Redukcia dimenzie blockchainových dát pre sieť Etherea

Adam Martinka

2023

1 Úvod

1.1 Intro do blockhainu

Pred samotným uvedením cieľu projektu alebo predstavením samotných dát, považujeme za potrebné si vysvetliť aspoň tie najnutnejšie koncepty v oblasti blockchainu pre porozumeniu dát a výsledkov v projekte.

Sieť Etherea sa dá pochopiť ako p2p - peer to peer (tj. decentralizovaná) sieť, ktorá validuje rôzne transakcie medzi ľubovolnými účastníkmi siete. Pre rýchle porozumenie môžeme uviesť jednoduchý príklad: užívateľ A chce poslať užívateľovi B svoje **prostriedky** (napr. 1 ETH) cez sieť Etherea. Ak užívateľ A má vytvorenú tzv. peňaženku a pozná **addresu** (tj. ID) peňaženky užívateľa B, tak po zaplatení gas fees (poplatkov) sa vykoná daná požiadavka na presun prostriedkov. Addresa peňaženky je voľne vyhladateľná ale je aj kompletne anonymná do bodu, pokiaľ vlastník peňaženky neprezradí jeho vlastníctvo danej peňaženky verejnosti (Tu zrejme A vie, že peňaženka patrí B). Potvrdenie a rôzne validácie zabezpečuje samotná sieť. Ak požiadavka prejde cez sieťové validácie, tak prenos prostriedkov bude do malej chvíle schválený a aktualizovaný.

Pod pojmom 'sieť' je potrebné poznamenať, že je tvorená 'nodmi', ktorí sú viacej mediálne známi ako **miners**. Mineri potvrdzujú, takéto transakcie, za čo sú patrične odmeňovaní v digitalnej mene ETH. Takýchto transakcií sa spracuje niekoľko stoviek v krátkom okamžiku a vložia sa do takzvaného bloku. Okrem toho, že sú mineri platení za poskytovanie služieb (napr. svoj počítač) a za validácie transakcií, tak medzi sebou aj súperia o to, kto zvaliduje celý takýto blok, pred tým ako sa prejde na nový. V takomto prípade sa **vyberie jeden náhodný miner** aby potvrdil blok, ktorý dostane ďalšiu patričnú odmenu v mene ETH.

Na sieti Etherea môže ľubovolný používateľ A poslať svoje prostriedky na peňaženku vlastnenú burzou, ktorá potom prekonvertuje digitálnu menu za skutočné fiat a tie pošle užívateľovi na jeho bankový účet.

Najpodstatnejľím rozdielom medzi sieťou Etherea a Bitcoinu sú jednoznačne, **smart kontrakty**. Jednoduchý scénar použitia smart kontraktu by bol ako v predošlom príklade, s tým rozdielom, že užívateľ A by chcel aby jeho prostriedky odišli hneď z jeho peňaženky ale B majú prísť presne o 10 dní. Tým pádom prostriedky najprv putujú na adresu kontraktu a o 10 dní automaticky na adresu užívateľa B. Pod smart kontraktom si môžeme predstaviť akýsi napísaný (open source) kód, ktorý vie vykonávať rôzne funkcionality od prevodov a tradingu, až po obchodovanie s derivátmi medzi rôznymi účastníkmi siete alebo pripísanie vlastenectva rôznych obrázkov (NFT). Ak sa kontrakt týka finančného charakteru, tak dostáva označenie **DeFi**.

Na záver podotknime, že adresa v Ethereum sieti môže byť peňaženkou alebo smart kontraktom. Majoritu však tvoria práve peňaženky (99,9%). V projekte však každú takúto entitu bez znalosti klasifikácie budeme označovať ako adresu.

1.2 Spracovanie dát

Úlohou v tomto projekte bude analýza adries v Ethereum sieti podľa ich atribútov. Keďže získanie nejakých relevantných dát bolo namáhavé, vhodnou cestou sme si ich získali sami. Na kratšie časové obdobie sme sa pomocou vlastného skriptu v jazyku TypeScript napojili na sieť Etherea a získali transakcie medzi rôznymi adresami z celkových 98 blokov, ponechali sme iba tie, v ktorých sa presúvala nejaká čiastka ETH, čo vo výsledku predstavovalo celkovo 10063 transakcií. V každej transakcií sme mali získané údaje ako **kto** (adresa) posielal **komu** (adresa) s **akým obnosom** (ETH). Z transakcií sme si vytvorili novú dátovú sadu unikátnych peňaženiek získaných za toto časové obdobie. Dáta sme uložili, tak že sme ku každej peňaženke vypísali dodatočné číselne atribúty, ktoré boli takisto získane zo samotných transakčných dát. V transakčných dátach sa nachádzala aj informácia, aké adresy zvalidovali rôzne bloky, ak sa takéto adresy nachádzali v našom vytvorenom datasete, tak dané adresy dostali klasifikáciu s označením 'Minning' a typom adresy ako 'Wallet'. Následne sme sa viacerými spôsobmi snažili olabelovať aj iné adresy, napr. pomocou datasetu v [1] alebo pomocou voľného vyhľadávania cez internet. Takto sme niektoré adresy vedeli označiť, (ďalej ako 'olabelovať') či sa jedná o typ peňaženky a či patrí burze (bude niesť ozn. 'Exchange') alebo je anonymná, alebo či sa jedná o typ smart kontraktu a či jeho konkrétny typ je finančného charakteru (bude niesť ozn. 'DeFi'). Treba poznamenať, že pri labelovaní sme úmyselne vyberali adresy, ktoré boli niečim z daných premenných oproti ostatným signifikantné, napr. za sledovaný čas vykonali čo najviac posielacich transakcií. Príklad rozdielu signifikantnosti môžeme hneď vidieť na Obr. 1, kde riadok s označením 'Exchange' ma oproti ostatným 4 riadkom veľký rozdiel v premennej 'totalSent', síce zatiaľ bez poznania významu premennej.

