**Delivery date prediction**

**Problem statement:**

The logistics team at Olist uses heuristics to provide an estimated delivery date for the orders placed. It is very conservative about the delivery dates. As a result, it is able to deliver the products much in advance. Although this is beneficial for the logistics team’s 'on time delivery' KPI, it is not favorable for the CMO. He found that on average, the estimated time to deliver products that are given to customers is twice that of the actual delivery time. Such a high expected delivery time is driving away Olist's customers. So, the CMO is looking to use ML to get a far more accurate expected delivery date.

**Proposed Solutions:**

1. ML solutions -   
   1. Use one regression model using features like destination location and source location, and historic delivery date, item size, proximity to distribution hubs, seller, etc to predict the estimated time of delivery. Add the delivery time to the order date and calculate the estimated delivery.
   2. Divide the entire delivery process into individual stages of transport.
      1. Estimate time for getting the goods from the vendor.
      2. Estimate the time that the item will stay in the warehouse.
      3. Estimate the time required to send the item from the warehouse to the delivery location.  
           
         Build individual models for estimating these times, add the time to get the overall delivery time and add it to the order time to get the estimated delivery date.   
         The advantage of building separate models is more fine control over the processes and better prediction.
2. Non - ML solutions
   1. A rule-based approach to predicting delivery times. There are set times for each step of the delivery process such as set time for getting goods from the warehouse to the transport hub, from the transport hub to the next transport hub etc.   
        
      Each of these times is mapped to the nearest shipping cutoffs wherever appropriate. Fixed additional processing times are added for weekends and holidays. Thus, the rule-based model doesn’t adapt based on recent performance changes and is designed based on heuristics.

Selecting First ML solution - because the data points for individual steps are not present.

**Benefits of Proposed Solution**

**Process improvements -**

Predicting delivery dates more accurately will not impact any process in the e-commerce setup.

Reduction in Inventory costs.

**Monetary benefits -**

The monetary benefits of accurate delivery can be divided into two streams

* Reducing customer churn because of accurate delivery date prediction. Hence, increasing the revenue of the company.   
  1. Reduction in the number of people who drop out - 10% (Assumed)
  2. Number of customers daily - 272.44   
     (total number of orders in 1 year = 99442 / number of days in a year 365)
  3. Average order value - 100 Brazilian real (Assumed)
  4. Revenue increased by = 0.1 \* 272.44 \* 100 = 2724.4 per day.
* Inventory management   
    
  Predicting the delivery date correctly will reduce the pressure on the supply chain. The throughout rate of deliveries of the entire system increases but predicting the delivery date correctly. The warehouses will have to hold lesser inventory.
  1. Reduction of time items stay in warehouse = 40% (assumed)
  2. In the same time frame warehouse will be able to hold more items = 40% (assumed)
  3. Warehouse cost per item delivered will reduce = 25% (assumed)
  4. Initial warehouse cost per item delivered = 5 BR (assumed)
  5. Total savings per day = 272.44 \* 0.75 \* 5 BR = 1021.65 BR

**Summarize the DS approach**

The estimation of accurate delivery dates is a regression problem to be solved. You use various data to estimate the time needed for delivery, then add the time to order date to get the right delivery date.

Delivery date prediction is also a kind of balancing act between competitiveness and accuracy.   
You can always have long delivery dates and always be accurate but might lose on sale to some competitor who can deliver quickly.

Or you can have extremely short delivery time promises and disturb the customer sentiment. Hence add a buffer  
  
[There is a constant trade-off between being accurate and being competitive and, of course, we would aim to optimize both.](https://towardsdatascience.com/delivery-date-estimation-5aff1a0ff8dc)

**Limitation**

1. RMSE might be high
2. High complexity model - non linearity, thus nonlinear model might be needed - is the team capable or not.
3. Think of edge cases - data for special cases is not available

**Define Appropriate Success Metrics**

1. Early delivery rates
2. Late delivery rates
3. Churn rate after estimated dates is shown

**Prioritizing Use Case**

Refer to the use case prioritization framework linked [here](https://docs.google.com/spreadsheets/d/1EGoVubzdetsv8YhrKeenSq_uaBSubtsgZm2WydL1U1I/edit#gid=704154787).

