

Jautrių duomenų apsauga naudojant DI įrankius

Justas Kalpokas

Kas blogo gali nutikti?



Kas blogo gali nutikti?



Korteles numeris : 1234 5678 9012 3456

Kitos korteles numeris: 1234 5678 9012 3457

Dar kitos korteles numeris: 1234 5678 9012 3458

Kas blogo gali nutikti?

Lookup

Type:

All

Name:

bin laden

ID # / Digital Currency Address:

Program:

All

Minimum Name Score:

100

Address:

City:

State/Province*:

Country:

All

List:

All

Search

Reset

Lookup Results: 7 Found

Name	Address	Type	Program(s)	List	Score ▼
BIN LADEN_Hamza		Individual	SDGT	SDN	100
BIN LADEN_Osama		Individual	SDGT	SDN	100
BIN LADEN_Sa'ad		Individual	SDGT	SDN	100
BIN LADEN_Sad		Individual	SDGT	SDN	100
BIN LADEN_Usama		Individual	SDGT	SDN	100
USAMA BIN LADEN NETWORK		Entity	FTO; SDGT	SDN	100
USAMA BIN LADEN ORGANIZATION		Entity	FTO; SDGT	SDN	100

Tikras atvejis. Bankų sektorius

Norime užkirsti kelią sukčiams, kurie vilioja pinigus iš banko klientų.

Du scenarijai:

1. Sukčiai įtikina banko klientą pervesti didelę pinigų sumą.
2. Sukčiai išvilioja prisijungimo duomenis ir patys inicijuoja pavedimą.

Tikras atvejis. Bankų sektorius

Norime užkirsti kelią sukčiams, kurie vilioja pinigus iš banko klientų.

Du scenarijai:

1. Sukčiai įtikina pervesti jiems didelę sumą pinigų.
2. Sukčiai išvilioja prisijungimo duomenis ir patys inicijuoja pavedimą.

Populiariausi sukčiavimo būdai:

1. Romantiniai santykiai
2. Investavimas ir greito uždarbio pažadas

Romantiniai santykiai

Penkiasdešimtmetė prancūzė Anne tapo sukčių auka. Sukčiai apsimetė sergančiu Bradu Pittu ir įtikino ją, kad Bradas ją myli. Naudodami dirbtiniu intelektu sukurtas nuotraukas, apgavikai išviliojo 830 000 eurų. Anne istorija buvo parodyta per Prancūzijos televizijos kanalą TF1, įspėjant apie tokias aferas.

Brad Pitt

This article is more than 2 months old

Brad Pitt reacts to 'awful' scammers who fooled French woman using his pictures

Anne, 53, was scammed out of hundreds of thousands after believing she was in a long-term relationship with the actor

Anna Betts
Thu 16 Jan 2025 20:19 CET

Share



Bankų motyvacija

Scams

● This article is more than **1 year old**

UK banks to reimburse fraud victims under new rules, regulator confirms

Requirement to refund people who have been tricked by scammers will be implemented in 2024

● [Business live - latest updates](#)

Jess Clark

Wed 7 Jun 2023 16:17 CEST



Share



Banking

● This article is more than **6 months old**

Explainer

Why is the UK slashing the maximum banks must refund to fraud victims?

Consumer groups hit out at regulator's 'outrageous' decision to cut limit from £415,000 to £85,000

