POLITECNICO
MILANO 1863

DIPARTIMENTO DI
ELETTRONICA,
INFORMAZIONE
E BIOINGEGNERIA

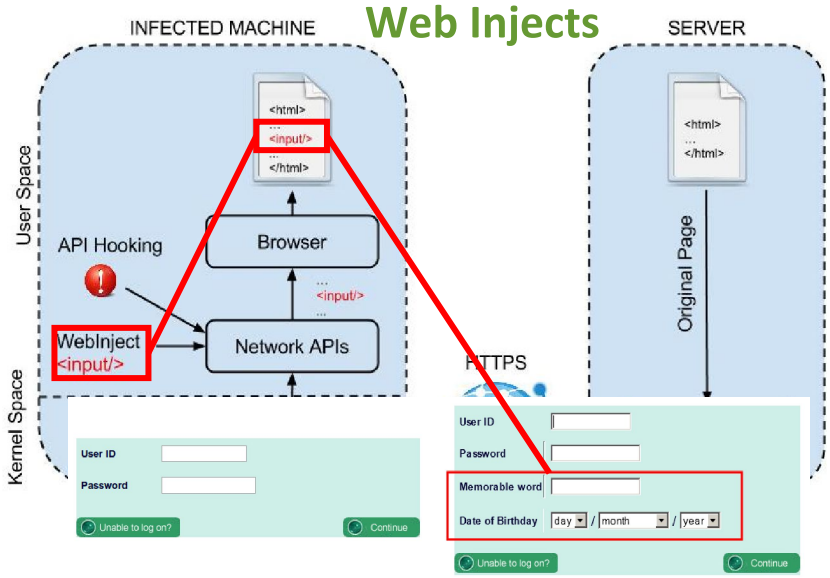
Financial Fraud Detection

Prof. Stefano Zanero

BANKSEALER

NECST laboratory

Threats and Anatomy of a Fraud



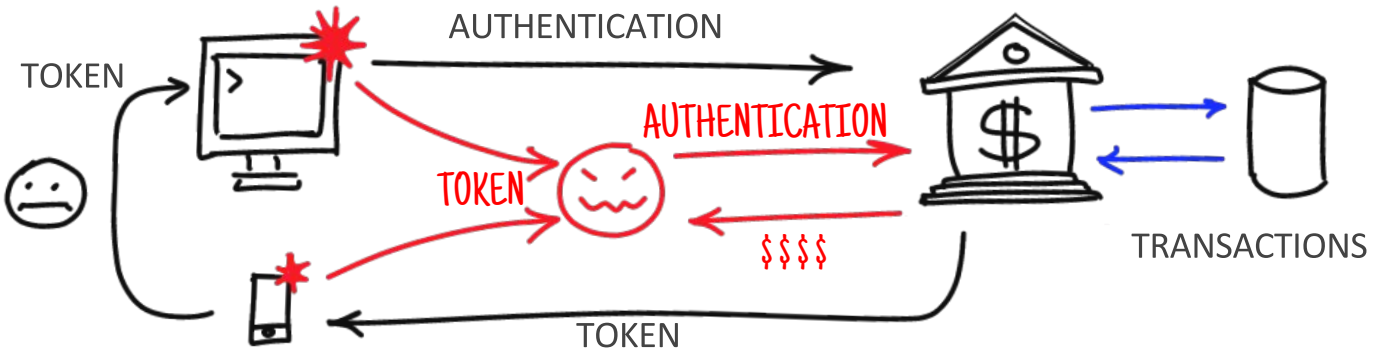
Traditional threats:

- Phishing
- Credentials Database Theft

Banking Trojans: Malware that aims to perform online financial frauds.

Man-In-The-Browser attack:

- steal credentials and private information
- Hijack browser session
- Infect Mobile Devices



Threats and Anatomy of a Fraud

Internet Banking Frauds Challenges

Internet banking frauds are difficult to analyze and detect:

- **Fraudulent behavior** is dynamic and dispersed in large and highly imbalanced datasets with different customer's profile
- Scarcity of available informations and data
- Most of the **existing approaches**:
  - Black box
  - Based on Synthetic data
  - Not adaptive baseline profiling

NECST laboratory

Goals

- Not focus on pure detection approach
- Support the analysis and the investigation of (novel) frauds and anomalies throught readable model and results.
- Decision support system able to model user behaviour and its evolution

