UPOTREBA TEHNOLOGIJE VELIKIH PODATAKA U SUVREMENOM BANKARSTVU

Pejić, Ivan

SVEUČILIŠTE U SPLITU

EKONOMSKI FAKULTET SPLIT

DIPLOMSKI RAD

UPOTREBA TEHNOLOGIJE VELIKIH PODATAKA U SUVREMENOM BANKARSTVU

Mentor:

Prof.dr.sc Marko Hell

Student:

Univ.bacc.oec Ivan Pejić

Split, kolovoz 2020.

Sadržaj

1. Uvod

1.1. Problem istraživanja

Živimo u vremenu gdje nam tehnologija strahovito brzo mijenja našu svakodnevnicu. Inovativna rješenja, u različitim područjima, omogućuju lakši pristup izazovima s kojima se neprestano susrećemo. Utjecaj razvoja tehnologije je vidljiv u izgledu privatnog života svakog pojedinca. Vidljiviji, čak i važniji u poslovnim procesima svakog poduzeća. Stupanj otvorenosti tehnologiji se razlikuje od poduzeća do poduzeća te od djelatnosti, do djelatnosti u kojoj se poduzeće nalazi.

Financijske institucije a posebice banke, kao najprisutniji predstavnici financijskog tržišta u Hrvatskoj, se susreću s izazovima promjene načina poslovanja, uzrokovani razvojem tehnologije. Osim interakcije s klijentima, u kojima se sve više primjenjuje digitalni način komunikacije, banke konstantno unaprjeđuju pozadinske poslovno procese raznim tehnološkim rješenjima.

Često čujemo kako je informacija – sve. [1]Uz pravovremenu i adekvatnu obradu, podatak postaje informacija, a informacija komparativna prednost određenog poslovnog subjekta. Banke imaju pristup ogromnim količinama podataka koji uz odgovarajuću obradu i primjenu postaju najvažniji resurs u suvremenom poslovanju banaka. Na tragu ovog zaključka, dolazimo do pojma 'Big Data', odnosno tehnologiji velikih podataka. Upravo je ova tehnologija problem istraživanja ovog rada. Točnije, načini primjene te (ne)svjesnost mogućnosti tehnologije među zaposlenicima samih banaka. S obzirom da u zadnje dvije godine je generirano više podataka nego u povijesti cijelog čovječanstva, a obrađeno je manje od 1%, jasno je koliki potencijal ima ova tehnologija.

U bankarstvu, već sada postoje razna područja primjene big data tehnologije. Velike kompanije sve više pozornosti pridaju ovoj temi. J.P. Morgan [2] osniva posebni odjel unutar kompanije, koji se bavi samo obradom velikih podataka. Navode kako je to odjel koji se bavi pravim svjetskim problemima.

[1] Informacija, *Hrvatska enciklopedija, mrežno izdanje*. Leksikografski zavod Miroslav Krleža(http://www.enciklopedija.hr/Natuknica.aspx?ID=27405)
[2] https://www.jpmorgan.com/global/technology/applied-AI-and-ML#key-initiatives

Danske Bank [3], banka s više od 5 milijuna klijenata, najveća je banka u Danskoj. Loš postotak otkrivanja prijevara su poboljšali uz korištenje alata big data tehnologije te tako poboljšali svoje poslovanje.

Bank of America, [4] je jedna od najvećih Američkih banaka. Baza korisnika im iznosi oko 70 milijuna klijenata, u 2008 je zabilježila ogroman pad korisničke baze i odljev klijenata. Ne znajući razlog tome, angažirali su tvrtku koja se bavi big data tehnikama. Analizirajući korisničke podatke iz različitih izvora došli su do rješenja.

J.P. Morgan, [5] također, koristi alate big date kako bi pratili zadovoljstvo svojih klijenta. Osim za praćenje zadovoljstva klijenata, ova institucija alate big data tehnologije koristi i za mogućnost virtualnog asistenta,[6] kako bi poboljšali komunikaciju s klijentima.

Upravljanje rizicima ili takozvani "risk managment" je jedno od najbitnijih područja poslovanja svake banke. Razvoj tehnologije te pojava alata big date je već utjecala na područje upravljanja rizicima banaka u sektoru stanovništva. Pojava alata big data tehnologije omogućila je da pri analizi pojedinog klijenta je moguće uzeti znatno veću količinu varijabli nego do sada. [7] Tradicionalni alati za credit scoring u bankama, obrađuju 300 pristupnih podataka koji su nepromjenjivi te ostaju u sustavu takvi duži niz godina.

Uz upotrebu big data tehnologiju ove brojke se znatno mijenjaju, tako postoji mogućnost korištenja 20.000 pristupnih podataka [8] koji su promjenjivi u realnom vremenu te se kreditni status klijenta mijenja iz minute u minutu. Također, pomoću alata big data tehnologije ne promatra se samo povijesne podatke klijenta već i predviđa buduće ponašanje. Ovakav način određivanja kreditnog scoringa klijenta, daje nam realniju, pravovremenu i fleksibilniju procjenu klijenta te kao takav poboljšava poslovanje suvremenih banaka. Upravo ovo je problem istraživanja ovog rada.

