

MBAS905 FINAL PROJECT

Loan repayment failure prediction using app behavior data

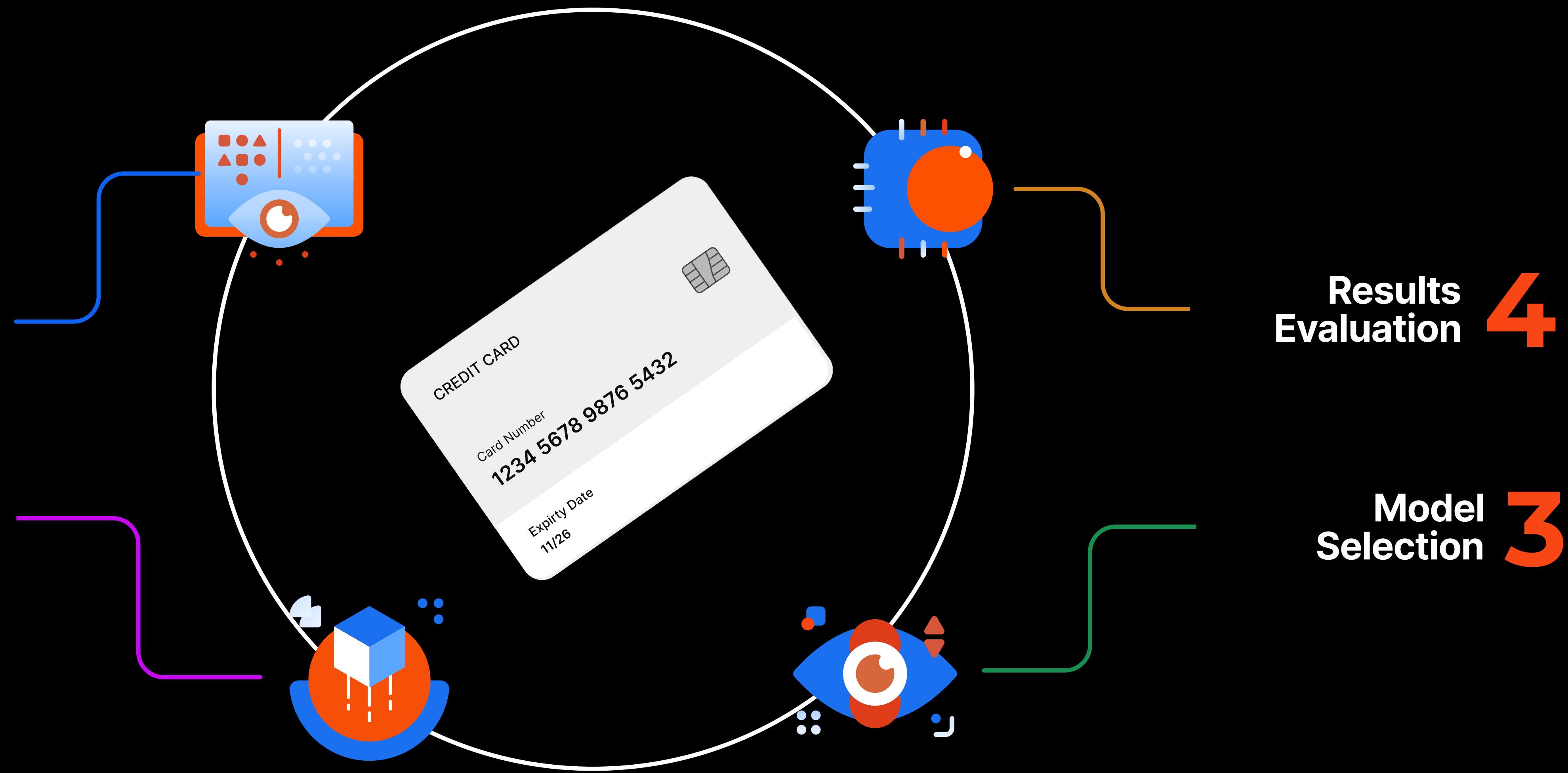
Agenda

1 Business Understanding

2 Data Understanding