Approach and System Description

Overview

User Profiles

Undertraining and Updating

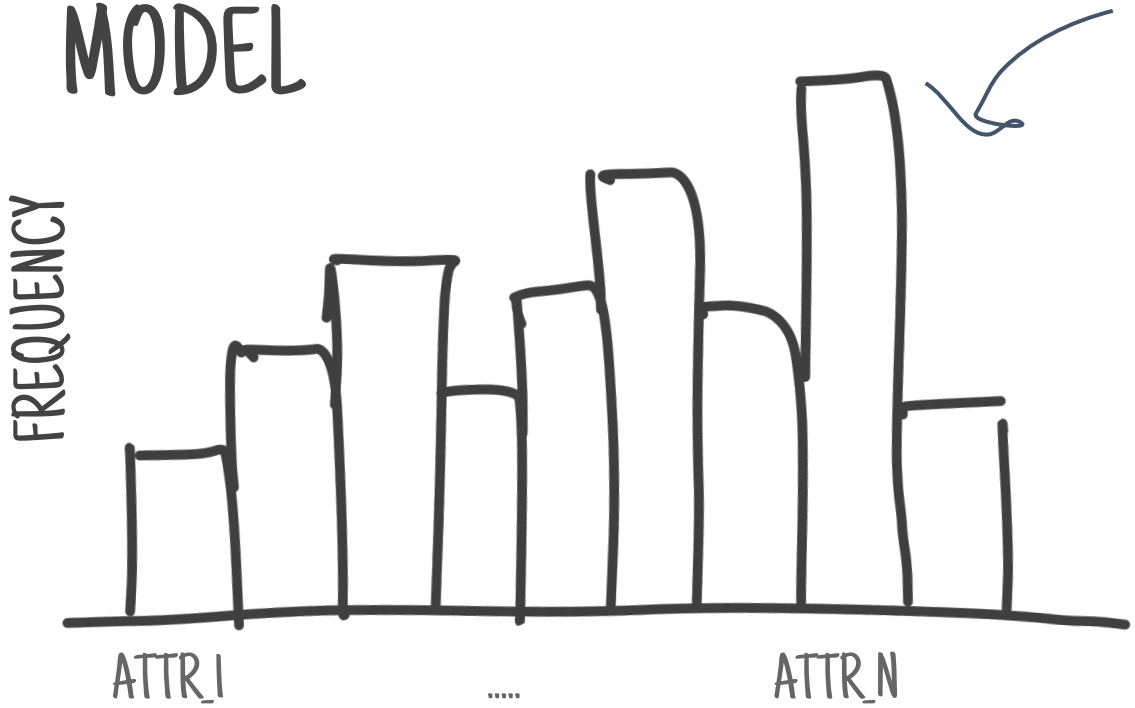
ATTRIBUTES

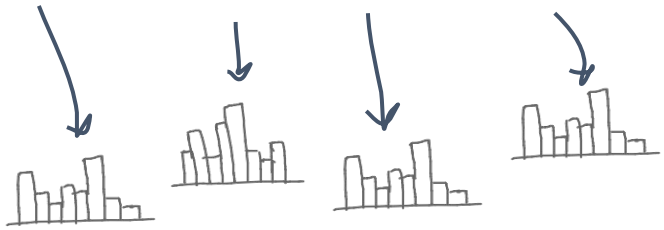
BANK TRANSFERS $$$ CC.IP IP IBAN NAZ_IBAN D:H:M:S

PHONE RECAHRGES $$$ CC.IP IP OP.TEL NUM.TEL D:H:M:S

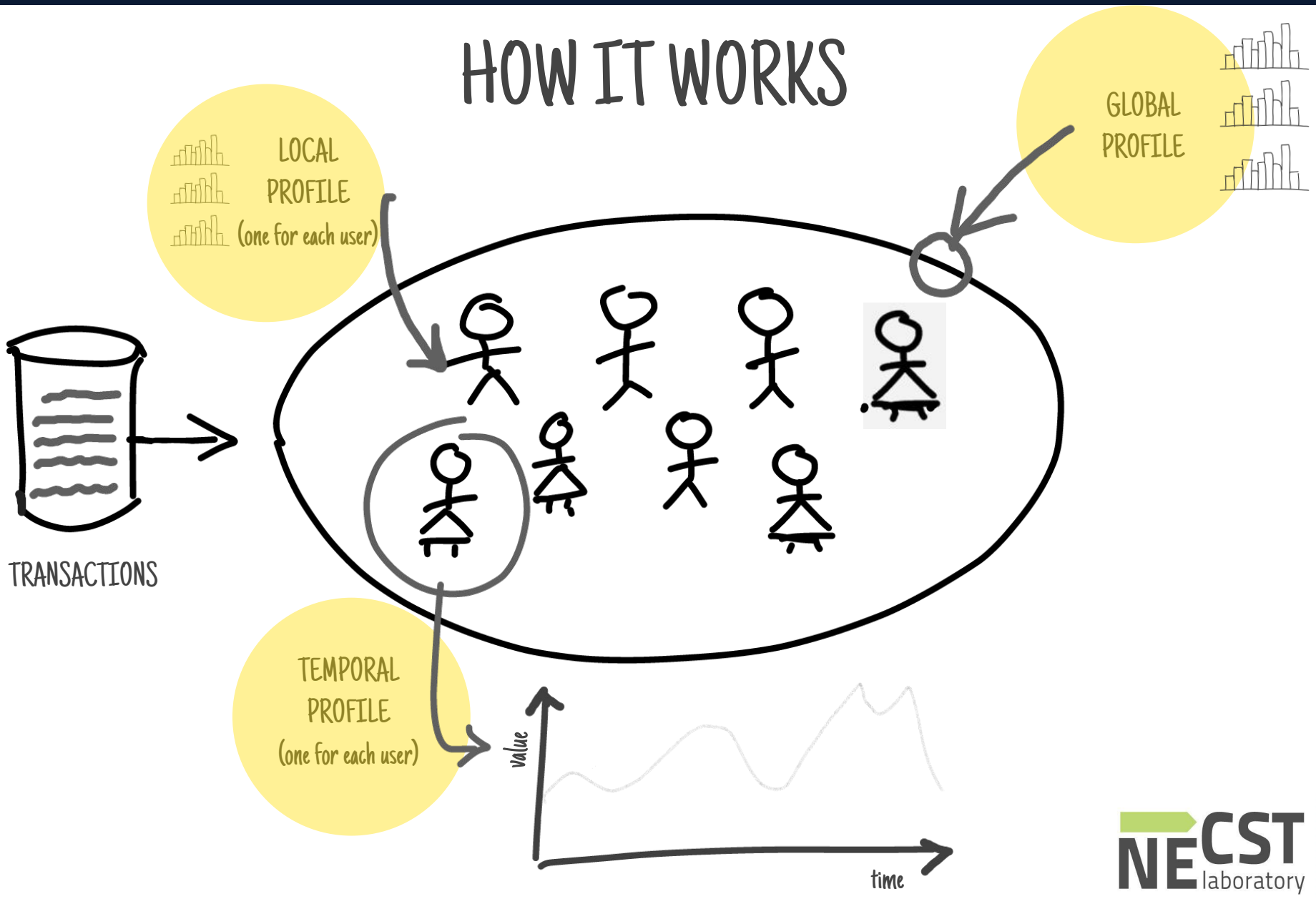
PREPAID CARDS $$$ CARD_TYPE CARD.NUMBER CC.IP D:H:M:S

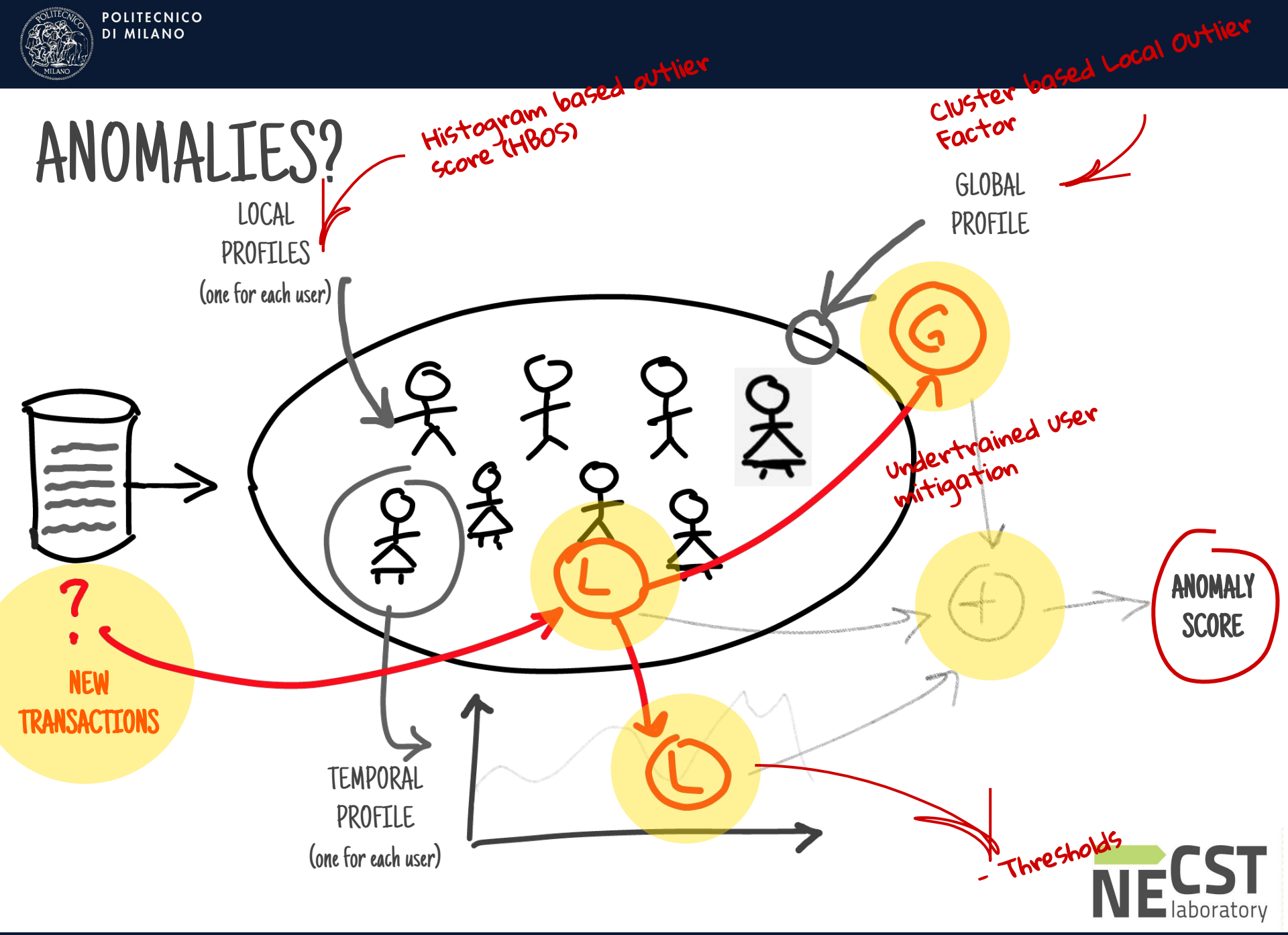
MODEL





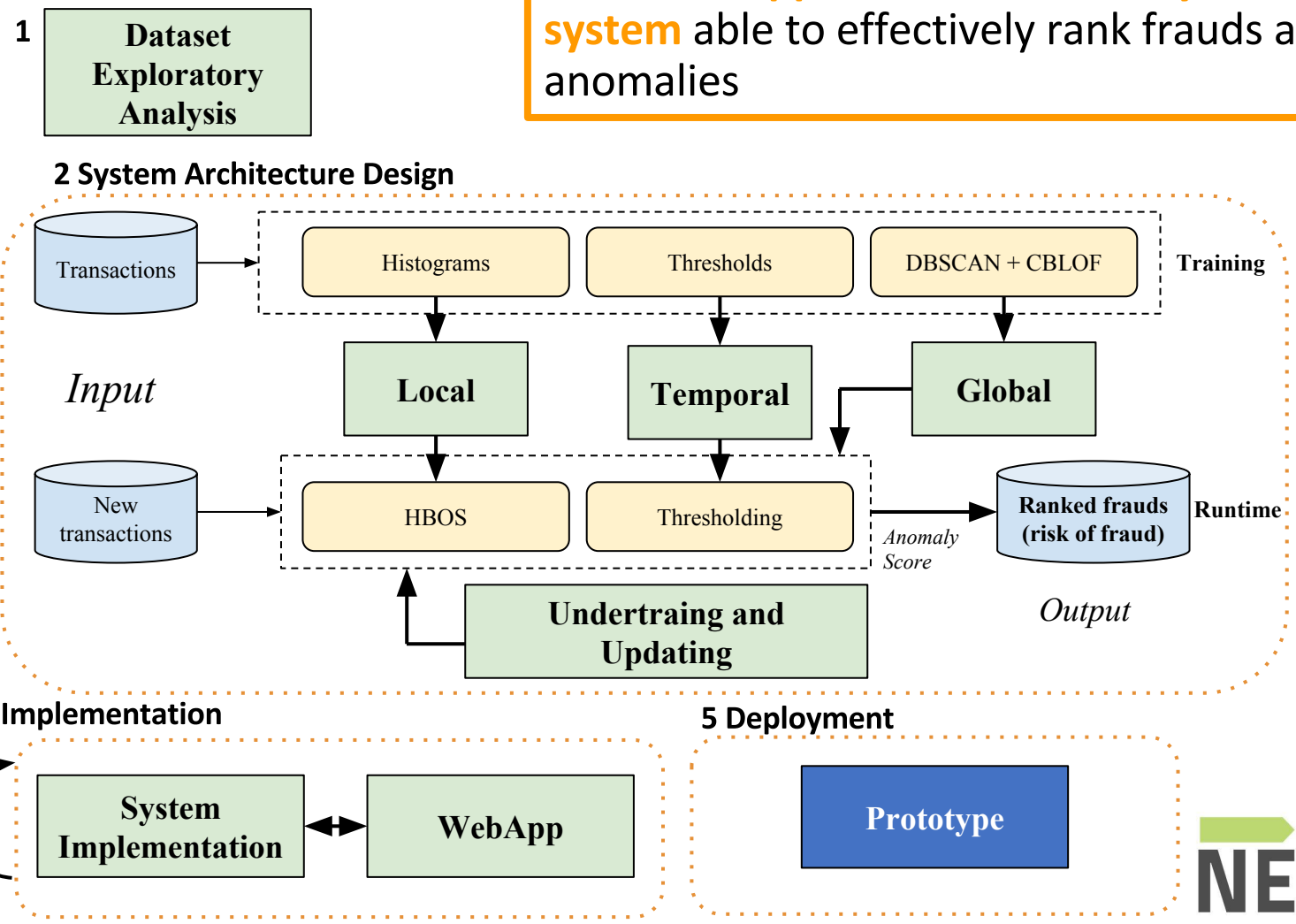
HOW IT WORKS


LOCAL
PROFILE
(one for each user)
GLOBAL
PROFILE
TRANSACTIONS
TEMPORAL
PROFILE
(one for each user)
value
time

