Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Data Quality, ATS & Fraud Detection

Ylenia Maruccia & Riccardo Crupi

Spoke 3 III Technical Workshop, Perugia 26-29 Maggio, 2025

Data Quality & ATS

Scientific Rationale

Anomaly detection plays a crucial role in ensuring data quality for time series datasets.
The objective of anomaly detection on time series data is to identify unusual patterns or outliers that deviate from the expected behavior. By detecting anomalies, we can ensure data quality, identify data integrity issues, and mitigate potential risks.

By identifying and managing anomalies, we can improve the quality of data in our systems by helping to identify:

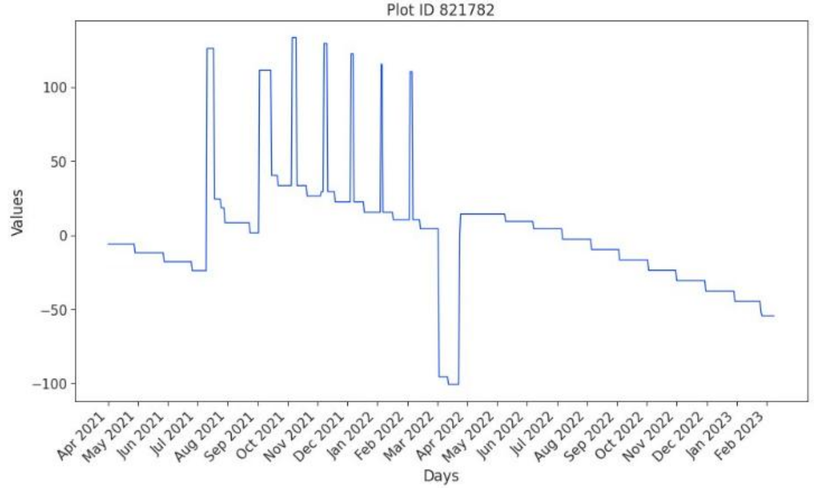
- **Data measurements errors**
- **Data entry mistakes**
- **Data corruption**
- **Network distruption**
- **Intercept fraudulent transactions.**

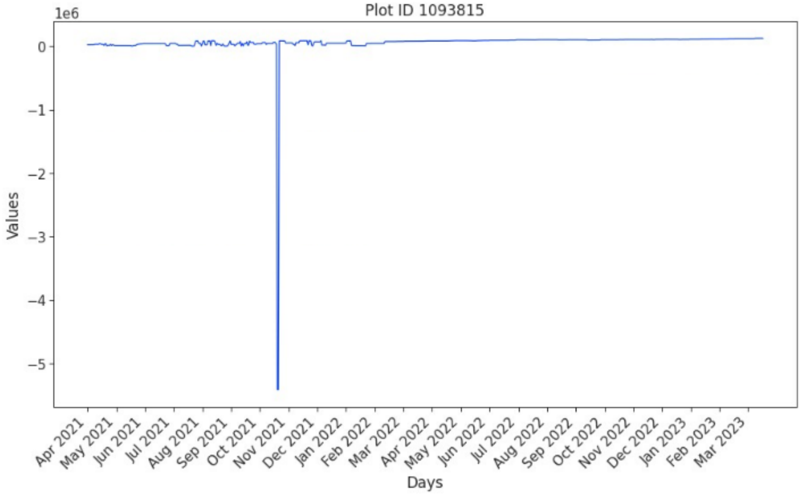
Technical Objectives, Methodologies and Solutions

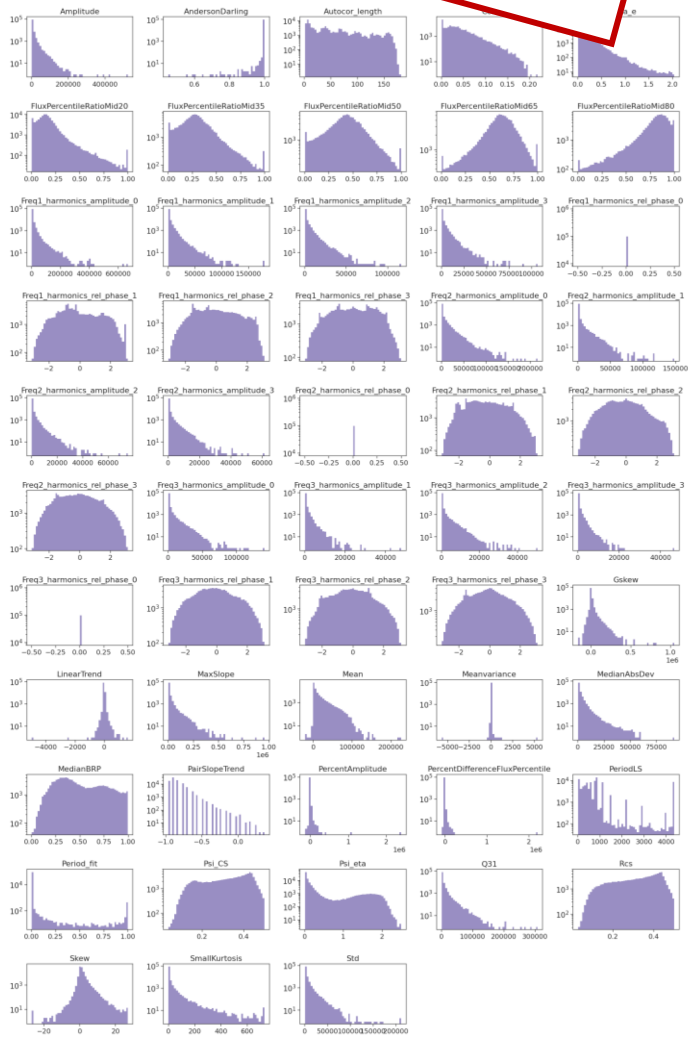
LAST YEAR WE STARTED FROM HERE...

Starting from the ISP dataset:

- **Building time series representing each ID (total users: 2,636,027)**
- **Feature engineering: extracting statistical parameters for each time series (FATS library)**
- **First attempt for the identification of anomalies (K-Nearest Neighbours)**

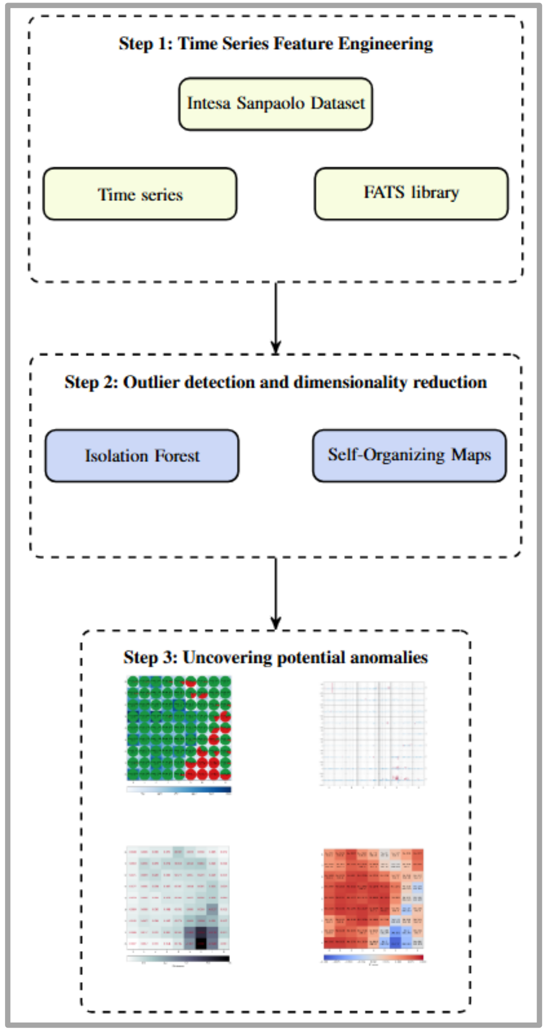






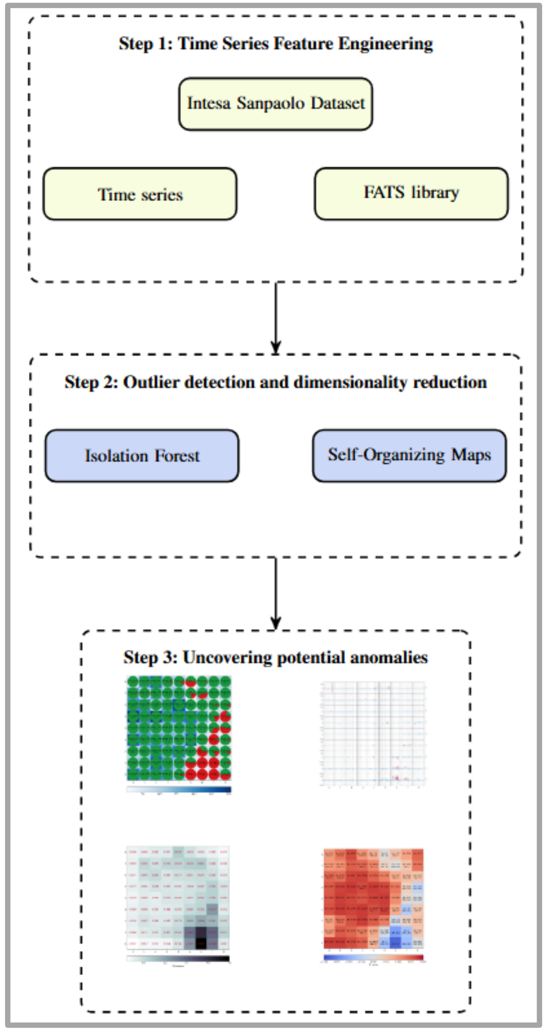
WHERE WE LEFT OFF IN DECEMBER...

Technical Objectives, Methodologies and Solutions



WHERE WE LEFT OFF IN DECEMBER...

Technical Objectives, Methodologies and Solutions



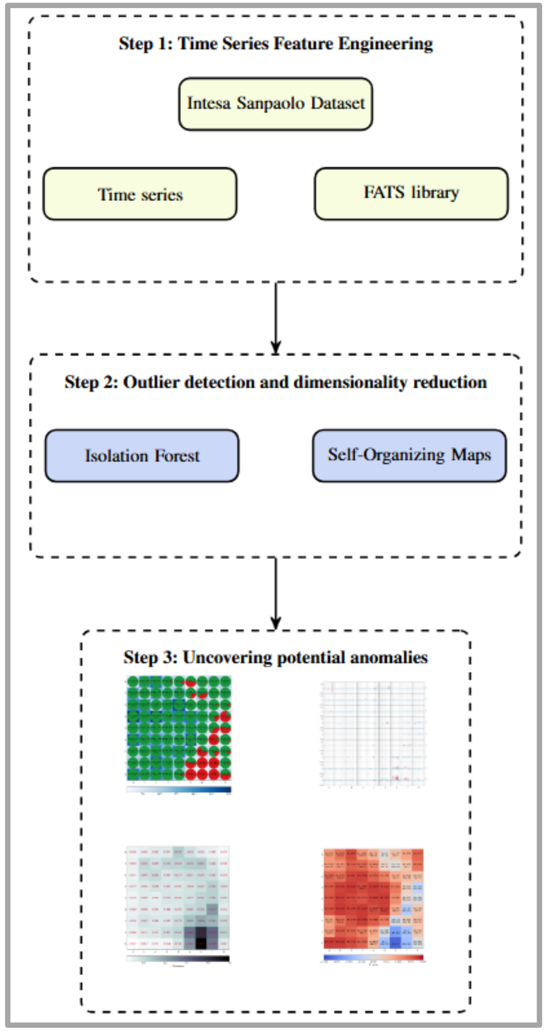
Description of the ISP dataset through Feature Engineering.

TABLE I
THE FEATURES COMPUTED ADOPTING THE FATS LIBRARY AND ADOPTED IN THIS WORK.

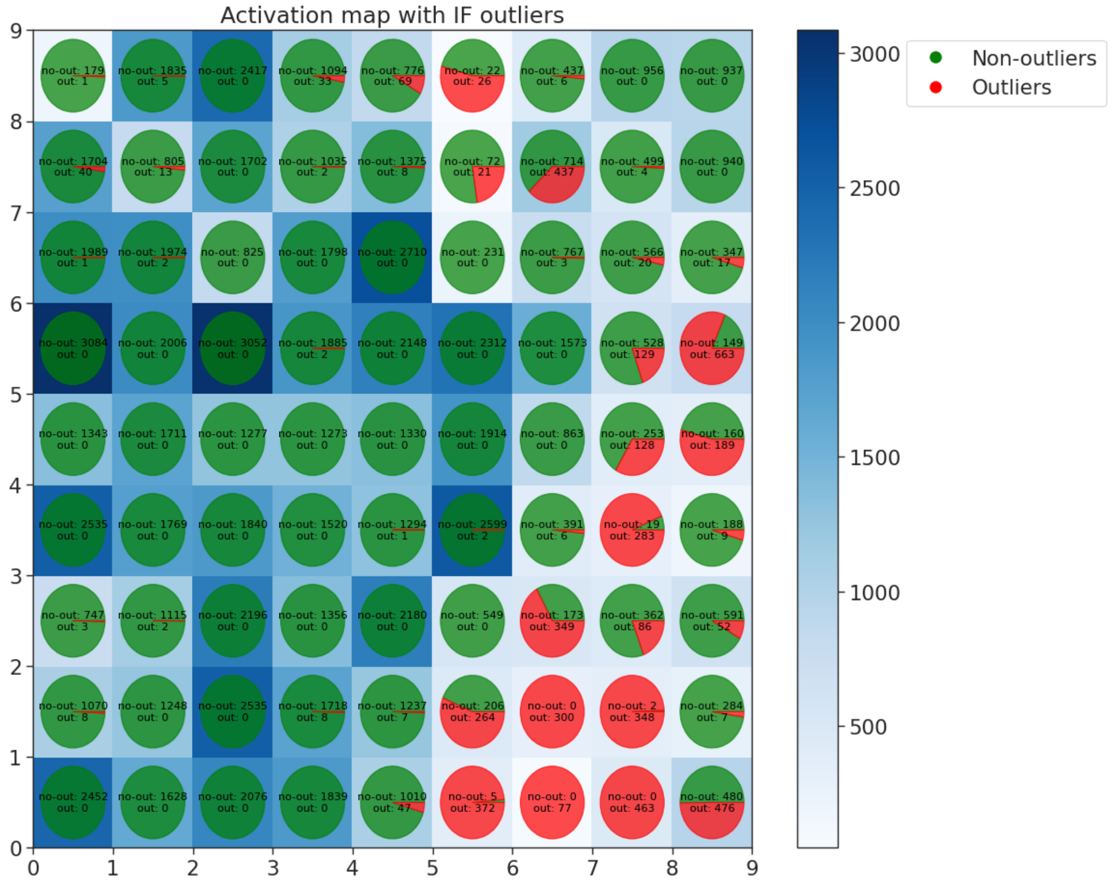
Category	Feature	Description
Amplitude and variation	Amplitude	The amplitude of the time series.
	PercentAmplitude	Percentage amplitude.
	PercentDifferenceFluxPercentile	Percentage difference between flux percentiles.
	FluxPercentileRatioMid20	Ratio of flux percentiles between 40^{th} and 60^{th} percentile.
	FluxPercentileRatioMid35	Ratio of flux percentiles between 32.5^{th} and 67.5^{th} percentile.
	FluxPercentileRatioMid50	Ratio of flux percentiles between 25^{th} and 75^{th} percentile.
	FluxPercentileRatioMid65	Ratio of flux percentiles between 17.5^{th} and 82.5^{th} percentile.
	FluxPercentileRatioMid80	Ratio of flux percentiles between 10^{th} and 90^{th} percentile.
	Meanvariance	Variance of the mean value.
Periodicity and harmonics	PeriodLS	Period derived from the Lomb-Scargle method.
	Period_fit	Fitted period value.
	Freq1_harmonics_amplitude_[0,1,2,3]	Amplitude of the [1^{st}, 2^{nd}, 3^{rd}, 4^{th}] harmonic component for the first frequency.
	Freq2_harmonics_amplitude_0	Amplitude of the [1^{st}, 2^{nd}, 3^{rd}, 4^{th}] harmonic component for the second frequency.
	Freq3_harmonics_amplitude_0	Amplitude of the [1^{st}, 2^{nd}, 3^{rd}, 4^{th}] harmonic component for the third frequency.
	Freq[1-3]_harmonics_rel_phase_[1-3]	Relative phase of the [2nd, 3rd, 4th] harmonic component for the [first, second, third] frequency.
Distribution statistics	Mean	Mean value of the time series.
	MedianAbsDev	Median absolute deviation from the median.
	Skew	Skewness of the light curve.
	SmallKurtosis	Kurtosis of the light curve.
	Std	Standard deviation of the time series.
	Gskew	A measure of the skewness of the light curve.
	Q31	Interquartile range (Q3-Q1).
Trends and local variations	LinearTrend	The slope of the linear fit to the time series.
	MaxSlope	Maximum slope between consecutive points in the time series.
	PairSlopeTrend	Slope of the pairwise differences.
Indicators of non-Gaussian or non-linear behavior	AndersonDarling	Anderson-Darling test statistic for normality.
	Con	A measure of the light curve's constancy.
	Eta_e	Ratio of mean square successive differences to the variance.
	Psi_CS	Clipped standard deviation of the light curve.
	Psi_eta	Eta index of the light curve.
	Rcs	Robustness of the light curve.
	MedianBRP	Median buffer range percentage.
Temporal structure	Autocor_length	The length of the autocorrelation function.

WHERE WE LEFT OFF IN DECEMBER...

