

# End-to-End Customer Satisfaction Prediction Using Machine Learning

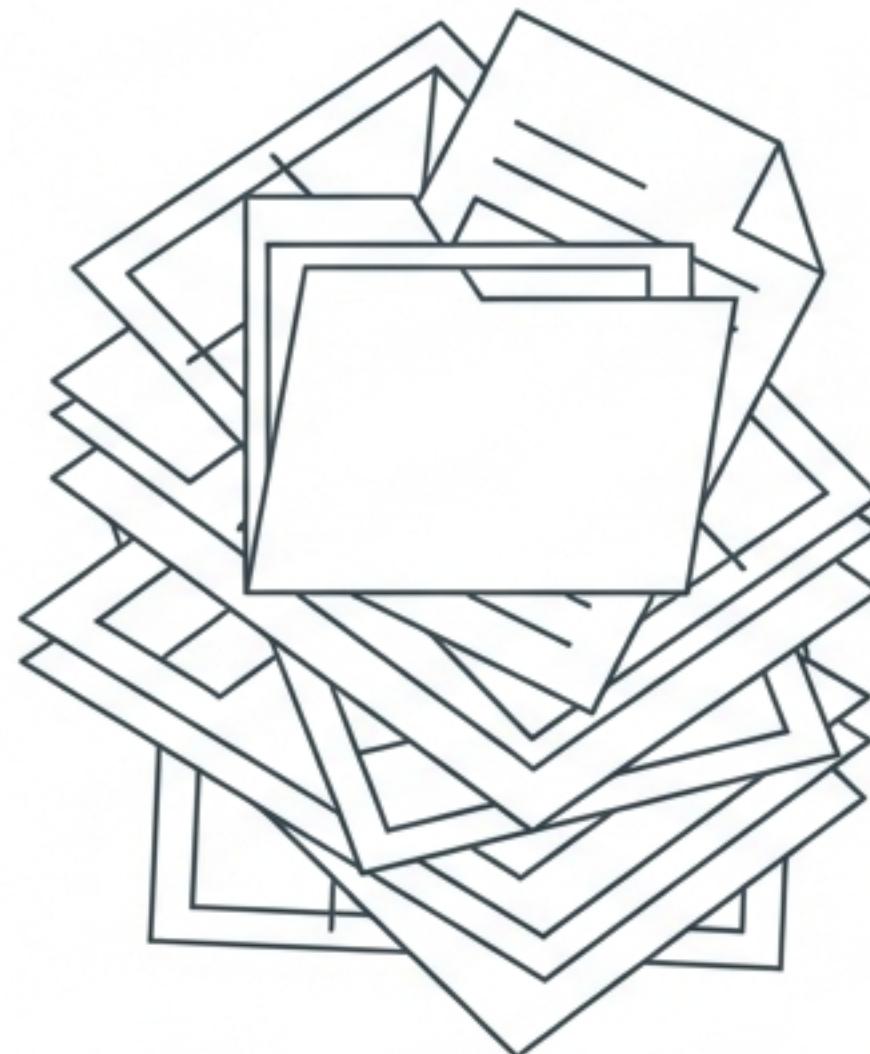
A Data Science Internship Project

Nasir Husain  
Data Science Intern  
UMID14112570495



# Manual Analysis Cannot Scale

## The Challenge



Organizations receive thousands of support tickets daily across email, chat, and social media.