1.3 Dáta

	minBalance	maxBalance	entity	countAsSender	totalSent	totalReceived	coununtAsReceiver	avgSent	avgReceived	currentBlock	id	avgBalance	addressType
1	8.196888e+03	8.336878e+03	Exchange	136	137.2153	0.000000	0	1.0089361	0.0000000	17196119	0x46340b20830761efd32832a74d7169b29feb9758	8.255174e+03	Wallet
2	1.044000e-02	1.044000e-02		0	0.0000	0.005210	1	0.0000000	0.0052100	17196022	0x2e2ccece147daf7f5c29baafa5357944008cd55b	1.044000e-02	
3	2.449601e-01	2.449601e-01		0	0.0000	0.190400	1	0.0000000	0.1904000	17196022	0x22aa53274523610600 4be7cd4afe348d471dfe9	2.449601e-01	
4	3.627691e-02	3.627691e-02		0	0.0000	0.020990	1	0.0000000	0.0209900	17196022	0x3556516c41edd322414d52cb1d29b94746d4956b	3.627691e-02	
5	7.271570e+02	7.294682e+02		21	12.7011	7.086499	64	0.6048145	0.1107266	17196119	0x6dfc34609a05bc22319fa4cce1d1e2929548c0d7	7.282677e+02	

Obr. 1: Vzorka dát, s ktorými sa bude pracovať v projekte . (Kvôli veľkosti šírky sa uprednostnil obrázok.)

Ak by nebolo z názvov stĺpcov v Obr. 1 zrejmé, ako ich interpretovať, tak nižšie popisujeme význam jednotlivých premenných. Pripomíname, že hodnoty premenných sú získané len z času sledovacieho horizontu a nie za celý čas existencie adries.

- minBalance/maxBalance - minimálny/maximálny balance adresy,
- entity - klasifikácia konkrétneho typu adresy, v prípade peňaženky, to môže byť burza, miner alebo unknown a v prípade smart kontraktu sa môže jednať o DeFi alebo unknown.
- countAsSender/countAsReceiver - koľkokrát daná adresa vykonala/prijala transakciu s ETH,
- totalSent/totalReceived - aký celkový objem ETH daná adresa odoslala alebo prijala,
- avgSent/avgReceived - aká bola priemerná výška (v ETH) jednej poslanej/prijatej transakcie
- id - samotná unikátna adresa,
- avgBalance - priemerný balance adresy,

- addressType - typ adresy, tj. či sa jedná o peňaženku alebo smart kontrakt.

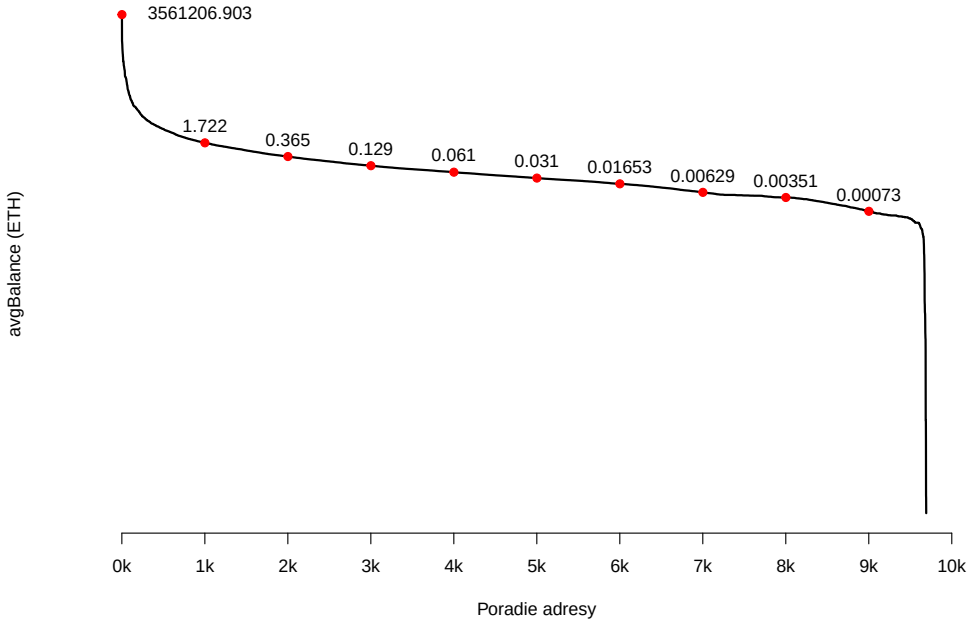
Následne sme pre naše dáta vypočítali nejaké základne štatistiky ako napríklad:

- celkový počet dát: 10 049,
- z toho máme iba 49 adries označených typom peňaženky a 10 ako smart kontrakt,
- ďalej zo 49 peňaženiek 13 patrí minerom a 19 burzám a zo spomínaných 10 smart kontraktov sa v prípade 7 jedná o DeFi. Je vidieť, že počet klasifikovaných je v porovnaní s celkovým počtom adries len maličkým zlomkom.
- Ďalej uvádzame výsledky aritmetických priemerov a mediánov pre naše premenné v Tabuľke 1.

Tabuľka 1: Priemer a medián premenných z našich dát

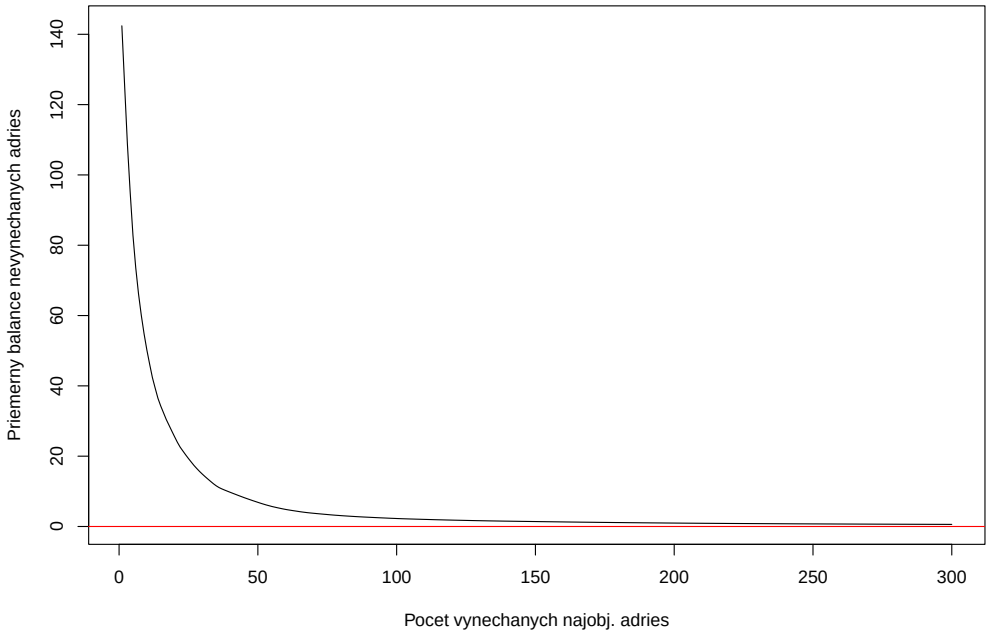
Premenná	Priemer	Medián
minBalance	528.219775	0.05300330
maxBalance	530.784603	0.08171367
countAsSender	1.001393	1.00000000
totalSent	3.919808	0.01780778
totalReceived	2.577443	0.00000000
countAsReceiver	1.001294	0.00000000
avgSent	1.435696	0.00000000
avgReceived	1.379314	0.00000000
avgBalance	496.785438	0.03116136

Z Tabuľka 1 je jasné vidieť rozdiel medzi výsledkami priemeru a mediánu. To môže zapríčiniť presne to, že väčšina adries má malé hodnoty skúmaných premenných ale existuje pár adries, ktoré majú pre tie isté premenné astronomické hodnoty. Ako ukážku si vykreslíme na Obr. 2 zoradené a logaritmicky preškálované priemerné veľkosti všetkých adries (v ETH).



Obr. 2: Distribúcia adries podľa veľkosti ETH.