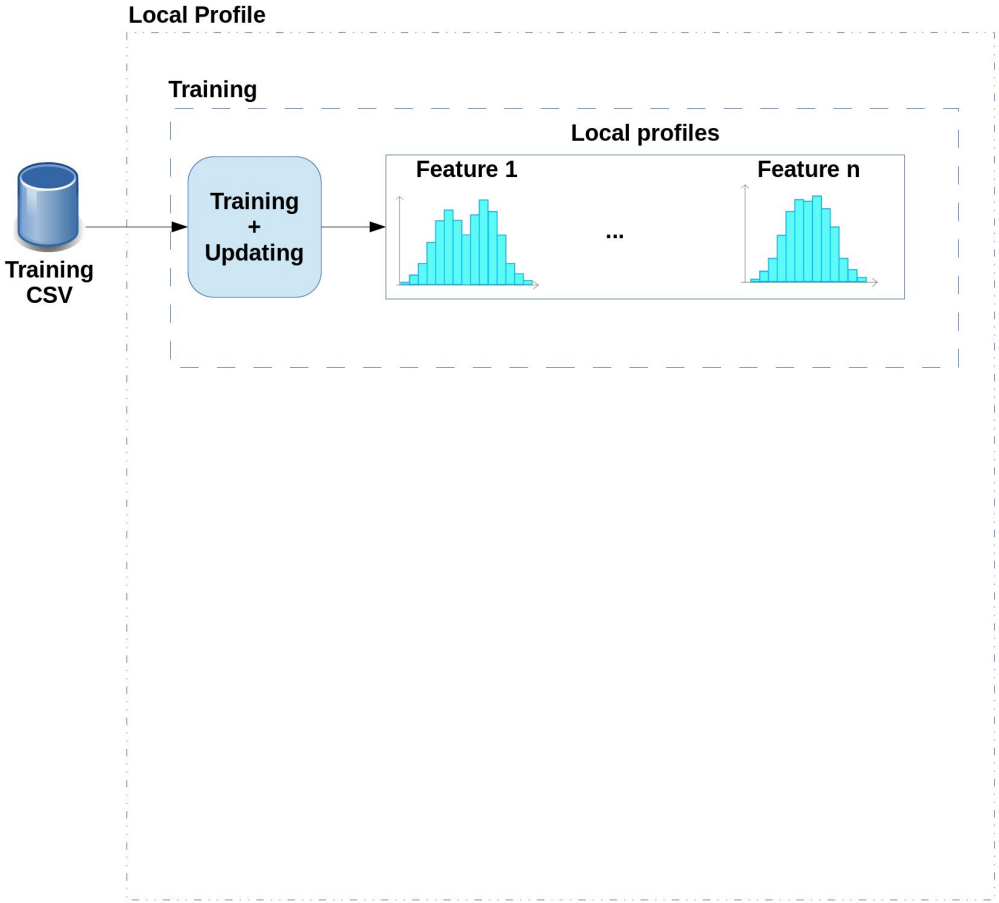

ANOMALIES?
Histogram based outlier score (HBOS)
Cluster based Local Outlier Factor
LOCAL PROFILES
(one for each user)
GLOBAL PROFILE
G
Undertrained user mitigation
?
NEW TRANSACTIONS
L
+
ANOMALY SCORE
TEMPORAL PROFILE
(one for each user)
L
- Thresholds

BankSealer

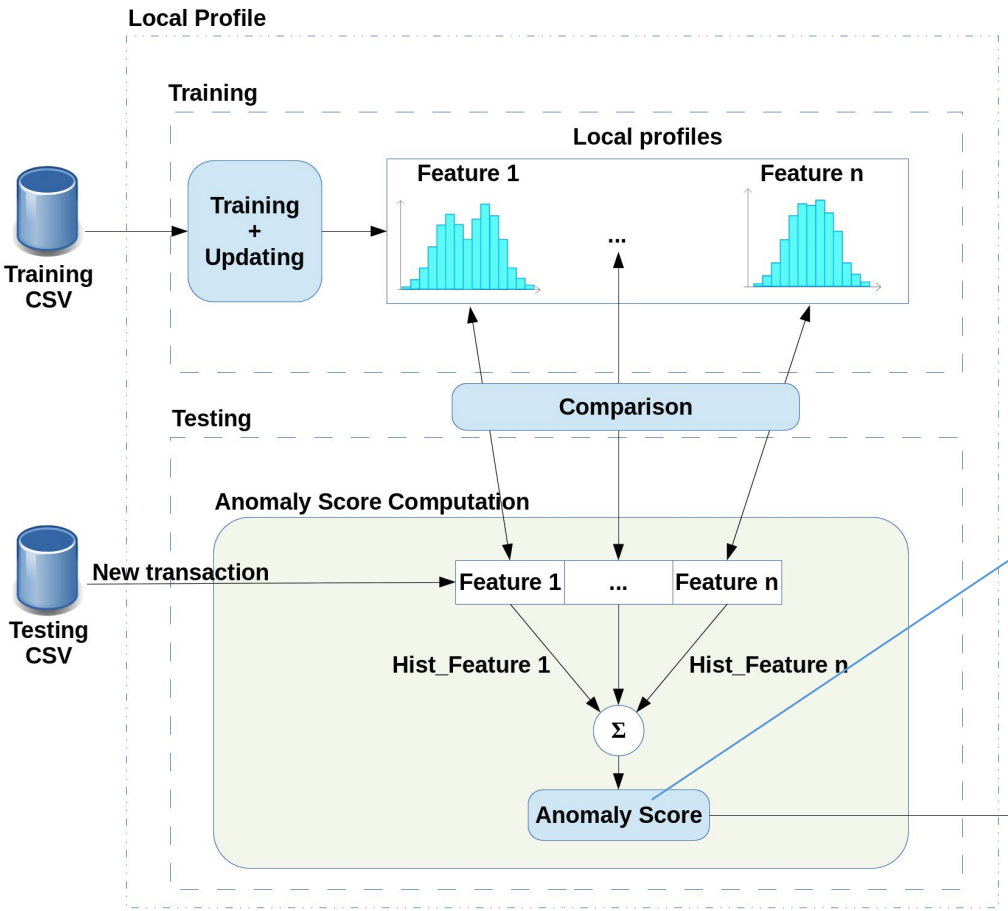
Decision-support and fraud-analysis system able to effectively rank frauds and anomalies



Local Profile



Local Profile



Histogram-Based Outlier Score

$$HBOS(t) = \sum_{0<i\leq d} w_i * \log \frac{1}{f(t_i)}; \qquad \sum_{0<i\leq d} w_i = 1$$