[3] https://data-flair.training/blogs/big-data-in-banking/
[4] https://data-flair.training/blogs/big-data-in-banking/
[5] https://data-flair.training/blogs/big-data-in-banking/
[6] https://www.jpmorgan.com/global/technology/applied-AI-and-ML#key-initiatives
[7] http://www.eurofinas.org/uploads/documents/Non-visible/Big%20data/Roundtable/Eurofinas_Kreditech.pdf
[8] https://www.hbs.edu/openforum/openforum.hbs.org/goto/challenge/understand-digital-transformation-of-business/kreditech-big-data-scoring-for-consumer-lending.html

1.2. Predmet istraživanja

Predmet istraživanja rada će biti ključni načini primjene Big Data tehnologije, gdje će se istraživati mogućnosti primjene u poslovanju suvremenog bankarstva. Istraživati će se na koji način i u kojim područjima banke pokušavaju odnosno primjenjuju Big Datu. Dodatno, će se obratiti pozornost utjecaja tih primjena na efikasnost poslovanja. Kroz jednostavnu simulaciju procesa obrade podataka alatima Big Data tehnologije, prikazati će se konkretni primjer u segmentiranju, profiliranju i klastreriranje klijenata koje omogućuje lakšu, točniju i bržu procjenu rizika promatranog klijenta u realnom vremenu. .

Drugi dio istraživanje se odnosi na svjesnost zaposlenika banaka o Big Data tehnologiji. S obzirom na važnost i aktualnost primjene u radu će se istražiti jesu li zaposlenici na relevantnim pozicijama unutar banaka upoznati s tehnologijom velikih podataka te u kolikoj mjeri.

1.3. Istraživačke hipoteze

Na temelju prethodno opisanog predmeta i problema istraživanja te analize različitih izvora podataka postavit ćemo hipoteze.

Nakon prikupljanja novih, analiziranja postojećih podataka te istraživanja hipoteza donijet će se zaključak o prihvaćanju ili ne prihvaćanju postavljenih hipoteza.

Hipoteza:

H 1: Big data tehnologiju moguće je koristiti u segmentaciji i profiliranju klijenata banke u svrhu upravljanja rizicima.

H 2: Zaposlenici na relevantnim pozicijama u promatranim bankama svjesni su prednosti Big Data tehnologije u procesima poslovanja s klijentima

1.4. Ciljevi Istraživanja

Cilj ovog istraživanja je analizirati dijelove ključnih procesa u poslovanju suvremenih banka, gdje je primjena Big Data tehnologije poboljšala efikasnost poslovanja banaka. Definiranjem i vrednovanjem utjecaja alata Big Date na već postojećim primjerima primjene, ukazati će na mogući razvoj poslovanja banaka u segmentu odnosa s klijentima.

U hipotezama koje su postavljene u ovom radu, definirana su aktualna i važna područja te načini primjeni Big Data tehnologije u bankarstvu. Istražiti će se načini obrade podataka koje banka posjeduje (in house) kao i eksternih podataka. Dodatno, njihov utjecaj na efikasnost poslovanja, odnosno da li obrada velike količine podataka, koristeći alate Big Data tehnologije može dovesti do točnijih, bržih i automatiziranijih procesa odlučivanja. Također, u kojim segmentima poslovanja banaka će ova tehnologija biti najlakše i najefikasnije primjenjiva. Dokazivanjem primjene istraživane tehnologije u segmentiranju i profiliranju korisnika kao i u detaljnijoj, točnijoj te bržoj procjeni kreditnog rizika, pokazat će potencijal Big Date kao neizostavnog alata suvremenog bankarstva.

U drugom djelu istraživanja, istraživati i analizirati će se svjesnost zaposlenika na primjeru odabrane banke, na relevantnim pozicijama, o tehnologiji velikih podataka. Cilj ovog djela istraživanja je definirati razinu informiranosti djelatnika o spomenutoj tehnologiji te definiranu razinu analizirati kao prihvatljivu ili neprihvatljivu.

Ciljevi istraživanja koje smo naveli će se realizirati metodama koje su navedene i objašnjene u nastavku. Rezultati analize hipoteza, koje su postavljene, ukazuju na važnost primjene istraživane tehnologije na području upravljanja podacima što doprinosi efikasnijem poslovanju banaka. [9]Biti efikasan znači slijediti proces koji koristi najmanju količinu energije i na najmanju mjeru svodi rasipanje energije. Efikasnost je posljedica poštovanja prave forme.

[9] http://www.efos.unios.hr/poduzetnicke-strategije/wp-content/uploads/sites/231/2017/03/Efektivnost_efikasnost_Adizes.pdf

Nema prostora za greške. Kada koristite sustav koji je osmišljen tako da bude efikasan, proces učenja nije uključen. Samo trebate pratiti programirani, propisani sustav, koji vam detaljno govori gdje, kada, kako i s kim što raditi.

1.5. Metode istraživanja

Prilikom ovog istraživanja koristit će se različite istraživačke metode. Metodologija istraživanja u ovom radu dijeli se na dva dijela- teorijska analiza i empirijska analiza.

U teorijskom dijelu rada koristit ćemo stručnu i znanstvenu literaturu koja se odnosi na problematiku ovog istraživanja. Ova literatura odnosi se na već postojeće podatke odnosno sekundarne podatke koje smo prikupili iz različitih izvora.