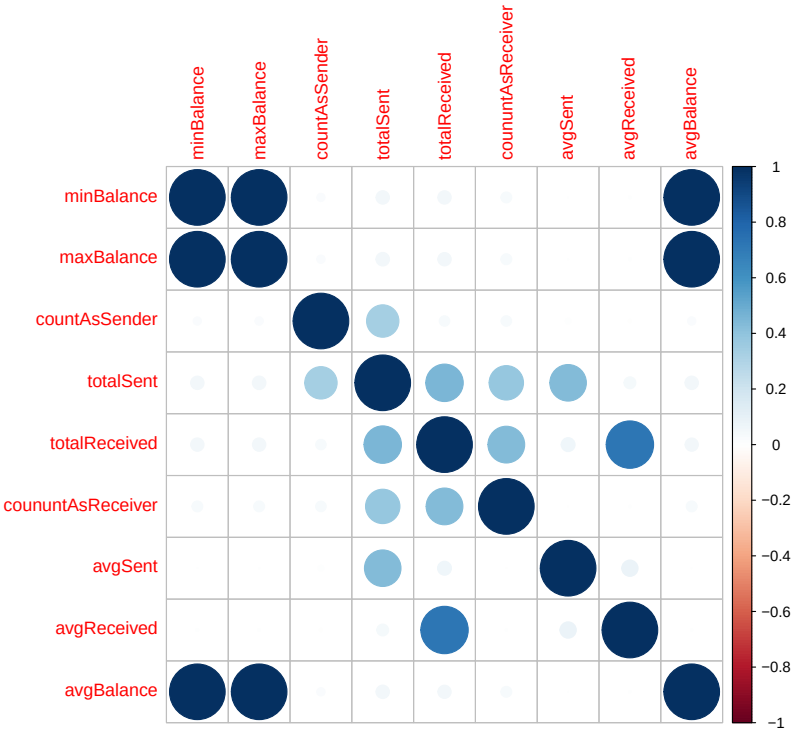
Na Obr. 2, máme logaritmicky preškálovanú krivku distribúcie hodnôt adries avšak s reálnymi vyznačeniami hodnôt premennej 'avgBalance'. Pre nejakú vyznačenú hodnotu (červenou farbou) si vieme na ľavej strane od daného bodu na x-ovej osi pozrieť koľko adries má väčší 'avgBalance' a analogicky pre pravú stranu, koľko adries má menší 'avgBalance' ako je zvolená hodnota 'avgBalance'. Z tohto grafu, môžeme napríklad vidieť, že najväčší obsah ETH ma nejaká peňaženka v celkovej hodnote 3 561 206,903 ETH, zároveň to, že 90% adries za sledované obdobie má menej ako 1,722 ETH alebo, že približne polovica adries vlastní menej ako 0,031 ETH, čo vlastne zodpovedá mediánu v Tabuľka 1. Môžeme vidieť, že naozaj adresy s veľkým počtom ETH, vyhodili $\overline{\text{avgBalance}} \approx 500$ aj keď polovica adries nemá hodnotu ani 0,03 ETH. Vykreslíme si ešte graf na Obr. 3, ktorý bude zobrazovať ako sa bude meniť $\overline{\text{avgBalance}}$ ak vynecháme prvých $n \in \{1, 2, \ldots, 300\}$ adries s najväčšou bilanciou ETH.



Obr. 3: Zmena $\overline{\text{avgBalance}}$ podľa počtu vynechania adries s najväčšou bilanciou.

Vidíme, že už po vynechaní prvej adresy máme z pôvodnej hodnoty 500 hneď nový $\overline{\text{avgBalance}} \approx 140$.

Na záver analýzy dát si ukážeme vzájomnú koreláciu dát na Obr. 4.



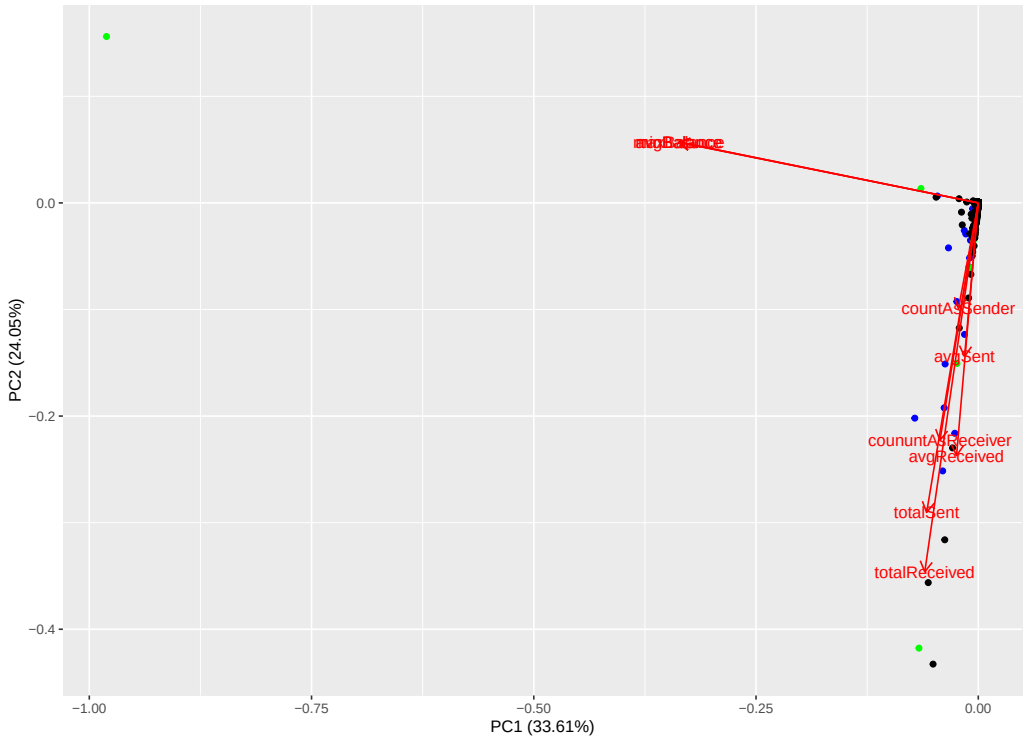
Obr. 4: Odhady korelácií medzi premennými.

