

Intelligent CRM: Integrating Customer Data, Analytics, and Machine Learning for Better Customer Decisions

Syeda Zoya¹, Insiya Maryam², Raju Kumar Allam³, Neha Kauser⁴

¹ Data Solutions Specialist, Falcon Informatics

² Platform Head, Falcon Informatics

³ Practice head, Falcon Informatics

⁴ Business Development Executive, Falcon Informatics

Abstract - Customer Relationship Management (CRM) systems collect customer information but often provide limited intelligence for identifying customer risks, behavioral patterns, and appropriate actions. This study proposes an Intelligent CRM framework that integrates customer data, Customer 360 profiling, analytics, machine learning, and rule-based recommendations.

The framework combines transaction, payment, product, delivery, and customer-feedback information into a unified customer view. Machine learning is applied to customer segmentation, dissatisfaction prediction, value classification, and delivery-risk analysis. The resulting intelligence is integrated into a working CRM proof of concept containing Dashboard, Customers, Customer 360, and ML Insights modules.

The results demonstrate that the proposed approach can support customer dissatisfaction identification and value classification, while customer segmentation shows limited separation and delivery-risk prediction remains challenging with the available features. These findings highlight the importance of data quality, coverage, and appropriate feature selection for effective CRM intelligence.

The study demonstrates how customer data and machine learning can be integrated into a practical CRM decision-support workflow, providing a foundation for extending CRM capabilities with additional customer and operational data.

Keywords: Intelligent CRM, Customer 360, Customer Analytics, Machine Learning, Customer Segmentation, Customer Satisfaction, Predictive CRM

1. INTRODUCTION

1.1 Background

Customer Relationship Management has become an important component of modern organizations because customer information is generated across multiple business processes, including purchases, payments, product interactions, service activities, and customer feedback. Traditional CRM systems primarily provide mechanisms for storing customer records and supporting operational activities. However, storing customer information alone does not necessarily provide organizations with sufficient intelligence to determine which customers require attention, which customers may be dissatisfied, or which customer groups exhibit different behavioral patterns.

The increasing availability of transactional and behavioral data has created opportunities for analytical CRM systems. Customer data can be transformed into structured features that describe purchasing behavior, customer value, service experience, and engagement. Machine learning can then be applied to these features to identify customer segments, estimate customer-related risks, and support operational decisions.

In this context, an Intelligent CRM system can be viewed as a combination of three capabilities: a unified customer data layer, an analytical and machine learning layer, and a decision-support layer. The first capability provides a consolidated view of the customer. The second derives descriptive and predictive intelligence from customer information. The third converts analytical outputs into actions that can be used by CRM users.

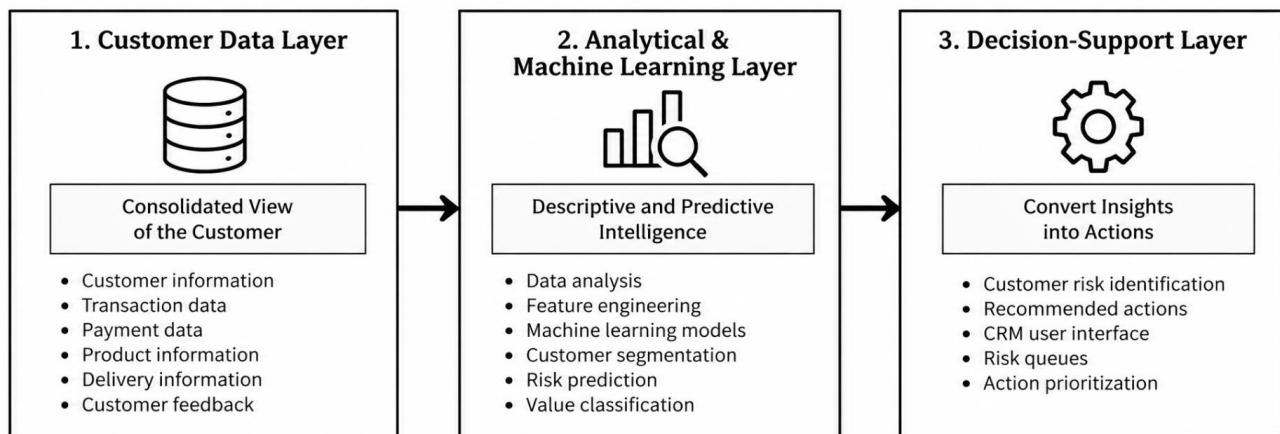


Figure 1: Conceptual Structure of an Intelligent CRM System

The objective of this work is therefore not to develop another conventional customer database. Instead, the objective is to investigate how customer data, analytics, and machine learning can be integrated into a practical CRM workflow.

1.2 Problem Statement

A major challenge in data-driven CRM is the transformation of heterogeneous customer information into useful customer decisions. Customer information may exist across order records, payment records, product information, service events, and customer feedback. These sources must first be integrated before meaningful customer-level analysis can be performed.

Furthermore, predictive models developed independently from CRM workflows may provide probabilities or classifications without explaining how those predictions should influence customer management activities.

This study addresses the following problem:

How can heterogeneous customer data be transformed into customer-level intelligence and integrated into CRM decision-making?

1.3 Motivation

The motivation for this research comes from the need to connect data analysis with practical CRM activities.

For example, identifying a customer with a high probability of dissatisfaction is more useful when the CRM system can present that customer to a service employee and provide an appropriate action. Similarly, customer segmentation becomes more useful when segments can be used to organize customer views or identify potential CRM actions.

Therefore, this work focuses on the complete path:

Customer Data → Intelligence → CRM Action

1.4 Research Objectives

The study aims to:

1. **Integrate** customer and transaction-related data into a unified CRM-oriented structure and develop a **Customer 360** representation.
2. **Engineer and evaluate** customer features for segmentation, dissatisfaction prediction, delivery analysis, and value classification.
3. **Convert ML outputs** into CRM risk queues and recommended actions.
4. **Implement** the proposed workflow as a working CRM proof of concept.
5. **Identify** key data limitations and requirements for extending the system to a broader enterprise CRM.

2. RELATED WORK

2.1 Traditional CRM

Traditional CRM systems primarily support the collection, storage, and management of customer information. Operational CRM typically supports activities such as customer records, sales transactions, and service interactions. These systems provide the foundation for customer management but may not automatically transform historical information into predictive customer intelligence.

Analytical CRM extends this concept by applying analytical techniques to customer information. Instead of only recording what happened, analytical CRM attempts to identify customer patterns and support business decisions.

This distinction motivates the architecture proposed in this study, where customer information is not only stored but also transformed into analytical features and predictive outputs.

2.2 Customer Intelligence

Customer intelligence refers to the process of transforming customer-related information into knowledge that can support decisions. Examples include customer segmentation, customer value analysis, satisfaction analysis, and behavioral profiling.

A unified customer profile is particularly important because individual transactions do not necessarily represent the overall relationship between an organization and a customer. Aggregating transactions at the customer level enables features such as purchase frequency, monetary value, recency, service performance, and review behavior to be considered together.

The Customer 360 component of the proposed system follows this principle by consolidating information associated with a customer into a single customer-level representation.

2.3 Data-Driven CRM

Data-driven CRM depends on the availability and quality of customer data. Transactional information can provide strong evidence of purchasing behavior, while reviews can provide information about customer experience. Payment information can provide additional behavioral information.

However, a CRM system intended for broader customer lifecycle management may require other data sources, including leads, campaigns, customer service tickets, website interactions, email engagement, and sales opportunities.

The dataset used in this study does not contain these sources. Therefore, the proposed system represents a CRM intelligence proof of concept rather than a complete enterprise CRM implementation.

2.4 Machine Learning in CRM

Machine learning has been applied to several CRM-related tasks, including customer segmentation, churn analysis, customer value estimation, recommendation, satisfaction prediction, and campaign optimization.

In this study, machine learning is applied to several CRM intelligence tasks. However, the experiments also evaluate whether the available data actually contains sufficient information for each task.

This distinction is important because the existence of a machine learning algorithm does not guarantee that the required business problem can be predicted accurately. For example, the late-delivery experiment produced performance close to random classification, indicating that the available features did not provide sufficient predictive information for this task.

2.5 Research Gap

Existing CRM and customer analytics approaches demonstrate the usefulness of customer data and machine learning for individual analytical tasks. However, a practical CRM workflow requires more than an isolated predictive model.

The focus of this study is therefore the integration of:

Data integration → Customer 360 → Feature Engineering → Machine Learning → CRM Intelligence → CRM action

The study also explicitly evaluates the limitations of the available data and reports negative or weak modelling results. This allows the research to distinguish between capabilities demonstrated by the current data and capabilities that require additional CRM data.

3. DATASET AND DATA PREPARATION

3.1 Dataset Description

The study uses an e-commerce dataset for CRM-oriented analytics.

