

NeuronAI

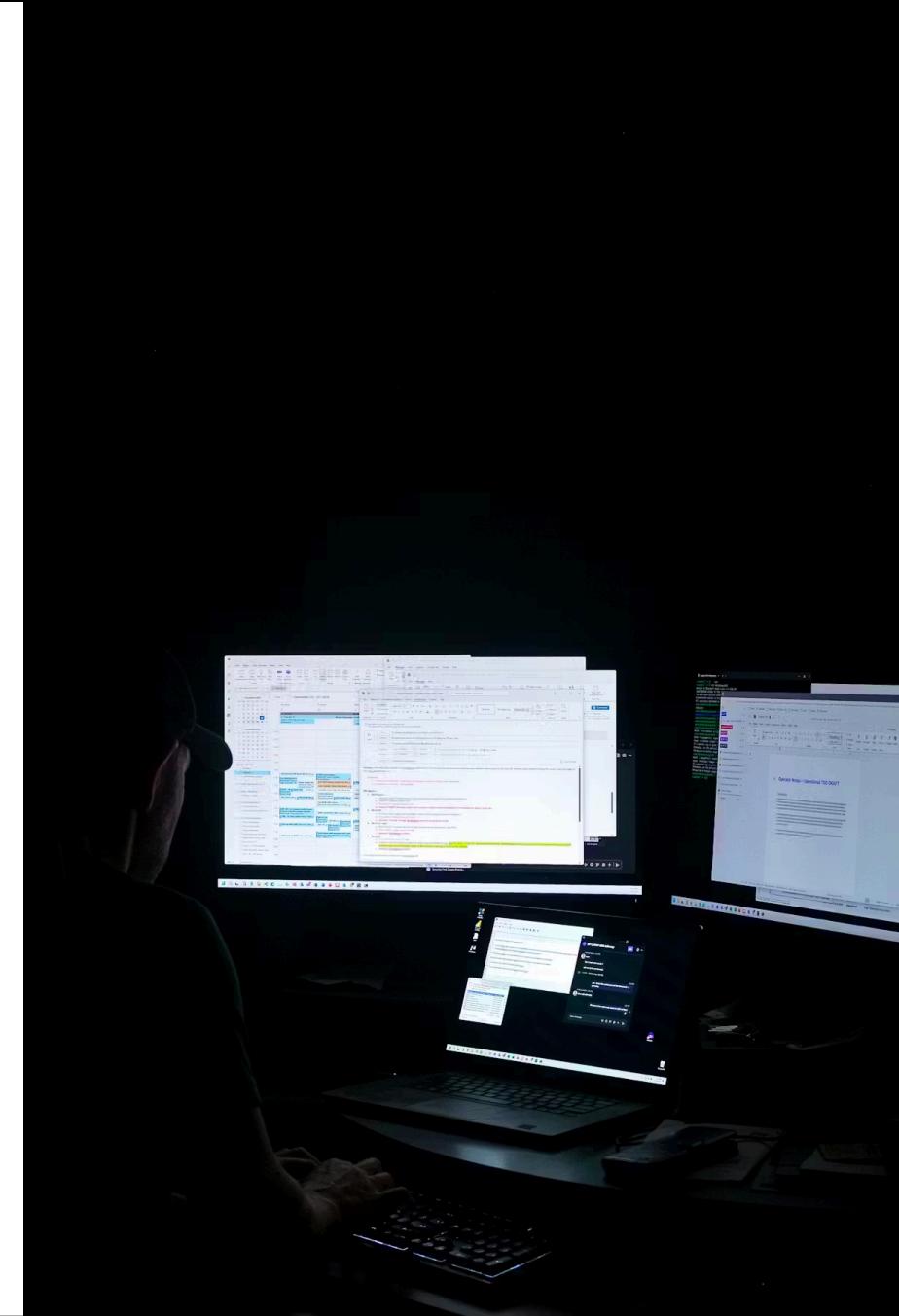
# Credit Default Prediction System



# Dataset & Merging

## Data Preprocessing Pipeline

- Merged application, credit history, demographics, financial ratios, geographic and load details data by unique identifiers.
- Removed unnecessary signs and noisy random columns
- Normalized values: formats
- Removed duplicates and handled missing values
- Filtered the ID columns



