**Problem Framing**

<table>
<tr><th></th><th>qualitative</th><th>quantitative</th><th>question</th></tr>
<tr><td>Current State</td><td>too many fraudulent transactions => less customers => less revenue => loss to the bank</td><td>10% fraudulent transactions => 5% less customers => 5% loss in revenue</td><td>what is the avg. number of fraudulent transactions in the current situation and what we can do to decrease the number?</td></tr>
<tr><td>Objectives</td><td><ul><li>build a model that can detect fraud transaction before it gets completed</li><li>decrease fraudulent transaction => improve customer experience => more revenue</li></ul></td><td>reduce fraudulent transaction by at least 5% => 5% more revenue</td><td>How we can detect the fraudulent transactions?</td></tr>
<tr><td>Benefit/ Cost Tradeoff and Prioratization</td><td><ul><li>cost of errors:</li></ul>FN => Fraudulent transaction marked as non-fraudulent => risk increases => bad user experience => loss of revenue<br><br>FP => non-fraudulent transaction marked as fraudulent => less revenue<br><ul><li>benefits of correct predictions:</li></ul>TP => correctly identified fraudulent transaction => customers are protected => better user experience => more revenue<br><br>TN => correctly marked non-fraudulent transactions => maintained user experience as expected => no significant impact on revenue</td><td>cost-benefit matrix<br><table><tr><td>c(TP)</td><td>c(FP)</td></tr><tr><td>c(FN)</td><td>c(TN)</td></tr></table><br>1% TP =><br>+ 0.5% revenue<br><br>1% FN => 1% very bad experiences => 0.1% risk of customers’ assets => 10% less revenue<br><br>1% FP => -0.5% revenue<br><br>1% TN => no significant impact on revenue</td><td>what is the cost of errors/benefits of correct predictions and why?</td></tr>
</table>

| | | | |
|---|---|---|---|
| Constraints | can only afford a small FN percent => very small percent of very bad user experience => limited risk of customers' assets => limited loss of revenue | at most 10% FN => 1% very bad experiences => 0.1% churn => 0.1% risk of revenue loss => acceptable risk for 5% potential upside in revenue | what are the acceptable risks/budgets and why? |
| Desired State | • benefit: significantly lesser fraudulent transactions => significantly better user experience => significantly more customers => significantly better revenue<br>• cost: very few false negatives => limited risk of very bad user experience => limited risk of customers' assets => limited risk to revenue | • at least 50% decrease in fraudulent transactions (from 20% to 10%) => 5% more revenue<br>• at most 10% false negatives => 1% very bad experience => 0.1% risk to revenue | what is the desired outcome (benefits/costs) that we want to see and why? |

**Why ML**

| | qualitative | quantitative | question |
|---|---|---|---|
| best non-ML alternative hypothesis | classify based on amount of money transaction => too many FP and FN => very bad user experience => lesser customer => loss of revenue | 50% FN 70% FP => not cleaning enough fraudulent transactions and causing more complaints for misplacing fraudulent transactions as genuine => 5% revenue loss risk | what are the non-ML alternatives and why are they problematic? (pains/missed gains)? |
| ML value proposition hypothesis | much fewer FP and FN => much better user experience => much better revenue | 10% FN 50% FP => 50% decrease in fraudulent transactions (from 20% to 10%) at the expense of 1% bad | what are the advantages (pain relievers/gain creators) of ML solution and why? |

| | | | |
|---|---|---|---|
| | | engagements => 5% increase in revenue at the expense of 0.1% risk | |
| ML feasibility hypothesis | • data: labelled samples of each person's transaction data<br>• model: state of the art review suggests promising candidates are available | • data: around five thousand samples<br>• model: state of the art claim solutions with 10% FN 20% FP | what data and model are good candidates and why? |

## ML Solution Design

| | choices | metrics | experiment |
|---|---|---|---|
| data | (labelled) money transaction data | • label imbalance | • randomized 70/15/15 train/validation/test split |
| model | pr(fraud) | • AUCPR (Precision recall curve) | • rule based heuristic<br>• tf-idf + logistic regression<br>• tf-idf + random forest<br>• BERT + logistic regression<br><br>train these benchmark models (from simpler to more complex) using train data. validate and tune using validation data. select the model with best AUCPR on test data |
| action | if pr(fraud) > | • precision | • choose a |

| | | | |
|---|---|---|---|
| | threshold: auto take down | • recall<br>• confusion matrix | threshold to maximize the recall (estimated reward) subject to recall > 90% |
| reward | • decrease in fraudulent transaction<br>• cost of misclassification | • % decrease in fraud<br>• % increase in daily active users | • shadow test<br>• A/B test |