AutoML Debt Scoring

Automated Machine Learning for More Efficient Collection

Business Challenge

- The mean collection rate in the debt collection industry is low
- Collection is still labour-intensive

→ Identifying the accounts most likely to be settled is key

- Our client <redacted> currently offers a prediction model to identify them, but it performs poorly

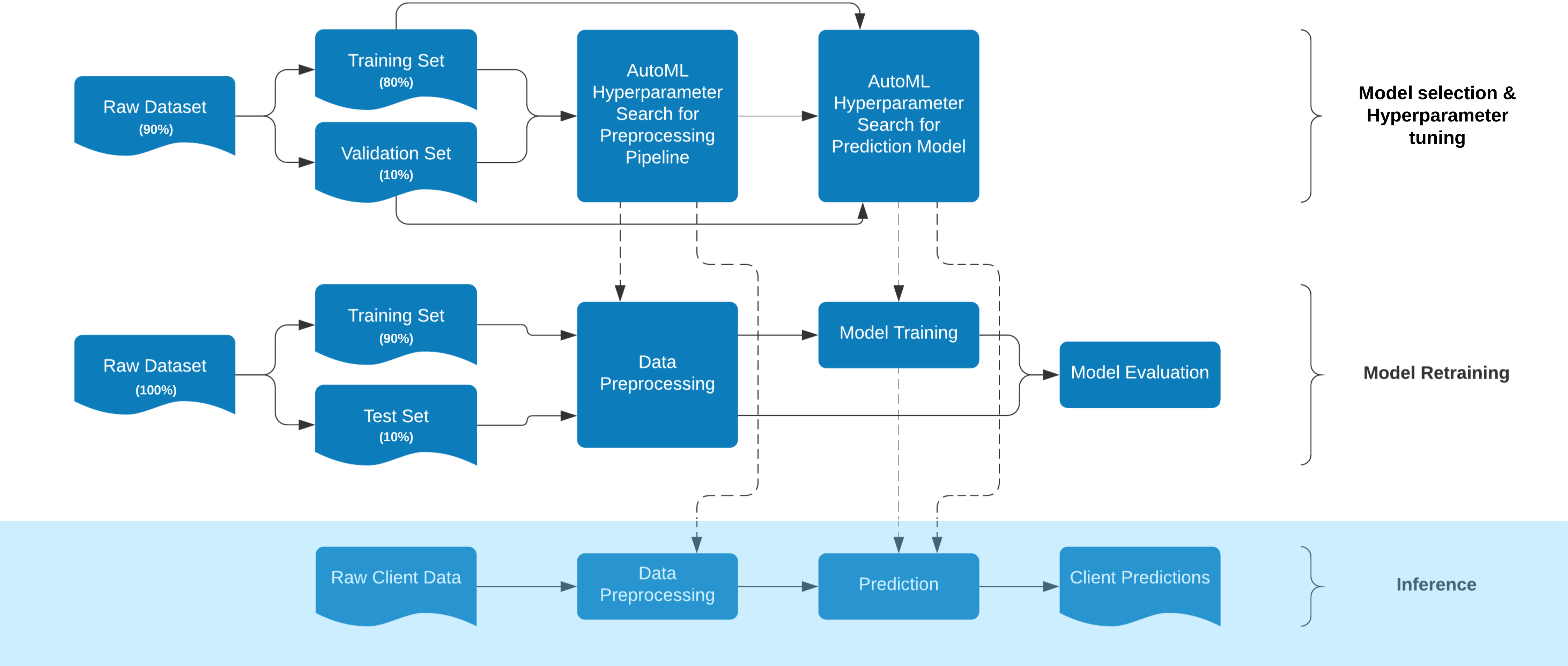
Project Deliverables

- A new prediction model to quickly identify priority accounts
  - <redacted>
- A complete system:
  - <redacted>
- Requirements:
  - Flexibility to adapt to each agency's operations (see next slide)
  - <redacted>
  - What performance metrics? See discussion later.

