**Machine Learning Implementation Feasibility Documentation**

**Project Name:** Predictive Modeling for Ghorerbazar Order Dataset  
**Date:** July 19, 2025  
**Prepared By:** Mehedi Hasan Tushar, on behalf of Ghorerbazar Data Science Team

**1. Objective**

This document presents the feasibility and planned approach for developing machine learning models on Ghorerbazar’s order dataset. The objective is to leverage the dataset to support operational efficiency, customer satisfaction, and revenue growth through predictive insights.

**Key goals include:**

* Predicting delivery success or failure to improve logistics and customer communication
* Forecasting product demand for inventory management and supply chain optimization
* Identifying orders at risk of return or cancellation to proactively mitigate losses
* Segmenting customers based on purchasing behavior for targeted marketing and personalized service
* Predicting payment defaults to minimize financial risk and optimize collections
* Enhancing sales and marketing through additional customer and sales predictions

**2. Dataset Summary**

The dataset comprises transactional order data collected from Ghorerbazar’s e-commerce platform. It contains fields such as customer details, order metadata, product information, financial transactions, and delivery tracking.

**Key columns include:**

* Customer Info: Customer Name, Phone, Address, Tag (e.g., regular, new, probashi)
* Order Info: Invoice ID, Creation Date, Shipping Date, Status, Sub-status
* Product Info: SKU, Name, Specifications, Quantity, Price, Discounts
* Financial Info: Sales Amount, Payments Received, Due Amount
* Delivery Info: Delivery Partner, Delivery Status, Delivered At

**Sample Size:** Currently, over 30 records are available, with monthly growth expected to exceed 10,000 records, ensuring enough volume for reliable modeling.

**Sample Data Excerpt:**  
(Refer to attached sample data table in Appendix or previous section.)

**3. Machine Learning Use Cases**

**A. Delivery Status Prediction**

**Description:**  
This model aims to predict whether an order will be successfully delivered, delayed, or flagged for issues based on factors like product type, customer location, and timing of order and shipment.

**Business Impact:**  
Accurate delivery predictions allow proactive management of logistics, timely customer notifications, and resource allocation to reduce delays and increase customer satisfaction.

**Target Variable:** Delivery status (e.g., DELIVERED, IN\_TRANSIT, FLAGGED)  
**Input Features:** Product category, geographic location of customer, order creation time, shipment delay, delivery partner  
**Model Type:** Classification (e.g., Random Forest, XGBoost)

**B. Return/Cancellation Risk Prediction**

**Description:**  
This use case identifies orders that have a high risk of being returned, cancelled, or damaged, enabling preemptive customer engagement or operational adjustments.

**Business Impact:**  
Reducing return rates lowers costs associated with reverse logistics, improves inventory planning, and increases overall profitability.

**Target Variable:** Binary indicator of return/cancellation (based on returned or cancelled quantity)  
**Input Features:** Product details, customer purchase history and behavior, delivery partner, shipment timings  
**Model Type:** Classification (e.g., Logistic Regression, Gradient Boosting)

**C. Demand Forecasting**

**Description:**  
Forecasting product demand at SKU-level on a weekly or monthly basis supports inventory replenishment and supply chain management.

**Business Impact:**  
Better demand forecasts prevent stockouts and overstock, reduce wastage, and improve customer satisfaction by ensuring product availability.

**Target Variable:** Quantity sold per SKU over defined periods  
**Input Features:** Historical sales data, SKU attributes, seasonality (time features), customer segmentation  
**Model Type:** Time Series Forecasting (e.g., Prophet, ARIMA, LSTM)

**D. Payment Completion Prediction**

**Description:**  
This model predicts the likelihood of an order’s payment being completed or remaining due/unpaid.

**Business Impact:**  
Timely identification of potential payment defaults helps prioritize collection efforts and optimize cash flow.

**Target Variable:** Binary indicator of due amount (paid vs unpaid)  
**Input Features:** Product pricing, discounts applied, payment method, customer type/tag  
**Model Type:** Binary Classification (e.g., XGBoost, Logistic Regression)

**E. Customer Segmentation**

**Description:**  
This unsupervised learning use case groups customers based on purchasing frequency, order values, and return behaviors.

**Business Impact:**  
Segmentation enables targeted marketing, personalized promotions, loyalty programs, and improved customer retention strategies.

**Features Used:** Purchase frequency, total spending, product categories bought, return rates  
**Model Type:** Clustering algorithms (e.g., K-Means, DBSCAN)

**4. Additional Machine Learning Use Cases for Sales and Customer Insights**

**F. Sales Conversion Prediction**

**Description:**  
Predict the likelihood that a website visitor or a cart user will complete a purchase, based on browsing behavior, product views, past purchases, and promotional campaigns.

**Business Impact:**  
Helps marketing teams focus on high-potential leads, optimize promotions, and increase conversion rates.

**Target Variable:** Whether a session leads to a purchase (yes/no)  
**Features:** Number of product views, time spent on site, referral source, customer history, discount usage  
**Model Type:** Binary Classification (e.g., Logistic Regression, Random Forest)

**G. Customer Lifetime Value (CLV) Prediction**

**Description:**  
Estimate the total revenue expected from a customer over their entire relationship with Ghorerbazar, using purchase frequency, average order value, and retention trends.

