Machine learning models for predicting customer decisions in motor claims settlements

ProService Finteco

Paweł Sujecki
ProService Finteco CEO

@SPARTUS

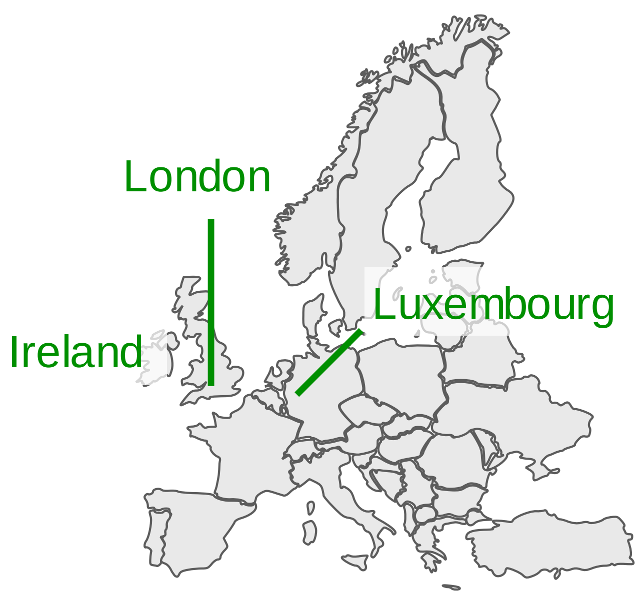
Adam Smółkowski
Aspartus CEO

ProService Finteco Group overview

Established 1994

674 EMPLOYEES

EUROPEAN REACH



100% Shares

Over 142 CLIENTS

Over 361 valued funds

OVER EUR15.3bn in administration

ProService Finteco Group – products and services overview

Proservice Finteco

- ✓ Leading Transfer Agent for investment founds in Poland, full business proces and operations outsourcing solutions

Aspartus

- ✓ Deployment of self-learning AI/ML architectures, delivery and maintenance of Insurance solutions including core and front-ends

make it right

- ✓ Test Automation, Robotic Process Automation, Quality Assurance and consulting, front services automation with chatbots and voicebots, cyber security audits and SOAR tools

Machine learning models in insurance

@SPARTUS

Risk assessment

- ✓ Although insurance tariffs need to be explainable, some insurance companies include parameters produced by machine learning models

Fraud detection

- ✓ Anomaly detection is widely used as a way to detect dishonest customers' behavior

Document classification, voice bots

- ✓ As a modern approach in area of customer contact

Our approach – The claim process

Why is the claim area so important?

- ✓ *A moment of truth* for every Insurance company customer
- ✓ Has a direct impact on financial state of every Insurer

