Zürcher Hochschule
für Angewandte Wissenschaften

School of
Management and Law

P2P lending: Employing explainable AI to optimize the return target function of a loan portfolio

Building Competence. Crossing Borders.

Dr. Thomas Gramespacher, Dr. Jan-Alexander Posth

ZHAW - Institute of Wealth & Asset Management

Contents

i. Defaults are expensive – a motivation

ii. P2P context & data

iii. AI, ML & profit optimization: Avoiding (potential) defaulters vs. maximizing economic payoff

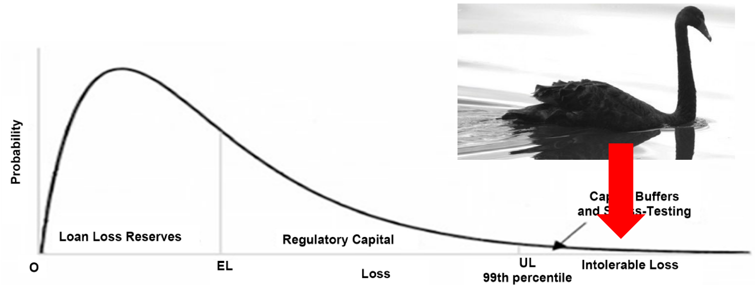
iv. XAI: Providing transparency without sacrificing accuracy

v. Conclusio

i. Defaults are expensive – a motivation

Remember: If you (or somebody else) can put a **price tag** on it, it's most probably a risk. If not, it's something else – e.g. uncertainty, bad luck or chaos.

Credit Risk is **asymmetric**: Credit risk is skewed towards large losses with low probability and small gains with high probability ("fat tail").



Economic payoff = target function:

Cost of defaults = (1-RR) * PD = "Fair" risk premium

"few" **"many"**

ii. P2P context & data: Overview

Data source:

Fintech-ho2020 project: www.fintech-ho2020.eu
Use case I: smaller_dataset.csv

Total data set:

Total # loans:	4514
Non-defaults:	4016 (88.97%)
Defaults:	498 (11.03%)

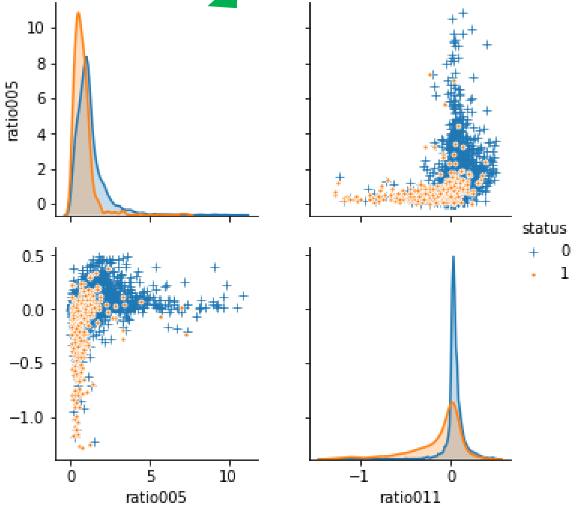
Split in-sample (train) **60%** / out-of-sample (test) **40%**: 2708 loans vs. 1806 loans
Split preserves ratio of defaults to non-defaults in each sub-set

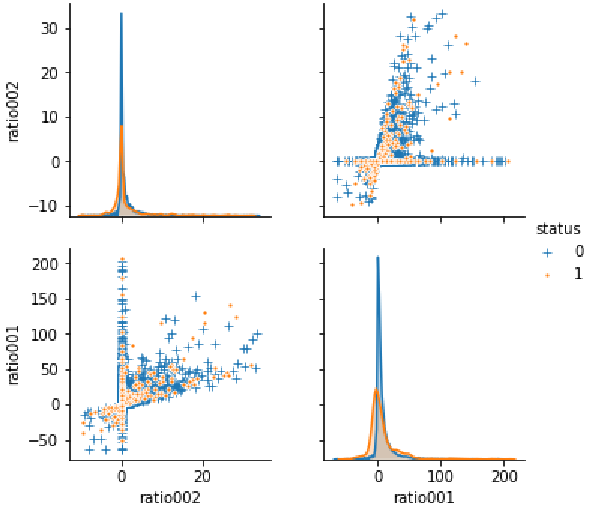
ii. P2P context & data: The good, the bad, and…

