

Algoritmi strojnog učenja u poslovnoj inteligenciji: Kako strojno učenje revolucionira rješenja za poslovnu inteligenciju?

Blažević, Mihael

Undergraduate thesis / Završni rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:137:368475>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-27**



Repository / Repozitorij:

[Digital Repository Juraj Dobrila University of Pula](#)



Sveučilište Jurja Dobrile u Puli

Fakultet Informatike u Puli

Mihael Blažević

**Algoritmi strojnog učenja u poslovnoj inteligenciji: Kako strojno učenje
revolucionira rješenja za poslovnu inteligenciju?**

Završni rad

Pula, Rujan 2023. godine

Sveučilište Jurja Dobrile u Puli

Fakultet informatike u Puli

Mihael Blažević

**Algoritmi strojnog učenja u poslovnoj inteligenciji: Kako strojno učenje
revolucionira rješenja za poslovnu inteligenciju?**

Završni rad

JMBAG: 0303094587, Mihael Blažević /redovni student

Studijski smjer: Informatika

Kolegij: Sustavi poslovne inteligencije

Mentor: doc.dr.sc. Goran Oreški

Pula, Rujan 2023. godine



IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Mihael Blažević, kandidat za prvostupnika Informatike ovime izjavljujem da je ovaj Završni rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljeni način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student/ica
Mihael Blažević

U Puli, Rujan 2023. Godine



IZJAVA O KORIŠTENJU AUTORSKOGA DJELA

Ja, Mihael Blažević, dajem odobrenje Sveučilištu Jurja Dobrile u Puli, nositelju prava korištenja, da moj završni rad pod nazivom „ Algoritmi strojnog učenja u poslovnoj inteligenciji: Kako strojno učenje revolucionira rješenja za poslovnu inteligenciju“ upotrijebi da tako navedeno autorsko djelo objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te preslika u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu sa Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama. Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

Student/ica

Mihael Blažević

U Puli, Rujan 2023. Godine

Sažetak:

Rad "Algoritmi strojnog učenja u poslovnoj inteligenciji: Kako strojno učenje revolucionira rješenja za poslovnu inteligenciju" istražuje integraciju strojnog učenja u područje poslovne inteligencije (BI) i kako ova kombinacija mijenja način analize i donošenja poslovnih odluka. U prvom dijelu rada analizira se osnovni koncept poslovne inteligencije, njezini procesi, arhitektura i ciljevi, kako bi se postavio temelj za razumijevanje uloge strojnog učenja u ovom kontekstu. Nakon toga, rad istražuje povijest i ključne koncepte strojnog učenja, uključujući duboko učenje (Deep learning), te najpopularnije algoritme strojnog učenja i njihovu primjenu u analizi podataka. Kroz primjere iz stvarnog svijeta, rad ilustrira kako se algoritmi strojnog učenja koriste u poslovnoj inteligenciji za analizu korisničkog odnosa, upravljanje zalihama, personalizirani marketing, analizu sentimenta na društvenim medijima i otkrivanje prijevara. Nadalje, rad istražuje integraciju strojnog učenja u postojeće BI alate, prednosti i nedostatke te izazove koji se javljaju u ovom procesu. Posebna pažnja posvećena je etičkim i sigurnosnim aspektima korištenja strojnog učenja u BI-u, uključujući razmatranje etičkih okvira i sigurnosnih strategija. Na kraju, rad raspravlja o budućnosti poslovne inteligencije uz strojno učenje i zaključuje s naglaskom na važnosti ove tehnološke revolucije za donošenje informiranih poslovnih odluka.

Ključne riječi: baze podataka, poslovna inteligencija, strojno učenje, duboko učenje, algoritmi, analiza podataka, integracija, etika, sigurnost, održivost.

Abstract:

The paper "Machine Learning Algorithms in Business Intelligence: How Machine Learning Revolutionized Business Intelligence Solutions" explores the integration of machine learning into the field of business intelligence (BI) and how this combination is reshaping the way data analysis and business decision-making are conducted. The first part of the paper examines the fundamental concept of business intelligence, its processes, architecture, and objectives, laying the groundwork for understanding the role of machine learning in this context. Subsequently, the paper delves into the history and key concepts of machine learning, including deep learning, as well as the most popular machine learning algorithms and their application in data analysis. Through real-world examples, the paper illustrates how machine learning algorithms are used in business intelligence for customer churn analysis, inventory management, personalized marketing, sentiment analysis on social media, and fraud detection. Furthermore, the paper explores the integration of machine learning into existing BI tools, highlighting the advantages, disadvantages, and challenges that arise in this process. Special attention is given to ethical and security aspects of using machine learning in BI, including the consideration of ethical frameworks and security strategies. Lastly, the paper discusses the future of business intelligence with machine learning and concludes by emphasizing the significance of this technological revolution for making informed business decisions.

Keywords: databases, business intelligence, machine learning, deep learning, algorithms, data analysis, integration, ethics, security, sustainability.

Sadržaj

1. Uvod.....	1
2. Poslovna Inteligencija (BI)	2
2.1. Uvod u Poslovnu Inteligenciju.....	2
2.2. Procesi poslovne inteligencije.....	2
2.3. Arhitektura i komponente BI sustava.....	4
2.4. Ciljevi i koristi BI-a	5
3. Strojno učenje	7
3.1. Povijest Strojnog Učenja.....	7
3.2. Uvod u strojno učenje	7
3.3. Ključni koncepti strojnog učenja	8
3.4. Duboko učenje (Deep learning)	9
4. Algoritmi strojnog učenja u analizi podataka	10
4.1. Klasifikacija algoritama strojnog učenja.....	10
4.2. Linearna regresija.....	12
4.3. Logistička regresija.....	14
4.4. Stablo odlučivanja.....	15
4.5. K-srednjih vrijednosti (K-Means Clustering)	17
4.6. Slučajne šume	18
4.7. K-najbližih susjeda (K-Nearest Neighbors).....	20
4.8. Naivni Bayes	21
5. Primjeri upotrebe algoritama strojnog učenja u BI.....	23
5.1. Linearna regresija u BI.....	23
5.2. Logistička regresija u BI.....	23
5.3. Stablo odlučivanja u BI.....	24
5.4. K-srednjih vrijednosti (K-Means Clustering) u BI	24
5.5. Slučajne šume (Random Forests) u BI.....	24
5.6. K-najbliži susjed (K-Nearest Neighbors, KNN) u BI.....	24
5.7. Naivni Bayes u BI.....	25
6. Integracija strojnog učenja u postojeće BI alate	26
6.1. Potrebni koraci za implementaciju ML algoritama u BI alate	26
6.2. Microsoft Power BI za integraciju Strojnog Učenja	27
6.3. Prikaz Power BI insights i Power BI Key Influencers	28
6.4. Prednosti Integracije Strojnog Učenja u BI alate.....	34
6.5. Nedostaci Integracije Strojnog Učenja u BI alate	35

6.6.	Izazovi Integracije Strojnog Učenja u BI alate	35
7.	Etika i sigurnost korištenja strojnog učenja u BI	37
7.1.	Etički Aspekti.....	37
7.2.	Sigurnosni Izazovi	37
7.3.	Etički Okviri i Sigurnosne Strategije	38
8.	Budućnost BI-a uz strojno učenje	39
9.	Zaključak.....	40
10.	Sadržaj Slika	41
11.	Reference	42

1. Uvod

Poslovna inteligencija (BI) predstavlja ključnu komponentu suvremenog poslovanja i informacijske tehnologije. U današnjem globaliziranom poslovnom okruženju, organizacije se suočavaju s ogromnim količinama podataka koje generiraju interni i eksterni izvori. Upravljanje ovim podacima i izvlačenje vrijednih informacija iz njih postalo je imperativ za uspješno donošenje poslovnih odluka. U tom kontekstu, poslovna inteligencija predstavlja sustavski pristup prikupljanju, obradi, analizi i interpretaciji podataka s ciljem pružanja uvida i informacija potrebnih za optimizaciju poslovnih procesa, poboljšanje performansi te osiguranje konkurentske prednosti.

U skladu s tim, ovaj završni rad temelji se na dubinskom istraživanju uloge strojnog učenja u okviru poslovne inteligencije. Strojno učenje, kao grana umjetne inteligencije, revolucioniralo je način na koji organizacije obrađuju i analiziraju podatke. Strojno učenje je ključno sredstvo za pronalaženje skrivenih uzoraka, identifikaciju trendova te predviđanje budućih događaja na temelju ogromnih datasetova.

Ovaj završni rad razmotrit će osnovne koncepte poslovne inteligencije, uključujući procese prikupljanja podataka, transformacije, pohrane i analize. Istovremeno, istražiti ćemo značaj integracije strojnog učenja u BI sustave, naglašavajući kako strojno učenje obogaćuje tradicionalne BI tehnike s naprednim algoritmima za obradu podataka. Razmotrit ćemo primjere primjene strojnog učenja u poslovnom okruženju, uključujući personalizaciju usluga, optimizaciju lanca opskrbe te upravljanje rizicima.

S obzirom na važnost ovog područja, istražiti ćemo i izazove koji prate integraciju strojnog učenja u BI, uključujući pitanja etičnosti, sigurnosti podataka i transparentnosti algoritama. Osim toga, analizirat ćemo prednosti koje donosi ovakva integracija, uključujući poboljšane prediktivne analize, brže donošenje odluka i veću konkurentska prednost na tržištu.

U konačnici, ovaj završni rad bavi se i budućnošću poslovne inteligencije uz strojno učenje. S obzirom na neprestani razvoj tehnologije i sve veću dostupnost podataka, ovo dinamično područje obećava nove mogućnosti i izazove. Rad će završiti pozivom na daljnje istraživanje i istraživanjem potencijala koji strojno učenje ima za transformaciju načina na koji organizacije analiziraju, razumiju i koriste svoje podatke.

2. Poslovna Inteligencija (BI)

2.1. Uvod u Poslovnu Inteligenciju

Poslovna inteligencija (BI), kao ključna komponenta suvremenog poslovanja, označava sveobuhvatni skup procesa, tehnika i alata čija svrha je omogućiti organizacijama da prikupe, analiziraju i interpretiraju podatke kako bi donosile informirane i strategijske poslovne odluke. Ovaj koncept temelji se na ključnoj pretpostavci da su podaci neprocjenjiva imovina organizacije te da njihova pažljiva analiza i učinkovito korištenje pružaju značajnu prednost u konkurenciji na tržištu. (Negash, S. 2004.)

U osnovi, BI pristup podrazumijeva sustavno prikupljanje podataka iz raznolikih izvora, uključujući interne izvore kao što su interni sustavi baza podataka, ERP (Enterprise Resource Planning) sustavi, aplikacije i vanjske izvore kao što su tržišni podaci, društveni mediji, web stranice i senzorski uređaji. Ovi podaci, kad se sakupljaju i centraliziraju, postaju sirov materijal koji se dalje obrađuje kako bi postao koristan informacijski resurs.

Proces BI-a uključuje čišćenje i transformaciju podataka kako bi se uklonile eventualne pogreške, standardizirali formati, te kako bi se osigurala dosljednost i točnost. Zatim se ti podaci pohranjuju u centralnom skladištu podataka (poznatom kao Data Warehouse) koje služi kao centralno mjesto za pohranu i upravljanje podacima. Ovdje podaci postaju dostupni za analizu, ažuriraju se redovito i čuvaju se na siguran način.

Ovaj sveobuhvatni BI pristup omogućuje organizacijama da stvore temelje za daljnje analize podataka. Analiza se može provoditi pomoću različitih tehnika, uključujući statističke metode, rudarenje podataka (data mining) i strojno učenje. Cilj analize je otkriti obrasce, trendove i ključne pokazatelje performansi koji mogu pomoći organizaciji u razumijevanju trenutnih poslovnih izazova i prilika. (Sherman, R. 2015.)

Kada organizacija ima uvid u svoje podatke i može interpretirati njihovu vrijednost, to joj omogućuje bolje donošenje odluka na svim razinama upravljanja. Odluke mogu obuhvaćati aspekte kao što su planiranje proizvodnje, upravljanje zalihama, marketing strategije, optimizacija lanca opskrbe, prilagodba proizvoda i usluga prema potrebama tržišta i mnoge druge poslovne aspekte.

2.2. Procesi poslovne inteligencije

Proces poslovne inteligencije predstavlja ključni mehanizam pomoću kojeg organizacije prikupljaju, analiziraju i interpretiraju podatke kako bi donosile informirane poslovne odluke. Ovaj kontinuirani proces igra ključnu ulogu u unaprjeđenju poslovanja i stvaranju konkurentske prednosti.

Početna faza procesa je **prikupljanje podataka**. Organizacije identificiraju relevantne izvore podataka, bilo da su interni (kao što su baze podataka s informacijama o prodaji, financijama i ljudskim resursima) ili eksterni (uključujući tržišne analize, društvene medije i konkurentske podatke). Cilj je stvoriti sveobuhvatnu sliku poslovanja, prikupljajući podatke iz različitih izvora.

Nakon prikupljanja, podaci prolaze kroz fazu **čišćenja i transformacije**. Ovdje se suočavamo s izazovima poput grešaka u unosu, nedosljednostima u formatima i dupliciranim zapisima. Proces čišćenja i transformacije osigurava točnost i konzistenciju podataka. Ovo uključuje korekciju pogrešaka, uklanjanje duplikata i standardizaciju formata.

Zatim se čisti podaci pohranjuju u **centralnom skladištu podataka** (Data Warehouse). Skladište podataka optimizirano je za brz pristup i analizu. Podaci se organiziraju u tablice i čuvaju na način koji olakšava buduće upite i analize.

Slijedi **analiza podataka**, što je srce poslovne inteligencije. Različite tehnike i alati koriste se za istraživanje podataka i identifikaciju obrazaca, trendova i ključnih pokazatelja performansi. To uključuje statističke analize, rudarenje podataka i primjenu strojnog učenja. Cilj je dublje razumijevanje podataka i otkrivanje skrivenih informacija koje mogu biti ključne za poslovne odluke.

Kako bi se omogućila bolja interpretacija podataka, organizacije primjenjuju **vizualizaciju i izvješćivanje**. Analitički rezultati često nisu intuitivni za sve korisnike, pa se podaci prikazuju putem grafova, tablica i atraktivnih izvješća. Ovo olakšava komunikaciju informacija unutar organizacije i omogućava donositeljima odluka da brže razumiju relevantne podatke.

Na kraju, proces poslovne inteligencije kulminira fazom **donošenja odluka**. Na temelju analitičkih rezultata i vizualizacija, organizacija donosi informirane odluke koje mogu obuhvaćati prilagodbu strategija, budžeta, marketinških kampanja ili upravljanje zalihama. Ova faza je ključna za postizanje konkurentne prednosti i optimizaciju poslovanja. (Sherman, R. 2015.)

Važno je napomenuti da su procesi poslovne inteligencije iterativni i kontinuirani. Organizacije redovito provode ove korake kako bi održale konkurentne prednosti i prilagodile se promjenjivim uvjetima u poslovnom okruženju te to je konstanti krug ponavljanja tih koraka da se ostane relevantan u tržištu (slika 1).



Slika 1 Krug poslovne inteligencije

2.3. Arhitektura i komponente BI sustava

Poslovna inteligencija (BI) sustav predstavlja sveobuhvatni okvir koji organizacijama omogućuje prikupljanje, analizu i interpretaciju podataka kako bi donosile informirane poslovne odluke. Ova kompleksna mreža komponenata surađuje kako bi organizacijama omogućila bolje razumijevanje njihovih poslovnih operacija i stvaranje konkurentske prednosti. Ključne komponente BI sustava čine temeljnu infrastrukturu za ostvarivanje tih ciljeva. (Sherman, R. 2015.)