Metode koje će se koristiti su :

- Komparativna metoda : Metoda koja omogućuje usporedbu sličnosti i razlike u načinima obrade podataka, poslovnim procesima te rezultatima kod banaka koje primjenjuju alate tehnologije velikih podataka. Koristeći ovu metodu ukazat ćemo na različite razine efikasnosti poslovanja banka koje koriste alate tehnologije velikih podataka te onih koje ih ne koriste ili koriste u manjoj mjeri.
- Induktivna i deduktivna metoda:[10]Pomoću ove metode, kroz pojedinačna zapažanja formirat ćemo opće zaključke. dok deduktivnom metodom ćemo izvesti posebne stavove iz općih zaključaka. Koristeći navedene metode doći ćemo do zaključka o razini efikasnosti u primjeni Big Data tehnologije na primjerima poslovanja suvremenih banaka.
- Metoda analize i sinteze: [11]Analiza je postupak znanstvenog istraživanja raščlanjivanjem složenih pojmova, sudova i zaključaka na njihove jednostavnije sastavne dijelove i elemente. [12]Sinteza je na čin sistematiziranja znanja po

[10] http://www.unizd.hr/portals/4/nastavni_mat/1_godina/metodologija/METODE_ZNANSTVENIH_ISTRAZIVANJA.pdf

[11] http://www.unizd.hr/portals/4/nastavni_mat/1_godina/metodologija/metode_znanstvenih_istrazivanja.pdf

[12] http://www.unizd.hr/portals/4/nastavni_mat/1_godina/metodologija/metode_znanstvenih_istrazivanja.pdf

zakonitostima formalne logike, kao proces izgradnje teorijskog znanja u pravcu od posebnog ka općem, odnosno od vrste prema rodu. Koristeći navedene metode ćemo definirati pojam tehnologije velikih podataka, kao i same alate te njihovu učinkovitost u poslovanju određenih subjekata.

- Metoda deskripcije: Opisujući određene pojave vezane uz poslovanje banaka, ukazat će se na važna obilježja tehnologije velikih podataka.
- Statistička metoda: U ovom radu ćemo provesti pojedinačna statistička istraživanja kojim ćemo dokazati zadanu hipotezu.
- Metoda modeliranja: Pomoću ove metode ćemo kroz aplikativni primjer pokazati primjer obrade podataka u programskog okruženju koje koristi tehnologiju velikih podataka.

Empirijski dio zapravo služi za praktičnu primjenu. U analizi empirijskog dijela rada koristit će se rezultati primarnih i sekundarnih podataka.

U posljednjoj fazi istraživanja će se sintetizirati teorijski i empirijski dio istraživanja. Dodatno će se interpretirati i prezentirati prikupljene podatke, nakon čega će se donijeti konačni zaključak istraživanja.

1.6. Doprinos Istraživanja

U današnjem, izuzetno promjenjivom, izazovnom te konkurentnom poslovnom okruženju detalji čine razliku među onima koji će preživjeti surovost otvorenog tržišta i onima koji neće uspjeti. Na tragu ove činjenice, zaključujemo kako menadžment poduzeća nema pravo na grešku ako želi opstati i rasti na tržištu. Općenito tehnologija, samim time i [13]Big Data, je važan alat koji omogućuje postavljanje ključnih parametara za donošenje odluka u realnom vremenu na svim razinama menadžmenta. U ovom radu ćemo pokušati skrenuti pozornost na važnost kohabitacije IT segmenta s tradicionalnim menadžmentom banaka.

U ovom radu će se objasniti Big Data tehnologiju, kako bi je približili svima onima koji se do sada nisu upoznali a planiraju se razvijati ili već razvijaju karijeru u bankarskom sektoru. Objasnit će se ključni načini primjene ove tehnologije na određena područja poslovanja

[13] https://www.investopedia.com/terms/b/big-data.asp

bankarskog sektora te ukazati na trenutnu svijest i otvorenost na primjeru jedne od većih financijskih institucija u Republici Hrvatskoj.

Ključni doprinos ovog rada je osvješćivanje, trenutnih i budućih stručnjaka s područja financija, marketinga te svih vrsta i razina menadžmenta o Big Data tehnologiji i njenim mogućnostima. Kroz upoznavanje i buđenje zainteresiranosti za navedene alate odlučivanja, u budućnosti, bi se mogla olakšati komunikacija na važnoj relaciji it sektor i ekonomski stručnjaci.

2.Big Data

Budući da se informacijska tehnologija u kontekstu inovacija razvija pod utjecajem suvremenog načina življenja, zbirka digitalnih podataka raste eksponencijalno. Danas postoji ogromna količina podataka koji se generiraju svakodnevno u sektorima proizvodnje, poslovanja, znanosti i našeg osobnog života. Pravilna obrada podataka mogla bi otkriti nova znanja o našem tržištu, društvu i okolišu i omogućiti nam pravovremeno reagiranje na nove mogućnosti i promjene. Međutim, rast obujma podataka u digitalnom svijetu unaprijedio je naše računalne infrastrukture. Konvencionalne tehnologije obrade podataka, poput baza podataka i skladišta podataka postaju neadekvatne za količinu podataka s kojima se bavimo.