	Non-Defalult	Default	t-Values	p-Values
ratio005	1.243633	0.757088	13.798067	4.900817e-39
ratio011	0.047816	-0.133996	14.139802	1.431940e-38
ratio012	0.008752	-0.694699	13.210514	9.767516e-35
ratio004	1.597079	1.041546	12.834535	2.662133e-34
ratio029	0.084345	-0.118976	13.017858	9.449645e-34
ratio030	0.091696	-0.119036	12.802153	7.326660e-33
ratio003	1.487742	1.086767	11.702571	8.294173e-29
ratio008	26.217702	-2.334116	9.623006	1.271021e-20
ratio027	40.178063	6.955663	9.116989	7.749628e-19
DPO	67.347361	145.180723	-8.947930	6.163742e-18
turnover	3542.273904	1749.405622	8.260879	4.223880e-16
DSO	91.071215	133.317269	-5.269663	1.955716e-07
ratio019	0.211768	0.050361	5.145753	3.699050e-07
DIO	100.609313	142.471888	-2.003893	4.555395e-02
ratio006	7.929163	6.089699	1.423779	1.550459e-01
ratio017	1.380640	1.301807	1.412098	1.584418e-01
ratio018	1.341287	1.287108	0.948039	3.434978e-01
ratio002	1.248618	1.389016	-0.616914	5.375488e-01
ratio001	8.852383	9.148855	-0.243575	8.076496e-01

```
Total number of samples: 4514
Share of Defaulters: 11.03%
Number of Features: 19
```

RATIO005:	(Current assets - Current assets: stocks)/Current liabilities
RATIO011:	(Profit (loss) before tax + Interest paid)/Total assets
RATIO012:	P/L after tax/Shareholders Funds
RATIO004:	Current assets/Current liabilities
RATIO029:	EBITDA/Operating revenues





iii. AI, ML & profit optimization: The naïve estimator & the economic target function

Naïve estimator: Accept ALL loans!

Total # loans:	4514
Non-defaults:	4016 (88.97%)
Defaults:	498 (11.03%)
Predicted defaults:	0
Accepted # loans:	4514 (100%)
Accuracy:	88.97%
True-positive rate:	0

Confusion matrix:

	0	1
0	4016	0
1	498	0

Economic target function

All loans: $1

Recovery rate: 20%

Cost of defaults:
498 x $1 x (1-20%) = $398.40

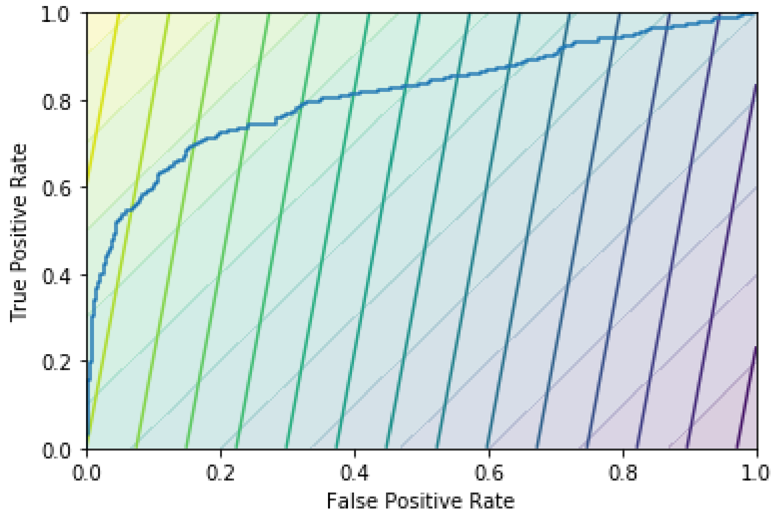
Necessary Risk premium:
Cost of defaults = $398.40
to be paid by non-defaults only:
$398.40 / (4016 x $1) = 9.92%

Spread: 500 bps

Income from Spread:
4016 x $1 x 500 bps = $200.80

iii. AI, ML & profit optimization: Logistic regression – In and out of sample

In sample:



True defaults: 299
Predicted defaults: 115
Accuracy: 91.36%
True-positive rate: 0.30

Threshold @ 0.5

Confusion matrix:

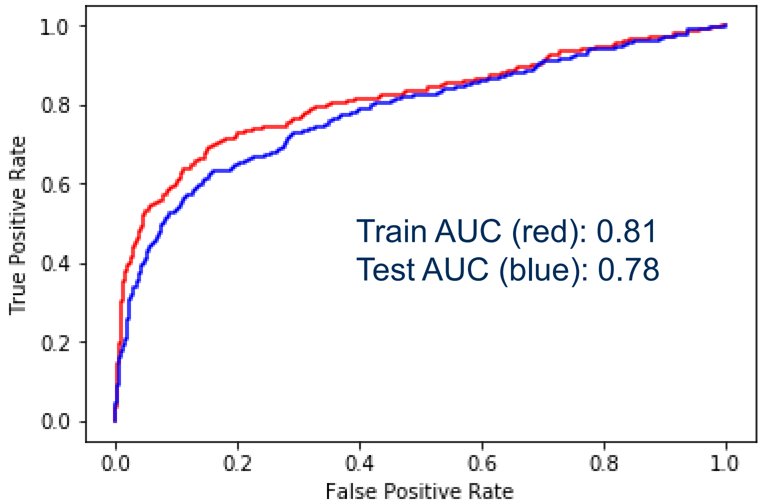
	0	1
0	2384	25
1	209	90

Out-of sample:

True defaults: 199
Predicted defaults: 53
Accuracy: 89.92%
True-positive rate: 0.18

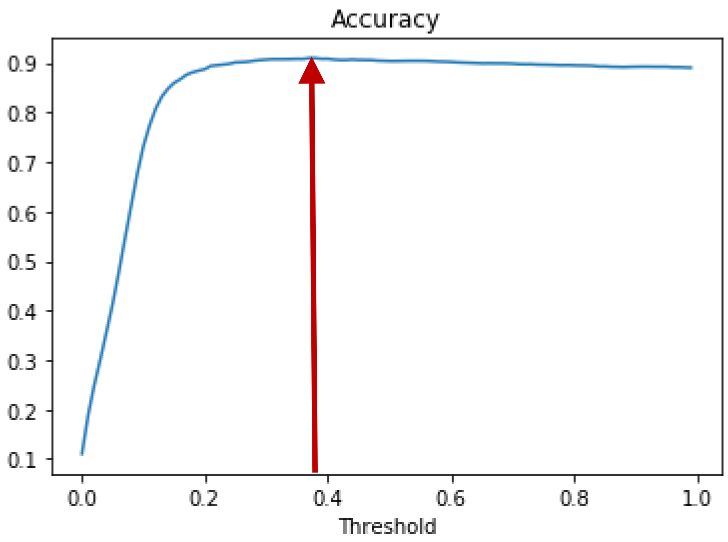
Confusion matrix:

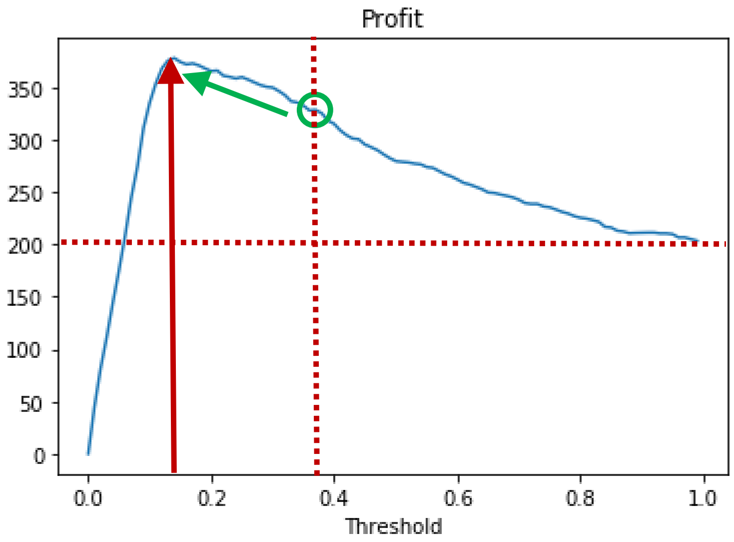
	0	1
0	1589	18
1	164	35



iii. AI, ML & profit optimization: Logistic regression – Accuracy vs. profit

Now: Vary **threshold** (“thresholding”) to find maximum accuracy or profit (**full data set**)





	Max. accuracy:	**Max. profit:**
True defaults:	498	498
Optimal threshold:	0.37	0.14
Predicted defaults:	260	**819**
Accuracy:	90.96%	84.78%
True-positive rate:	0.35	0.63

