Predictive Premium Subscription Adoption at Website XYZ

Building a predictive model to identify adopters of premium subscription

- Aakash Patil
- Harshal Sable
- Charlotte Wang
- Rahma Hassan
- Abraham Perunthekary George

What we'll talk about

1. The problem statement and the key pain-points we intend to address
2. Our solution and the process of narrowing to this solution
3. Interpretation of the results and robustness of our solution
4. Recommendations aligned with XYZ's business and goals

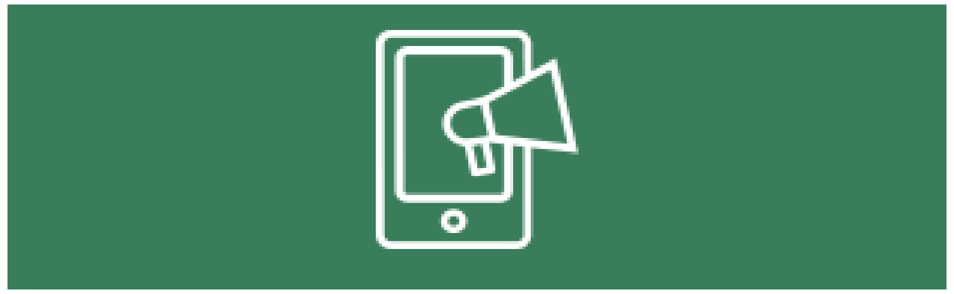
The Problem and The Plan

What's the problem?

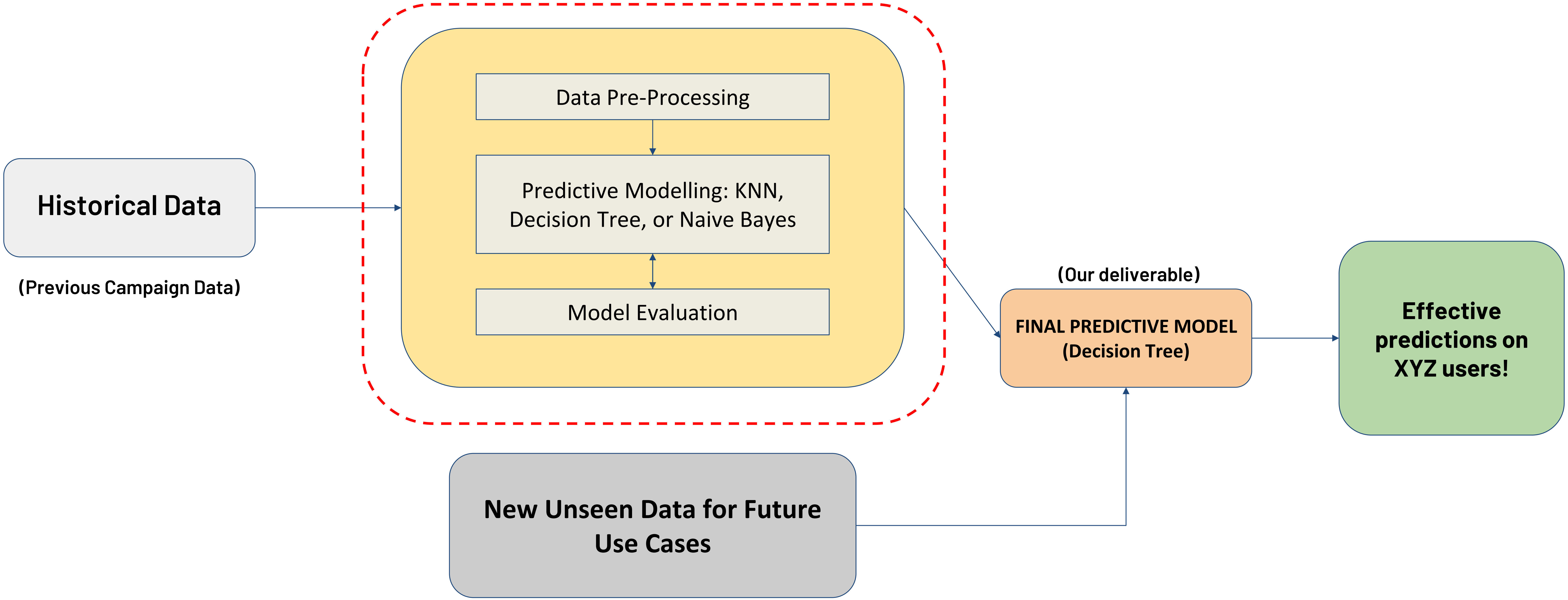
- Identifying users who will likely convert to premium subscriber.
- High volume of expected non-adopters in the user-base.
- Need for a prediction tool that can predict effectively with historical information.