Ovaj novi izazov naziva se tehnologija Velikih podataka. Veliki podaci su podaci koji nadmašuju kapacitete procesuiranja uobičajenih sustava baza podataka. Podaci su preveliki, kreću se prebrzo ili se ne uklapaju u strukturu arhitekture baze podataka. Da bi ovakvi podaci postali korisni, potrebna je suvremena tehnologija procesuiranja istih. Velika količina podataka, zbog svoje veličine i kompleksnosti, traži nove mogućnosti, alate i modele za upravljanje informacijama, te njihovim vanjskim i unutarnjima tokovima. Transformacija velike količine podataka u strateške resurse preduvjet je za zadovoljenje potreba budućih korisnika. S druge strane, izazovi koje velika količina podataka postavlja zahtjeva promjenu poslovnih modela i ljudskih resursa.

2.1. Karakteristike Velikih podataka

Velike podatke je potrebno razlikovati od velike količine podataka. Kako bi se određeni skup podataka mogao smatrati Velikim podacima mora posjedovati karakteristike popularno zvane „3V“ prema početnim slovima značajki na engleskom jeziku *volume, velocity* i *variety* .

2.1.1. Volumen

Volumen (eng. *volume*) je prva dimenzija, a odnosi se na dostupnost izuzetno velike količine podataka kojoj se može pristupiti i koja se može pohraniti putem interneta. Zbog prirode nastajanja velikih podataka, njihova količina obično se kreće od nekoliko gigbajta pa do zetabajta. Primjerice, procjenjuje se da Facebook trenutno skladišti oko 300 PB (10^{15} bajta)

podataka. Broj novih komentara i lajkova na Facebook-u dnevno iznosi 3 milijarde. Broj pregleda videa na YouTube-u dnevno prelazi milijardu. Twitterovi korisnici dnevno objave preko 500 milijuna tvitova. Sve te podatke potrebno je negdje pohraniti te omogućiti da njihova obrada bude brza i jedinstvena. Upravo[14] iz tog razloga volumen predstavlja jedan od najvećih izazova u tehnološkom smislu .

Za razliku od statističkih metoda pomoću kojih bi odgovor na pitanje tražili na uzorku dijela populacije, tehnologija velikih podataka omogućava obradu cijele populacije. Taj pomak u količini obrađenih podataka otvara vrata novim perspektivama iz kojih se mogu promatrati veze među podacima.

2.1.2. Brzina

Brzina (eng. *velocity*) je druga dimenzija, a odnosi se na dinamiku velike količine podataka. Izazov brzine dolazi s potrebom za rješavanjem brzine kojom se stvaraju novi podaci, ili ažuriraju postojeći. Procjenjuje se da će broj povezani internet stvari u svijetu ove godine doseći 6,4 milijarde, a njihov broj dnevno raste za 5,5 milijuna. Novi izvori donose i novu brzinu kojom se podaci prikupljaju. Senzori koji se koriste u prometu prikupljaju podatke u realnom vremenu. Web stranice prikupljaju podatke o svakom kliku svojih posjetitelja. Takvi podaci mogu se iskoristiti u analizi korisnikovog ponašanja kako bi se unaprijedila prodaja, odnosno bilo kakav ciljni događaj. Još jedan od primjera je prijenos podataka generiranih na stroju, poput onih koje generiraju senzori ili mobilni uređaji. U tim aplikacijama, velika količina novih i ažuriranih podataka neprestano odlazi u sustav, dok mi zahtijevamo od tog istog sustava smisao tih podataka u realnom vremenu. Brzina podataka donosi izazove svakom sloju platforme za upravljanje podacima. I sloj za pohranu i sloj za obradu moraju biti iznimno brzi i skalabilni. Tehnologija prijenosa podataka nekoliko je godina istraživana za obradu velike brzine, međutim kapacitet postojećih strujnih sustava i dalje je ograničen, pogotovo kada se bave povećanim količinama ulaznih podataka.

2.1.3. Raznolikost

[14] Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu: Big data Challenge: a data management perspective, Key Laboratory of Data Engineering and Knowledge Engineering, School of Information, Renmin University of China, Beijing 100872, February 22, 2013, Front. Comput. Sci., 2013, 7

Aplikacije u realnom svijetu i vremenu dobivaju podatke koji često ne dolaze iz jednog izvora. Velike implementacije podataka zahtjevaju njihovu obradu iz različitih izvora u kojima podaci mogu biti različitih formata i modela. Raznolikost (eng. *variety*) podataka pruža više informacija za rješavanje problema ili za bolju uslugu. Izazov ove dimenzije je kako „uhvatiti“ različite vrste podataka na način koji omogućuje povezivanje njihovih značenja.

Podaci se mogu klasificirati u tri opće vrste podataka: strukturirane, polu-strukturirane i nestrukturirane.

Podaci u relacijskim bazama podataka strogo su strukturirani. Smješteni su u tablice sa unaprijed definiranim stupcima u kojima je poznata maksimalna duljina i tip podataka. Svaki redak tablice predstavlja jedan zapis. Takvi podaci su primjerice podaci o transakcijama, podaci koje prikupljaju senzori, web logovi i slično.

Kada skup podataka djelomično poštuje neku strukturu, ali ne u potpunosti, radi se o skupu polu-strukturiranih podataka. Oni mogu sadržavati parove oznaka i vrijednosti, poput XML datoteka.