The dataset covers transactions of about 2 years. After integration and processing, the dataset contains:

Measure	Value
Orders	99,441
Unique customers	96,096
Order items	112,650
Payments	103,886
Orders with reviews	98,673
Customer repeat rate	3.12%

The low repeat rate is an important characteristic of the dataset. Only 3.12% of customers made more than one purchase. Consequently, customer lifecycle modelling such as true churn prediction and dense recommendation modelling is constrained.

3.2 Source Tables

The original data contains nine CSV sources:

- Customers
- Orders
- Order Items
- Payments
- Reviews
- Products
- Sellers

- Geolocation
- Product Category Translation

The main relationships are represented through customer, order, product, seller, payment, and review identifiers.

Table: Main CRM Data Entities

Entity	Role
Customer	Customer identity and location
Order	Transaction and delivery lifecycle
Order Item	Product, seller, price and freight
Payment	Payment method and installments
Review	Customer review score and comments
Product	Product category
Seller	Seller information
Geolocation	Geographic information

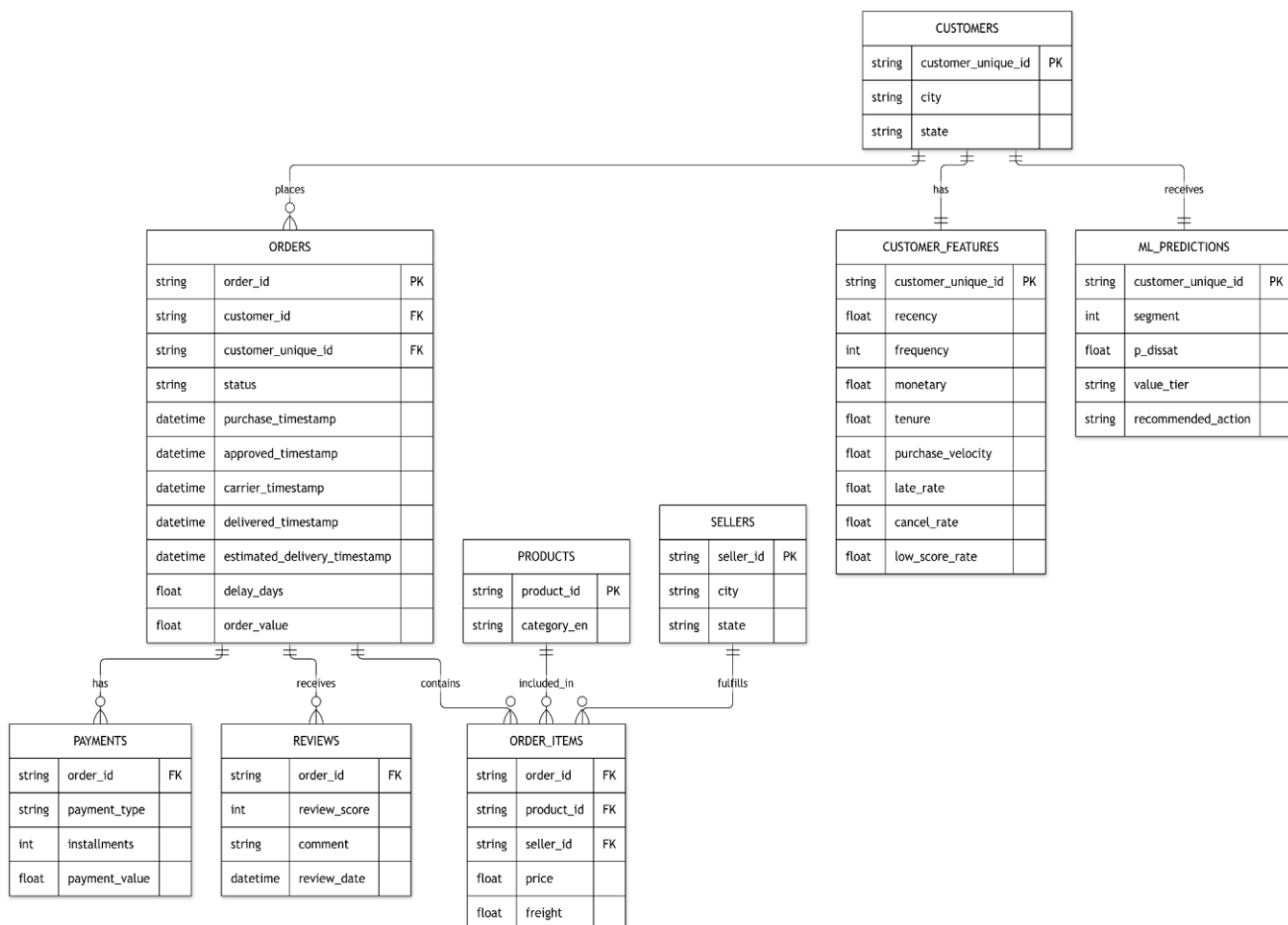


Figure 2: Entity-Relationship Diagram of the CRM Data Model

3.3 Data Integration

The raw datasets were joined around the order and customer relationships.

The resulting enriched order table contains:

- Order status

- Purchase timestamp
- Approval timestamp
- Carrier delivery date
- Customer delivery date
- Estimated delivery date
- Product information
- Seller information
- Payment information
- Review score
- Delivery delay
- Delivery duration
- Approval lag
- Order value
- Freight ratio
- Cancellation status
- Dissatisfaction indicator

The enriched order dataset contains **99,441 orders**.

Customer-level aggregation was then performed using *customer_unique_id* as the customer grouping key.

3.4 Data Cleaning and Feature Preparation

The preprocessing stage calculated several derived variables.

Transaction features

- Number of orders
- Monetary value
- Average order value
- Number of items
- Number of sellers
- Product-category diversity

Temporal features

- Recency
- Tenure
- Purchase velocity
- Delivery time
- Approval lag

Service features

- Late delivery
- Late-delivery rate
- Cancellation rate
- Average delivery delay

Customer feedback features

- Average review score
- Low-score rate

- Comment availability
- Comment length

Payment features

- Dominant payment type
- Installment behavior
- Credit-card ratio
- Average installments

These variables form the basis of the Customer 360 feature table.

3.5 Exploratory Data Analysis

The exploratory analysis revealed that reviews are heavily skewed toward positive ratings.

- 5-star: 57.8%
- 4-star: 19.3%
- 3-star: 8.2%
- 2-star: 3.2%
- 1-star: 11.5%

Review text is incomplete. Review titles are missing for approximately **88.34%** of records, while review messages are missing for **58.7%**. Approximately **7.87%** of all orders are classified as late using the study's overall-order calculation.

