

# Ako algoritmy sociálnych médií získavajú informácie z používateľov a následne ich využívajú vo svoj prospech

David Vach

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií

xvachd@stuba.sk

Vedenie: Mirwais Ahmadzai

16. december 2023

## Abstrakt

Tento článok sa bude venovať podrobnej analýze získavaniu informácii používateľov sociálnych médií pomocou algoritmov. Práca bude pozostávať z toho ako sa tieto algoritmy snažia získať čo najviac informácií z používateľov, ktoré následne môžu využiť na rôzne marketingové ciele. Jadro tohto článku sa sústreďuje na algoritmy a ich efektivitu v udržiavaní pozornosti používateľov na základe získaných dát a následné zobrazovanie špecifického obsahu za účelom získať čo najviac presnejších dát a následné vymazávanie nepotrebných dát. Ďalšia časť tejto práce ukazuje rôzne spôsoby zneužívania týchto dát vo svoj prospech z dlhodobého hľadiska a ako tieto algoritmy vedia pomocou získaných dát určiť s veľkou pravdepodobnosťou osobnosť používateľa a iné citlivé informácie. Záver práce tvoria dôvody prečo by sme si mali tieto súkromné dáta chrániť a aké spôsoby sa na to využívajú.

**Kľúčové slová**— sociálne médiá, zber dát, ochrana súkromia, algoritmy, marketing

# 1 Úvod

V úvode dnešnej digitálnej éry predstavujú sociálne médiá viac než len platformy na zdieľanie osobných príbehov a fotografií. Sú to sofistikované ekosystémy využívajúce algoritmy, ktoré transformujú spôsob, akým komunikujeme, angažujeme sa a dokonca ako sa rozhodujeme. Algoritmy sociálnych médií sú navrhnuté tak, aby boli nenápadnými, no mocnými kustódmi našej online pozornosti.

Tieto algoritmy pracujú na základe predpokladu, že každá interakcia na platforme - lajk, zdieľanie, komentár, dokonca aj dĺžka zdržania sa pri príspevku - je cenným údajom, ktorý odhaľuje náš vkus, preferencie a správanie. Tieto údaje sú potom analyzované a transformované na mieru, ktorá nás drží zapojených - a často prilepených - na naše obrazovky dlhšie než sme plánovali.

No táto personalizácia má aj svoju cenu. Algoritmy sociálnych médií sú takisto navrhnuté na podporu obchodných modelov ich prevádzkovateľov, čo znamená, že informácie získané z používateľských interakcií sú často využívané na cieľovú reklamu a obsahové odporúčania, ktoré podporujú angažovanosť a zisky. Tento proces môže používateľov nevedomky manipulovať k tomu, aby konzumovali obsah a produkty, ktoré možno nepotrebujú alebo dokonca nechcú.

Jedným z hlavných problémov týchto algoritmov je ich neschopnosť porozumieť ironii a sarkazmu. To vedie k tomu, že dáta, ktoré tieto algoritmy vyprodukuje, môžu obsahovať nepresné informácie a interpretácie. Používatelia často využívajú ironiu a sarkazmus vo svojich príspevkoch na sociálnych médiách, čo môže algoritmy zmiatať a viesť k nesprávnym záverom.

Tento úvod bude skúmať, ako algoritmy sociálnych médií zbierajú a využívajú používateľské informácie a aké sú implikácie tohto procesu na súkromie, spoločenské správanie a dokonca na demokraciu. Dôležitou témou bude aj vplyv týchto algoritmov na manipuláciu s používateľmi a formovanie ich názorov prostredníctvom personalizovaného obsahu a reklamy.

Na záver, bude sa táto práca venovať aj možnostiam riešenia tohto problému a zlepšeniu efektívnosti algoritmov sociálnych médií v interpretácii komplexných používateľských prejavov vrátane ironie a sarkazmu. [2] [7]

## 2 Súvisiaca práca

V tejto sekcií sa venujeme prehľadu literatúry, ktorá sa zaoberá rôznymi aspektmi analýzy a klasifikácie dát sociálnych médií, ako aj predpovedaním a vizualizáciou spotrebiteľských nálad. Táto analýza literatúry nám pomôže lepšie pochopiť komplexnosť a rôznorodosť prístupov k riešeniu problémov spojených so sociálnymi médiami.