Nestrukturirani podaci su primjerice blog postovi, novinski članci, slike, itd. Takvi podaci ne poštuju određeni format pa ih je do razvoja tehnologija za obradu velikih podataka bilo moguće samo prikupljati te naknadno ručno analizirati. Nestrukturirani podaci čine većinu od ukupne količine podataka. Procjenjuje se da je taj udio oko 80%. Nestrukturirane podatke također možemo podijeliti na podatke generirane od strane čovjeka (tekstualni podaci unutar neke tvrtke, podaci društvenih mreža, mobilnih telefona, web stranica i sl.) te od strane računala (meteorološke satelitske snimke, znanstveni podaci, fotografije i video zapisi, podaci prikupljeni sonarom ili radarom, itd.).

Navedena svojstva čine velike podatke neprikladnim za obradu i pohranu u standardnim relacijskim bazama podataka pa se javila potreba za razvojem novih tehnologija i alata za njihovu obradu .

Mnogi autori koriste još nekoliko V-ova pri definiranju velikih podataka. Neki od uobičajenih dodataka su vjerodostojnost (eng. *Veracity*), vrijednost (eng. *Value*), promjenjivost (eng.

Variability), nestalnost (eng. *Volatility*), vizualizacija (eng. *Visualisation*). To su također svojstva velikih podataka, no ona ne čine razliku između velikih podataka i ostalih podataka.

2.2. Prednosti i nedostaci tehnologije velikih podataka

Prema istraživanju TDWI (*Transforming Data With Intelligence*)[15] o prednostima i nedostacima tehnologije velikih podataka, ali i infrastrukture i ljudi koji su potrebni za uspješnu implementaciju ove tehnologije, kao najveći razlog naveden je nedostatak vještina i osoblja potrebnog za *Big Data* analitiku . Ovaj razlog se u brojnim člancima navodi kao najveći problem za uspjeh, pa čak i odustajanje od implementacije ove tehnologije. Obzirom da se radi o relativno novim tehnologijama, veliki je i nedostatak za kadrom koji je osposobljen za rad s velikim podacima, stoga su tvrtke primorane na improvizaciju, prekvalifikaciju već postojećih radnika, dodatno obrazovanje, a shodno tome i dodatne troškove kako bi imali eventualne dobitke u budućnosti. Iako se radi o programima otvorenog koda (eng. *open-source software*) i dalje postoje troškovi koji mogu biti poprilično veliki. Tu se naravno govori o troškovima edukacije osoblja, a ne samo o troškovima hardvera i softvera. Još neki od razloga ističu se i nedostatak podrške od strane upravitelja i nedostatak poslovnih slučajeva. Na kraju se postavlja pitanje: da li je tehnologija velikih podataka svojevrsni problem za jednu organizaciju? Prema istraživanju TDWI-a 30% ispitanika smatraju ovu tehnologiju problemom, a kao jedan od ključnih razloga ističe se količina podataka. No, s druge strane 70% ispitanika u ovoj tehnologiji vidi priliku za napretkom .

Velika količina podatka može imati veliku ulogu u razvijanju marketinške strategije, zadržavanja postojećih kupaca i poboljšanje prodaje. Uz pomoć novih tehnika poboljšanja odnosa s korisnicima usluga, tvrtke više ne moraju razdvajati tržišta u velike demografske skupine. Umjesto toga, tvrtke mogu koristiti nove analitičke alate, te uz pomoć njih analizirati velike količine podataka, otkriti nove niše, ili čak dodatno, sa većom preciznošću, segmentirati već postojeće skupine u manje i bliskije. Primjerice, velika količina podataka marketingu obećava masovnu prilagodbu. S obzirom da je većina objava spontano mišljenje na društvenim mrežama, takvi su podaci puno vrjedniji od ispitivanja putem upitnika i slično. Analizom takvih podataka mogu se otkriti zahtjevi za nove proizvode i/ili usluge. Marketing

[15] Russom, P. (2011). TDWI Best Practices Report – Big Data Analytics. TDWI Research. Renton

želi predvidjeti promjene u kupčevim željama prije nego se one dogode, a fina analiza velike količine informacija o mišljenjima kupaca rezultira upravo takvim podacima.

Tvrtke koje ulažu u analitiku velike količine podataka mogu učiniti pogrešku i posljedično izgubiti povrat na investiciju ukoliko ne znaju uključiti podatke u kompleksno donošenje odluka. Temeljni koncept kaže da se ekstrakcija korisnih znanja iz podataka, sa ciljem rješavanja poslovnih problema, može provoditi sistematski praćenjem procesa čije su faze dobro definirane. Dodatno, smatra se da rezultati analitike podataka zahtijevaju pažljiv odabir konteksta unutar kojeg će se upotrijebiti. Informacijske tehnologije mogu se upotrijebiti za pronalaženje takvih informativnih podataka unutar velike količine podataka.

Prema istraživanju Economist Intelligence Unit (2015)[16] 85% ispitanika smatra da je problem kod upotrebe tehnologije velikih podataka nije količina, već sposobnost analitike i reagiranja u realnom vremenu, te hardverske mogućnosti koje tvrtka ima na raspolaganju. Većina ljudi ima kao pokretač stare koncepte povećanja vrijednosti, želeći pritom raditi stare stvari efikasnije i efektivnije, a korištenjem novih koncepata i analitikom velike količine podataka. Nedostatak vještina, kulturološke prepreke, procesi i struktura, te raspoloživa tehnologija, glavni su problemi na koje organizacije dolaze prilikom migracije prema korištenju tehnologije velikih podataka u svojem poslovanju.