## The Pain Point

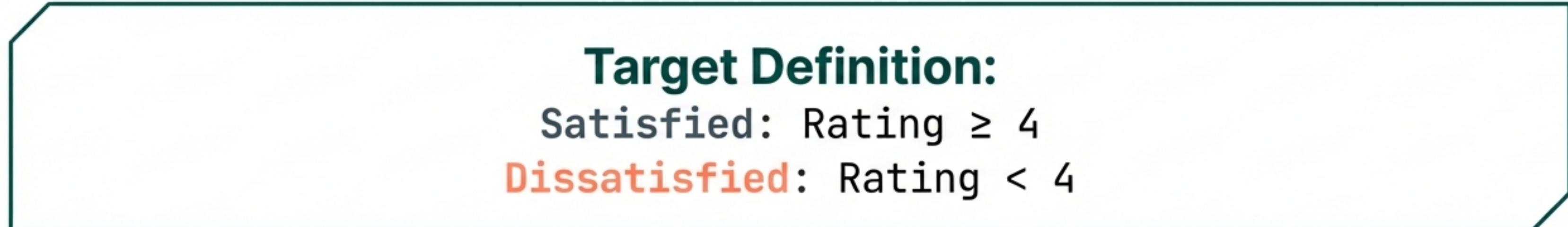
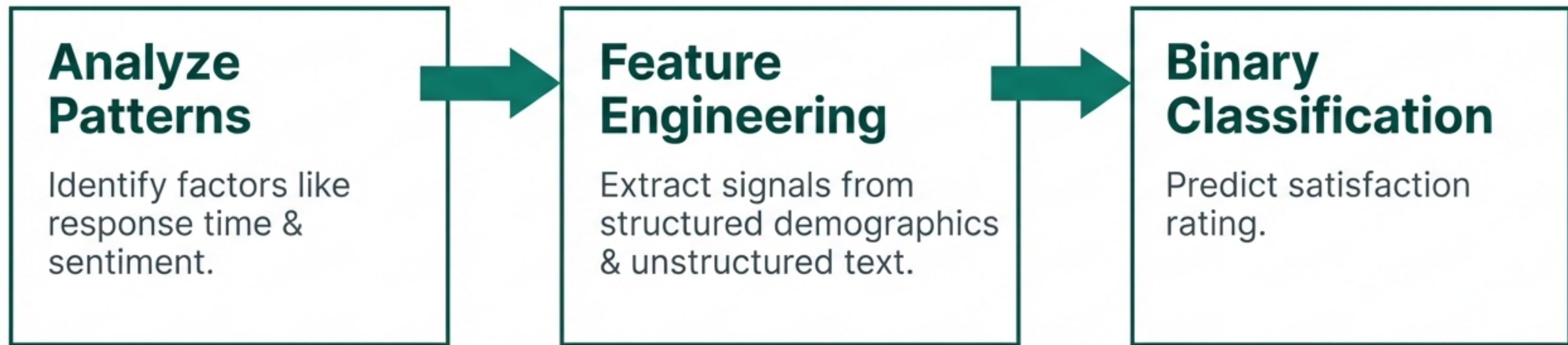
Manual review is reactive, error-prone, and inefficient.

## The Opportunity



Transforming support from a cost center to an intelligence hub through automated prediction.

# From Reactive Support to Predictive Intelligence



# The Dataset: Customer Support Tickets

**8,469**

Total Rows

**17**

Feature Columns

Ticket ID	Customer Name	Product	Description	Resolution	Rating
T-0012	Alice Chen	Laptop	Screen flickering....	Refunded	1
T-0045	Bob Smith	Headphones	Audio cuts out...	Replaced	4
T-0087	Charlie Kim	Software	Installation failed...	Patched	2

# Feature Taxonomy



## Structured Data

Customer Age

Gender

Product Purchased

Ticket Priority



## Unstructured Data

Ticket Subject

Ticket Description  
(NLP Required)