Modelling Decisions

- Treat as a Binary Classification problem
  - Will pay? Y/N + probability
- Agencies (our client's clients) all operate and use our client's product differently
  - Different operations and data nomenclature
  - Therefore, we must fit a different model for each agency
- Automated Machine Learning pipeline:
  - Develop an AutoML pipeline able to fit different models on different data
  - Promotes quick deployment of new clients and retraining of existing ones

Scoring Model Pipeline



Performance Evaluation: What Metric?

- Confusion Matrix

Real life / Model	Actual Yes (will pay)	Actual No (won’t pay)
Predicted Yes	True Positives (TP)	False Positives (FP)
Predicted No	False Negatives (FN)	True Negatives (TN)

Performance Evaluation: What Metric? (2)

- You said: “*We want positive predictions to be correct…*”

→ Emphasis on predicted Ys rather than predicted Ns
→ Emphasis on Precision:

- Proportion of Actual Yes among the Predicted Yes
- Our confidence that a positive prediction is correct
- Mathematically:

$$Precision = \frac{TP}{TP + FP}$$

Model \ Real life	Actual Yes (will pay)	Actual No (won't pay)
Predicted Yes	True Positives (TP)	False Positives (FP)
Predicted No	False Negatives (FN)	True Negatives (TN)

Performance Evaluation: What Metric? (3)

- But Precision alone is not sufficient
  - Trade-off between Precision and number of Predicted Yes
  - Remember, the model computes *probabilities* of payment

  → Predicted outcome (Y/N) depends on an arbitrary decision threshold
  - Extreme example:
    - To be very confident in her model's positive predictions, Alice decides only to consider "Predicted Yes" above 0.99 probability
    - Her model is good, so almost all these cases are correctly predicted (Actual Yes), leading to Precision $\approx$ 1.0
    - But she will have very few cases to work with, because very few have a probability > 0.99

Real life / Model	Actual Yes (will pay)	Actual No (won't pay)
Predicted Yes	True Positives (TP)	False Positives (FP)
Predicted No	False Negatives (FN)	True Negatives (TN)

Performance Evaluation: What Metric? (4)

- You said: *"…while having enough positive predictions to work on"*

  → Precision must be balanced with another metric to make sure we are not too selective

  → Recall:

  - Proportion of Actual Yes that have been correctly identified
  - If high, means that we do not leave many Actual Yes on the table
  - Mathematically:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Real life / Model	Actual Yes (will pay)	Actual No (won't pay)
Predicted Yes	True Positives (TP)	False Positives (FP)
Predicted No	False Negatives (FN)	True Negatives (TN)

Why Not Use Accuracy?