Tu môžeme vidieť, že premenné spojené s balancom adresy (minBalance, maxBalance, avgBalance) medzi sebou silno korelujú, čo intuitívne dáva zmysel, pretože ak sledujeme náhodne vybratú adresu s veľkým obsahom ETH nejaké krátke obdobie, tak zrejme aj jej maximálny aj minimálny balance za toto obdobie nebude až tak radikálne odlišné. To znamená, že je menej pravdepodobné, že by sa z chudobnej adresy stala v priebehu momentu bohatá adresa alebo opačne. Takisto je vidieť, že značné korelácie môžu byť aj medzi transakčnými premennými spojené s posielaním a prijímaním prostriedkov.

2 PCA

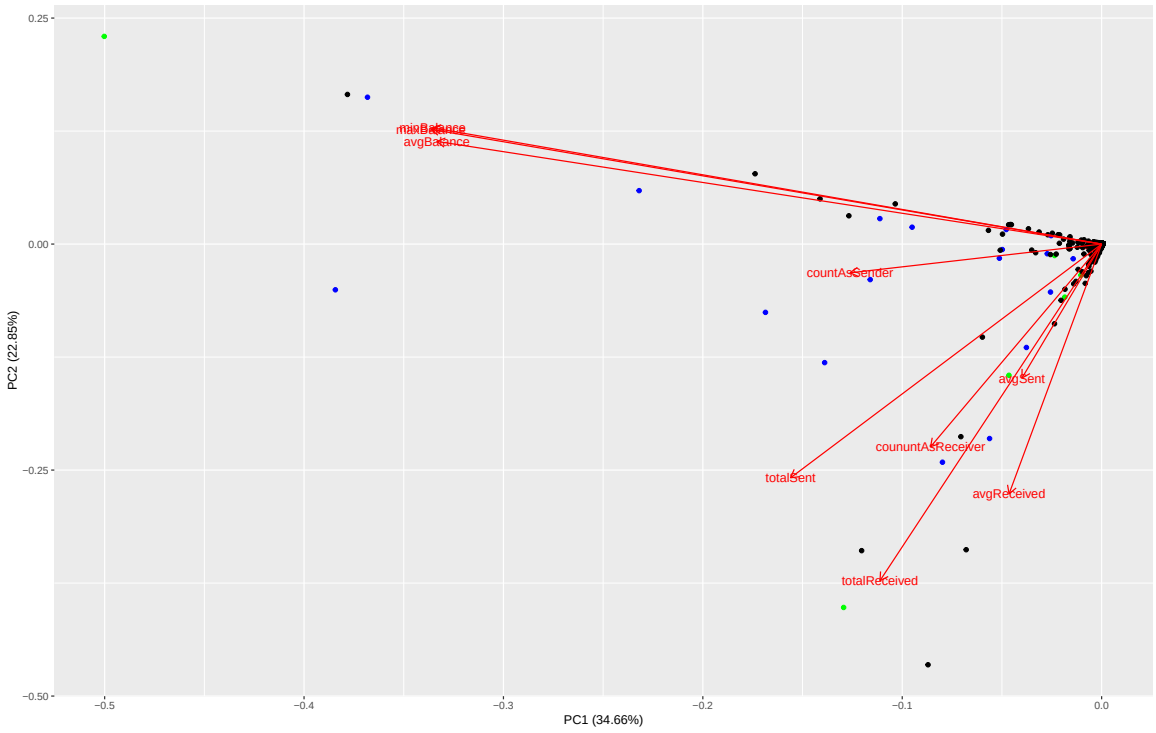
V tejto časti aplikujeme metódu hlavných komponentov na spomínané číselné premenné. Samotné dáta v 2D PCA budú vykresľované 4 rozličnými farbami.

- **zelenou** budú vykreslované dáta smart kontraktov s labelom **DeFi**,
- **modrou** budú vykresľované dáta peňaženiek s labelom **Exchange**,
- **červenou** budú vykreslované dáta peňaženiek s labelom **Mining**,
- **čiernou** budú vykreslované všetky **ostatné adresy**, ktoré sa nám nepodarilo olabelovať aj keď väčšina z nich zrejme budú peňaženky súkromných osôb.