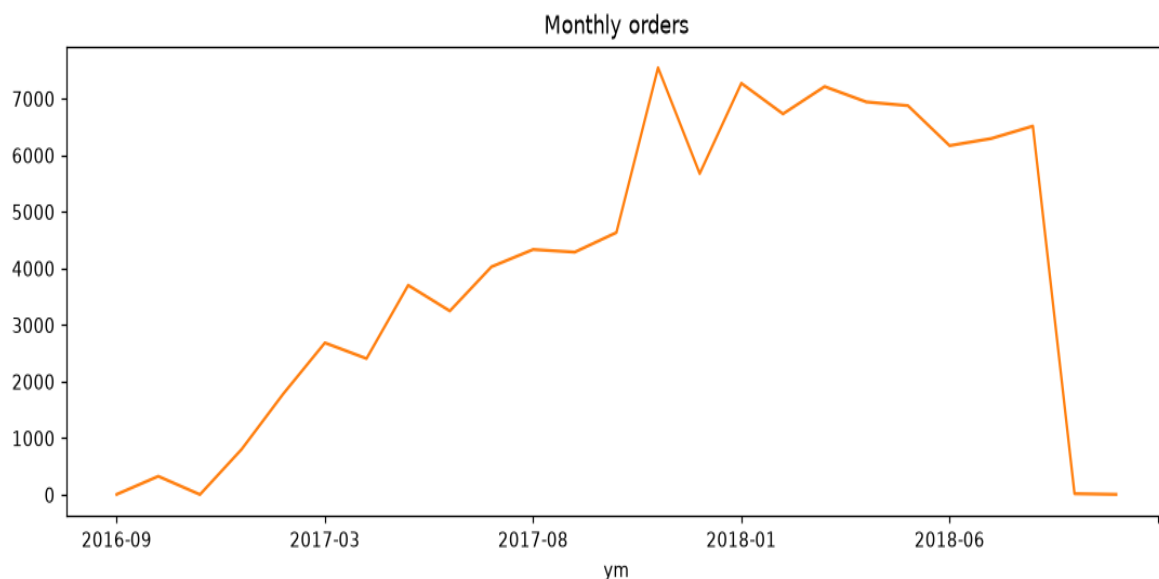


Figure 3: Monthly Orders

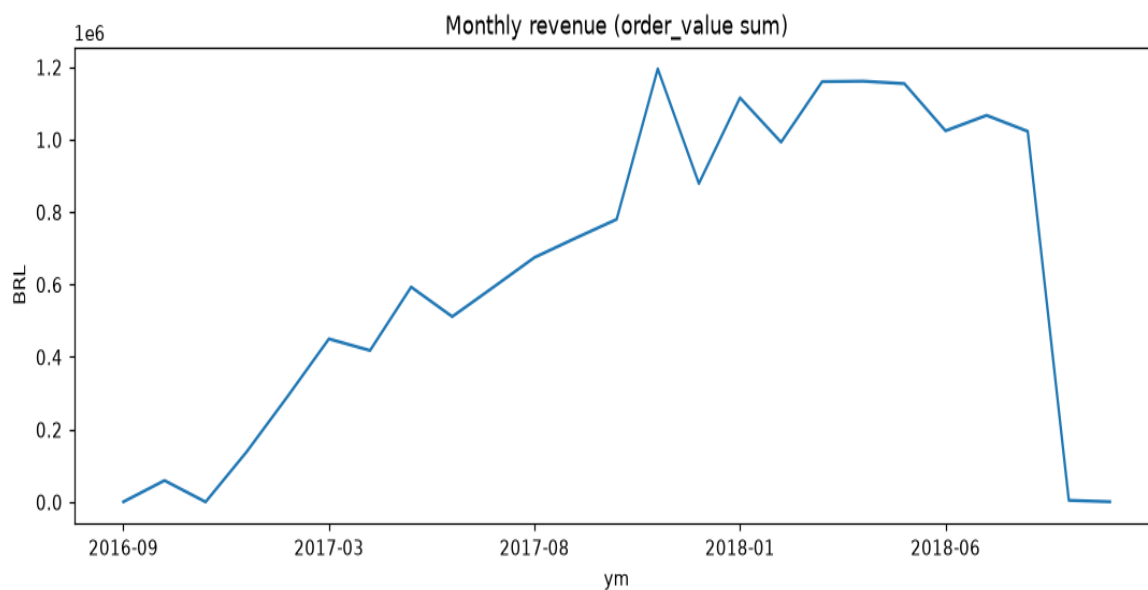


Figure 4: Monthly Revenue

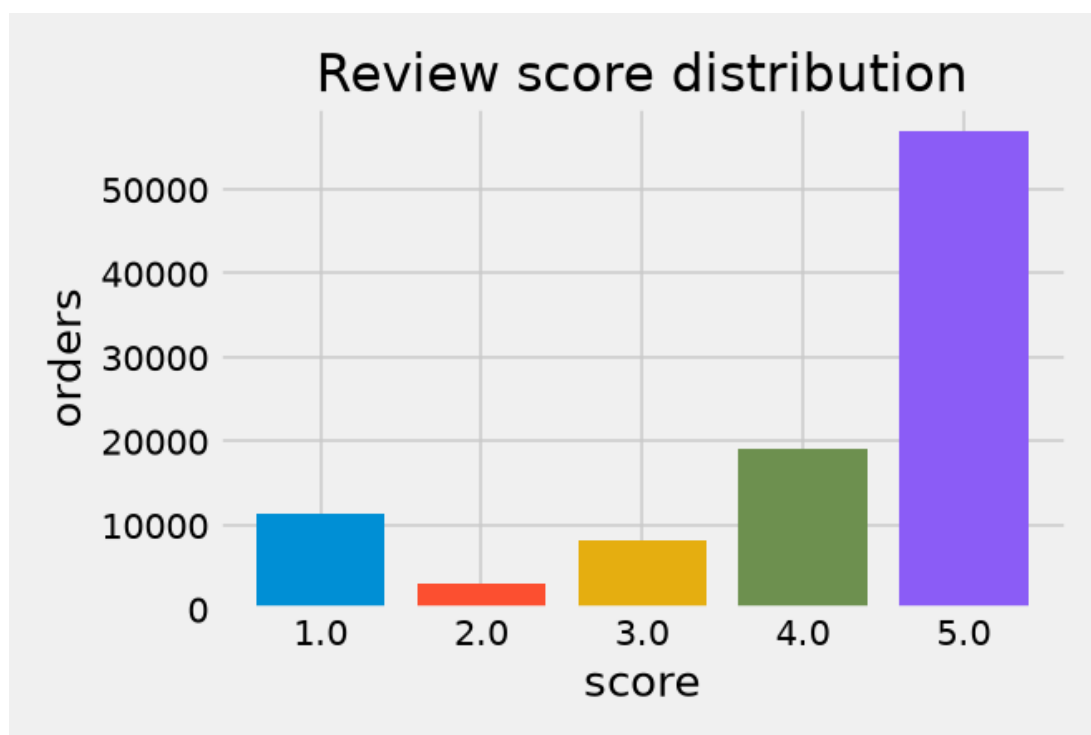


Figure 5: Review Score Distribution

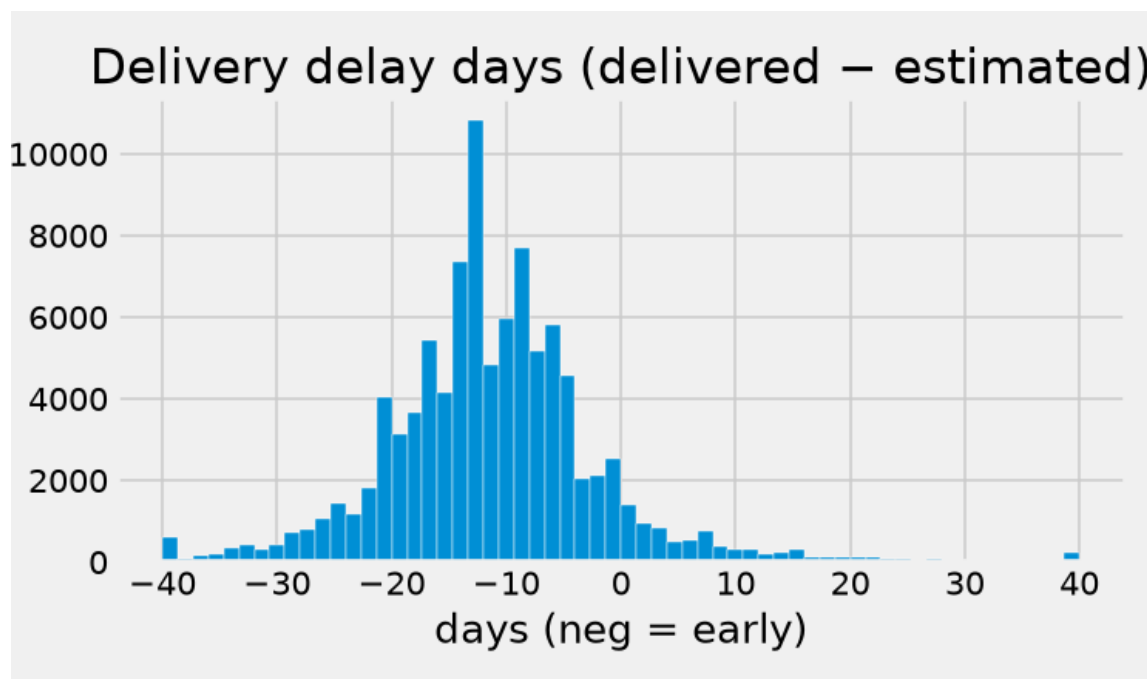


Figure 6: Delivery Delay Distribution

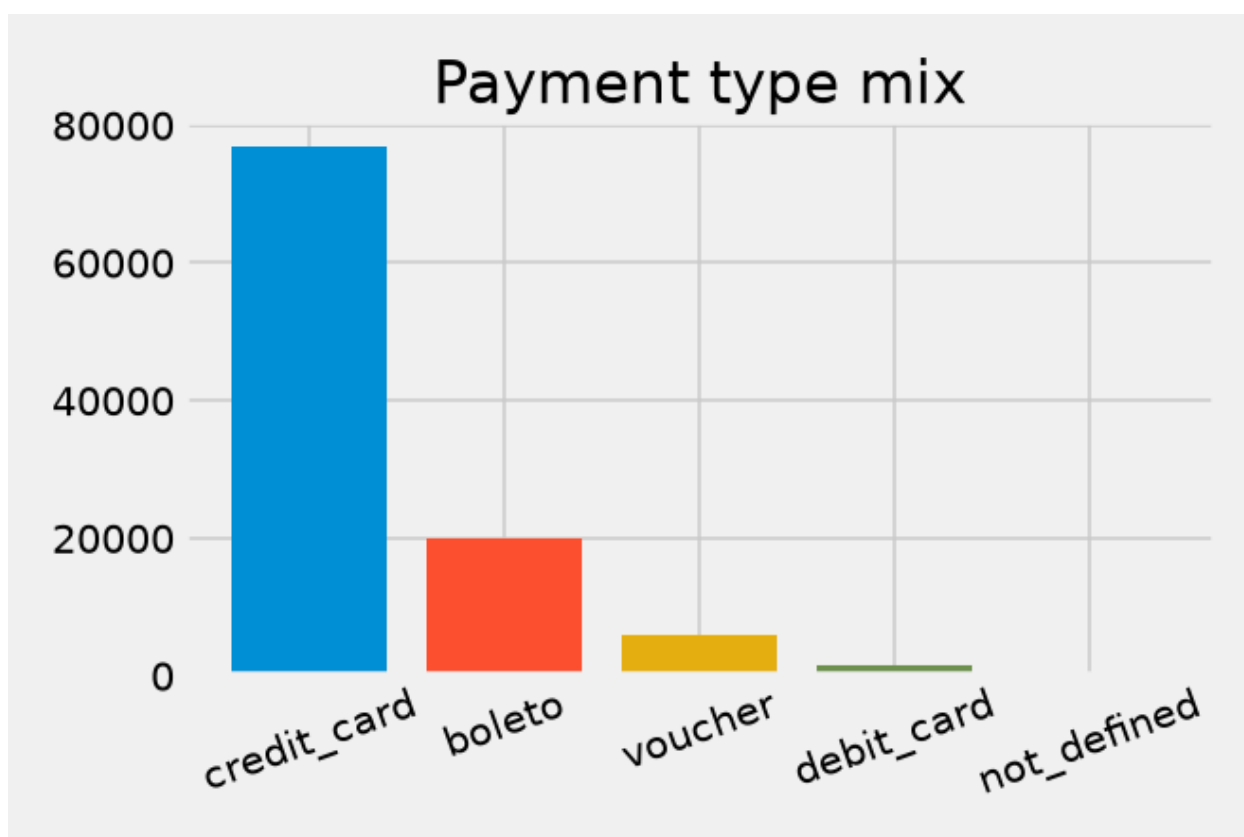


Figure 7: Payment Mix

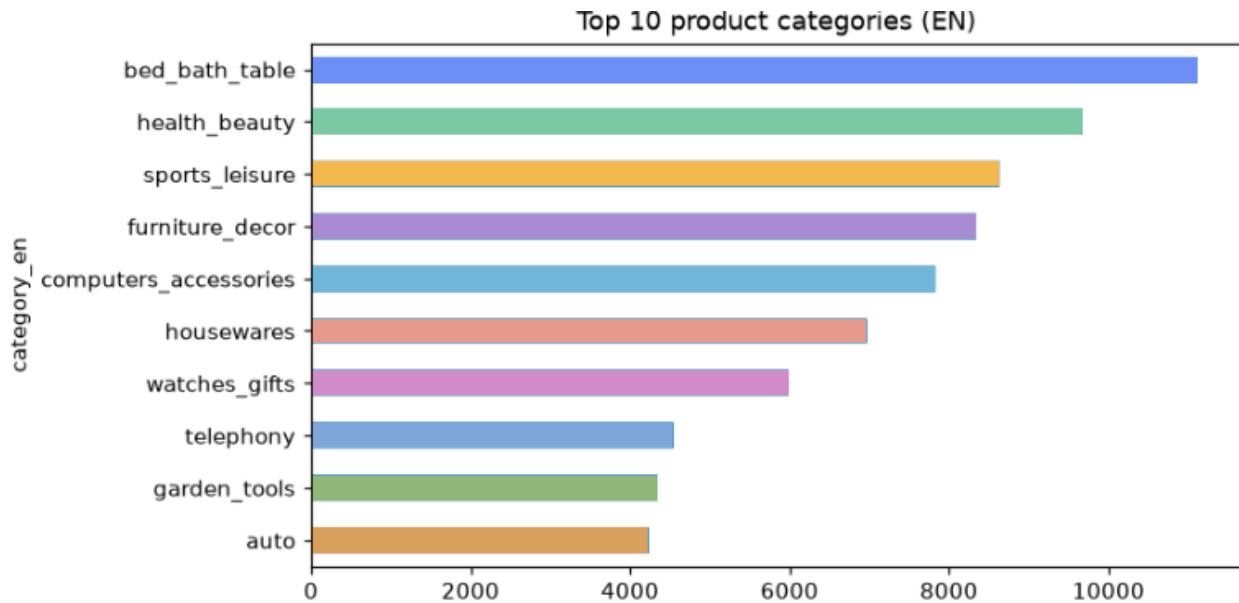


Figure 8: Top Product Categories

These figures establish the characteristics of the data before the machine learning experiments.

4. Proposed Intelligent CRM Framework

4.1 Overall Architecture

The proposed framework consists of six major layers:

1. Raw customer and transaction data
2. Data processing
3. Customer 360
4. Feature engineering
5. Machine learning and customer intelligence
6. CRM actions

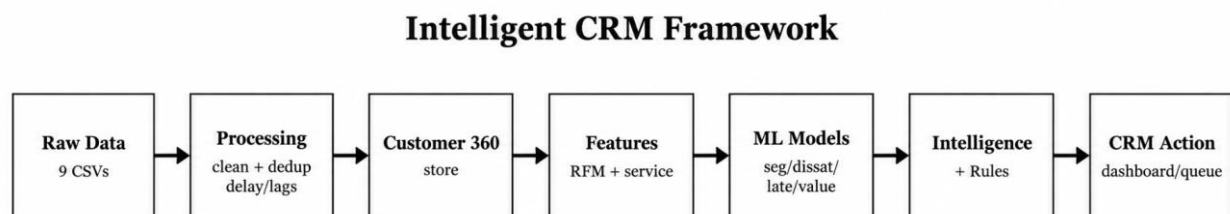


Figure 9: Proposed Intelligent CRM Framework Pipeline

4.2 Customer Data Layer

The data layer combines information from customer, order, item, payment, product, seller, and review datasets.

The purpose of this layer is to establish a common customer identifier and connect transaction-level information to customer-level information.

The primary customer grouping key used in the system is *customer_unique_id*.

4.3 Customer 360

The Customer 360 layer transforms individual transaction records into a consolidated customer profile.

Each customer profile contains:

- Purchase history
- Monetary value
- Recency
- Frequency
- Tenure
- Delivery performance
- Review behavior
- Payment behavior
- Category preferences
- Customer segment
- Predicted dissatisfaction probability
- Value tier
- Recommended CRM action

For example, the implemented Customer 360 structure can show a customer's order history together with their review score, delivery information, segment, dissatisfaction probability, value tier, and recommended action.

4.4 Feature Engineering Layer

The feature engineering layer transforms transactional data into customer-level analytical variables.

The major features include:

$$Recency = T_0 - LastPurchase$$

where $T_0 = 2018 - 10 - 18$.

Frequency is represented by the number of purchases made by the customer.

Monetary value is based on the customer's transaction value.

Additional service and behavioral features include:

- Late rate
- Cancellation rate
- Average review score
- Average delivery delay
- Average delivery time
- Payment behavior
- Category diversity

These features are subsequently used for segmentation and prediction.

4.5 Intelligence Layer

The intelligence layer contains four principal components:

Customer segmentation

K-Means clustering was used to identify behavioral customer groups.

Dissatisfaction prediction

Machine learning was used to estimate the probability that a customer's order would receive a review score of 3 or lower.

Value classification

Customer value tiers were evaluated using leakage-free behavioral features.

Risk queues

Predictions were converted into customer and order risk queues that can be displayed within the CRM interface.

4.6 CRM Action Layer

The final layer converts analytical outputs into operational recommendations.

The implemented rules include:

Condition	CRM Action
$P(dissatisfaction) > 0.6$	Escalate to service queue
Order cancelled	Consider alternative seller
Late rate > 0.3	Use conservative delivery promise
High-value + dormant	VIP win-back

This layer is important because the purpose of predictive CRM is not only to generate predictions but to support customer-facing decisions.

5. EXPERIMENTAL METHODOLOGY

5.1 Experimental Setup

The modelling experiments use a temporal split rather than a random split.

This setup is intended to reduce temporal leakage and better represent the situation where historical information is used to predict future customer outcomes.

5.2 Customer Segmentation

K-Means clustering was evaluated using $k = 3,4,5,6$.

The silhouette scores were:

k	Silhouette
3	0.221
4	0.224
5	0.227
6	0.224

The Davies-Bouldin values were:

k	Davies-Bouldin
3	1.371
4	1.345
5	1.265
6	1.211

Five clusters were selected for the operational CRM PoC.

