AMERICAN EXPRESS DEFAULT PREDICTION

A Transformer-Based Approach

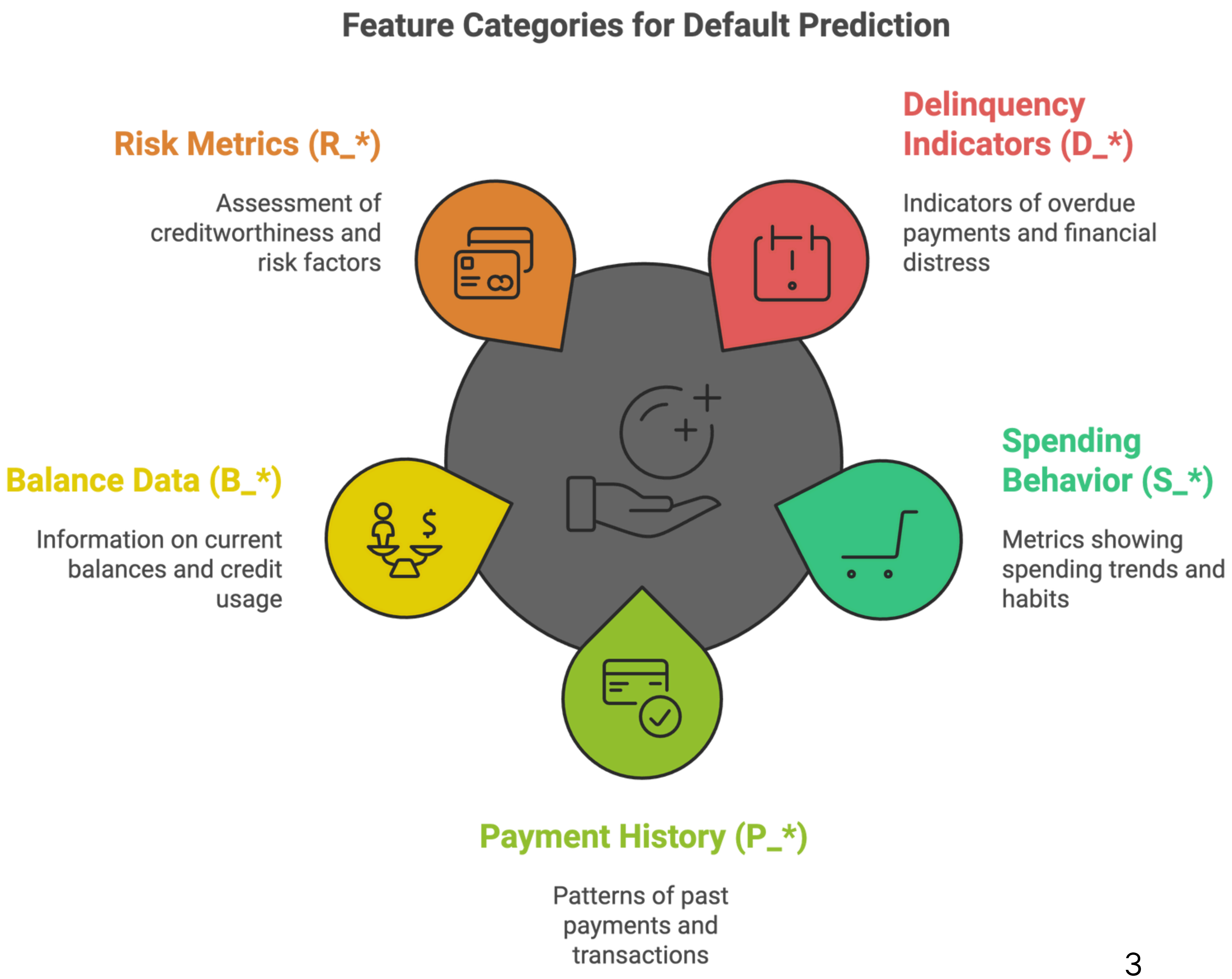
By: Mohamed Naoufal Mahfoudi

PREDICTING CREDIT CARD DEFAULTS

Objective:
Develop a predictive model to estimate the likelihood of a customer defaulting on their credit card payments