Izvori podataka predstavljaju ključnu polazišnu točku u BI procesu i igraju ključnu ulogu u obogaćivanju informacija koje organizacije koriste za donošenje odluka. Izvori podataka mogu biti izuzetno raznoliki i obuhvaćati širok spektar informacija. Unutar organizacije, interni izvori podataka često uključuju interne baze podataka koje sadrže informacije o transakcijama, zalihama, klijentima, zaposlenicima i drugim aspektima poslovanja. Osim toga, aplikacije koje organizacija koristi za upravljanje različitim poslovnim funkcijama, poput ERP (Enterprise Resource Planning) sustava ili CRM (Customer Relationship Management) sustava, također služe kao ključni interni izvori podataka. Ti sustavi generiraju obilježja o poslovnim aktivnostima i međusobnim interakcijama.

S druge strane, organizacije sve više prepoznaju vrijednost senzora u svojim operacijama. Senzori, uključujući IoT (Internet of Things) uređaje, prikupljaju podatke o različitim aspektima, kao što su temperatura, vlažnost, brzina, lokacija i još mnogo toga. Ovi podaci mogu biti korisni u razumijevanju performansi proizvoda, praćenju uvjeta u stvarnom vremenu i optimizaciji procesa.

Također, ne smijemo zaboraviti izvore podataka izvan organizacije, koji također imaju značajan utjecaj na BI. To uključuje društvene medije, gdje organizacije prate povratne informacije i sentiment kupaca putem platformi kao što su Facebook, Twitter i LinkedIn. Analiza društvenih medija omogućuje organizacijama da razumiju kako ih percipira šira javnost i prilagode svoje strategije. Web stranice su također izvor podataka koji može pružiti vrijedne informacije o posjetiteljima, njihovim preferencijama i ponašanju. Analiza web prometa može pomoći organizacijama u optimizaciji iskustva korisnika i marketinških kampanja. Tržišne analize uključuju prikupljanje podataka o konkurenciji, tržištu i trendovima u industriji. Ovi podaci omogućuju organizacijama da ostanu informirane o promjenama u okruženju i prilagode svoje strategije kako bi ostale konkurentske.

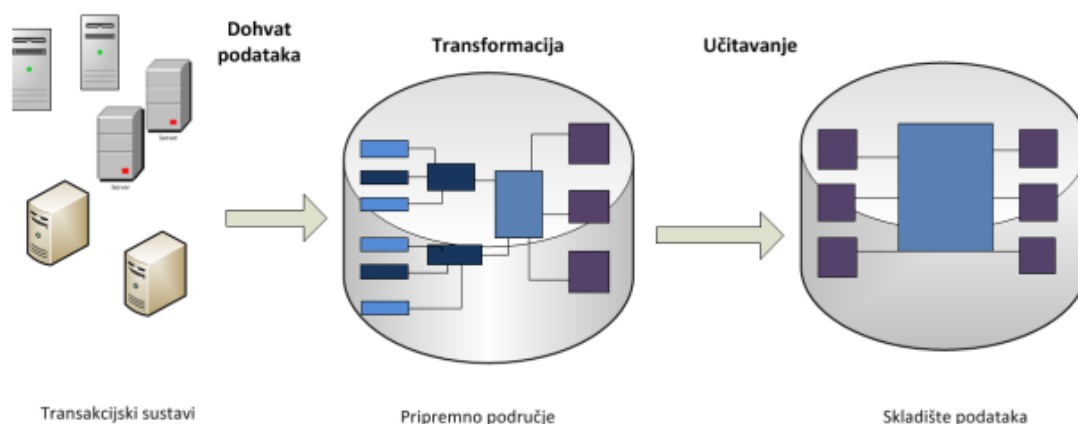
Nakon prikupljanja podataka, sljedeći logični korak je njihovo spremanje, odnosno skladištenje. Skladište podataka (Data Warehouse) predstavlja srce i temelj poslovne inteligencije. Ovo centralno skladište podataka igra ključnu ulogu u organizaciji, upravljanju i pripremi podataka za analizu i donošenje odluka. Ovdje se prikupljeni podaci ne samo pohranjuju, već prolaze kroz procese čišćenja, transformacije i organizacije kako bi postali korisni i upotrebljivi informacijski resurs.

ETL (Extract, Transform, Load) proces je bitan korak u pripremi podataka za analizu. Ova procedura omogućuje prijenos podataka iz izvora u skladište podataka. Prvo, podaci se "izvlače" (extract) iz različitih izvora. Zatim se podvrgavaju procesu "transformacije" (transform), gdje se čiste, standardiziraju i oblikuju kako bi se osigurala konzistentnost i kvaliteta podataka. Konačno, podaci se "učitavaju" (load) u skladište podataka. Ovaj proces osigurava da podaci u skladištu budu dosljedni, pouzdani i spremni za analizu. (slika 2)

BI alati su ključni za analizu podataka i stvaranje izvješća. Ovi alati omogućuju korisnicima da pristupe podacima, postavljaju upite, provode analize i stvaraju vizualne prikaze podataka. Ovi alati često uključuju

možnosti za izvješćivanje, grafove, dashboarde i mogućnosti za dubinsku analizu podataka. Oni su ključni za pretvaranje podataka u korisne informacije.

Poslovni sloj (Business Layer) pruža infrastrukturu za definiranje poslovnih pravila, logike i metapodataka. Ovo omogućuje organizacijama da precizno definiraju kako će se podaci koristiti u analizi i donošenju odluka. Poslovni sloj često sadrži definicije ključnih pokazatelja performansi i pruža strukturu za organiziranje podataka prema poslovnim potrebama te zadnji korak je izrada korisničkog sučelja, korisničko sučelje predstavlja most između korisnika i podataka. Ovo sučelje omogućuje korisnicima da pristupe podacima, postavljaju upite, pregledavaju izvješća i vizualizacije te donose odluke na temelju analize podataka. Intuitivno i interaktivno korisničko sučelje ključno je za olakšavanje korištenja BI sustava od strane različitih korisničkih profila unutar organizacije. (Sherman, R. 2015.)



Slika 2 Proces skupljanja, transformacije i skladištenja podatka

2.4. Ciljevi i koristi BI-a

Implementacija poslovne inteligencije (BI) predstavlja ključan korak za organizaciju, donoseći niz ciljeva i koristi koje oblikuju njezinu operativnu izvrsnost i konkurentske prednosti. Ovdje ćemo detaljno razmotriti te ciljeve i koristi. (Sherman, R. 2015.)

- **Bolje odlučivanje.** BI se ističe kao moćan alat za unapređenje procesa donošenja odluka u organizaciji. Analiza podataka i izvještavanje omogućuju donositeljima odluka da temelje svoje odluke na činjenicama i dubinskoj analizi. Time se smanjuje rizik od donošenja loših odluka koje bi mogle imati negativne posljedice za poslovanje. Kroz BI, organizacija stvara jasnu sliku svog trenutnog stanja i budućih mogućnosti.
- **Povećana efikasnost i produktivnost.** Jedna od ključnih koristi BI-a je mogućnost analize poslovnih procesa. Organizacije koriste BI alate kako bi identificirale potencijalna poboljšanja i efikasnost u operacijama. To uključuje smanjenje nepotrebnih troškova, optimizaciju procesa, smanjenje vremena potrebnog za obavljanje zadataka te bolje raspoređivanje resursa. Kroz ovo poboljšanje efikasnosti, organizacije postaju konkurentnije i fleksibilnije.

- **Bolje razumijevanje kupaca.** Poslovna inteligencija omogućuje duboko razumijevanje potreba, preferencija i ponašanja kupaca. Analiza podataka o kupcima pomaže organizacijama da stvore segmentirane profile kupaca i prilagode svoje proizvode i usluge prema tim specifičnim potrebama. Ovo personalizirano iskustvo često rezultira povećanom lojalnošću kupaca i povećanim prihodima.
- **Povećana konkurentska prednost.** BI omogućuje organizacijama da prate tržište, konkurenciju i trendove u stvarnom vremenu. Analizom konkurencije, organizacije identificiraju prilike za inovaciju i razvoj novih proizvoda ili usluga. Također mogu bolje razumjeti promjene u potražnji i prilagoditi svoju strategiju tržištu, stvarajući tako održivu konkurentske prednosti.
- **Poboljšana kontrola nad troškovima.** Praćenjem finansijskih podataka i detaljnom analizom troškova, organizacije postižu bolju kontrolu nad svojim resursima i finansijskim planiranjem. Ovo je ključno za održavanje stabilnosti i održivosti poslovanja. BI alati omogućuju upravama i finansijskim timovima da identificiraju potencijalne troškovne uštede, prepoznaju nepotreban gubitak i bolje planiraju buduće investicije.

U zaključku, implementacija poslovne inteligencije nije samo tehnološka investicija, već strateški potez koji omogućuje organizaciji da postigne brojne ciljeve i ostvari ključne koristi. Od boljeg odlučivanja do povećane konkurentske prednosti, BI postaje temelj uspješnog poslovanja u modernom poslovnom okruženju.

3. Strojno učenje

3.1. Povijest Strojnog Učenja

Strojno učenje ima svoje korijene u polju umjetne inteligencije (AI) i razvijalo se tijekom više desetljeća. Povijest strojnog učenja možemo podijeliti na nekoliko ključnih faza (Foote, K. D. 2021):

Rani Temelji (1940-1950): U ranim danima računarstva, istraživači su eksperimentirali s konceptima strojnog učenja, uključujući neuronske mreže. Alan Turing, čuveni britanski matematičar i računalni znanstvenik, doprinio je razvoju teoretskog okvira za strojno učenje svojim radom na "Turingovom testu".

Simboličko Učenje (1950-1960): U ovoj fazi, dominiralo je simboličko učenje koje se temeljilo na logičkim pravilima i simbolima. Programi su bili usmjereni na manipulaciju simbola kako bi se donosile odluke.

Pioniri u Neuralnim Mrežama (1960-1970): Istraživanja neuronskih mreža, inspiriranih biološkim mozgom, dovela su do pokušaja razvoja modela za učenje. Frank Rosenblatt je 1957. godine predstavio perceptron, jednostavan model neuronske mreže.

Zastoj u Financiranju (1970-1990): U razdoblju od 1970-ih do 1990-ih, interes za strojno učenje opao je zbog ograničenih resursa i neuspjeha u postizanju očekivanih rezultata. Ovo je razdoblje nazvano "zastojem u financiranju" za AI i strojno učenje.

Razvoj Algoritama (1990-2000): Razvoj novih algoritama i tehnika za strojno učenje ponovno je privukao pažnju istraživača. Algoritmi poput support vector machines (SVM) i decision trees postali su popularni za različite zadatke.

Uspon Dubokog Učenja (2010- danas): Duboko učenje, posebno duboke neuronske mreže, doživjelo je revoluciju učinkovitosti i primjena. Ovi modeli postigli su izvanredne rezultate u zadacima prepoznavanja slika, obradi prirodnog jezika i drugim područjima. Razvoj grafičkih procesora (GPU) pridonio je ubrzanju treniranja dubokih modela. (Foote, K. D. 2021.)

3.2. Uvod u strojno učenje

Strojno učenje, jedna od najdinamičnijih i najutjecajnijih grana umjetne inteligencije, predstavlja izazovno i napredno područje koje ima dubok utjecaj na mnoge aspekte našeg svakodnevnog života. Ova disciplina ima potencijal transformirati način na koji razumijemo, analiziramo i rješavamo različite probleme, od medicinske dijagnoze do vožnje autonomnih vozila i personaliziranih preporuka na internetu.

Strojno učenje se temelji na osnovnoj ideji da računalni sustavi trebaju biti sposobni učiti i poboljšavati svoje performanse iz iskustva, umjesto da budu programirani za obavljanje specifičnih zadataka. Ovaj pristup postavlja temeljni preokret u načinu na koji računalni sustavi procesiraju informacije i donose odluke, omogućujući im da se prilagode raznolikim situacijama i rješavaju probleme za koje nisu bili unaprijed programirani. (Flach, P. A. 2012.)

Jedna od ključnih karakteristika strojnog učenja je njegova sposobnost generalizacije. To znači da modeli razvijeni pomoću strojnog učenja nisu dizajnirani samo za rješavanje konkretnih problema na kojima su

trenirani, već su sposobni primjenjivati svoje naučeno znanje na nove i nepoznate situacije. Ova sposobnost generalizacije omogućuje strojnom učenju da se primijeni u širokom spektru primjena, od prepoznavanja rukopisa do analize tržišta dionica. (Flach, Peter A. 2012)

3.3. Ključni koncepti strojnog učenja

Strojno učenje se temelji na fundamentalnim konceptima i terminima koji omogućuju računalima da uče, razumiju podatke i donose odluke. Ovaj multidisciplinarni pristup obuhvaća različite aspekte, a nekoliko ključnih pojmova ističe se kao temeljni.

Podaci su srž strojnog učenja. Oni predstavljaju informacije koje računalno koristi za stvaranje modela i donošenje zaključaka. Podaci mogu biti strukturirani, poput tablica s redcima i stupcima, što se često nalazi u bazama podataka ili Excel tablicama. Također mogu biti nestrukturirani, u obliku slika, tekstualnih dokumenata, zvuka ili videa.

Modeli su računalni algoritmi koji se treniraju na temelju podataka kako bi prepoznali obrasce i donosili predviđanja ili odluke. Ovi modeli mogu biti jednostavni, kao što su linearna regresija ili logistička regresija, ili složeni, poput dubokih neuronskih mreža, karakterističnih za duboko učenje. (Goodfellow, I., Bengio, Y., & Courville, A. 2015)

Trening je ključan proces u razvoju modela. Tijekom treninga, model analizira podatke i prilagođava svoje unutarnje parametre kako bi minimizirao grešku u predviđanjima. Trening omogućuje modelu da "uči" i stječe sposobnost za obavljanje specifičnih zadataka.

Evaluacija se provodi kako bi se ocijenila sposobnost modela da donosi točne odluke. Nakon treninga, model se testira na nepoznatim podacima, a rezultati se mjere različitim metrikama performansi kao što su točnost, preciznost i odziv.

Strojno učenje se također klasificira prema različitim paradigama koji su nadzirano učenje, nenadzirano te pojačano. Nadzirano učenje uključuje modele koji uče na temelju označenih primjera, što znači da podaci uključuju i ulazne podatke i odgovarajuće izlazne podatke. Nenadzirano učenje se koristi kada modeli samostalno otkrivaju obrasce u podacima bez označenih informacija. Primjeri uključuju klasteriranje podataka i smanjenje dimenzionalnosti. Pojačano učenje temelji se na interakciji s okolinom i nagrađivanju ili kažnjavanju modela za njegove akcije. To se često koristi u autonomnim vozilima i raznim igrama.

Dodatno, duboko učenje se ističe kao područje unutar strojnog učenja koje koristi duboke neuronske mreže kako bi naučilo reprezentacije podataka na različitim razinama apstrakcije. Ovo područje postiglo je značajne uspjehe u zadacima kao što su prepoznavanje slika, obrada prirodnog jezika i mnoge druge aplikacije.

Razumijevanje ovih ključnih koncepta i terminologije ključno je za uspješno primjenjivanje strojnog učenja u različitim domenama, od medicinske dijagnoze i financijske analize do analize teksta i autonomne navigacije.