Zhang et al. [11] sa zameriavajú na predpovedanie a vizualizáciu spotrebiteľských sentimentov v online sociálnych médiách. Tento prístup je dôležitý pre pochopenie správania sa spotrebiteľov a ich postojov voči rôznym produktom a službám.

Desai a Patil [2] skúmajú efektívne regresné algoritmy pre klasifikáciu dát sociálnych médií. Ich práca poskytuje užitočné informácie o tom, ako efektívne spracovať a analyzovať veľké množstvá dát.

Reddy a Parvathy [4], ako aj Reddy a Kumar [5], poskytujú inovatívne metódy na predpovedanie úrovne znečistenia ovzdušia, resp. na určovanie presnosti cien akcií s využitím algoritmov náhodných lesov a gradient boosting.

Rizky Pribadi et al. [6] sa zameriavajú na analýzu angažovanosti používateľov sociálnych médií na trhoviskách podľa tém, čo je kľúčové pre marketing a reklamu.

Shahare [8] sa venuje analýze sentimentu pre dáta z novín na základe sociálnych médií, čo je dôležité pre pochopenie verejnej mienky a trendov.

Zhang et al. [10] a Zhang [12] skúmajú výber funkcií a algoritmy učenia založené na hlbokých neurónových sieťach pre dáta zo sociálnych médií, čo je dôležité pre efektívne spracovanie a analýzu týchto dát.

Arambepola a Munasinghe [1] sa zameriavajú na empirickú analýzu faktorov používateľov ovplyvňujúcich dizajn používateľského rozhrania v mobilných aplikáciách, čo je dôležité pre optimalizáciu používateľského zážitku.

Rohani et al. [7] poskytujú praktický prístup k modelovaniu tém pre obsah sociálnych médií, čo je kľúčové pre porozumenie a analyzovanie diskusií a trendov na sociálnych sieťach.

Napokon, Fan et al. [3] sa zaoberajú otázkou, či je náhodný model lepší z hľadiska presnosti a efektívnosti, čo je zásadný príspevok k diskusií o optimálnych modeloch pre dáta sociálnych médií. Táto sekcia nám ukazuje rozmanitosť výskumu v oblasti analýzy sociálnych médií a jeho dôležitosť pre rôzne odvetvia vrátane marketingu, financií a verejnej mienky.

### 3 Metodológia

V rámci tejto štúdie sme sa zameriavali na analýzu algoritmov sociálnych médií a ich vplyvu na správanie používateľov. Naša metodológia zahŕňala niekoľko dôležitých krokov.

Prvým krokom bolo výber relevantných dátových sád zo sociálnych médií, pričom sme sa špecifickejšie zameriavali na obdobie posledných dvoch rokov. Tieto dáta zahŕňali rôzne používateľské interakcie, ako sú lajky, komentáre a zdieľanie obsahu.

Následne sme prešli k analýze konkrétnych algoritmov používaných na týchto platformách, vrátane algoritmov ako NaiveBayes a Adaboost. Naším cieľom bolo posúdiť ich efektívnosť v personalizácii obsahu a skúmať, aký vplyv majú na používateľské správanie.

Pri analýze dát sme využívali techniky spracovania prirodzeného jazyka a analýzy sentimentu, aby sme získali hlbší pohľad na povahu týchto dát. Počas celého procesu sme venovali osobitnú pozornosť etickým aspektom spojeným so zberom a analýzou používateľských dát.

Na vyhodnotenie účinnosti rôznych algoritmov sme použili špecifické metriky, ako sú miera zapojenia používateľov a čas strávený na platforme. Táto metodológia nám umožnila poskytnúť podrobnejšie pochopenie toho, ako algoritmy sociálnych médií ovplyvňujú používateľské správanie a súkromie.

### 4 Ako tieto algoritmy fungujú

Algoritmy sociálnych médií sú navrhnuté tak, aby efektívne zbierali a analyzovali údaje generované užívateľmi pri ich interakciách online. Ich schopnosť poskytovať personalizovaný obsah je založená na dômyselnom sledovaní a vyhodnocovaní činností používateľov. Keď používateľ lajkuje príspevok, komentuje video alebo strávi určitý čas čítaním článku, tieto akcie sú zaznamenané a stávajú sa časťou veľkého údajového modelu, ktorý tento algoritmus používa na pochopenie a predvídanie preferencií používateľa.[6]