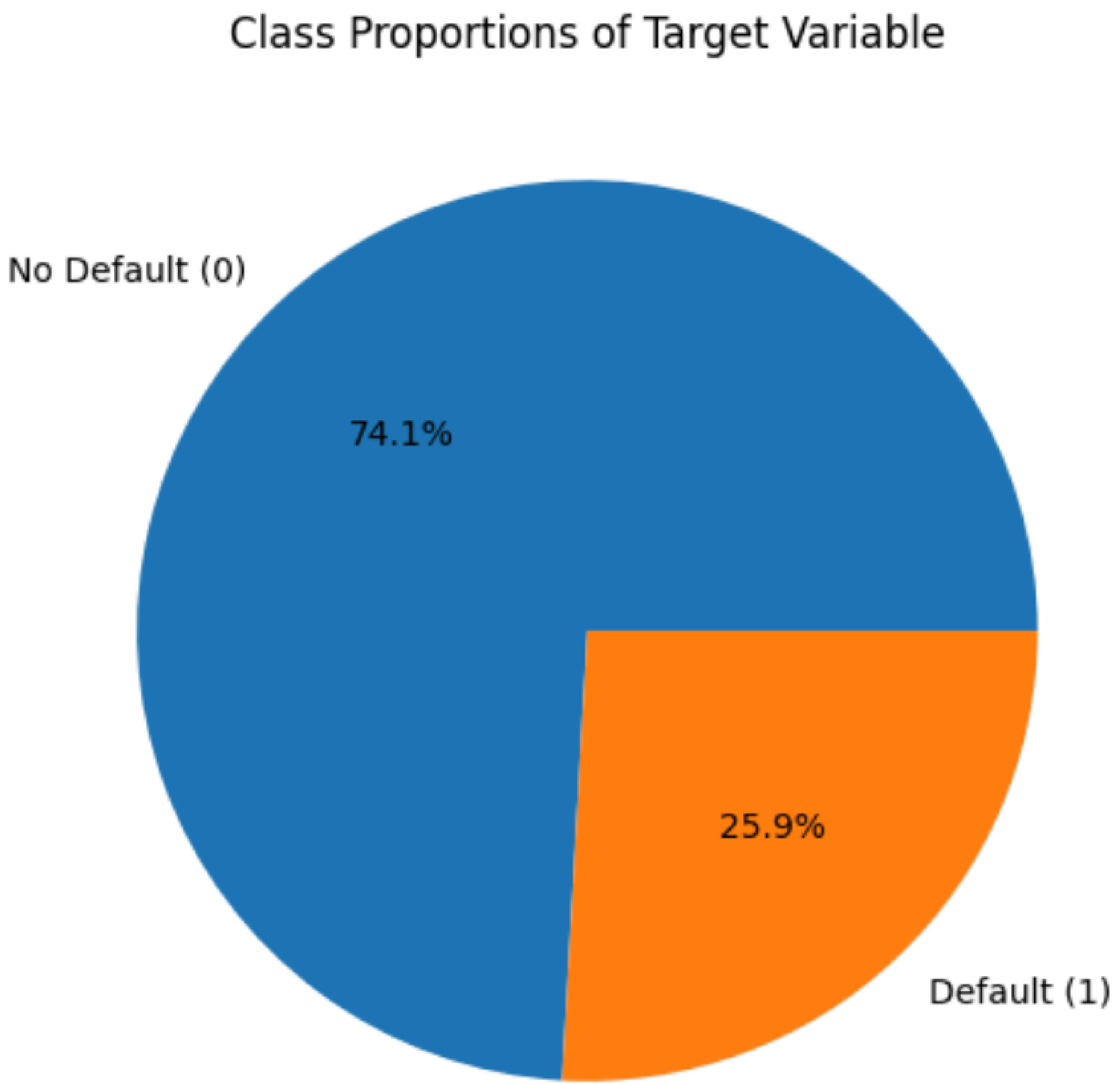
DATASET OVERVIEW FOR DEFAULT PREDICTION

- **Customer Count:** 458,913 unique customers.
- **Target Variable:**
  - 1 = Default
  - 0 = No Default
- **Analysis Period:** 18 months of customer activity.
- **Default Definition:** No payment within 120 days of the latest statement.



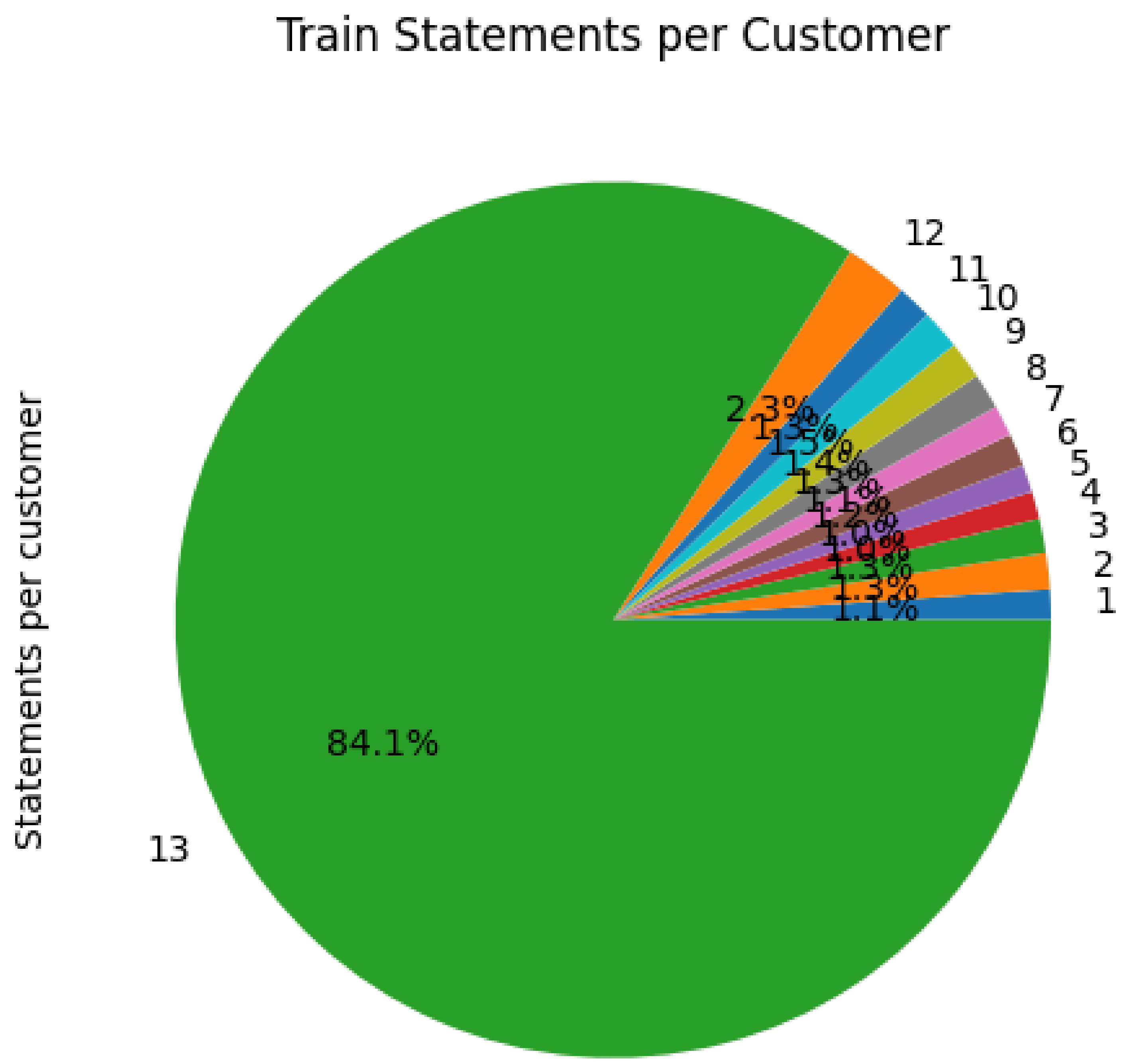
DATASET IMBALANCE

- Bias towards predicting non-defaults.
- **Potential Solutions:**
  - Focus on precision, **recall**, and **F1-score** to evaluate performance.
  - Use techniques like resampling, or class weights to address imbalance.