Results Evaluation 4

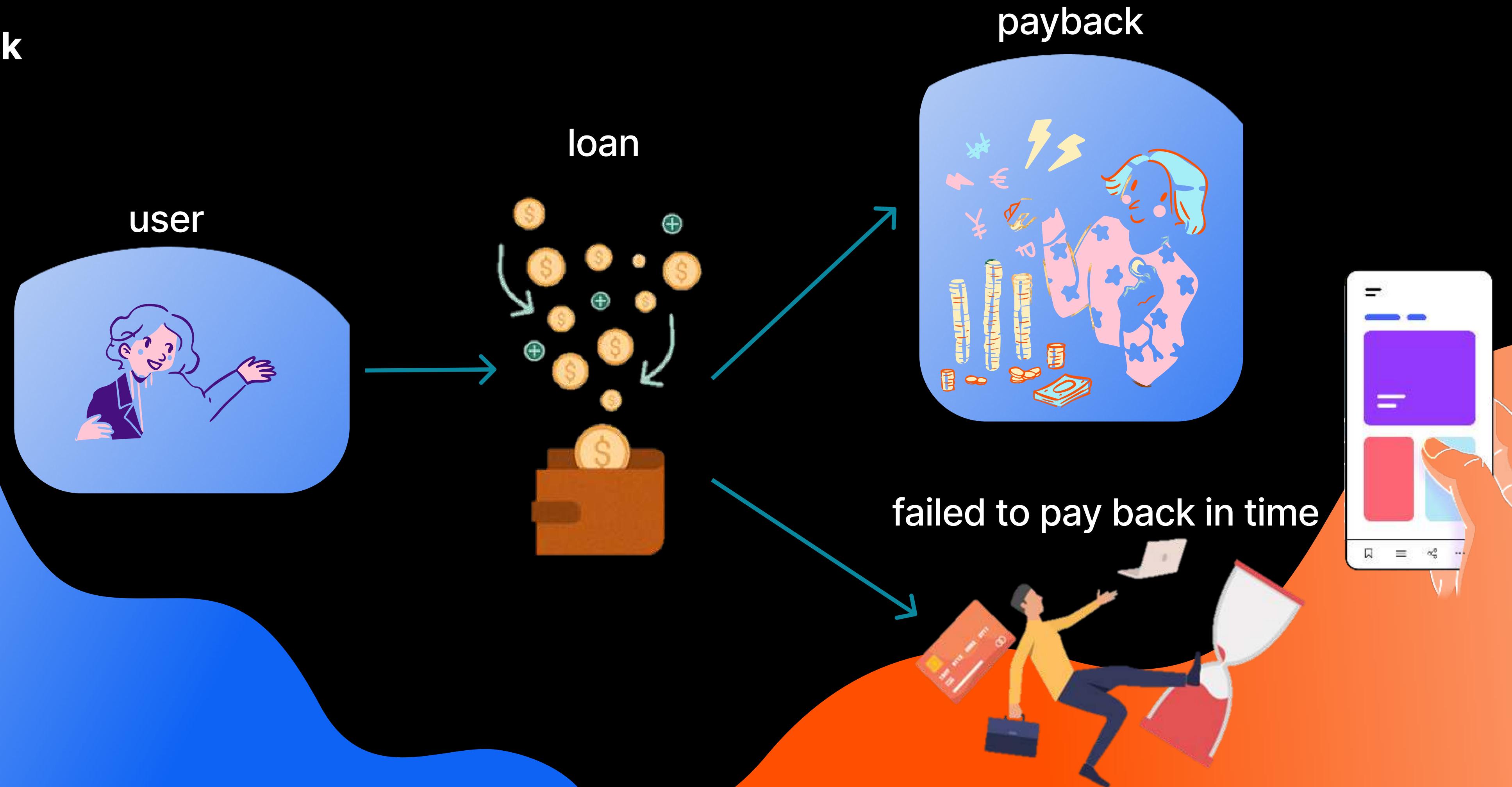
Model Selection 3



Business Understanding

Which behavioral/app interaction variables are the most predictable of **repayment overdue**?