Feature Normalization

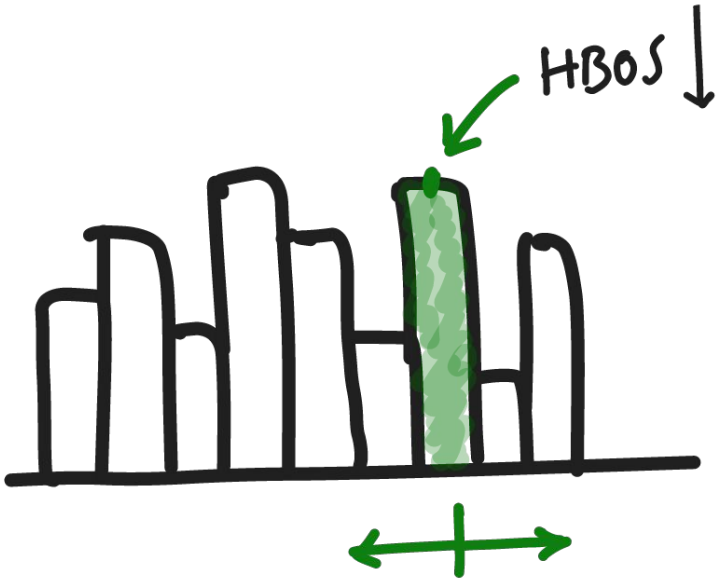
$$f(t_i) = \frac{hist_i(t_i)}{\max_{j \in feature_i} hist_i(j)}$$

Rare Values

$$\beta = \frac{k}{1-f}$$

$$RISK(t) = t_{transaction\_amount}{}^{w_{amount}} * HBOS(t)^{w_{HBOS}}$$

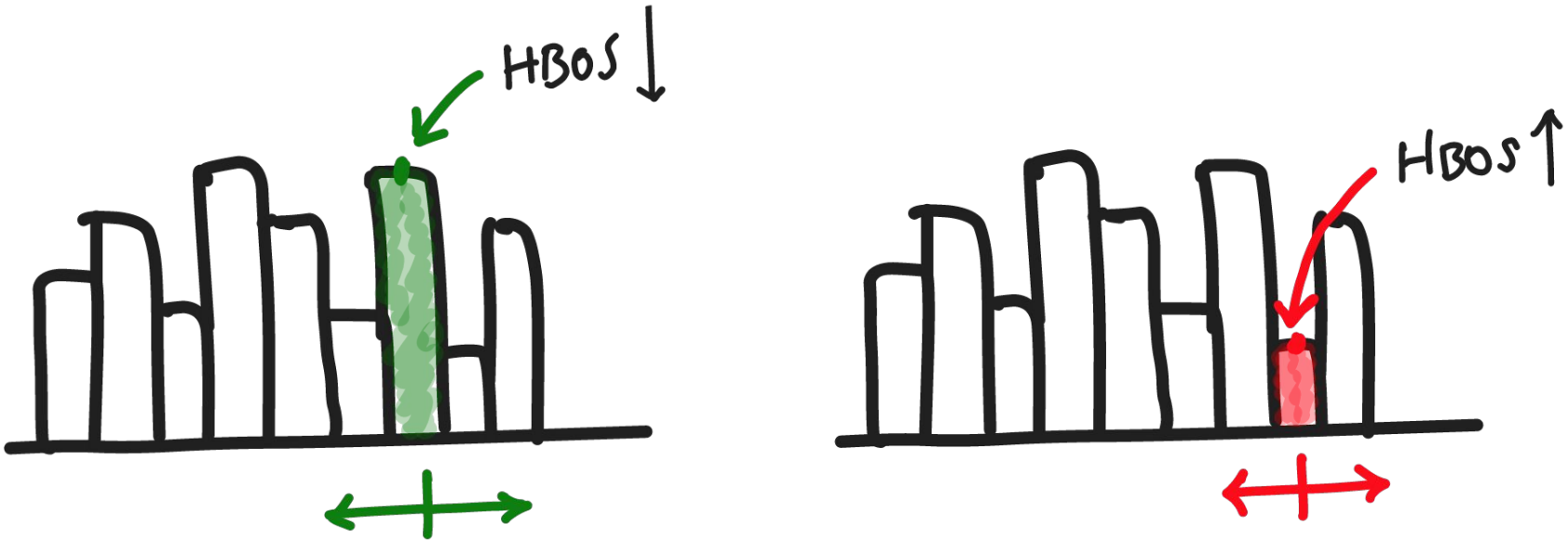
HBOS = Histogram Based Outlier Score



Each feature is weighted → Prioritizes relevant features

Discount factor → Forget old data & Model concept drifting

HBOS = Histogram Based Outlier Score



Each feature is weighted → Prioritizes relevant features

Discount factor → Forget old data & Model concept drifting

HBOS visualization

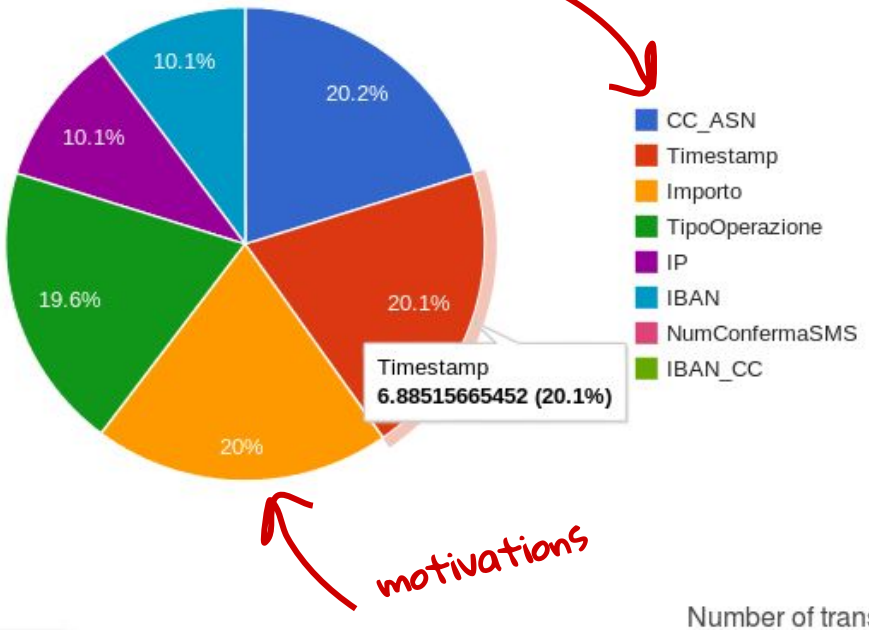
	index	IP	Timestamp	TipoOperazione	Importo	UserID	IBAN	Numero Conferma SMS	IBAN_CC	CC_ASN	HBOS locale	Undertrained	New User
1	92580	3d64e9f4a188aa034659d1409f90456a	08/feb/2013 21:06:30	Giroconto	20000	dcfc15d4d65e05ebafde6ac9383062aa	e6c2a617f55090de28a24c67dfbedf40	✓	IT	IT,8612	29.402365073	✓	✗
2	91133	99ca402ce2299ecd72e2ebe269b5d35f	06/feb/2013	Bonifici per detrazione	9900	4a4ee6e2ac1b17e20958ad7a8221c1b4	11e6c83f92f065b9776	✓	IT	IT,3269	27.0032775111	✓	✗
3		3c9dad3b2601c929e2cece7	✓		IT	IT,30722	24.5660880611	✓		✗	5475914	✗	✗
4											5765096	✗	✗
5	101355	86419f50fbda2742c1dba87cd34294	28/feb/2013 17:46:46	Giroconto	12000	49zee7251c425c36c82cd3241c563f79	484fe271f1804b3e4291a537bb65279a	✓	IT	IT,44957	25.0995700682	✓	✗
6	92502	dd5d85da0532104875e18e4e32bc152c	08/feb/2013 16:11:24	Bonifici per detrazione fiscale	3863.29	a6f088c1fae1085def1308e532082cb7	73b5047423c9dad3b2601c929e2cece7	✓	IT	IT,30722	24.5660880611	✓	✗
7	99074	cd002daddde353900cc24e4ffc3b235c	21/feb/2013 18:40:02	Bonifici per detrazione fiscale	5643	a7b7a36b2769a1be86d1a544b67007a9	2626bfbc3376dababb639201c9b8ff67	✓	IT	IT,21056	24.4493640116	✓	✗
8	99827	0bdda3afaf28f049483d89d53f021c11	25/feb/2013 09:36:12	Bonifici Italia e SEPA	31000	be2b61118c081429cfbbc0c3d948743b	831687c224f781f106604f984e14f414	✓	IT	IT,12874	24.4175445119	✓	✗
9	89586	2aeddb8850ae946914285eb3bcd28d55	04/feb/2013 16:42:53	Giroconto	10000	41efb45d969e9511b7df6504840cc572	40c200429a2c2a4c7268b3300681e5e3	✓	IT	IT,12874	23.6879134642	✓	✗
10	101627	70c765c7265d92f96a05d91eebb4eb64	28/feb/2013 19:04:37	Bonifici Italia e SEPA	6529.6	8b7ed02e24a297a7ad7b91d28a5b35e1	3ada9624925ed42838bd4b8fab9eae81	✗	IT	IT,50809	23.6370584204	✓	✗
11	98401	d00d1939b4f71eaa199a57fff9cf0c19	20/feb/2013 14:59:10	Bonifici Italia e SEPA	50000	9bc3d0e6065284891a42ce6f9d828c38	65ecb9d1169b23049ec018d31c27af0a	✗	IT	IT,3269	23.5882551789	✓	✗
12	95342	2c2c6f325c547ee1fb0efc01475bc7d6	14/feb/2013 09:17:53	Bonifici Italia e SEPA	50000	f2a7341750c1cc6dc8bea45185a7fe26	60414014d030aa24b4cef90c32fac61f	✓	PT	IT,16232	23.5439265229	✓	✓
13	92842	ba3664bb7ebbf9e8bf4ac0664d65e239	10/feb/2013 19:21:11	Bonifici Italia e SEPA	20000	2a17ed71d9e2c82f39e174e424bf7eb9	e9987193889c72a6dcb94bbd47e35699	✗	IT	IT,3269	23.5233646465	✓	✗
14	92551	d7ab9d7839eb60ff6e606c496e1c848a	08/feb/2013 19:23:46	Bonifici Italia e SEPA	25266.8	435b8226966d2fb40d52bafd6aaa8a93	92a91621f34b668401e8c26050f4e0c6	✗	IT	IT,3269	23.4602697196	✓	✗
15	97221	6da1465327246224216c1c929c339c6f	18/feb/2013 15:09:11	Bonifici Italia e SEPA	50000	f2a7341750c1cc6dc8bea45185a7fe26	60414014d030aa24b4cef90c32fac61f	✓	PT	IT,16232	23.2738702047	✓	✓