Obr. 5: Výsledok PCA pre všetky dáta

Na Obr. 5 môžeme vidieť výsledok aplikácie metódy hlavných komponentov, avšak vizuálne nedostávame niečo uchopiteľné a to kvôli tomu, že krivka PCA1 zahŕňa v sebe hlavne premenné spojené s balancom a zároveň z Obr. 2 a 3 vieme, že jeden účet ma oproti ostatným astornomickú hodnotu ETH, tak že dokázal posunúť $\overline{\text{avgBalance}}$ celej dátovej sady o niekoľko stoviek ETH vyššie. Z tohto hľadiska budeme dané dáto považovať ako outliera a vyradíme ho z dátovej sady. Pre zaujimavosť môžeme poznamenať, že dané dáto interpretuje smart kontrakt DeFi charakteru. Konkrétne sa jedná o Wrapped Eth kontrakt, ktorého transakcie ako aj stav balancu si vieme pozrieť na `https://etherscan.io/address/0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2`. *Jedná sa o kontrakt, ktorý zabezpečí výmenu povôdnej meny ETH za tzv. ERC-20 token WETH. Niektoré burzy alebo aj DeFi platformy dokážu pracovať iba s tokenovou verziou ETH (tj. ERC-20 tokenom - WETH) a tak ak chce užívateľ využívať takéto služby pre jeho ETH je potrebná najskôr zámena.



Obr. 6: Výsledok PCA pre dáta bez outliera

Na Obr. 6 dostávame jednoznačne lepší vizuálny výsledok. Na osi PCA1 vidíme, že minBalance, maxBalance, avgBalance idú takmer identickým smerom, čo by sedelo podľa Obr. 4 keďže majú vysoké vzájomné korelačné koeficienty. Keď sa pozrieme na stranu PCA2 tak vidíme, že posielacie premenné sú naklonené akýmsi vlastným smerom viacej doľava a prijímacie premenné sú zgrupené viacej doprava, čo nám dáva akési dobré informatívne zobrazenie, či daná adresa viacej (alebo silnejšie) participuje ako odosielateľ alebo prijímateľ ETH. Vo výsledku by sme dáta posunuté na osi PCA1 viacej doľava mohli interpretovať ako adresy, ktoré majú väčší stav účtu a dáta, ktoré sú posunuté na osi PCA2 smerom dole, môžeme interpretovať ako adresy, ktoré sú na blockhaine aktívnejšími. Tiež si môžeme všimnúť, že adresy, ktoré vykonávajú viacej posielacich úkonov majú podľa výsledku PCA väčší objem ETH ako tí, ktorí viacej prijímajú, čo je jednoznačne zaujímavý ukazovateľ. Ukazovateľ si môžeme aj jednoducho prešetriť, tak že si dané adresy budeme nejakým spôsobom porovnávať, napr. necháme si balance adries, ktoré odoslali aspoň $i \in \{0, 1, \ldots, 33\}$ transakcií a súbežne s nimi si necháme adresy, ktoré prijali ten istý počet i transakcií. Potom pre oba prefiltrované datasety spočítame $\overline{\text{avgBalance}}_{\text{send,receive}}$ a vyhodnotíme, ktorý z nich je väčší. To isté spravíme párovo pre ďalšie premenné odosielacieho a prijímacieho charakteru - (avgSent, avgReceived) a (totalSent, totalReceived). V párových prípadoch spočítame koľkokrát bol $\overline{\text{avgBalance}}_{\text{send}} > \overline{\text{avgBalance}}_{\text{receive}}$ a predelíme celkovým počtom porovnávaní. Dostali sme, že v 70% prípadoch porovnávania došlo k javu: $\overline{\text{avgBalance}}_{\text{send}} > \overline{\text{avgBalance}}_{\text{receive}}$. *O správnosti postupu porovnávania by sa dalo diskutovať, jednalo sa však o prvotný nápad pre overenie pozorovania.

Na záver 2D PCA výsledku si môžeme všimnúť, že síce sme mali zlomok zaklasifikovaných adries, aj tak je jasne vidieť, že Exchange a DeFi, tvoria najaktívnejšiu časť adries, ktoré najčastejšie posielajú/prijímajú transakcie spolu s najväčšími hodnotami ETH a zároveň predstavujú aj signifikantnú časť účtov, ktoré majú najväčší balance. Mineri zrejme podľa výsledkov PCA nie sú až tak aktívni a nemajú ani veľký balance a sú v nejakom spoločnom 'zhluku' s ostatnými nezaklasifikovanými adresami.

Keďže malo 2D pca podľa ľaktového diagramu cca. 57,51 % výpovednu hodnotu, tak vykreslíme ešte 3D PCA, ktorého výpovedná hodnota bude 72,12 %. Nižšie prikladáme link na

interaktívny grafický výstup PCA s využitím prvých troch hlavných komponentov: `https://6458e56a0554c82be443ed6d--shimmering-chaja-d1d40b.netlify.app`. V 3D zobrazení môžeme vidieť podobné správanie ako v zobrazení 2D a však s tým rozdielom, že posielajúce premenné a prijímacie premenné su viacej prehľadne od seba oddelené a je jasnejšie vidieť, že posielajucé premenné v sebe zahŕňajú o niečo väčší balance adresy ako prijímacie premenné.

