BPI CHALLENGE 2017

A Goal-Oriented Approach to Loan Prioritization using Process Mining

Process Oriented Data Science
Academic Year 2025/2026

Ilaria Boschetto, Federico Clerici, Raffaele D'Agostino, Gabriele Villa, Davide Volpi

Loan Applications Log

The log contains 31,509 loan applications and comprises approximately 1.2 million events executed by 149 different resources. A distinctive feature of this log is its multi-object structure.

APPLICATION STATES

A_events

- They represent the lifecycle of the case from submission to the final decision.
- A_Submitted, A_Concept, A_Accepted, A_Validating, A_Pending

OFFER STATES

O_events

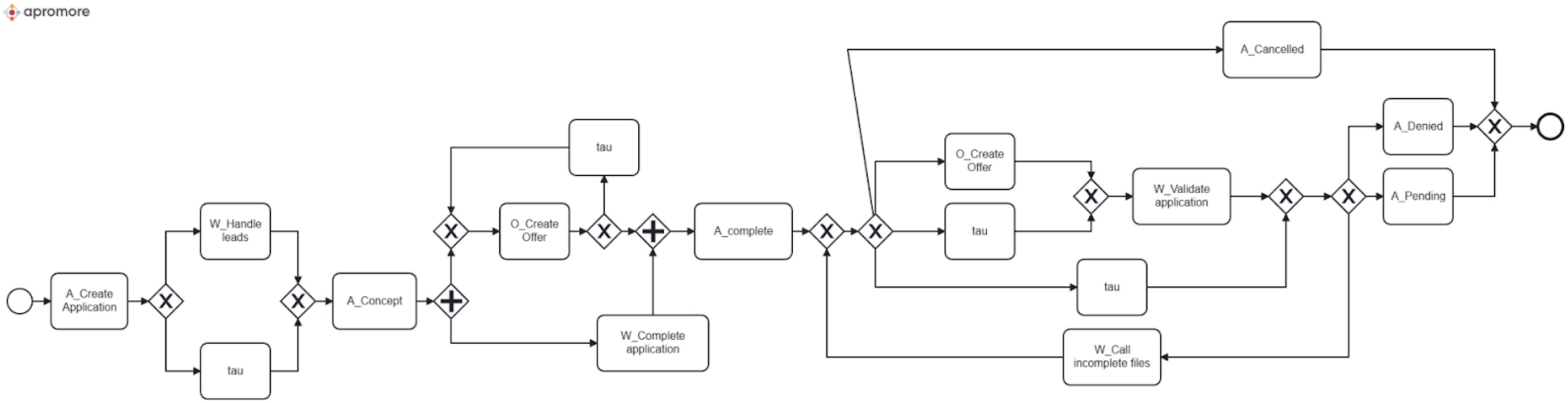
- These events track the lifecycle of the specific credit offers created for the customer.
- O_Create Offer, O_Sent, O_Returned, O_Accepted

WORKFLOW ACTIVITIES

W_events

- These represent the actual work items performed by bank resources.
- W_Handle leads, W_Validate application, W_Call_incomplete file

Baseline BPMN model

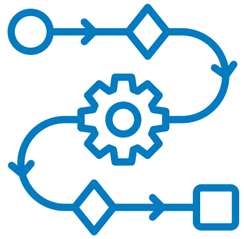


Academic paper model
Clustered activities
Complete cases

→

simple but well fitting and precise baseline model

Business question



How to reduce workload and speed up the process?

A way of reducing the workload on resources and speed up this process is by prioritizing applications that are likely to be successful. Moreover we performed a detailed analysis in order to identify activities which is possible to automate making the process lighter.

How can we identify and prioritize applications with the highest probability of success?

We implemented a ML model specifically designed to predict the likelihood of an offer being selected by the client.
By analyzing early process features, the model assigns a success probability score to each running application and sends applications in two different branches of the process.

