FRAUDPANDA -- COUNTERACTING FRAUD USING PRO-ACTIVE (NETWORK) DETECTION AND ANALYSIS

A Data Management Plan created using DMPonline.be

Creators: Bruno Deprez, Wouter Verbeke

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Bruno Deprez

Project Administrator: Wouter Verbeke

Grant number / URL: 1SHEN24N

ID: 204431

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:
Financial transaction fraud remains an important societal challenge, despite decades of research on automatic fraud detection systems. In light of these struggles, I propose to tackle fraud from a different perspective: rather than reactively detecting fraud after it occurs, I aim to proactively prevent fraud before it even happens. The envisioned research project, FraudPANDA, aims to counter-act fraudsters by using Pro-Active (Network) Detection and Analysis. Firstly, I aim to proactively detect who is at risk of becoming targeted before the attack happens. I will develop methods to find the relevant customer characteristics. Secondly, I aim to analyse different preventive interventions and analyse their efficacy at decreasing a customer's fraud risk. This will be based on insights from causal inference. Thirdly, for both methods, I will also explore the possibility of using network information to improve prevention performance. This is motivated by recent success of network analysis in fraud detection. In the end, the ultimate goal of FraudPANDA is an innovative solution that operationalizes these insights by optimally allocating preventive interventions and minimizing fraudulent activities. A field test will be conducted at an industrial partner to evaluate FraudPANDA's ability to proactively prevent fraud in real-world scenarios.

Last modified: 12-02-2024

FRAUDPANDA -- COUNTERACTING FRAUD USING PRO-ACTIVE (NETWORK) DETECTION AND ANALYSIS

DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

FRAUDPANDA -- COUNTERACTING FRAUD USING PRO-ACTIVE (NETWORK) DETECTION AND ANALYSIS

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

FRAUDPANDA -- COUNTERACTING FRAUD USING PRO-ACTIVE (NETWORK) DETECTION AND ANALYSIS

Application DMP

QUESTIONNAIRE

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

For the testing of methods, mostly open-source and synthetic data will be used. These consist of either fully anonymised or fictive monetary transactions. For the field test in WP5, I will collaborate with financial institutions to test the newly introduced methods to real-world transaction data.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

The data used coming from collaborations with third party financial institutions will not be stored on KU Leuven side. These will be extensions of existing research chairs and/or collaborations already present in the research group. Here, access to the data is only granted by IT equipment of the corresponding institution.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

The data used coming from collaborations with third party financial institutions will not be stored on KU Leuven side. These will be extensions of existing research chairs and/or collaborations already present in the research group. Here, access to the data is only granted by IT equipment of the corresponding institution.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

At the moment, all personal or sensitive data will not be stored on KU Leuven side. Access to the data is only granted via IT equipment of the collaborating institution. These financial institutions have very stringent data protection policies in place, and all data/results leaving the institution needs to be screened first by their legal departments to guarantee that no personal or sensitive data will be leaked.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

n/a

FRAUDPANDA -- COUNTERACTING FRAUD USING PRO-ACTIVE (NETWORK) DETECTION AND ANALYSIS

FWO DMP (Flemish Standard DMP)

1. RESEARCH DATA SUMMARY

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
Transaction data	Full set of transaction data over different time periods of clients of a financial institution.	Generate new data Reuse existing data	Digital	*Please choose from the following options:* • Observational • Compiled/ aggregated data	*Please choose from the following options:* • .parquet	*Please choose from the following options:* • <1 TB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

This data is provided via de project conducted under the BNP PF research chair in fraud analytics (https://www.kuleuven.be/fondsenwerving/hoe-steunen/leerstoelen/groepsoverschrijdend/bnp-paribas-fortis-chair-in-fraud-analytics)

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

In the final work package, transaction data is used to identify who is most susceptible to fall victim to fraud. The application of interventions might include bias against certain people groups, due to historical bias in the dataset.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

The data contains a person's transaction history at the bank. This may include a name and surname of both the sending and receiving persons.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

The bank that is owner of the data provided might also implement the developed methods for commercial exploitation, to prevent fraud cases for their clients.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

The data used and the data we plan to use will be provided via the BNP PF research chair (https://www.kuleuven.be/fondsenwerving/hoe-steunen/leerstoelen/groepsoverschrijdend/bnp-paribas-fortis-chair-in-fraud-analytics). Access is only granted via their IT equipment.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- Yes

The owner of the data is the financial institution. The access to the data is restricted, and will not be made available for reuse.

2. DOCUMENTATION AND METADATA

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Accompanying information will consist of descriptive statistics, including average amounts people transfer, number of transfer etc. There will be no detailed documentation about the data, given the sensitive nature of the data.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

3. DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT

Where will the data be stored?

On the side of the financial institution.

How will the data be backed up?

On the side of the financial institution.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

On the side of the financial institution.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

This is done via the security that is in place at the financial institution. Access will be via their IT systems, and no sharing of data outside the company will be done.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

On the side of the financial institution.

4. DATA PRESERVATION AFTER THE END OF THE RESEARCH PROJECT

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

We will not store that data as part of the project. This will be done on the side of the financial institution, as we only will be given access to the data under the BNP PF research chair.

Where will these data be archived (stored and curated for the long-term)?

On the side of the financial institution.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

On the side of the financial institution.

5. DATA SHARING AND REUSE

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- No (closed access)

If access is restricted, please specify who will be able to access the data and under what conditions.

On the side of the financial institution.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Privacy aspects

Highly sensitive transaction data.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

N/A

When will the data be made available?

N/A

Which data usage licenses are you going to provide? If none, please explain why.

Fully restricted data

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

What are the expected costs for data sharing? How will these costs be covered?

N/A

6. RESPONSIBILITIES

Who will manage data documentation and metadata during the research project?

Bruno Deprez

Who will manage data storage and backup during the research project?

On the side of the financial institution.

Who will manage data preservation and sharing?

On the side of the financial institution.

Who will update and implement this DMP?

Bruno Deprez