Technical Objectives, Methodologies and Solutions



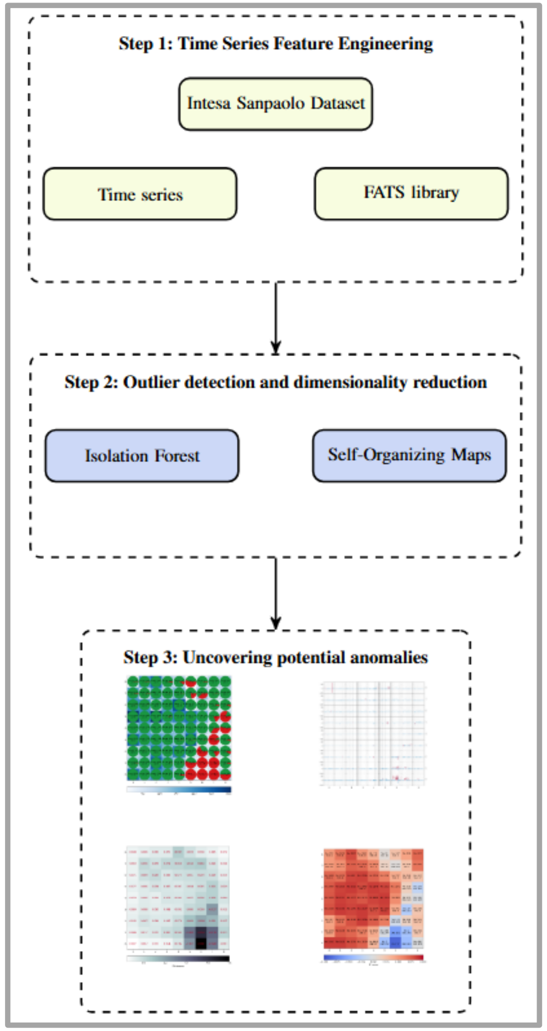
Outlier detection through the use of Isolation Forest, mapped on the Self-Organizing Map.



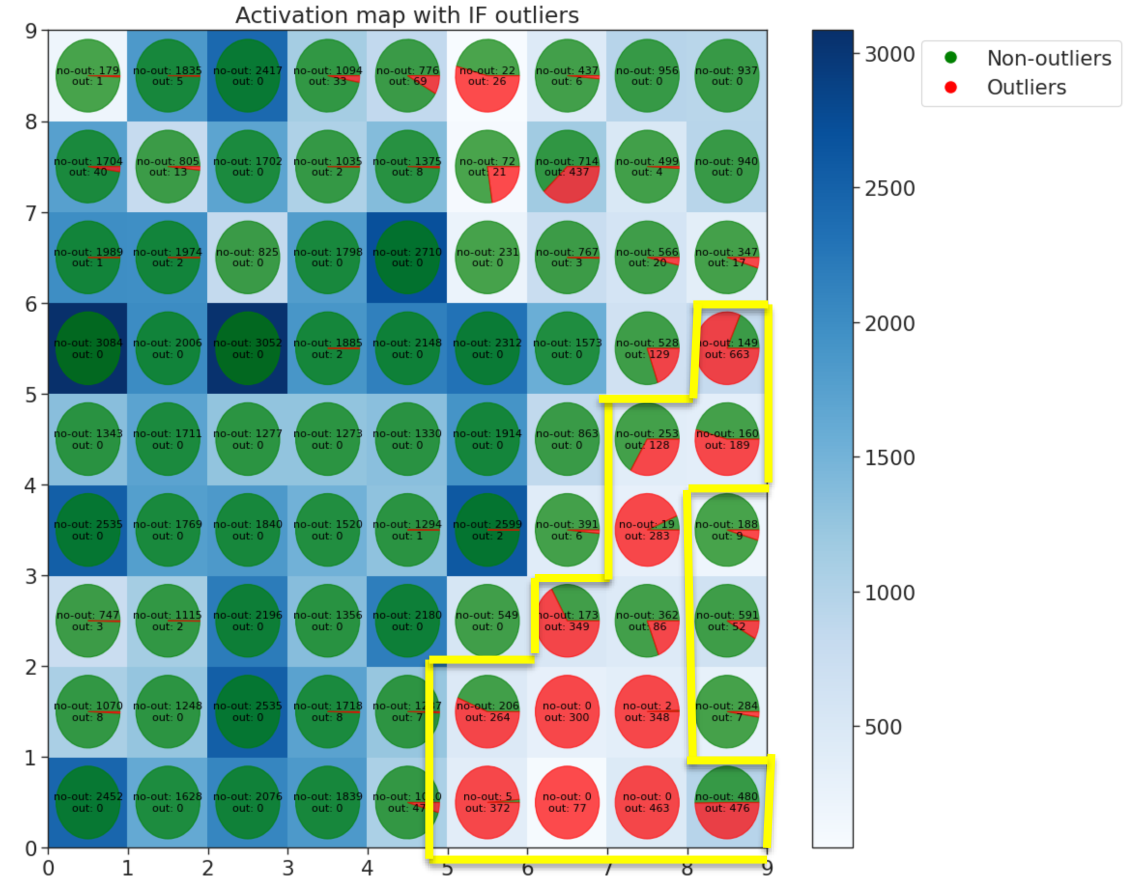
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

WHERE WE LEFT OFF IN DECEMBER...

Main Results

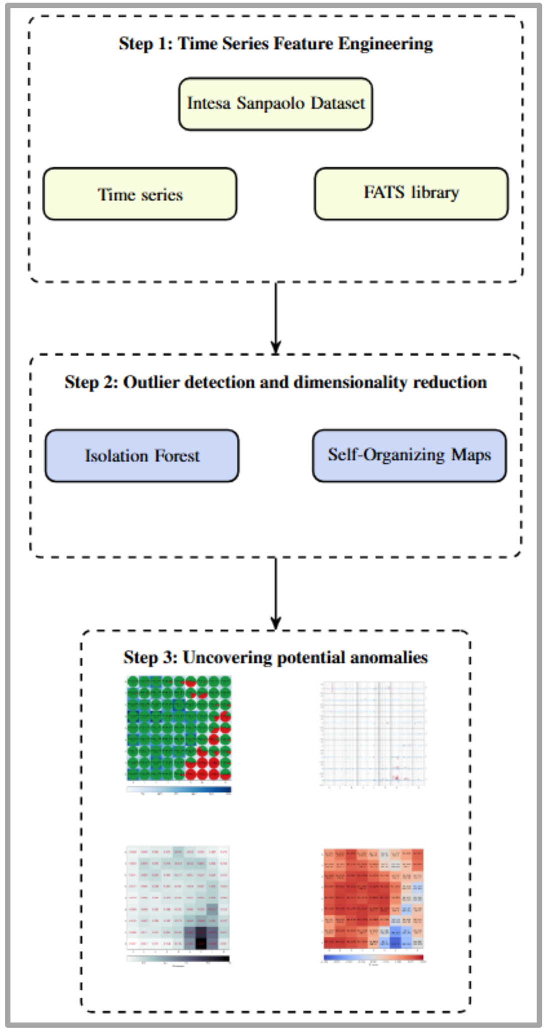


Uncovering similar behaviours among users and potential anomalies.

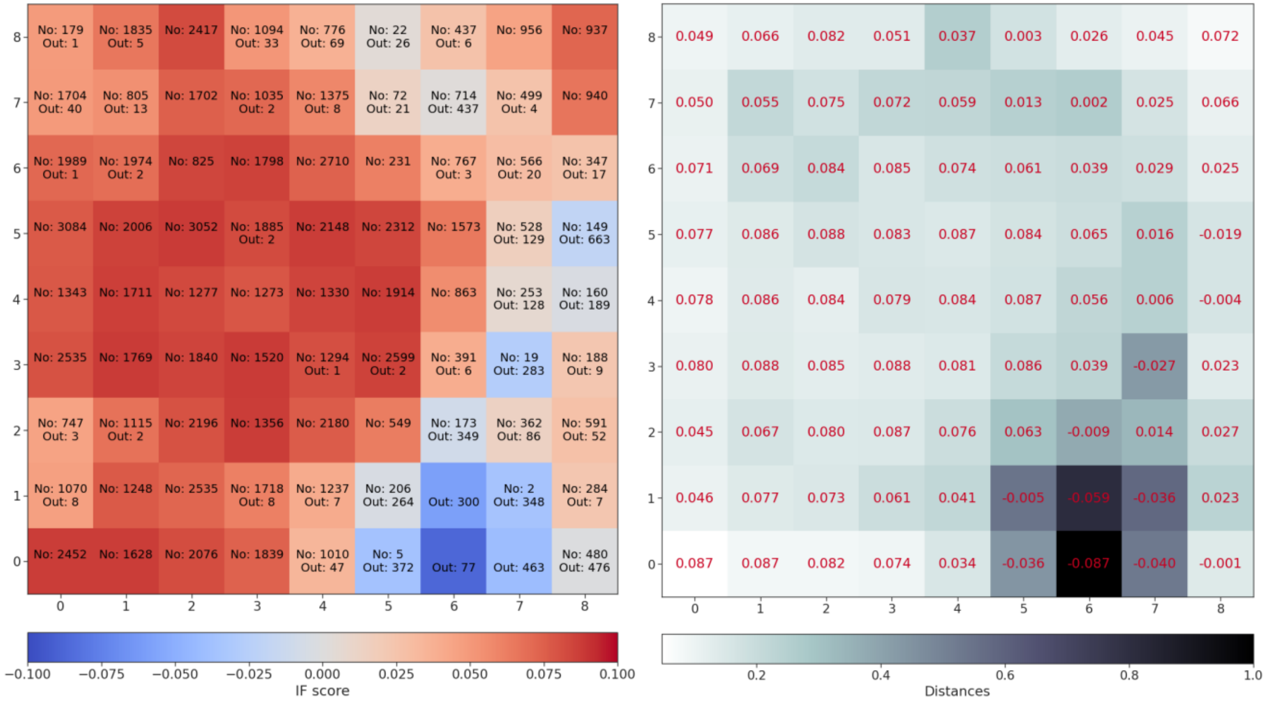


WHERE WE LEFT OFF IN DECEMBER...

Main Results

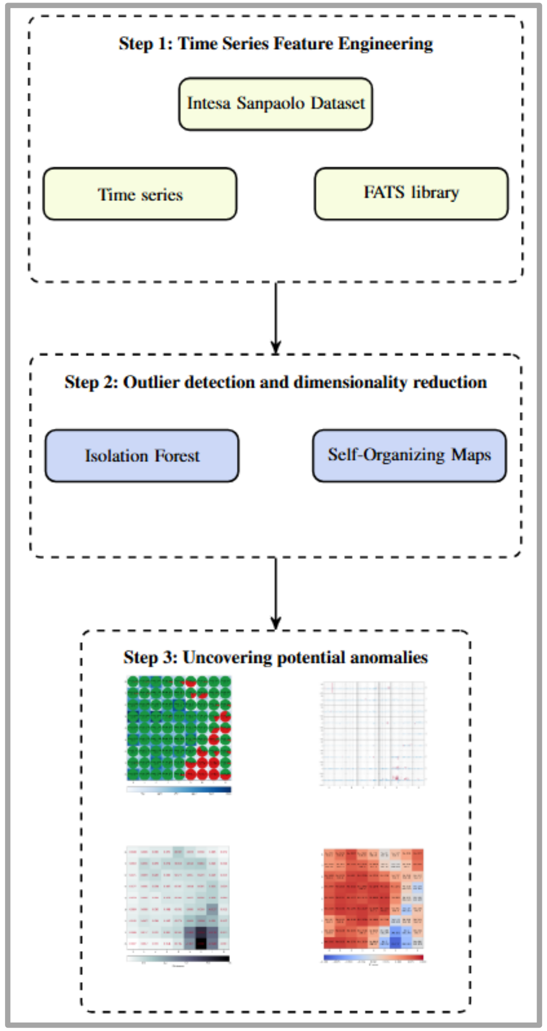


Uncovering similar behaviours among users and potential anomalies.

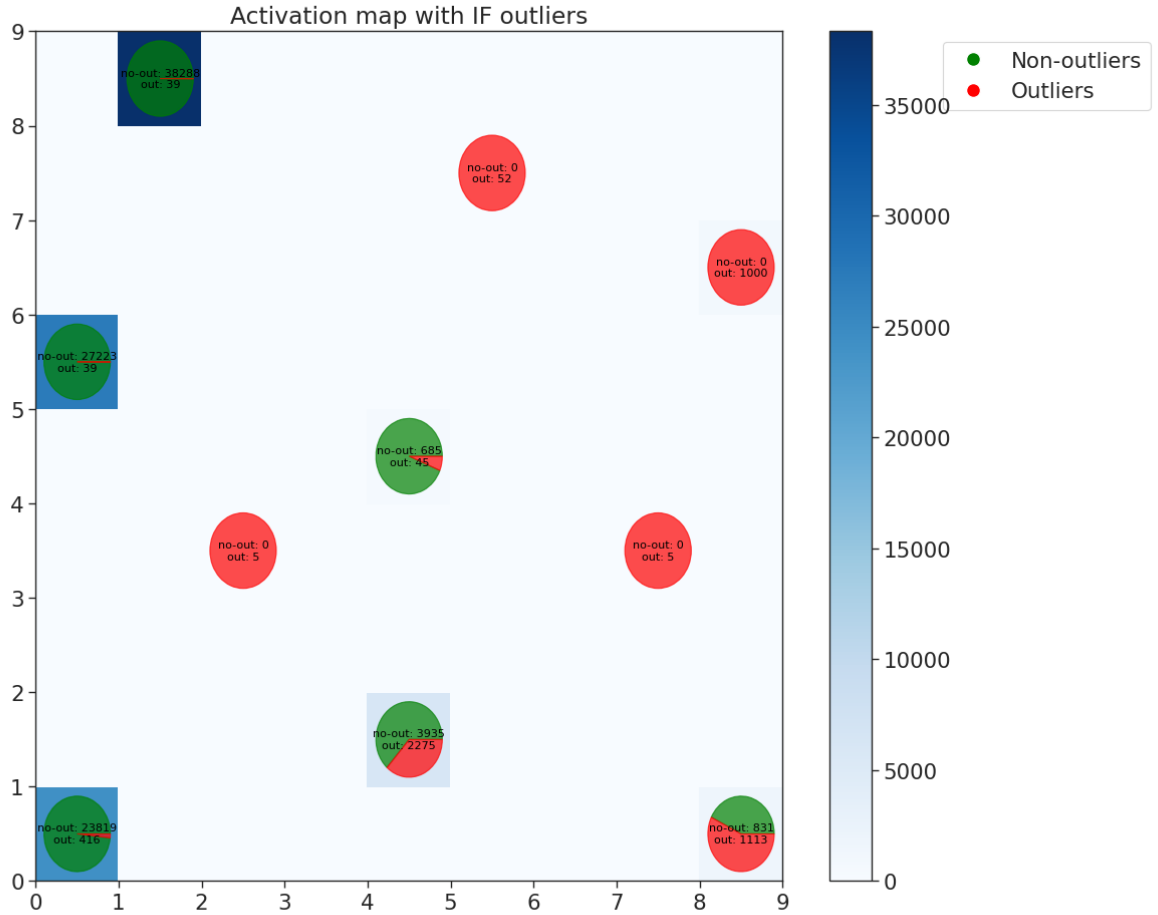


WHERE WE LEFT OFF IN DECEMBER...

Main Results



Uncovering similar behaviours among users and potential anomalies.



Final Steps

The work package in the Spoke 3 corresponds to WP3.

Deliverable:

- **M6 Data delivery – completed 100%**
- **M7 Preliminary results on data exploration and how to format the dataset, preprocessing the data – completed 100%**
- **M8 first prototype application with ML to one use case – completed 100%**
- **M9 Prototype with explainability – completed 100%**
- **M10 Python library that works on the banking dataset and environment – ongoing 70%**

Final Steps

Anomaly detection in the banking sector using feature engineering

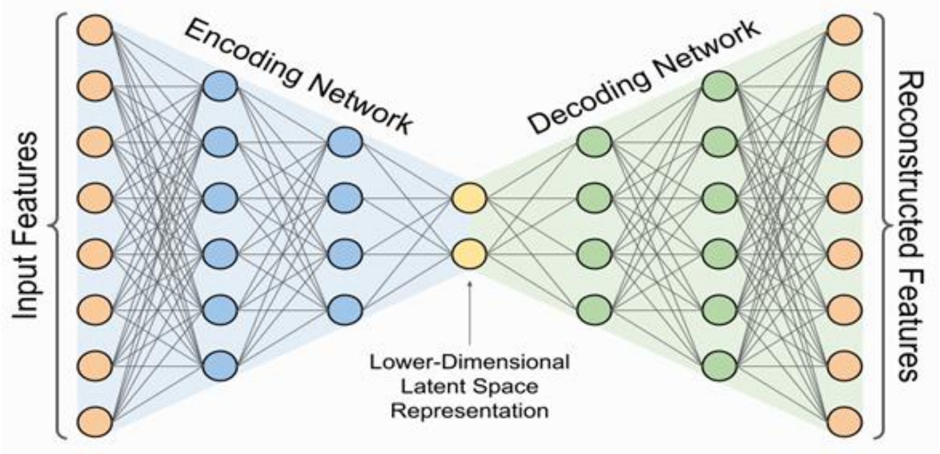
Ylenia Maruccia[1*], Stefano Cavuoti[1,2*], Riccardo Crupi[3*], Giuseppe Riccio[1], Demetra De Cicco[4], Paola Barone[3], Stella Rubini[3], Adriano Koleci[3], Paolo Pavanelli[3], Alessandro D. Sabatino[3], Rachele Desiante[3]

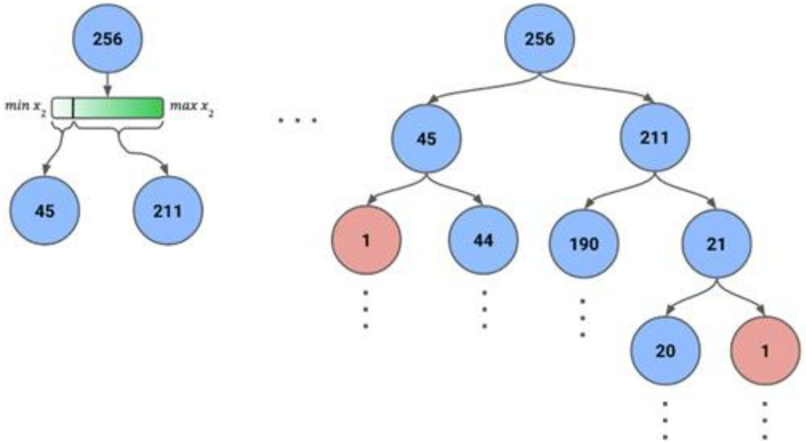
[1]INAF - Astronomical Observatory of Capodimonte,Via Moiariello 16, Napoli, I-80131, Italy.
[2]INFN section of Naples,via Cinthia 6, Napoli, I-80126, Italy.
[3]Intesa Sanpaolo S.p.A.,Corso Inghilterra 3, Turin, I-10138, Italy.
[4]Department of Physics "E. Pancini", University Federico II of Napoli,Via Cinthia 21, Napoli, I-80126, Italy.