# Dataset Overview

## Key Features

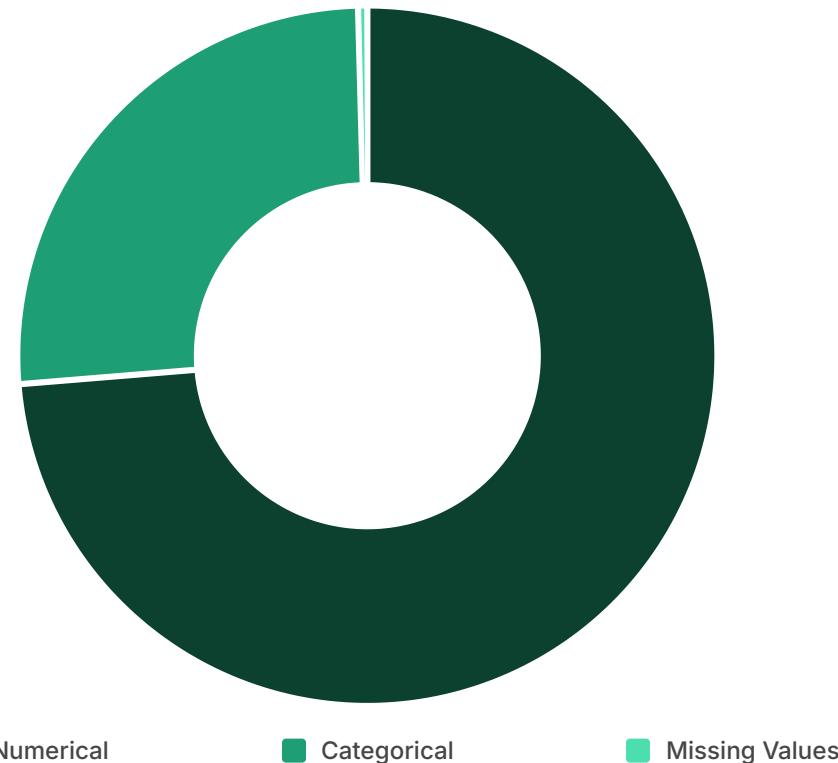
Application hour, income levels, age demographics, account metrics, login behaviour, and customer service interactions.

## Data Structure

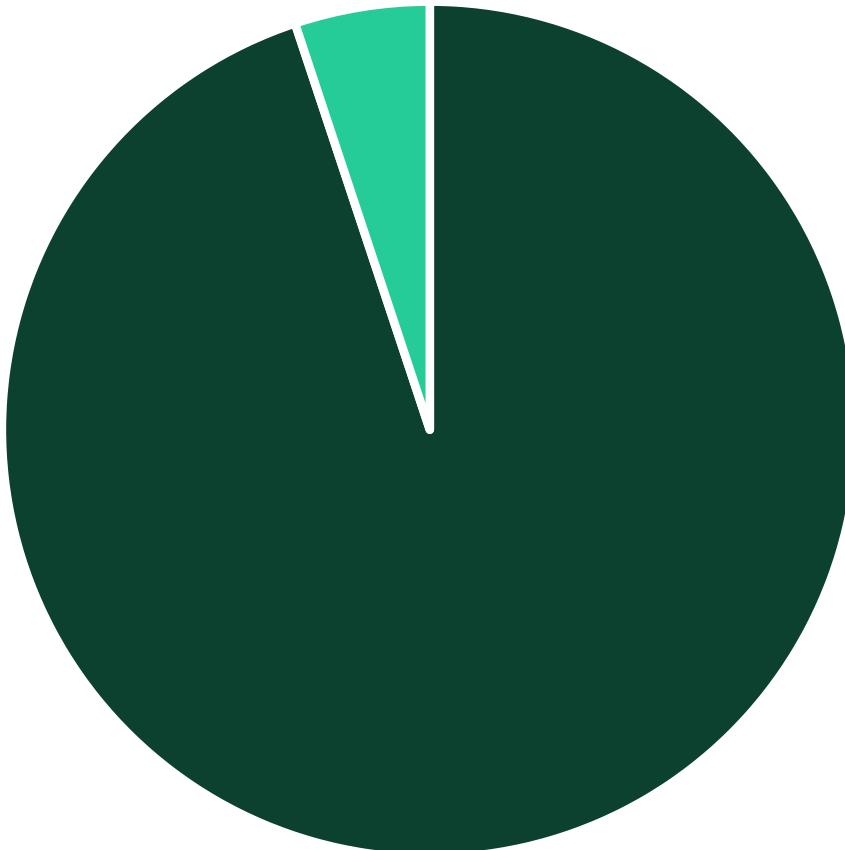
Comprehensive dataset with categorical and numerical variables capturing customer financial behaviour and demographics.