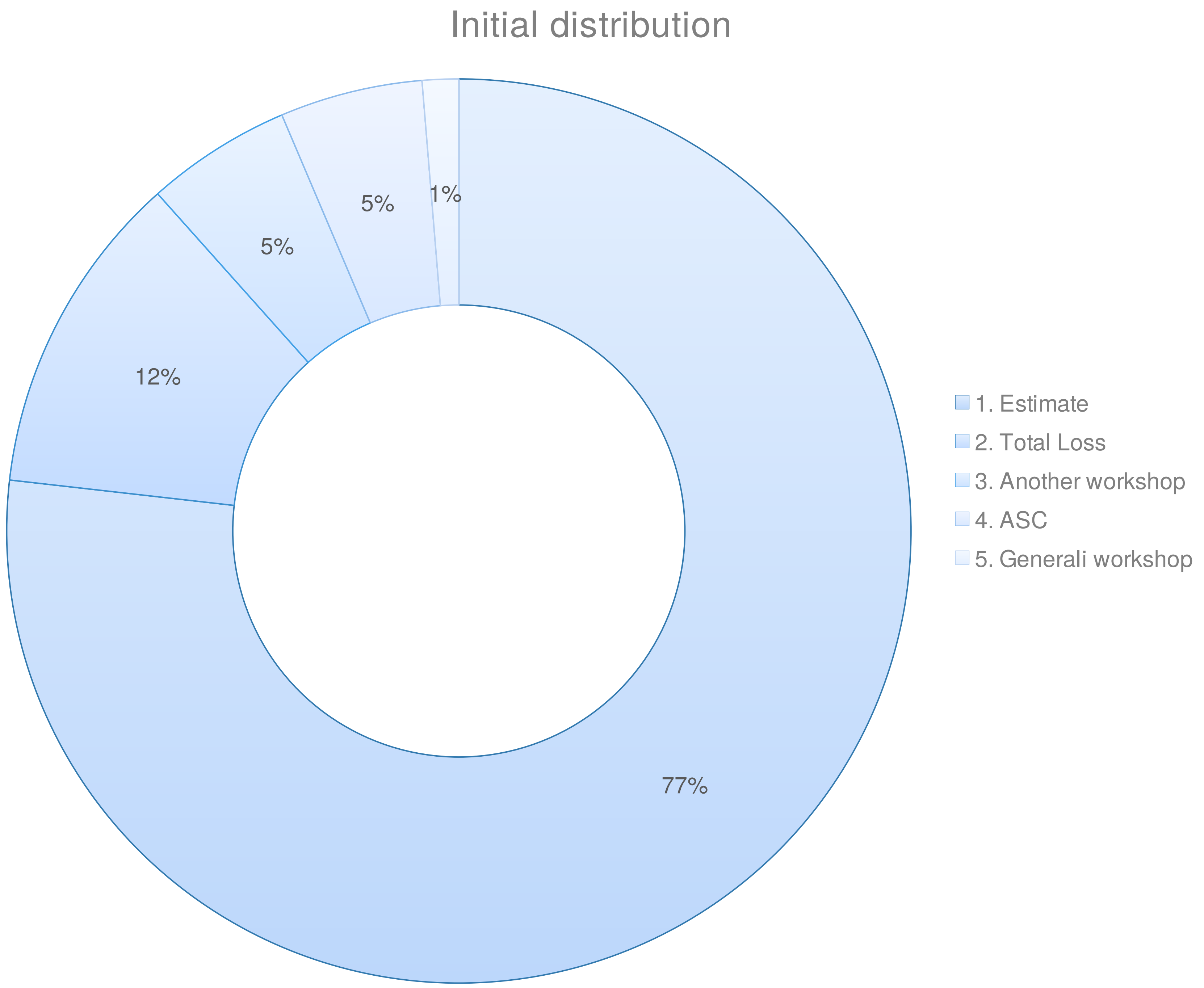
Different paths

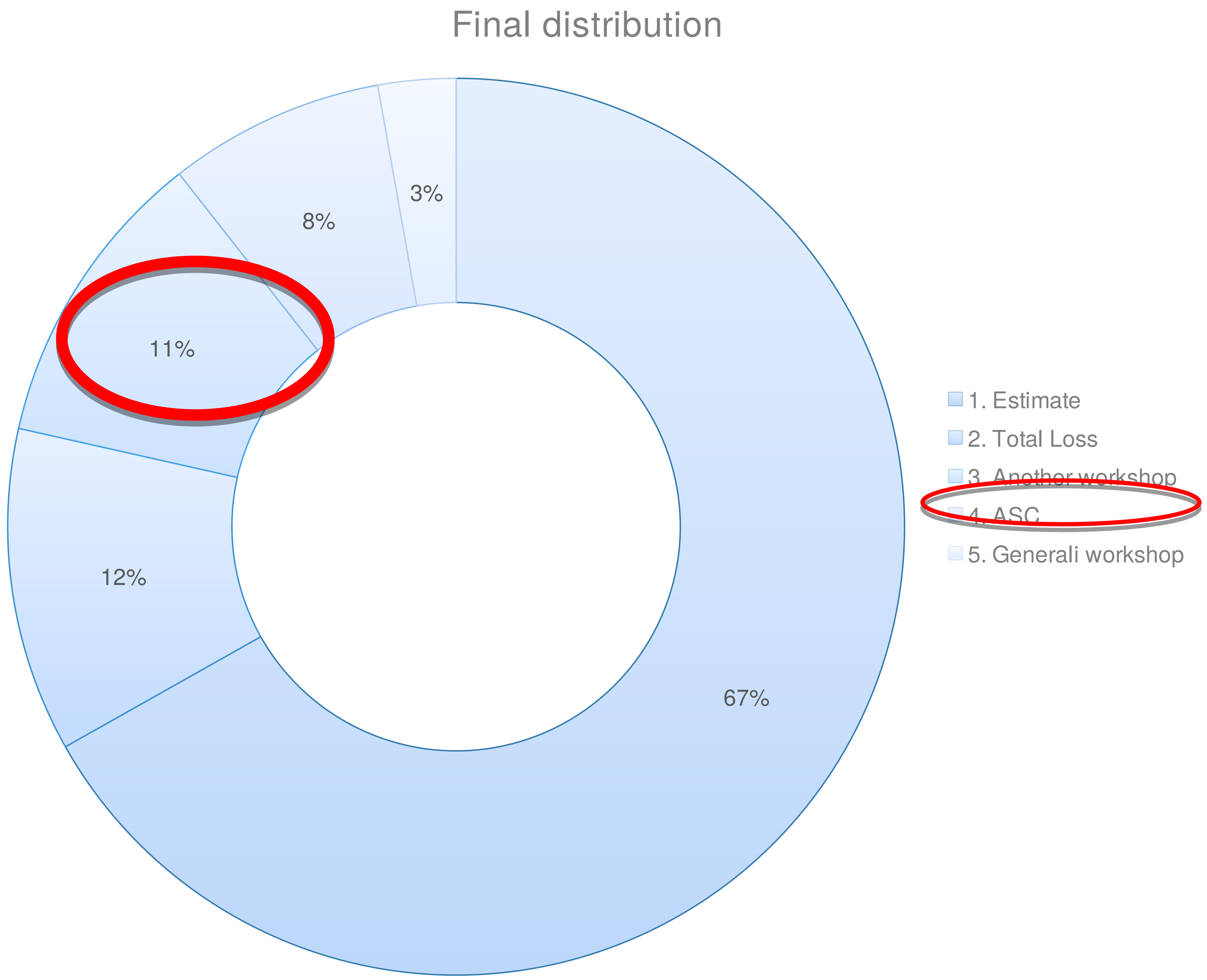
- ✓ the final result in the form of insurance payout can be achieved in various ways