NECST laboratory

HBOS visualization

Local profile details

User feature contributions



Number of transactions: 3
Final HBOS: 25.8683920672
Show user profile

Close

	Timestamp	TipoOper						IBAN_CC	CC_ASN
c11	25/feb/2013 09:36:12	Bonifici Ita						IT	IT,12874
456a	08/feb/2013 21:06:30	Giroc						IT	IT,8612
051d	06/feb/2013 20:07:26	Giroc						IT	IT,1267
152c	08/feb/2013 16:11:24	Bonifici per fisc						IT	IT,30722
9470	28/feb/2013 17:46:46	Giroc						IT	IT,44957
7c82	15/feb/2013 14:28:10	Bonifici per fisc						IT	IT,12874
348a	08/feb/2013 19:23:46	Bonifici Ita						IT	IT,3269
9cfb	13/feb/2013 10:33:06	Bonifici Ita						IT	IT,35612
9cfb	13/feb/2013 10:51:22	Bonifici Ita						IT	IT,35612
d35f	06/feb/2013 19:28:26	Bonifici per fisc						IT	IT,3269
e239	10/feb/2013 19:21:11	Bonifici Ita						IT	IT,3269
b580	28/feb/2013 18:16:52	Bonifici Ita						IT	IT,12874
05e3	12/feb/2013 11:24:04	Bonifici Ita						IT	IT,8612
235c	21/feb/2013 18:40:02	Bonifici per fiscale	5643	a7b7a36b2769a1be86d1b544b67007a9	2626bfbc3376dabab639201c9b8ff67	✓		IT	IT,21056
7d6	14/feb/2013 09:17:53	Bonifici Italia e SEPA	50000	f2a7341750c1cc6dc8bea45185a7fe26	60414014d030aa24b4cef90c32fac61f	✓	PT		IT,16232

Dataset

	# Transactions	# Users
Bank Transfer	371,137	47,650
Prepaid phone	54,141	16,093
Debit cards	34,986	8,415

- Three months
- Skewed and unbalanced distribution of the attribute values
- High cardinality
- Prevalence of users perform a low number of transactions

Global Profile

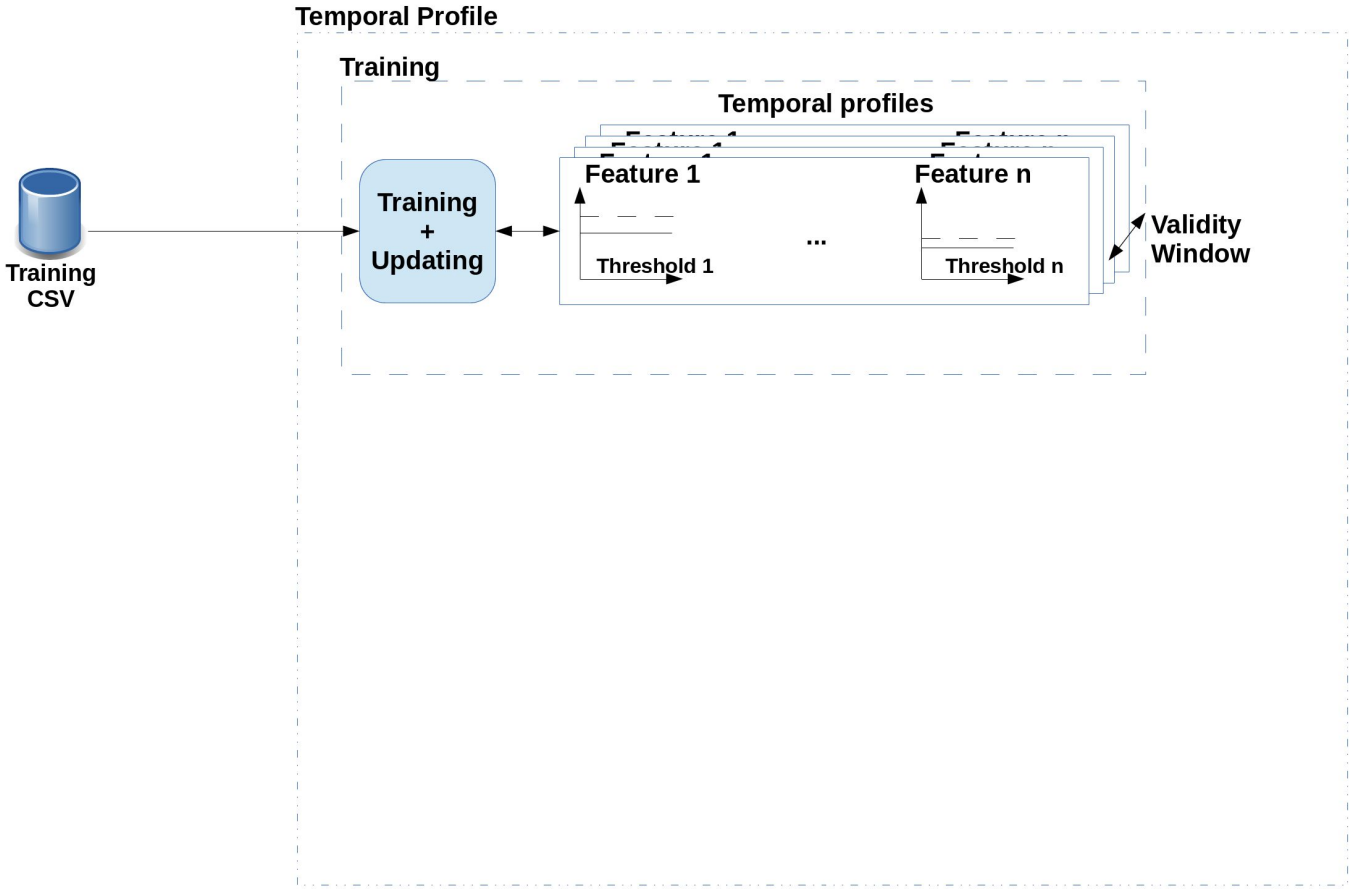
Two phases:

1. **Clustering**
   a. **Algorithm:** incremental DBSCAN with decreasing epsilon
   b. **Distance Metrics:** Mahalanobis
2. **Unweighted CLUSTER BASED LOCAL OUTLIER FACTOR:** anomaly score based on the distinction between:
   A. LARGE Clusters LC
   B. SMALL Clusters SC