**Business Impact:**  
Supports customer segmentation, loyalty program design, and budget allocation for marketing campaigns.

**Target Variable:** Predicted total spend or revenue per customer over a period  
**Features:** Historical purchase frequency, average order value, recency of last purchase, customer segment  
**Model Type:** Regression (e.g., Gradient Boosting Regressor, Neural Networks)

**H. Churn Prediction**

**Description:**  
Identify customers who are likely to stop buying or become inactive, by analyzing their recent purchase patterns, browsing behavior, and engagement metrics.

**Business Impact:**  
Enables targeted retention campaigns, personalized offers, and improved customer support to reduce churn.

**Target Variable:** Churn (yes/no) within a future time window (e.g., next 30 days)  
**Features:** Time since last purchase, frequency drop, complaints or return history, customer demographics  
**Model Type:** Binary Classification (e.g., XGBoost, SVM)

**I. Cross-Sell and Upsell Recommendation**

**Description:**  
Recommend additional products or upgrades to customers based on their purchase history, browsing data, and similarity to other customers’ buying patterns.

**Business Impact:**  
Boosts average order value and sales by increasing product basket size.

**Target Variable:** Next product(s) likely to be purchased  
**Features:** Customer purchase history, product categories, seasonality, price sensitivity  
**Model Type:** Recommendation Systems (e.g., Collaborative Filtering, Association Rule Mining)

**J. Seasonal Sales Trend Analysis and Promotion Planning**

**Description:**  
Predict sales fluctuations during holidays, festivals, or special promotions by analyzing historical sales and external factors like weather or events.

**Business Impact:**  
Helps optimize inventory and marketing spend to maximize revenue during peak seasons.

**Target Variable:** Sales volume by product category during specific time frames  
**Features:** Historical sales, event calendar, promotions, weather data (if available)  
**Model Type:** Time Series Forecasting with external regressors (e.g., Prophet with holidays)

**5. Data Preprocessing Requirements**

* **Missing Values:** Address nulls in critical date fields (Delivered At), status flags, and return quantities.
* **Feature Engineering:** Create new features such as delivery delay (Delivered At minus Shipping Date), payment ratios (Paid Amount vs Total Amount), and time-based features (month, day of week, hour).
* **Categorical Encoding:** Convert textual fields like Product Name, Payment Method, and Delivery Partner into machine-readable numeric representations.
* **Outlier Detection:** Identify anomalous records in prices, quantities, or dates that may skew modeling.

**6. Evaluation Metrics**

| **Use Case** | **Metrics** |
| --- | --- |
| Delivery Prediction | Accuracy, F1-Score, AUC |
| Cancellation Forecasting | Precision, Recall |
| Payment Default | AUC, Accuracy |
| Demand Forecasting | RMSE, MAE |
| Customer Segmentation | Silhouette Score |
| Sales Conversion Prediction | Accuracy, Precision, Recall |
| Customer Lifetime Value (CLV) | RMSE, MAE |
| Churn Prediction | Precision, Recall, F1-Score |
| Cross-Sell / Upsell Recommendation | Precision@K, Recall@K |
| Seasonal Sales Trend Analysis | RMSE, MAPE |

**7. Tools and Technologies**

* **Programming Language:** Python
* **Libraries:** pandas, scikit-learn, xgboost, Prophet, matplotlib, seaborn, surprise (for recommender systems)
* **Deployment:** Flask or FastAPI for API endpoints (optional)
* **Data Storage:** SQL Server, CSV, or Excel
* **Visualization:** Power BI dashboards, matplotlib plots

**8. Challenges and Risks**

* **Data Quality:** Missing delivery timestamps, inconsistent product naming conventions, and incomplete financial records.
* **Class Imbalance:** Low proportion of cancellations/returns may affect classifier performance.
* **Privacy:** Customer information requires anonymization to comply with data privacy regulations.
* **Operational Integration:** Aligning model outputs with existing business processes and systems.
* **Feature Availability:** Some advanced features like browsing behavior or external weather data may require integration with other systems.

**9. Next Steps**

1. Confirm and acquire at least 6 months of continuous historical order data.
2. Execute detailed data cleaning, normalization, and preprocessing steps.
3. Develop and validate the delivery status prediction model as an initial pilot.
4. Extend modeling to other use cases iteratively based on business priorities.
5. Visualize insights through Power BI or custom dashboards for business stakeholders.
6. Plan integration and automation for regular model updates and monitoring.
7. Explore additional data sources (e.g., web analytics) to enhance sales and customer models.

**Conclusion**

With a rich and growing order dataset, Ghorerbazar has a strong foundation to apply predictive analytics and machine learning. These models will enhance decision-making, reduce operational inefficiencies, increase customer satisfaction, and boost revenue — driving measurable business value.

**Prepared By:**  
Mehedi Hasan Tushar  
Data Engineer, Ghorerbazar