3 Vyberanie vhodného predikčného modelu s využitím najoptimálnejšieho LASSA

V tejto časti by sme chceli nájsť vhodný model na predikciu premennej 'entity' pomocou ostatných číselných premenných. Síce premenná entity nadobúda celkovo 4 rôzne hodnoty a mohlo by sa zdať praktické použiť multinomickú logisticku regresiu, je potrebné dodať že množstvo v jednotlivých kategóriach je veľmi malé a tak premennú 'entity' skôr budeme modelovať binárnym spôsobom, tj. ak entity je 'Exchange' alebo 'Miner' alebo 'DeFi, tak prislúchajúcu hodnotu budeme označovať 1 a zároveň takúto adresu budeme nazývať zaujímavou, v opačnom prípade označujeme hodnotou 0 a adresu nazývame nezaujímavou adresou. Dáta rozdelíme v pomere 70% na trénovacie účely a 30% na testovacie účely (v oboch setoch zabezpečíme dostatočnú početnosť olablovaných dát, teda neprázdnosť 'entity') a následne na trénovacich dátach vytvoríme logistický model, ktorý používa všetky ostatné číselné premenné a rovno v Tabuľke 2 zobrazíme jeho úspešnosť v klasifikácií na testovacích dátach.

Tabuľka 2: Kontingenčná tabuľka, zobrazujúca úspechy a neuspéchy klasifikácie modelu

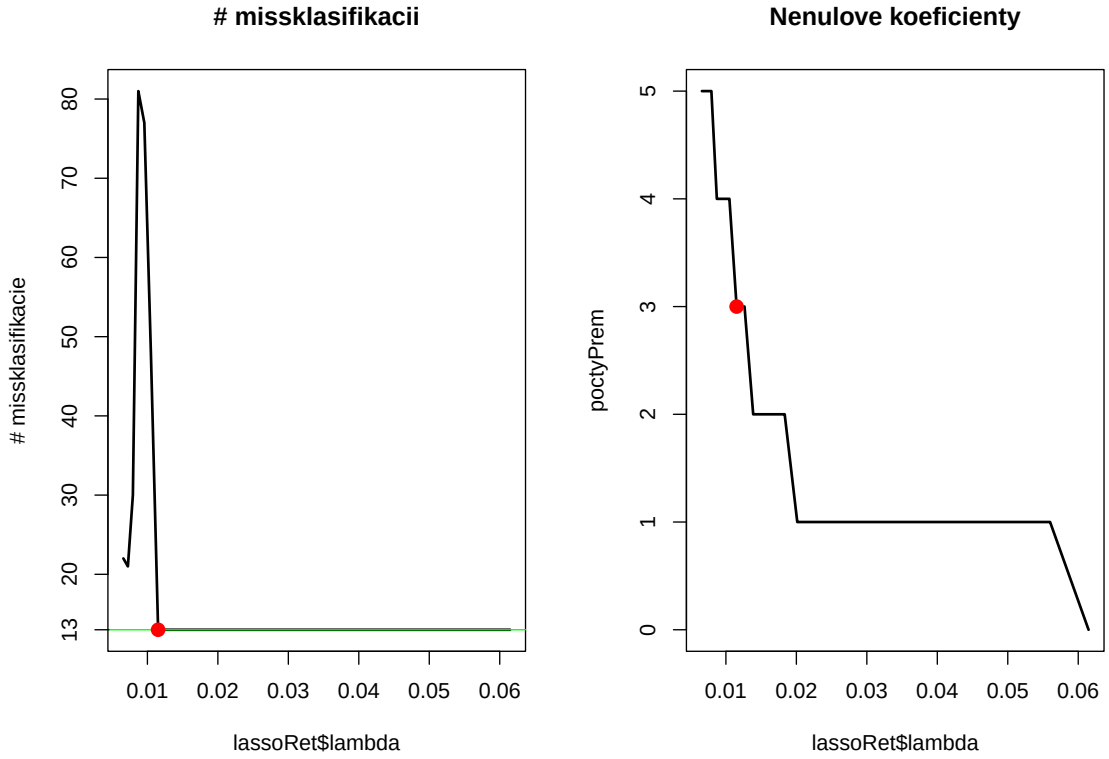
Aktuálne / Predikované	0	1
0	3002	0
1	9	4

Celkovo model nesprávne zaklasifkoval 9 zaujímavých adries a priradil ich ku kategórií nezaujímavých, čo nie je najhorší výsledok, keďže sa mu stále podarilo odhaliť 4 zaujímavé adresy. Čo teraz vyskúšame je selekcia iba niektorých premenných s použitím metódy LASSO. Dôvod prečo chceme urobiť takýto krok, je tušenie, že by LASSO mohlo niektoré premenné označiť ako zbytočné, keďže medzi sebou niektoré silno korelujú.