Na kraju koliko god bilo nedostataka, prednosti tehnologije velikih podataka više se ističu. Nedostaci su premostivi, a koristi ove tehnologije višestruki. Ono što se očekuje od poslovne organizacije u budućnosti kako bi iskoristili ovu tehnologiju je obuka ljudi, kontrola troškova, ali najvažnije je prepoznavanje tehnologije velikih podataka te njenih mogućnosti u pogledu napretka. Također, važno pitanje je što ova tehnologija može učiniti poslovanju ako je iskoristi konkurentska tvrtka .

2.3. Programsko okruženje Hadoop

[16] Economist Intelligence Unit (2015). The Deciding Factor. Big Data & Decision Making.Capgemini. http://capgemini.com/thought-leadership/the-deciding-factor-big-data-decision-making

Kada se govori o tehnologijama velikih podataka, najčešće se misli na pojam Hadoop. Hadoop[17] je programsko okruženje otvorenog koda koje, koristeći jednostavne programske modele, služi za raspodjeljenu pohranu i obradu velikih skupova podataka na računalnim nakupinama (eng. cluster). Nastao je 2005. godine, a kreirali su ga Doug Cutting i Mike Cafarella u programskom jeziku Java. Hadoop osigurava pohranu velike količine podataka (bilo koje vrste) uz iznimno snažnu procesorsku obradu virtualno neograničenog broja zadataka ili poslova.

Hadoop [18]je dizajniran kako bi pružio usluge prema pojedinačnim poslužiteljima (do tisuće uređaja), osiguravajući pritom lokalne proračune i pohranu. Kako bi pružio isporuku visoke raspoloživosti, Hadoop se ne oslanja na sklopovlje, već na same biblioteke koje su dizajnirane za otkrivanje i otklanjanje kvarova na aplikacijskom sloju i isporuku visoko raspoloživih usluga na vrhu nakupine računala. Jezgra Hadoop-a sastoji se od dijela za pohranu – HDFS (eng. Hadoop Distributed File System) i dijela za obradu – MapReduce. Hadoop dijeli datoteke u velike blokove i distribuira ih među čvorovima unutar nakupine računala. Za obradu podataka Hadoop MapReduce prenosti zapakirane kodove čvorova kako bi se paralelno obradili prema principu da se svaki čvor mora obraditi. Takav pristup daje prednost u odnosu na lokalnost podataka (čvorovi manipuliraju podacima koje imaju) što omogućuje bržu obradu podataka i veću učinkovitost u odnosu na konvencionalne super-računalne arhitekture koje se oslanjaju na paralelni datotečni sustav gdje su podaci i proračuni povezani mrežom velike brzine (eng. high-speed network).

Hadoop se sastoji od četiri komponente. To su:

- Hadoop Common paket koji sadrži biblioteke i uslužne programe za druge module,
- Hadoop raspodijeljeni datotečni sustav (HDFS) – sustav koji pohranjuje podatke na strojevima za pričuvu,
- Hadoop MapReduce – programski model za obradu velikih skupova podataka, i
- Hadoop YARN (eng. Yet Another Resource Negotiator) – platforma odgovorna za upravljanje računalnih resursa u nakupinama računala koja ih koristi kod raspoređivanja korisničkih aplikacija.

[17] Julio, P. (2009) Big Data Analytics with Hadoop. LinkedIn Corporation. http://www.slideshare.net/PhilippeJulio/hadoop-architecture

[18] Lockwood, G. K. (2014). Conceptual Overview of Map-Reduce and Hadoop. http://www.glennklockwood.com/data-intensive/hadoop/overview.html

2.3.1. Prednosti korištenja *Hadoop*-a u poslovanju

Jedan od glavnih razloga zašto tvrtke uvode Hadoop u poslovanje je[19] njegova sposobnost pohrane i obrade velike količine podataka bilo koje vrste velikom brzinom. Uz konstantno povećanje obujma i izvora podataka (raznolikost) s društvenih mreža i internet objekata, brzina obrade je od ključnih osobina koje se uzimaju u obzir. Uz brzinu, drugi bitni epiteti su:

- računalna snaga – Hadoop-ov raspodijeljeni računalni model brzo obrađuje podatke. Što se veći broj računalnih čvorova koristi, veća je i računalna snaga,
- fleksibilnost – za razliku od tradicionalnih relacijskih baza podataka, podaci se ne moraju predobraditi prije spremanja. Korisnik može pohraniti gotovo neograničen broj podataka i kasnije odlučiti što želi s tim podacima. To mogu biti tekstualni podaci, slike, video zapisi i slično,
- toleriranje kvarova – obrada podataka je zaštićena od potencijalnog kvara sklopovlja. U slučaju kvara jednog čvora, poslovi se automatski preusmjeravaju na druge čvorove kako bi se osiguralo raspodijeljeno računarstvo od kvara. Uz to, u slučaju kvara/ispada, sustav automatski sprema kopije svih podataka.
- niska cijena – okruženje otvorenog koda je besplatno i koristi poslužitelje za pohranu velikih skupova podataka, i
- skalabilnosti – nadogradnja sustava se provodi jednostavnim dodavanjem više čvorova, tablica ili datoteka u sustav gdje nisu potrebne velike administrativne promjene.