TIME SERIES INSIGHTS: CUSTOMER STATEMENTS

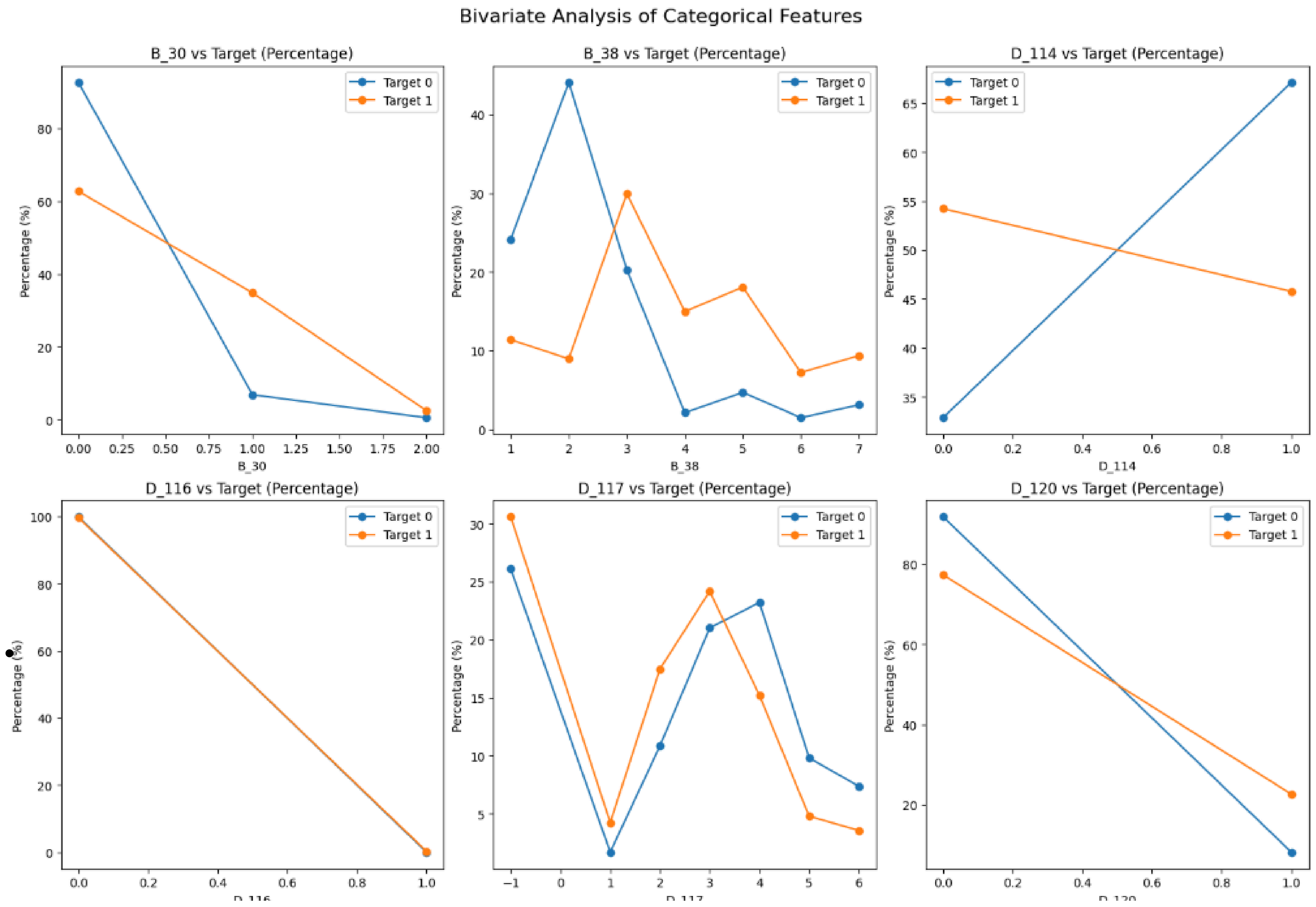
- **84.1% of customers have 13 statements.**
- Irregular data lengths can disrupt training for time-sensitive models like transformers.
- Homogenization using Padding.



EXPLORING CATEGORICAL PATTERNS IN DEFAULTS

Relationships between features and the target variable.
Examples:

- B_30, B_38, and D_114 clear separations, indicating strong predictive potential.
- D_117 presents overlapping shapes offers moderate discriminatory power.
- D_116 displays limited distinction between targets -> low predictive utility.