Our paper has been submitted on Financial Innovation and it is under review.

Technical Objectives, Methodologies and Solutions

- Training Dataset Size: ~130k points.
- Features FATS extracted <u>separately</u> for two available years: april 2021 – march 2022, april 2022 – march 2023.



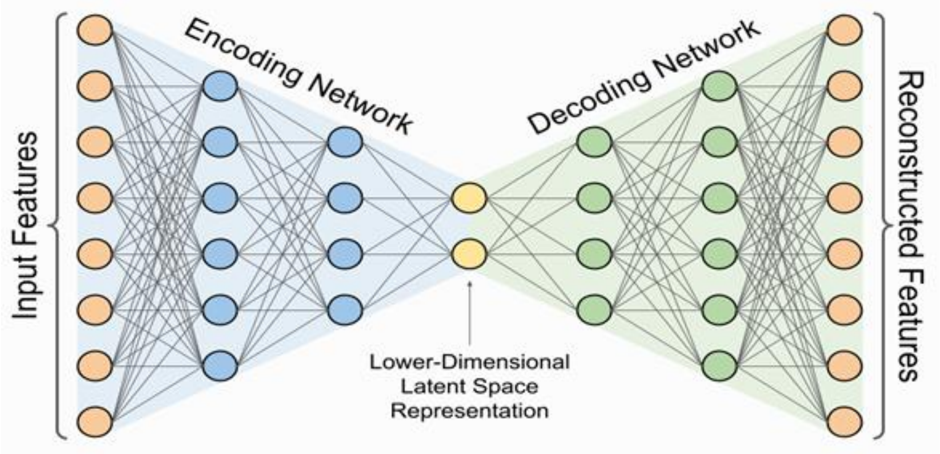


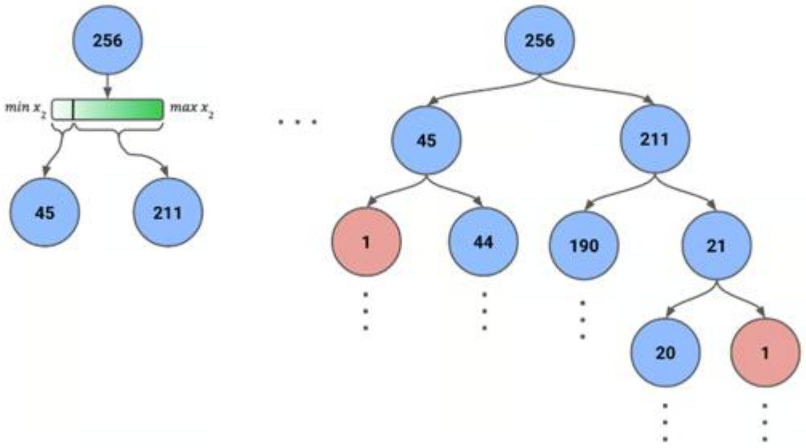
AUTOENCODER
- Latent space dimension: 10;
- Input/Output space dimension: 50;
- Points are classified as anomalies if they are reconstructed with significant error by the autoencoder.

ISOLATION FOREST
- Outlier contamination set to 5%

Technical Objectives, Methodologies and Solutions

- Training Dataset Size: ~130k points.
- Features FATS extracted <u>separately</u> for two available years: april 2021 – march 2022, april 2022 – march 2023.





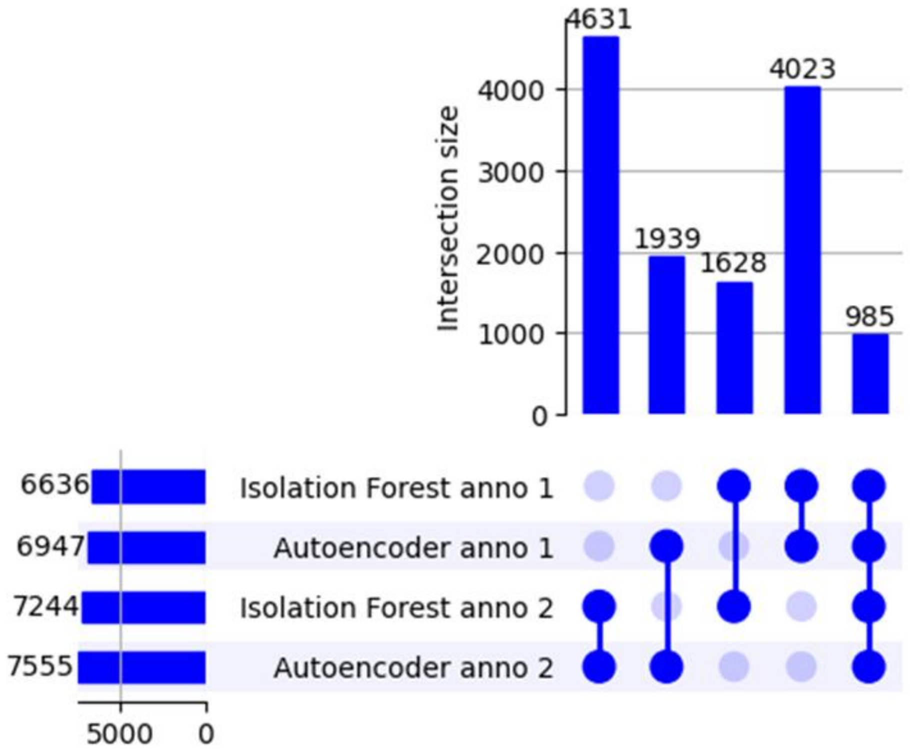
AUTOENCODER
- Latent space dimension: 10;
- Input/Output space dimension: 50;
- Points are classified as anomalies if they are reconstructed with significant error by the autoencoder.

WHO IS ANOMALOUS?

ISOLATION FOREST
- Outlier contamination set to 5%

Main Results

Common anomalies of both the models between the two years.



Exclusive anomalies.

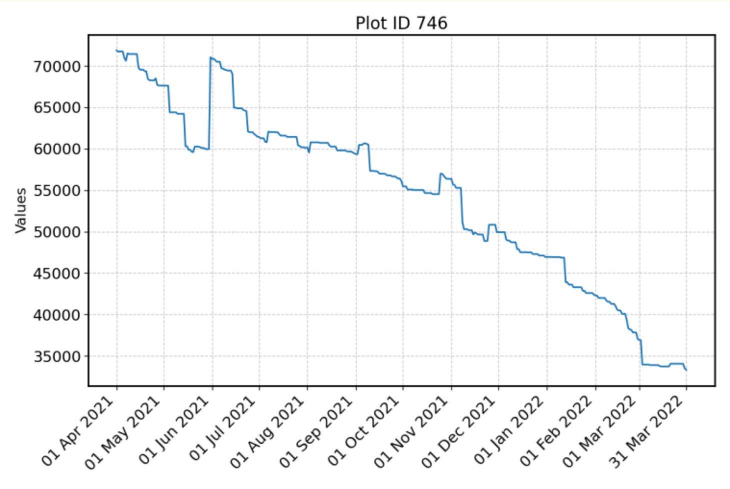
Anomalies	Counts
Exclusive to year 1 autoencoder	2244
Exclusive to year 2 autoencoder	1823
Exclusive to year 1 isolation forest	2115
Exclusive to year 2 isolation forest	1650

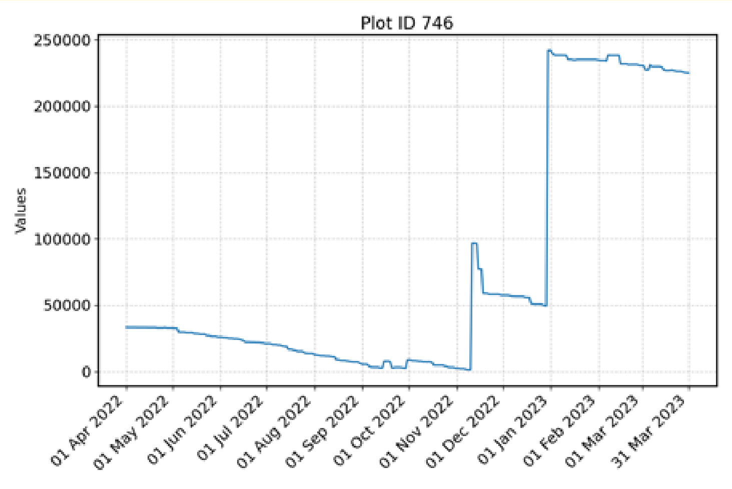
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

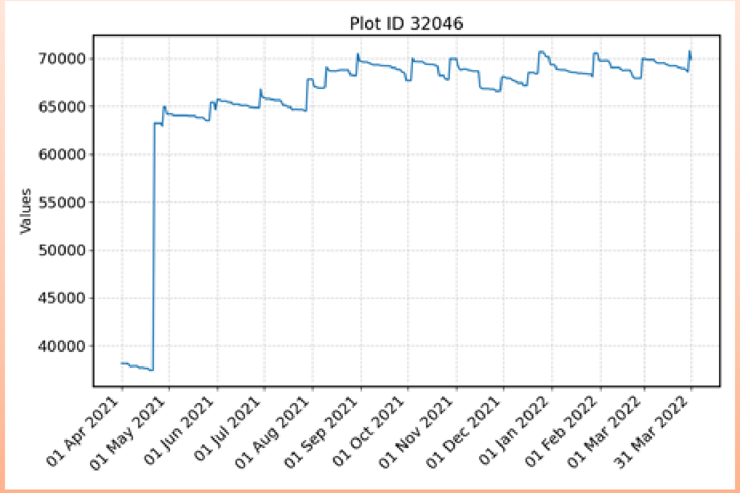
Main Results

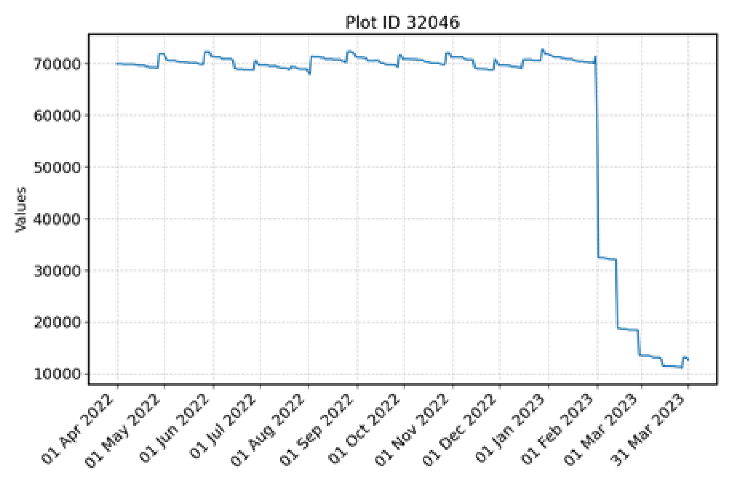
Year 1

Year 2









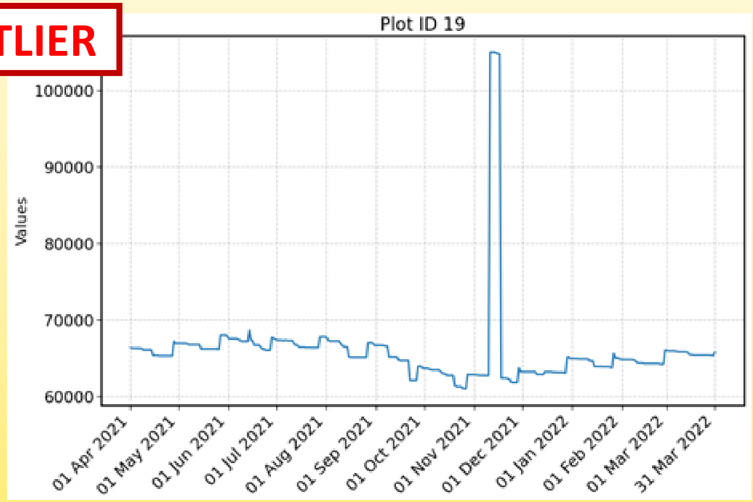
Both years outliers

Main Results

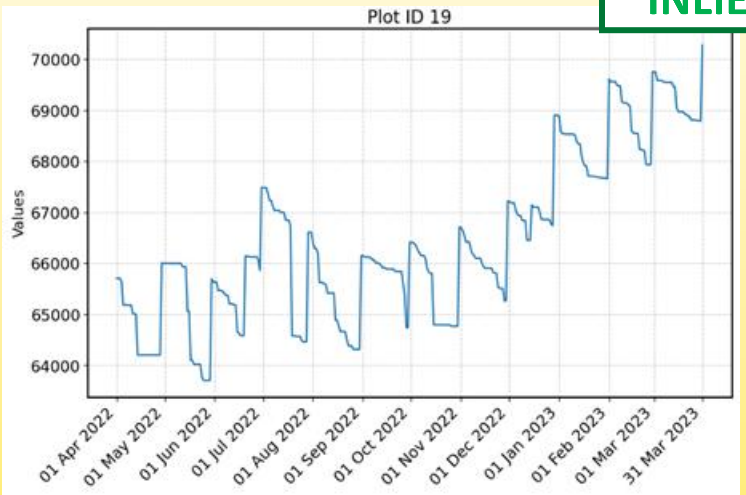
Year 1

Year 2

OUTLIER

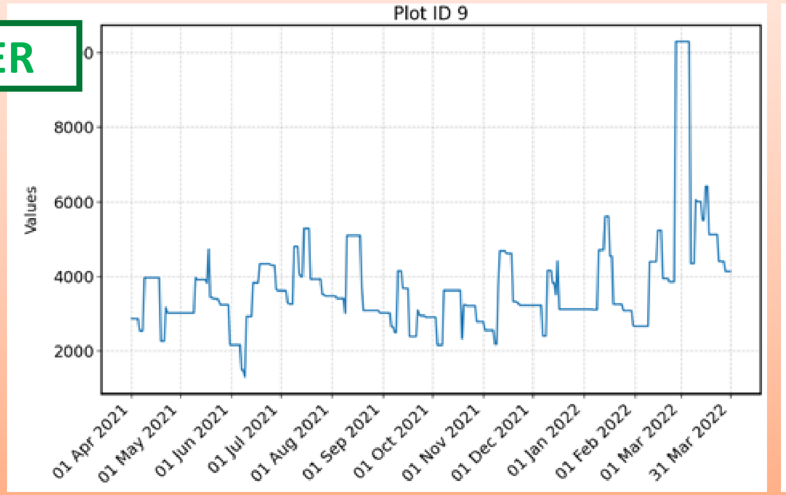


INLIER

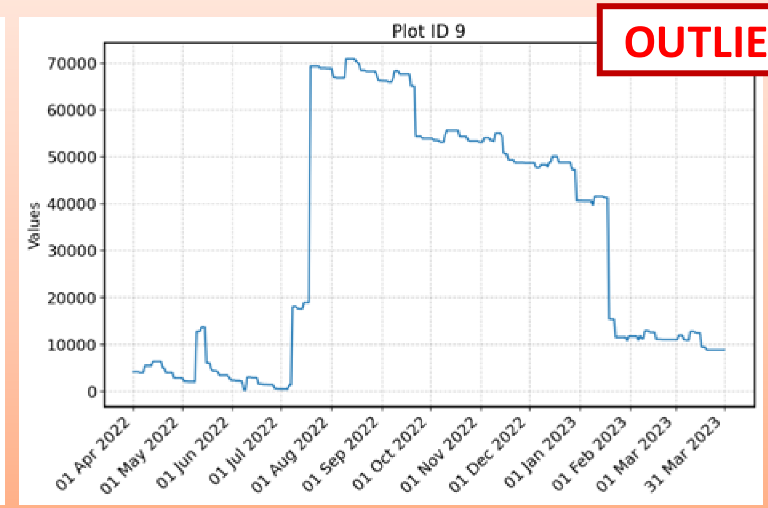


Single year outliers

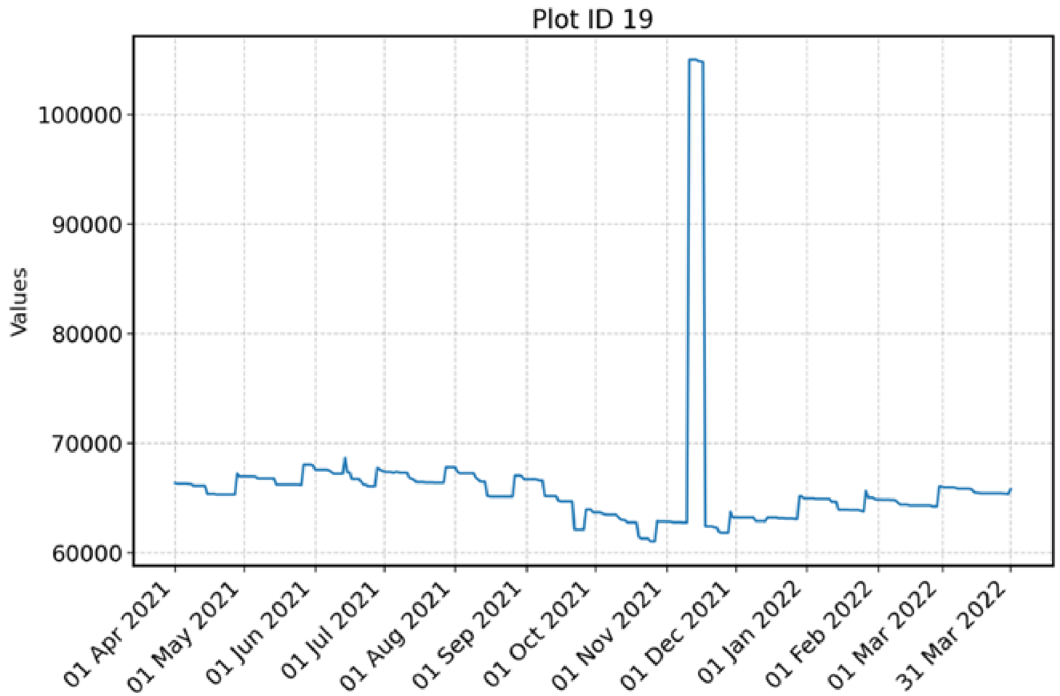
INLIER



OUTLIER



Ongoing activities & Next steps



We are working on the time series using windows of different length for each user.