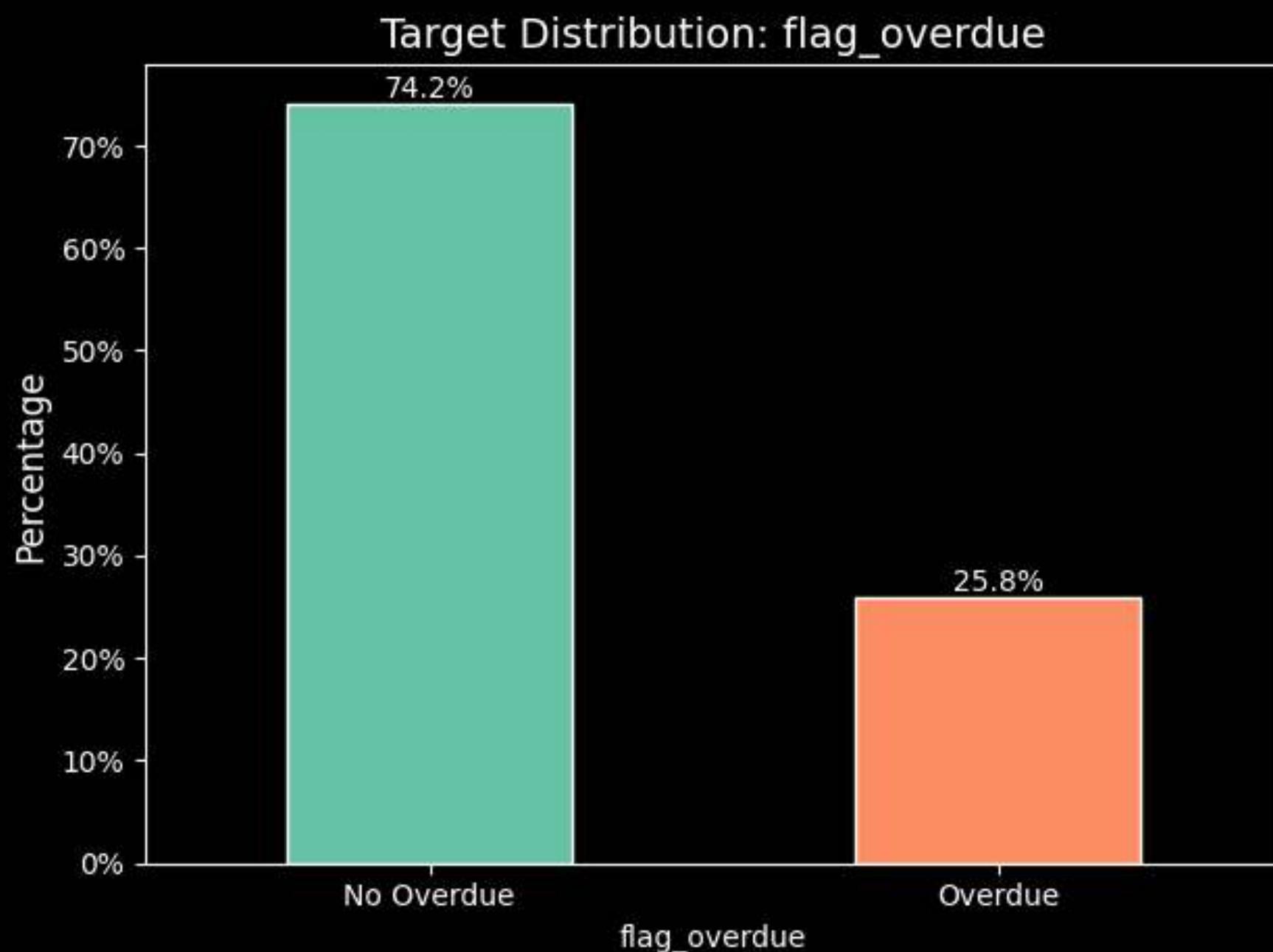
Credit risk



Data Understanding

Variable Selection and Exploration