- Accuracy: proportion of all predictions that were correct

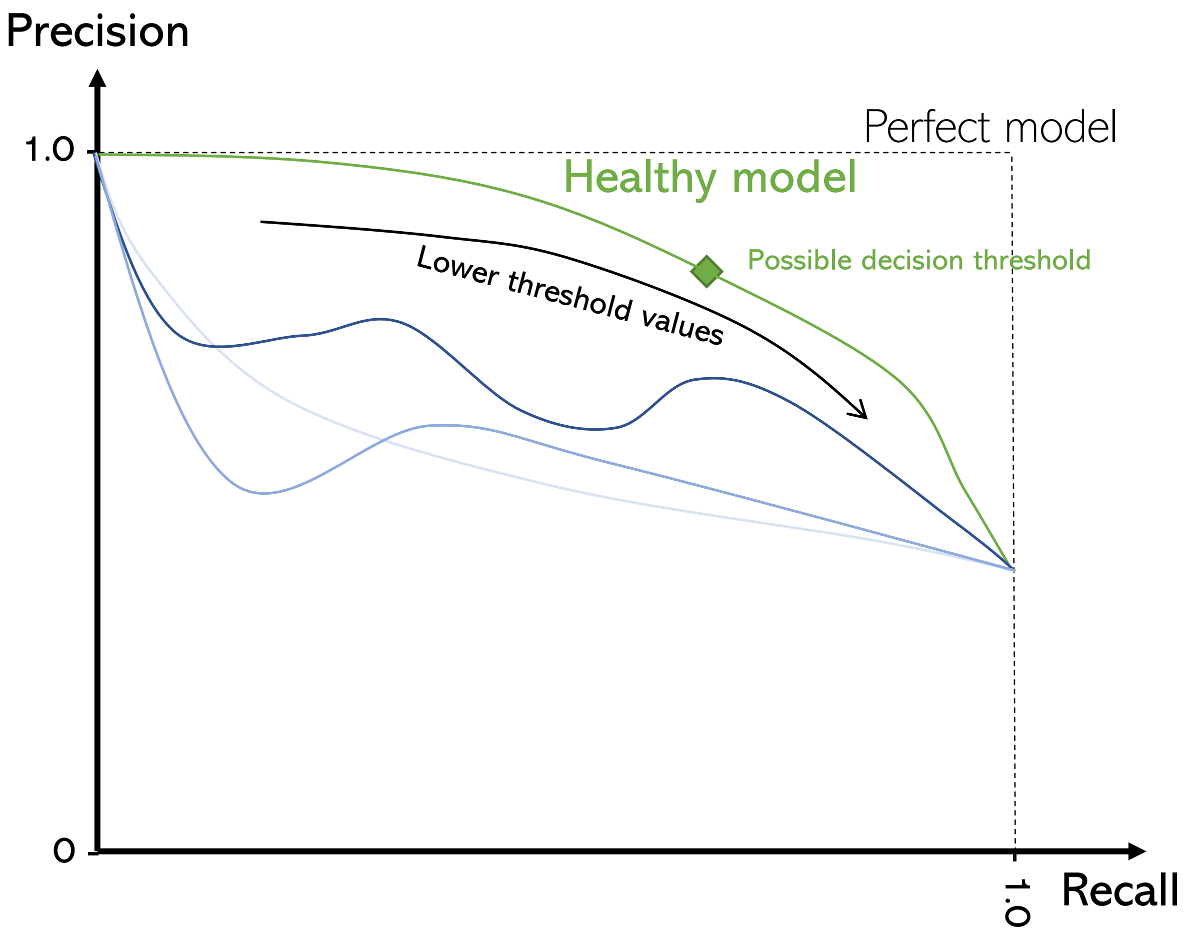
$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

- Pros:
  - Easy to interpret
  - Good when both classes (Y/N) are equally important
  - Good with balanced dataset
- Cons:
  - Doesn't assign more importance to Predicted Yes
  - Datasets are imbalanced