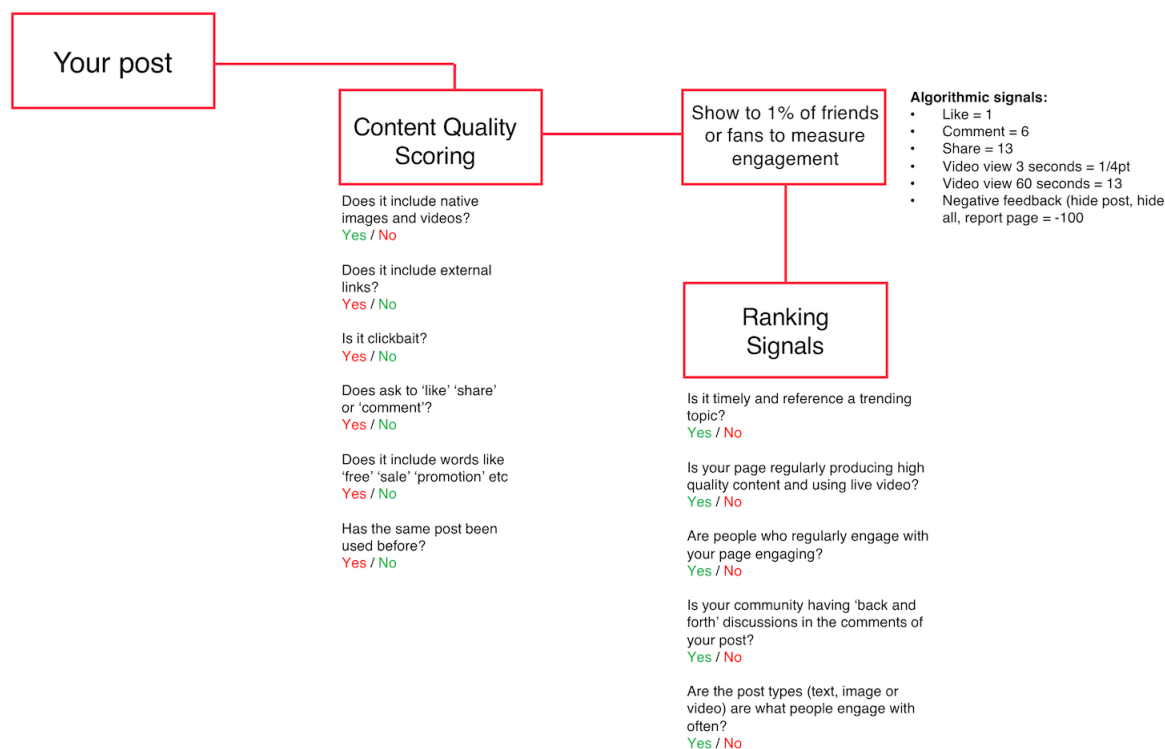
Na základe textu, obrázkov a videí, ktoré používatelia zdieľajú alebo komentujú, algoritmy využívajú pokročilé metódy ako spracovanie prirodzeného jazyka a rozpoznávanie obrazu na to, aby lepšie porozumeli obsahu, ktorý používatelia preferujú. Algoritmy zároveň skúmajú vzťahy a interakcie medzi používateľmi, aby zistili, aké sociálne siete sú pre nich najdôležitejšie a ako sa tie môžu prepojiť s ich obsahovými preferenciami.

Na základe získaných dát tieto algoritmy vytvoria obsah, ktorý sa používateľom zobrazuje vo feedoch, aby bol čo najrelevantnejší a najzaujímavejší. Tento personalizovaný prístup sa nezameriava len na zvýšenie užívateľskej spokojnosti a zapojenia, ale je tiež základom pre efektívnu cieľovú reklamu, kde sú reklamy prispôbené konkrétnym záujmom a správaniu používateľov.

Súčastou procesu je aj predpovedanie správania. Algoritmy sociálnych médií sa pokúšajú predvídať, čo by mohlo používateľov zaujímať v budúcnosti, a prispôbujú svoje stratégie, aby bol obsah, ktorý zobrazujú, vždy krok vpred. Vďaka neustálej spätnej väzbe a iterácii sa algoritmy neustále učia a zdokonaľujú, aby boli ešte lepšie v predvídaní a napĺňaní očakávaní užívateľov.