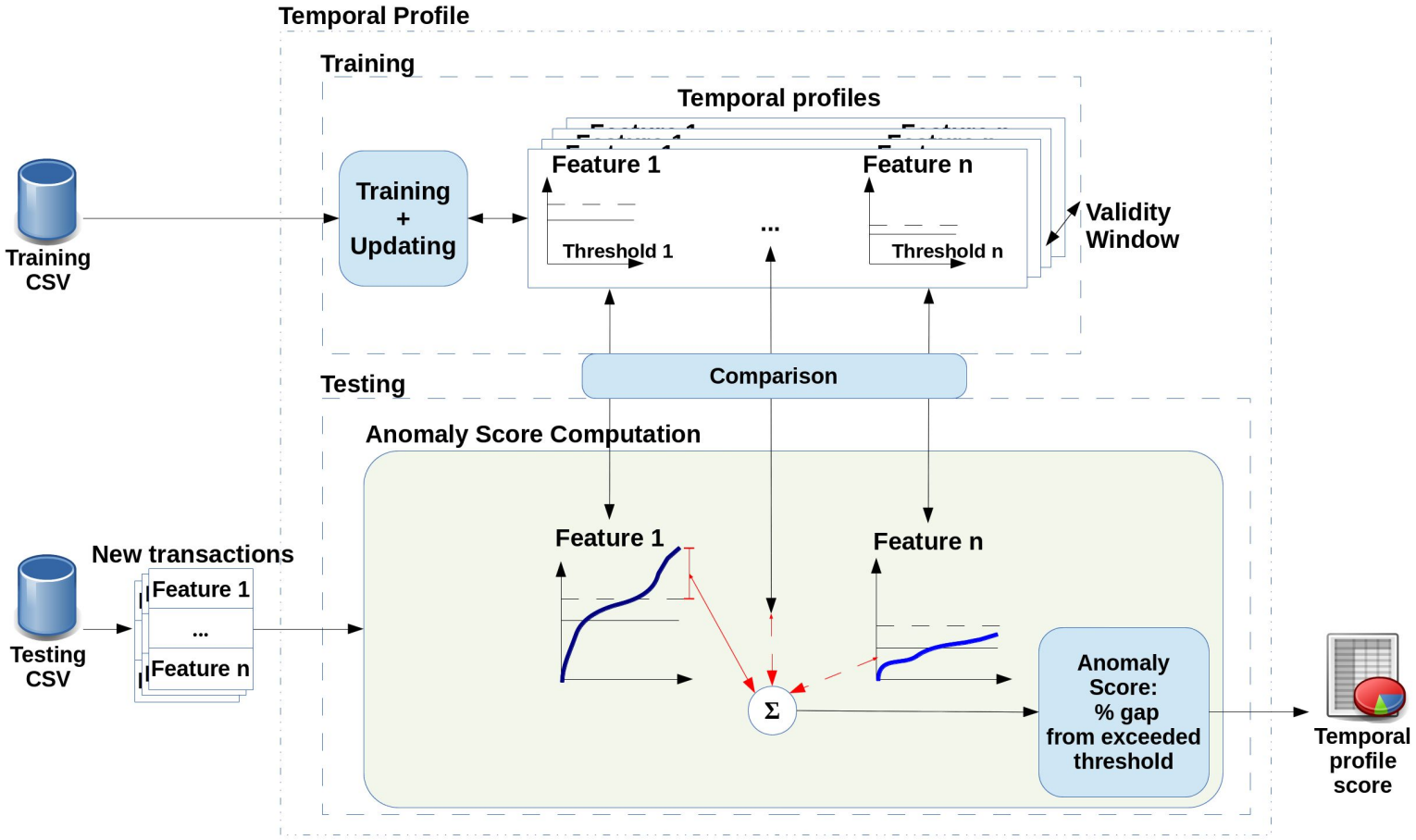
CBLOF = Min distance of a point from the centroid of the nearest LC

NECST laboratory

Temporal Profile



Temporal Profile



Testing and Evaluation

Time and Resource Performances

Fraud Scenarios

Evaluation Approach and Metrics

Overall Results

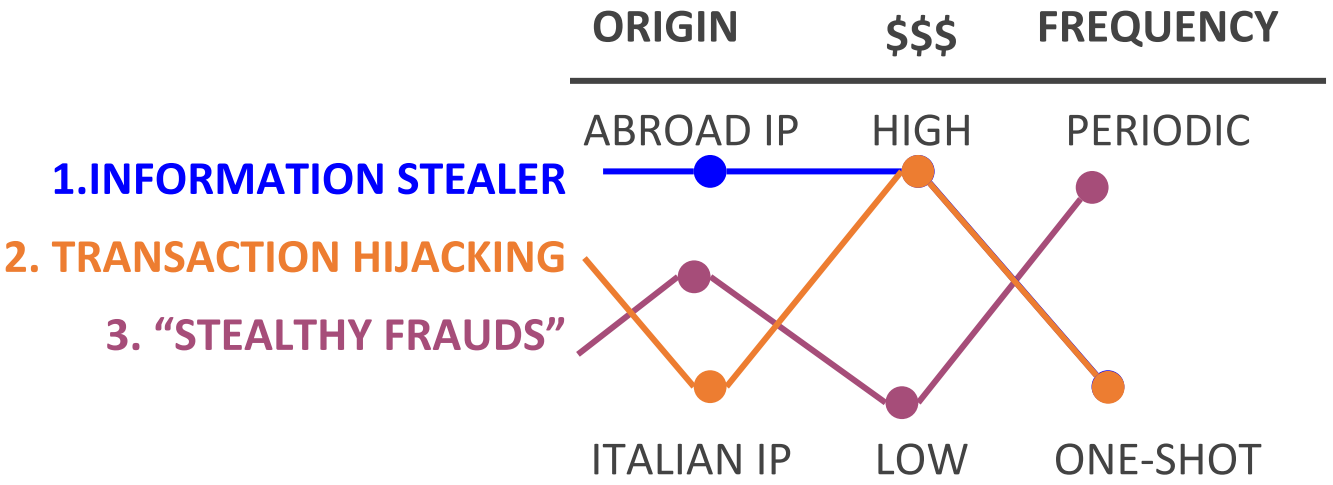
Time Performances: Testing

Domain	Test	Time of runtime execution
Bank Transfers	1 day filtered	1 min.
	1 day unfiltered	4 min.
	1 month filtered	6 min.
	1 month unfiltered	93 min.
Phone Recharges	1 day filtered	18 sec.
	1 day unfiltered	25 sec.
	1 month filtered	30 sec.
	1 month unfiltered	2.5 min.
Prepaid Cards	1 day filtered	7 sec.
	1 day unfiltered	10 sec.
	1 month filtered	12 sec.
	1 month unfiltered	1 min.

NECST laboratory

Evaluation of BankSealer: Performances

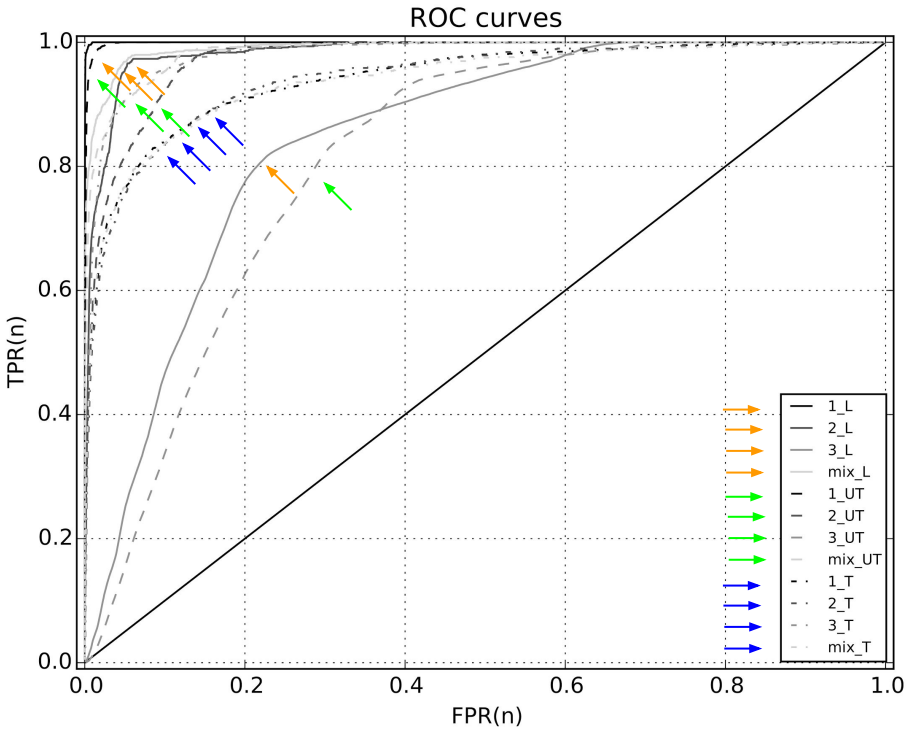
Generate synthetic frauds based on scenarios built with the collaboration of bank experts that replicate the typical real attacks performed against online banking users