Two hypotheses

- ✓ Cost estimate -> External workshop (A)
- ✓ External workshop -> Insurer’s repair network (B)

Business Case background – where is the money?





Data

- ✓ Training data set
  - ✓ Time span: 45 months, records: 335075, attributes: 120 + 28
  - ✓ Attributes for car/customer/claim from insurer’s systems plus external ones
  - ✓ For hypothesis A: 19683 positive records out of 257346 total
  - ✓ For hypothesis B: 46 positive records out of 17287 total

Final set

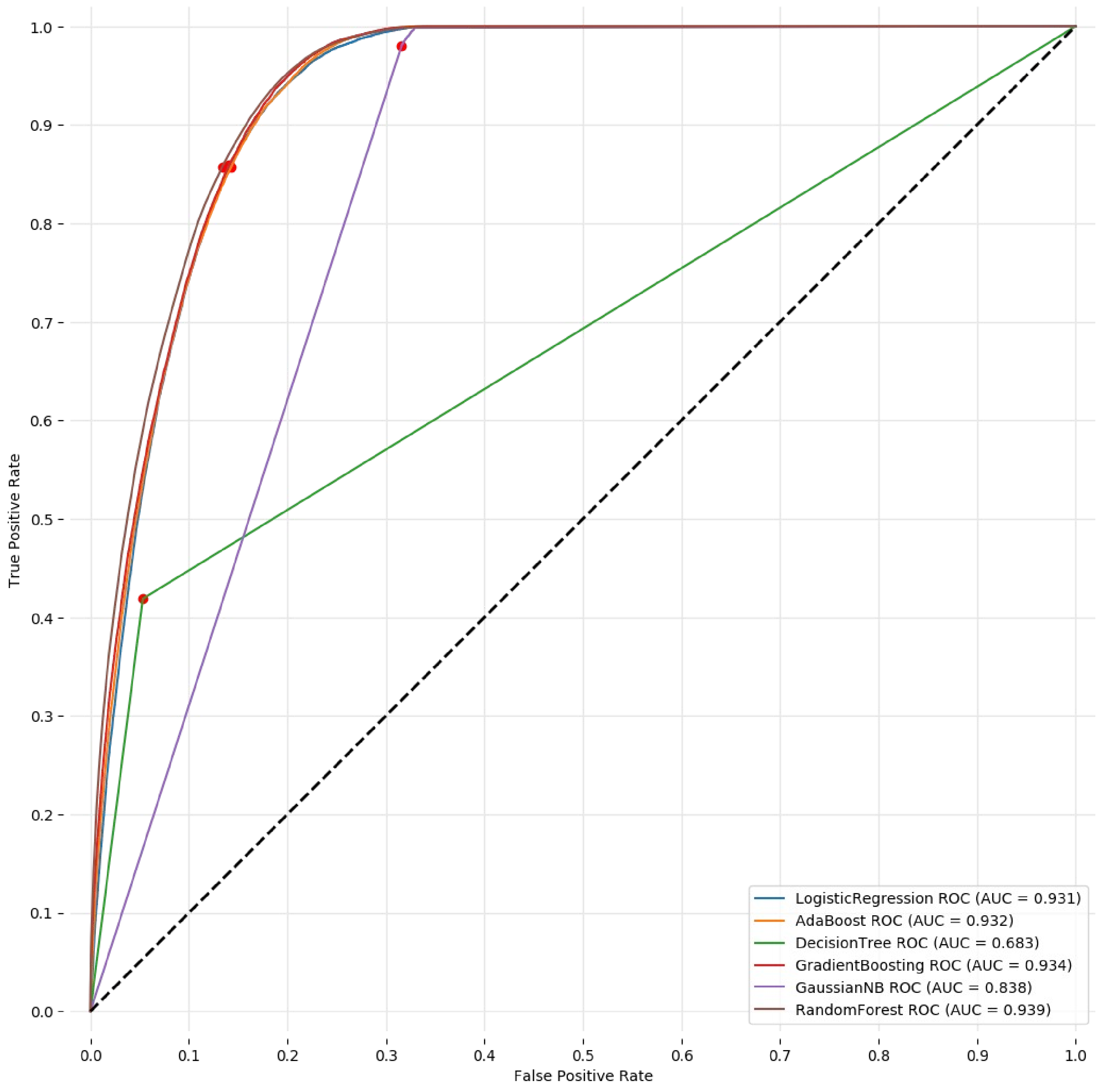
- ✓ 61 best ranked attributes evaluated with recursive feature elimination and cross-validation (RFECV) method

Results

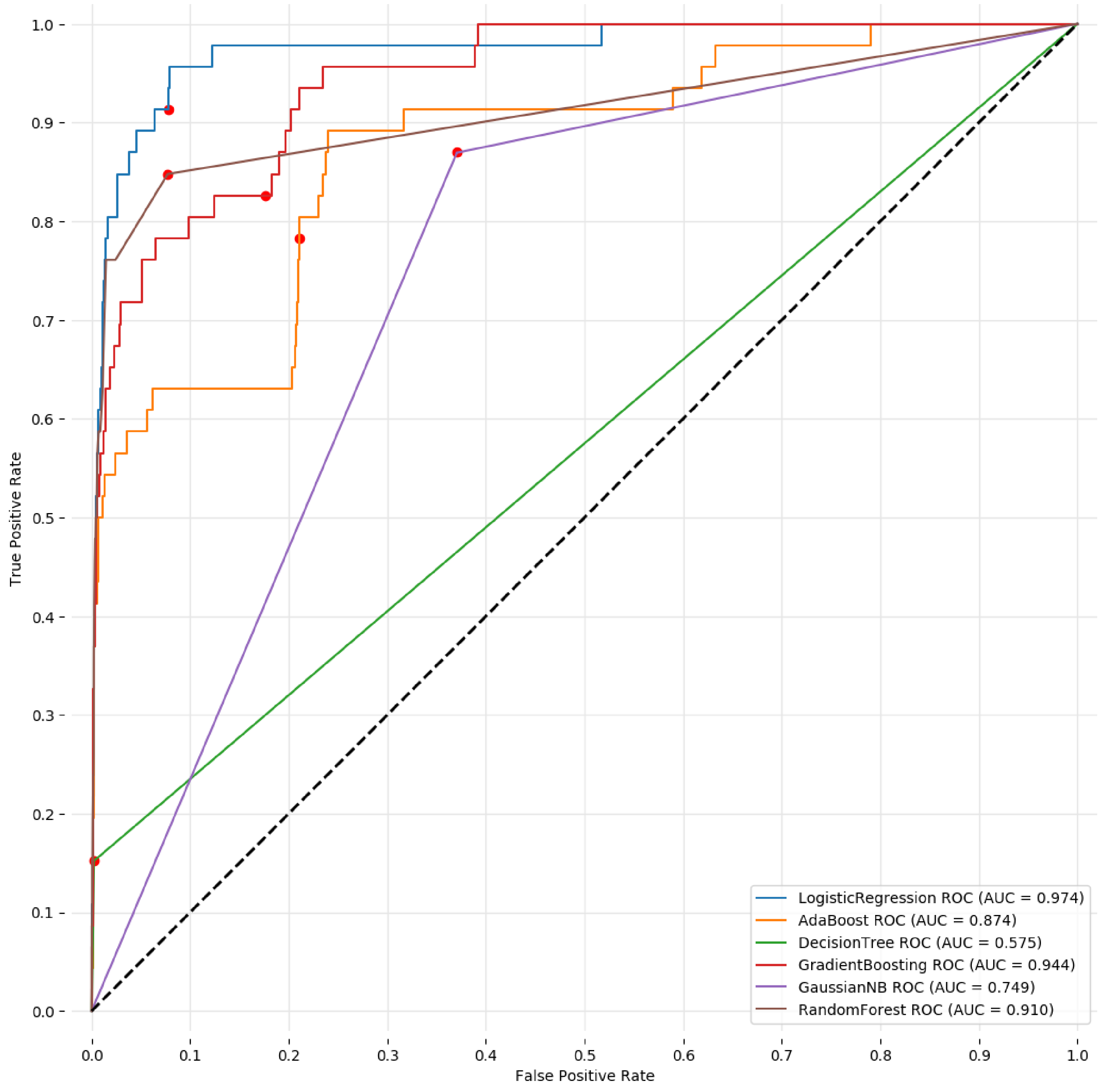
- ✓ Two set of models for hypotheses A and B
  - ✓ LogisticRegression, AdaBoost, DesicionTree, GradientBoosting, GaussianNB, RandomForest
  - ✓ 10-fold cross validation, the same ratio of examples of each class, each experiment was repeated 5 times and average values reported