Rupert Jones

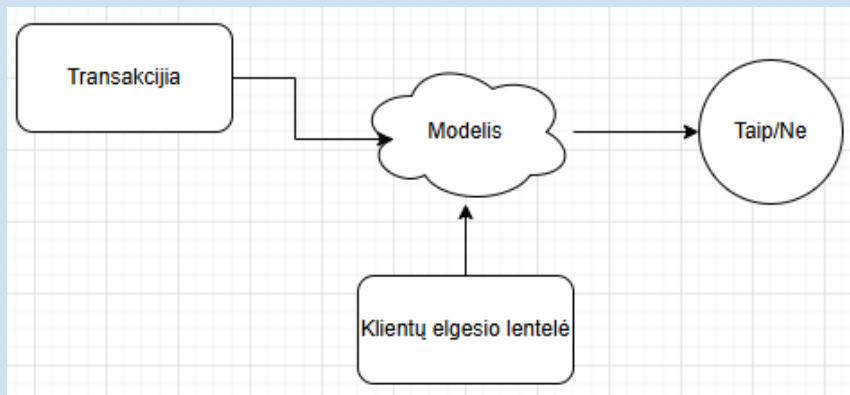
Wed 4 Sep 2024 16:04 CEST



Dirbtinis intelektas

Tikslas. Norime užkirsti kelią sukčiams, kurie vilioja pinigus iš banko klientų.

Sprendimas. Automatinis duomenų klasifikavimas. T.y. Machine Learning modelis, kuris automatiškai klasifikuoja transakcijas.



Duomenys

1. Klientai (Customers)

ID	Vardas	Pavardė	El. paštas	Telefonas	Adresas	Sukūrimo data
1	Jonas	Petrauskas	jonas@email.lt	+37060000001	Vilnius, LT	2023-01-01

2. Sąskaitos (Accounts)

ID	Kliento ID	Sąskaitos numeris	Tipas	Balansas	Sukūrimo data	Statusas
1	1	LT6010100123456789	Einamoji	5000.00	2023-01-01	Aktyvi

3. Operacijos (Transactions)

ID	Sąskaitos ID	Data	Suma	Operacijos tipas	Gavėjo sąskaita	Aprašymas	Statusas
1	1	2024-03-22	-200.00	Pavedimas	LT6010100987654321	Sąskaitos apmokėjimas	Vykdyta

Modelio kūrimas

Modelio treniravimui skirtos lentelės

1. Transakcijų lentelė (Transactions)

ID	Kliento ID	Sąskaitos ID	Data	Suma	Operacijos tipas	Gavėjo sąskaita	Aprašymas	Statusas
1	1	1	2024-03-22	-200.00	Pavedimas	LT6010100987654321	Sąskaitos apmokėjimas	Vykdyta
2	2	2	2024-03-21	300.00	Pavedimas	LT6010101122334455	Dovanų pervedimas	Vykdyta

2. Klientų elgesio lentelė (Customer Behavior)

ID	Kliento ID	Paskutinė prisijungimo data	Prisijungimo dažnumas	Transakcijų skaičius	Bendros transakcijos suma	Vidutinė transakcijos suma	Sumu variacija	Didžiausia transakcija	Mažiausia transakcija	Tarpinis suma per paskutinius 30 d.	Paskutinė operacija	Prekės/Paslaugos dažniausiai perkamų
1	1	2024-03-20	10	25	10,000.00	400.00	1200.00	1000.00	50.00	1500.00	1	Kredito kortelės, paskolos
2	2	2024-03-15	5	15	4,500.00	300.00	800.00	800.00	20.00	1200.00	3	Taupymo sąskaitos, pervedimai

Jautrūs duomenys

- Vardas
- Pavardė
- Telefono numeris
- El. Paštas
- Adresas
- Sąskaitos numeris
- Transakcijos aprašymas
- Pavedimo gavėjas
- ...

Pagrindiniai iššūkiai

1. Jūsų teisininkų leidimas naudoti duomenis.
 - a. Modelio naudojimo pagrindinimas.
 - b. Klientų sutikimas naudoti duomenis.
 - c. Asmens duomenų apsaugos atitikimas.