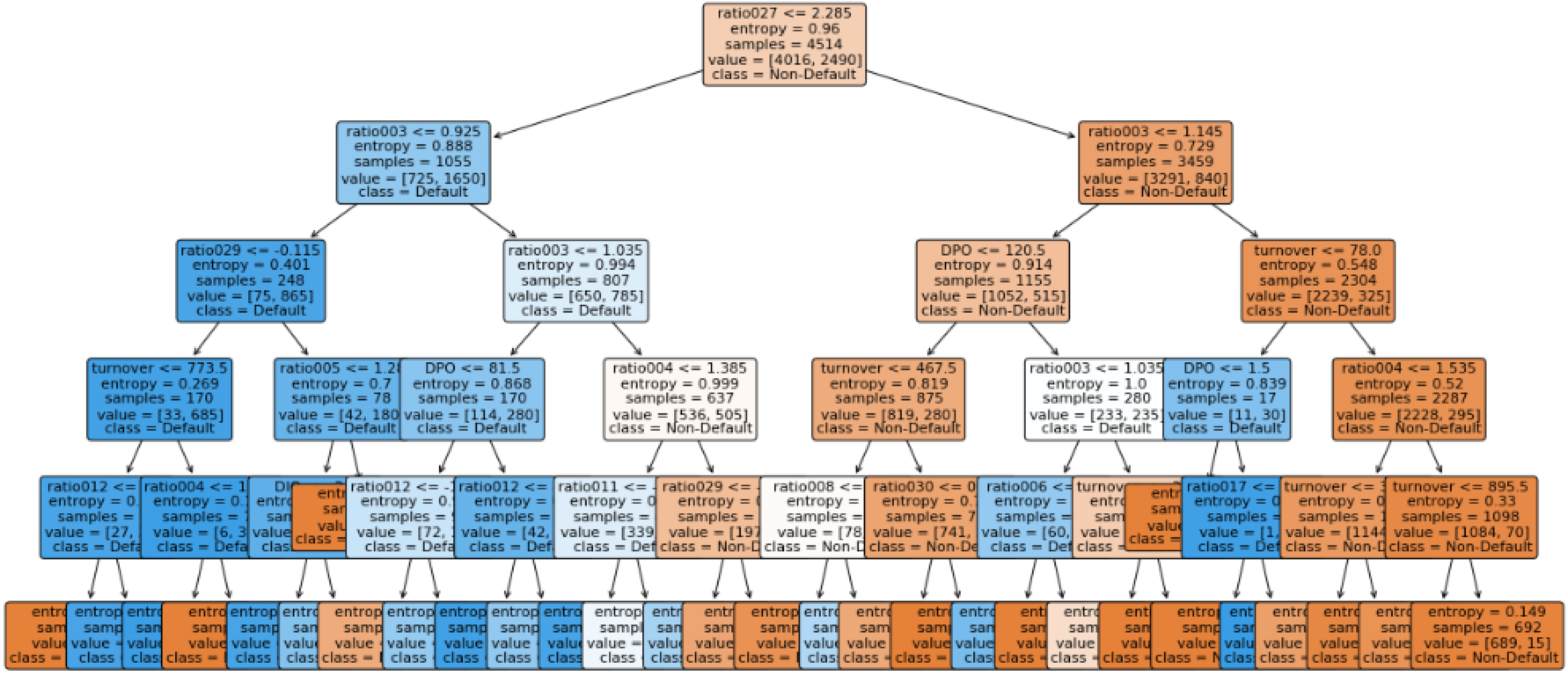
Confusion matrix:

	0	1
0	3931	85
1	323	175

	0	1
0	3512	504
1	183	315

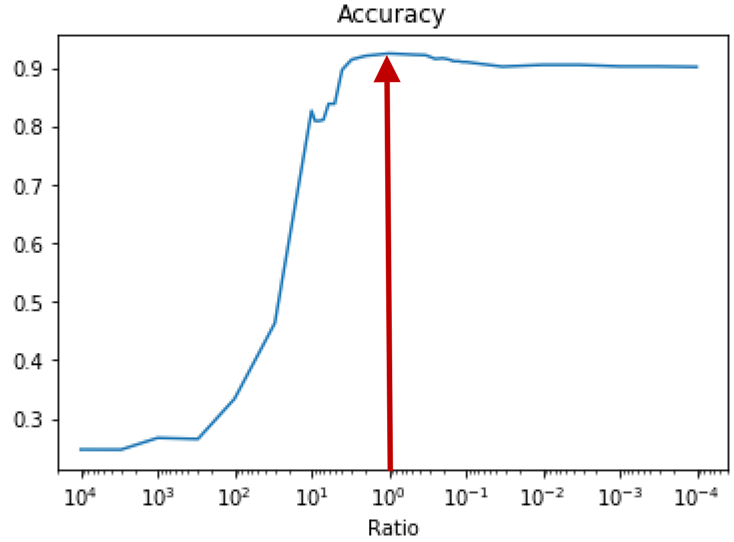
Max. profit @ 0.14 ≈ **$375** > profit @ 0.37 > $201

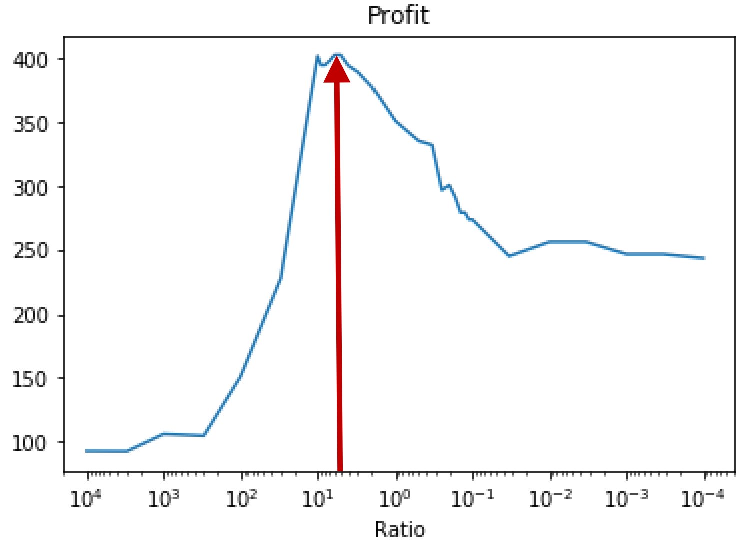
iv. XAI: Decision Trees – Explainable at a glance



iv. XAI: Decision Trees – Rebalancing: Accuracy vs. profit

Here: Vary **ratio of weight** “defaults / non-defaults” to find maximum accuracy or profit





	Max. accuracy:	**Max. profit:**
True defaults:	498	498
Optimal ratio:	1	**5**
Accuracy:	92.45%	83.83%

Accepted loans:	78.73% (3554)
Defaults in full data set:	11.03%
Defaults in accepted loans:	3.77%
Non-defaults in rejected loans:	62.08%

Max. profit @ 5 ≈ **$403** > profit @ 1 ≈ $351

Max. profit @ 5 ≈ **$403** > $375 (logistic regression)

iv. XAI: Decision Trees – Method generalizes nicely, too

Out-of-sample: Apply in-sample profit-optimization to test data set

Stats @ profit maximum	In sample:	Out-of sample:
True defaults:	299	199
Optimal ratio:	**8**	**8**
Accuracy:	81.94%	78.29%
Accepted loans:	74.00%	75.03%
Defaults in full train data set:	11.04%	11.02%
Defaults in accepted loans:	2.10%	5.17%
Non-defaults in rejected loans:	63.49%	71.40%

Why the fuss? – Out-of sample profit @

naïve estimator:	$ 80.32	
max. accuracy (ratio = 0.5, in sample):	$128.19	⇨ **59.60% more profit**
max. profit (ratio = 8 , in sample):	$135.73	⇨ **69.00% more profit**

v. Conclusio

- Credit risk is asymmetric and defaults are expensive: with P2P loan selection, this skews the cost of errors in prediction

- An economic target function is needed to correctly state the true optimization problem: Maximize profit, and not prediction accuracy

- ML techniques operating under the constraint of profit maximization yield a substantial pick-up in profitability while prediction accuracy may be inferior to even the naïve estimator

- XAI techniques, like Decision Trees, provide easy and intuitive explainability of the prediction process while being as accurate as their non-explainable peers. They, too, can be subjected to profit maximization constraints

Thank you very much

https://www.linkedin.com/in/thomas-gramespacher-cfa-69145693/

grat@zhaw.ch

www.linkedin.com/in/dr-jan-alexander-posth-5324b152

posh@zhaw.ch