3.4. Duboko učenje (Deep learning)

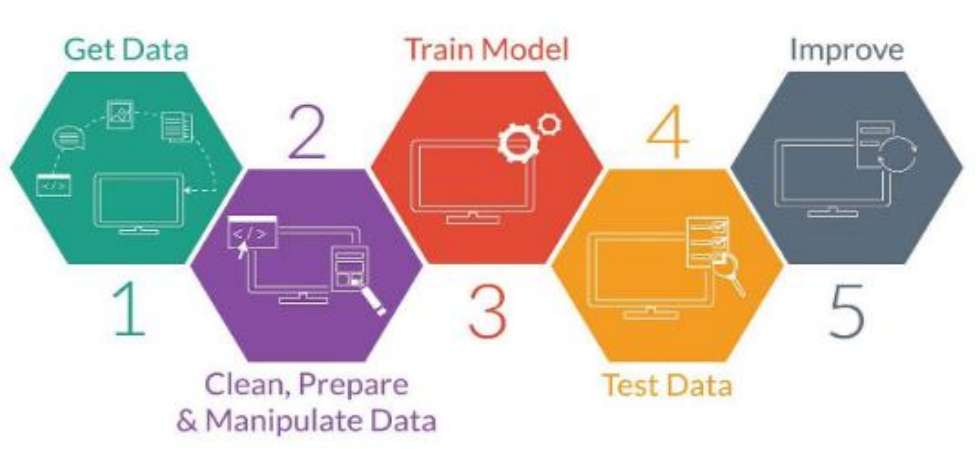
Duboko učenje, poznato i kao deep learning, predstavlja iznimno dinamično i utjecajno područje u okviru strojnog učenja i umjetne inteligencije. Ono se ističe kao snažan alat s dubokim utjecajem na različite aspekte našeg svakodnevnog života i industrije. Ovo područje ima potencijal transformirati način na koji razumijemo, analiziramo i rješavamo širok spektar problema, bilo da se radi o medicinskoj dijagnozi, autonomnoj vožnji ili personaliziranim preporukama na internetu. (Goodfellow, I., Bengio, Y., & Courville, A. 2015)

U srži dubokog učenja nalaze se neuronske mreže, složeni matematički modeli koji su inspirirani biološkim neuronima i sposobni su za obradu velikih količina podataka. Ovi modeli sadrže različite slojeve, uključujući ulazne, skrivene i izlazne slojeve. Skriveni slojevi omogućuju modelima da nauče dublje reprezentacije podataka, čineći ih sposobnima za izvlačenje složenih obrazaca.(slika 3)

Konvolucijske mreže (CNN) su poseban tip neuronskih mreža razvijen za obradu slika i matrica podataka. One koriste konvolucijske slojeve kako bi naučile lokalne značajke, kao što su rubovi, oblici i teksture, što ih čini izuzetno korisnima u zadacima poput prepoznavanja objekata u slikama ili analizi medicinskih snimaka.

Rekurentne neuronske mreže (RNN) su, s druge strane, posebno prilagođene za sekvencijalne podatke poput vremenskih nizova i prirodnog jezika. One koriste rekurentne slojeve koji omogućuju modelima da zadrže unutarnje stanje i razumiju kontekst sekvencijalnih podataka. (Goodfellow, I., Bengio, Y., & Courville, A. 2015)

Primjene dubokog učenja su raznolike i obuhvaćaju mnoga područja. To uključuje prepoznavanje slika, medicinsku dijagnozu, autonomna vozila, obradu prirodnog jezika, analizu vremenskih nizova i generiranje sadržaja kao što su slike, glazba i tekst. Duboko učenje kontinuirano napreduje, a razumijevanje ovih ključnih koncepata i njihovih primjena igra ključnu ulogu u rješavanju izazovnih problema i stvaranju inovativnih tehnoloških rješenja.



Slika 3 Proces Deep learning-

4. Algoritmi strojnog učenja u analizi podataka

4.1. Klasifikacija algoritama strojnog učenja

Klasifikacija algoritama strojnog učenja predstavlja temeljni korak u razumijevanju različitih pristupa u ovoj dinamičnoj disciplini. Ova sekcija će istražiti raznolike kategorije algoritama strojnog učenja i njihove ključne karakteristike kako bismo dobili dublji uvid u njihovu primjenu.

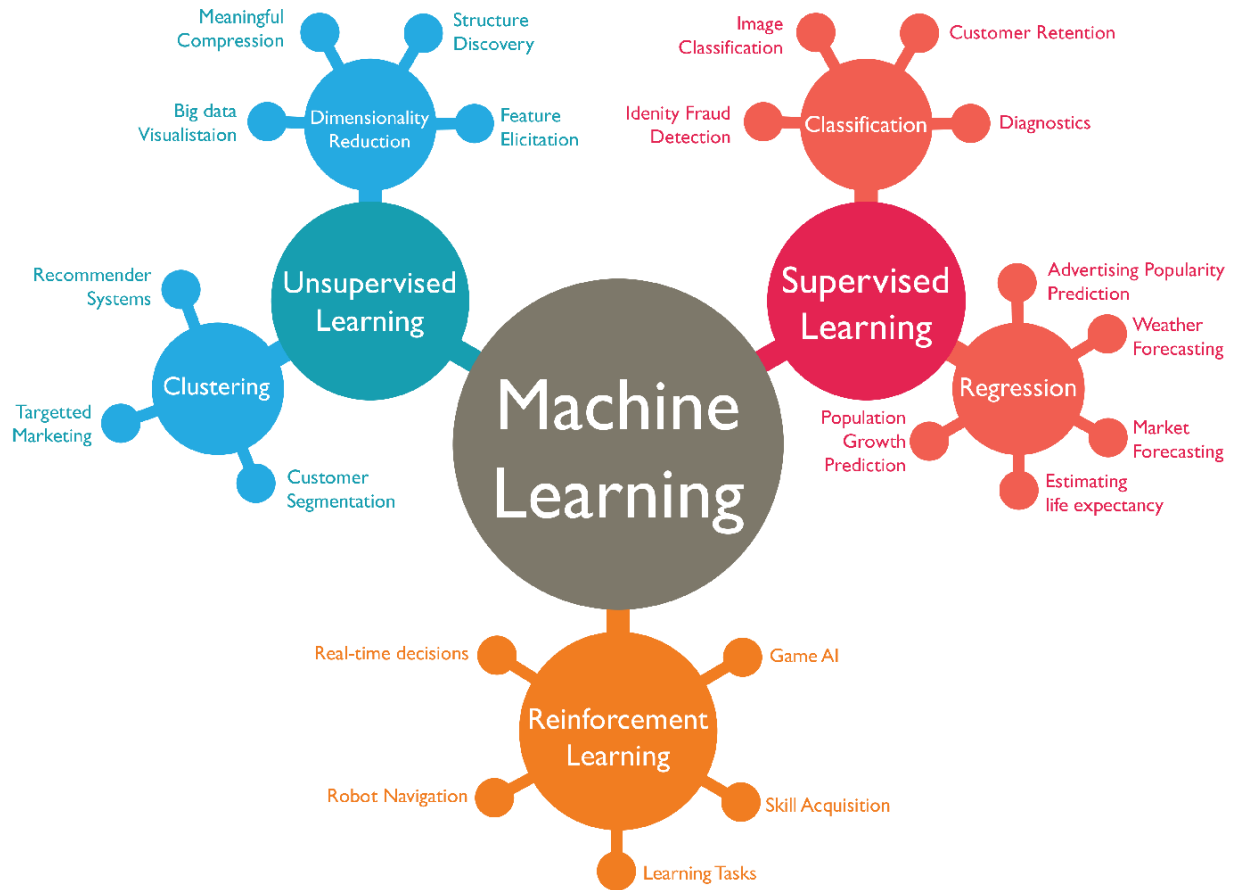
Regresija je ključna klasa algoritama strojnog učenja koja se koristi za predviđanje kontinuiranih vrijednosti. Ova grana uključuje niz tehnika koje omogućuju modeliranje i predviđanje numeričkih podataka. Na primjer, linearna regresija se koristi za identifikaciju linearnih odnosa između varijabli, dok polinomijalna regresija proširuje ovu sposobnost na nelinearne odnose u podacima (Bishop, C. M. 2006.)

Klasifikacija (slika 5) je proces kategorizacije podataka u diskretne razrede ili grupe. U ovoj kategoriji algoritama, istražujemo metode koje omogućuju računalu da donese odluke o pripadnosti podataka određenim klasama. Logistička regresija, koja se često koristi za binarnu klasifikaciju, i support vector machines (SVM), koji traže optimalnu razdjelnicu između klasa, samo su neki od algoritama koji će biti detaljnije razmotreni. (Bishop, C. M. 2006.)

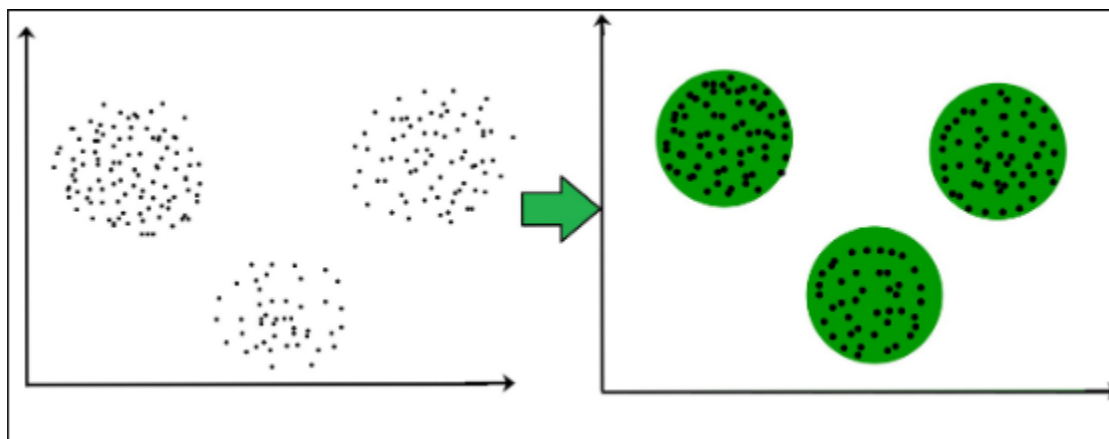
Grupiranje je tehnika koja se koristi za identifikaciju sličnih entiteta ili uzoraka u skupu podataka. K-means klasterizacija, jedna od najpoznatijih metoda, grupira podatke u određeni broj klastera, dok hijerarhijsko grupiranje stvara hijerarhijsku strukturu grupa. Ova kategorija algoritama omogućuje nam otkrivanje skrivenih obrazaca u podacima, što je ključno za razumijevanje njihove strukture. (Bishop, C. M. 2006.)

Smanjenje dimenzionalnosti je važna tehnika u analizi podataka koja pomaže u obradi velikih i složenih skupova podataka. Smanjenjem broja značajki olakšava se analiza i interpretacija podataka, a istovremeno se može poboljšati performansa modela. Metode kao što je Principal Component Analysis (PCA) omogućuju projekciju podataka u manje dimenzije, čime se zadržava većina varijabilnosti u podacima. Također, tehnike za selekciju značajki omogućuju odabir najvažnijih značajki za analizu.

Reinforcement učenje predstavlja paradigmu strojnog učenja koja se bavi procesom odlučivanja i kontrolom. Ovaj pristup omogućuje računalima da nauče kako donositi odluke na temelju iskustva i interakcije s okolinom. Primjene ovog tipa uključuju autonomna vozila koja uče kako voziti kroz iskustvo i igre u kojima agenti uče kako postupiti kako bi maksimizirali svoje nagrade.



Slika 4 Vrste Algoritama strojnog učenja



Slika 5 Klasterziacija u ML

4.2. Linearna regresija

Linearna regresija je statistička metoda koja se koristi za modeliranje linearnog odnosa između nezavisnih i zavisne varijable, s ciljem predviđanja ili objašnjavanja vrijednosti zavisne varijable. Može biti jednostavna, s jednom nezavisnom varijablom, ili višestruka, s više nezavisnih varijabli. Naučava se određivanjem parametara koji minimiziraju grešku između stvarnih i predviđenih vrijednosti, a evaluira se pomoću metrika poput MSE i R-squared. Model se koristi za predviđanje vrijednosti na temelju novih podataka, no zahtijeva pretpostavke o linearnoj neovisnosti i normalnoj distribuciji greške. Regularizirane verzije modela se koriste za suočavanje s izazovima poput overfittinga, čineći linearnu regresiju ključnom tehnikom u mnogim disciplinama, uključujući ekonomiju, medicinu i inženjering. (Bishop, C. M. 2006.)

U početku, imamo dvije skupine podataka - nezavisnu varijablu (X) i zavisnu varijablu (Y). Osnovna svrha linearne regresije jest pronaći odnos između ovih varijabli i omogućiti predviđanje vrijednosti Y na temelju poznate vrijednosti X. Sljedeći korak je inicijalizacija parametara, parametri modela su nagib (a) i odsjek (b). Nagib (a) predstavlja brzinu promjene zavisne varijable u odnosu na nezavisnu varijablu, dok odsjek (b) označava vrijednost Y kada je X jednako nuli. U početku postavljamo ove parametre na nulu. Sljedeći korak je naučiti model. To znači da ćemo pronaći optimalne vrijednosti za parametre a i b koji minimiziraju razliku između stvarnih vrijednosti Y i predviđenih vrijednosti modela. Onda moramo izračunati parametre a i b oni se računaju pomoću matematičkih formula koje se temelje na statističkim izračunima. Ovisno o skupu podataka, koristi se formula za nagib (a) i odsjek (b) kako bi se pronašle najbolje vrijednosti koje najbolje odgovaraju podacima. Nakon izračuna parametara a i b, model je naučen i spreman za korištenje. Predviđanje: Kada imamo novu vrijednost za nezavisnu varijablu (novi_X), možemo koristiti naučeni model da bismo predvidjeli odgovarajuću vrijednost za zavisnu varijablu (predviđena_Y). Ovo nam omogućuje da donesemo predviđanja na temelju naučenog modela. Važno je napomenuti da stvarne primjene linearne regresije često uključuju složenije postupke kao što su analiza podataka, provjera pretpostavki modela i obrada outliersa kako bi se osigurala pouzdanost i točnost predviđanja. Linearne regresije su moćna tehnika u mnogim područjima, kao što su ekonomija, medicina i znanost, jer omogućuju modeliranje i predviđanje odnosa između varijabli. (slika 6)

```

# Ulazni podaci
X - Nezavisna varijabla
Y - Zavisna varijabla

# Inicijalizacija parametara
a = 0      # Nagib (slope) početno postavljen na 0
b = 0      # Odsjek (intercept) također početno postavljen na 0

# Naučavanje modela
N = broj_uzoraka(X) # Broj uzoraka u skupu podataka
suma_X = suma(X)
suma_Y = suma(Y)
suma_XY = suma(X * Y)
suma_X_kvadrat = suma(X * X)

# Računanje parametara a i b
a = (N * suma_XY - suma_X * suma_Y) / (N * suma_X_kvadrat - suma_X * suma_X)
b = (suma_Y - a * suma_X) / N

# Model je naučen, a i b su pronađeni

# Sada možemo koristiti model za predviđanje
novi_X = novi_podaci()
predviđena_Y = a * novi_X + b

```

Slika 6 Pseudokod za Linearnu regresiju (Bishop, C. M. 2006.)

4.3. Logistička regresija

Logistička regresija je statistička metoda koja se koristi za modeliranje odnosa između kategoričke zavisne varijable i jedne ili više nezavisnih varijabli kako bismo klasificirali primjere u različite kategorije. Iako se često naziva "regresija," logistička regresija se primarno koristi za klasifikaciju, posebno za binarnu klasifikaciju, gdje se primjeri razvrstavaju u jednu od dvije kategorije, kao što su pozitivno/negativno ili prisutan/odsutan. Osnovna ideja je modeliranje vjerojatnosti da će zavisna varijabla (Y) pripadati jednoj od klasa kao funkciju nezavisnih varijabli (X) pomoću logističke funkcije, koja transformira linearnu kombinaciju nezavisnih varijabli u raspon vrijednosti između 0 i 1. Parametri modela (a i b) se uče iz podataka pomoću metode maksimalne izglednosti. Logistička regresija je široko korištena u problemima klasifikacije, kao što su medicinska dijagnoza, spam filtriranje, analiza rizika i mnogi drugi, zbog svoje efikasnosti i jednostavnosti interpretacije rezultata. (Bishop, C. M. 2006.)