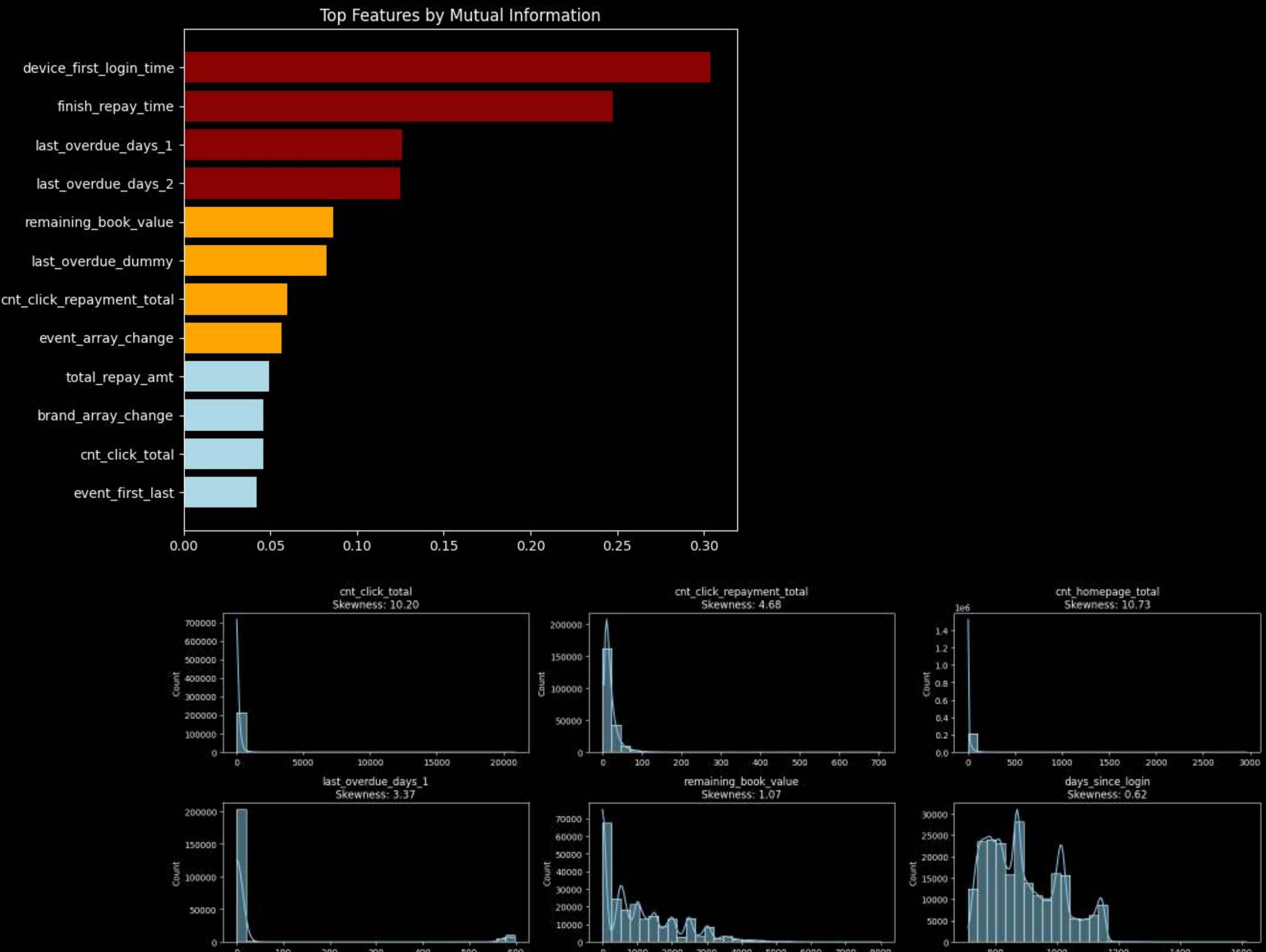
Target variable is imbalanced ⇒
class-weight adjustment