We firstly consider a window large one day (N=1).

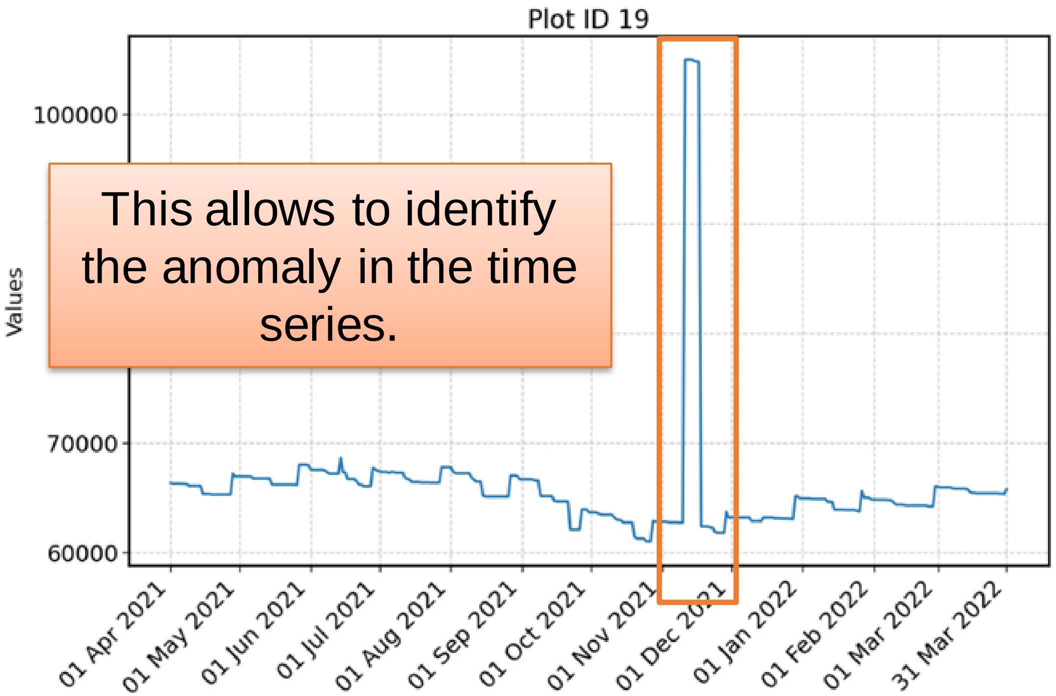
- **We remove the 1st day, then we calculate FATS.**
- **Then, we remove the 2nd day and calculate FATS.**
- **…**
- **Finally, we remove the 365th day and calculate FATS.**

Then, we consider a window large two days (N=2).

- **We remove the 1st and 2nd days, then we calculate FATS.**
- **Then, we remove the 2nd and 3rd days and calculate FATS.**
- **…**
- **Finally, we remove the 364th and 365th day and calculate FATS.**

And so on, for N = 3, 4, ..., 7.

Ongoing activities & Next steps



We are working on the time series using windows of different length for each user.

We firstly consider a window large one day (N=1).

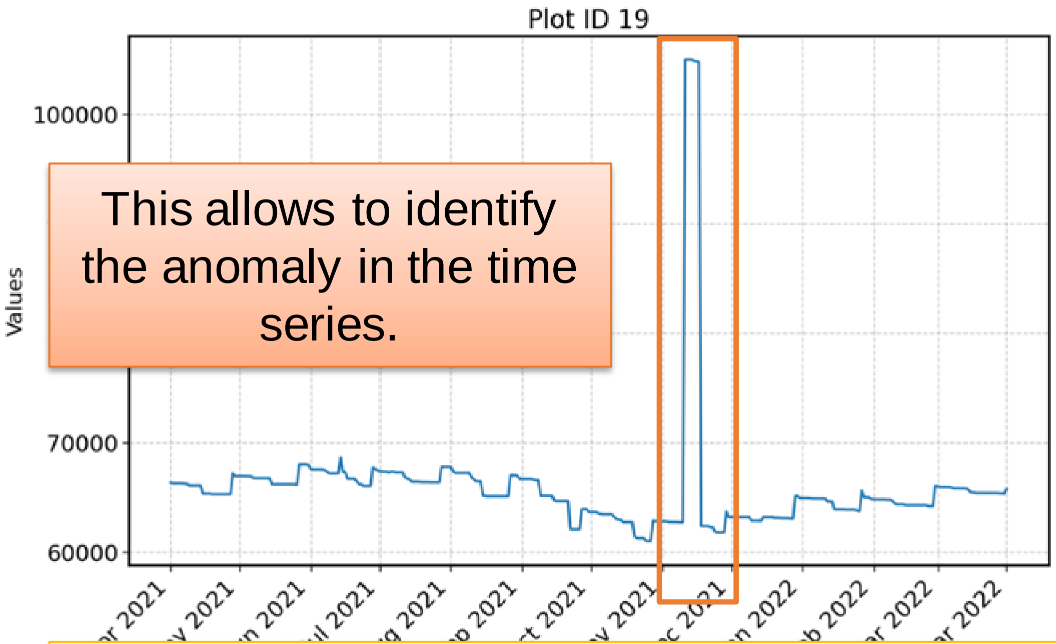
- **We remove the 1st day, then we calculate FATS.**
- **Then, we remove the 2nd day and calculate FATS.**
- **…**
- **Finally, we remove the 365th day and calculate FATS.**

Then, we consider a window large two days (N=2).

- **We remove the 1st and 2nd days, then we calculate FATS.**
- **Then, we remove the 2nd and 3rd days and calculate FATS.**
- **…**
- **Finally, we remove the 364th and 365th day and calculate FATS.**

And so on, for N = 3, 4, ..., 7.

Ongoing activities & Next steps



This allows to identify the anomaly in the time series.

Computational expensive!
For each user, this process needs ~35 minutes to complete the process.
Our dataset has 2 million of users!

We are working on the time series using windows of different length for each user.

We firstly consider a window large one day (N=1).

- **We remove the 1st day, then we calculate FATS.**
- **Then, we remove the 2nd day and calculate FATS.**
- **…**
- **Finally, we remove the 365th day and calculate FATS.**

Then, we consider a window large two days (N=2).

- **We remove the 1st and 2nd days, then we calculate FATS.**
- **Then, we remove the 2nd and 3rd days and calculate FATS.**
- **…**
- **Finally, we remove the 364th and 365th day and calculate FATS.**

And so on, for N = 3, 4, ..., 7.

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

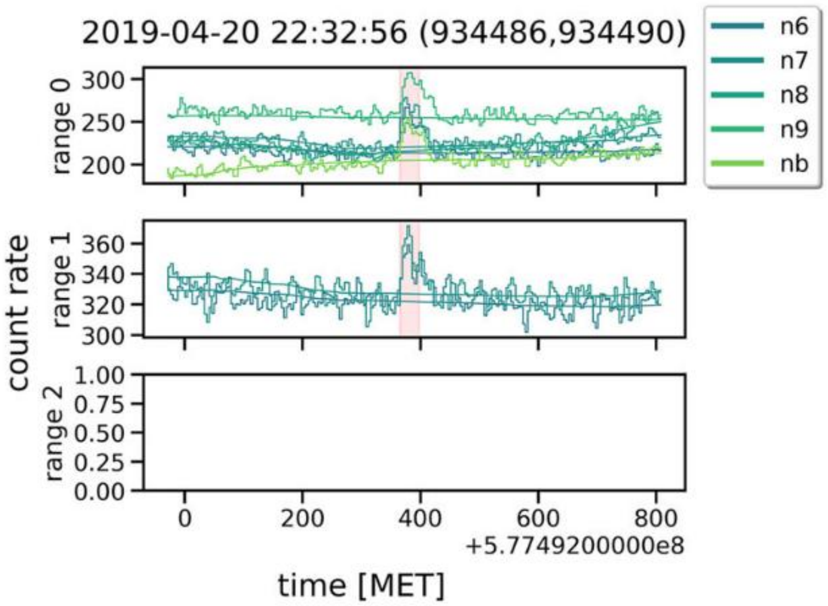
Anomaly Detection with FOCuS on Banking Time Series

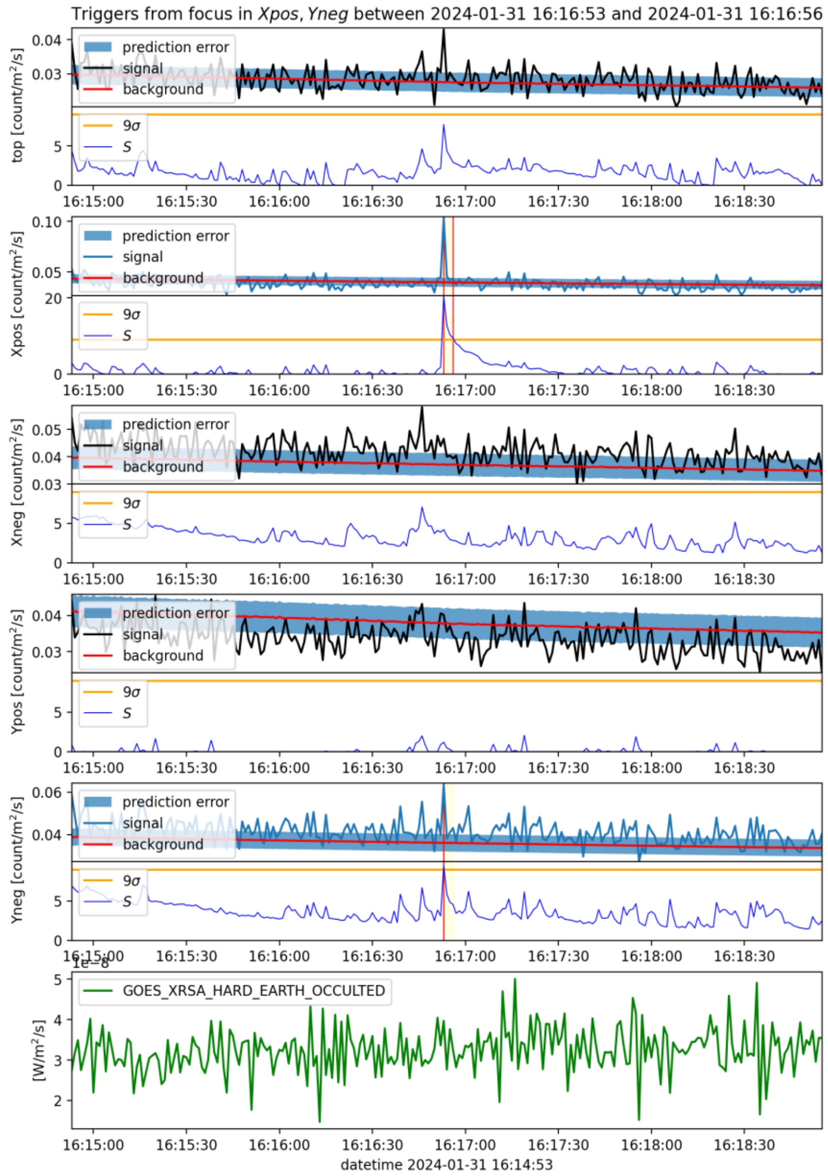
Andrea Adelfio (INFN-Pg) and Riccardo Crupi (Intesa Sanpaolo)

Technical Meeting, Perugia 26-05-2025 – 29-05-2025

FOCuS used on the Fermi Data

The FOCuS algorithm has been used with Fermi data, specifically the GBM data and the ACD data, to evaluate the efficiency of identifying new astrophysical transients.





Triggering Algorithm: FOCuS

The Functional Online CuSUM (FOCuS) is a fast and efficient algorithm based on the computation of the cumulative sum of the significance of the data.

***Efficient*: computes the sum of score statistics and compares it to a threshold. Efficient at identifying change points in the data set.**

$$S(s,n) = \sum_{i=s+1}^{N} H(x_i, \mu_0)$$

***Fast*: only records score statistics of data points that deviate from the distribution.**

Very powerful to find the start of an anomaly, before the cumulative sum is above the threshold.

Triggering Algorithm: FOCuS

The Functional Online CuSUM (FOCuS) is a fast and efficient algorithm based on the computation of the cumulative sum of the significance of the data.

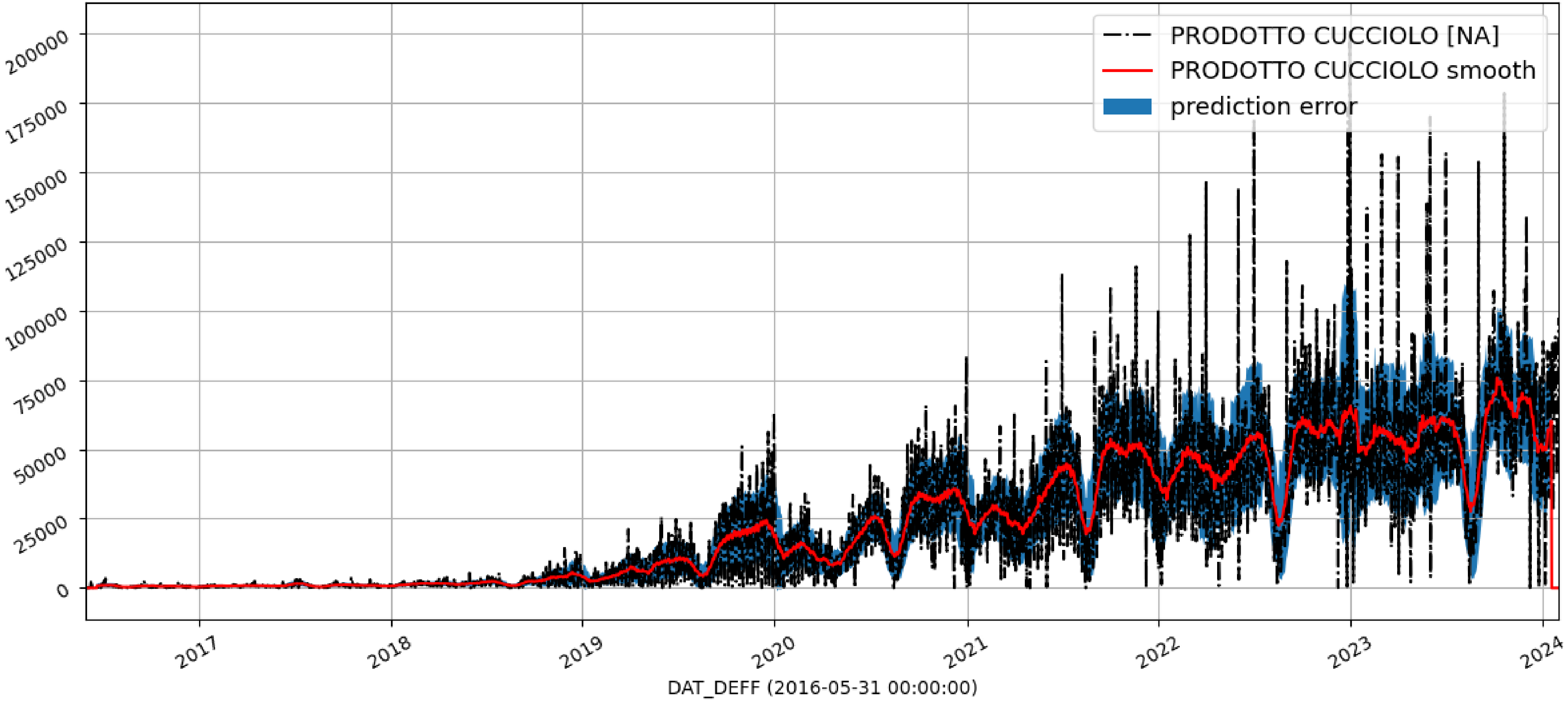
Can be used in *flavours*:
-Poisson-FOCuS: assumes a Poisson-like distribution of data; can be used for count rates data.

-Gaussian-FOCuS: assumes a Gaussian distribution; can be used for continuous signals (flux...)

-Non-parametric-FOCuS: no assumptions on the type of data.