Na početku imamo skup podataka s nezavisnim varijablama (X) i kategoričkom zavisnom varijablom (Y). Cilj logističke regresije je predvidjeti klasu Y (0 ili 1) na temelju nezavisnih varijabli. Parametri modela se inicijaliziraju na početne vrijednosti. Parametri uključuju a i b koji se odnose na linearnu kombinaciju nezavisnih varijabli i stopu učenja (learning_rate) koja kontrolira veličinu koraka u optimizaciji parametara. Također, postavljamo broj iteracija (broj_iteracija) za proces učenja. Logistička regresija koristi sigmoidnu funkciju kako bi se transformirala linearna kombinacija nezavisnih varijabli u vjerojatnost koja se nalazi u rasponu između 0 i 1. Sigmoidna funkcija omogućava da dobijemo vjerojatnost da će Y biti jednako 1. Nakon inicijalizacije, slijedi faza učenja modela. U petlji koja se izvršava kroz broj iteracija, izračunava se linearna kombinacija, vjerojatnost, te gradijent za a i b pomoću funkcije sigmoid i stvarnih vrijednosti Y. Gradijent se koristi za ažuriranje parametara a i b putem postupka gradijentnog spusta kako bi se minimizirala greška modela. Nakon završetka iteracija, model je naučen, a optimalni parametri a i b su pronađeni. Kada imamo novi set nezavisnih varijabli (novi_X), možemo koristiti naučeni model za predviđanje pripadajuće klase. Prvo izračunavamo linearnu kombinaciju, zatim vjerojatnost pomoću sigmoidne funkcije, te konačno klasu (0 ili 1) ovisno o vjerojatnosti. Ovaj algoritam je osnova za implementaciju logističke regresije, koja se često koristi u problemima klasifikacije zbog svoje jednostavnosti i dobre interpretabilnosti rezultata. Važno je napomenuti da stvarna implementacija obično uključuje i dodatne korake kao što su provjera konvergencije, regularizacija i validacija modela kako bi se osigurala njegova pouzdanost i performanse. (slika 7) (Bishop, C. M. 2006.)

```

# Ulazni podaci
X - Nezavisne varijable
Y - Kategorička zavisna varijabla (0 ili 1)

# Inicijalizacija parametara
a - Inicijalizirati parametre na neku početnu vrijednost
b - Inicijalizirati parametre na neku početnu vrijednost
learning_rate - Stopa učenja (korak za ažuriranje parametara)
broj_iteracija - Broj iteracija za učenje

# Logistička funkcija (sigmoid)
sigmoid = funkcija(x):
    povrat 1 / (1 + e^(-x))

# Naučavanje modela
za svaku iteraciju u rasponu broj_iteracija:
    # Izračunavanje linearne kombinacije
    linearna_kombinacija = a*X + b

    # Izračunavanje vjerojatnosti
    vjerojatnost = sigmoid(linearna_kombinacija)

    # Izračunavanje gradijenta za a i b
    gradijent_a = (1/broj_primjera) * suma((vjerojatnost - Y) * X)
    gradijent_b = (1/broj_primjera) * suma(vjerojatnost - Y)

    # Ažuriranje parametara
    a = a - learning_rate * gradijent_a
    b = b - learning_rate * gradijent_b

# Model je naučen, a i b su pronađeni

# Sada možemo koristiti model za predviđanje
novi_X - Novi skup nezavisnih varijabli
linearna_kombinacija = a*novi_X + b
vjerojatnost = sigmoid(linearna_kombinacija)
predviđena_klasa = (vjerojatnost >= 0.5) ? 1 : 0

```

Slika 7 Logistička regresija pseudokod (Bishop, C. M. 2006.)

4.4. Stablo odlučivanja

Stablo odlučivanja je popularan algoritam u strojnom učenju koje se koristi za klasifikaciju i regresiju. Osnovna ideja je izgraditi stablo koje se sastoji od čvorova i grana, gdje svaki čvor predstavlja odluku temeljenu na vrijednosti određenih značajki, a svaka grana predstavlja mogući ishod te odluke. Proces izgradnje stabla počinje od korijenskog čvora, podijeliti podatke na temelju kriterija poput Gini indeksa ili entropije za klasifikacijske probleme te srednje kvadratne greške za regresijske probleme. Nakon podjele, proces se ponavlja za svaki podskup podataka na sljedećem nivou čvorova, uz mogućnost zaustavljanja prema definiranim kriterijima. Stablo odlučivanja je visoko interpretabilno, omogućavajući ljudima razumijevanje odluka koje su donesene, a također se često koristi kao osnovni građevni blok u ansambl metodama za poboljšanje učinkovitosti i generalizacije. Ova metoda ima široku primjenu u različitim područjima, uključujući medicinu, financije i sustave preporuka. (Bishop, C. M. 2006.)


```

Funkcija izgradnje_stabla(X, Y, dubina):
    Ako su svi primjeri u Y iste klase ili dubina doseže maksimalnu:
        Kreiraj list čvora sa najčešćom klasom u Y
    Inače:
        Odaberi najbolju značajku za podjelu (npr. Gini indeks, entropija)
        Kreiraj unutarnji čvor za odabranu značajku
        Podijeli primjere na temelju vrijednosti značajke
        Za svaki podskup primjera:
            Pozovi funkciju izgradnje_stabla sa (X_podskup, Y_podskup, dubina)
            Dodaj rezultat kao djecu trenutnom čvoru

Stablo = izgradnje_stabla(X_trening, Y_trening, dubina=0)

Funkcija predviđanja_stabla(X_test, čvor):
    Ako čvor je list:
        Vрати klasu čvora
    Inače:
        Na temelju vrijednosti značajke u X_test odaberi dijete čvora
        Pozovi funkciju predviđanja_stabla sa (X_test, odabrano_dijete)

Za svaki primjer u X_test:
    Predviđena_klasa = predviđanja_stabla(primjer, Stablo)
    Zabilježi Predviđena_klasa

# Model je naučen i spreman za predviđanje

```

Slika 8 Pseudokod za Stablo odlučivanja (Bishop, C. M. 2006.)

Pseudo kod za stablo odlučivanja prikazuje osnovni algoritam za izgradnju stabla odlučivanja i njegovu primjenu u predviđanju. Prvo, važno je razumjeti da stablo odlučivanja je hijerarhijska struktura koja se sastoji od čvorova i grana, a svaki čvor predstavlja odluku na temelju odabrane značajke, dok grane vode do podskupova podataka. Algoritam počinje s funkcijom `izgradnje_stabla`, koja je rekurzivna. U koraku 1, provjerava se da li su svi primjeri u trenutnom čvoru iste klase ili je dosegnuta maksimalna dubina stabla. Ako je to slučaj, stvara se list čvora s najčešćom klasom među primjerima u tom čvoru. Inače, odabire se

najbolja značajka za podjelu podataka, obično uz pomoć kriterija kao što su Gini indeks ili entropija. Nakon odabira značajke, stvara se unutarnji čvor i primjeri se dijele na temelju vrijednosti te značajke. Algoritam zatim rekurzivno poziva sam sebe za svaki podskup primjera, stvarajući podstabla i čvorove na dubljoj razini. Ovaj postupak se ponavlja dok se ne dosegnu kriteriji za zaustavljanje, kao što su maksimalna dubina stabla ili minimalni broj uzoraka u čvoru. Konačni rezultat je stablo odlučivanja koje se sastoji od čvorova i grana. Nakon izgradnje stabla, možemo koristiti funkciju predviđanja _stabla kako bismo predviđali klase novih primjera. Ta funkcija prolazi kroz stablo od korijena do lista, donoseći odluke na temelju vrijednosti značajki novih primjera. Stablo odlučivanja je visoko interpretabilna metoda koja se često koristi u strojnom učenju za klasifikaciju i regresiju zbog svoje jednostavnosti i sposobnosti za razumijevanje donesenih odluka. (slika 8) (Bishop, C. M. 2006.)

4.5. K-srednjih vrijednosti (K-Means Clustering)

K-srednjih vrijednosti (K-Means Clustering) je popularan algoritam za grupiranje podataka u skupine ili klase na temelju sličnosti među primjerima. Ovaj algoritam pripada obitelji neposrednih grupiranja (eng. "centroid-based clustering"), a osnovna ideja je grupirati podatke tako da primjeri unutar grupe budu što sličniji, dok se različite grupe razdvajaju na temelju sličnosti. Algoritam počinje s inicijalizacijom K centara, gdje je K broj grupa koje želimo identificirati. Centroidi su početne točke oko kojih će se formirati grupe. Zatim se iterativno izmjenjuju primjeri između grupa i centroidi se ponovno izračunaju tako da predstavljaju srednje vrijednosti primjera unutar svake grupe. Proces se ponavlja sve dok se centroidi i pripadnosti primjera konvergiraju. K-srednjih vrijednosti ima široku primjenu, uključujući segmentaciju tržišta, analizu slika, te smanjenje dimenzionalnosti podataka, a omogućuje razumijevanje skupova podataka i identifikaciju skrivenih uzoraka unutar njih. (Bishop, C. M. 2006.)

Pseudokod za K-srednjih vrijednosti (K-Means Clustering) predstavlja osnovni algoritam za grupiranje podataka. Algoritam počinje s inicijalizacijom K centara, što su početne točke oko kojih će se formirati grupe. Zatim slijedi iterativni proces u kojem se primjeri iz skupa podataka dodjeljuju grupama na temelju udaljenosti od centara, a zatim se centri grupa ažuriraju kako bi bolje odražavali srednje vrijednosti primjera unutar svake grupe. Ovaj proces se ponavlja dok se centroidi ne konvergiraju, tj. dok njihove pozicije prestanu značajno mijenjati. Konačni rezultat su formirane grupe i njihovi centroidi. K-srednjih vrijednosti je često korišten algoritam za grupiranje podataka zbog svoje jednostavnosti i efikasnosti, ali inicijalizacija centara i potreba za višestrukim pokretanjem algoritma kako bi se izbjegle zamke lokalnog minimuma su važni aspekti za uspješno grupiranje podataka. (Slika 9) (Bishop, C. M. 2006.)

```

# Inicijalizacija K centara
centroids = slučajna_inicijalizacija(K)

ponavljaj dok se centroidi ne konvergiraju:
    # Dodjela primjera grupama
    za svaki primjer u skupu podataka:
        odaberi najbliži centar iz centroids
        dodijeli primjer toj grupi

    # Ažuriranje centara grupa
    za svaku grupu:
        izračunaj novi centroid kao srednju vrijednost primjera u grupi

# Rezultat su formirane grupe i njihovi centroidi

```

Slika 9 Pseudokod za K-srednjih vrijednosti (Bishop, C. M. 2006.)

4.6. Slučajne šume

Slučajne šume (Random Forests) su moćan ansambl algoritam strojnog učenja koji se temelji na stablima odlučivanja. Ovaj algoritam radi tako da kombinira predikcije više stabala odlučivanja kako bi postigao bolju stabilnost i generalizaciju. Temelji se na tehnikama poput bagginga (bootstrap aggregating), gdje se uzorkuju različiti podskupovi podataka s povratkom iz skupa za treniranje kako bi se izgradila raznolika stabla odlučivanja. Slučajnost je ključna komponenta Slučajnih šuma, a to se ogleda u nasumičnom odabiru značajki tijekom izgradnje stabala, što pomaže u sprečavanju neprilagodbe. Konačna predikcija za novi primjer dobiva se većinskim glasanjem svih stabala u ansamblu. Slučajne šume su poznate po svojoj robusnosti, visokoj sposobnosti generalizacije i sposobnosti smanjenja varijance modela. Koriste se za različite zadatke kao što su klasifikacija, regresija, detekcija anomalija i smanjenje dimenzionalnosti, čineći ih široko primjenjivim algoritmom u strojnog učenju. (Shalev-Shwartz, S., & Ben-David, S. 2014)

Pseudokod za Slučajne šume (slika 10) predstavlja ansamblski algoritam strojnog učenja koji kombinira više stabala odlučivanja kako bi postigao bolje performanse i generalizaciju. Počinjemo s funkcijom "Slučajne_šume", koja prima trening podatke X_{trening} i odgovarajuće oznake Y_{trening} , broj stabala u ansamblu broj_stabala i broj značajki koji će se nasumično odabrati za svako stablo broj_značajki. Unutar petlje, za svako stablo odlučivanja u ansamblu, koristimo tehniku bagginga za stvaranje slučajnih podskupova X_{podskup} i Y_{podskup} iz trening skupa. Nakon toga, svako stablo se gradi na temelju ovih podskupova.

Nakon izgradnje svih stabala, imamo funkciju "Predviđanje_Slučajnih_šuma", koja prima ansambl stabala Slučajne_šume i testne primjere X_test. Unutar ove funkcije, za svako stablo u ansamblu, izračunavamo predikcije za svaki testni primjer. Konačna predikcija za svaki testni primjer dobiva se većinskim glasanjem među svim stablima u ansamblu.

Slučajne šume su popularne zbog svoje robusnosti, visoke sposobnosti generalizacije i smanjenja varijance. Ansambl ovih stabala često daje bolje performanse od pojedinačnih stabala odlučivanja i koristi se za različite zadatke strojnog učenja, uključujući klasifikaciju, regresiju i druge. Ova tehnika omogućava modelima da budu stabilniji i bolje se nose s raznolikim vrstama podataka. (Bishop, C. M. 2006.)

```
Funkcija Slučajne_šume(X_trening, Y_trening, broj_stabala, broj_značajki):  
    Za svako stablo od 1 do broj_stabala:  
        # Izbor slučajnih podskupova  
        X_podskup, Y_podskup = slučajni_uzorak(X_trening, Y_trening)  
  
        # Izgradnja stabla odlučivanja  
        Stablo = izgradnja_stabla(X_podskup, Y_podskup, broj_značajki)  
  
        Dodaj Stablo u Slučajne_šume  
  
    Vрати Slučajne_šume  
  
Funkcija Predviđanje_Slučajnih_šuma(Slučajne_šume, X_test):  
    Predviđenje = prazna_lista()  
    Za svako stablo u Slučajne_šume:  
        # Predviđanje za svako stablo  
        Predviđen_primjer = predviđanje_stabla(stablo, X_test)  
        Dodaj Predviđen_primjer u Predviđenje  
  
    # Konačna predikcija je većinsko glasanje  
    Konačna_predikcija = većinsko_glasanje(Predviđenje)  
  
    Vрати Konačna_predikcija
```

Slika 10 Pseudokod za slučajne šume (Bishop, C. M. 2006.)