## Status Codes

Binary target variable indicating default vs non-default outcomes for credit risk assessment.



# Feature Distributions



■ Default: 0 ■ Default: 1

# Analytics & Insights

## Feature Importance Analysis



### Behavioural Metrics Lead

Customer service calls and login patterns show strongest correlation with default risk, outperforming traditional demographic factors.



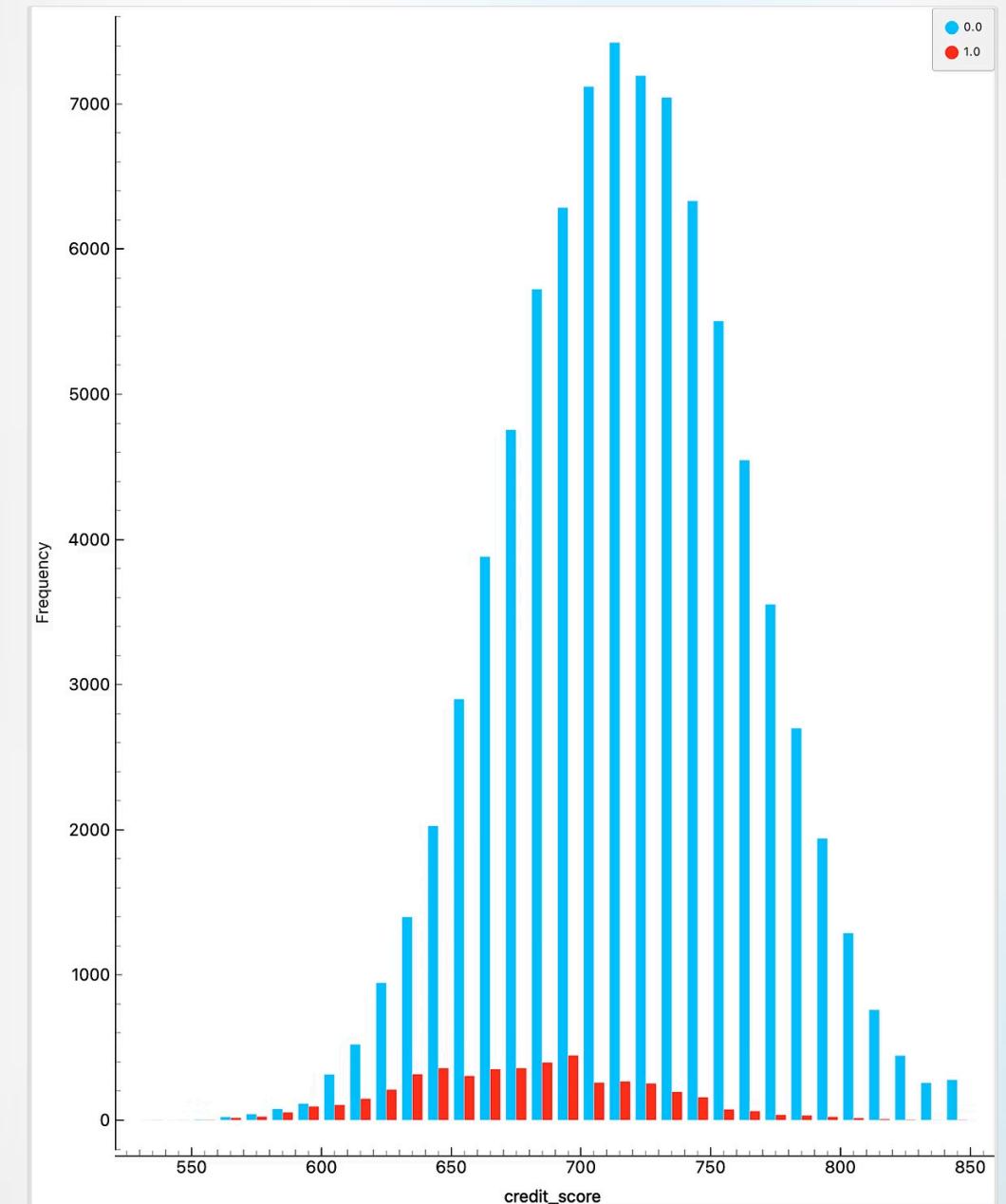
### Income Significance

Annual income demonstrates moderate predictive power when combined with account age and transaction behaviour.



### Account Activity

Account age and transaction frequency provide crucial context for assessing creditworthiness and default probability.





# Risk Pattern Discovery

## Low Account Age Risk

Customers with accounts younger than 6 months show 35% higher default rates compared to established accounts.

## Service Call Indicator

More than 8 customer service calls monthly correlates with 42% increased default probability.

## Login Irregularity

Irregular login patterns indicate moderate risk, suggesting financial stress or disengagement.

# Customer Segmentation

## Key Patterns

- High-risk cluster: young accounts, frequent service calls
- Stable cluster: established accounts, regular activity
- Moderate cluster: mixed behavioural signals

Customer behaviour is the strongest predictor of credit default risk.

# AI Model Architecture



1

## Feature Engineering

Created interaction features and temporal patterns from raw data.



2

## Encoding

Applied target encoding for categorical variables.