Názorný príklad pre Facebook:

# The Facebook algorithm



Obr. 1: Truncation of SVD for LSI. [2]

Výsledkom je, že algoritmy sociálnych médií tvoria dynamický a neustále sa vyvíjajúci systém, ktorý je mimoriadne citlivý na potreby a chovanie používateľov, a to všetko v úsilí udržať ich zapojenie a poskytnúť im maximálne prispôbený zážitok. Ekonomický model sociálnych sietí, založený predovšetkým na reklamných príjmoch, je priamo ovplyvnený úspešnosťou týchto algoritmov v predpovedaní a ovplyvňovaní používateľského správania.

## 5 Analýza

Analýza algoritmov strojového učenia v kontexte sociálnych médií je kľúčovou časťou našej štúdie. Tieto algoritmy majú zásadný vplyv na to, ako sú prezentované informácie používateľom a ako sú tieto informácie personalizované. Na túto analýzu sme vybrali viacero algoritmov strojového učenia, ktoré budeme dôkladne analyzovať z hľadiska ich výhod, nevýhod a najlepších použití.

### 5.1 NaiveBayes[11]

### 5.2 Adaboost[2]

### 5.3 Random Forest[4]

### 5.4 Gradient Boosting Machines (GBM)[5]

Algoritmus	Výhody	Nevýhody	Najlepšie použitie
Adaboost	Zlepšuje slabé modely, prispôsobivý	Citlivý na šum, náchylný na preučenie	Klasifikácia 'náročných' prípadov
NaiveBayes	Jednoduchý, rýchly, efektívny pre textové dáta	Predpokladá nezávislosť atribútov, môže byť menej presný	Klasifikácia textu, filtrovanie spamu
GBM	Vysoká presnosť, efektívny pre komplexné vzory	Náchylný na preučenie, výpočtovo náročný	Klasifikácia a regresia s komplexnými vzormi
Random Forest	Odolný voči preučeniu, efektívny pre veľké dáta	Menej interpretabilný, výpočtovo náročný	Klasifikácia a regresia, veľké dáta

Tabuľka 1: Porovnanie algoritmov strojového učenia podľa článku Desai a Patil [2] a ďalších prác [11, 4, 5]

## 6 Výsledok

## 7 Záver

V závere našej štúdie sme dospeli k dôležitým zisteniam týkajúcim sa algoritmov sociálnych médií a ich vplyvu na používateľov. Naše analýzy ukázali, že tieto algoritmy hrajú kľúčovú úlohu v personalizácii obsahu a vytváraní zapojenia používateľov. Sú schopné efektívne zbierať, analyzovať a predvídať používateľské preferencie na základe ich interakcií na platformách sociálnych médií.

S týmto prispôbovaním však prichádza aj riziko zneužitia a narušenia súkromia používateľov. Algoritmy môžu byť zneužitie na manipuláciu s používateľskými správaniami a na cieľovú reklamu, ktorá niekedy prekračuje hranice etiky. Je preto nevyhnutné, aby sa

vývoj a implementácia týchto algoritmov riadila prísnyimi etickými normami a normami týkajúcimi sa ochrany súkromia.

Transparentnosť a regulácia v oblasti algoritmov sociálnych médií sú kľúčovými faktormi na zaistenie bezpečnosti a súkromia používateľov. Je dôležité, aby platformy sociálnych médií aktívne pracovali na zlepšovaní svojich algoritmov a na ochrane súkromia svojich používateľov.

Celkovo nás naša štúdia núti zamyslieť sa nad tým, ako sú naše údaje spracovávané a využívané na sociálnych médiách, a zdôrazňuje potrebu ochrany súkromia a etického využívania algoritmov. Len tak môžeme zabezpečiť, že budúcnosť sociálnych médií bude prospešná a bezpečná pre všetkých používateľov.

## 8 Reakcia na prednášky

### 8.1 Kreatívne písanie (Prednáška 8)

Kreatívne písanie je umenie, ktoré umožňuje autorom vyjadriť svoje myšlienky, emócie a predstavy prostredníctvom slov. Je to proces, ktorý spočíva v tvorbe originálnych textov, či už ide o poéziu, prózu, drámu alebo iné literárne formy. Pri kreatívnom písaní je dôležitá nielen schopnosť gramaticky správneho a zrozumiteľného vyjadrovania, ale predovšetkým originalita a kreativita. Autor by mal byť schopný preniesť čitateľa do svojho sveta, ukázať mu nové perspektívy a vyvolať v ňom emócie.