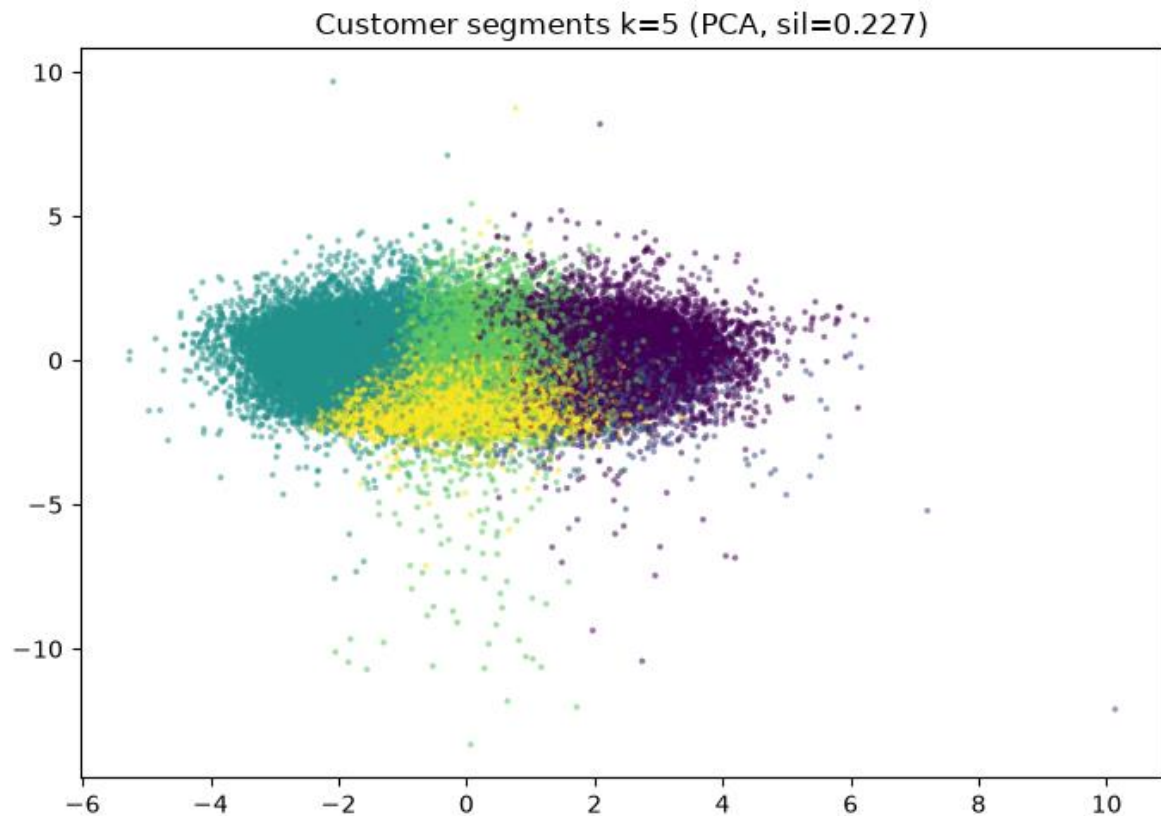


Figure 10: Customer Segments (k = 5) Projected with PCA

5.3 Dissatisfaction Prediction

The test dataset contains **46,434 observations**, with a dissatisfaction base rate of **22.54%**.

Three models were evaluated:

- Logistic Regression
- Random Forest
- HistGradientBoosting

RESULTS

Model	ROC-AUC	PR-AUC	F1
Logistic Regression	0.781	0.584	0.453
Random Forest	0.798	0.657	0.581
HistGradientBoosting	0.811	0.668	0.578

HistGradientBoosting produced the highest ROC-AUC and PR-AUC and was therefore used as the primary predictive model in the CRM intelligence workflow.

The most influential features included delivery delay, delivery time, and freight ratio.

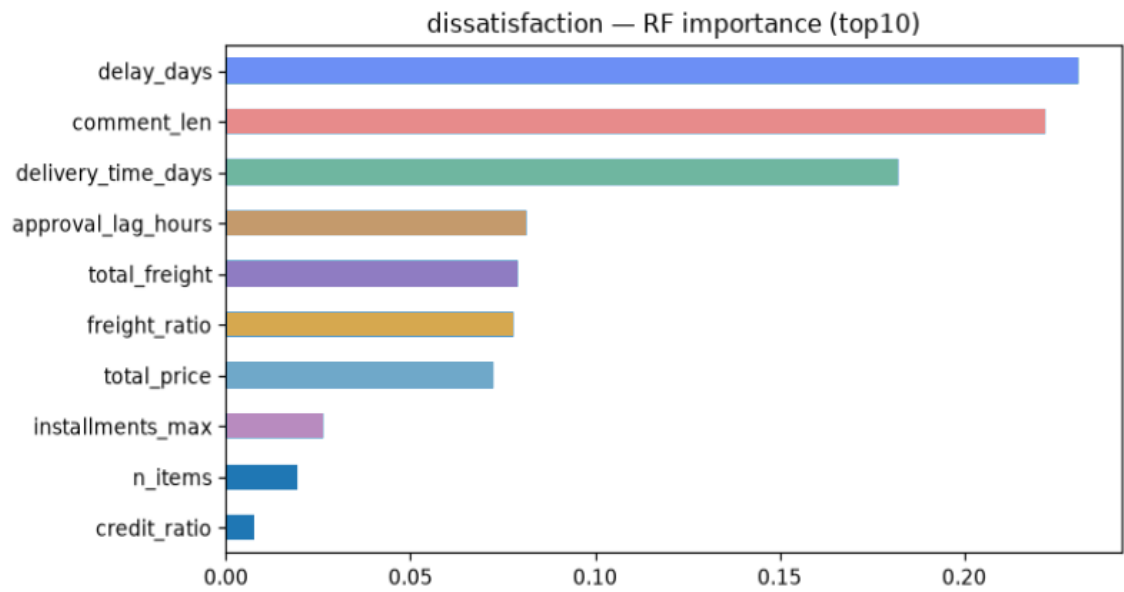


Figure 11: Top 10 Random Forest Feature Importances for Dissatisfaction Prediction

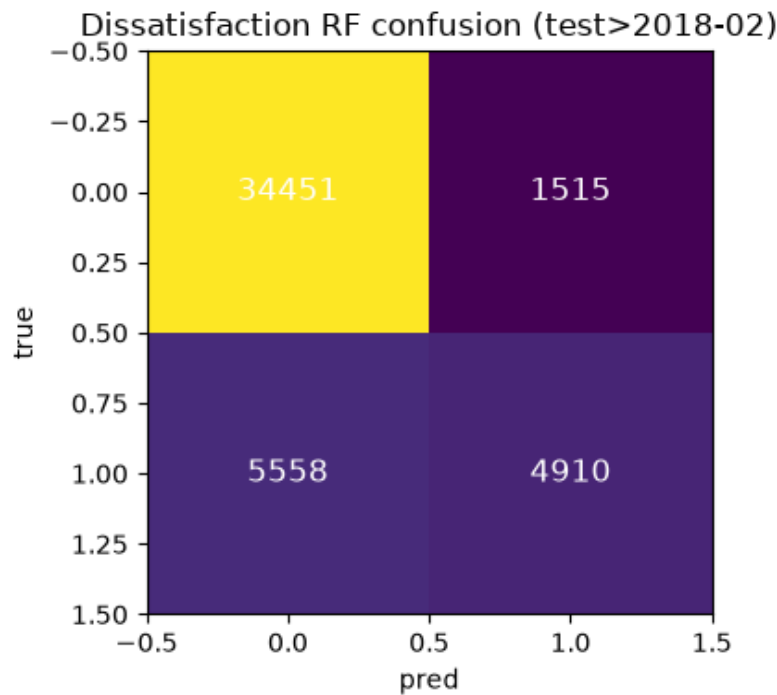


Figure 12: Random Forest Confusion Matrix for Dissatisfaction Prediction

5.4 Late Delivery Prediction

Late delivery was evaluated as another prediction task.

The test set contains **45,708 observations**, with a base rate of approximately **9.8%**.

The models performed poorly:

Model	ROC-AUC	PR-AUC	F1
Logistic Regression	0.537	0.107	0.001
Random Forest	0.510	0.100	0.000
HistGradientBoosting	0.530	0.104	0.000

These results indicate that the available features do not provide sufficient information for reliable late-delivery prediction.

5.5 Customer Value Prediction

An initial value-tier model produced an apparently extremely high F1-macro score of 0.9996.

However, inspection revealed that the features included variables derived directly from monetary value, such as AOV and order count/value relationships. These variables introduced target leakage.

Therefore, the value-tier experiment was repeated using leakage-free behavioral features:

- Recency
- Tenure
- Average delay
- Average freight ratio
- Average installments
- Category diversity
- Late rate
- Average review score

The corrected model achieved:

- F1-macro = **0.680**
- Balanced accuracy = **0.682**

The leakage-free result is used in the final interpretation.

5.6 Order Value Prediction

Similarly, an initial order-value prediction model produced an R^2 of approximately 1.0.

This result was identified as leakage because price and freight were directly used to construct the target order value.

After removing the leaking variables, the HistGradientBoosting model achieved:

- RMSE = **204.26**
- MAE = **84.84**
- R^2 = **0.146**

This indicates relatively weak predictive performance.

This experiment demonstrates the importance of leakage detection when evaluating CRM machine learning models.