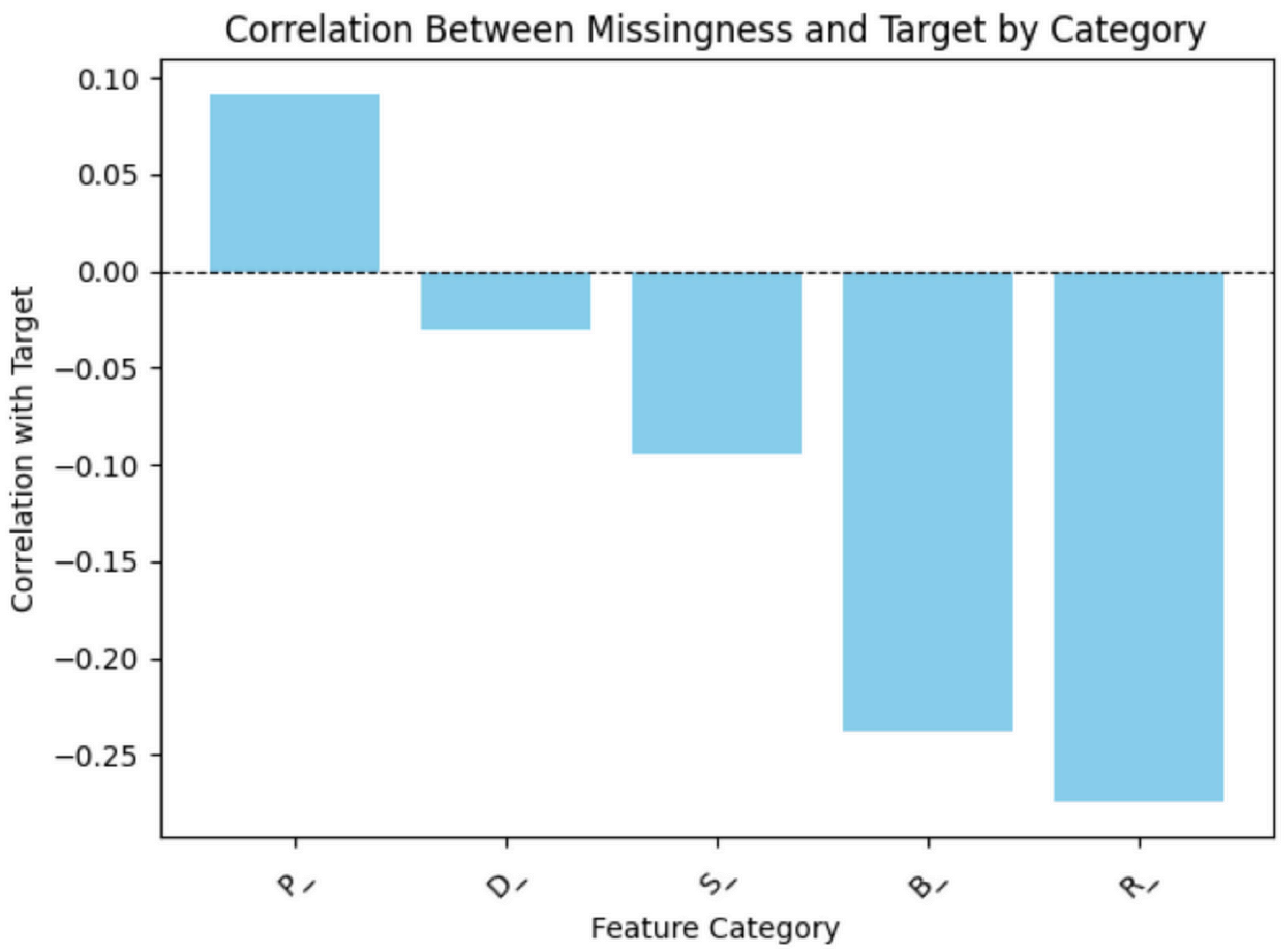


WHAT DOES MISSING DATA TELL US ABOUT DEFAULTS?

Missingness of data potential impact on prediction:

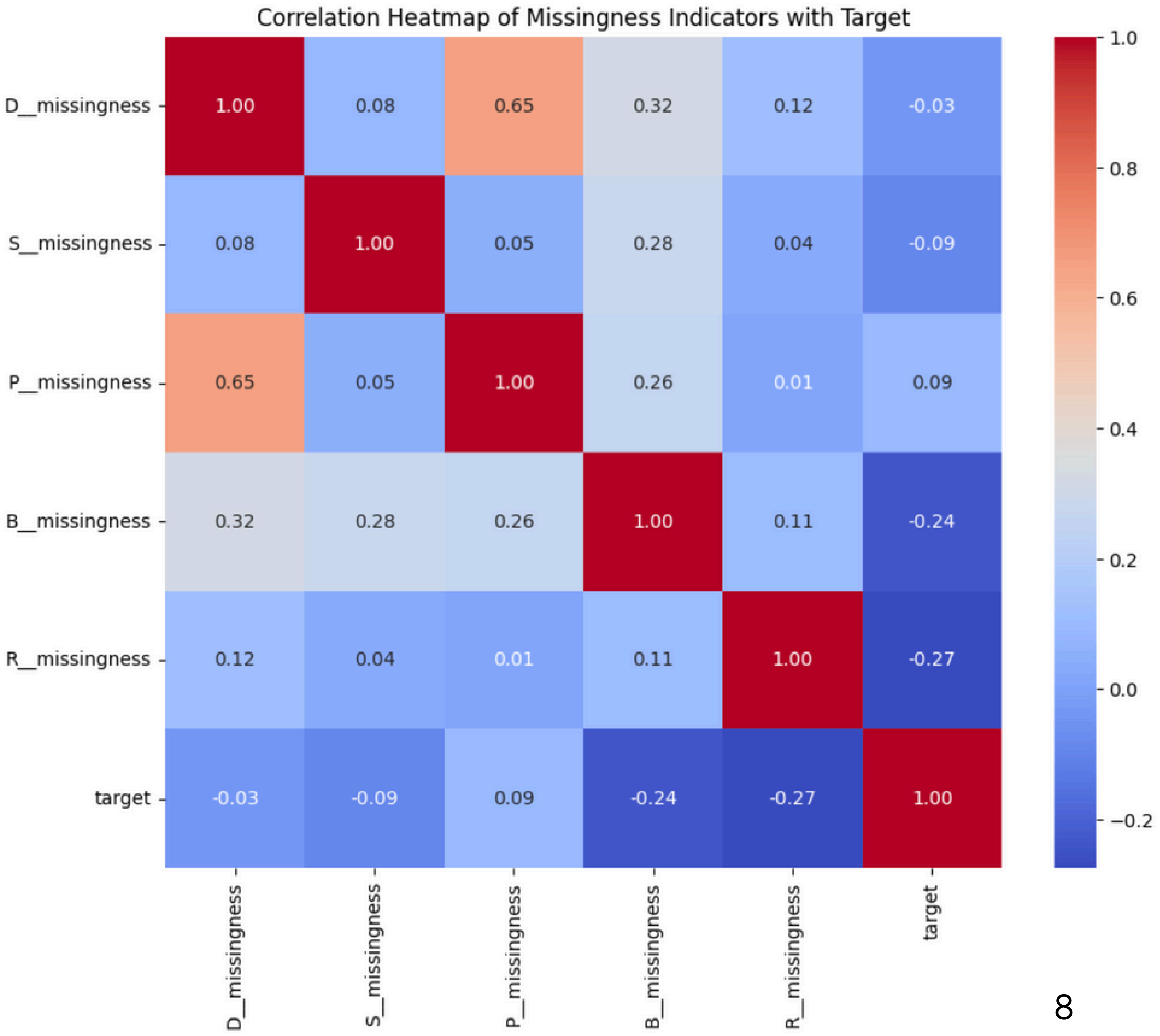
- **Delinquency (D_*):** Minimal impact on predictions.
- **Spend (S_*):** Slightly linked to non-defaults.
- **Payment (P_*):** Indicates a small risk of default.
- **Balance (B_*):** Moderately tied to non-defaults.
- **Risk (R_*):** Strongly signals non-defaults.