What's the plan?

- Use predictive analytics to enrich the target list for marketing so that there is a higher proportion of adopters targeted.
- Choose best prediction model that identifies adopters with high accuracy while filtering out non-adopters.
- Draft an implementation plan and relay best practices.



Solution Lifecycle

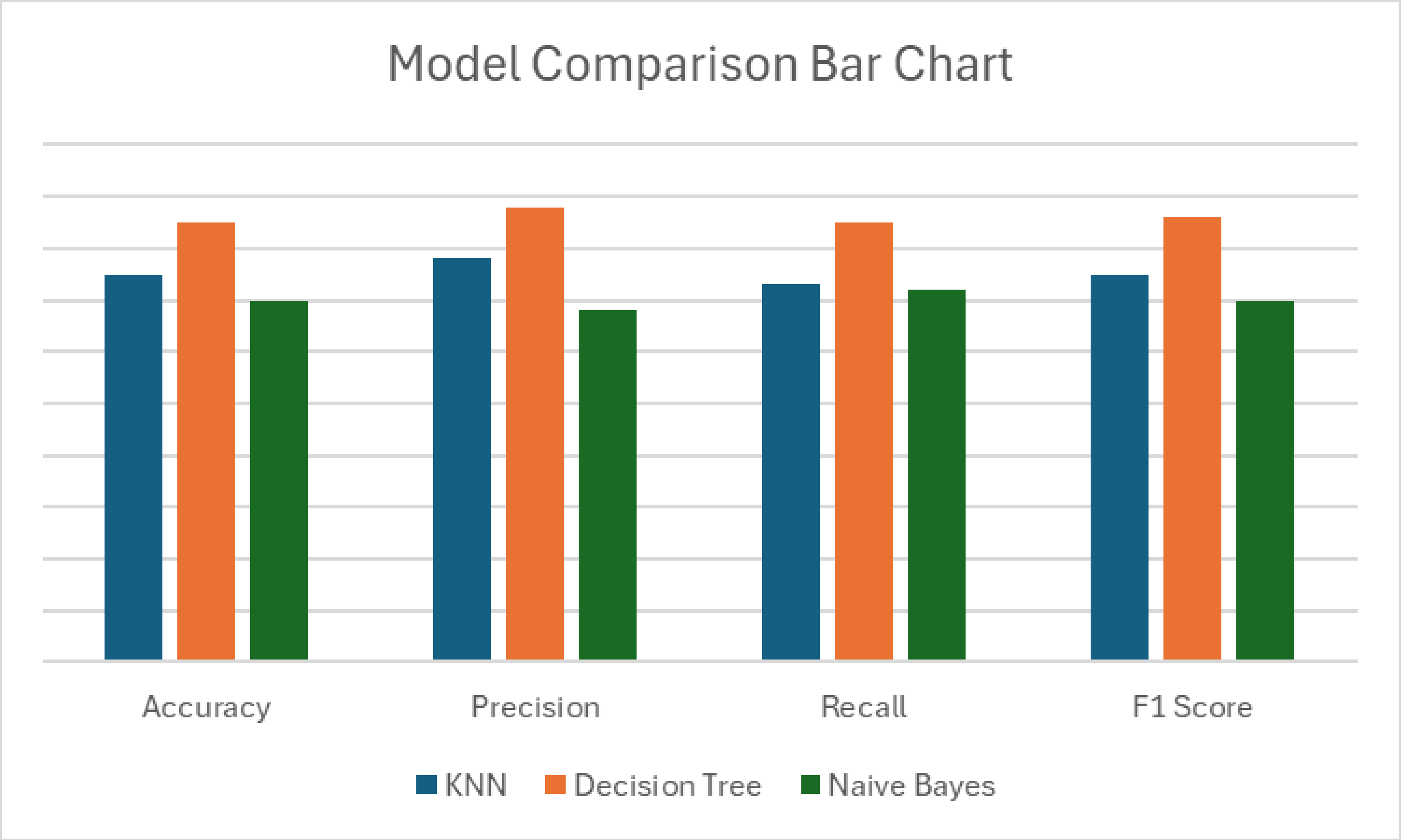


Our Predictive Model

- We trained our model by increasing 'adopter' users from the historical data to increase the quality and accuracy.