Selecting features out of 84:

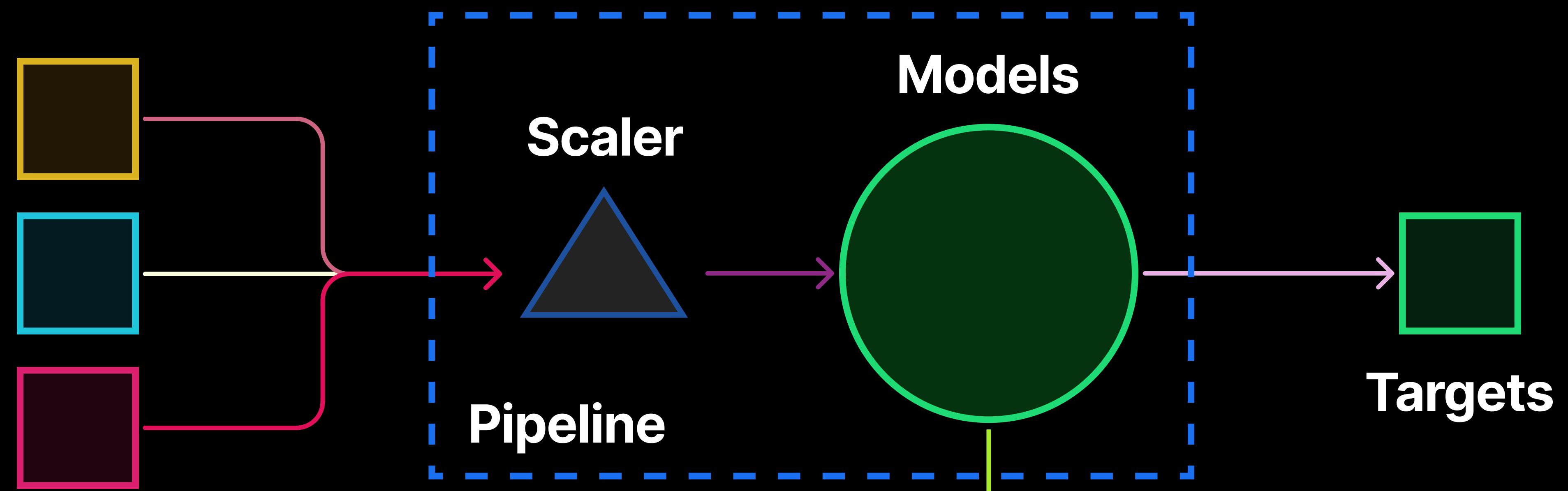
- 1) MI score for features
 - 2) Check skewness
 - 3) Apply log-transformations

Selected features = 6 (transformed into 8)