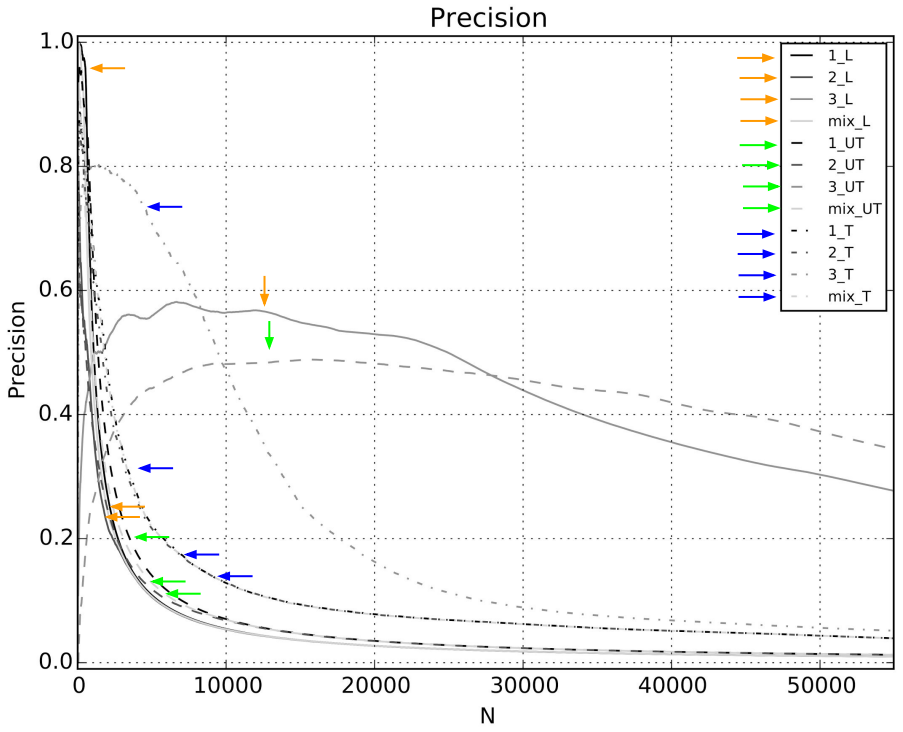


Inject n fraudulent transactions (or users) in the testing dataset and analyze the top n transactions (or users) in the ranking: ROC Curve - Precisions

NECST laboratory

Evaluation of BankSealer: Performances





MICHELE CARMINATI, R. Caron, I. Epifani, F. Maggi, S. Zanero.
"BankSealer: A decision support system for online banking fraud analysis and investigation" Computers & Security, Volume 53, September 2015, Pages 175-186, ISSN 0167-4048, http://dx.doi.org/10.1016/j.cose.2015.04.002

Evaluation of Banksealer: Security and Quality

Influence of the granularity at which the spending habits are modeled on the detection quality.

- System-Centric:
  - less computational requirements
  - more generalization (New Users)
- User-Centric:
  - better catch user peculiarities

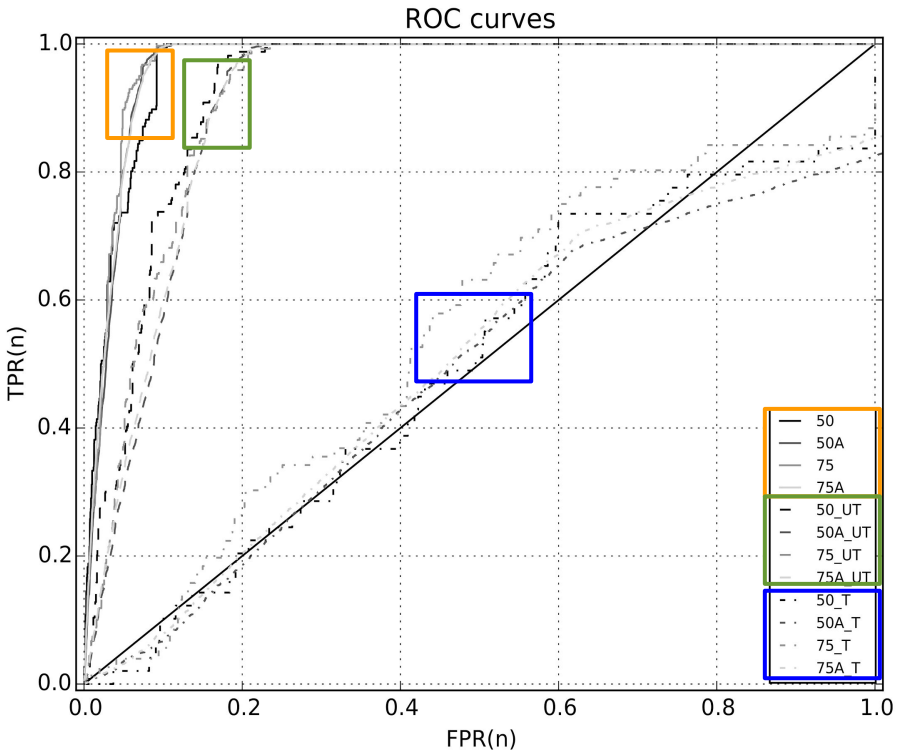
1	2	3
97 %	45 %	1%
98%	**75%**	**65%**

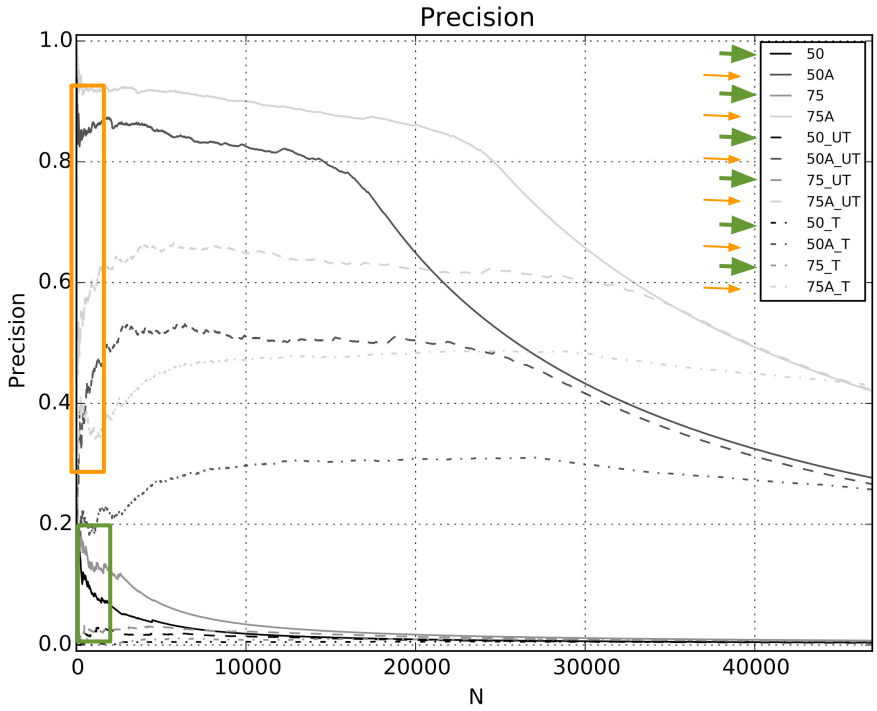
Security against evasion attacks: assess the robustness to malicious attackers that are aware of how spending habits are modeled.

- Design and implement of an attacking tool that synthesizes **mimicry attacks**: allow a sophisticated attacker to cloak frauds to avoid detection.

NECST laboratory

Evaluation of Banksealer: Security and Quality





MICHELE CARMINATI, Mario Polino, Andrea Continella, Andrea Lanzi, Federico Maggi, Stefano Zanero - *"Security Evaluation of BankSealer: An Internet Banking Frauds Analysis System"* - ACM Transactions on Information and System Security (TISSEC)
(Writing in Progress)

FraudBuster: Temporal Analysis 2.0

The Problem

Improve detection of frauds that exploits the repetition of legitimate-looking transactions

Objective

Find a model able to describe the user's transactions in term of their periodicity and detect frauds as "deviations" from the learnt temporal model

Proposed Solution

- Study **transactions distribution in the time domain** and classify user periodicity
  - Auto-correlation