- Through the iterative process of selecting and evaluating predictive models, we chose Decision Tree to predict adopters.

- The decision tree performed the best, giving high quality results in line with our goals to optimize efficiency of campaigns.



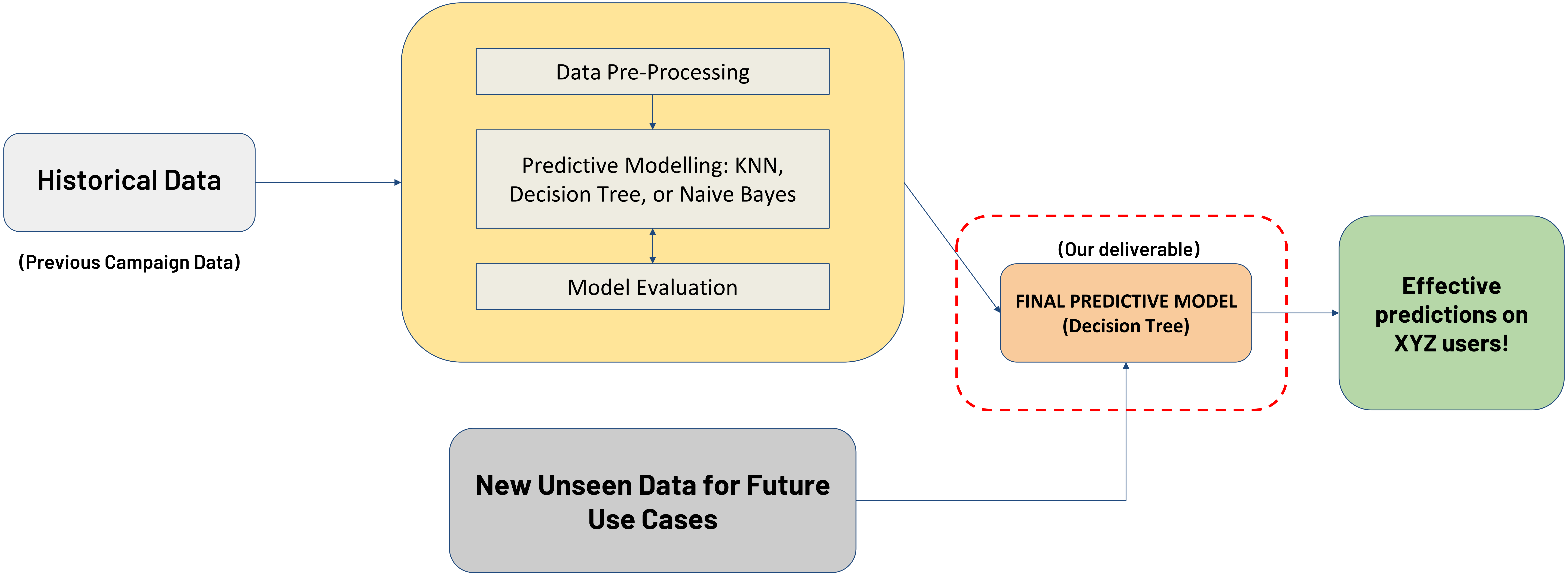
Our Predictive Model

- ➔ We have chosen **Recall** (a metric that gives the percentage of adopters correctly predicted) as our primary metric to increase, as we want to capture as many adopters as possible.

- ➔ **F-Measure** (a combination of precision and recall) holds the performance of the model in check.

- ➔ A tradeoff between our adopter capturing ability (recall) and overall quality of model (F-Measure) exists. As recall increases, F-measure decreases.

Solution Lifecycle



Model Results

Running our model on testing data:

	Actual Adopter	Actual Non-Adopter
Predicted Adopter	262	3,996
Predicted Non-Adopter	46	4,004

- Here, we can see that model was able to predict 262 adopters correctly, while only incorrectly predicting 46. This shows us the recall measure of 85% (262/308).

- The recall of 85% shows the ability of our model to identify actual adopters effectively, while a random selection marketing would've had around a 50% chance of identifying adopters.

- We can also see that the model was also able to identify more than half non-adopters correctly.

Flexibility of the Model

- The model also can be tuned to strategy and budget - threshold value in the model code can be changed to increase recall (ability to identify adopters) while disproportionately also increases users to target.