Skalabilnost je izrazito važna komponenta ovog sustava, a njegove instance mjere se korištenjem slijedećih faktora:

1. Datoteke – Hadoop-ova početna arhitektura sastoji se od jednog NameNod-a. To ograničava Hadoop grozd na 100 do 150 milijuna datoteka. Taj broj ovisi i o dostunoj memoriji za metapodatke. U malim grozdovima, maksimalan broj blokova na svakom čvoru dodatno ograničava broj dostupnih datoteka. Stoga je važno odabrati Hadoop platformu koja izbjegava zagušenje koje nastaje zbog jednog NameNod-a, te koja ima

[19] Mitchell, R.L. (2014). 8 big trends in big data analytics. Computerworld, http://www.computerworld.com/article/2690856/8-big-trends-in-big-data-analytics.html

distribuiranu arhitekturu za metapodatke. Na taj način može se proširiti na količinu koja prelazi miljardu datoteka i tablica,

2. Broj čvorova – druga dimezija skalabilnosti je broj fizičkih čvorova. Ovisno o zahtjevima procesuiranja ili podatkovnog pohrambenog prostora, odabrana Hadoop distribucija može zahtjevati i tisuće fizičkih čvorova, i
3. Kapacitet/gustoća čvorova – u slučajevima pohrambeno intenzivnih korištenja, potrebno je napraviti proširenje putem čvorova koji imaju veći kapacitet.[20] Takav pristup služi smanjenju općeg broja čvorova koji su potrebni za pohranu date količine podataka.

2.4. Najčešći alati Big Date u primjeni

Big data ima mogućnost promijeniti načine na koje se donose poslovne odluke u svakodnevnom poslovanju. Količina podataka raste eksponencijalnom brzinom, prema IDC International Data Corporation[21]ukupna količina podataka iz 2018te od 33 zetabajta će, do 2025, narasti na 175 zetabajta. Kako bi se stekao dojam veličine 1 zetabajt iznosi 1 trilion gigabajta. Ovakvim rastom količine podataka, jasno je da tradicionalni alati za obradu podataka već sada nisu dovoljni, a posebice neće biti u budućnosti.

Mnoge velike kompanije su pridale veliku važnost big data tehnologiji te tako osnivaju posebne timove, unutar svojih organizacija, koji se bave big data tehnologijom. Big Data alati, prema Binu Mathewu [22], dovode do smanjenja troškova te povećanja produktivnosti poduzeća. To je smjer u kojem svaka organizacija želi ići, zbog čega big dana kao tehnologija, a samim time i alati big dana tehnologije, postaju sve važniji faktor poslovanja. Na tržištu postoji veliki broj alata tehnologije velikih podataka, u daljnjem tekstu ćemo navesti 10 najpopularnijih alata, prema Dataflair timu[23]

[20] Schneider, R. D. (2013). Hadoop Buyer's Guide. Ubuntu. http://insights.ubuntu.com/wp-content/uploads/HadoopBuyersGuide_sm.pdf
[21] https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf
[22] https://www.worldoil.com/news/2016/12/13/how-big-data-is-reducing-costs-and-improving-performance-in-the-upstream-industry
[23] https://data-flair.training/blogs/top-big-data-tools/

2.4.1. Apache Hadoop

O ovom alatu je detaljno pisano u prethodnim poglavljima. Definitivno jedan od najvažnijih alata Big Data tehnologije. Apacheov program otvorenog koda, koji koristi klasteriranu arhitekturu te omogućuje paralelnu obradu podataka jer radi na više strojeva odjednom.

Slika 1. : Logo Apache Hadopp alata

Izvor: Slide Share https://www.slideshare.net/pothiq/introduction-to-apache-hadoop-ecosystem

2.4.2. Apache Spark

Apache Spark se može smatrati Hadoopovim nasljednikom. Razlog tome je jer prevladava nedostatke koje Hadoop ima. Spark podržava i realno vrijeme kao serijsku obradu. Također podržava izračune u memoriji, što ga čini 100 puta bržim od Hadoopa. To je omogućeno smanjenjem broja operacija čitanja/ pisanja na disk. Pruža više fleksibilnosti i svestranosti u odnosu na Hadoop jer djeluje s različitim spremištem podataka kao što su HDFS, OpenStack i Apache Cassandra.

Slika 2.: Logo Apache Spark

Izvor: Analytics Vidhya https://www.analyticsvidhya.com/blog/2020/06/22-tools-data-science-machine-learning/apache_spark_logo/

2.4.3. Apache Storm

Prema Clouderi, [24]sustav za obradu streaming podataka u stvarnom vremenu Apache ™ Storm dodaje pouzdane mogućnosti obrade podataka u stvarnom vremenu u Enterprise Hadoop. Storm on YARN moćan je za scenarije koji zahtijevaju analitiku u stvarnom vremenu, strojno učenje i stalno praćenje operacija. Oluja se integrira s YARN-om putem Apache Slider-a, YARN upravlja Storm-om, dok također razmatra klasterske resurse za upravljanje podacima, sigurnosne i operativne komponente moderne arhitekture podataka.