Model Selection

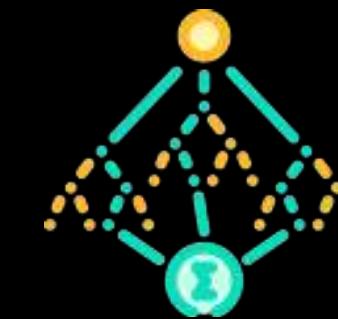
Training Set



Logistic Regression



Random Forest



XGBoost



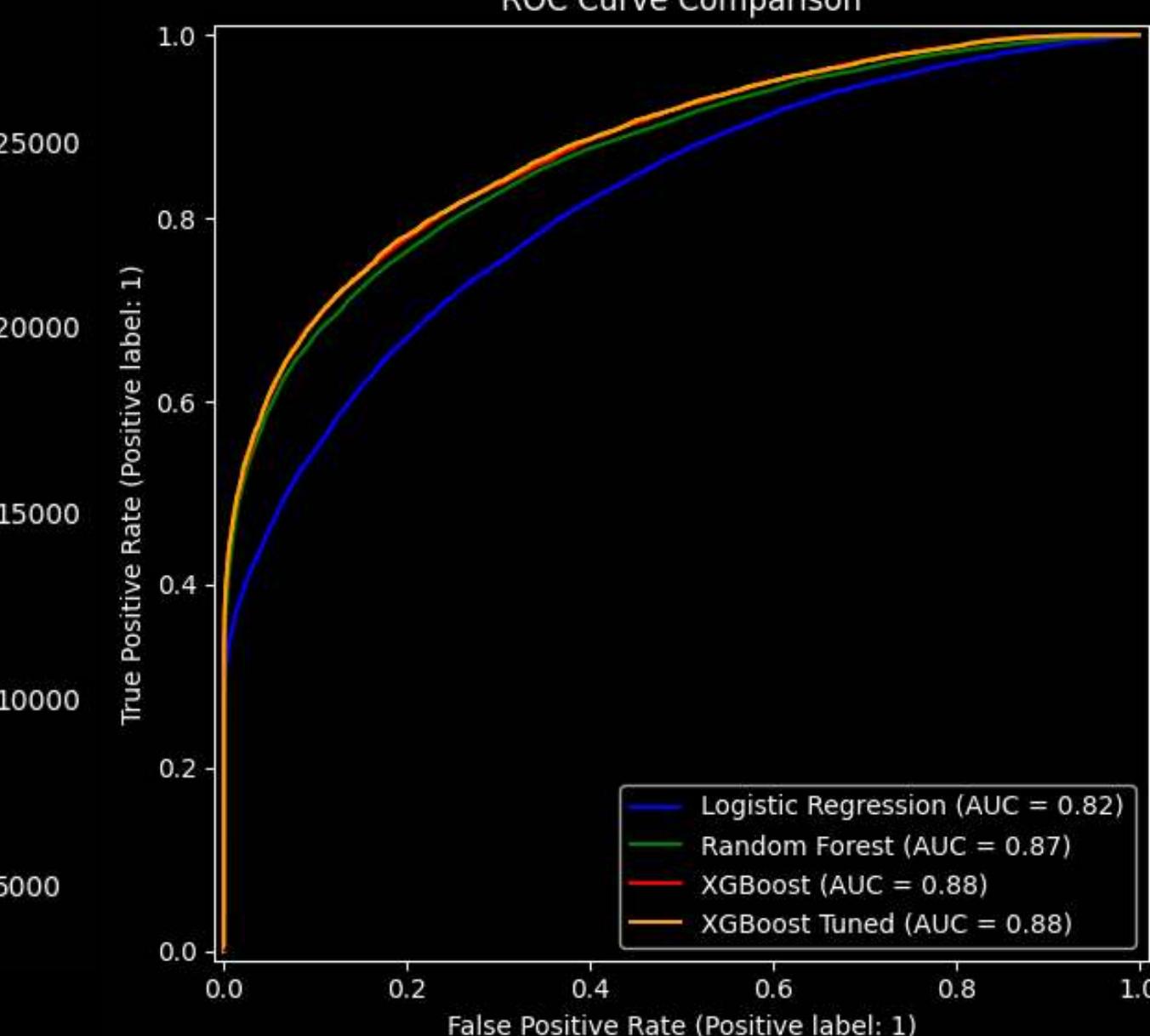