- **Leverage missingness to capture these patterns effectively.**

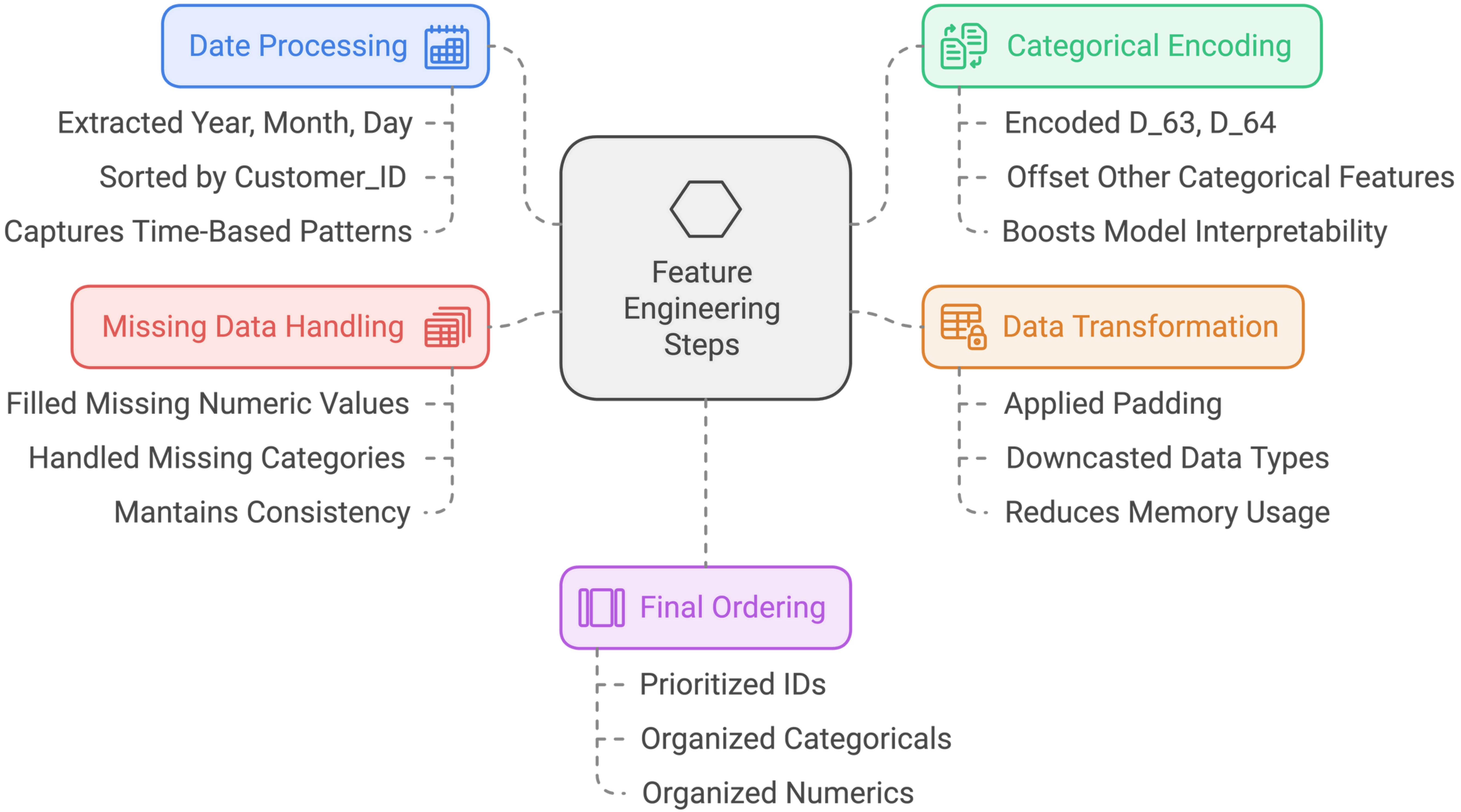


MAPPING CONNECTIONS IN MISSING DATA PATTERNS

- **Payment & Delinquency (0.65)** missed payments <-> delinquent behavior.
- **Balance & Delinquency (0.32)** could indicate shared account challenges.
- **Balance & Spend (0.28) and Balance & Payment (0.26)** could help understand patterns in spending, payments, and balances reveal common financial habits
- **Balance (-0.24) and Risk (-0.27)** data are strongly correlated with default risk.