Prioritization and Automation

ML model

- Random Forest
- Process-derived features + application attributes

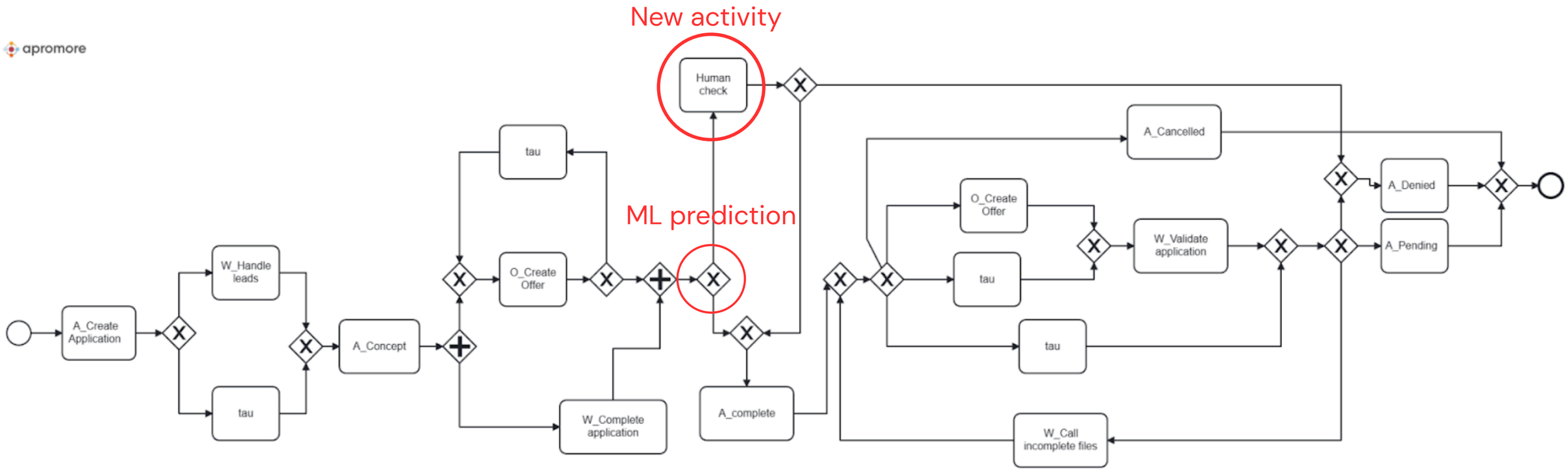
Framework Dimension Analysis

- Frequency
- Variability
- Resource
- Domain knowledge

Likelihood that an user accepts an offer

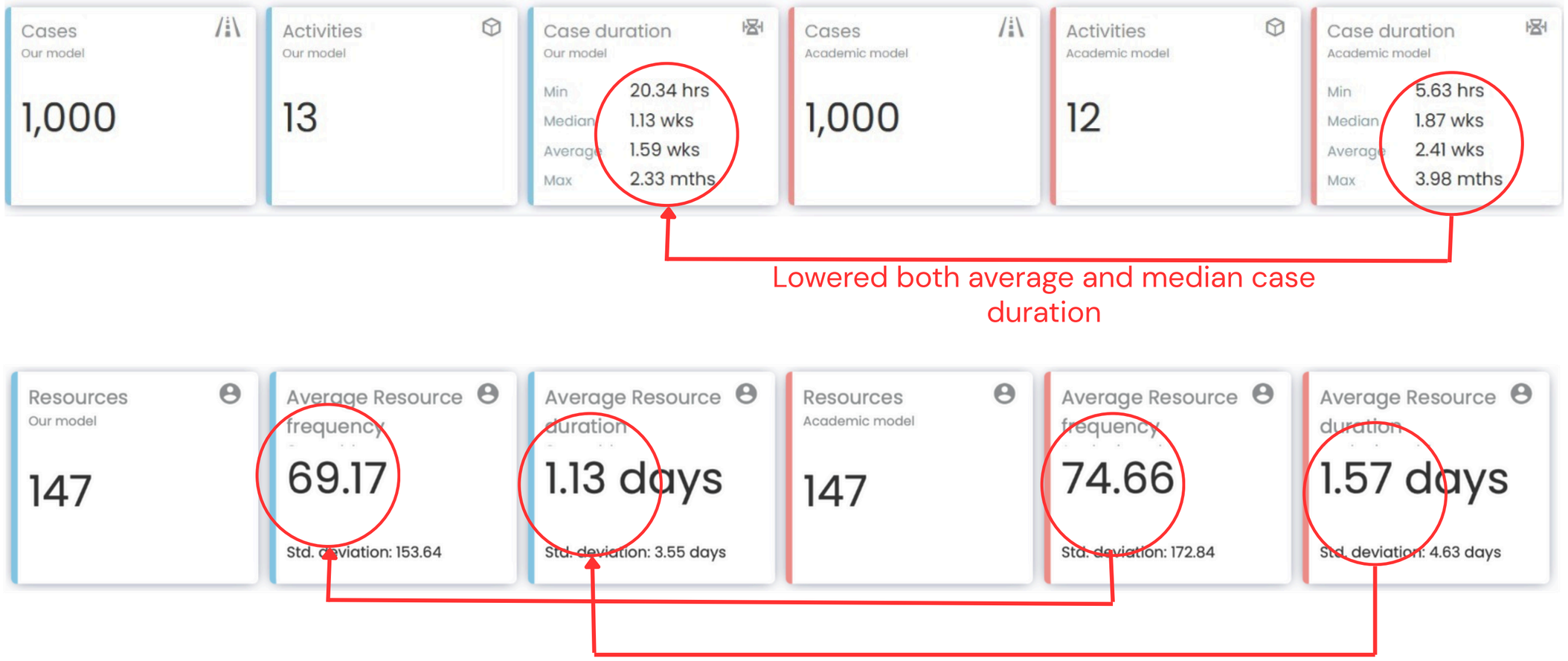
Top 5 scores

Final BPMN model

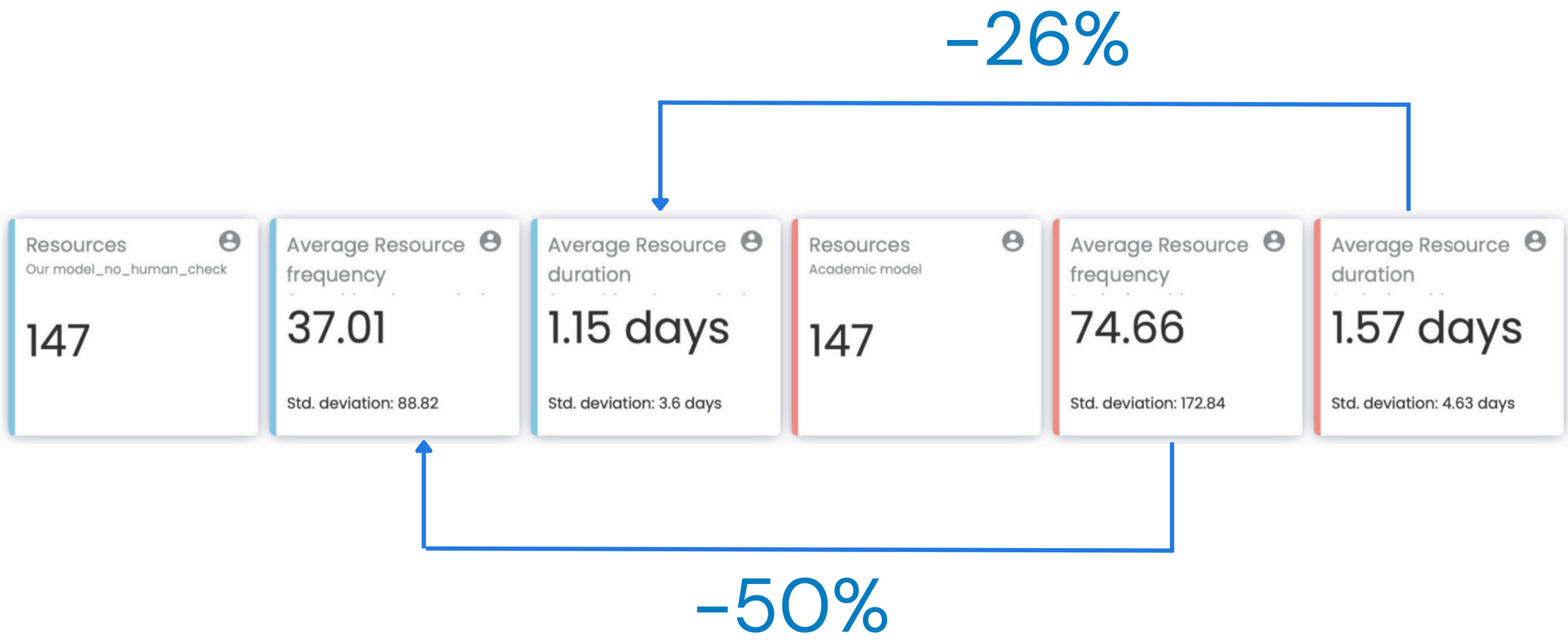


New proposed model
→ +1 activity
→ +2 splits
→ Redesigned workflow

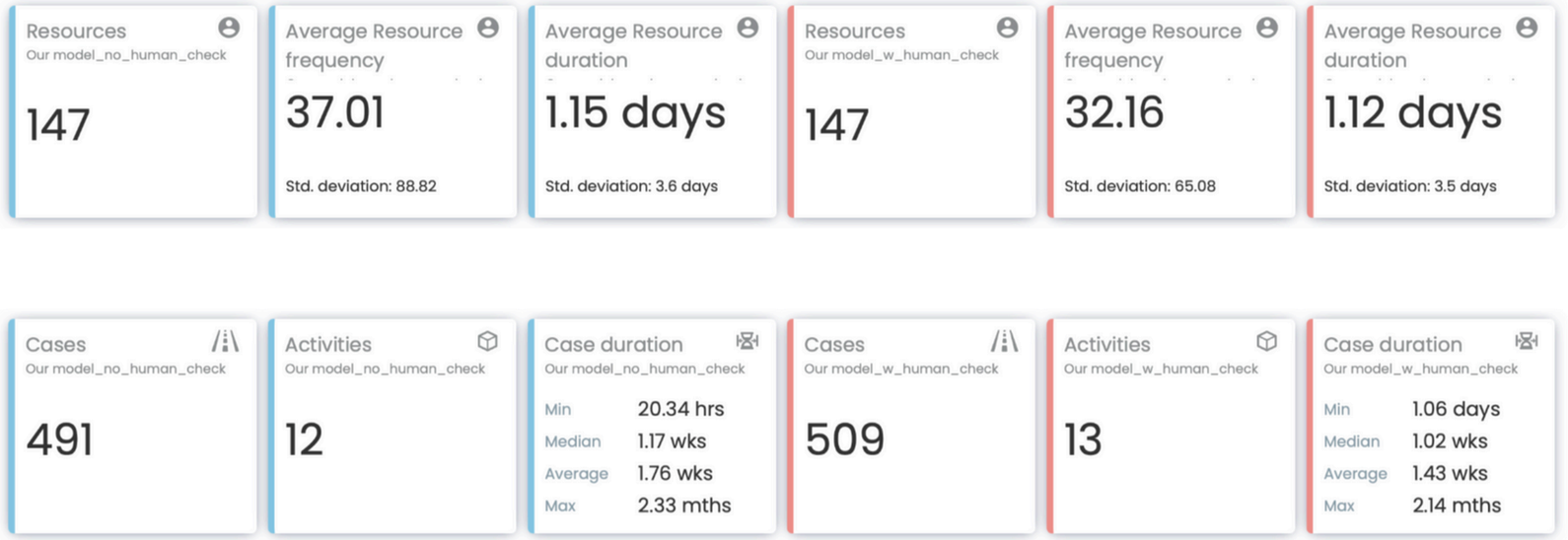
Overall Process Analysis vs Baseline



Resource Efficiency in Standard Cycle



What happens at the splitting?



- Almost balanced Resource usage
- One of the two branches slightly faster

Business Impact:
Two Key Questions Answered

By predicting outcomes early, our ML model immediately rejects likely-to-fail applications, allowing resources to focus on high-potential cases without delays from hopeless applications clogging the system.

Process automation and ML-based early filtering reduce manual workload by immediately rejecting low-probability cases, allowing resources to focus on successful applications with significantly higher efficiency in both time and resources.

Future improvements

- Ensure data quality and digitization rates meet requirements for automated validation
- Collaborate with domain experts to design the "Human check" activity without introducing delays or bias
- Conduct cost-benefit analysis with ROI estimates to quantify economic impact and guide investment decisions

Thanks
for your attention!