Postup je následovný:

- Dáta predelíme po novom na 2 kategórie, trénovacie (70%) a validačné (30%) a kópia validačných s označením testovacie (vysvetlené nižšie)
- Vytvoríme viacero LASSO modelov, pre rôzne λ_i. Každý z nich bude získaný z trénovacích dát, z 1. kroku.
- Na validačnej vzorke pomocou LASSO modelov budeme predikovať premennú 'entity' a následne zvolíme LASSO model s najmenším počtom nesprávnej klasifikácie.
- Pozrieme sa na výsledné optimálne LASSO a na premenné, ktoré ponechal nenulové. Z takýchto nenulových premenných vytvoríme opäť logistický model na tých istých trénovacich dátach.
- Vyhodnotíme jeho úspešnosť na testovacích dátach z prvého kroku (validačných dátach pre LASSO), presne tak ako bol vyhodnocovaný prvý logistický model so všetkými premennými a vzájomne ich porovnáme.

Priebeh všetkých vytvorených LASSO modelov v závislosti od zvolenej lambdy a výsledného počtu nesprávnej klasifikácie pre validačnú sadu (ľavá časť obrázku) ako aj závislosť zvolenej lambdy a výsledného počtu nenulových premenných (pravá časť obrázku) si vieme, pozrieť na Obr. 3 .



Najoptimálnejšie LASSO sme zobrali prvé v poradí s najnižším počtom nesprávnej klasifikácie, ktoré má celkovo 3 nenulové premenné. Prekvapivo nenulovými premennými sú 'countAsSender', 'avgSent' a 'avgReceived' a teda daný LASSO model označil všetky premenné úzko súvisiace s balancom ako nepodstatné. Ako možný dôvod môže byť nahliadnutie na Obr. 6, kde môžeme vidieť, že smer 'countAsSender' nesie v sebe informáciu vyššieho balancu ako aj informáciu spojenú s odosielaním a prijímaním. Takže premennú 'countAsSender' možno vyhodnotil ako nejaký zlatý stred, ktorý vie popísať aj informácie ostatných dát.

Teraz si vytvoríme podľa výsledkov LASSA logistický model len s použitím daných 3 premenných a zhodnotíme spomínané výsledky na testovacej vzorke. Už iba z výsledkov PCA na 6 by niekto mohol očakávať, že vynechanie premenných spojených s balancom dokážu radikálne zhoršiť model ak ich premenná 'countAsSender' nedokáže dostatočne nahradiť, avšak prekvapivo, keď si pozrieme výsledky klasifikácie v Tabuľke 3 na testovacích dátach, tak môžeme vidieť, že celkové vynechanie premenných spôsobilo nárast v nesprávnej klasifikácií len o 2 jednotky.

Tabuľka 3: Kontingenčná tabuľka, zobrazujúca úspechy a neuspéchy klasifikácie modelu

Aktuálne / Predikované	0	1
0	3002	0
1	11	2

4 Zhodnotenie výsledkov

V prvej časti projektu sme si vysvetlili s akými dátami pracujeme, čo znamenajú a ako sme ich získali. Následne sme pre ne vykreslili nejaké prvotné pozorovania ako vzájomne korelačne koeficienty alebo priemery a mediány a ukázali akú signifikantnosť zohráva najsilnejší outlier v určovaní premennej avgBalance.

V druhej časti sme zobrali naše viacdimenzionálne dáta skúsili sme ich s použitím metódy PCA vykresliť v menej rozmere - v 2D a 3D. V oboch vykresleniach sme podľa nás dostali veľmi rozumné výsledky, z ktorých sa dá veľa vyčítať. Jednotlivé PCA môžu mať aj praktické využitie pre vstup do rôznych zhlukovacích metód.

Keďže sme si dali menšiu námahu a pár desiatok adries sme olabelovali (premenná 'entity'), tak sme si vytvorili dva logistické modeli na predpovedanie toho či daná adresa je zaujímavá alebo nie. Jeden model bral na vstupe všetky číselné premenné a druhý model bol akýmsi výsledkom nášho vlastného postupu s využitím metódy LASSA a validačnej vzorky dát. Prvotným dôvodom prečo sme chceli takúto alternatívu modelu použiť bola možnosť zanedbania nepotrebných premenných. Očakávali sme, že vyhodí niektoré z premenných, ktoré medzi sebou silno korelovali, avšak vo výsledku boli zanedbané pre nás intuitívne dôležité premenné. Vo výsledku je diskutabilné či je model s využitím LASSA dobrá voľba. Síce bol horši v klasifikácií iba o 2 nesprávne zaradenia v porovnaní s prvým modelom ale koniec koncom nesprávne klasifikoval 2 zaujímavé adresy a označil ich za nezaujímavé aj keď naším hlavným cieľom je odhaľovanie presne takých adries. Čiže vo výsledku by sme zrejme siahli po plnom modeli, ktorému sa lepšie darilo na testovacej sade dát.

5 Referencie

[1] Použité dáta k čiastočnému labelovaniu `https://www.kaggle.com/datasets/hamishhall/labelled-ethereum-addresses`

[2] Všetky vytvorené dáta, skript na parsing blochainových dát, Rkovský skript ako aj obrázkové výstupy použité v prezentácií a samotné pdf `https://github.com/devAdam117/Data-dimension-reduction`