	hypothesis	result [AUC ROC]
LogisticRegression	A	0.931
LogisticRegression	B	0.974
GradientBoosting	A	0.934
GradientBoosting	B	0.944
RandomForest	A	0.939
RandomForest	B	0.910

Results – ROC



Hyphotesis A



Hyphotesis B

Evaluation

Additional dataset not seen previously with 12966 records

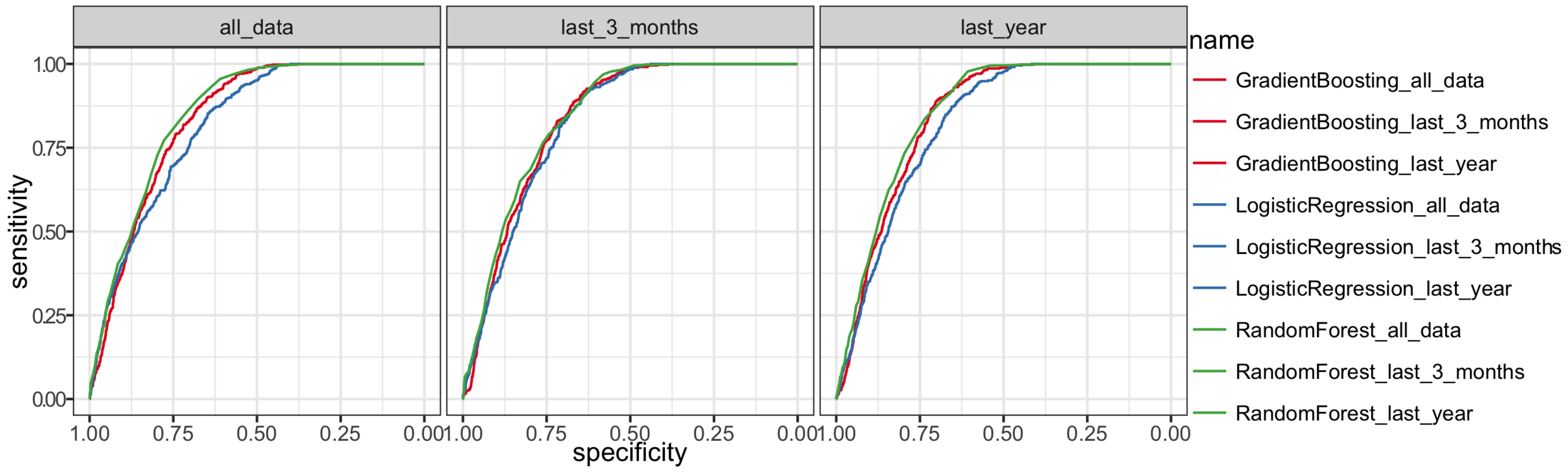
- ✓ Three models: Logistic Regression, Gradient Boosting, Random Forest
- ✓ Three training dataset subsets: (1) whole dataset, (2) data from last year, (3) data from last three months