**Cite references**

1. Similar delivery date estimation application form example on amazon blogs. <https://aws.amazon.com/blogs/industries/how-to-predict-shipments-time-of-delivery-with-cloud-based-machine-learning-models/>
2. Paper suggesting various solutions for delivery date production <https://arxiv.org/pdf/2105.00315.pdf>

**Sentiment Analysis**

**Problem statement:**

The Chief Marketing Officer at Olist wanted to understand the experience of the customers based on the reviews received after the delivery of the orders. He also wanted to identify the areas of improvement based on these reviews. He had heard that NLP can be used for sentiment analysis and topic modeling, which will be useful in finding topics in customer reviews. However, he was also cognizant of the fact the customer reviews are in Portuguese, whereas the NLP algorithms are not so sophisticated in Portuguese.

**Proposed Solutions:**

1. ML Solution:

Sentiment analysis is NLP problems where various algorithms can be used to identify problems from customer reviews; It can be addressed in timely manner. Most of the cloud providers (For ex: AWS/GCP/Azure) provide language neutral solutions. It can be used at very large scale.

Some of the machine learning models that can be used in additional to NLPS are BERT model, GRU model, Deep learning models

1. Non-ML solution:

Rule based solution can also be used for sentiment analysis where rules can be configured in back end or certain keywords can be looked into reviews to categorize it. It requires huge operation costs and manpower to maintain the solution. Data collection, processing messaging is another laborious intensive task.

**Benefits of Proposed Solution**

**Process improvements –**

Sentiment analysis will not impact any process in the e-commerce setup.

**Monetary benefits –**

Sentiment analysis and addressing the problem can result in huge monetary benefits. These benefits are not limited by just selling the product but over the period it can result in significant improvements in many ways, few of them are listed below:

* Improve customer experience
* Open new markets and build a new brand
* Predict the future/Valuable business intelligence.
* Gain competitive advantage edge

This may lead to an increase in average volume of orders per day/month, reduced warehouse managements.

**Summarize the Solution**

Building a home grown or manual solutions can be very costly. Most of the cloud providers do provide an automated solution where reviews can be analyzed, and customer problems can be addressed in timely manner.

With the advent of technological advantages customers are based all around the world and reviews can be in different languages. It’s advisable to move to automated solution. Most importantly this does not require data to predict every day; reviews can be monitored periodically and from the cost perspective solutions can be optimized.

**Prioritizing Use Case**

Refer to the use case prioritization framework attached in spreadsheet

**Cite references**

<https://risnews.com/how-sentiment-analysis-can-power-your-e-commerce-website>

<https://medium.com/federatedai/sentiment-analysis-in-e-commerce-e8a06a498a75>

**Customer churn**

**Problem statement:**

Customer churn is a critical metric for a CMO at an e-commerce company. Olist wants to develop customer churn models to identify 'at-risk' customers so that appropriate retention strategies can be built. This will provide insights into the factors driving customer churn, thus reinforcing its retention efforts.

**Proposed Solutions:**

1. ML Solution

* Based on important attributes from the historical data (for ex: average volume, geographic location, sex, age, source of income if available) various machine learning classification models can be deployed to predict the customer churn.
* User behavior can also be monitored over a period and based on demographic/behavioral pattern regression models can be used to predict certain promotions/discount rates to retain the customers.

1. Non-ML solution

* Based on customer average volume analysis over a period, customer representatives can give a periodical call to customer to see if they are having any issues. However, this solution may not be optimistic as they need to invest a significant amount of time with consumers to understand the grievances and address in timely manner.

**Benefits of Proposed Solution**

**Process improvements -**

* Predicting customer churns more accurately will not impact any process in the e-commerce setup.
* Improvement in revenue over a period.

**Monetary Benefits:**

From market research, it’s obvious that customer acquisition cost i.e., attracting new customers are very difficult and retaining existing customers may result in greater profits. Using machine learning models in customer churn can benefits in multiple ways:

* Increase in company’s revenue
* Improving customer experience and better understanding their behavior and preferences automatically results in greater profits.
* By analyzing past historical data companies may cross-sell products which can significantly improve the average volume per customers. Companies may use this data to analyze new market segments by taking regular feedback.
* Early intervention in customers who are likely to churn and retaining on the platform which can result in revenue.

**Summarize the Solution**

* Predicting the customer churn is a classification problem to be solved. We use historical data for analysis and list down the important attributes that may indicate a customer is more likely to churn.
* Customer relationship management can use this data to offer some discount/coupon; give a call to understand the grievances and address in timely manner.

**Limitation**

* Using machine learning solutions may churn in high number of false-positive & false-negative. So, data and solutions need to monitor over a period.
* Companies may burn some cash in offering the incentives/coupons to customer who will not be churning and customer who will churn anyways.