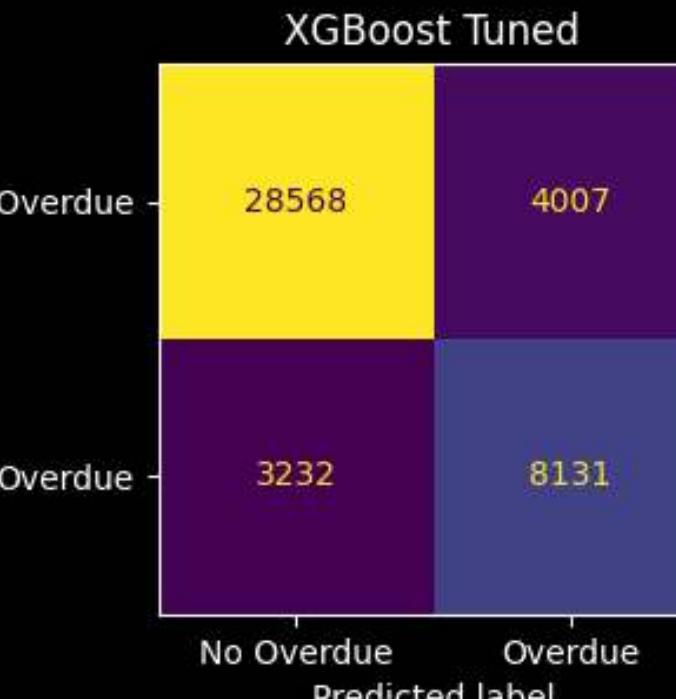
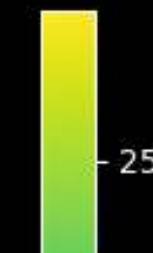
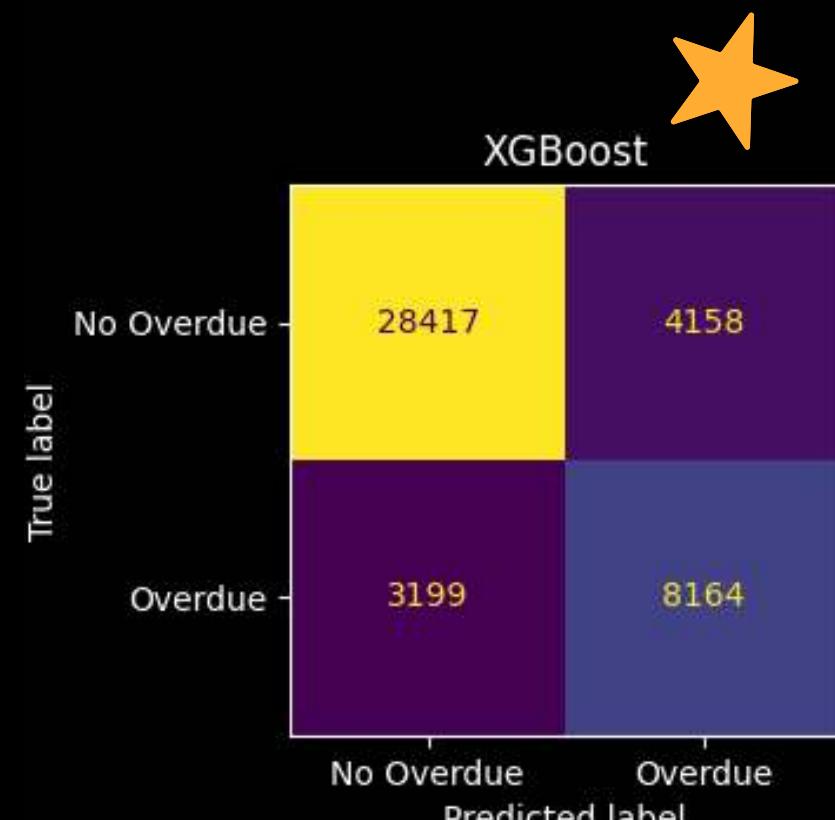
+ GridSearchCV