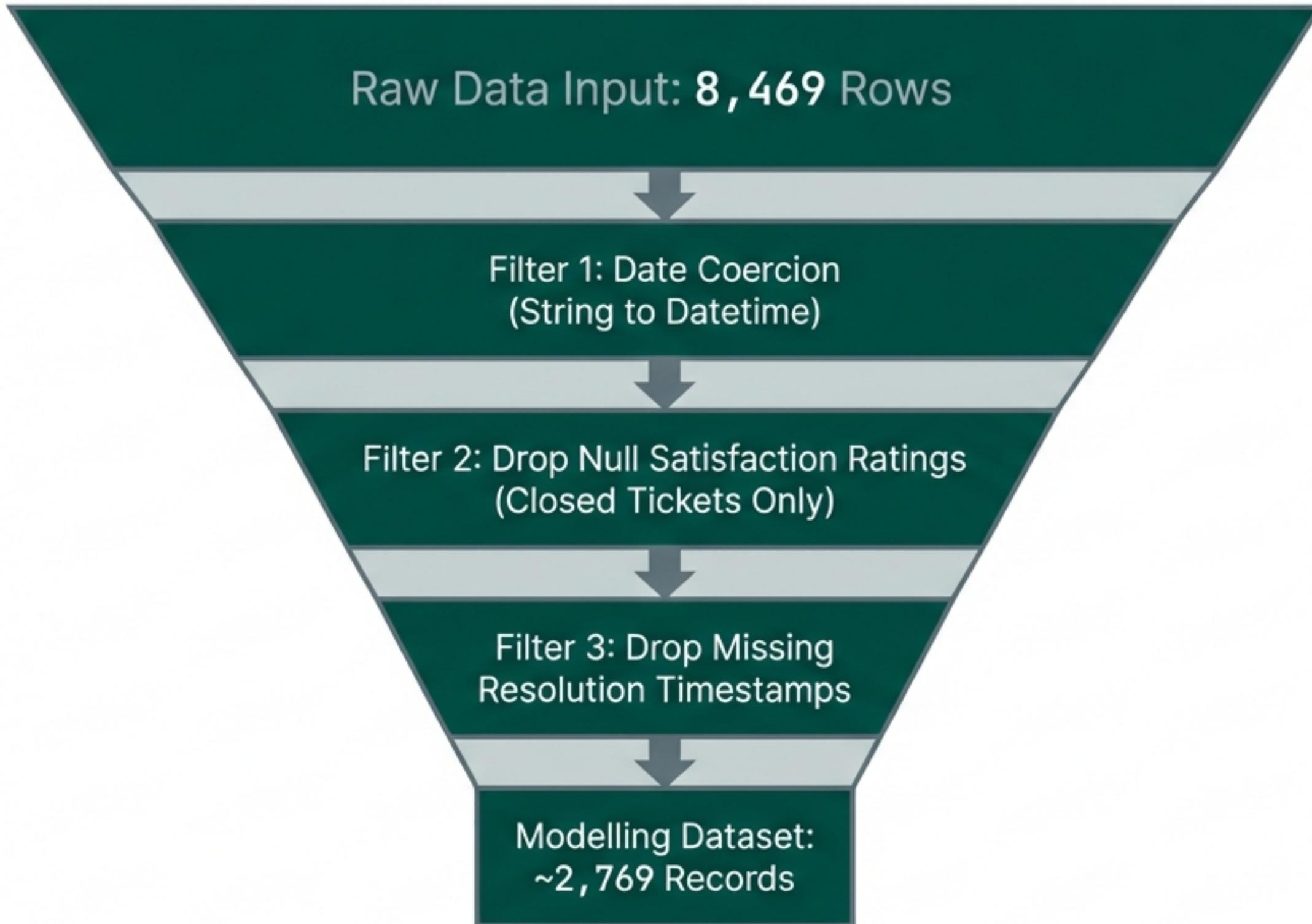
## Temporal Data

Date of Purchase

First Response Time

Time to Resolution

# Preprocessing Strategy: Refining the Input



**Objective:** Ensure Ground Truth accuracy by training only on closed tickets with complete time metrics.

# Feature Engineering: Signal from Noise

## Temporal Engineering

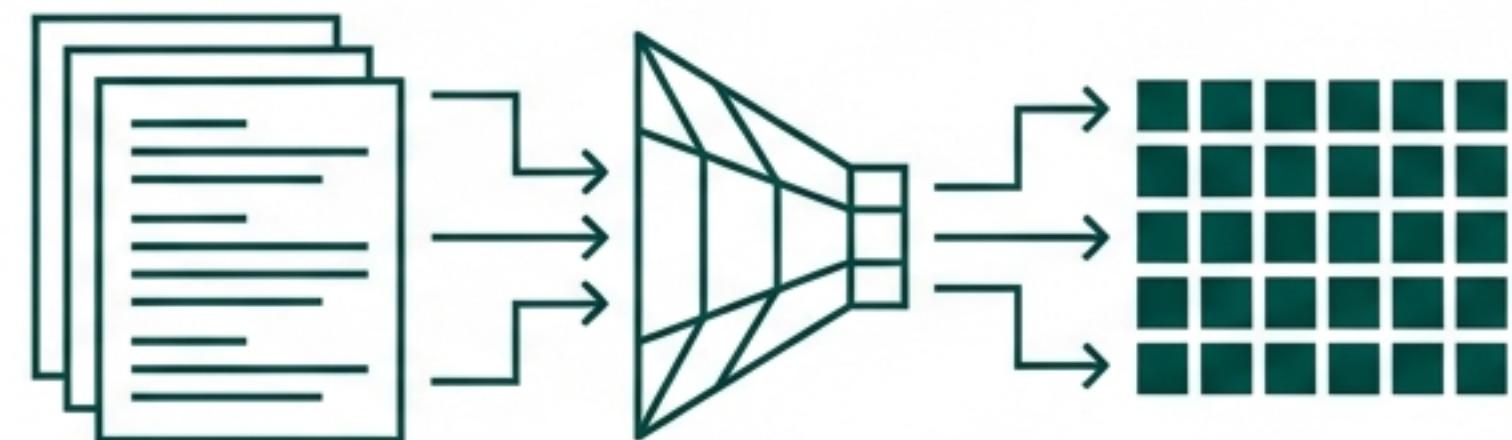
Inter Regular

```
response_delay_hours = First Response Time  
                         - Date of Purchase →  
JetBrains Mono           Inter Regular
```

```
resolution_time_hours = Time to Resolution  
                         - First Response Time →  
JetBrains Mono           Inter Regular
```