Among the eligible customers, the single-purchase rate is approximately **96.1%**.

6. RESULTS

6.1 Overall Findings

The experiments demonstrate different levels of success across CRM intelligence tasks.

Task	Result
Customer segmentation	Operationally usable but weak separation
Dissatisfaction prediction	Strongest predictive result
Late-delivery prediction	Poor
Value classification	Moderate after leakage removal
Order-value prediction	Weak
Churn	Not suitable for ML using current data

The results demonstrate that the suitability of machine learning depends strongly on the information available in the underlying CRM data.

6.2 Segmentation Results

The five customer segments produced by the clustering model can be summarized as follows:

Segment	General profile
S0	High-spend single-purchase customers
S1	Repeat-oriented customers
S2	Low-value customers
S3	Dormant/lower-score customers
S4	Recent and relatively satisfied customers

The most important segment is S1 from a behavioral perspective because it is the only segment with substantial repeat purchasing. S1 contains approximately **2,997 customers**, with an average frequency of approximately **2.12 purchases**.

The weak overall clustering separation indicates that the dataset does not contain enough repeated customer behavior to form highly distinct lifecycle groups.

6.3 Dissatisfaction Results

Dissatisfaction prediction produced the strongest machine learning result.

The HistGradientBoosting model achieved:

$$\text{ROC-AUC} = 0.811$$

and

$$\text{PR-AUC} = 0.668.$$

The dissatisfaction base rate was 22.5%.

The Random Forest confusion matrix contained:

- True negatives: 34,451
- False positives: 1,515
- False negatives: 5,558
- True positives: 4,910

The results indicate that service-related information, particularly delivery-related variables, contains useful information for identifying customers at higher risk of dissatisfaction. Approximately **21.8% of late orders still received a five-star review**, while approximately **51.2% of one-star orders were classified as on-time**. This demonstrates that customer satisfaction cannot be reduced to delivery delay alone.

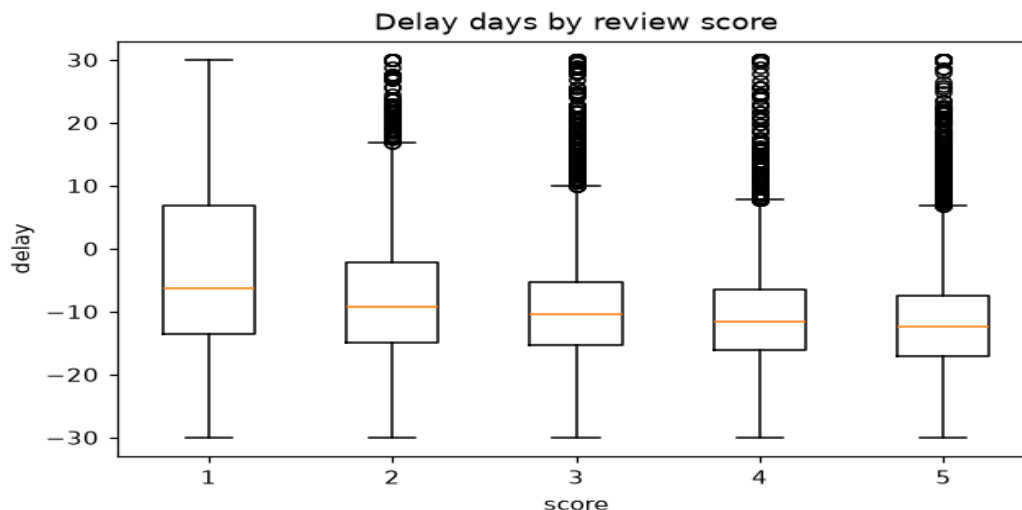


Figure 13: Delivery Delay by Review Score

6.4 Late Delivery Results

Late-delivery prediction produced ROC-AUC values between approximately 0.51 and 0.54.

This is close to random discrimination.

The result suggests that variables available in the dataset, such as price, freight, installments, and approval-related information, do not sufficiently explain future late delivery.

This finding is important for the proposed CRM framework because it demonstrates that not every operational problem can be solved through machine learning using the current data.

6.5 Value Results

The leakage-free value-tier experiment achieved an F1-macro score of 0.680.

This represents a substantially more realistic result than the original F1-macro of 0.9996.

The comparison demonstrates the importance of separating predictive inputs from variables used to construct the target.

Similarly, the leakage-free order-value model achieved an R^2 of only 0.146 compared with the original near-perfect result.

These findings support the use of leakage checks as part of the experimental methodology.

6.6 CRM Action Results

The ML outputs were connected to CRM rules.

For example:

A customer with a predicted dissatisfaction probability above 0.6 is added to the service escalation queue.

Other actions include:

- cancellation → alternative seller action
- high late rate → conservative delivery expectation
- high value + dormancy → VIP win-back action

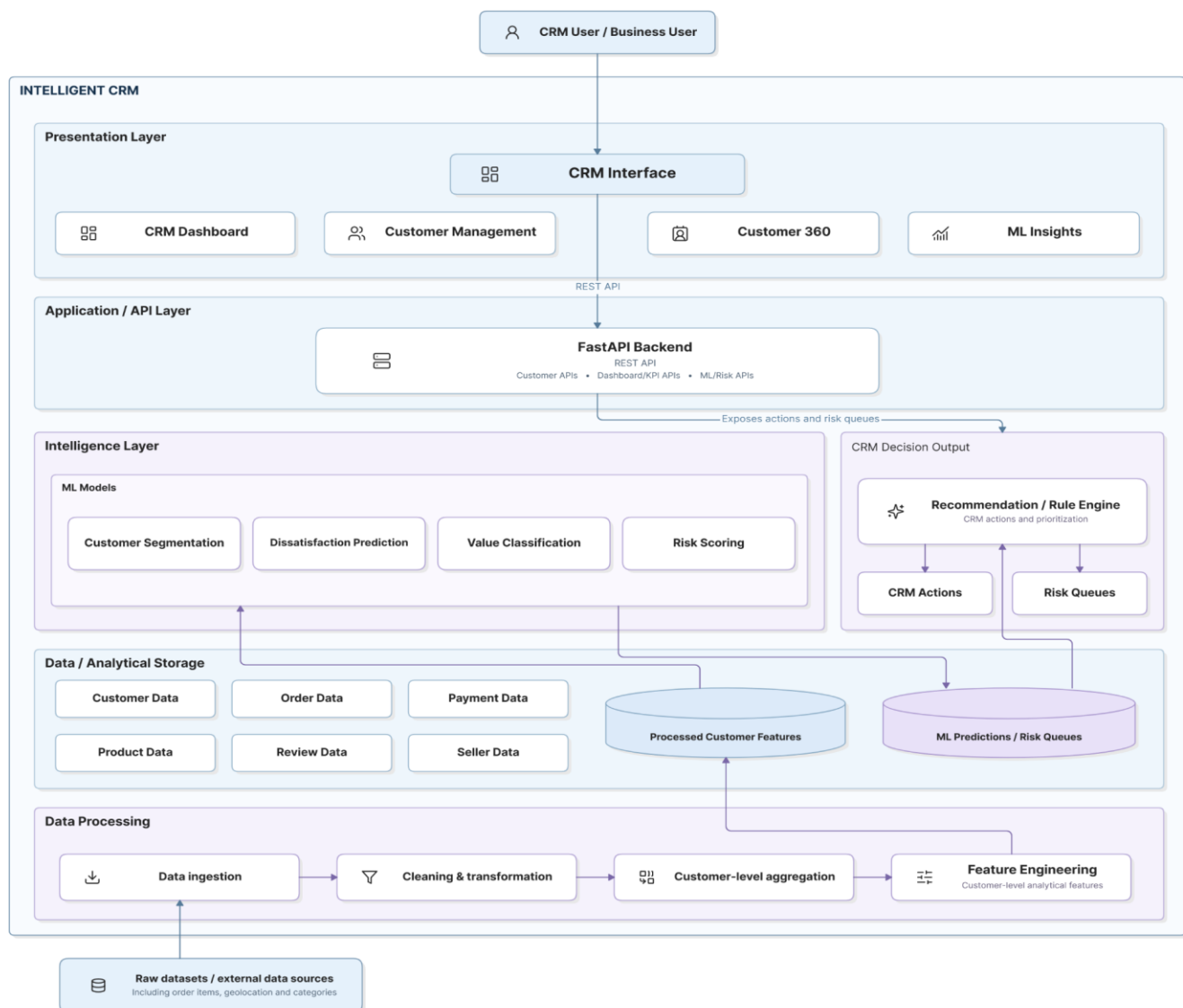
Thus, the final system transforms ML output into an operational CRM recommendation rather than presenting the prediction as an isolated score.