Kľúčom k úspechu v kreatívnom písaní je pravidelná prax a ochota experimentovať so slovami a štýlom. Dobrý autor by mal byť otvorený novým nápadom a neprestajne rozvíjať svoje písacie zručnosti. Kreatívne písanie nie je len o výroku umeleckých diel, ale aj o sebaujadrení a osobnom rozvoji. Táto forma písania môže byť terapeutická a umožňuje autorom vyrovnávať sa s osobnými záležitosťami alebo spracovať svoje myšlienky a pocity. Preto je dôležité vytvárať prostredie, kde sa autori cítia inšpirovaní a majú priestor na kreativitu.

### 8.2 Plagiátorstvo (Prednáška 9)

Plagiátorstvo je téma, o ktorej by mal byť každý tvorca článku, eseje, prezentácie alebo autor akéhokoľvek verejného prejavu oboznámený. Dá sa to nazvať krádežou, ale niekedy sa môže človek dopustiť neúmyselnému plagiátorstvu. Spôsob, akým sa mu vyhnúť je jednoduchý, a to odkazovať sa na použité zdroje.

### 8.3 Technológia a ľudia: Scrum. doc. Valentino Vranič (Prednáška 11)

Scrum predstavuje cenný nástroj aj v akademickom prostredí. Pre mňa osobne znamená efektívny spôsob organizácie a riadenia mojich študijných projektov a úloh. Princípy Scrumu, ako je rozdelenie práce do krátkych iterácií a pravidelné kontroly, mi pomáhajú udržať si prehľad nad rozsiahlymi úlohami. Zahŕnutie flexibility na reagovanie na nečakané zmeny v pláne mi umožňuje byť pripravený na všetko, čo sa počas semestra môže stať.

Okrem toho, Scrum podporuje silnú spoluprácu tímu, čo je dôležité v rámci skupinových projektov a prezentácií. Dokážeme lepšie komunikovať, sledovať postup jednotlivých členov tímu a včas identifikovať prípadné problémy. Týmto spôsobom sa stávame efektívnejšími a dosahujeme lepšie výsledky.

# Literatúra

- [1] N. Arambepola and L. Munasinghe. Empirical analysis of user factors that affect the user interface design in mobile applications. In *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 166–171, 2020.
- [2] S. Desai and S. Patil. Efficient regression algorithms for classification of social media data. In *2015 International Conference on Pervasive Computing (ICPC)*, pages 1–5, 2015.
- [3] W. Fan, H. Wang, P. Yu, and S. Ma. Is random model better? on its accuracy and efficiency. In *Third IEEE International Conference on Data Mining*, pages 51–58, 2003.
- [4] P. D. Reddy and L. R. Parvathy. Prediction analysis using random forest algorithms to forecast the air pollution level in a particular location. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1585–1589, 2022.
- [5] P. V. Reddy and S. M. Kumar. A method for determining the accuracy of stock prices using gradient boosting and the support vector machines algorithm. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1596–1599, 2022.
- [6] M. Rizky Pribadi, E. Puji Widiyanto, D. Hermanto, D. Iba Ricoida, D. Pibriana, and Rusbandi. Analysis of marketplace social media user engagement by topic. In *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pages 35–39, 2022.
- [7] V. A. Rohani, S. Shayaa, and G. Babanejaddehaki. Topic modeling for social media content: A practical approach. In *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, pages 397–402, 2016.
- [8] F. F. Shahare. Sentiment analysis for the news data based on the social media. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1365–1370, 2017.
- [9] J. Vaidya, B. Shafiq, W. Fan, D. Mehmood, and D. Lorenzi. A random decision tree framework for privacy-preserving data mining. *IEEE Transactions on Dependable and Secure Computing*, 11(5):399–411, 2014.
- [10] H. Zhang, G. Hu, Z. Pan, and H. Yang. A feature selection algorithm based on pairwise constraints for linked social media data. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 548–553, 2017.
- [11] L. Zhang, H. Yuan, and R. Y. K. Lau. Predicting and visualizing consumer sentiments in online social media. In *2016 IEEE 13th International Conference on e-Business Engineering (ICEBE)*, pages 92–99, 2016.
- [12] R. Zhang. Research on social media feature learning algorithm based on deep neural network. In *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*, pages 590–594, 2022.