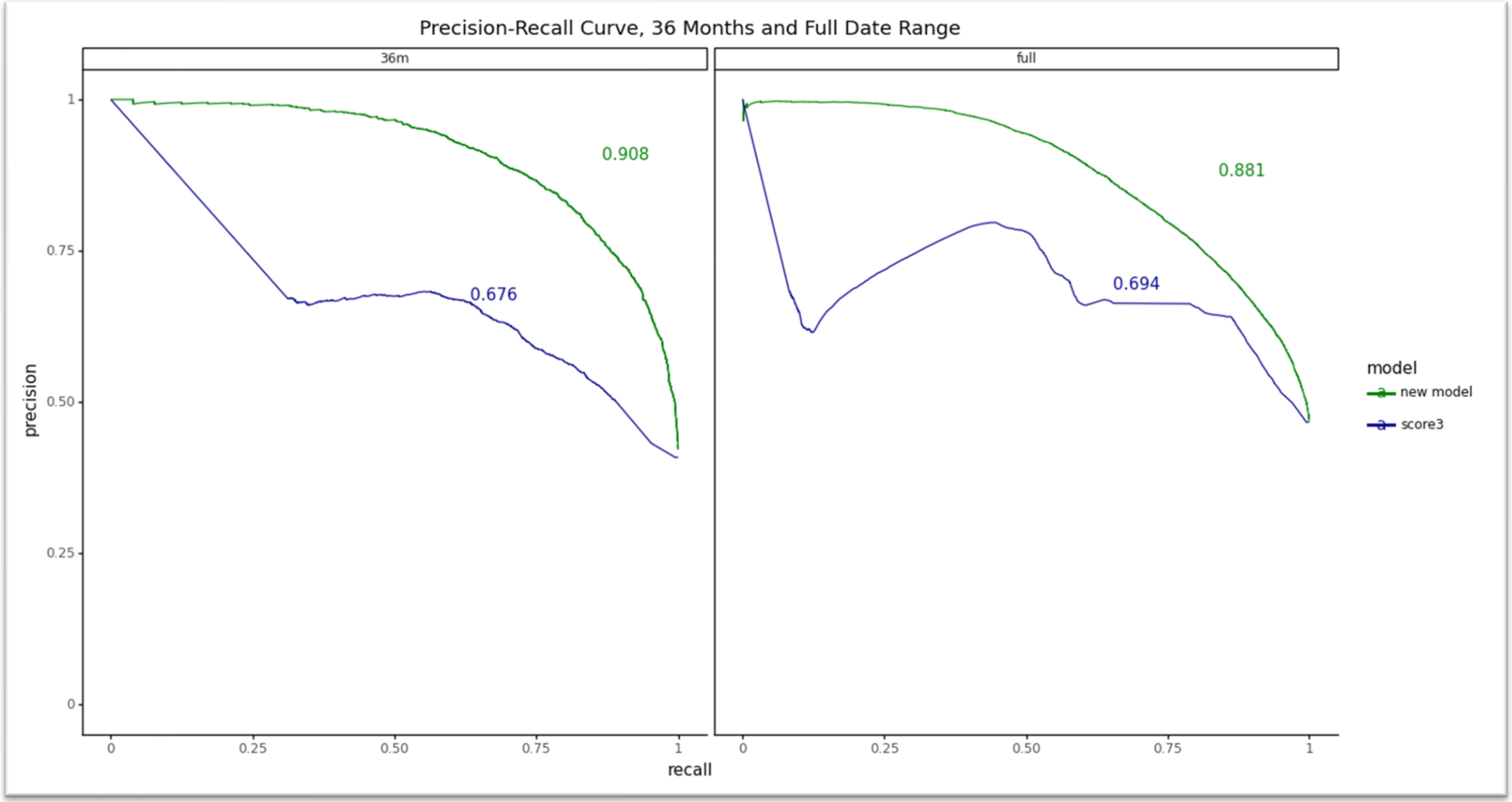
Model \ Real life	Actual Yes (will pay)	Actual No (won't pay)
Predicted Yes	True Positives (TP)	False Positives (FP)
Predicted No	False Negatives (FN)	True Negatives (TN)

Precision-Recall Curve

- Shows the P/R trade-off as we move the decision threshold
- We want the line to be towards the top-right corner
  - Maximises precision AND recall for each threshold value
- In other words, we want to maximise the area under the PR curve (AUC-PR)