4.7. K-najbližih susjeda (K-Nearest Neighbors)

K najbližih susjeda (K-Nearest Neighbors ili K-NN) je jednostavan i intuitivan algoritam strojnog učenja koji se temelji na principu da slični primjeri imaju slične oznake ili vrijednosti. Algoritam radi tako da za svaki nepoznati primjer traži K najbližih poznatih primjera u skupu podataka, koristeći mjere udaljenosti poput Euklidske udaljenosti. Za klasifikaciju, primjer se dodjeljuje klasi koja je većinska među K najbližih primjera, dok se za regresiju primjenjuje prosjek ili ponderirani prosjek vrijednosti K najbližih primjera. Važan hiperparametar je K, koji određuje broj susjeda korišten za donošenje odluka, pri čemu veći K čini model manje osjetljivim na šum, dok manji K omogućuje preciznije prilagodbe lokalnim varijacijama u podacima. K-NN je jednostavan za primjenu, ali ima ograničenja učinkovitosti kod velikih skupova podataka i osjetljiv je na nebalansirane skupove, stoga zahtijeva pažljiv izbor parametara i pripremu podataka. (Shalev-Shwartz, S., & Ben-David, S. 2014)

Pseudokod za K najbližih susjeda (K-NN) opisuje korake ovog algoritma za klasifikaciju i regresiju. Za svaki primjer u testnom skupu X_{test} , algoritam prvo prolazi kroz svaki primjer u trening skupu X_{training} kako bi izračunao udaljenost između testnog primjera i svih trening primjera. Ovaj korak uključuje primjenu odabrane mjere udaljenosti, kao što je Euklidska udaljenost, kako bi se kvantificirala razlika između primjera. Nakon što su sve udaljenosti izračunate, one se sortiraju u rastućem redoslijedu.

Zatim se odabiru K najbližih susjeda, tj. K primjera iz trening skupa s najmanjim udaljenostima prema testnom primjeru. U klasifikacijskom slučaju, K-NN određuje većinsku klasu među tim K najbližim susjedima i dodjeljuje je testnom primjeru kao predikciju. Za regresiju, umjesto toga, K-NN izračunava prosječnu vrijednost ciljne varijable među tim K najbližih susjeda i koristi je kao predviđenu vrijednost za testni primjer. (Bishop, C. M. 2006.)

Važno je napomenuti da je ključna hiperparametar ovog algoritma K, koji predstavlja broj susjeda koji se uzimaju u obzir. Veći K čini model manje osjetljivim na lokalni šum, dok manji K omogućuje modelu bolje prilagodbe lokalnim varijacijama u podacima. K-NN je jednostavan za razumjeti i primijeniti, ali zahtijeva pažljiv odabir vrijednosti K i prikladnu normalizaciju značajki kako bi pružio pouzdane rezultate za različite tipove podataka i zadatke.

```
Funkcija K_NN(X_trening, Y_trening, X_test, K):
Za svaki primjer u X_test:
Za svaki primjer u X_trening:
# Izračunaj udaljenost između trenirajućih primjera i testnog primjera
izračunaj udaljenost između primjera i testnog primjera

# Sortiraj primjere prema udaljenosti
sortiraj primjere prema udaljenosti u rastućem redoslijedu

# Odaberi K najbližih susjeda
K_najbližih_susjeda = prvih K primjera iz sortiranog redoslijeda

# Izračunaj većinsku klasu (za klasifikaciju) ili prosječnu vrijednost (za regresiju) među K najbližih susjeda
klasa_ili_vrijednost = većinska_klasa(K_najbližih_susjeda) ili prosječna_vrijednost(K_najbližih_susjeda)

# Dodaj klasa_ili_vrijednost u rezultate_predikcije

Vrati rezultate_predikcije
```

Slika 11 Pseudokod K najbližih susjeda (Bishop, C. M. 2006.)

4.8. Naivni Bayes

Naivni Bayes je probabilistički algoritam strojnog učenja koji se koristi za klasifikaciju i analizu teksta. Temelji se na Bayesovom teoremu i pretpostavci o neovisnosti značajki (stoga "naivni" u nazivu) te se često koristi u problemima klasifikacije teksta, kao što su filtriranje spam poruka ili analiza sentimenta. Ključne karakteristike Naivnog Bayesa uključuju sljedeće:

- Bayesov teorem: Algoritam koristi Bayesov teorem za izračun vjerojatnosti pripadnosti primjera određenoj klasi, na temelju vjerojatnosti značajki i vjerojatnosti klase. To se izražava kao: $P(\text{klasa}|\text{značajke}) = (P(\text{klasa}) * P(\text{značajke}|\text{klasa})) / P(\text{značajke})$
- Neovisnost značajki: Naivna pretpostavka algoritma je da su sve značajke neovisne jedna o drugoj, što rijetko vrijedi u stvarnom svijetu, ali pojednostavljuje računanje i često daje dobre rezultate.
- Klasifikacija: Naivni Bayes se često koristi za klasifikaciju, gdje se primjer dodjeljuje klasi s najvećom posterior vjerojatnosti. Na primjer, u analizi sentimenta, možete odrediti je li recenzija pozitivna ili negativna.
- Varijacije: Postoje različite varijacije Naivnog Bayesa, uključujući Bernoullijev Naivni Bayes (za binarne značajke), Multinomialni Naivni Bayes (za diskretne značajke, često korišten u analizi teksta) i Gaussov Naivni Bayes (za kontinuirane značajke).
- Primjena: Osim analize teksta, Naivni Bayes se koristi u različitim područjima kao što su klasifikacija e-mailova, dijagnostika bolesti, filtriranje informacija i više.

Unatoč svojoj "naivnoj" pretpostavci o neovisnosti značajki, Naivni Bayes često daje solidne rezultate i brzo se trenira. Međutim, nije uvijek najbolji izbor za složene zadatke gdje značajke nisu stvarno neovisne, ali može biti dobar izbor za početak analize podataka ili kao dio ansambla modela.

Naivni Bayes je probabilistički klasifikacijski algoritam koji se temelji na Bayesovom teoremu i na pretpostavci da su sve značajke neovisne, iako to često nije istinito u stvarnim podacima. Algoritam se koristi za klasifikaciju primjera u različite klase ili kategorije na temelju vjerojatnosti. Tijekom treninga, algoritam prvo izračunava apriornu vjerojatnost svake klase, što predstavlja omjer broja primjera te klase u trening skupu i ukupnog broja primjera. Također, izračunava uvjetne vjerojatnosti za svaku značajku za svaku klasu, što je vjerojatnost da će se ta značajka pojaviti za dani razred.

Zatim, tijekom testiranja, za svaki primjer iz testnog skupa, algoritam koristi Bayesovu formulu kako bi izračunao posterior vjerojatnosti za svaku klasu. To uključuje množenje apriorne vjerojatnosti klase s produktom svih uvjetnih vjerojatnosti značajki za tu klasu. Primjer se dodjeljuje klasi s najvećom posterior vjerojatnošću kao predikcija.

Naivni Bayes je često korišten za klasifikaciju teksta, kao što je analiza sentimenta (određivanje pozitivnog, negativnog ili neutralnog sentimenta teksta) i filtriranje spam poruka. Unatoč svojoj "naivnoj" pretpostavci o neovisnosti značajki, algoritam često daje zadovoljavajuće rezultate, posebno kada postoji velik broj značajki. Ipak, može biti neprecizan u slučajevima gdje su značajke međusobno zavisne, pa je važno razumjeti njegove ograničenja i provoditi adekvatnu pripremu podataka kako bi pružio pouzdane rezultate. (Bishop, C. M. 2006.)

```

Funkcija Treniraj_Naivni_Bayes(X_trening, Y_trening):
# Inicijalizacija varijabli za pohranu brojača i vjerojatnosti
Inicijaliziraj brojače i vjerojatnosti za svaku klasu i značajku

Za svaku klasu c:
# Izračunaj apriornu vjerojatnost klase P(c)
Izračunaj P(c) kao broj primjera klase c podijeljen s ukupnim brojem primjera u skupu za treniranje

Za svaku značajku X_i:
# Izračunaj uvjetnu vjerojatnost P(X_i|c)
Izračunaj P(X_i|c) kao broj primjera klase c s vrijednošću X_i podijeljen s ukupnim brojem primjera klase c

Vrati brojače i vjerojatnosti

Funkcija Predviđanje_Naivni_Bayes(X_test, brojači, vjerojatnosti):
Za svaki primjer u X_test:
Za svaku klasu c:
# Inicijaliziraj posterior vjerojatnost za klasu c
Inicijaliziraj posterior vjerojatnost P(c|X_test) na 1

Za svaku značajku X_i:
# Izračunaj posterior vjerojatnost za klasu c koristeći Bayesovu formulu
Pomnoži P(c|X_test) s P(X_i|c)

# Pomnoži posterior vjerojatnost za klasu c s apriornom vjerojatnosti P(c)
Pomnoži P(c|X_test) s P(c)

# Dodjeli primjer klasi s najvećom posterior vjerojatnošću
Dodijeli primjer klasi s najvećom posterior vjerojatnošću

Vrati predikcije za sve primjere u X_test

```

Slika 12 Naivni Bayes Pseudokod (Bishop, C. M. 2006.)

5. Primjeri upotrebe algoritama strojnog učenja u BI

5.1. Linearna regresija u BI

Linearna regresija je jedan od najosnovnijih i najčešće korištenih algoritama u statistici i strojnom učenju. Ulazni podaci, poznati kao značajke, obično su kontinuirane varijable (kao što je površina nekretnine ili godina izgradnje). Prije treniranja modela, ovi se podaci često standardiziraju kako bi imali slične skale. Kroz ovu metodu, nastojimo modelirati i razumjeti odnos između dvije ili više varijabli. Primjer primjene ovog pristupa može se vidjeti u predviđanju prodajnih cijena nekretnina. Zamislite da želimo predvidjeti prodajnu cijenu nekretnine temeljem njenih karakteristika: površine nekretnine, broja soba, godine izgradnje i lokacije. Model linearnog regresije pokušat će pronaći linearni odnos između ovih ulaznih podataka (značajki) i izlazne vrijednosti, koja je kontinuirane prirode, u ovom slučaju, prodajna cijena. Osnovna ideja je da se, s promjenom jedne značajke, predviđena prodajna cijena mijenja u određenom pravcu i za određeni iznos. Prednosti linearnog modela uključuju njegovu jednostavnost i interpretativnost. Model je lako razumljiv i implementirati ga. Međutim, glavni nedostatak je što pretpostavlja linearni odnos između ulaznih i izlaznih varijabli, što u stvarnom svijetu nije uvijek slučaj. Također, linearna regresija može biti osjetljiva na "outliere" ili ekstremne vrijednosti.

5.2. Logistička regresija u BI

Logistička regresija, iako se zove 'regresija', zapravo je klasifikacijski algoritam koji se koristi za predviđanje kategoričkih ishoda. Ulazni podaci za ovaj algoritam mogu biti kombinacija kontinuiranih i kategoričkih varijabli, poput demografskih informacija, povijesti ponašanja korisnika ili medicinskih mjerenja. Kako bi se postigla uniformnost i poboljšala točnost modela, ulazni podaci često se kodiraju (npr. "one-hot encoding" za kategoričke varijable) ili normaliziraju. Umjesto kontinuirane izlazne vrijednosti, kao što je to slučaj s linearnom regresijom, logistička regresija daje vjerojatnost pripadnosti određenoj klasi. Primjerice, želimo li modelirati vjerojatnost da će korisnik kliknuti na određeni oglas na temelju njegove starosti, prethodnog ponašanja na webu i demografskih podataka, logistička regresija je idealan izbor. Algoritam analizira navedene ulazne podatke i daje vjerojatnost, koja je kontinuirana vrijednost između 0 i 1, koja predstavlja vjerojatnost pripadnosti nekoj kategoriji. Jedna od ključnih prednosti ovog pristupa je što izlaz, kao vjerojatnost, nudi kvantitativnu mjeru sigurnosti predviđanja. Međutim, model također nosi neke izazove, kao što je pretpostavka o linearnom odnosu između logIT transformacije ciljne varijable i ulaznih značajki, kao i potencijalna multikolinearnost među prediktorima.

5.3. Stablo odlučivanja u BI

Stabla odlučivanja su grafički algoritmi koji segmentiraju skup podataka na osnovu različitih kriterija odluke. Ulazni podaci mogu biti i kategorički i kontinuirani, ali često se moraju transformirati ili kodirati kako bi se pravilno koristili. Primjer njegove upotrebe je kreditni scoring. Uzimajući u obzir različite faktore kao što su prihodi, starost, zaposlenje i povijest kredita, stablo odlučivanja kategorizira pojedince temeljem rizika kredita, rezultirajući ocjenom kreditnog rizika koja može biti kontinuirana ili kategorička vrijednost. Prednosti ovog pristupa uključuju intuitivan grafički prikaz, sposobnost rada s kategoričkim i numeričkim značajkama, te relativnu jednostavnost za razumijevanje. No, nedostaci su da može biti sklon neprilagođavanju i da može postati nestabilan ako se podaci samo malo promijene.

5.4. K-srednjih vrijednosti (K-Means Clustering) u BI

K-Means je metoda grupiranja koja pokušava podijeliti skup podataka u 'k' broj skupina na temelju sličnosti. Ulazni podaci su tipično kontinuirane varijable koje opisuju karakteristike primjera u skupu podataka. Ovi se podaci često normaliziraju kako bi imali jednake skale prije korištenja algoritma. Na primjer, za segmentaciju klijenata, algoritam analizira demografske podatke klijenata, povijest kupnje i ponašanje na web stranicama kako bi pronašao skupine klijenata s sličnim obilježjima. Izlaz ovog algoritma su labela koje predstavljaju kategorije ili klaster. Prednosti ovog pristupa uključuju njegovu efikasnost za velike skupove podataka i relativnu jednostavnost za implementaciju. Međutim, nedostatak je što unaprijed moramo odabrati broj klastera (k), a rezultati mogu varirati ovisno o početnoj randomizaciji.

5.5. Slučajne šume (Random Forests) u BI

Slučajne šume su ansambl tehnika koje kombiniraju višestruka stabla odlučivanja kako bi dobile stabilniji i točniji model. Ulazni podaci mogu biti kombinacija kontinuiranih i kategoričkih varijabli, slično kao u stablu odlučivanja. Model radi tako da svako stablo "glasuje" za ishod, a krajnja odluka se temelji na većinskom glasu. Na primjer, ako želimo predvidjeti hoće li klijent napustiti banku, slučajne šume uzimaju u obzir različite varijable poput povijesti transakcija, demografije i ponašanja klijenta. Izlaz ovog modela obično je kategorička varijabla (npr. da/ne). Prednosti slučajnih šuma uključuju smanjenu vjerojatnost neprilagođavanja (u usporedbi s pojedinačnim stablom odlučivanja) i robustnost u prisustvu šuma u podacima. Nedostaci mogu uključivati veću složenost modela i potencijalno duže vrijeme treniranja.

5.6. K-najbliži susjed (K-Nearest Neighbors, KNN) u BI

KNN je algoritam temeljen na instanci koji klasificira objekte na temelju sličnosti s njegovim K najbližim susjedima u skupu podataka. Ulazni podaci su tipično kontinuirane varijable, koje opisuju karakteristike primjera u skupu podataka, ali mogu biti i kategorički ako se pravilno kodiraju. Na primjer, da bismo klasificirali vrstu voća na temelju boje, težine i oblika, KNN će usporediti nepoznato voće s voćem u skupu podataka i dodijeliti klasu temeljem sličnosti s najbližim primjerima. Izlaz ovog algoritma je kategorička labela. Prednosti KNN-a uključuju jednostavnost i fleksibilnost. Nedostaci su mu osjetljivost na irelevantne

značajke i potreba za pohranom cijelog skupa podataka (što može biti neefikasno za vrlo velike skupove podataka).

5.7. Naivni Bayes u BI

Naivni Bayes je statistički klasifikacijski algoritam temeljen na Bayesovom teoremu s naivnom pretpostavkom o neovisnosti među značajkama. Ulazni podaci mogu biti kategorički ili kontinuirani, ali često se moraju transformirati ili kodirati kako bi se pravilno koristili u modelu. Primjerice, u analizi sentimenta, algoritam se koristi za klasifikaciju teksta kao pozitivnog, negativnog ili neutralnog temeljem frekvencije pojavljivanja riječi. Izlaz ovog algoritma je kategorička labela. Prednosti Naivnog Bayesa uključuju efikasnost, jednostavnost i performanse koje su često usporedive s naprednijim tehnikama. Glavni nedostatak je naivna pretpostavka o neovisnosti značajki, što može biti netočno u stvarnom svijetu.

