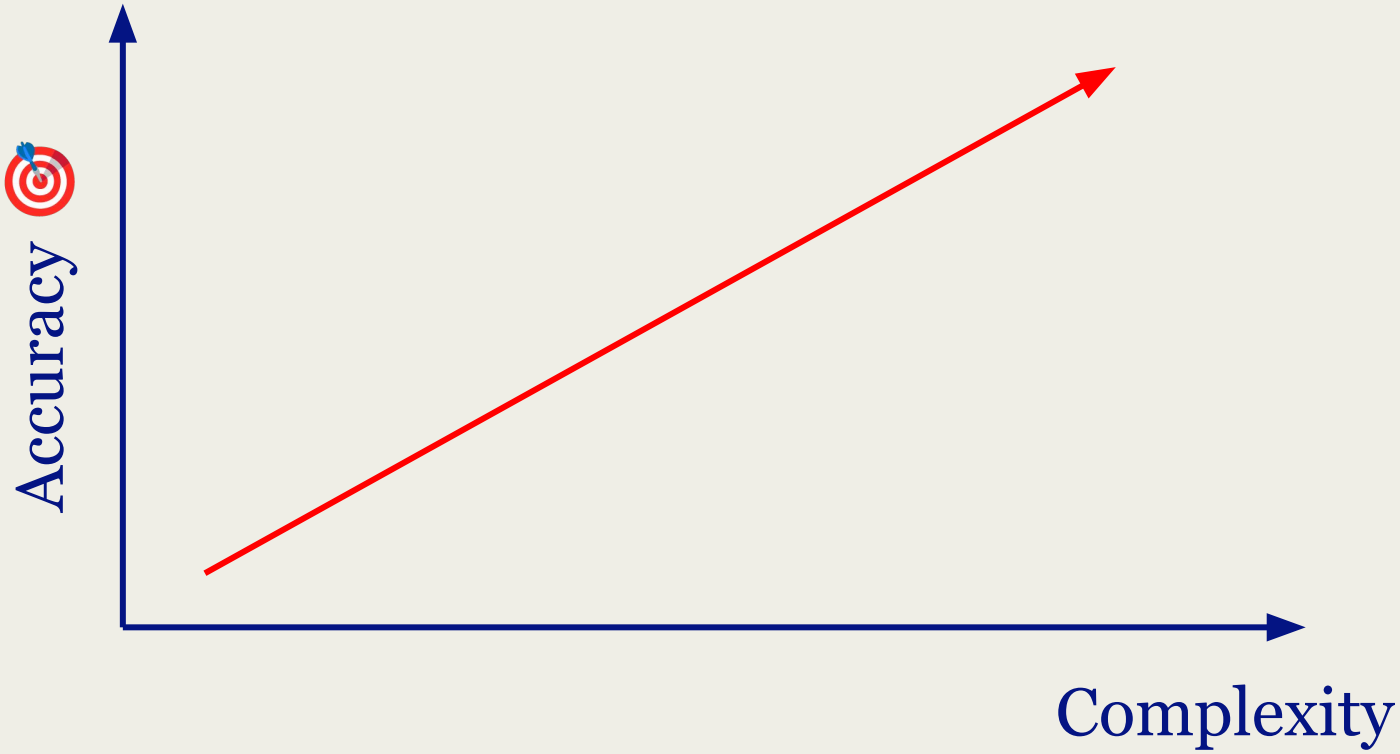
Julius Bär

Onboarding Quest

Team Based Bayes - Shakir Yousefi, Igor Pradhan, Rahul Steiger, Joel André

How to approach a challenge like this?

Gradually Increase Complexity



What can we do without Machine Learning?