**Prioritizing Use Case**

* Refer to the use case prioritization framework attached in spreadsheet.

**Cite references**

* <https://www.analyticsvidhya.com/blog/2022/06/e-commerce-customer-churn-prediction/>
* https://scholarworks.rit.edu/cgi/viewcontent.cgi?article=12319&context=theses

**Customer Acquisition Cost**

**Problem statement:**

The Marketing team at Olist runs multiple promotional campaigns to acquire new customers. However, the CFO believes that the marketing team is burning significant cash by offering deep discounts on products and other benefits, which is inflating the customer acquisition cost. The CFO wants to initiate a new process to measure the effectiveness of the acquisition campaigns by comparing them against the lifetime value of customers.

**Proposed Solutions:**

Customer acquisition cost can be reduced by understanding the user personas and their behavior. As the user data collection has become easy over the period targeted approaches mentioned below can be used to reduce customer costs:

1. Customer segmentation
2. Targeted messaging
3. Recommendation engines
4. Marketing mix modeling
5. Targeted lead generation
6. ML Solution: Mix of machine learning models can be used to solve this problem. Some of the machine learning models used are:
7. Customer segmentation: Supervised/un-supervised both models can be used to segment customer based on their geographic, age, sex, income, behavior.
8. Targeted messaging: By understanding the user behavior and what they are looking for NLP based machine learning models can be used to solve this problem.
9. Recommendation engine: NLP based machine learning models & AI driven powered content can be used to solve this problem.
10. Targeted lead generation: Supervised/un-supervised machine learning models can be used to segment the leads into separate groups based on behaviors, demographics, interaction, and many other factors.
11. Marketing mix modeling: Mix of ml solutions can be designed in incremental manner to solve stage by stage basis or channels specific.

1. Non-ML solution
2. By running specific marketing campaign data can be collected to understand user behavior; Companies can use this data to generate specific leads or run demographic based promotion to attract new customers.

**Benefits of Proposed Solution**

**Process improvements -**

Increasing customer acquisition cost will require improvement over some of the existing processes. As most of the organizations are moving to AI driven solutions it requires one-time significant investment. It may require the change in organization approach over the period.

**Monetary benefits -**

Monetary benefits can be divided into four streams:

1. Better outreach to new customers in timely manner by understanding the dynamic behavior
2. Reduction in cost by generating customer specific contents as data collection and prediction is easy in AI driven solutions.
3. Inventory/human power management which is required to organize these campaigns.
4. Significant investment over a period to main infrastructure.

**Summarize the Solution**

Customer acquisition cost is mainly linked to attracting and retaining customers. In current dynamic world its very important to understand user behavior and their personas linked to demographic. As the data is freely available or it can be collected through various sources companies can use different machine learning models to solve this problem by generating customer/demographic/gender-based behavior.

**Limitation**

* Data collection could be very tricky and costly.
* Understanding the user behavior which changes frequently over a period may be of an issue.
* As user behavior is quite dynamic a significant investment may be required to monitor the AI driven solutions and it could be a huge investment and ROI may be negative.

**Prioritizing Use Case**

Refer to the use case prioritization framework attached in spreadsheet.

**Cite references**

* <https://www.infopulse.com/blog/ai-for-customer-acquisition>
* <https://medium.com/swlh/creating-a-prediction-model-for-customer-acquisition-3517f538ef66>
* https://whites.agency/blog/how-to-use-machine-learning-for-customer-segmentation/
* <https://aithority.com/guest-authors/machine-learning-is-your-secret-weapon-for-customer-acquisition-2/>

**Fraud detection**

**Problem statement:**

Fraud is one the most challenging areas to deal with in an e-commerce industry, as it can result in huge financial losses. There can be fraud in the areas of merchant identity, advanced fee, and wire transfer scams, chargeback fraud, etc. The CFO wants to use the power of analytics to identify fraudulent transactions to help guard the organization against such actions.

**Proposed Solutions:**

1. ML Solution:

Some of the most prevalent machine learning solutions that can be used to prevent fraud activities are listed below. Models can be used to train in isolation or in combination along with others to build machine learning solutions. It’s a classification problem that need to be solved. Some of

* Supervised Decision Tree: Classification or prediction model where fraudulent and normal transactions can be fed.
* Support Vector Machine
* Anomaly detection: autoencoder

1. Non-ML solution

* Fraudulent activities in finance can be detected by looking at on surface and evident signals. Unusually large transactions or the ones that happen in atypical location requires additional verification.
* Firm may adopt rules-based system to detect fraudulent activities but maintaining the rules can be very cumbersome as
* Catching obvious fraudulent scenarios requires manual work
* Requires much manual work to enumerate all possible detection rules.
* Multiple verification steps that harm user experience.
* Long-term processing or maintenance.