## Textual Engineering (NLP)

Inter Regular



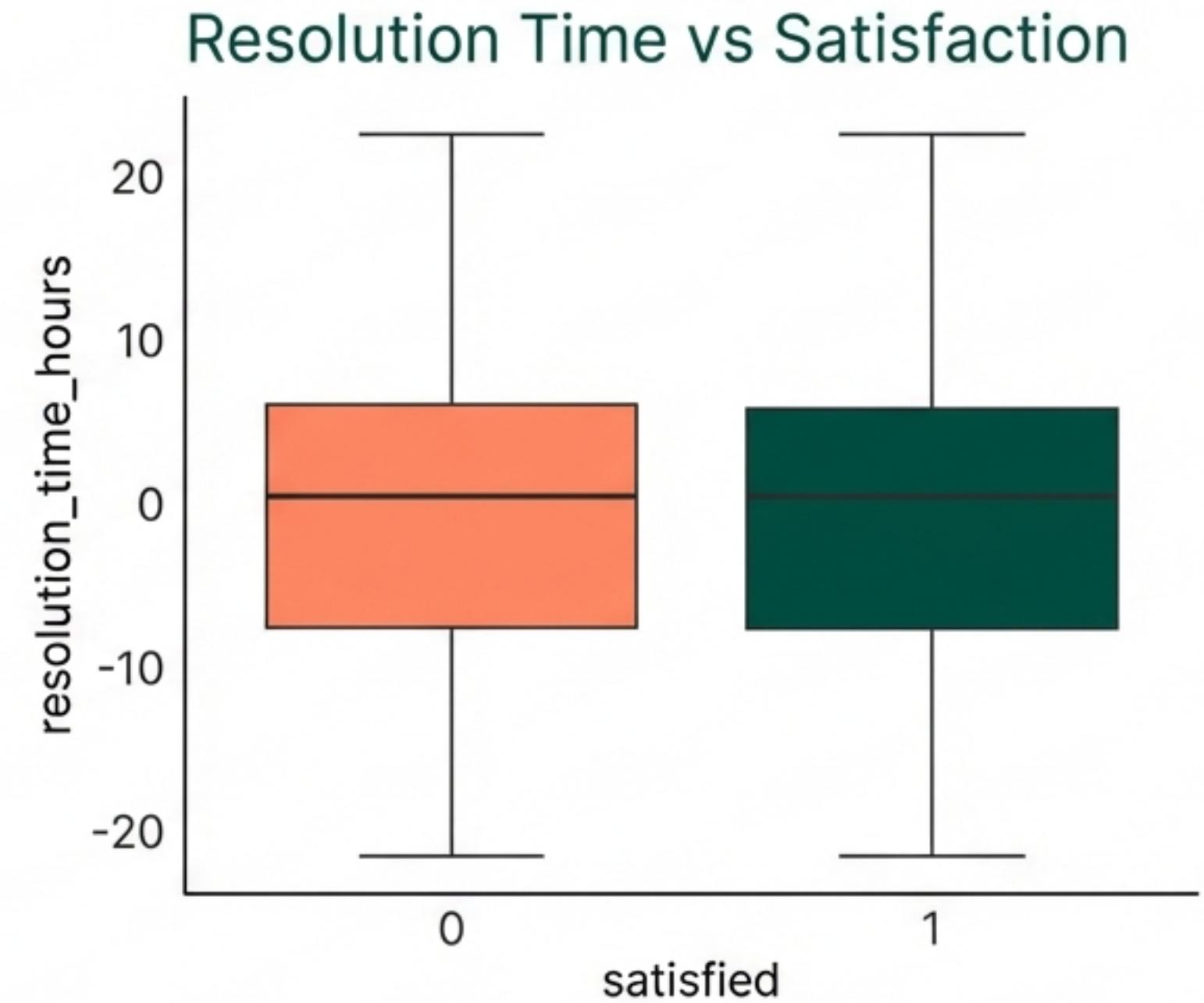
Input: 'Ticket  
Description'  
Inter Regular

