# THE MACHINE LEARNING CANVAS Designed for: Armani Designed by: E. Molinari, N. Pinto, E. Tanzi Date: 09/10/24 Iteration: 1 .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PREDICTION TASKThe task is a **multi-output prediction problem** aiming to predict whether a product will be returned and the likely return time frame. The model operates at the product level, with outcomes including both return probability and estimated return window (e.g., within 7 or 30 days). It analyses product, customer, and order data to predict return likelihood and timing.A challenge arises due to the delay in returns, which often occur weeks after purchase. A 90-day window is needed to observe returns, delaying feedback on model accuracy. The model updates monthly with recent data, but this lag must be considered in performance evaluations. | DECISIONSThe model's predictions offer detailed insights to guide business decisions. Logistics and inventory teams will not only know **which** products are likely to be returned but also **when**, enabling better stock adjustments, return handling, and inventory management. Marketing and product teams can use this data to refine product descriptions or recommend alternatives for high-return items, especially those expected to be returned quickly, reducing return rates.These insights can trigger alerts or automate workflows, flagging high-risk products and those likely to be returned soon. This enables more proactive and dynamic management of returns across the supply chain and sales strategies. | VALUE PROPOSITIONThe system’s primary users—logistics, product management, and marketing—aim to reduce product returns, which incur financial and operational costs. Machine learning predictions can enhance decision-making, improving inventory management and product recommendations.Logistics benefit from lower reverse logistics costs and better warehouse efficiency. Product management can use return data to refine designs and improve product descriptions. Marketing can offer more accurate recommendations, increasing customer satisfaction and retention. | | DATA COLLECTIONTo train the machine learning models, a large historical dataset of product orders and returns will be required. This includes information on each product purchased, whether it was returned, and any associated metadata, such as return reasons. The initial dataset will consist of past order data, while the system must be updated continuously with new information as new orders and returns come in.To ensure the model remains robust and accurate, data will need to be refreshed periodically, and holdout sets can be used to evaluate performance on products that have yet to be encountered. | DATA SOURCESThe order database will provide key information on each purchase, including customer details, product attributes, and whether a return occurred. Product images will also be provided by the company; these can be sourced from the company’s product catalog or image repository.Additionally, feedback from customers - such as reviews or return reasons - can offer valuable insights into why certain products are being returned. |
| IMPACT SIMULATIONBefore deployment, the model's impact on operations must be simulated using historical data. This involves training the model and validating it on a test set to assess how well it predicts both return likelihood and timing.Incorrect predictions can lead to costly errors like unnecessary stock adjustments or misguided marketing. Accurate predictions, however, will improve inventory management, reduce costs, and enhance customer service. It's also crucial to apply fairness constraints to avoid bias, ensuring all products are treated equitably.These predictions can integrate with decision systems, triggering alerts for high-return or imminent-return products, enabling timely quality checks, marketing adjustments, and customer communications to reduce returns. | MAKING PREDICTIONSThe ML system will generate predictions for each product shortly after new customer orders are placed, estimating both the likelihood and timing of returns. While real-time processing is not needed, the system will promptly analyse order data to support inventory and marketing decisions. Given the need to process product images for return risk patterns, substantial computing resources will be required, with batch processing used to manage predictive modelling and image analysis. | Additionally, managing high-return users could involve restrictions or added verification for future purchases, further reducing returns. | | BUILDING MODELSThe project will likely use a single machine learning model or a combination of models. One component will predict the probability of a product being returned based on order data (e.g., product category, price, customer history). The second will estimate the time frame of the return (e.g., within 7 or 30 days). Additionally, computer vision techniques will analyse product images to identify patterns (such as color, material, or design) that correlate with return likelihood and timing.The model will require monthly updates with new data, and the image analysis component will need substantial computational resources to process large image sets. | FEATURESThe input features for these models will come from a combination of structured and unstructured data. For the return prediction model, features will include information from the orders, such as product characteristics, customer demographics, price points, and past return behaviours. For the image analysis model, features will be extracted from the product images themselves, looking for visual attributes that may be associated with high return rates. Additional inputs could come from customer feedback, such as reviews or the reasons given for returning the product. |
|  | MONITORINGThe performance of this machine learning system will be measured by its accuracy in predicting product returns and its overall impact on business operations. Key metrics to monitor include the precision and recall of the return prediction model, which will show how effectively it identifies products likely to be returned. Additionally, the system's capability to recognize users with a high return rate will serve as a qualitative metric, helping to mitigate potential fraud by blocking or limiting the activities of fraudulent users.Moreover, the model will assess the importance of the return by defining a risk factor proportional to the cost of the returned product. This will enable the system to recommend less expensive products to users who frequently return high-cost items, thus aligning product recommendations with the likelihood of returns. The reduction in overall return rates and improvements in inventory management efficiency will provide tangible evidence of the model’s value.Over time, analysing the visual patterns identified by the image model will help to determine whether there are common traits among returned products. Monitoring these patterns could lead to insights that influence product design and development, ultimately reducing the rate of returns at the source. | |  |  |  |