Tabular Consistency Checks

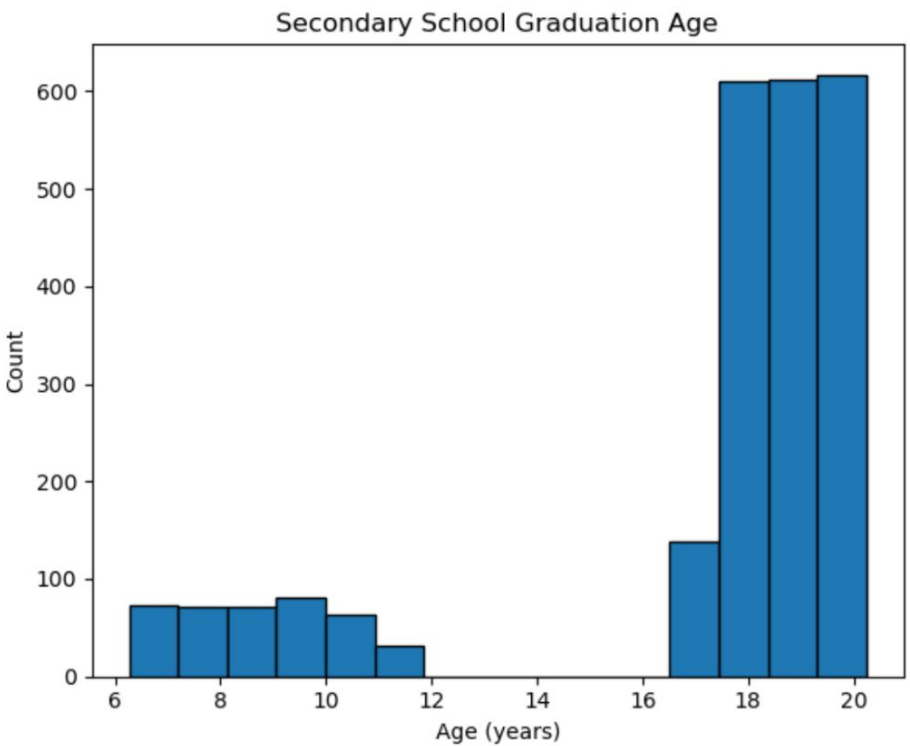
- Tabular consistencies across the datasets ‘passport’, ‘client_profile’, and ‘account_form’
- Reject onboarding in cases where fields such as ’first_name’, ‘middle_name’, ‘last_name’ do not match.

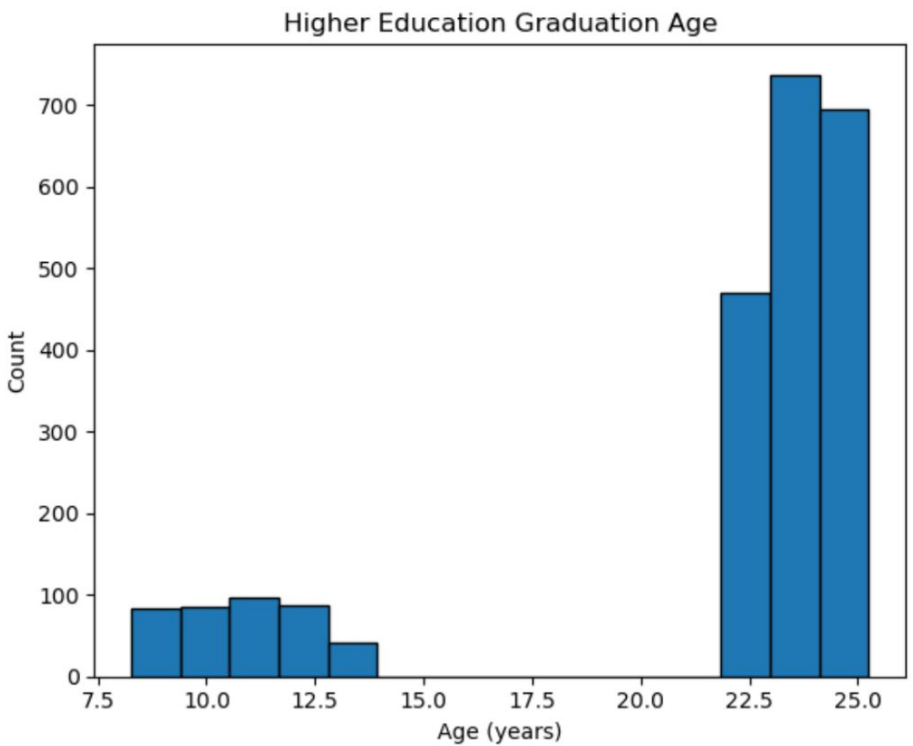
- This already gives ~75% accuracy without false negatives