Feature Engineering & Preprocessing



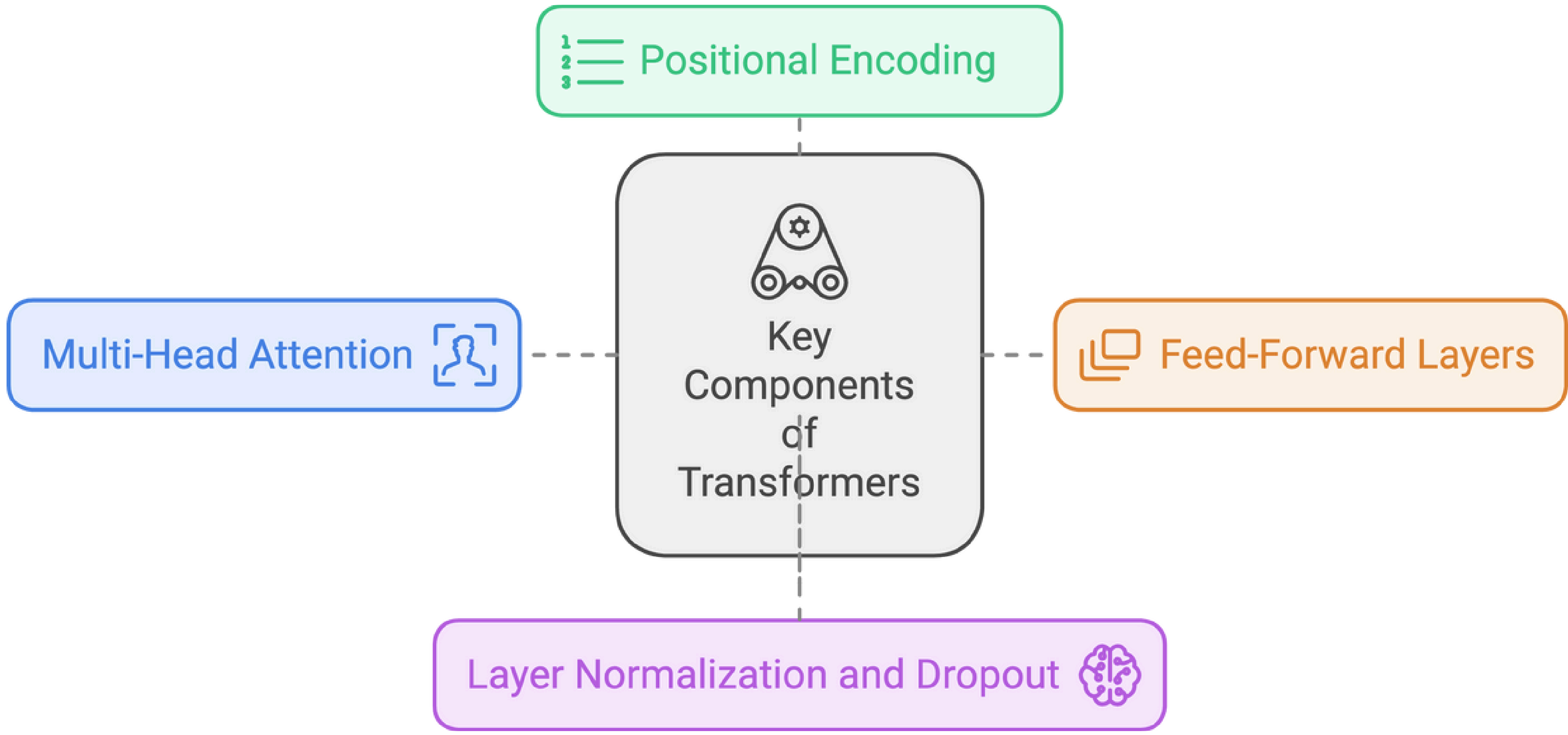
UNDERSTANDING TRANSFORMERS

What Are Transformers?

- A type of neural network designed to handle sequential data efficiently.
- They are particularly good at capturing long-range relationships in data.

Why Use Them?

- Ideal for modeling temporal dependencies in customer behavior.
- Excellent at capturing complex relationships in high-dimensional data.



MODEL DESIGN

Transformer Architecture:

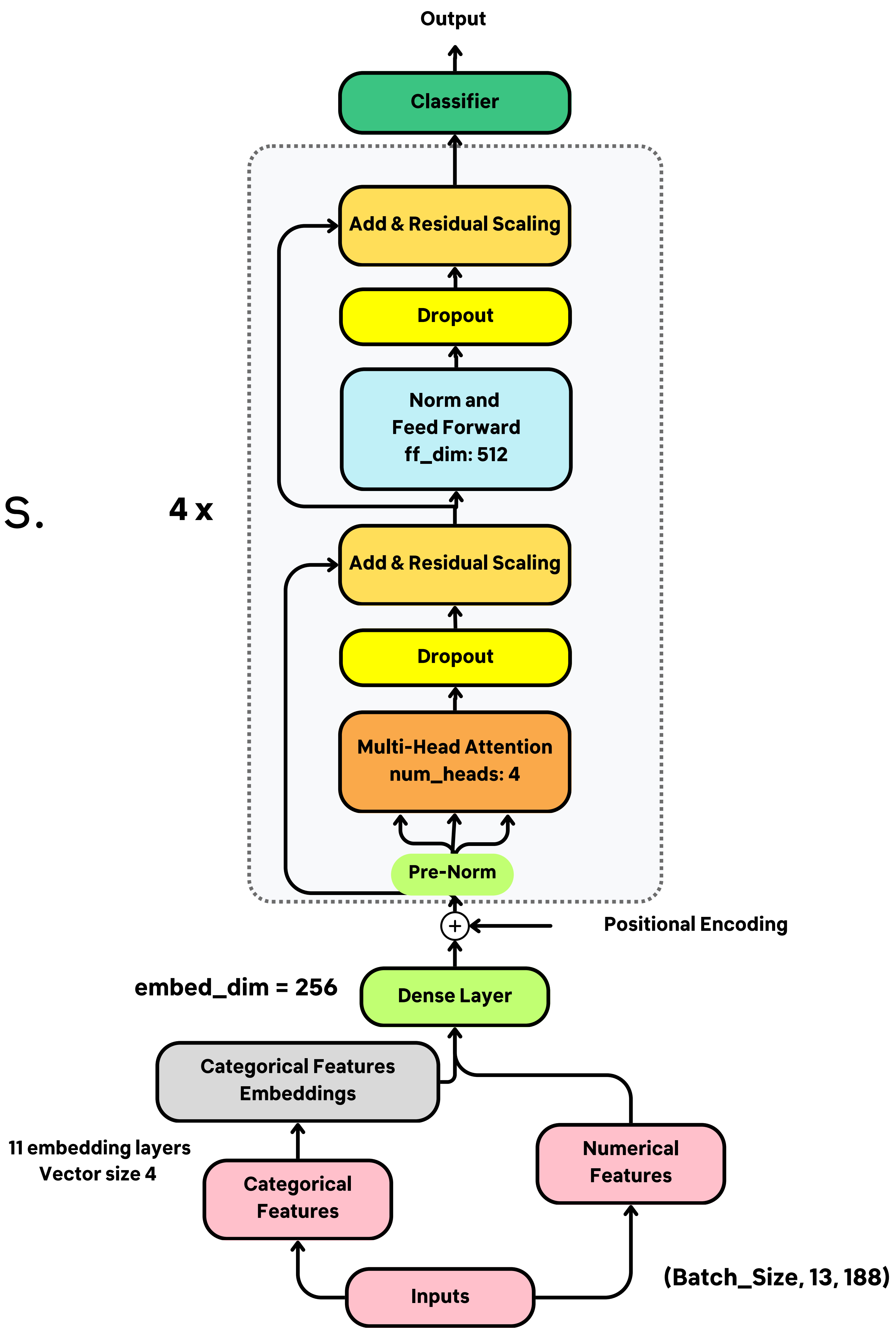
- **Transformer Blocks:** Multi-head attention with feed-forward networks for capturing sequential dependencies.
- **Positional Encoding:** Adds temporal context for time-series data.