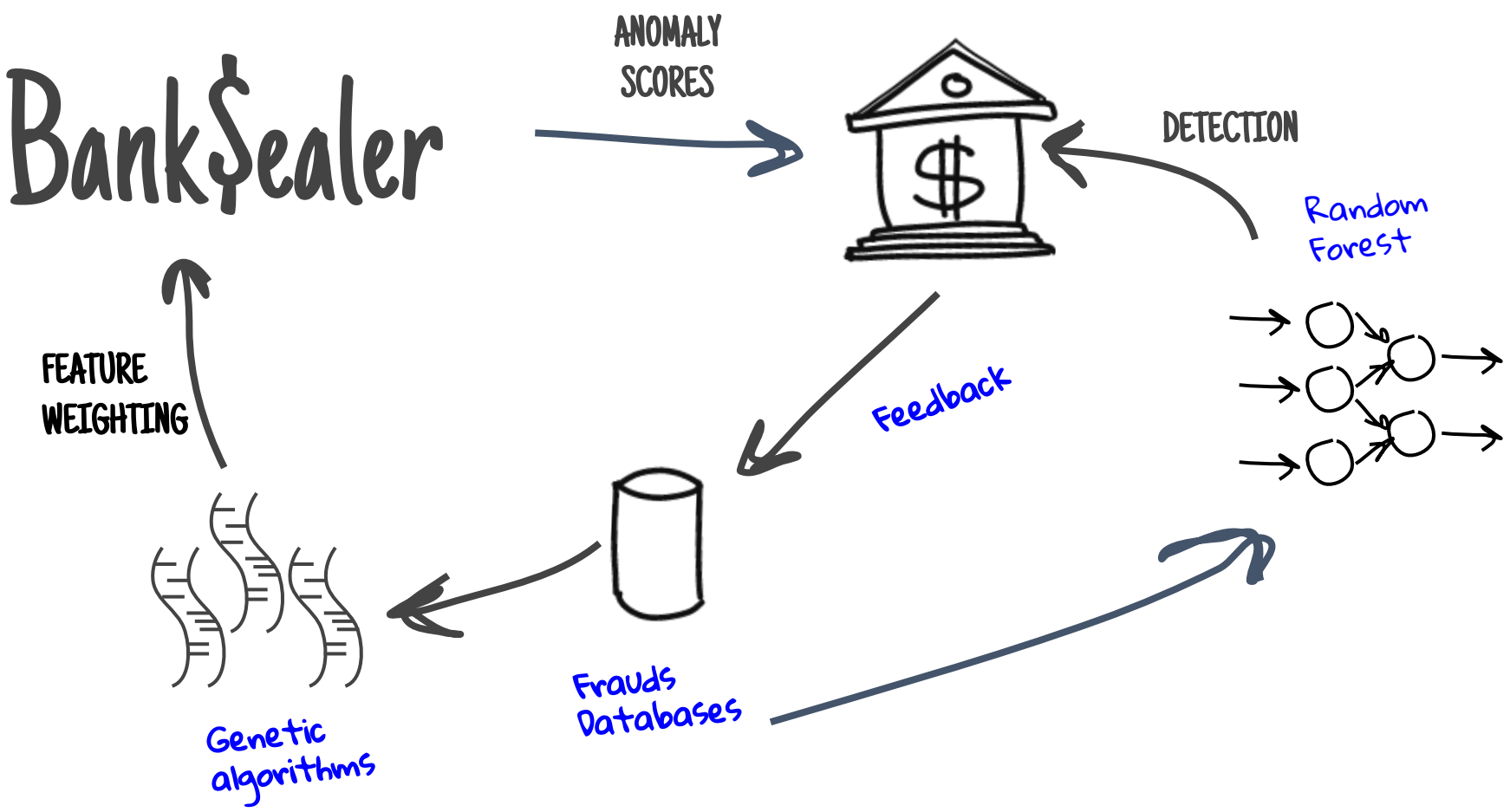
→ Almost 60% of user shows a monthly periodic behaviour

- **Detection**
  - Improved thresholds based on mean and variance
  - Histogram based distribution on monthly basis
  - *Dynamic Time Warping* on Transaction Time series

→ + 30% detection rate improvement

NECST laboratory

Supervised Analysis

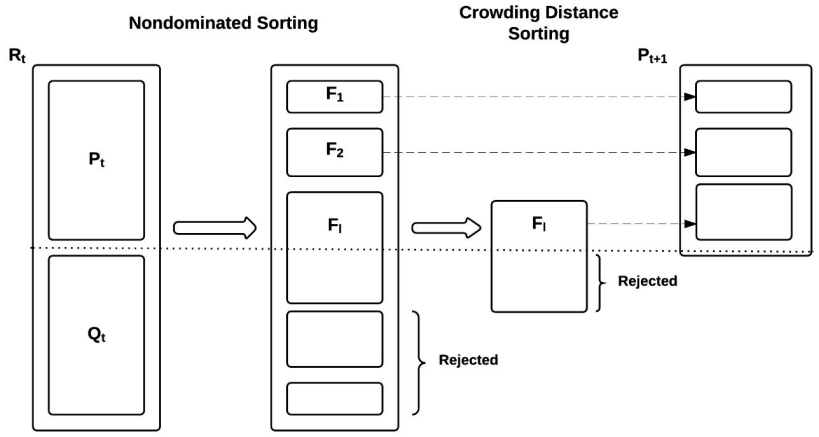


Genetic Algorithms Optimization

The problem (Feature Weighting)

- not receive a direct feedback from the analyst
- not completely adaptive
- requires manual parameter setting

Objective: Automatically learn best weight configuration exploiting analyst's feedback

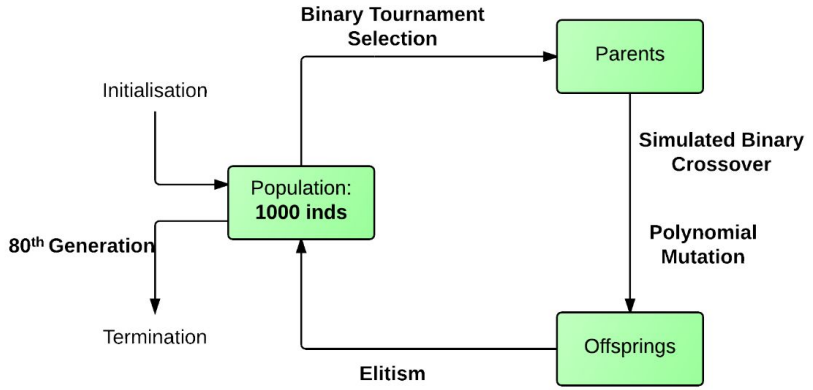


Proposed Solution

NSGA-II: MOGA based on an ordering algorithm that preserve non dominated solutions (Pareto Optimal).

Fitness Functions:

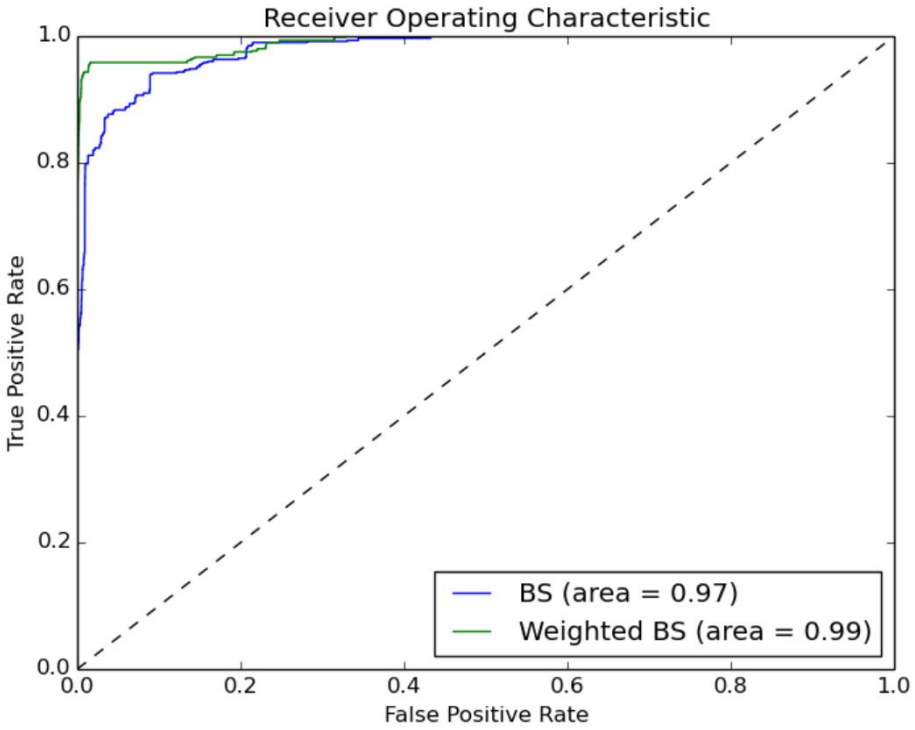
- True Positive Rate
- Average Precision
- Frauds Penalty

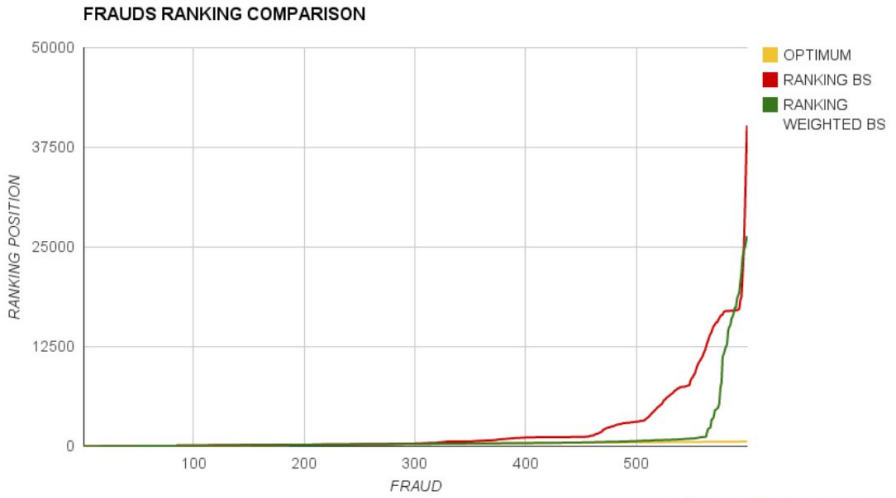


*Implemented with **Deap***

Genetic Algorithms Optimization

	BS TPR	Weighted BS TPR	Improvements
Scenario Misto	58%	81%	**+23%**





NECST laboratory

Thanks!

For further information:

stefano.zanero@polimi.it
michele@banksealer.com