Intesa Sanpaolo application

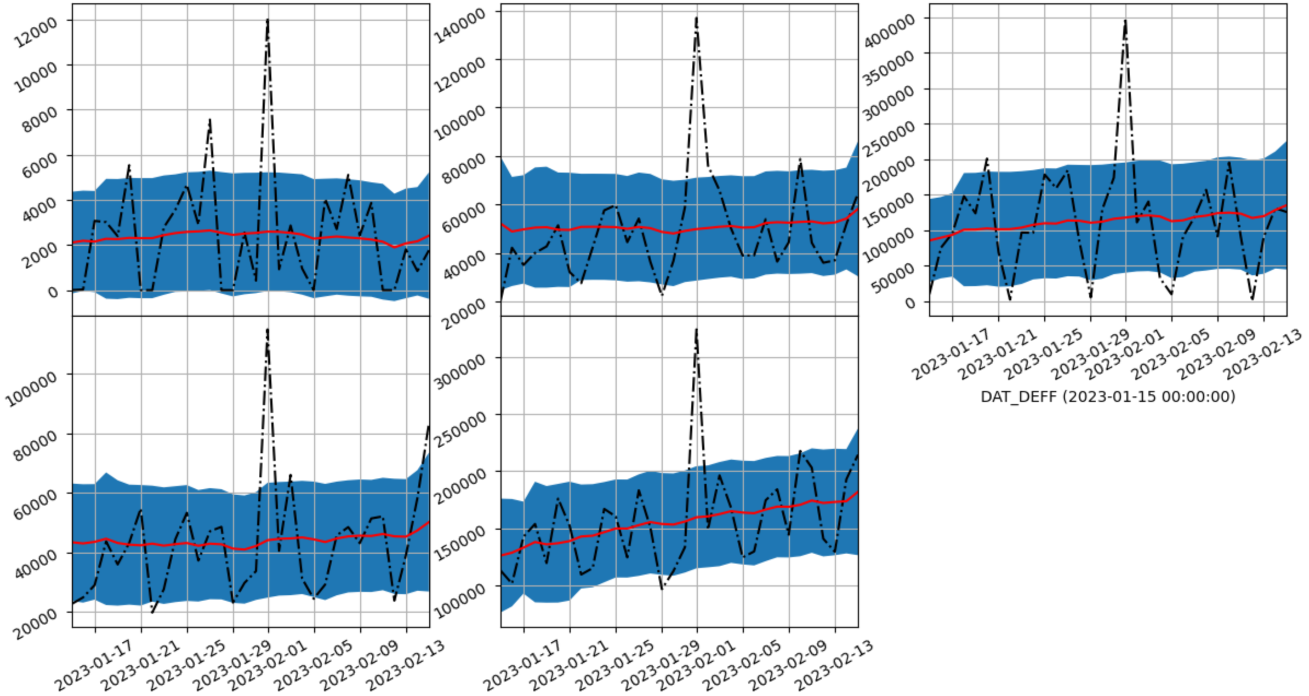
The idea is to apply it to the counts of daily policies subscriptions.



Poisson-FOCuS Results

Given a nominal behaviour taken with a moving average of 30 days and a standard deviation calculated on the same period:

37 anomalies found above a threshold of 2 standard deviations.




```
start_datetime,stop_datetime,triggered_faces
2023-03-29 00:00:00,2023-03-30 00:00:00,PRODOTTO MAMMOLO
2023-02-28 00:00:00,2023-03-01 00:00:00,PRODOTTO PAPERINO/PRODOTTO CUCCIOLO
2023-01-31 00:00:00,2023-02-01 00:00:00,PRODOTTO PAPERINO/PRODOTTO CUCCIOLO/PRODOTTO PIPPO/PRODOTTO DOTTO/PRODOTTO BRONTOLO
2023-01-01 00:00:00,2023-01-02 00:00:00,PRODOTTO PAPERINO
2022-12-27 00:00:00,2022-12-28 00:00:00,PRODOTTO EOLO
```

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Data Quality, ATS & Fraud Detection

Ylenia Maruccia & Riccardo Crupi

Spoke 3 III Technical Workshop, Perugia 26-29 Maggio, 2025

Fraud Detection

Technical Objectives, Methodologies and Solutions

WHERE WE STARTED

Starting point

- folder «*punto_2*», containing all product transactions and all events from January to September 2023 relating to BT that have had at least one fraudulent bank transfer.

Features exploration

- Analyze the distribution and correlation of features to identify possible relevant predictors for fraudulent bank transfers.

Random forest experiment

- Implement and evaluate a Random Forest model to classify fraudulent bank transfers

WHERE WE LEFT OFF

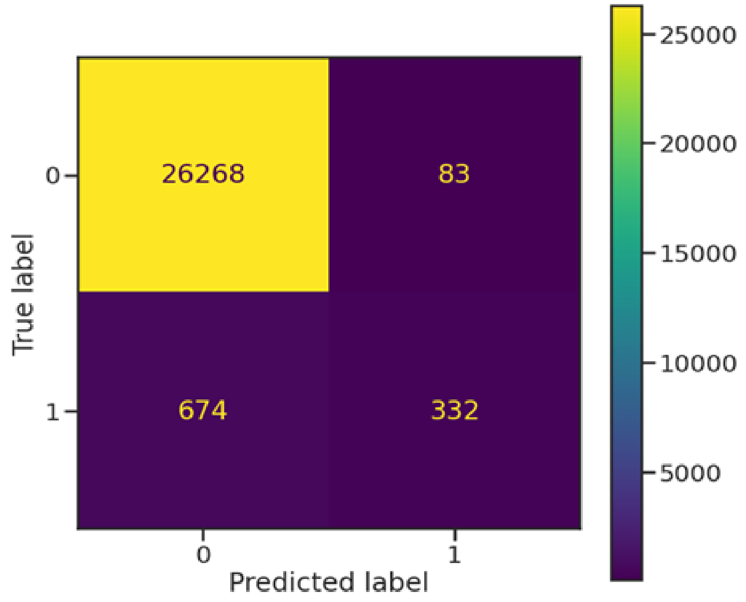
Random Forest

Features

- *Date_numeric*
- *Time_in_seconds*
- *appsicura_bin*
- *fingerprint_bin*
- *bonifico.prodotto_latitude*
- *bonifico.prodotto_longitude*
- *CountryCodeBIC_bin*
- *instantpayment_bin*
- *bonifico.prodotto_importooperazione*
- *divisibile_per_2_bin*
- *divisibile_per_3_bin*
- *divisibile_per_5_bin*
- *divisibile_per_10_bin*
- *parte_decimale_00_bin*
- *parte_decimale_50_bin*
- *parte_decimale_99_bin*

Target

- *bonifico.last_status*
  (**0**→ Genuine, **1**→ Fraud)



	precision	recall	f1-score	support
0	0.97	1.00	0.99	26351
1	0.80	0.33	0.47	1006
accuracy			0.97	27357
macro avg	0.89	0.66	0.73	27357
weighted avg	0.97	0.97	0.97	27357

Class weight:
1) None.
2) balanced.
3) balanced_subsample.

Preliminary results:
Using any of the three methods, up to **33%** of fraudulent transactions can be recovered, with an increase of just **0.3%** in false positives.

Technical Objectives, Methodologies and Solutions

Problem: Existence of repeated operations for the same user, reporting the same amount (bonifico.prodotto_importooperazione) in a short time interval - a few minutes/seconds.

ID	ID_operazione	Date	Time	bonifico.prodotto_importooperazione	bonifico.last_status
4	135	2023-03-20	10:33:53.081000	5963.99	0
4	134	2023-03-17	10:51:02.272000	150.40	0
4	133	2023-03-16	18:19:52.942000	201.86	0
4	132	2023-03-16	15:09:53.121000	488.00	0
4	131	2023-03-15	18:31:56.385000	987.56	0
4	130	2023-03-15	18:31:52.963000	987.56	0
4	129	2023-03-09	16:28:08.559000	5963.99	0

Action on the dataset:

We eliminate for each user the transactions that report the same amount and are repeated in a time interval of 60 minutes, keeping the last one.

- **before cleaning: 90,314**
- **after cleaning: 77,699 (2,695 frauds)**

Technical Objectives, Methodologies and Solutions

Problem: Existence of repeated operations for the same user, reporting the same amount (bonifico.prodotto_importooperazione) in a short time interval – not only a few minutes!

ID	ID_operazione	Date	Time	bonifico.prodotto_importooperazione	bonifico.last_status
2	65	2023-09-05	12:09:15.495000	101.04	0
2	64	2023-09-03	13:02:36.090000	701.62	0
2	63	2023-09-01	10:44:00.751000	668.21	0
2	62	2023-08-09	09:33:48.257000	840.55	0
2	61	2023-08-09	09:28:34.394000	840.55	0
2	60	2023-08-08	20:01:35.914000	840.55	0
2	59	2023-07-30	00:11:14.842000	173.84	0

Technical Objective … Solutions

Problem: Existence of r… , reporting the same amount (bonifico.prodotto_imp… val – not only a few minutes!

ID	ID_operazione	Date	Time	…	…o.last_status
2	65	2023-09-05	12:09:15.49500…		0
2	64	2023-09-03	13:02:36.09000…		0
2	63	2023-09-01	10:44:00.75100…		0
2	62	2023-08-09	09:33:48.25700…		0
2	61	2023-08-09	09:28:34.39400…		0
2	60	2023-08-08	20:01:35.91400…		0
2	59	2023-07-30	00:11:14.84200…		0

Technical Objectives, Methodologies and Solutions

Features temporali
- hour, day, month, day_of_the_week, is_night, is_weekend.

Features geografiche
- latitude, longitude.

Importo
- importooperazione.

Causale
- causale_1, causale_2, causale_3, causale_4, causale_5, causale_6, causale_7, causale_8, causale_9, causale_10.

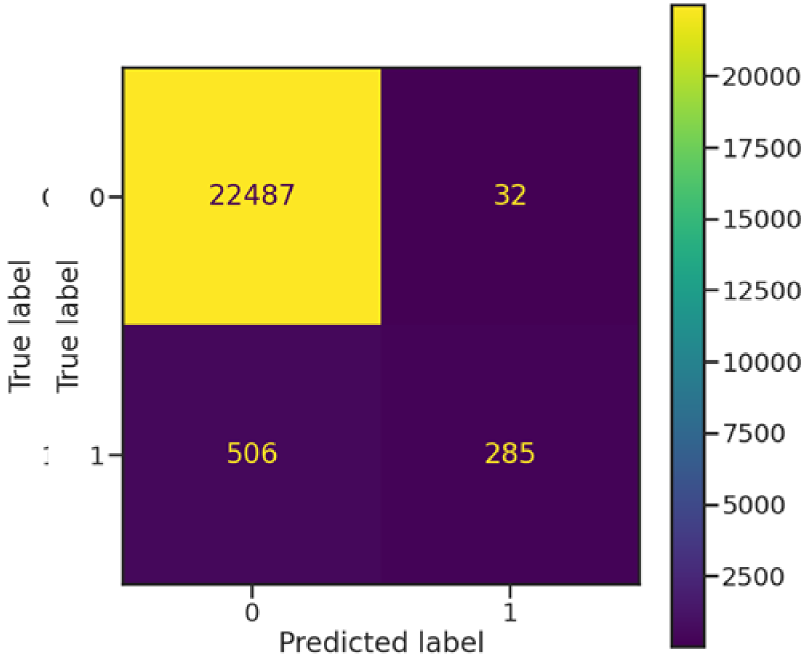
IP
- IP ottenuto dalla conversione e combinazione di octet1, octet2, octet3, octet4.

Features convertite
- accessmode, applicationname, bank_code, clientversion, device_fingerprint, deviceidrba, deviceids, devicemodel, flag_appsicura, flag_fingerprint_certificato, operationsystem, useragent, x_isp_device_id, x_isp_signature, x_isp_tipocliente, x_sec_session_id, InstitutionCode, CountryCodeBIC, LocationCode, BranchCode, codicedivisa, descrizionebanca, divisibile_per_2, divisibile_per_3, divisibile_per_5, divisibile_per_10, parte_decimale_00, parte_decimale_50, parte_decimale_99, instantpayment, carrier_country, carrier_mobile_coutry_code, wifi_is_connected, devicefingerprint.

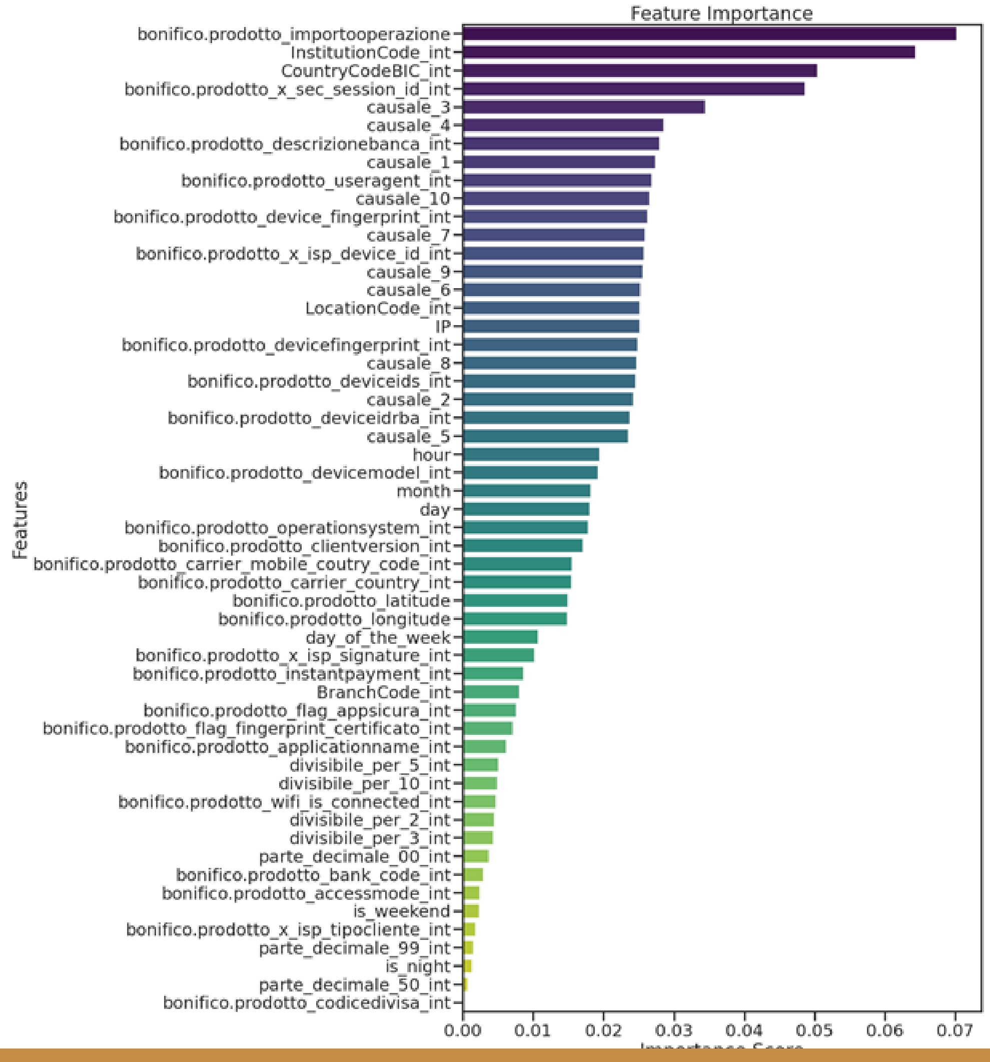
Features escluse
- x_isp_chiaveantifrode, nominativo, numerorapporto, carrier_name.

Target
- bonifico.last_status.

Technical Objectives, Methodologies and Solutions



	precision	recall	f1-score	support
0	0.98	1.00	0.99	22519
1	0.90	0.36	0.51	791
accuracy			0.98	23310
macro avg	0.94	0.68	0.75	23310
weighted avg	0.98	0.98	0.97	23310



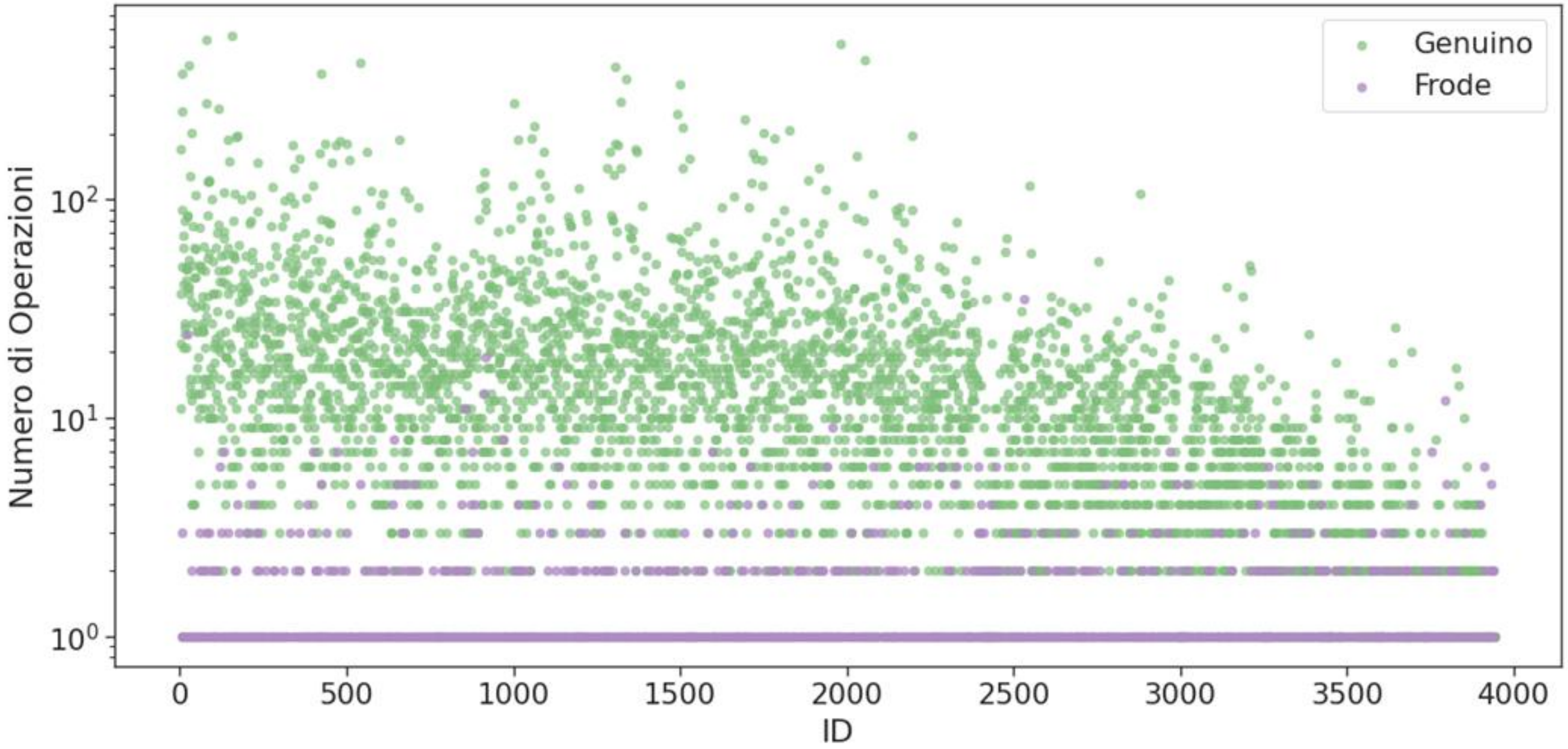
Class weight: None.