6. Integracija strojnog učenja u postojeće BI alate

U današnjem digitalnom dobu, sposobnost organizacija da prikupljaju, analiziraju i iskoriste svoje podatke igra ključnu ulogu u konkurenciji na tržištu. Integracija strojnog učenja (ML) u postojeće poslovno-informatičke (BI) alate predstavlja napredni korak prema ostvarivanju punog potencijala podataka. U ovoj sekciji, dublje ćemo istražiti ovu integraciju i istaknuti ključne koristi koje donosi, zajedno s primjerima BI alata koji su otvoreni za takvu integraciju

6.1. Potrebni koraci za implementaciju ML algoritama u BI alate

Integracija algoritama strojnog učenja u poslovno-informatičke (BI) alate predstavlja složen proces koji zahtijeva pažljivo planiranje i tehničku ekspertizu. Ovaj proces obuhvaća nekoliko ključnih koraka kako bi se osiguralo uspješno uvođenje strojnog učenja u okvir poslovne inteligencije.

Prvi korak u integraciji je povezivanje BI alata s izvorima podataka. Ovi izvori podataka mogu uključivati internu pohranu podataka, različite baze podataka, vanjske datoteke ili čak cloud-based resurse. Ova povezanost omogućava BI alatu pristup podacima potrebnim za analizu i treniranje modela strojnog učenja.

Nakon uspostavljanja veze s izvorima podataka, podaci se moraju pažljivo pripremiti za analizu i modeliranje. Ovaj korak uključuje čišćenje podataka, rješavanje nedostajućih vrijednosti, normalizaciju i druge transformacije kako bi podaci bili prikladni za uporabu u algoritmima strojnog učenja. Kvaliteta podataka igra ključnu ulogu u ovom koraku, budući da nedostaci ili nekonzistentnosti u podacima mogu dovesti do nepouzdanih rezultata i odluka.

Nakon pripreme podataka, korisnik ili analitičar treba odabrati odgovarajući algoritam strojnog učenja za svoj konkretan problem. Ovisno o vrsti analize koja se provodi, mogu se koristiti različiti algoritmi. Na primjer, za klasifikaciju podataka, mogu se primijeniti algoritmi kao što su Random Forest, Naive Bayes ili Support Vector Machines, dok se za regresiju mogu koristiti algoritmi poput linearnih regresija ili neuronskih mreža. Odabir odgovarajućeg algoritma ključan je za postizanje uspješnih rezultata u analizi podataka.

Nakon odabira algoritma, model strojnog učenja mora biti treniran na pripremljenim podacima. Treniranje uključuje postupak učenja algoritma na povijesnim podacima kako bi model mogao naučiti i razumjeti obrasce u podacima. Kvaliteta treniranja modela ključna je za točnost i pouzdanost rezultata analize.

Integracija s BI alatom ostvaruje se putem sučelja ili API-ja, koji omogućavaju korisnicima da uvezu svoje trenirane modele u BI alat. Većina modernih BI alata podržava popularne programski jezike kao što su Python ili R, koji se često koriste za implementaciju algoritama strojnog učenja.

Nakon integracije, modeli strojnog učenja mogu se koristiti za generiranje naprednih izvještaja i vizualizacija. Na primjer, model regresije može se koristiti za predviđanje budućih poslovnih rezultata, dok model klasifikacije može kategorizirati podatke i stvarati vizualizacije temeljene na tim rezultatima. Ova integracija omogućava organizacijama dublje razumijevanje svojih podataka i donošenje informiranih poslovnih odluka.

6.2. Microsoft Power BI za integraciju Strojnog Učenja









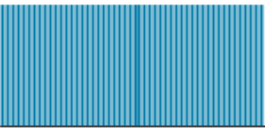
Microsoft Power BI, koji je uveo Microsoft 2014. godine kao dio svoje obitelji poslovnih proizvoda, postao je snaga na tržištu zahvaljujući svojoj integraciji s drugim Microsoftovim proizvodima i dubokoj cloud funkcionalnosti. Osim toga, Power BI donosi snažne mogućnosti za povezivanje s različitim izvorima podataka, kreiranje bogatih interaktivnih izvještaja i distribuciju tih izvještaja kroz organizacije, čineći analizu podataka dostupnom i razumljivom na svim razinama poslovanja.

- Power BI Insights automatski pretražuje dataset tražeći zanimljive trendove ili anomalije. Koristi algoritme strojnog učenja da bi detektirala i istaknula bitne informacije koje korisnik možda propusti.
- Power BI Key Influencers Visual služi za to kada želite znati koji faktori najviše utječu na određeni ishod, ova vizualizacija koristi logističku regresiju. Recimo da želite znati koji faktori utječu na smanjenje prodaje, vizualizacija će vam pokazati rangirane faktore.
- Azure Machine Learning integracija je platforma za izradu, trening i implementaciju modela strojnog učenja. Kada je model gotov, možete ga uvesti u Power BI i koristiti ga za predviđanja izravno unutar vaših izvještaja.

Neki primjer korištenja ovog alata te njegovih funkcionalnosti koje koriste ML bi bio u sklopu trgovinske tvrtke koja želi poboljšati korisničko iskustvo na svojoj web stranici, tim je odlučio analizirati feedback korisnika pomoću Power BI-ja i Azure ML-a. Nakon uvoza skupa podataka koji sadrži tisuće korisničkih komentara i ocjena u Power BI, tim je kreirao model analize sentimenta koristeći Azure Machine Learning. Pomoću ovog modela, analizirali su sentiment svakog komentara, kategorizirajući ga kao pozitivan, negativan ili neutralan. Uz pomoć Power BI-ja, vizualizirali su rezultate, prikazujući distribuciju sentimentata kroz različite kategorije proizvoda, te identificirali proizvode koji su dobili najviše negativnih ocjena. Osim toga, koristili su Power BI Insights kako bi automatski detektirali ključne teme ili riječi koje se često pojavljuju u negativnim recenzijama, što im je omogućilo da brzo identificiraju problematične značajke web stranice i poduzmu korake za njihovo rješavanje. U ovome primjeru strojno učenje pomaže donosiocima odluka da dobiju dublje uvide iz svojih podataka i donose informirane odluke na temelju tih analiza. Integracija BI alata s strojnim učenjem kombinira vizualnu analizu s naprednim analitičkim sposobnostima, čime se povećava vrijednost i učinkovitost BI rješenja.

6.3. Prikaz Power BI insights i Power BI Key Influencers

Prvi korak kod korištenja ML alata je naći dobar database u ovome slučaju ja sam odabrao Amazon reviews dataset. Prvi collum je "Id" (slika 13) on je jedinstveni identifikator za svaku recenziju, dok "Record ID" također predstavlja identifikator, ali može imati različit format ili svrhu. "ProductId" označava identifikator proizvoda koji je predmet recenzije, omogućujući povezivanje recenzija s odgovarajućim proizvodima. "UserId" identificira korisnika koji je napisao recenziju, dok "ProfileName" sadrži ime korisničkog profila koji je recenziju ostavio. "HelpfulnessNumerator" i "HelpfulnessDenominator" predstavljaju brojčane vrijednosti koje se koriste za izračun postotka korisnih ocjena za recenziju, ukazujući na korisnost recenzije za druge korisnike. "Score" sadrži ocjenu koju je korisnik dao proizvodu u recenziji, dok "Time" predstavlja vremenski žig koji označava kada je recenzija napisana. "Summary" je sažetak recenzije, pružajući brzi pregled njenih glavnih točaka, dok "Text" sadrži puni tekst recenzije. Svi ovi stupci su ključni resursi za analizu i modeliranje podataka o recenzijama proizvoda, omogućujući dublje razumijevanje korisničkog mišljenja, analizu sentimenta i izradu različitih modela strojnog učenja za različite svrhe, kao što su predviđanje ocjena proizvoda i analiza korisničkog ponašanja.

 Id 	 ProductId 	 UserId 	 ProfileName 
Record ID	Product ID	User ID who is reviewing	User Profile Name who is reviewing
 1 568k	74258 unique values	256059 unique values	218418 unique values
1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian
2	B00813GRG4	A1D87F6ZCVE5NK	d11 pa

Slika 13 Prikaz malog dijela tablice dataseta

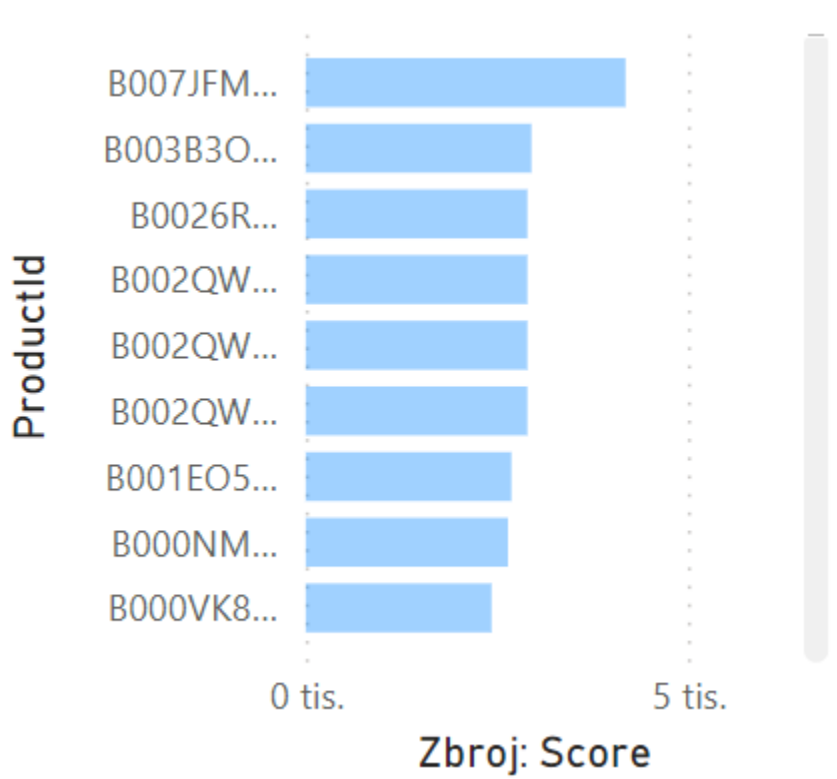
Sljedeći ključni korak u procesu je učitavanje podataka iz dataseta u alat Power BI kako bismo mogli koristiti informacije iz CSV datoteke. Ovaj proces se obavlja putem opcije "Get Data" unutar Power BI, pružajući nam širok raspon opcija za učitavanje podataka, uključujući izvore poput MySQL baze podataka, različitih hostinga i drugih. U našem slučaju, radimo s CSV datotekom. Nakon uspješnog učitavanja

podataka, ključno je stvoriti relevantne vizualizacije i uvide. Mi smo kreirali nekoliko ključnih uvida u našim analizama.

Prvo, izradili smo vizualizaciju koja prikazuje broj recenzija po proizvodu (slika 14), omogućujući nam bolje razumijevanje popularnosti i interesa za različite proizvode. Također smo stvorili tablicu koja detaljno prikazuje svaku recenziju, uključujući komentare, ocjene, korisnost recenzije i povezanost s određenim proizvodom (slika 15). Ovo nam pruža mogućnost dubljeg analiziranja pojedinačnih recenzija i njihovih karakteristika.

Osim toga, implementirali smo "heatmap" ili mapu topline, koja nam omogućuje vizualizaciju najčešćih ocjena korisnosti i njihovu povezanost s različitim vrstama recenzija (slika 16). Ovaj alat pomaže identificirati dominantne obrasce ocjena i razumjeti kakve vrste recenzija su najprisutnije u našem datasetu. Sve ove vizualizacije i uvidi doprinose boljem razumijevanju i interpretaciji podataka o recenzijama proizvoda unutar Power BI okoline, čime olakšavaju donošenje informiranih odluka i analizu ključnih aspekata dataset-a.

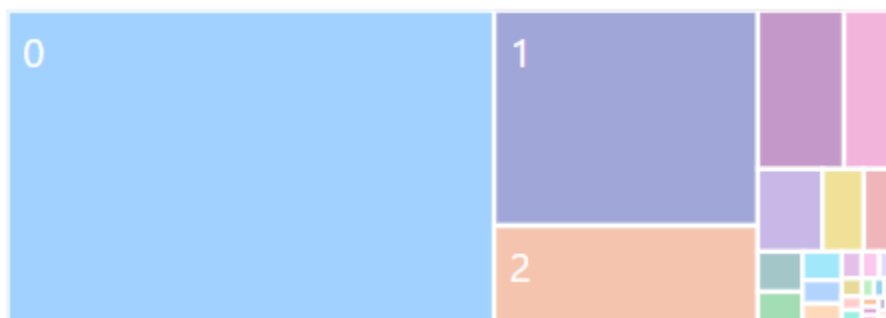
Također još smo implementirali Tortni ili "Pi" grafikon da možemo prikazati koliki je udio koje ocjene u našim recenzijama (slika 17).



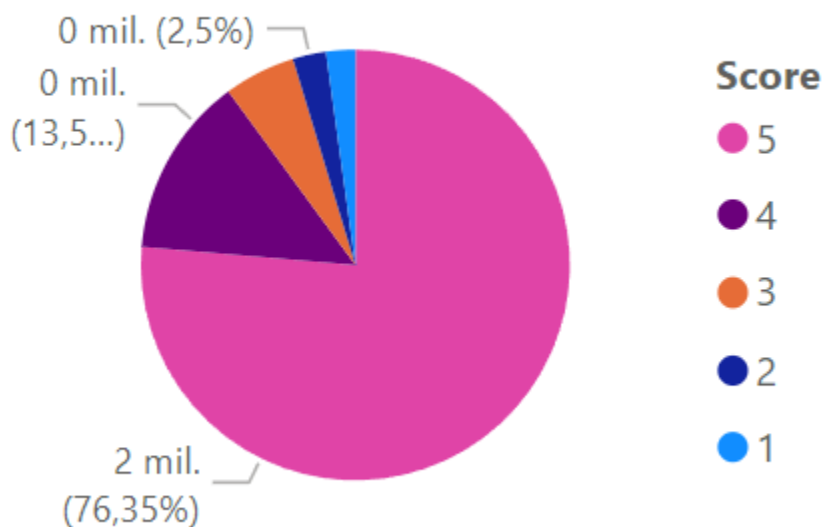
Slika 14 broj ocjena po proizvodu

Zbroj: Helpfulness	Numerator	ProductId	Zbroj: Score	Text
0	0006641040		4	A charming, rhyming book that describes the ci with rice, month-by-month. This sounds like the recess and sing over and over until they drive th childlike but is skillfully written.
1	0006641040		4	A very entertaining rhyming story--cleaver and However, the paperback is somewhat small and
3	0006641040		5	All of my children love this book. My first grade

