require fine-tuning the pre-trained models.
3. Test your models on a dataset. BLEU, METEOR, and ROUGE are translation metrics, while sentiment analysis uses accuracy, precision, recall, and F1 score.
4. In a pipeline, the translation model translates customer reviews into a common language (e.g., English) and the sentiment analysis model predicts sentiment scores or categories for the translated text.
5. Install the combined NLP pipeline in production to let users enter customer reviews and receive sentiment analysis results in real time.

Advantages:

- Process: Accurate sentiment analysis with minimal manual intervention
- Process: Faster processing of large volumes of data
- Process: Continual improvement with more data
- Monetary: Monetary benefits from reduced human resource requirements

Disadvantages:

- Process: Limited NLP models available for Portuguese language
- Monetary: High initial cost for model training and deployment
- Process: Potential difficulties in interpreting nuanced language, sarcasm, or idiomatic expressions

2. Non-Machine Learning Solutions:

a. Manual Review and Analysis of Customer Feedback

Implementation Strategy:

- 1) Create a team of support/customer relation officers to go through reviews.
- 2) Have them read each review and translate it
- 3) After translation read again and mark as positive or negative

Advantages:

- Process: Better understanding of nuanced language, sarcasm, and idiomatic expressions
- Process: No reliance on sophisticated NLP models
- Process: Immediate implementation without technical expertise

Disadvantages:

- Monetary: Labor-intensive and time-consuming
- Monetary: High ongoing costs for human resources
- Process: Inefficiency in processing large volumes of data
- Process: Potential inconsistencies in interpretation

Conclusion:

To achieve the desired outcomes of understanding customer experiences and identifying improvement areas, OLIST should consider investing in a pre-built NLP model tailored analysis whilst utilizing a translation for the Portuguese language. While this solution requires upfront investment and expertise, it offers long-term benefits in terms of accuracy and efficiency in sentiment analysis, ultimately leading to better customer experiences and increased business value.

Known Limitations:

- Data collection and annotation for training can be time-consuming and expensive
- Model performance may vary based on the quality and representativeness of the training data
- Maintaining and updating the model requires ongoing investment and expertise

Success Metrics:

- Improved accuracy of sentiment analysis compared to pre-trained models
- Reduction in time spent on manual review and analysis
- Increased customer satisfaction, as measured by an upward trend in average review ratings
- Identification of key areas for improvement and implementation of targeted interventions
- Positive impact on revenue growth and customer retention as a result of improved customer experience

***** End *****