Cross check

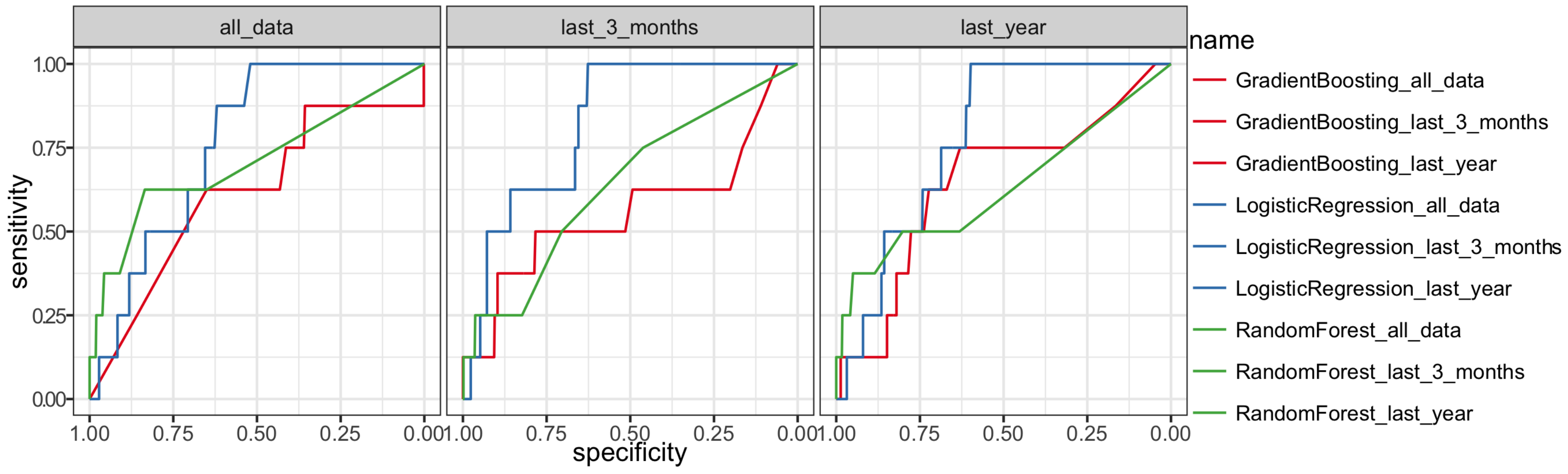
- ✓ Insurer sent us dataset without labels
- ✓ Predicted labels were send to insurer
- ✓ After predictions insurer sent ground truth for final evaluation

Evaluation results

	training subset	Hypothesis A	Hypothesis B
LogisticRegression	(1)	0.817	0.765
LogisticRegression	(2)	0.819	0.781
LogisticRegression	(3)	0.825	0.824
GradientBoosting	(1)	0.836	0.387
GradientBoosting	(2)	0.841	0.646
GradientBoosting	(3)	0.834	0.562
Random Forest	(1)	0.851	0.708
Random Forest	(2)	0.854	0.630
Random Forest	(3)	0.844	0.640



Hyphotesis A



Hyphotesis B

Conclusions

- ✓ Results achieved for main hypothesis (A) for all models are in range 0.817 - 0.841 ROC AUC which is far beyond customer expectation and allowed to go into pilot phase on production process.
- ✓ The goal of the pilot phase is to confirm that having knowledge about the customers’ preferred claim path allows management of the process in a cost effective manner

Thank you

Adam Smółkowski
CEO Aspartus
Adam.Smolkowski@aspartus.pl