Slika 15 Tablica za proizvod review

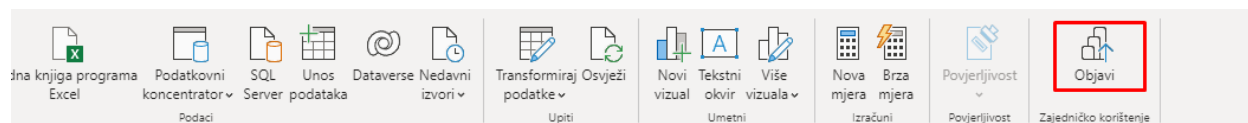


Slika 16 "heat map" najčešće ocjene korisnosti

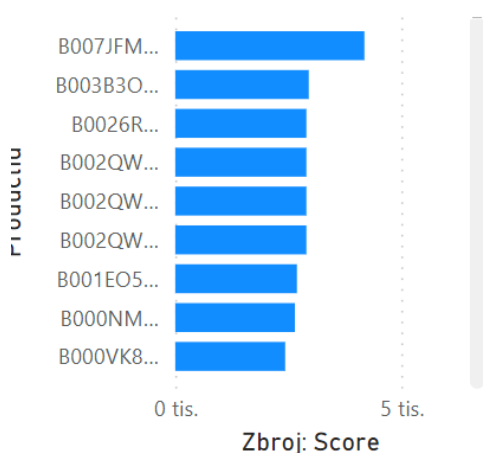


Slika 17 Tortni grafikon za udio cjenaa

Nakon izrade ciljanih vizualizacija, kako biste mogli iskoristiti funkcionalnost "Get Insights," prvi korak je osigurati objavljivanje datoteke (slika 18) u Power BI servisima. Objavljivanje datoteke omogućuje vam pristup dodatnim naprednim funkcionalnostima koje su dostupne unutar Power BI servisa, gdje možete proširiti analizu i izraditi dublje uvide temeljene na vašem izvornom datasetu. Objavljivanjem datoteke omogućujete daljnju suradnju, dijeljenje i pristup analizama i vizualizacijama unutar Power BI ekosustava, čime se pruža prilika za dublje razumijevanje podataka i donošenje informiranih poslovnih odluka.



broj: Score kategorije ProductId



Zbroj: Score kategorije HelpfulnessNumerator

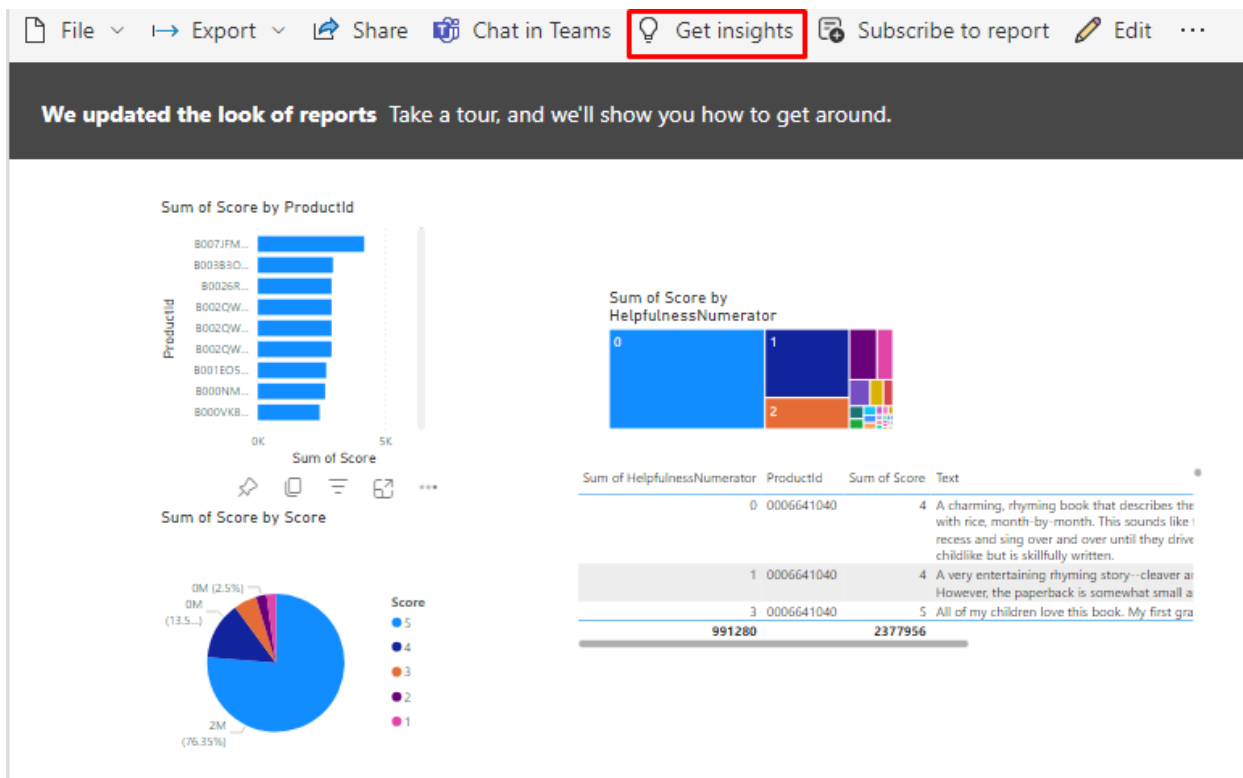


Zbroj: HelpfulnessNumerator ProductId
0 0006641040

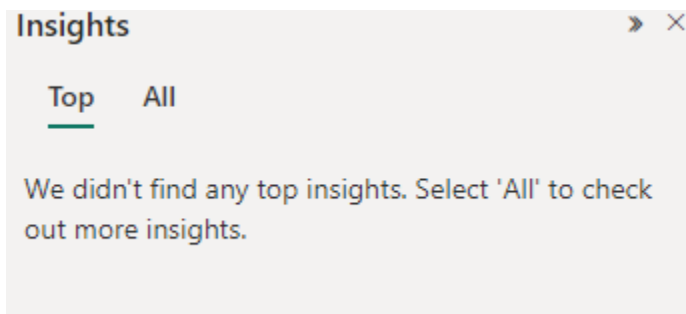
broj: Score kategorije Score

Slika 18 objava datoteke

Nakon što uspješno objavite datoteku, bit ćete upućeni na dobiveni link koji vas preusmjerava na Power BI servise kako biste koristili odabrane funkcionalnosti. U ovom koraku, bit će potrebno prijaviti se na svoj prethodno registrirani korisnički račun. Na vrhu ekrana bit će dostupna opcija "Get Insights" (slika 19). Po pokretanju ove opcije, slijedi faza u kojoj će algoritam analizirati vaše podatke. Ovo može potrajati neko vrijeme, budući da algoritam provodi dubinsku analizu vašeg dataset-a. Na kraju tog procesa, bit ćete obaviješteni o dobivenim saznanjima. Na slici 21 možemo vidjeti uspjeh algoritma u pronalasku uvida. Važno je napomenuti da postoji mogućnost da strojno učenje (ML) algoritam ne pronađe relevantne uvide (slika 20). Ovo se može dogoditi iz više razloga, uključujući veličinu dataset-a koja može biti prevelika za temeljitu analizu algoritma ili zbog ograničenja u dataset-u samom. U takvim slučajevima, algoritam možda neće uspjeti pronaći značajne uvide ili može generirati ograničene rezultate. Ovo naglašava važnost pripreme podataka i razmatranje njihove kvalitete prije nego što se pokrene proces izvlačenja uvida kako bi se osigurala što preciznija i korisnija analiza.

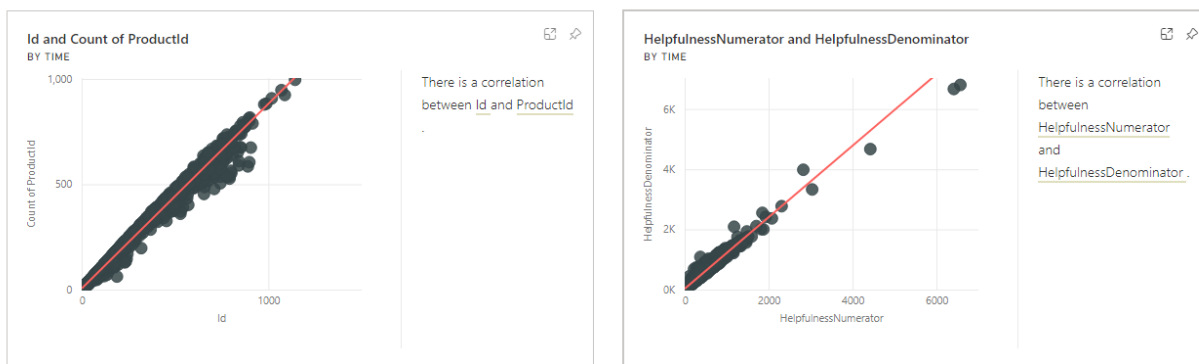


Slika 19 prikaz Power BI services



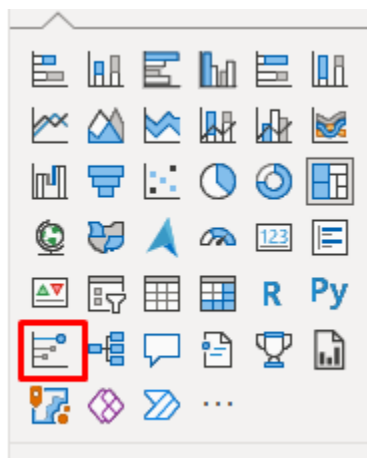
Slika 20 prikaz da nema uvida

A subset of your data was analyzed and the following insights were found. [Learn more](#)

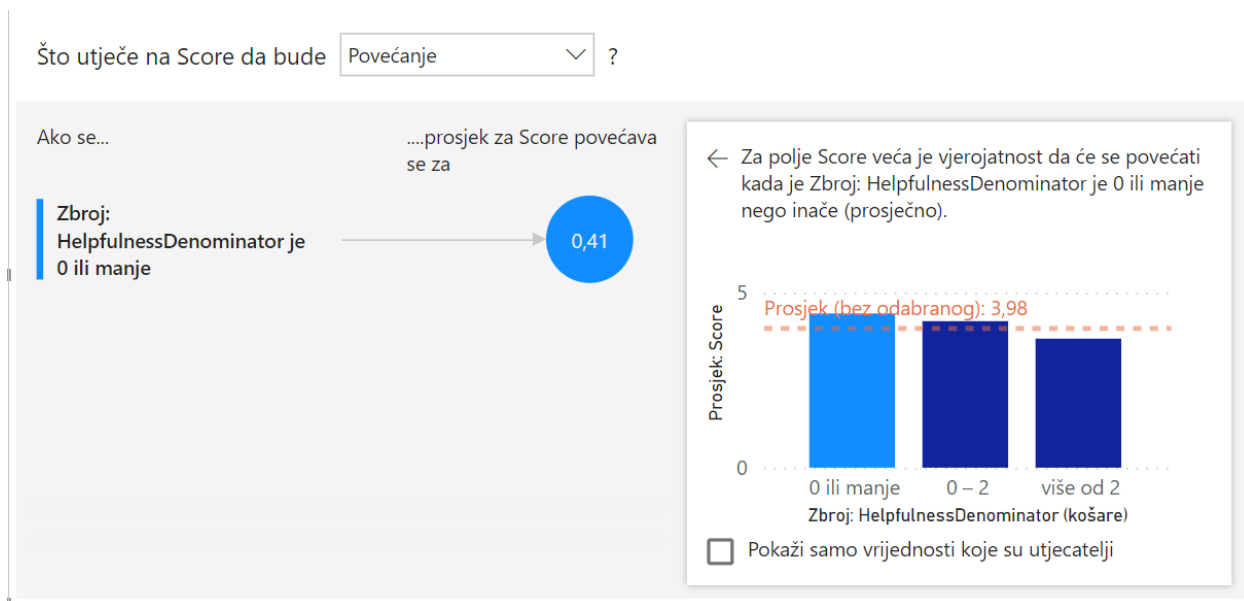


Slika 21 Prikaz uvida za Amazon Review dataset

U našoj analizi dalje prelazimo na implementaciju modula 'Ključni Utjecatelji' (Slika 22). Ovo je algoritam temeljen na tehnologiji strojnog učenja i umjetne inteligencije, koji pruža uvide temeljene na analizi dostavljenih podataka. Konkretni primjer njegove primjene prikazan je na Slici 23. Ovdje je cilj razumjeti koji faktori najviše pridonose rastu ocjene u recenzijama proizvoda. Iz rezultata je vidljivo da recenzije koje nisu označene kao "korisne" (Helpful) imaju tendenciju ostvarivanja viših prosječnih ocjena.



Slika 22 Ključni utjecatelji



Slika 23 prikaz Ključnih utjecatelja za Amazon Reviews

Alati kao što su Power BI 'Insight' i 'Key Influencers' nude izvanredne analitičke sposobnosti. Međutim, njihova stvarna snaga i prednost postaju očiti kada se primjenjuju na podacima većeg obima. U kontekstu manjih datasetova, njihova potreba možda nije uvijek primarna. Važno je napomenuti da se njihov puni potencijal često manifestira u obradi obimnih datasetova, koji su, zbog svoje osjetljive prirode ili strateškog značaja, često rezervirani za internu upotrebu i nisu široko dostupni javnosti.

6.4. Prednosti Integracije Strojnog Učenja u BI alate

Integracija strojnog učenja u poslovno-informatičke (BI) alate pruža organizacijama niz značajnih prednosti. Prvo, ML algoritmi unaprjeđuju analitiku i preciznost analize podataka. Ovo omogućava organizacijama da istraže dublje slojeve informacija, prepoznaju kompleksne obrasce u podacima i dobiju bolje razumijevanje poslovnih trendova te korisničkih preferencija. Nadalje, automatizirani procesi strojnog učenja ubrzavaju analitičke zadatke, omogućujući organizacijama da brže reagiraju na dinamične promjene u poslovnom okruženju. Osim toga, integracija strojnog učenja omogućava personalizirane preporuke za korisnike temeljene na njihovim navikama i ponašanju, poboljšavajući korisničko iskustvo i poticanje veće angažiranosti. Automatizirani procesi strojnog učenja također doprinose smanjenju potencijalnih ljudskih grešaka u analitičkim zadacima, što povećava točnost i pouzdanost rezultata. Konačno, ML algoritmi su korisni za proaktivno upravljanje rizicima jer mogu identificirati potencijalne rizike i prijetnje unaprijed, što organizacijama omogućava bolje planiranje i strategije za minimiziranje rizika i očuvanje stabilnosti poslovanja. Ova integracija stvara temelj za napredne analitičke sposobnosti koje potiču donošenje pametnih i proaktivnih poslovnih odluka temeljenih na podacima.