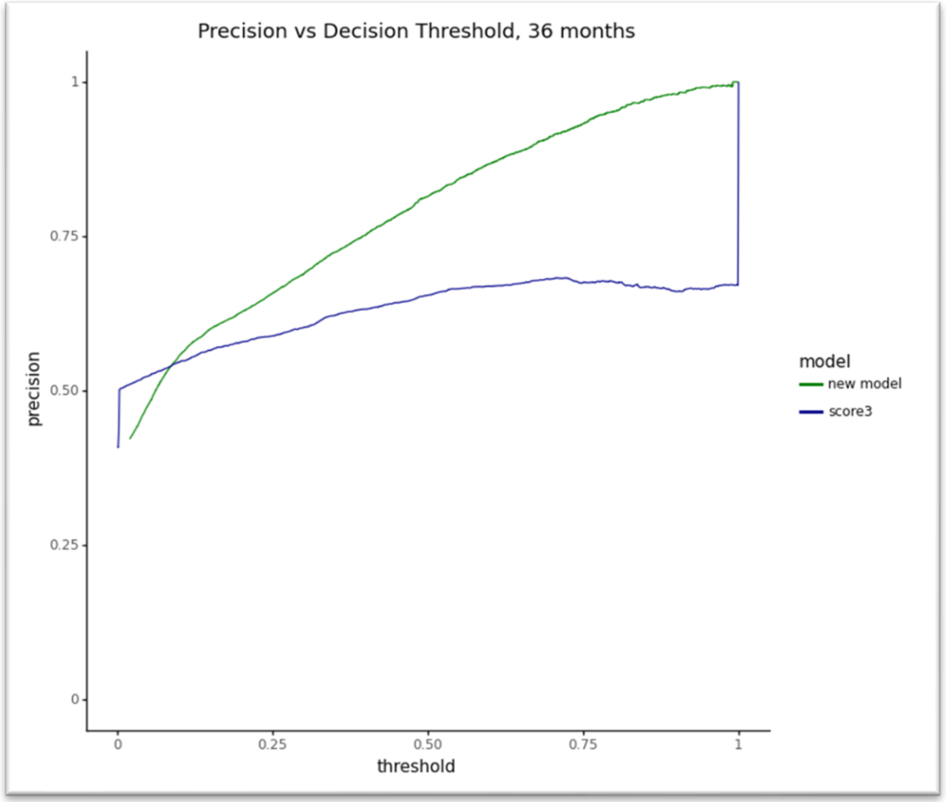


Results (agency <X>): Precision-Recall Curves

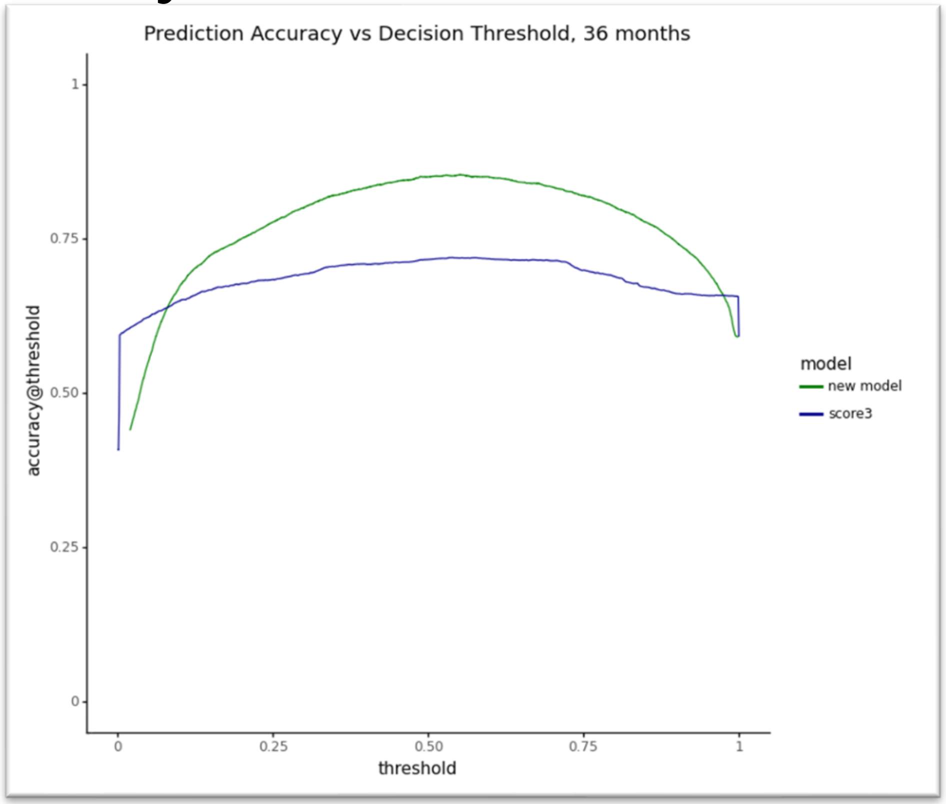


Results (agency <X>): Precision and Accuracy vs Threshold

Precision



Accuracy



Business Implications

5 hours per account @ <X> per hour

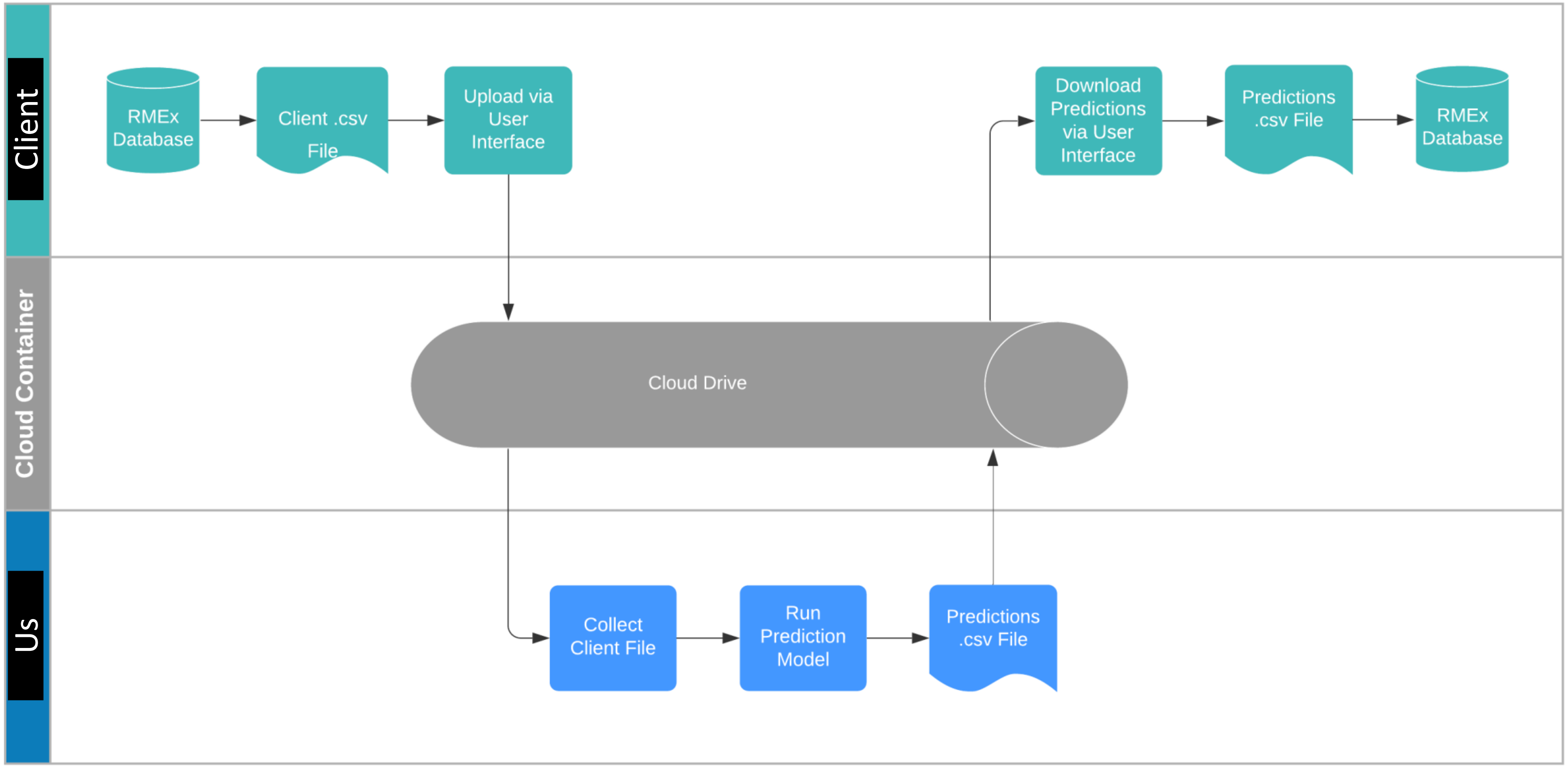
- Profit per account worked:
  - x1.4
- ROI per account worked:
  - x1.4
- Opportunity cost[1]:
  - -1/3

10 hours per account @ <X> per hour

- Profit per account worked:
  - x5
- ROI per account worked:
  - x5
- Opportunity cost[1]:
  - -1/3

[1] from “Actual Yes” accounts not chased because they were misclassified as “Predicted No”

Model Deployment (Handover Phase)



Proposed Future Developments

- <Redacted>
- <Redacted>
- <Redacted>
- <Redacted>