Pagrindiniai iššūkiai

1. Jūsų teisininkų leidimas naudoti duomenis.
 - a. Modelio naudojimo pagrindinimas.
 - b. Klientų sutikimas naudoti duomenis.
 - c. Asmens duomenų apsaugos atitikimas.
2. Jeigu duomenys bus pavogti programišių, turi būti neįmanoma indentifikuoti klientų.
 - a. Saugoti asmeninę klientų informaciją atskirose lentelėse. Treniravimo duomenyse naudoti tik atsitiktinai sugeneruotus ID.
 - b. Iš laisvo teksto pašalinti asmenų vardus, adresus ir kitą jautrią informaciją.
 - c. Jeigu naudojami geolokacijos duomenys, galima įterpti triukšmo, kad nebūtų galima tiksliai indentifikuoti kliento buvimo vietos.

Asmeninės informacijos šalinimas

- Reikia pašalinti - Personally Identifiable Information (PII).
- Naudojam pasirinktą atvirą biblioteką, kuri geba atlikti named entity recognition (NER).

Triukšmo įterpimas - Kaip tai išsprendė BPC

Bendrojo pagalbos centro (BPC) gauti pranešimai numeriu 112.

Bendrosios pagalbos centro (BPC) registruoti įvykiai, apie kuriuos gauti pranešimai pagalbos telefonu 112.

Pateikiamos "triukšmintos" įvykių koordinatės (paslenkant tikrąsias koordinates atsitiktine kryptimi taikant normalųjį triukšmo skirstinį su standartiniu nuokrypiu nuo 100 iki 1000 metrų, priklausomai nuo gyventojų tankio: mažesnis gyventojų tankis – didesnis nuokrypis

Didieji kalbos modeliai

Padėtų sukurti papildomų kokybinių stulpelių.

Pvz.

- Ar transakcijos aprašymas išsamus?
- Ar suma atitinka transakcijos aprašymą?
- ...

Ką galime naudoti:

1. Offline modeliai, kurie veikia jūsų serveriuose.
2. Dedikuoti modeliai, kurie parduodami kaip paslauga. Pvz. OpenAI chatGPT Azure debesyje.

DEV ir PROD aplinkų atskyrimas

1. Pirminę duomenų analizę ir modelio prototipo kūrimą atlikti dedikuotoje aplinkoje
 - a. AWS SageMaker
 - b. Azure Machine Learning
 - c. Azure DataBricks
2. Modelių kūrimui naudoti paruoštus treniravimui duomenis.
3. Nepradėti naudoti sintetinių duomenų modelio kūrimui.

Kuriant modelius svarbu

Pasirinkti teisingą metriką, t.y. ką mes norime maksimizuoti?

Teoriniai skaičiai:

- Transakcijų skaičius per dieną 1000 000.
- Apgaulės statistika 0.1% → 1000 transakcijų yra apgaulingos.
- Dėl turimų darbuotojų skaičiaus galime patikrinti 500.

Kuriant modelius svarbu

Pasirinkti teisingą metriką, t.y. ką mes norime maksimizuoti?

Teoriniai skaičiai:

- Transakcijų skaičius per dieną 1000 000.
- Apgaulės statistika 0.1% → 1000 transakcijų yra apgaulingos.
- Dėl turimų darbuotojų skaičiaus galime patikrinti 500.

Maksimizuojam sustabdytų apgaulingų transakcijų pinigų sumą.

Geriau sustabdyti 1 transakciją vertą 100 000 negu 10 vertų po 5000.

Nuorodos

1. Sankcijuotų asmenų sąrašai <https://sanctionssearch.ofac.treas.gov/>
2. Bradas Pitas <https://www.theguardian.com/film/2025/jan/16/brad-pitt-french-woman-scam>
3. Bankų isipareigojimai apgaulės atveju <https://www.theguardian.com/business/article/2024/sep/04/why-is-the-uk-slashing-the-maximum-banks-must-refund-to-victims>
4. 112 iškvietimų statistika <https://data.gov.lt/datasets/1763/>
5. NER biblioteka, named entity recognition - <https://github.com/flairNLP/flair>
6. Azure Machine Learning <https://azure.microsoft.com/en-us/products/machine-learning>