Results:

With the clean dataset and the new set of features, it is possible to recover up to 36% of fraudulent transactions, while we reveal only 0.14% of false positives.

Technical Objectives, Methodologies and Solutions

Users & Outliers: Exploration



Average number of operations per user: 19.7

Minimum number of operations per user: 1

Maximum number of operations per user: 565

Filter the dataset by deleting users who have fewer than 20 operations

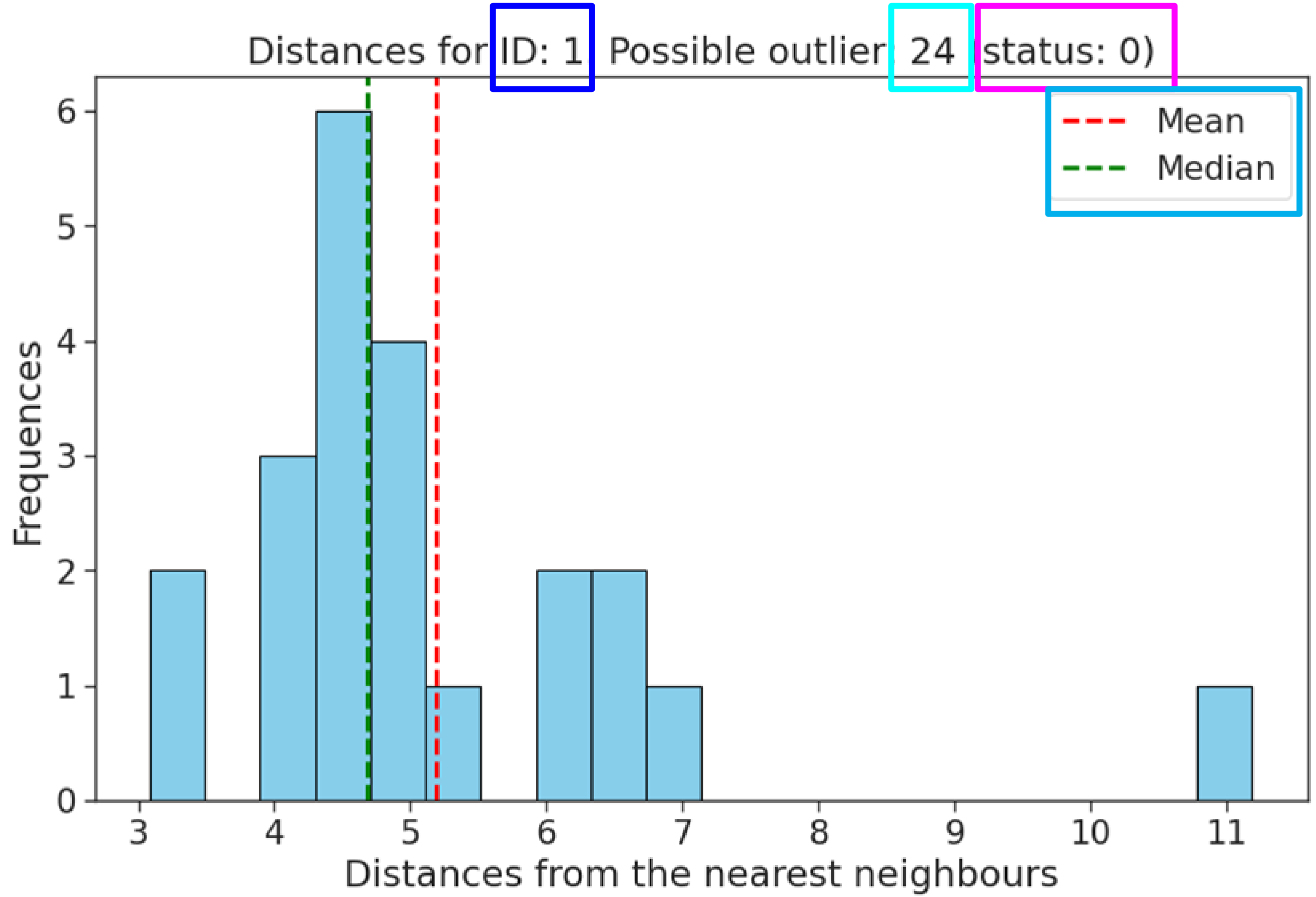
Number of initial users: 3,944

Number of End Users: 1,132

Number of final transactions: 55,804 (878 frauds)

Technical Objectives, Methodologies and Solutions

We use the KNN on each user to identify "anomalous" transactions compared to their "usual" behavior.



User ID

"Anomalous» operation ID: it is the furthest from the other operations of the same user.

Status ISP:

- 0 → genuine
- 1 → fraud

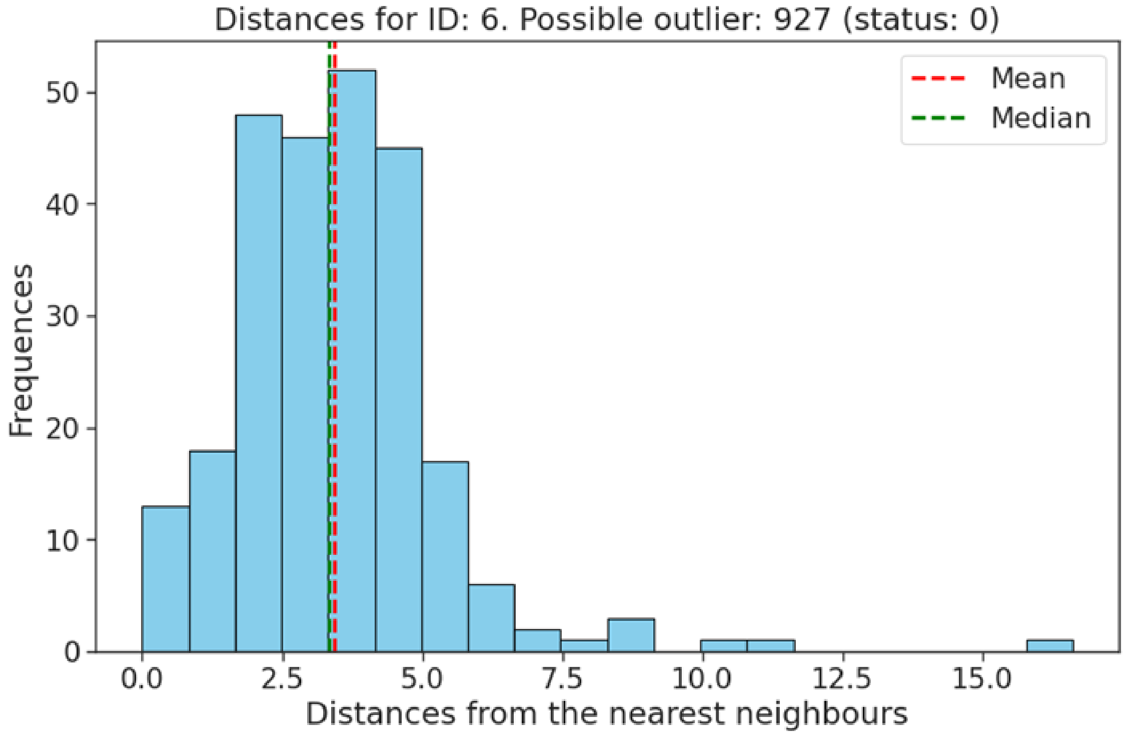
Mean/Median: Threshold to be used to identify outliers (with standard deviation/NMAD).

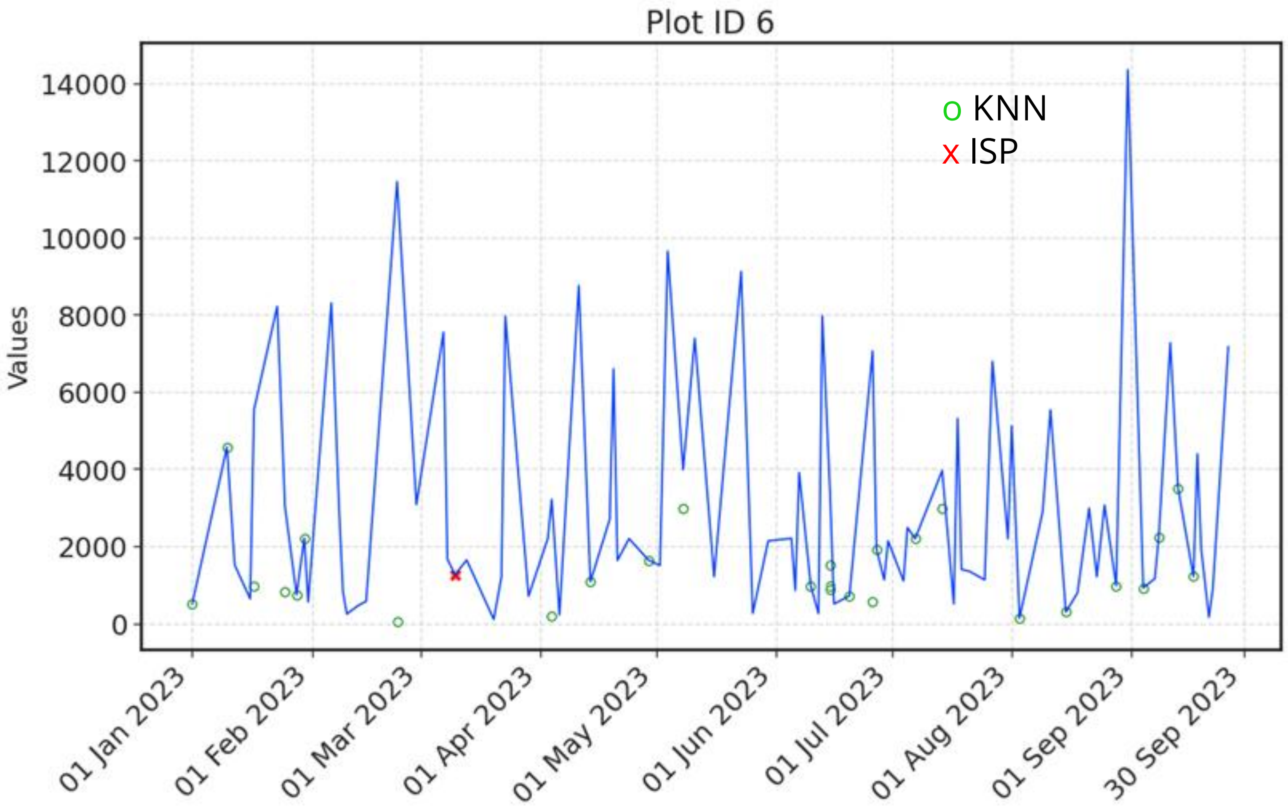
Main Results

We use the KNN on each user to identify "anomalous" transactions compared to their "usual" behavior.

- With the maximum distance:
  - **Fraud identified: 166 (→ 18.9%)**
  - **False negatives: 966**
- With the average+ 1/2/3*std:
  - **Fraud identified: 591 / 393/ 222 / 103 (→ 67.3% / 44.8% / 25.3% / 11.7%)**
  - **False negatives: 23,729/ 6,813/ 1,858 / 594**
- With the median+ 1/2/3*nmad:
  - **Fraud identified: 599 / 429 / 274 / 176 (→ 68.2% / 48.9% / 31.2% / 20.1%)**
  - **False negatives: 26,752 / 10,066 / 3,849 / 1,713**

Main Results





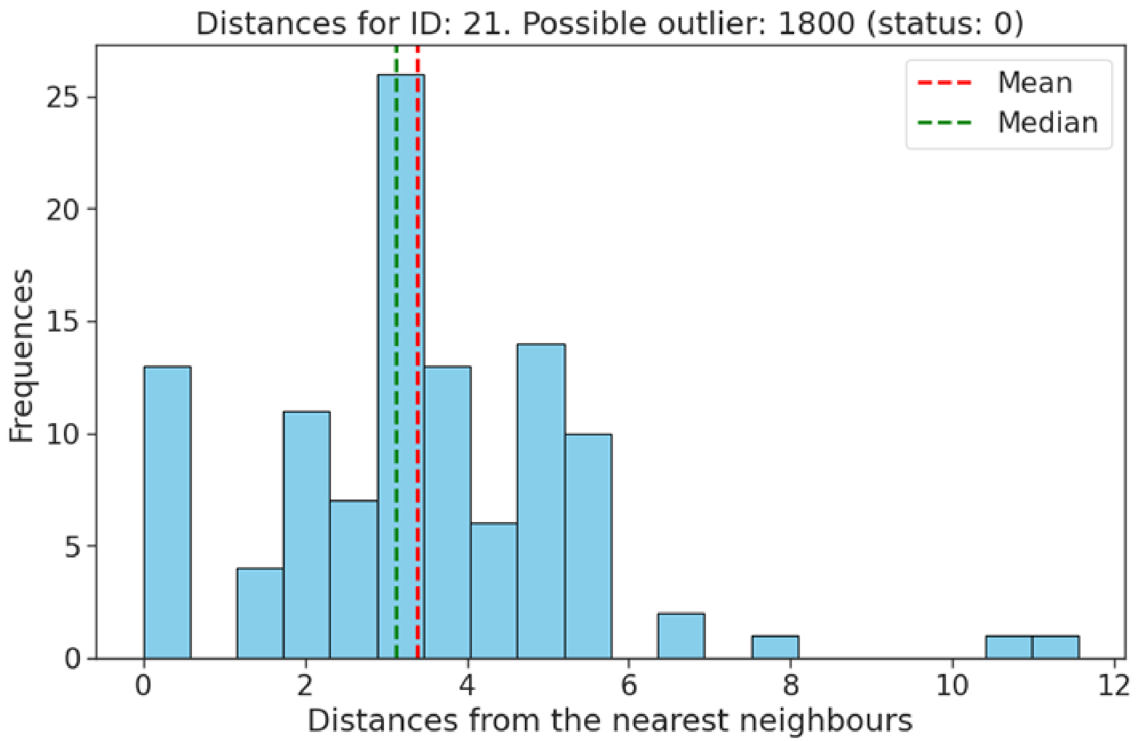
Fraud identified for user 6:
ID_operazione: 768

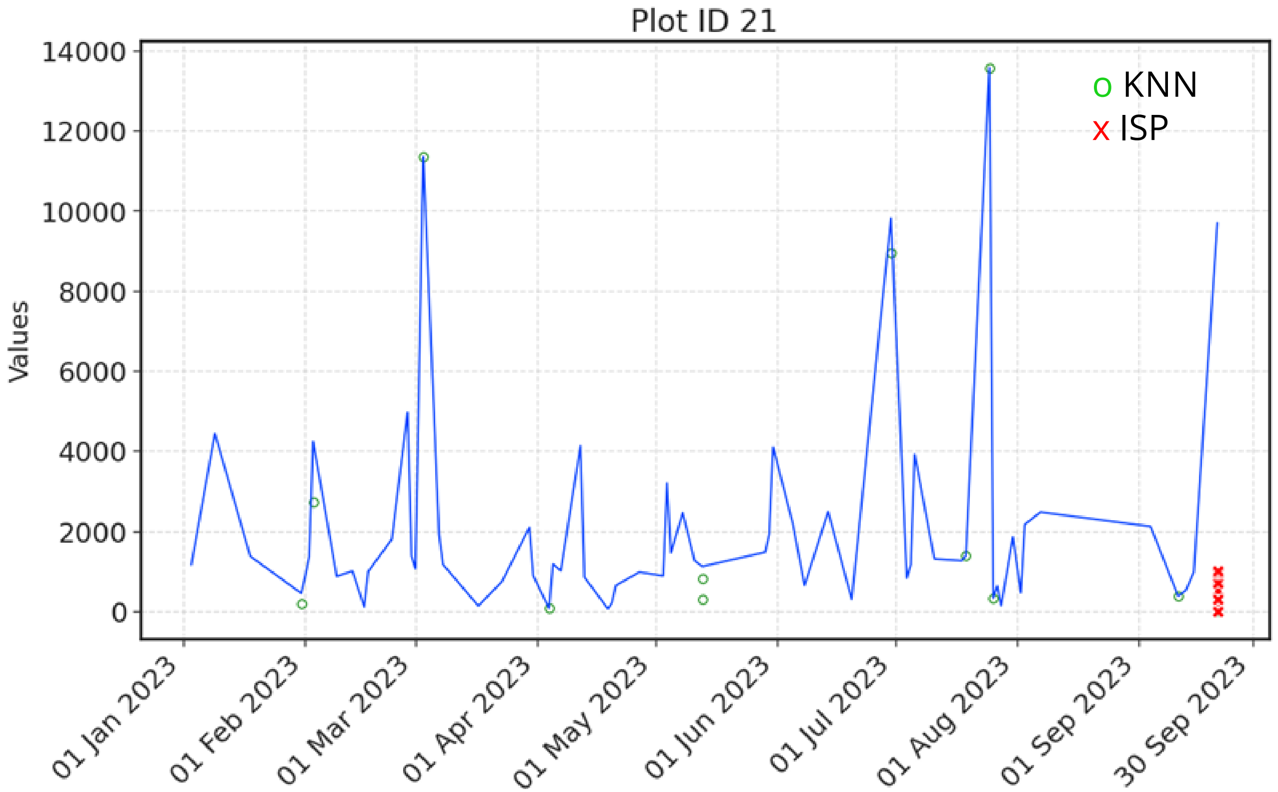
Fraud NOT identified for user 6:
None

Main Results

ID	ID_operazione	Date	Time	bonifico.prodotto_importooperazione	bonifico.last_status	bonifico.prodotto_iban_int	bonifico.prodotto_ibanordinante_int
6	775	2023-03-23	12:48:05.160000	243.08	0	213	258
6	774	2023-03-23	12:38:56.831000	476.02	0	236	258
6	773	2023-03-23	12:32:17.585000	1628.79	0	212	258
6	772	2023-03-23	12:30:34.430000	476.02	0	236	31424
6	771	2023-03-22	10:22:56.484000	1208.60	0	211	258
6	770	2023-03-20	16:18:07.035000	101.04	0	239	219
6	769	2023-03-13	13:09:55.634000	1639.17	0	238	31423
6	768	2023-03-10	17:23:08.411000	1258.45	1	237	219
6	767	2023-03-08	12:53:52.901000	456.42	0	230	258
6	766	2023-03-08	09:52:06.782000	1208.60	0	225	31424
6	765	2023-03-07	11:19:59.639000	2984.27	0	224	31424
6	764	2023-03-07	11:17:51.117000	4565.57	0	205	258
6	763	2023-02-28	12:24:18.097000	1499.30	0	219	31422
6	762	2023-02-28	12:22:40.761000	701.62	0	218	31422