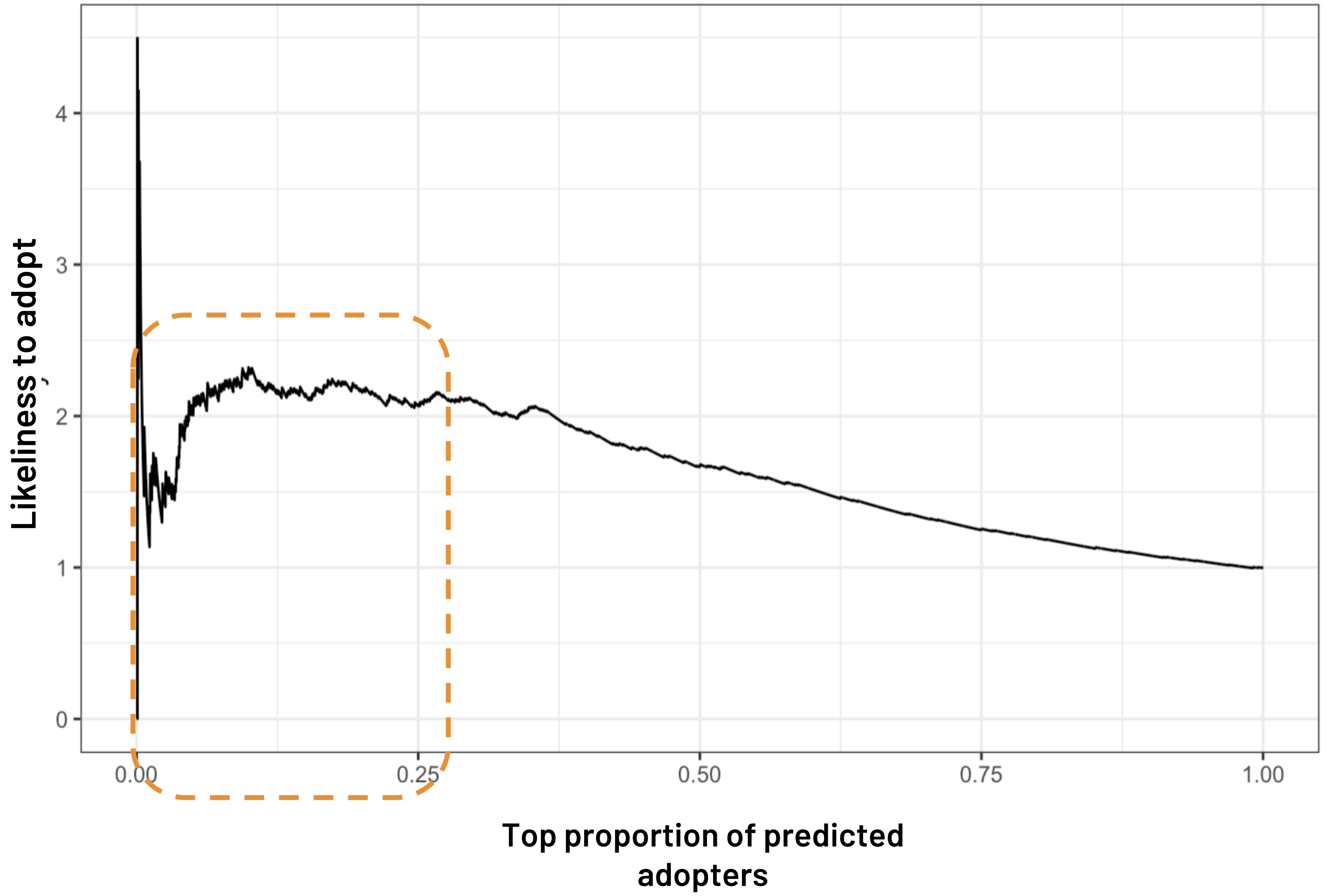
- Such a strategy can be used if there is no budget constraint, and XYZ does not want to miss any adopters.

- Same applies to cost-saving strategies by lowering the threshold values and observing lower capture of adopters.

- We have set a default threshold value, which gives the observed results. We believe that this is the best specification to get the most effective results.

Recommendations

- We believe that current specification of the model will give strong results.
- We would suggest targeting the top 25% users identified by from our model to be adopters.
- Interpreting the curve on the right, the top 25% users are twice (y-axis = 2) as likely to adopt the premium subscription than other users.
- To utilize marketing resources effectively, focusing on the top 25% of users identified by the model - rather than randomly selecting potential customers - can increase the likelihood of conversions and reduce resource waste on users who are less likely to adopt.




THANK YOU