Embedding Layers:

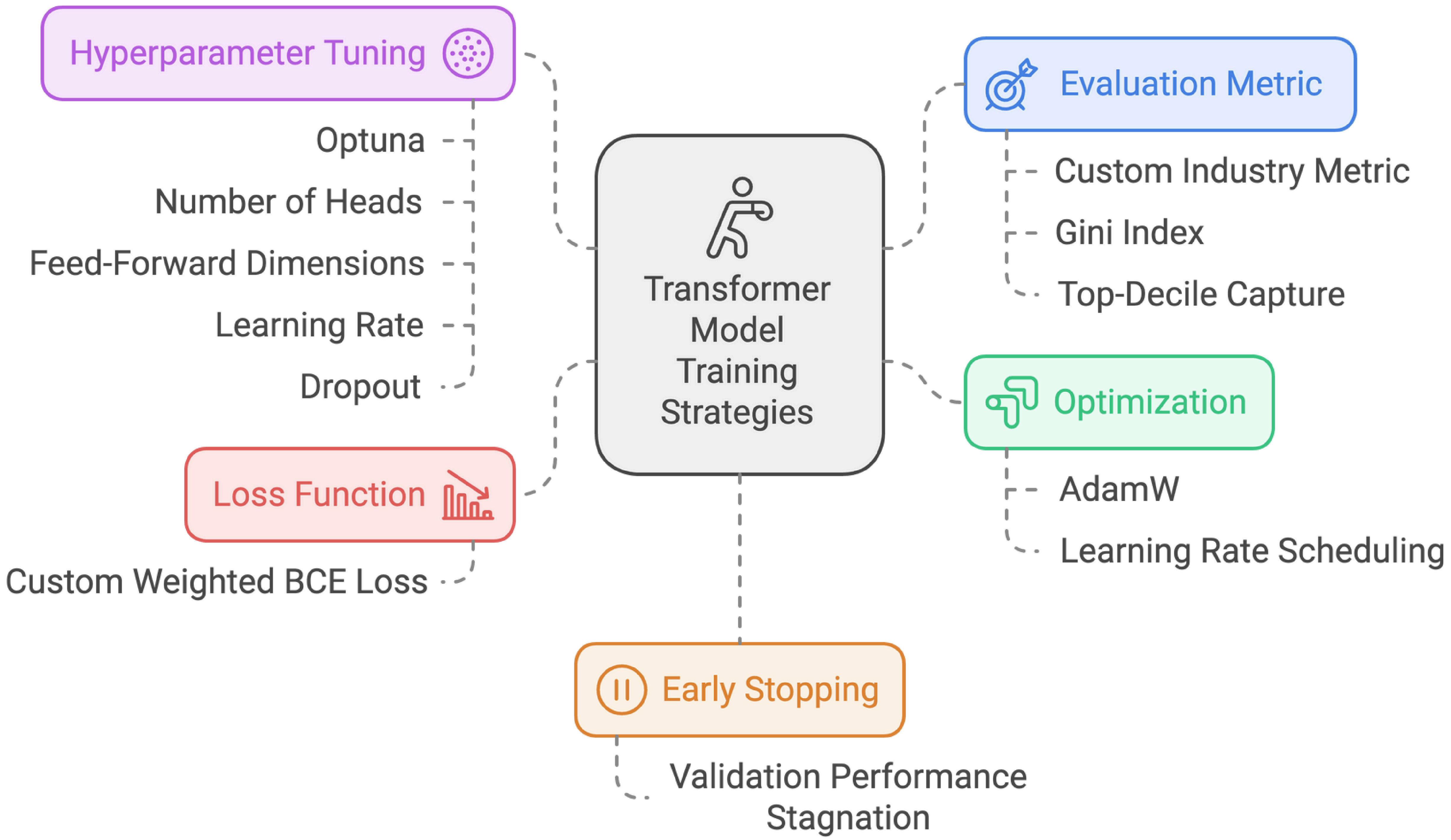
- To convert categorical features into numerical representations.

Classifier Network:

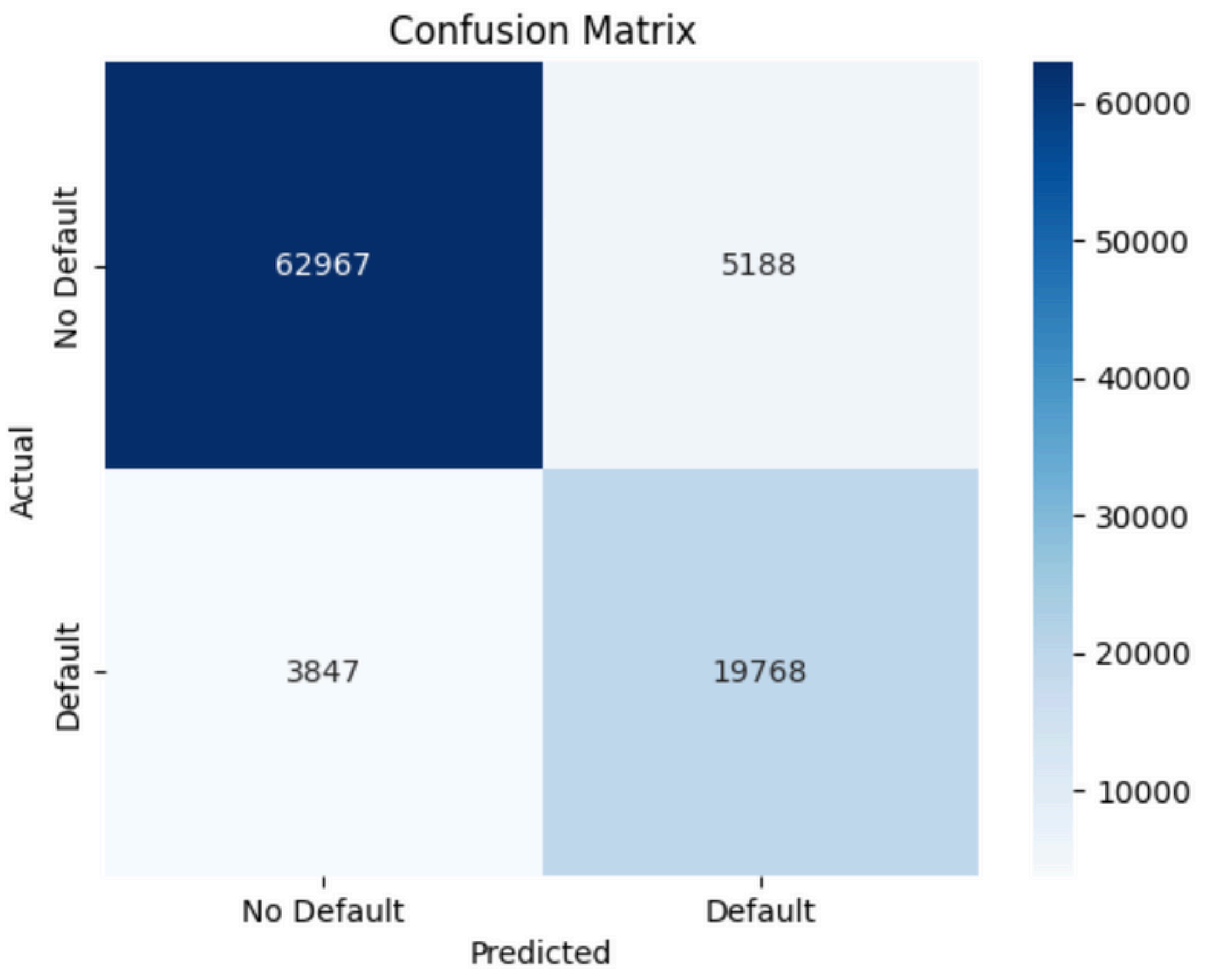
- Fully connected layers with ReLU activations and dropout for regularization.
- Output: A **sigmoid** layer predicting default probabilities.



