

Running use case: Fraud detection

ML Use Case Design

Problem Framing

	qualitative	quantitative	question				
Current State	Fraud create loss to the bank	5% fraud => bank can't tolerate	what is the current situation (pains/desires) that we want to address and why?				
Objectives	<ul style="list-style-type: none">● Build a model that can detect fraud● Less fraud => less loss to the bank	Fraud >= 5% bank may loose a huge revenue	what is that we want to do and why? (to improve topline/bottomline?)				
Benefit/ Cost Tradeoff and Prioratization	<ul style="list-style-type: none">- Cost of error: If we decrease the fraud tolerance much we may loose potential customer	<div>cost-benefit matrix</div> <table><tr><td>c(TP)</td><td>c(FP)</td></tr><tr><td>c(FN)</td><td>c(TN)</td></tr></table> <div>1% TP => 0.5% less fraud => 10% less loss for bank</div> <div>1% FP => 1% more risk of loosing customer => 1% loss for the bank</div> <div>1% FN => 0.1% more fraud => 2% more loss for the bank</div> <div>1% TN => no significant impact on revenue</div>	c(TP)	c(FP)	c(FN)	c(TN)	what are the cost of errors/benefits of correct predictions and why?
c(TP)	c(FP)						
c(FN)	c(TN)						

Constraints	can only afford a small FP percent => we may loose the customer => less profit to the bank	At most 5% FP=> 5% revenue loss for 10% less fraud	what are the acceptable risks/budgets and why?
Desired State	<ul style="list-style-type: none"> benefit: significantly lesser fraaud=> significantly less money loss => significantly better revenue cost: very few false positives => limited risk of losing cutomer => limited risk to revenue 	<ul style="list-style-type: none"> at least 50% decrease in Fraud=> 10% less loss of money => 5% more revenue at most 10% false positives => 1% loss of customer =>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 a income of person => too many FP and FN => more customer loss as well fraud	50% FP 70% FN => loose 10% of customer and 10% more fraud	what are the non-ML alternatives and why are they problematic? (pains/missed gains)?
ML value proposition hypothesis	much fewer FP and FN => less customer loss and less fraud => much better revenue	10% FP 50% FN => 50% less fraud in expense of 10% less customer => 5% increase in revenue at the expense of 0.1% risk	what are the advantages (pain relievers/gain creators) of ML solution and why?
ML feasibility hypothesis	<ul style="list-style-type: none"> data: labeled dataset of each person bank history 	<ul style="list-style-type: none"> data: around five thousand samples model: state 	what data and model are good candidates and why?

	<div>with their fraud history</div> <ul style="list-style-type: none">• model: state of the art review suggests promising candidates are available	<div>of the art claim</div> <div>solutions with 10% FP 20% FN</div>	
--	--	---	--