3

## Balancing

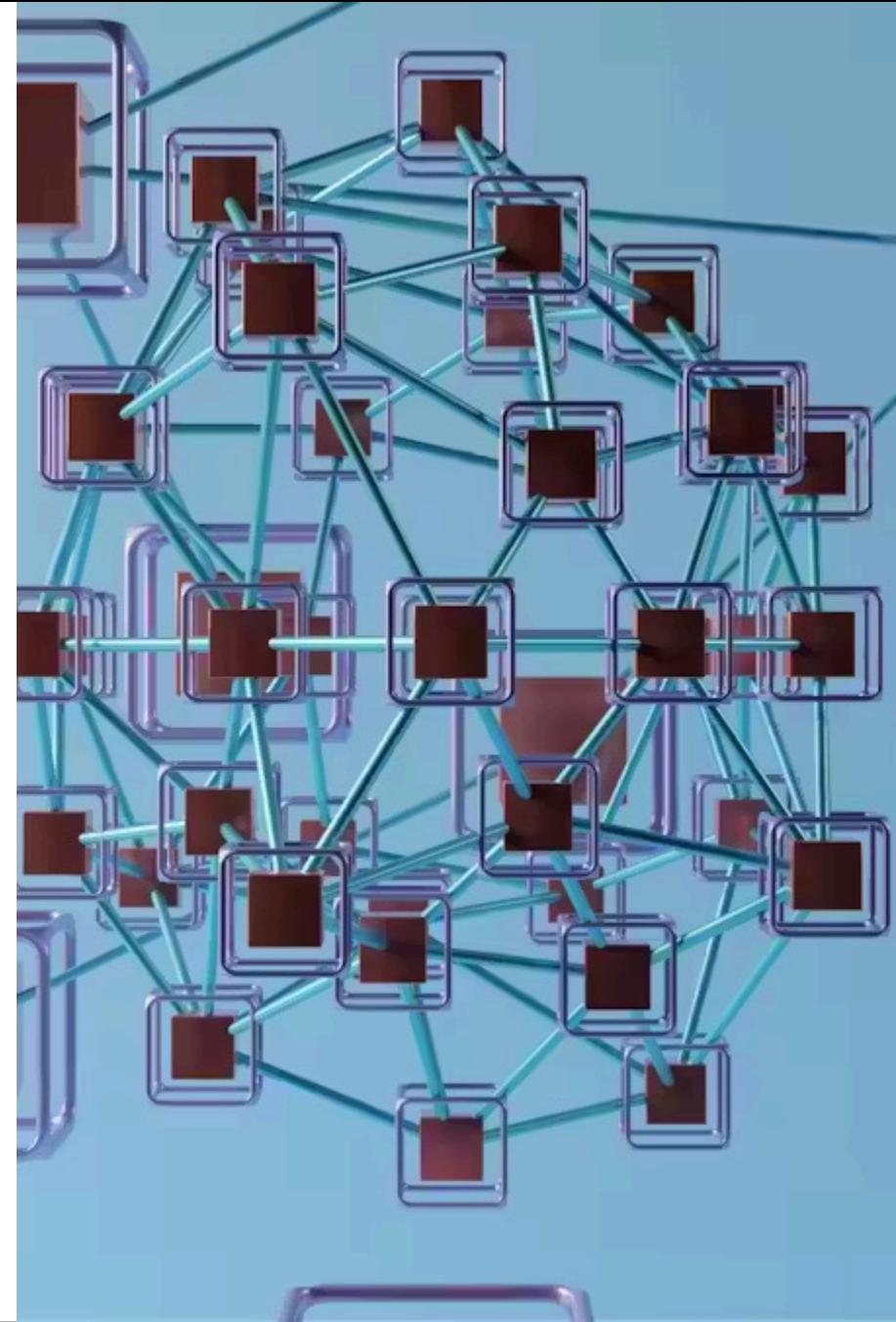
Used SMOTE to address class imbalance.