**Benefits of Proposed Solution**

**Process improvements -**

Predicting fraudulent activities more accurately will not impact any process in the e-commerce setup.

**Monetary benefits -**

* Reduce operation costs
* No impediment to the customer journey

**Summarize the Solution**

Machine learning allows for creating algorithms that process large datasets with many variables and help find these hidden correlations between user behavior and the likelihood of fraudulent actions. ML solutions allows:

* Real-time processing
* Reduced number of verification measures
* Automatic detection of possible fraud scenarios
* Finding hidden and implicit correlations in data.

Limitation:

* The number of fraudulent activities compared to real activities are very small so data collection will be one of the major problems to deal with.
* False negative as it may leads to significant loss in firms’ revenue and cause reputational damages.

**Prioritizing Use Case**

Refer to the use case prioritization framework attached in spreadsheet.

**Cite references**

* <https://ecommercefastlane.com/how-machine-learning-is-used-in-fraud-prevention-for-ecommerce/>
* <https://www.altexsoft.com/whitepapers/fraud-detection-how-machine-learning-systems-help-reveal-scams-in-fintech-healthcare-and-ecommerce/>

**Price optimization**

**Problem statement:**

Pricing is one of the most important pieces of business for an e-commerce organization. It has a direct and profound impact on revenue, sales, profit, and demand. Price optimization is performed using several factors such as the location, the attitude of the customer, competitor’s pricing, etc. and the data science algorithm predicts the customer’s segmentation to make a response to the change of price. OLISTs sales team wants to build a price optimization algorithm to maximize the sales and revenue.

**Proposed Solutions:**

A firm or company may choose difference pricing strategy depending on market conditions; some of them are listed below

1. Dynamic pricing
2. Cost-plus pricing
3. Competition-based pricing
4. Value-based pricing
5. Price skimming
6. Penetration pricing

Company needs to choose right plan and strategies depending on their specific needs.

1. ML Solution

* Many nonlinear, classification and regression models can be used to build this solution. Classification problem can be used for segmenting the market and customer. Machine learning models can be optimized for right plan and strategies for ex: Dynamic pricing rules or parameters can be derived from other market conditions which may not be feasible for other plans.
* For ex: In pricing related optimization problem such as sales vs pricing, one can first train a multiple regression model to get the pricing co-efficient. Then using these pricing co-efficient one can use linear regression models to solve pricing optimization problems such as revenue maximization while minimizing discount levels.

Machine learning models:

Random Forest, Boosting (GBM & XSBOOST), Deep learning,

Classification, Regression

1. Non-ML solution

* In traditional e-commerce platforms, pricing software is built into backend. However, the efficiency with which you can change prices at scale is limited. This also limits how much flexibility you have setting unique price points for different customer segments and sales channels.
* Rule based pricing strategies can also be chosen but it’s also limited.

**Benefits of Proposed Solution**

**Process improvements -**

Predicting price more accurately will not impact any process in the e-commerce setup.

Reduction in Inventory costs.

**Monetary benefits -**

Choosing the right pricing strategy and automating the solution by machine models can result in significant benefits

* Inventory management if firm is keeping stock for sale
* Increased revenue
* By bundling products togethers and offering discounts can result in daily average volume
* Reduced inventory management
* Better outreach to customers and hence resulting in boosting the average volume.

**Summarize the Solution**

Price optimization can be solved by using a combination of machine learning depending on firm strategies to attract new customers.

**Limitation**

* Cold Start problem (Data collection): When there is not much historical data on pricing, we cannot make accurate predictions about future prices.
* Price elasticity of demand: If you decrease the price, the demand increases and vice-versa.
* Cannibalization (cross-pricing effects): If you decrease the discount of a product, it can lead to an increase in sales or other products.
* Changing market conditions and customer behavior over the period.

**Prioritizing Use Case**

Refer to the use case prioritization framework attached in spreadsheet.

**Cite references**

* <https://7learnings.com/blog/price-optimization-with-machine-learning-what-every-retailer-should-know/>
* <https://vitalflux.com/pricing-optimization-machine-learning-techniques/>
* <https://hal.archives-ouvertes.fr/hal-01942038/document>