TF-IDF  
Vectorisation  
Inter Regular

Output: 'Top  
300 Features'  
Inter Regular

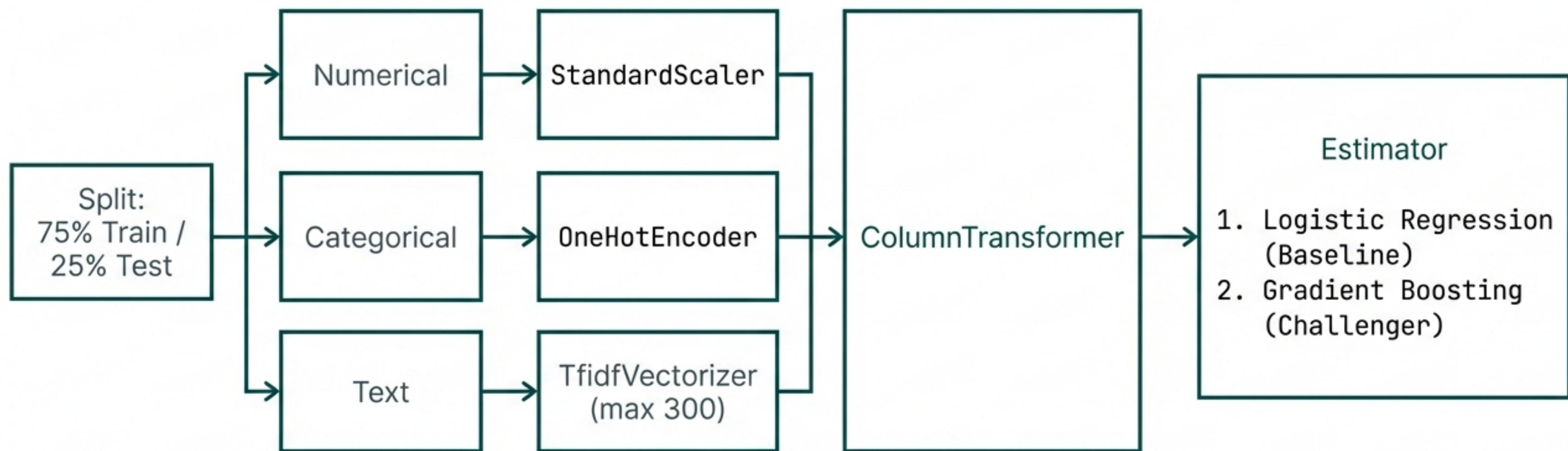
Purpose: Capture issue complexity and sentiment.  
Inter Regular

# Exploratory Analysis: The Impact of Time



# Modelling Architecture

## Scikit-Learn Pipeline Implementation



# Performance Evaluation

Baseline: Logistic Regression

**57.2%**

Accuracy

Linear model struggled with complex time/priority interactions.

Challenger: Gradient Boosting

**59.3%**

Accuracy

Non-linear ensemble captured complex patterns.



**Winner**

# Deep Dive: Gradient Boosting

Confusion Matrix - Gradient Boosting

		Predicted	
		0	1
Actual	0	391	30
	1	252	20
		Predicted	
0	1	391	30
1	0	252	20

## Model Behavior

- High Recall for Dissatisfied Customers (Class 0).
- Low Recall for Satisfied Customers (Class 1).
- Insight: The model is conservative. It prioritizes identifying unhappy customers but creates False Negatives (252) due to class imbalance.

# Business Insights: Speed is the Driver

## Time Matters



`response_delay_hours` &  
`resolution_time_hours` are the  
most significant predictors.

Optimize routing to reduce First  
Response delta.

## Complexity Signals



Ticket descriptions (TF-IDF)  
correlate with satisfaction.

Flag complex tickets for senior  
intervention

**Recommendation:** Prioritize tickets predicted as 'Dissatisfied' (Class 0) to prevent churn.

# Conclusion: A Scalable Framework



## Scalable Pipeline

Successfully converted raw text and timestamps into structured intelligence.



## Non-Linear Success

Gradient Boosting proved superior for complex support data patterns.



## Production Ready

Modular architecture (Pipelines) ensures reproducibility and prevents leakage.

The foundation is laid for a real-time decision-support system.

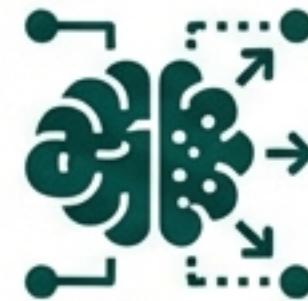
# Future Enhancements

## Optimising for Accuracy



### Address Imbalance

Implement *SMOTE* (Synthetic Minority Over-sampling) to improve *Recall* on Satisfied customers.



### Advanced NLP

Upgrade from *TF-IDF* to *BERT* embeddings for deeper sentiment understanding.



### Model Tuning

Execute *GridSearch* for hyperparameter optimization on *Gradient Boosting*.

# Thank You

## Q&A Session

Nasir Husain  
Data Science Intern  
UMID14112570495

[github.com/nasirhusain/customer-satisfaction-prediction](https://github.com/nasirhusain/customer-satisfaction-prediction)