6.5. Nedostaci Integracije Strojnog Učenja u BI alate

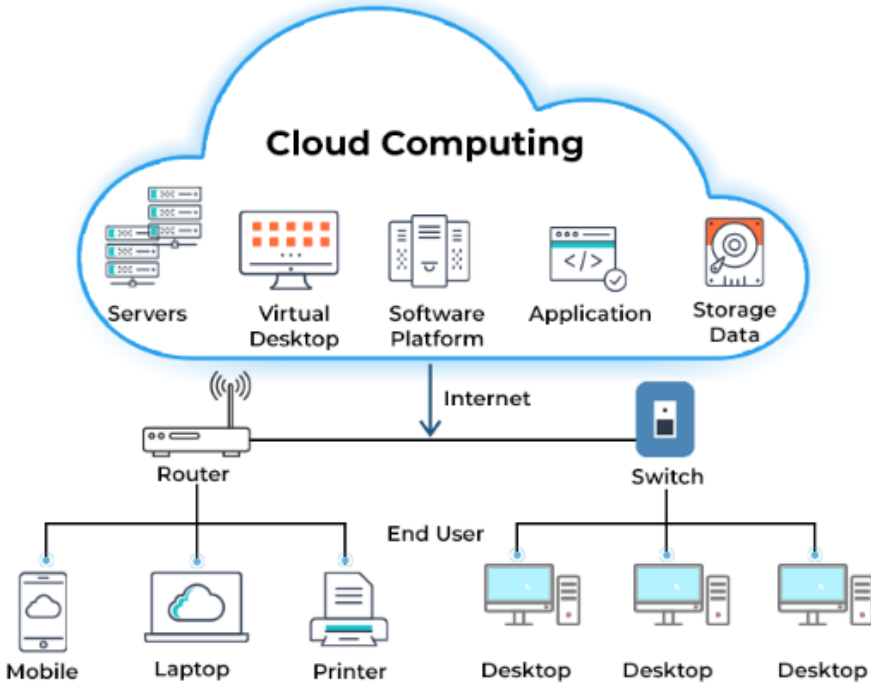
Integracija strojnog učenja u BI alate, iako donosi brojne prednosti, nosi sa sobom određene nedostatke. Prvi izazov predstavlja složenost implementacije. Ova integracija zahtijeva od organizacija da razumiju i primijene složene tehnološke komponente, što može predstavljati izazov za manje tehnički orijentirane timove. Također, potrebno je naglasiti da uspješna integracija strojnog učenja ovisi o kvaliteti podataka. ML algoritmi su izrazito osjetljivi na nedostatke ili nekonzistentnosti u podacima, što može rezultirati nepouzdanim rezultatima i odlukama. Osim toga, obuka osoblja unutar organizacije za rad s ML algoritmima i razumijevanje njihove primjene u poslovnom kontekstu zahtijeva dodatne napore i resurse. Naposljetku, integracija strojnog učenja često izaziva pitanja o etičnosti i privatnosti podataka, te organizacije moraju aktivno rješavati te izazove kako bi osigurale odgovorno korištenje tehnologije.

6.6. Izazovi Integracije Strojnog Učenja u BI alate

Izazovi integracije strojnog učenja u BI alate obuhvaćaju niz aspekata. Tehnički izazovi su među najznačajnijima, jer povezivanje različitih tehnoloških tehnologija može biti izrazito kompleksno. Učinkovita povezanost između BI alata i strojnog učenja zahtijeva stručno znanje i iskustvo u implementaciji ovih tehnologija. Nadalje, kvaliteta podataka ostaje ključnim izazovom, budući da nedostaci u podacima mogu dovesti do nepouzdanih rezultata i odluka. Razumijevanje složenih ML algoritama također može predstavljati prepreku, jer osoblje unutar organizacije može imati izazova u njihovom razumijevanju i praktičnoj primjeni. Pitanja etike i privatnosti također su neizbježna, a organizacije moraju provoditi pažljivo upravljanje i poštivati zakonske regulative. Konačno, prilagodba kulturi organizacije i radnim procesima može biti izazovna, jer uvođenje strojnog učenja može zahtijevati promjene i prilagodbe koje treba pažljivo planirati i implementirati.

Sada kad znamo sve prednosti, nedostatke i moguće izazove glavno pitanje je kako bi to trebali implementirati u neko poslovno okruženje. Prvi korak je pravilno planiranje i identifikacija specifičnih poslovnih problema koji bi se mogli rješavati primjenom strojnog učenja. Fokusirajte se na konkretne izazove i ciljeve kako biste osigurali usmjerenost i svrhovitost implementacije. Nadalje, organizacije mogu iskoristiti prednost cloud-based platformi (slika 6) i alata za strojno učenje, kao što su Amazon Web Services (AWS), Google Cloud, Microsoft Azure, i IBM Watson. Ovi pružatelji usluga nude gotove alate i resurse za stvaranje, treniranje i implementaciju ML modela bez potrebe za dubokim tehničkim znanjem. Osim toga, suradnja s vanjskim stručnjacima i konzultantima za strojno učenje može biti koristan korak. Ovi stručnjaci mogu pružiti potrebno znanje i iskustvo za brzu i učinkovitu integraciju strojnog učenja u postojeće poslovne procese. Dodatno, organizacije bi trebale ulagati u obuku svojeg osoblja kako bi se osiguralo razumijevanje i kompetencije u radu s ML algoritmima i alatima. Ovo može uključivati organiziranje radionica, online tečajeva i edukaciju koja je prilagođena specifičnim potrebama tima. Najvažnije je razumjeti da implementacija strojnog učenja ne mora biti izuzetno kompleksan proces. Koristeći cloud-based resurse, suradnju s eksternim stručnjacima i fokus na konkretnim poslovnim izazovima, organizacije mogu postići uspješnu integraciju strojnog učenja i ostvariti prednosti koje donosi ova tehnologija u poslovnom okruženju.

CLOUD COMPUTING ARCHITECTURE



Slika 24 Cloud based platforme

7. Etika i sigurnost korištenja strojnog učenja u BI

U eri sveprisutnih podataka i tehnološkog napretka, razvoj i primjena strojnog učenja u okviru poslovne inteligencije donose ne samo brojne prednosti, već i važna pitanja o etici i sigurnosti. U ovom poglavlju, dublje ćemo istražiti ključne etičke izazove i sigurnosne aspekte koji proizlaze iz korištenja strojnog učenja u BI kontekstu. Razumijevanje ovih pitanja ključno je za organizacije koje žele maksimizirati koristi strojnog učenja dok istovremeno očuvaju etičke norme i zaštitu podataka.

7.1. Etički Aspekti

Korištenje strojnog učenja u BI može izazvati niz etičkih dilema. Primjerice, pitanja privatnosti podataka postaju još važnija jer se strojno učenje temelji na analizi velike količine informacija. Organizacije moraju pažljivo razmotriti kako prikupljaju, pohranjuju i koriste korisničke podatke, posebno ako se koriste za personalizirane preporuke ili analize. Također, transparentnost je ključna - korisnici i dionici trebaju razumjeti kako se koriste njihovi podaci i kako algoritmi donose odluke. Diskriminacija i pristranost u modelima strojnog učenja također su važne teme. Algoritmi mogu reflektirati implicitne predrasude prisutne u podacima, što može rezultirati nepravednim ili diskriminirajućim rezultatima. Stoga je važno prilagoditi algoritme kako bi se minimizirali takvi efekti. (Bracanović, T. 2019)

7.2. Sigurnosni Izazovi

U kontekstu sigurnosti, integracija strojnog učenja u BI donosi nove rizike. Pogreške u algoritmima mogu rezultirati pogrešnim odlukama i potencijalno ozbiljnim poslovnim posljedicama. Osim toga, modeli strojnog učenja postaju meta potencijalnih napada. Odabrani i trenirani algoritmi mogu biti osjetljivi na manipulaciju podacima i pokušaje varanja. Zaštita ovih modela i osiguranje njihove sigurnosti postaju ključni zadaci za organizacije. Također, organizacije moraju razmisliti o etičkom i zakonskom aspektu korištenja strojnog učenja u sigurnosne svrhe, posebno u područjima kao što su praćenje i nadzor zaposlenika. (Bracanović, T. 2019)

7.3. Etički Okviri i Sigurnosne Strategije

Da bi se suočile s rastućim etičkim i sigurnosnim izazovima koji proizlaze iz upotrebe strojnog učenja u okviru poslovne inteligencije (BI), organizacije moraju posvetiti pažnju razvoju sveobuhvatnih etičkih okvira i sigurnosnih strategija. Ovo nije samo odgovornost tehničkih timova, već i ključna inicijativa na višem nivou unutar organizacije.

- **Etički Okviri:** Organizacije trebaju razmisliti o uspostavi etičkih okvira koji jasno definiraju principe i smjernice za rukovanje podacima i primjenu strojnog učenja. Na primjer, kada organizacija prikuplja i koristi osobne podatke korisnika za personalizirane preporuke, etički okvir trebao bi osigurati da se ovi podaci tretiraju s najvišim standardima privatnosti i sigurnosti. To može uključivati stroge politike pristupa, anonimizaciju podataka i jasnu komunikaciju s korisnicima o svrsi i postupcima obrade njihovih podataka.
- **Transparentnost Algoritama:** Transparentnost je ključna komponenta etičke primjene strojnog učenja. Organizacije bi trebale osigurati da algoritmi korišteni u BI procesima budu transparentni i interpretabilni. To znači da bi korisnici i relevantne strane trebali moći razumjeti kako algoritmi donose odluke i zašto dolaze do određenih rezultata. Na primjer, u bankarskom sektoru, transparentnost algoritama koji odobravaju ili odbijaju zajmove može pomoći u sprječavanju nepravednih ili diskriminirajućih odluka.
- **Redovita Revizija Modela:** S obzirom na dinamičku prirodu poslovnog okruženja, organizacije trebaju implementirati sustav redovite revizije modela strojnog učenja. To uključuje praćenje performansi modela u stvarnom vremenu i identifikaciju potencijalnih problema. Na primjer, u kontekstu e-trgovine, model koji pruža preporuke proizvoda trebao bi se kontinuirano nadzirati kako bi se otkrile promjene u ponašanju korisnika ili nove tendencije na tržištu. Ako se otkriju problemi, modeli se mogu prilagoditi kako bi se očuvala njihova točnost i etička prihvatljivost.
- **Obuka Zaposlenika:** Etički okviri i sigurnosne strategije neće biti učinkovite ako osoblje unutar organizacije nije svjesno njihove važnosti i primjenjuje ih u praksi. Organizacije bi trebale uložiti u obuku svojih zaposlenika o etičkim standardima i sigurnosnim postupcima povezanim sa strojnim učenjem. Primjerice, u financijskom sektoru, obuka osoblja o rizicima i sigurnosnim protokolima može pomoći u zaštiti podataka i očuvanju povjerenja klijenata.

Ove strategije i pristupi zajedno čine temelj za odgovorno i etičko korištenje strojnog učenja u BI. Kroz pažljivu implementaciju etičkih okvira, transparentnost algoritama, redovitu reviziju modela i obuku osoblja, organizacije mogu istovremeno ostvariti prednosti strojnog učenja i zaštititi se od etičkih i sigurnosnih rizika.

8. Budućnost BI-a uz strojno učenje

Napredak u umjetnoj inteligenciji (AI) donosi značajne i duboke inovacije u domenu poslovne inteligencije (BI), preoblikujući temeljne aspekte procesa donošenja poslovnih odluka. Ova integracija AI-a u BI okruženje ima potencijal promijeniti način na koji organizacije prikupljaju, analiziraju i interpretiraju svoje podatke radi postizanja konkurentske prednosti. (Russell, S. J., & Norvig, P. 2021.).

Jedna od ključnih prednosti AI-a u BI je sposobnost automatskog otkrivanja složenih i često skrivenih obrazaca u ogromnim bazama podataka. Ovo omogućava dublju analizu podataka nego što je to bilo moguće prije, što znači da organizacije mogu dublje razumjeti ponašanje svojih klijenata, tržišne trendove te identificirati prilike i rizike koje bi inače možda propustile.

Dalje, AI modeli su ključni za preciznu predikciju tržišnih trendova. Kombinirajući različite izvore podataka i algoritme strojnog učenja, organizacije mogu stvoriti modele koji su u stanju predvidjeti buduće događaje s iznimnom preciznošću. To pomaže organizacijama da bolje planiraju svoje aktivnosti i resurse te minimiziraju rizik. (Russell, S. J., & Norvig, P. 2021.).

Personalizacija je također ključna komponenta AI-a u BI. Kroz analizu ponašanja korisnika, AI može pružiti personalizirane preporuke i sadržaj. Na primjer, e-trgovine koriste AI kako bi svojim kupcima ponudile proizvode i usluge koji najbolje odgovaraju njihovim potrebama i preferencijama. To povećava angažman korisnika i potiče prodaju.

Nadalje, AI napredak omogućuje analizu sentimenta, što je ključno za razumijevanje kako korisnici i javnost percipiraju proizvode ili usluge. AI alati mogu automatski analizirati velike količine teksta i identificirati pozitivne ili negativne komentare te tako pomoći organizacijama da bolje razumiju svoj ugled na tržištu.

Što se tiče brzine i preciznosti, AI ubrzava proces analize podataka i omogućuje brže reagiranje na promjene na tržištu. To je ključno u brzopoteznom okruženju u kojem se mnoge odluke moraju donositi u stvarnom vremenu.

U konačnici, integracija AI-a u BI alate donosi brojne nove mogućnosti i potiče inovacije. Organizacije koje usvoje ove tehnologije bit će bolje opremljene za donošenje informiranih poslovnih odluka, što će ih postaviti u poziciju da ostvare konkurentske prednosti na dinamičkom tržištu.

9. Zaključak

U eri sveprisutnih podataka i tehnološkog napretka, primjena strojnog učenja u poslovnoj inteligenciji (BI) predstavlja ne samo izvanredne mogućnosti za dublju analizu i personalizaciju usluga, već i ključna pitanja o etici i sigurnosti. Strojno učenje omogućava organizacijama bolje razumijevanje korisničkih preferencija, ali istovremeno postavlja pitanja o privatnosti podataka i transparentnosti algoritama. Osim toga, sigurnosni izazovi postaju neizbježni, s potencijalnim pogreškama u algoritmima i prijetnjama od napada na modele strojnog učenja. Za suočavanje s tim izazovima, organizacije moraju implementirati etičke okvire, osigurati transparentnost algoritama, redovito revidirati modele strojnog učenja i educirati svoje osoblje. Sve to omogućuje ostvarivanje prednosti strojnog učenja u BI dok se istovremeno čuvaju etičke norme i sigurnost podataka. Budućnost BI-a sastoji se u integraciji umjetne inteligencije (AI), što će transformirati način donošenja poslovnih odluka i omogućiti organizacijama da budu bolje opremljene za brzo reagiranje na promjene na tržištu.

10. Sadržaj Slika

Slika 1 Krug poslovne inteligencije	3
Slika 2 Proces skupljanja, transformacije i skladištenja podatka	5
Slika 3 Proces Deep learning-	9
Slika 4 Vrste Algoritama strojnog učenja	11
Slika 5 Klasterziacija u ML	11
Slika 6 Pseudokod za Linearnu regresiju (Bishop, Christopher M. 2006).....	13
Slika 7 Logistička regresija pseudokod (Bishop, Christopher M. 2006)	15
Slika 8 Pseudokod za Stablo odlučivanja (Bishop, Christopher M. 2006)	16
Slika 9 Pseudokod za K-srednjih vrijednosti (Bishop, Christopher M. 2006).....	18
Slika 10 Pseudokod za slučajne šume (Bishop, Christopher M. 2006).....	19
Slika 11 Pseudokod K najbližih susjeda (Bishop, Christopher M. 2006)	20
Slika 12 Naivni Bayes Pseudokod (Bishop, Christopher M. 2006)	22
Slika 13 Prikaz malog dijela tablice dataseta	28
Slika 14 broj ocjena po proizvodu	29
Slika 15 Tablica za proizvod review	30
Slika 16 "heat map" najčešće ocjene korisnosti	30
Slika 17 Tortni grafikon za udio cjenaa	30
Slika 18 objava datoteke	31
Slika 19 prikaz Power BI services	32
Slika 20 prikaz da nema uvida	32
Slika 21 Prikaz uvida za Amazon Review dataset	33
Slika 22 Ključni utjecatelji.....	33
Slika 23 prikaz Ključnih utjecatelja za Amazon Reviews	34
Slika 24 Cloud based platforme	36

11. Reference

- Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems*, 13, Article 15. DOI: 10.17705/1CAIS.01315.
- Sherman, R. (2015). *Business Intelligence Guidebook: From Data Integration to Analytics*.
- Foote, K. D. (2021). *A Brief History of Machine Learning*.
- Flach, P. A. (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2015). *Deep Learning: An MIT Press book*.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*.
- Cickaji, I. (2020). Pregled business intelligence alata i primjena u poslovnim sustavima za potrebe kontrolinga.
- Bracanović, T. (2019). *Etički izazovi umjetne inteligencije i robotike*.
- Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*.