Rule-Based Modelling and Sanity Checks

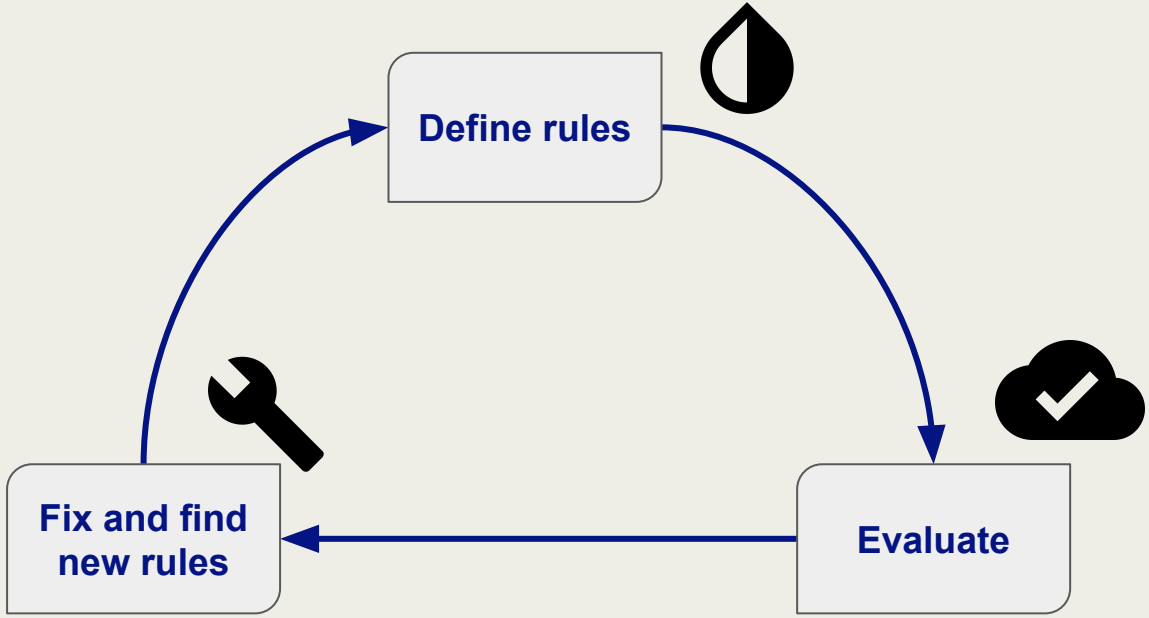
- Find Inconsistencies in the client's timeline.
- Use birth_date with graduation_year fields from both higher_education and secondary_school to verify chronological plausibility.

- Around ~85% accuracy without false negatives.


Secondary School Graduation Age
Count
Age (years)
0
100
200
300
400
500
600
6
8
10
12
14
16
18
20


Higher Education Graduation Age
Count
Age (years)
0
100
200
300
400
500
600
700
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0

Iterative Approach



Interesting Examples

Kuipers Snijders Spaans
Client 79

- €3 310 000 AuM
- Born 1965
- First job at Rijksmuseum Amsterdam
- Started working in 1961

Johanna Voutilainen
Client 712

- Finnish
- Hedge Fund Manager
- Single

Client description:
“Voutilainen is married to Laaksonen”

Using LLMs

- Use LLMs to reject clients based on their description
- Host Qwen 2.5 72B on our own with VLLM
- Too many false negatives