Main Results





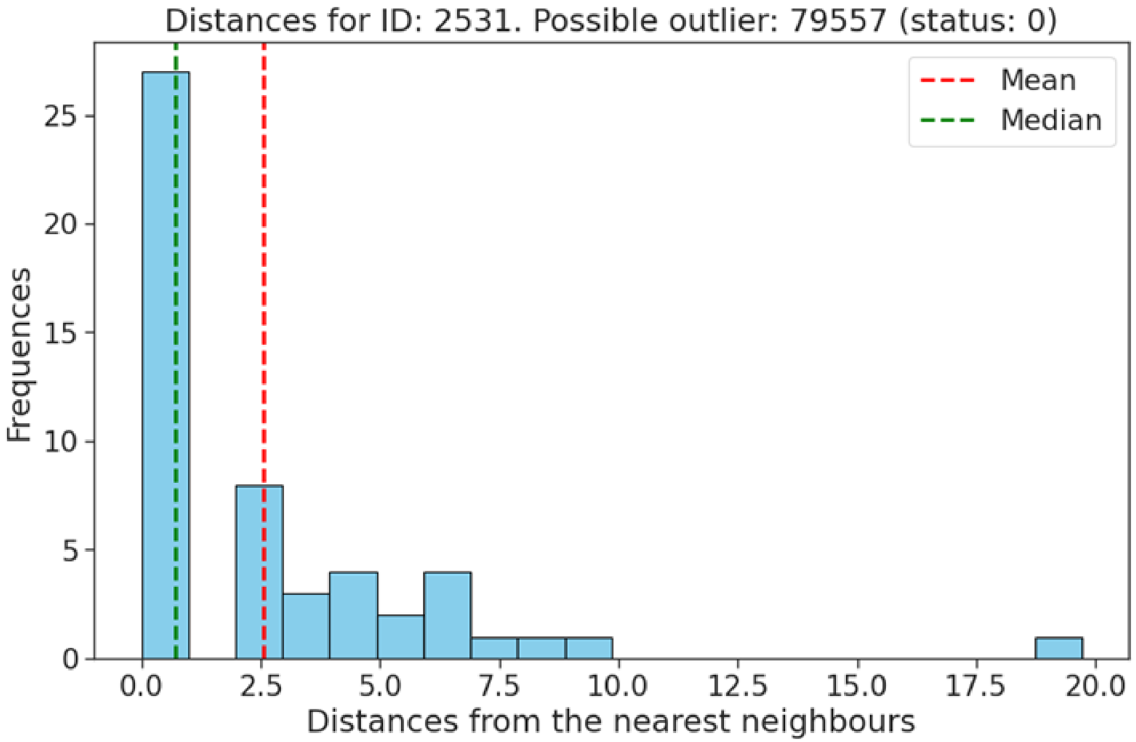
Fraud identified for user 21:
ID_operazione: 1826, 1842

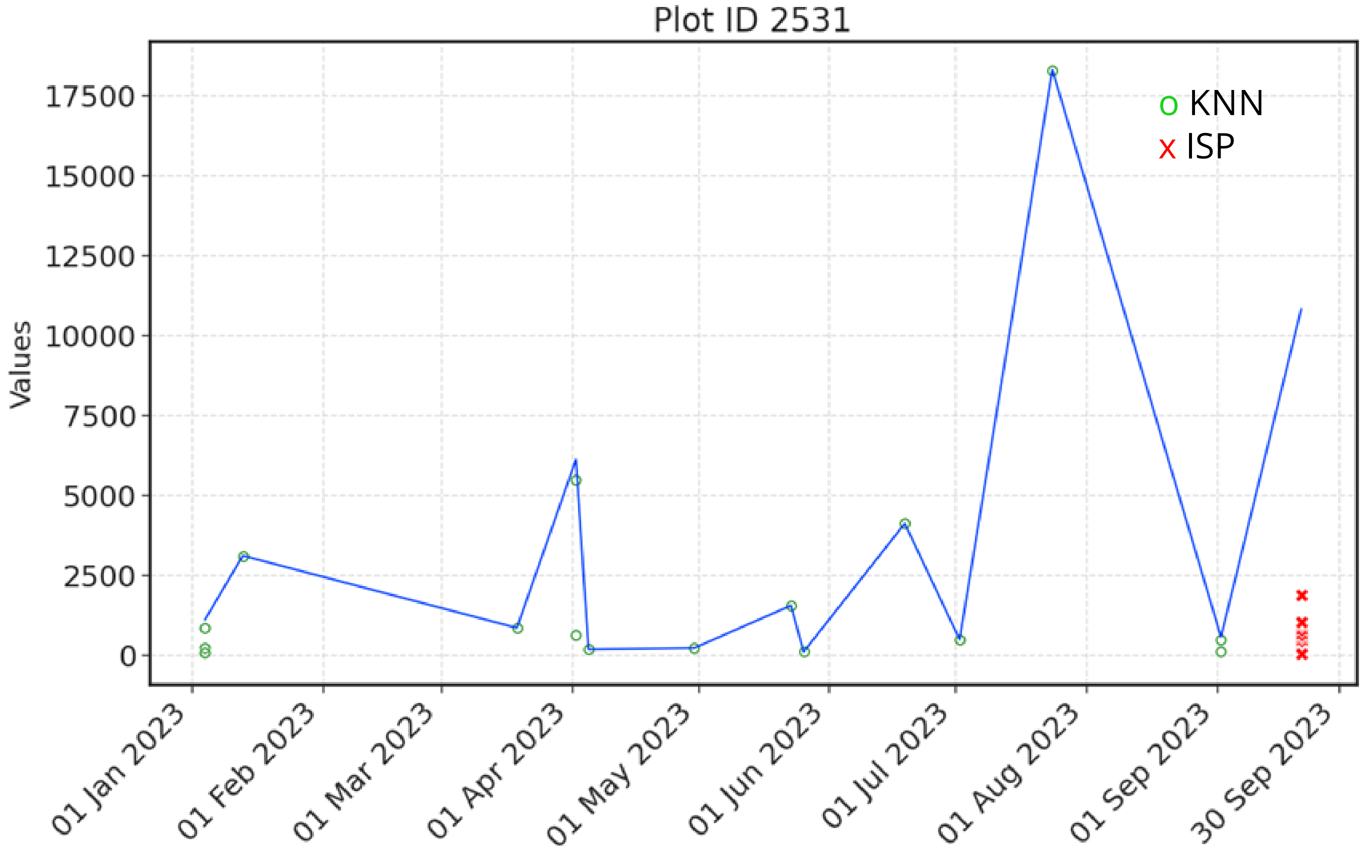
Fraud NOT identified for user 21:
1827, 1831, 1832, 1833, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1843, 1844, 1845, 1847, 1848, 1851, 1852, 1853, 1854, 1855, 1856

Main Results

ID	ID_operazione	Date	Time	bonifico.prodotto_importooperazione	bonifico.last_status	bonifico.prodotto_iban_int	bonifico.prodotto_ibanordinante_int
21	1848	2023-09-21	13:17:16.287000	751.00	1	596	31444
21	1847	2023-09-21	13:17:06.552000	291.06	1	596	31444
21	1845	2023-09-21	13:14:08.152000	504.24	1	594	31444
21	1844	2023-09-21	13:13:53.706000	291.06	1	595	31444
21	1843	2023-09-21	13:12:19.813000	504.24	1	594	31444
21	1842	2023-09-21	13:10:39.433000	1.01	1	552	31444
21	1841	2023-09-21	13:07:23.838000	291.06	1	595	31444
21	1840	2023-09-21	12:57:17.108000	504.24	1	594	31444
21	1839	2023-09-21	12:57:00.550000	291.06	1	592	31444
21	1838	2023-09-21	12:56:51.250000	504.24	1	592	31444
21	1837	2023-09-21	12:54:12.438000	291.06	1	592	31444
21	1836	2023-09-21	12:53:49.645000	10.72	1	594	31444
21	1835	2023-09-21	12:53:34.925000	504.24	1	594	31444
21	1833	2023-09-21	12:48:33.495000	103.70	1	594	31444
21	1832	2023-09-21	12:46:00.244000	10.72	1	593	31444
21	1831	2023-09-21	12:44:23.717000	291.06	1	592	31444
21	1827	2023-09-21	12:40:31.663000	53.04	1	592	31444
21	1826	2023-09-21	12:35:24.565000	291.06	1	591	31444

Main Results

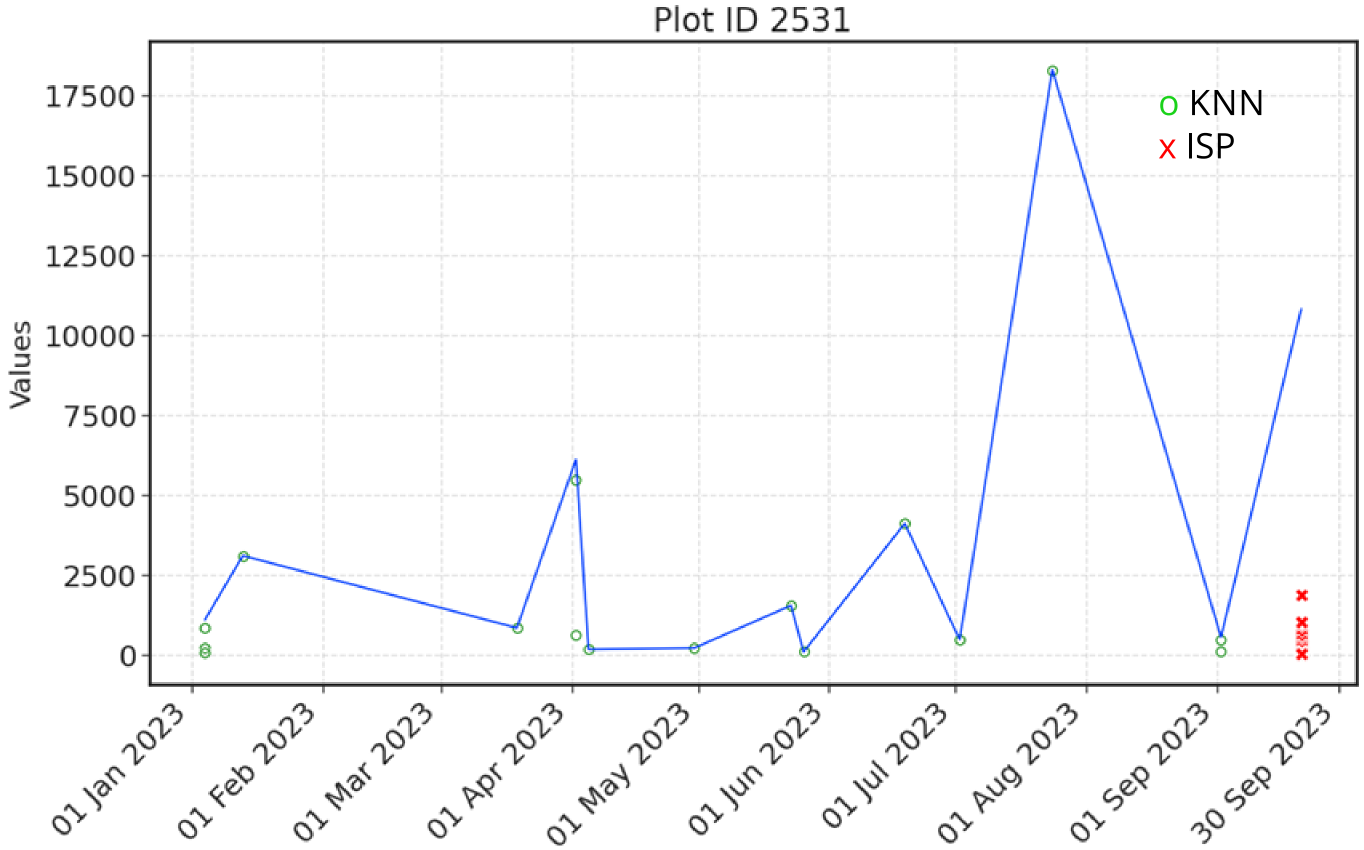




Fraud identified for user 2531:
ID_operazione: None

Fraud NOT identified for user 2531:
79571, 79572, 79573, 79574, 79575, 79576, 79577, 79578, 79579, 79580, 79581, 79582, 79583, 79584, 79585, 79586, 79587, 79588, ...

Main Results



Fraud identified for user 2531:
ID_operazione: None

Fraud NOT identified for user 2531:
79571, 79572, 79573, 79574, 79575, 79576, 79577, 79578, 79579, 79580, 79581, 79582, 79583, 79584, 79585, 79586, 79587, 79588, ...

Main Results

ID	ID_operazione	Date	Time	bonifico.prodotto_importooperazione	bonifico.last_status	bonifico.prodotto_iban_int	bonifico.prodotto_ibanordinante_int
531	79588	2023-09-21	05:02:30.796000	291.06	1	593	34574
531	79587	2023-09-21	05:01:57.678000	1011.25	1	593	34574
531	79586	2023-09-21	05:01:48.157000	10.19	1	593	34574
531	79585	2023-09-21	04:59:33.525000	90.84	1	593	34573
531	79584	2023-09-21	04:59:02.675000	10.19	1	591	34573
531	79583	2023-09-21	04:56:57.805000	504.24	1	594	34573
531	79582	2023-09-21	04:56:14.940000	526.51	1	594	34573
531	79581	2023-09-21	04:50:19.885000	556.15	1	593	34573
531	79580	2023-09-21	04:49:33.261000	367.34	1	591	34574
531	79579	2023-09-21	04:48:34.801000	499.76	1	593	34573
531	79578	2023-09-21	04:47:33.858000	367.34	1	593	34573
531	79577	2023-09-21	04:47:06.975000	668.21	1	591	34574
531	79576	2023-09-21	04:45:29.358000	362.07	1	591	34573
531	79575	2023-09-21	04:44:57.449000	1879.75	1	593	34574
531	79574	2023-09-21	04:44:31.012000	362.07	1	593	34574
531	79573	2023-09-21	04:44:17.520000	17.54	1	593	34574
531	79572	2023-09-21	03:50:29.845000	156.59	1	592	34574
531	79571	2023-09-21	03:25:14.981000	291.06	1	591	34573

Main Results

By adopting the Random Forest on the full feature set, it is possible to recover up to **36%** (compared to 33% in past experiments) of fraudulent transactions, while we reveal only **0.14%** (0.3%) of false positives.

By adopting the KNN restricted to the individual user on the complete set of features, it is possible to recover up to **48%** of fraudulent transactions. Nevertheless, the number of false positives is high.

Final Steps

Our work in progress:	Adoption of more specific features to describe the user.
	Detailed analysis of false positives to improve the performance of the method.
	Combination of more approaches.

Thank you for your attention!