Best Accuracy, F1, ROC AUC

→ Next Slides

Results Evaluation

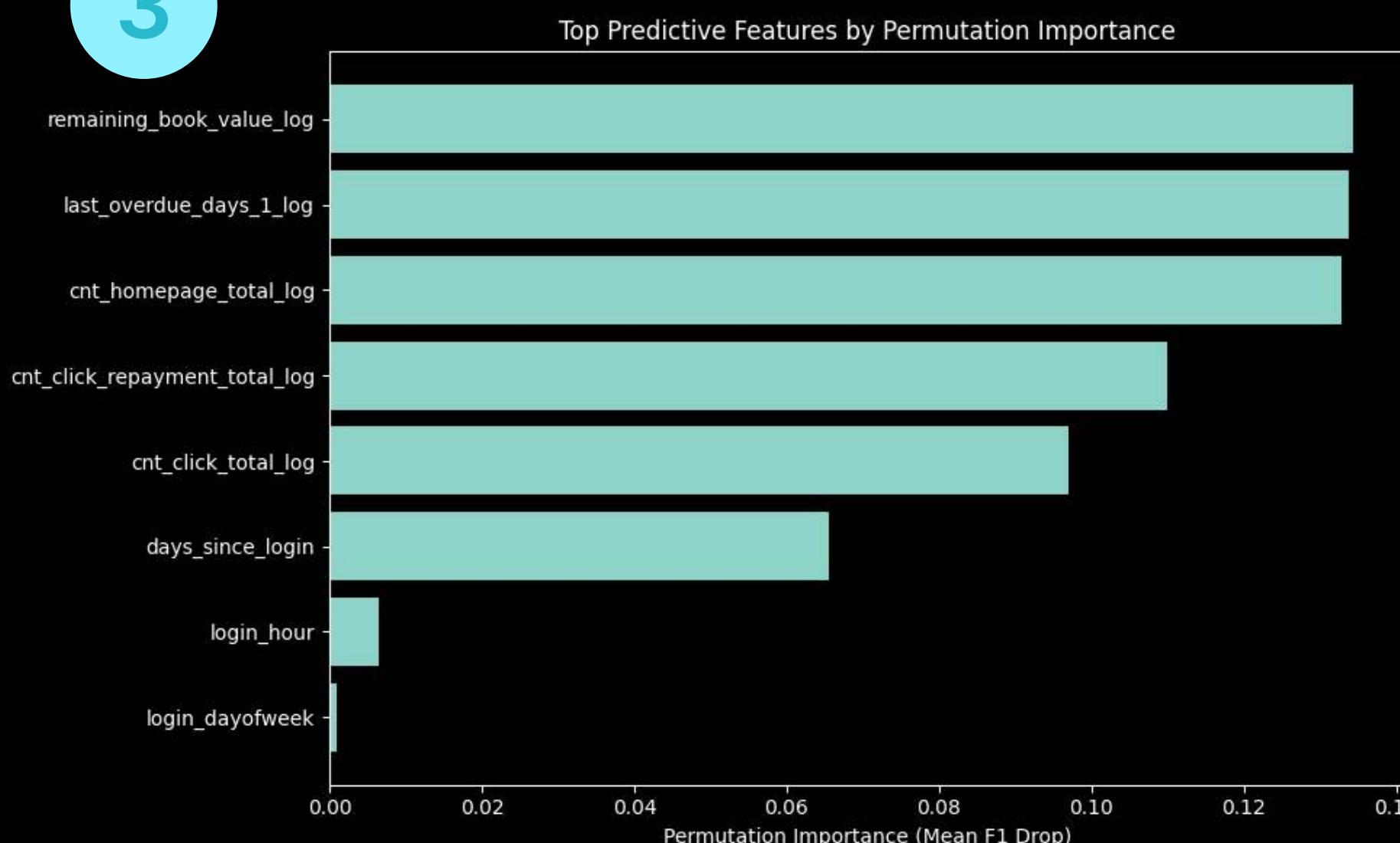
2
XGBoost



1
XGBoost_tuned

Model	Accuracy	F1 Score	ROC AUC
Logistic Regression	0.781123	0.600374	0.816831
Random Forest	0.858869	0.667881	0.867635
XGBoost	0.832560	0.689381	0.877089
XGBoost_tuned	0.835245	0.691971	0.878048

3



The **outstanding value**, **past overdue behavior** and **in-app behavior** (such as homepage and repayment page visits) are strong indicators of credit risk. Also **total click activity** and **recency** of app usage are meaningful predictors

The background features a dynamic, abstract design. On the left, a vibrant blue color serves as a backdrop for a series of fluid, white-greyish swooshes that resemble liquid or light trails. These swooshes curve elegantly across the frame. The right side of the image is dominated by a deep, solid black color, creating a strong contrast with the blue and the white swooshes.

**THANK YOU FOR
LISTENING!**

Q&A SESSION