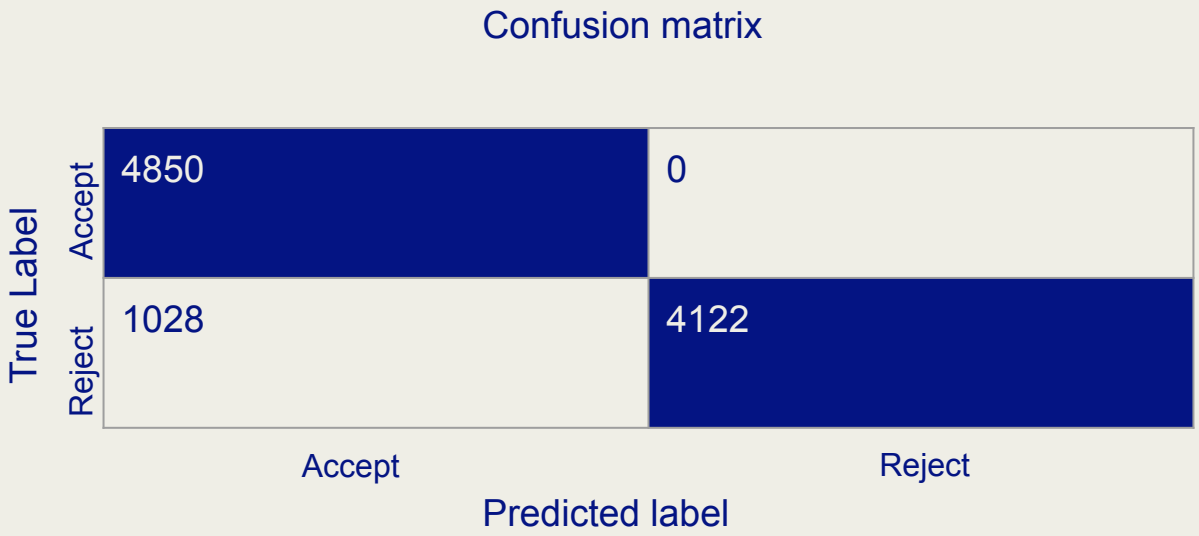
```
f"Here is the client profile {all_clients[i]['client_profile']}\n"
f"Here is the description {all_clients[i]['client_description']}\n"
"Does the description contradict the client profile net worth?\n"
"If this is the case say YES, otherwise say NO."
```

Classifier

Classification

of the original

10 000 training dataset



Implications

- High Accuracy
- Current model low false negatives
- Only do human checks on accepted through model
- Reduces human labour by 41.2%

Classifier

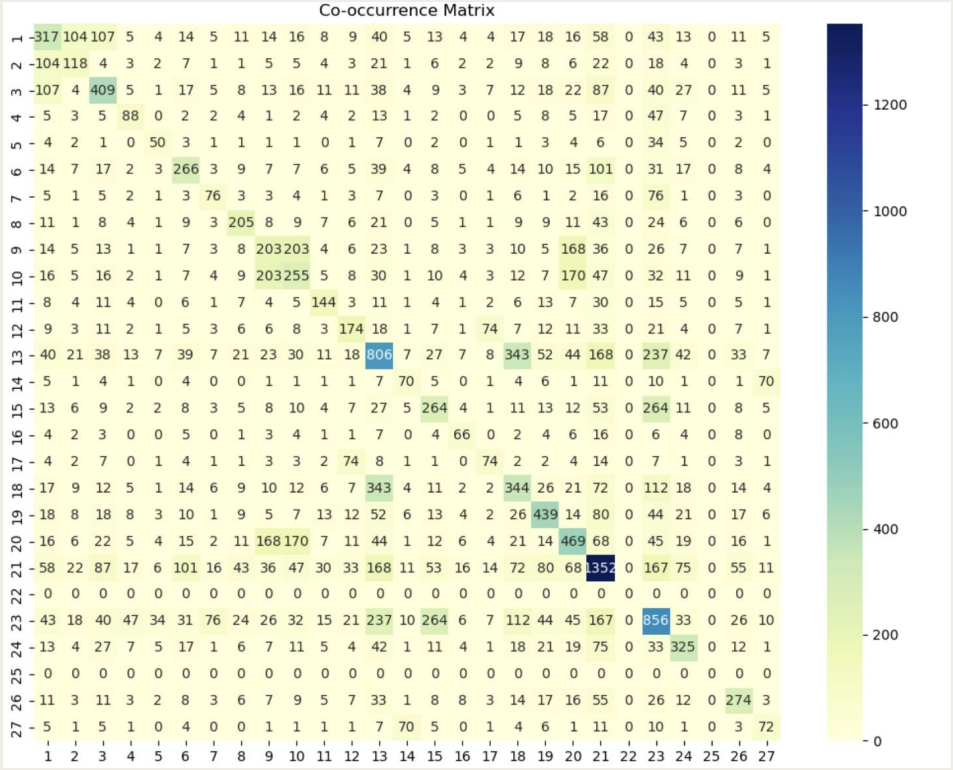
Every rule assigned a number

↓

Client Flagged based on rules

↓

Add to co-occurrence matrix



Evaluation

- Evaluation on 1000 points