7. CRM PROOF OF CONCEPT

7.1 System Architecture

The working PoC was implemented using:

- **FastAPI** for the backend
- **HTML, CSS, JavaScript** for the interactive CRM interface
- Precomputed customer features and ML outputs
- CSV-based analytical storage for the experimental implementation



7.2 CRM Dashboard

The Dashboard provides a high-level view of the customer and transaction environment. It presents information such as:

- Customer volume
- Order volume
- Revenue
- Customer segments
- Review distribution
- Risk information

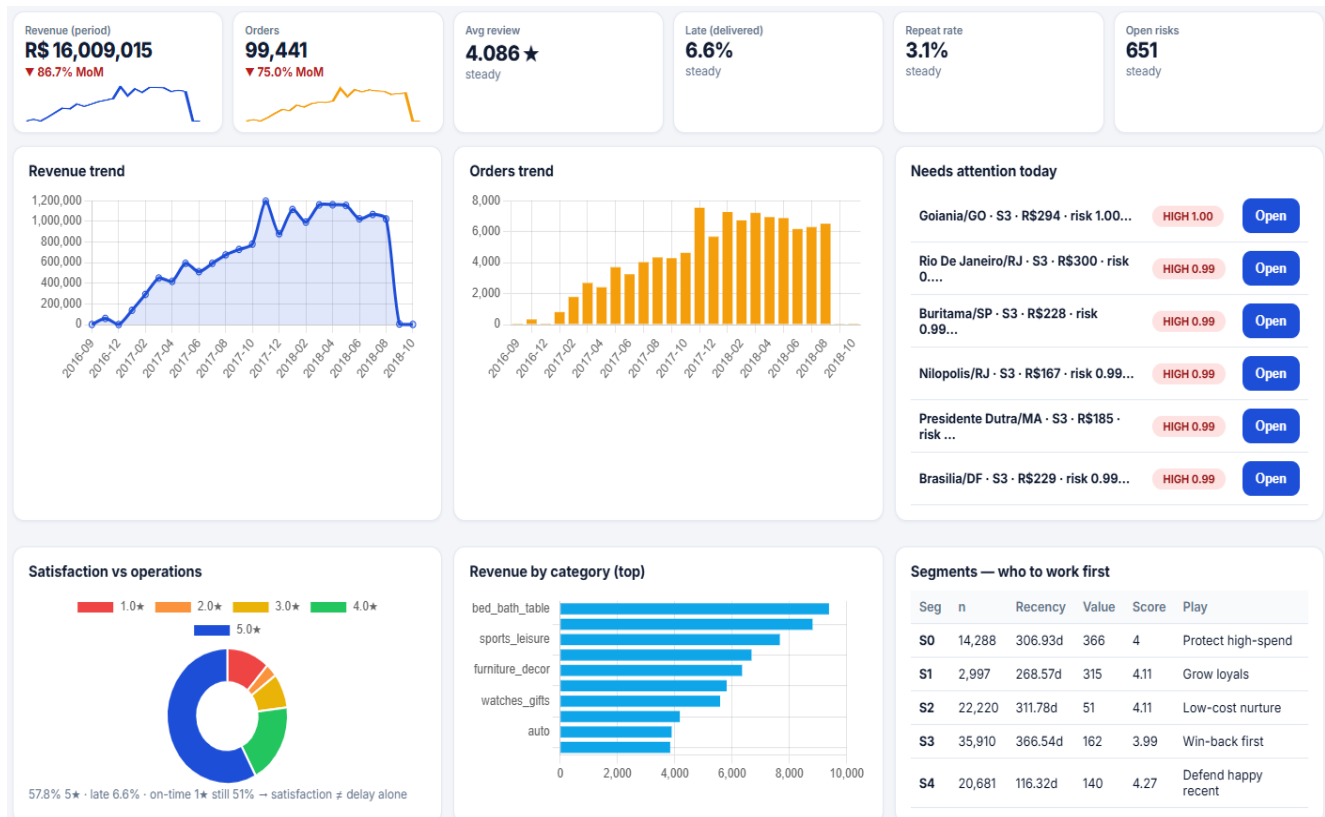


Figure 15: CRM Dashboard

The purpose of the dashboard is to provide CRM users with a starting point before moving to individual customer investigation.

7.3 Customer Management

The Customers page provides access to customer-level records.

Users can identify customers using the consolidated customer feature table rather than navigating across separate transaction tables.

The page acts as the operational bridge between aggregate CRM analytics and Customer 360 investigation.

Customers 1,994 match

Customer	Segment	Orders	Spent	Recency	Score	Risk
Nova Odessa/SP · S3 · R\$177 · risk 1.00...	S3	1	R\$ 177	327d	1.0	HIGH 1.00
Goiania/GO · S3 · R\$294 · risk 1.00...	S3	1	R\$ 294	329d	1.0	HIGH 1.00
Rio De Janeiro/RJ · S3 · R\$300 · risk 0.99...	S3	1	R\$ 300	337d	1.0	HIGH 0.99
Buritama/SP · S3 · R\$228 · risk 0.99...	S3	1	R\$ 228	254d	1.0	HIGH 0.99
Nilopolis/RJ · S3 · R\$167 · risk 0.99...	S3	1	R\$ 167	228d	1.0	HIGH 0.99
Presidente Dutra/MA · S3 · R\$185 · risk 0.99...	S3	1	R\$ 185	242d	2.0	HIGH 0.99
Brasilia/DF · S3 · R\$229 · risk 0.99...	S3	1	R\$ 229	364d	1.0	HIGH 0.99
Bertioga/SP · S3 · R\$606 · risk 0.99...	S3	1	R\$ 606	309d	1.0	HIGH 0.99
Belem/PA · S0 · R\$384 · risk 0.99...	S0	1	R\$ 384	328d	1.0	HIGH 0.99
Santarem/PA · S3 · R\$354 · risk 0.99...	S3	1	R\$ 354	322d	2.0	HIGH 0.99
Sao Luis/MA · S3 · R\$244 · risk 0.99...	S3	1	R\$ 244	303d	1.0	HIGH 0.99
Rio De Janeiro/RJ · S1 · R\$204 · risk 0.99...	S1	2	R\$ 204	264d	3.0	HIGH 0.99
Niteroi/RJ · S3 · R\$203 · risk 0.99...	S3	1	R\$ 203	455d	1.0	HIGH 0.99
Rio De Janeiro/RJ · S3 · R\$180 · risk 0.99...	S3	1	R\$ 180	268d	1.0	HIGH 0.99
Chapada Do Norte/MG · S3 · R\$289 · risk 0.99...	S3	1	R\$ 289	252d	3.0	HIGH 0.99
Rio De Janeiro/RJ · S2 · R\$46 · risk 0.99...	S2	1	R\$ 46	318d	1.0	HIGH 0.99

Figure 16: Customer Management Page

7.4 Customer 360

The Customer 360 page provides a consolidated view of an individual customer. The implemented view includes:

- Customer profile
- Order history
- Products
- Reviews
- Segment
- Value tier
- Dissatisfaction probability
- Recommended action

This is one of the main outputs of the proposed framework because it combines descriptive and predictive information within a single customer context.

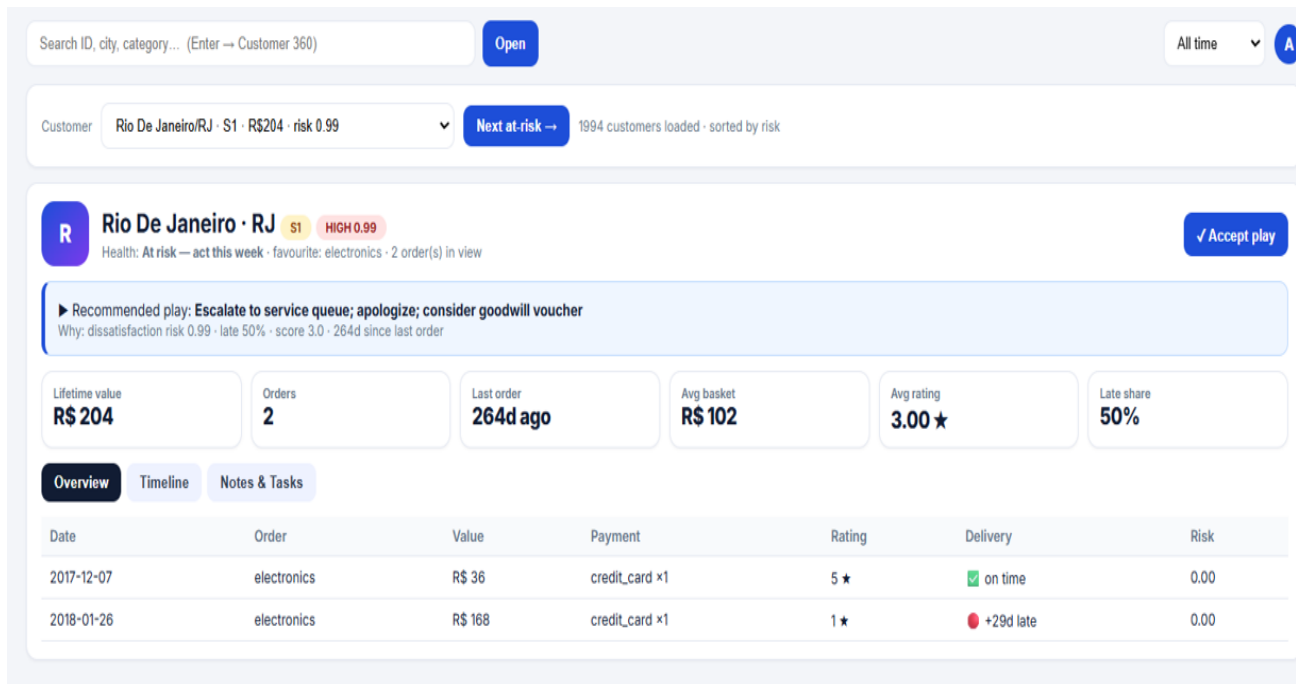


Figure 17: Customer 360 View

7.5 ML Insights

The ML Insights page presents the predictive outputs generated by the machine learning pipeline.