TRAINING AND TUNING THE TRANSFORMER MODEL



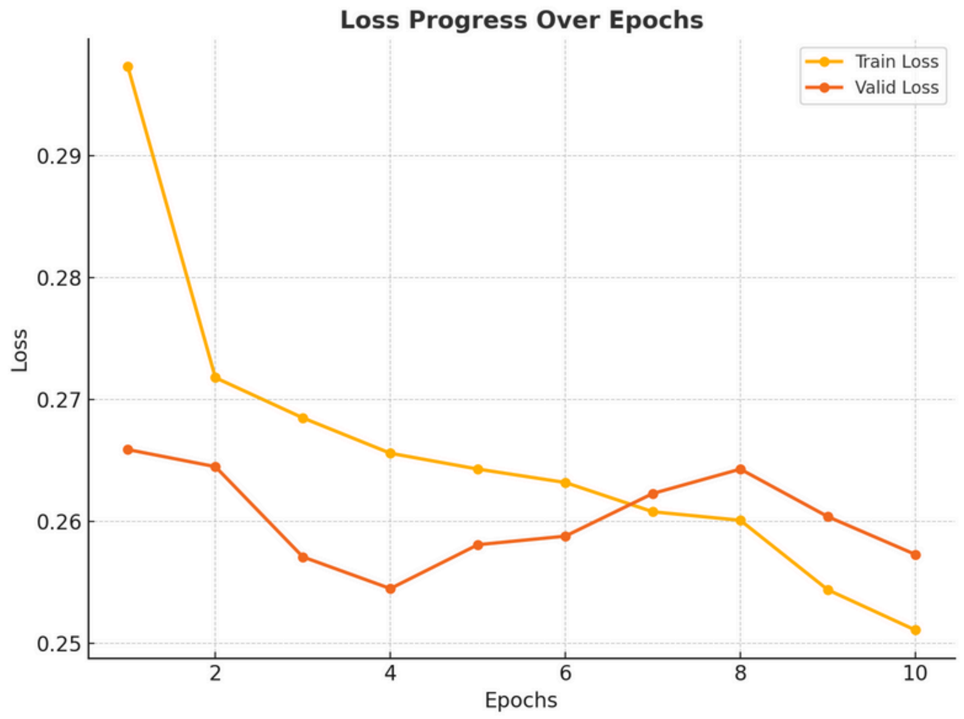
RESULTS

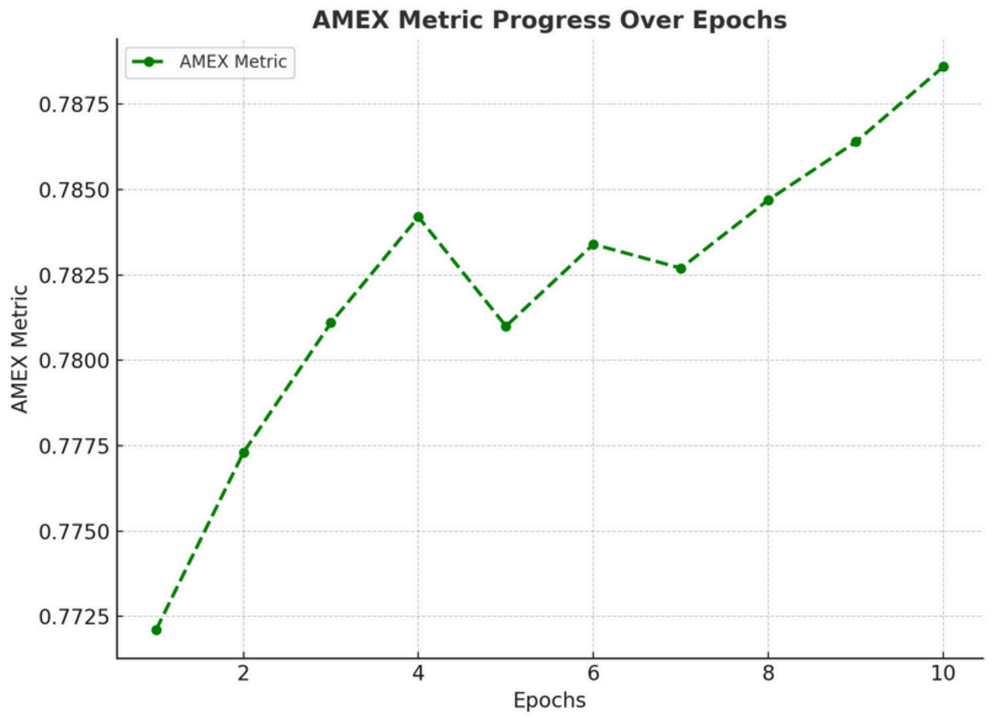


Performance Metrics

- **AMEX Metric: 0.7886**
- **Precision: 79.21%**
- **Recall: 83.71%**
- **F1-Score: 81.40%**

Possibility to improve recall and capture more missed defaulters.





CONCLUSION

We Built a **Transformer**-based model for credit card default prediction.

We achieved **0.7846** in AMEX Metric.

We captured data patterns, handled class **imbalance**, and optimized performance.

We plan to enhance recall by either by integrating data augmentation or trying ensemble methods.