4

## Model Selection

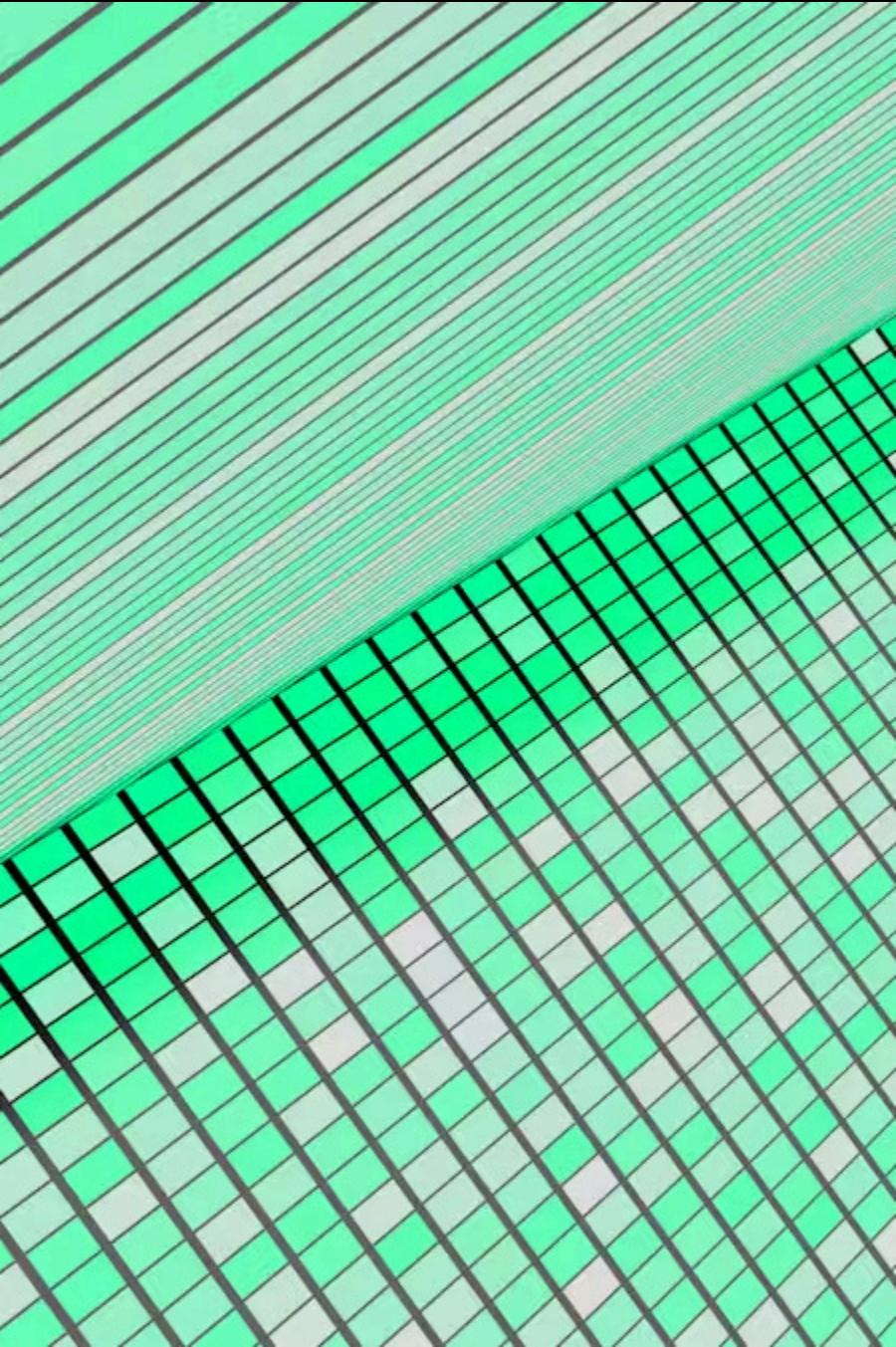
Evaluated CatBoost, XGBoost, and LightGBM.



# Model Performance Comparison

A detailed comparison of various machine learning models used in our analysis, highlighting their performance across key metrics.

Method/Model	Precision	Recall	F1 Score	AUC	GINI
CatBoost	0.824	0.85	0.639	0.805	0.84
XGBoost	0.332	0.149	0.206	0.749	0.498
LightGBM	0.152	0.669	0.248	0.804	0.609
Random Forest	0.180	0.325	0.231	0.741	0.482



# Model Performance

## Metrics Achieved

**0.84**

AUC Score

Excellent  
discrimination  
between default and  
non-default cases.

**86%**

Accuracy

Overall prediction  
accuracy across all  
cases.

**82%**

Recall

Successfully  
identifies 82% of  
actual default cases.

# Our Team



**Mahmudov Sunnat**

AI/CV/ML Engineer

**~10 years experience**

NeuronAI LLC



**Ergashev Adizbek**

Full-Stack Developer

**PHD at TUIT**

**~10 years experience**

NeuronAI LLC



**G'ayratov Muhriddin**

Project Manager

**MSc at Turin(TTPU)**

**~5 years experience**

NeuronAI LLC



**Mamaroziqov Kamron**

AI Engineer +3 years experience

**BSc at New Uzbekistan university**

NeuronAI LLC