The system provides:

- Customer segments
- Dissatisfaction probabilities
- Risk queues
- Recommended actions

This allows CRM users to identify customers requiring attention rather than manually examining every customer record.

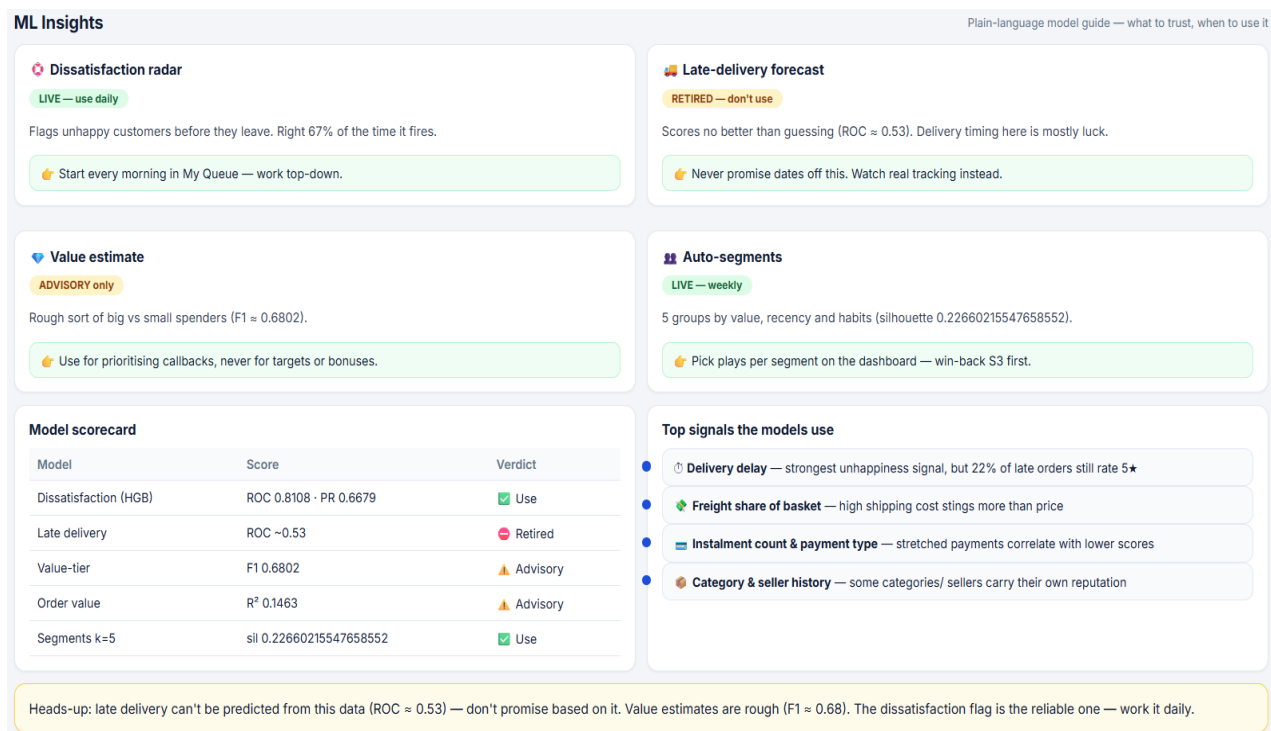


Figure 18: ML Insights Page

7.6 CRM Action Layer

The action layer converts model predictions into practical CRM recommendations.

For example:

Customer risk score → Service escalation

Cancellation → Alternative seller

High late rate → Conservative promise

High-value dormant customer → Win-back

This demonstrates the central principle of the proposed Intelligent CRM framework:

Machine learning becomes useful to CRM when predictive information is connected to an operational decision.

8. DISCUSSION

8.1 What Worked

The study demonstrates that customer transaction and review data can be transformed into a unified Customer 360 representation.

Customer segmentation provides operationally useful customer profiles despite weak cluster separation.

The strongest machine learning result was dissatisfaction prediction, with ROC-AUC of 0.811 and PR-AUC of 0.668.

The integration of these outputs into risk queues and recommended CRM actions demonstrates that the analytical layer can be connected to an operational CRM workflow.

8.2 Effect of Dataset Characteristics

The dataset has an extremely low repeat-purchase rate of 3.12%.

This creates a major limitation for customer lifecycle modelling.

Because most customers purchase only once, it is difficult to construct robust behavioral histories for:

- churn prediction
- lifetime value modelling
- recommendation systems
- customer journey modelling

Similarly, the dataset does not contain actual:

- sales leads
- opportunities
- marketing campaigns
- customer service tickets
- call interactions
- email engagement
- website behavior

Therefore, these capabilities are outside the scope of the current PoC.

8.4 Data Leakage Lessons

One of the important methodological findings is the effect of data leakage.

The initial value-tier model produced an F1-macro of 0.9996, while the initial order-value model produced R^2 of approximately 1.0.

After removing target-derived features, performance decreased substantially:

- Value tier: F1-macro = 0.680
- Order value: R^2 = 0.146

The corrected results are more realistic.

This demonstrates why CRM machine learning experiments require careful examination of how target variables are constructed and how features are generated.

8.5 CRM Implications

The findings suggest that an intelligent CRM system should not be considered only as a machine learning application.

Its effectiveness depends on three connected components:

1. **Data integration**
2. **Customer intelligence**
3. **Decision support**

If the underlying CRM does not contain information about customer interactions, campaigns, service cases, and engagement, the intelligence layer will also be limited.

Therefore, improving the data foundation is as important as improving the machine learning models.

9. LIMITATIONS AND FUTURE WORK

9.1 Limitations

The limitations are:

1. Very low repeat-purchase rate.
2. No direct sales lead information.
3. No campaign-response information.
4. No customer service ticket data.
5. No call-center interaction data.
6. No detailed website or clickstream behavior.
7. Incomplete review text.
8. Weak late-delivery prediction.
9. Limited support for true churn modelling.
10. Limited support for dense recommendation modelling.

9.2 Future Work

Future versions can extend the proposed framework by incorporating:

- Sales leads
- Opportunities
- Marketing campaigns
- Email interactions
- Website activity
- Customer service tickets
- Call-center interactions
- Complaints
- Customer feedback text
- Campaign response
- Real-time customer events

With these additional sources, the system could support more advanced capabilities such as:

- True churn prediction
- Customer lifetime value prediction
- Lead conversion prediction
- Campaign response prediction
- Recommendation systems
- Next-best-action models
- Real-time customer risk scoring
- More complete customer journey analytics

10. CONCLUSION

This study presented an Intelligent CRM proof of concept that integrates customer data processing, Customer 360 profiling, analytics, machine learning, and CRM actions.

The proposed workflow transforms heterogeneous customer and transaction information into customer-level features and predictive intelligence. Customer segmentation was implemented using K-Means clustering, while dissatisfaction prediction achieved the strongest predictive performance with a HistGradientBoosting ROC-AUC of 0.811 and PR-AUC of 0.668.

The study also demonstrated the importance of reporting unsuccessful experiments. Late-delivery prediction produced performance close to random, while leakage-free order-value prediction achieved an R^2 of only 0.146. These results show that predictive CRM capabilities depend strongly on the quality and scope of available customer data.

The resulting CRM proof of concept demonstrates how analytical outputs can be incorporated into Dashboard, Customer Management, Customer 360, ML Insights, risk queues, and rule-based CRM actions.

Overall, the work demonstrates that intelligent CRM is not solely a machine learning problem. It requires an integrated approach in which **customer data is consolidated, transformed into meaningful customer intelligence, evaluated using appropriate machine learning methods, and finally connected to practical CRM decisions.**

The current implementation provides a foundation for future CRM projects while clearly identifying the additional customer interaction, service, marketing, and engagement data required to develop a more comprehensive enterprise CRM intelligence system.