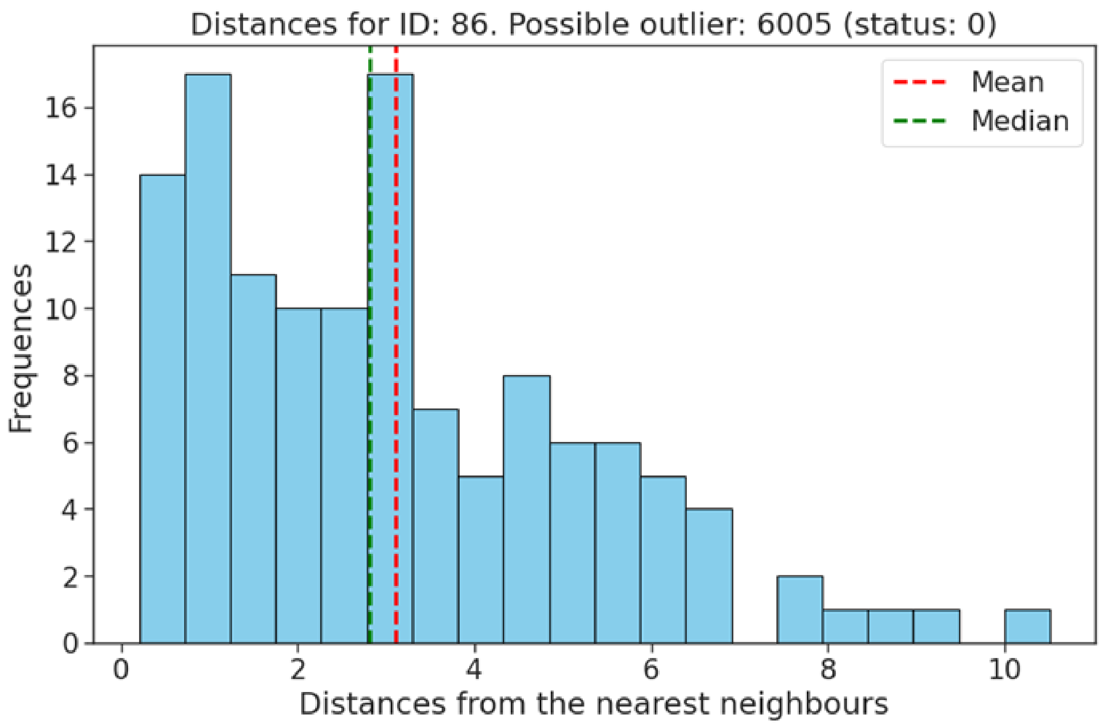


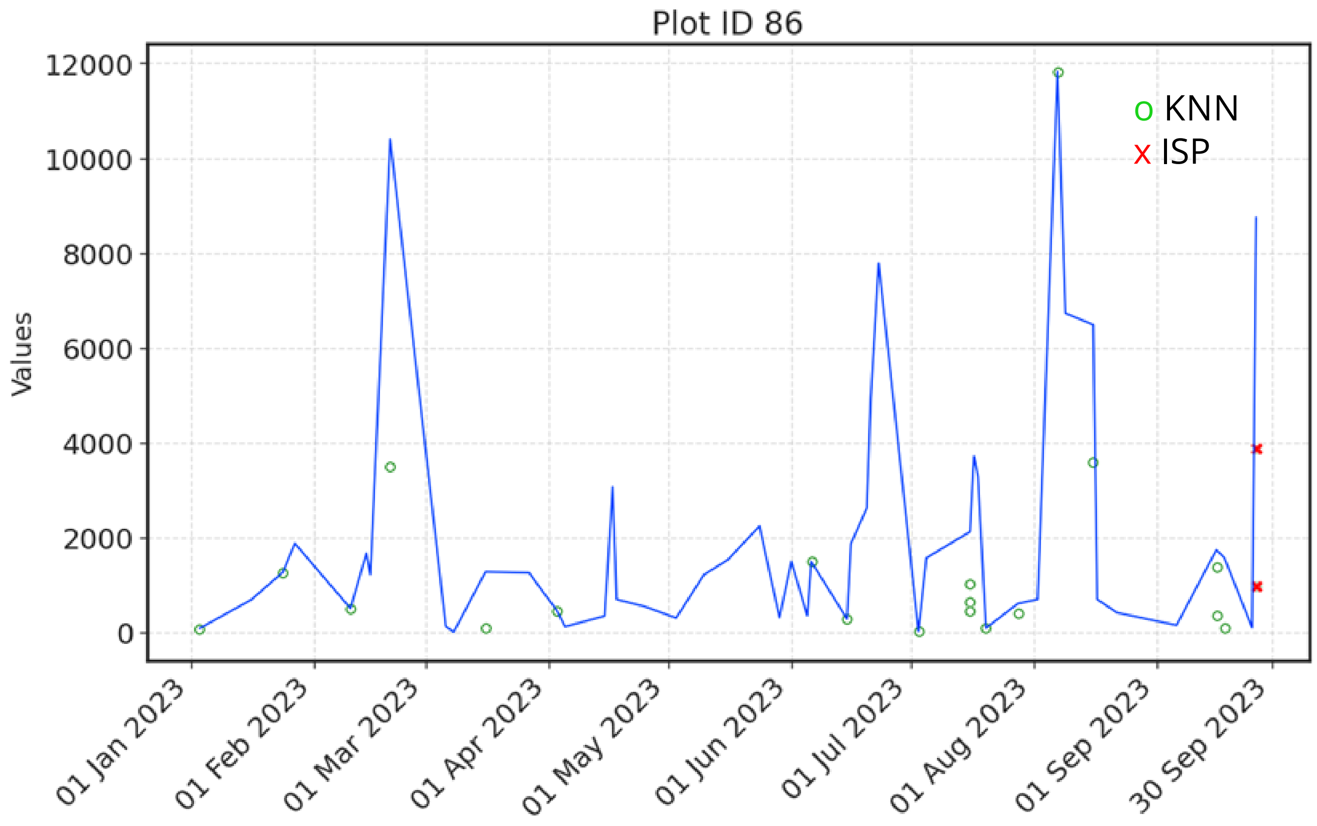
ylenia.maruccia@inaf.it
riccardo.crupi@intesasanpaolo.com

Thank you for your attention!

ylenia.maruccia@inaf.it
riccardo.crupi@intesasanpaolo.com

Main Results





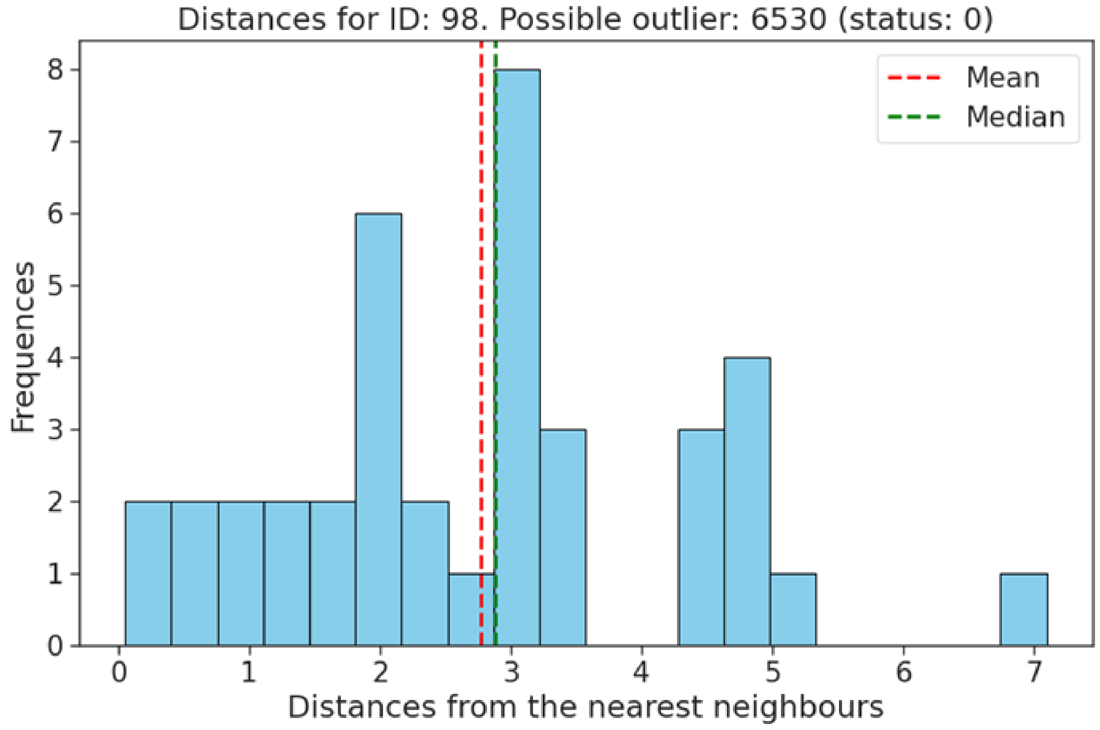
Fraud identified for user 86:
ID_operazione: 6070, 6071, 6072

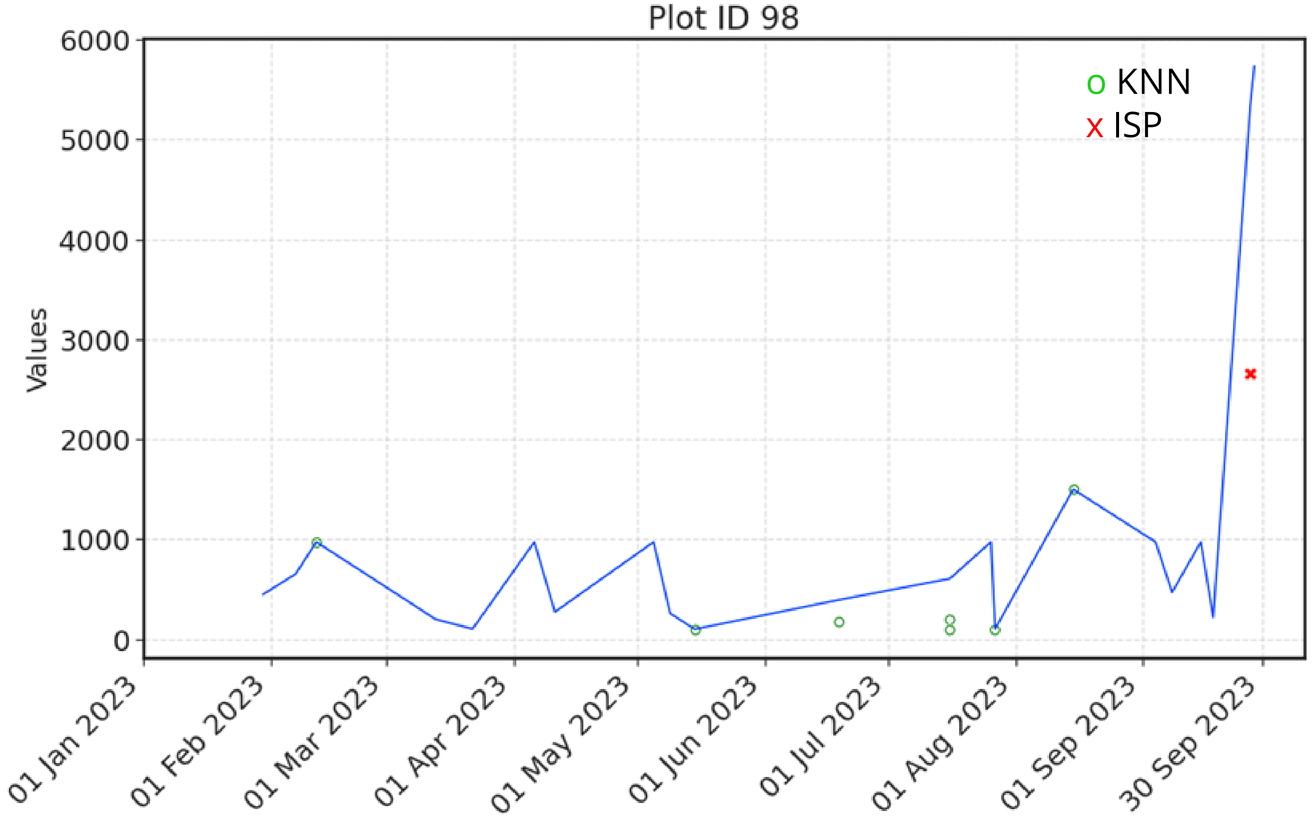
Fraud NOT identified for user 86:
None

Main Results

ID	ID_operazione	Date	Time	bonifico.prodotto_importooperazione	bonifico.last_status	bonifico.prodotto_iban_int	bonifico.prodotto_ibanordinante_int
86	6072	2023-09-26	16:43:51.239000	3890.95	1	1969	1961
86	6071	2023-09-26	16:38:39.547000	970.90	1	1968	31542
86	6070	2023-09-26	16:37:06.230000	3893.05	1	1967	1961
86	6069	2023-09-25	11:33:08.709000	105.53	0	1966	1936
86	6068	2023-09-18	20:55:10.754000	262.52	0	1931	1961
86	6067	2023-09-18	20:37:45.250000	243.17	0	1931	31543
86	6066	2023-09-18	20:34:03.431000	455.83	0	1940	31543

Main Results

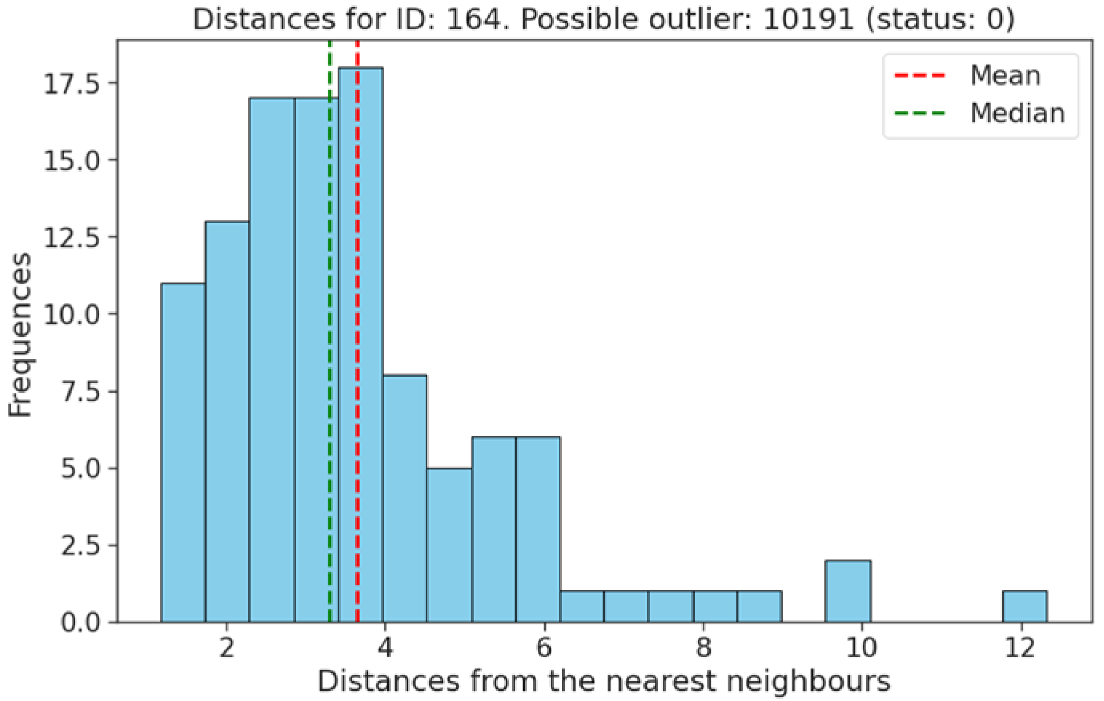


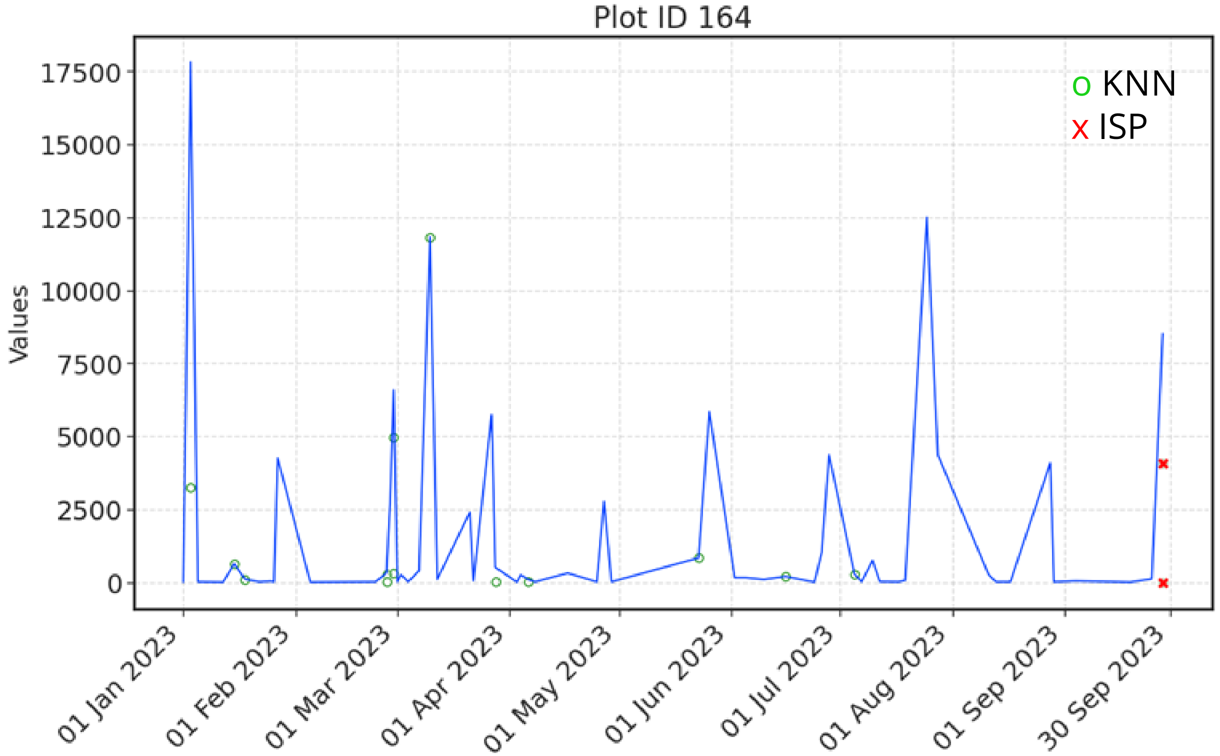


Fraud identified for user 98:
ID_operazione: 6551, 6552

Fraud NOT identified for user 86:
None

Main Results





Fraud identified for user 164:
ID_operazione: 10281, 10282

Fraud NOT identified for user 164:
None