Storm je distribuirani računski sustav u realnom vremenu za obradu velikih količina podataka velike brzine. Storm je izuzetno brz, s mogućnošću obrade preko milijun zapisa u sekundi po čvoru na klaster skromne veličine. Poduzeća koriste ovu brzinu i kombiniraju je s drugim aplikacijama za pristup podacima u Hadoopu kako bi se spriječili neželjeni događaji ili optimizirali pozitivni ishodi. Neke od novih poslovnih prilika uključuju: upravljanje uslugama za korisnike u stvarnom vremenu, unovčavanje podataka, operativne nadzorne ploče ili analizu cyber sigurnosti i otkrivanje prijetnji.

[24] https://www.cloudera.com/products/open-source/apache-hadoop/apache-storm.html

Slika 3.: Logo Apache Storm

Izvor: Datanami https://www.datanami.com/2015/01/09/lockheed-aims-make-apache-storm-easier-use/

2.4.4. Apache Cassandra

Apache Cassandra je distribuirani sustav za upravljanje bazama podataka koji je izgrađen za obradu velikih količina podataka u više podatkovnih centara i oblaku. Vrlo je skalabilan, nudi visoku dostupnost i nema nijednu točku pogreške. Napisano na Javi, to je NoSQL baza podataka koja nudi mnogo toga što druge NoSQL i relacijske baze podataka ne mogu. Cassandra je izvorno razvijena na Facebooku za značajku pretraživanja u pristigloj pošti. Facebook ju je otvorio 2008. godine, a Cassandra je postala dio Apache inkubatora 2009. godine. Od početka 2010. Apache je bio na najvišoj razini. Trenutno je ključni dio Apache Software Foundation i može ga koristiti svatko tko želi imati koristi od toga.

Stephen Watts, [25] smatra kako Cassandra je jedna od najučinkovitijih i najčešće korištenih NoSQL baza podataka. Jedna od ključnih prednosti ovog sustava je ta što nudi visoko dostupnu uslugu i nema ni jedne točke kvara. Ovo je ključno za tvrtke koje si mogu priuštiti da im se sistem pokvari ili izgubi podatke. Bez ijedne točke pogreške, nudi uistinu dosljedan pristup i dostupnost. Još jedna ključna prednost Cassandra je ogromna količina podataka s kojom se sustav može baviti. Može učinkovito i efikasno obraditi ogromne količine podataka na više poslužitelja.

[25] https://www.bmc.com/blogs/apache-cassandra-introduction/ Stephen Watts 2020

Slika 4. : Logo Apache Cassandra

Izvor: Wikipedia https://en.wikipedia.org/wiki/Apache_Cassandra

2.4.5. Mongo DB, Apache Flink, Kafka

Mongo DB po DataFlairu,[26] je alat za analizu podataka otvorenog koda, NoSQL baza podataka koja pruža mogućnosti platformi. To je primjer za posao kojem su potrebni brzorastući podaci u stvarnom vremenu za donošenje odluka. MongoDB je savršen za one koji žele rješenja na temelju podataka. Jednostavan je za korisnika jer nudi jednostavniju ugradnju i održavanje. MongoDB je pouzdan, kao i isplativ. MongoDB To je napisano na C, C ++ i JavaScript. To je jedna od najpopularnijih baza podataka za velike podatke jer olakšava upravljanje nestrukturiranim podacima ili podacima koji se često mijenjaju. MongoDB koristi dinamičke sheme. Stoga podatke možete brzo pripremiti. To omogućava smanjenje ukupnih troškova. Izvodi se na MEAN stoku softvera, NET aplikacijama i, Java platformi. Fleksibilan je i u oblačnoj infrastrukturi, no primjećen je određeni pad brzine obrade kod nekih slučajeva upotrebe.

Apache Flink je otvoreni izvorni alat za analizu podataka koji distribuira okvir za obradu za ograničene i neograničene tokove podataka. Piše na Javi i Scali. Omogućuje rezultate visoke točnosti, čak i za podatke koji stižu kasno. Flink je izvanredan i otporan na greške, tj. Ima mogućnost lakog oporavka od grešaka. Omogućuje visoku učinkovitost visokih performansi,

[26] https://data-flair.training/blogs/top-big-data-tools/ DataFlair: Big Data tools

izvodeći na tisućama čvorova. Daje motor s niskim kašnjenjem, visokom propusnom protočnošću i podržava upravljanje vremenom događaja i upravljanjem državom.

Kafka je platforma otvorenog koda koju je LinkedIn stvorio 2011. godine. Apache Kafka je distribuirana platforma za obradu ili streaming događaja koja omogućuje visoku propusnost za sustave. Dovoljno je učinkovit za obradu trilijuna događaja dnevno. To je strujna platforma koja je visoko skalabilna i pruža veliku toleranciju na greške. Proces strujanja uključuje objavljivanje i pretplatu na tokove zapisa slične sustavima za razmjenu poruka, trajno pohranjivanje tih zapisa, a zatim njihovu obradu. Ti se zapisi pohranjuju u skupinama pod nazivom teme. Apache Kafka nudi streaming velike brzine i jamči nulta zastoja.

2.4.6. Tableau

Tableau je jedan od najpopularniji alata za vizualizaciju podataka i u industriji poslovne inteligencije općenito. To je alat koji oslobađa snagu vaših podataka. Pretvara vaše neobrađene podatke u vrijedne uvide i poboljšava proces odlučivanja u poslovanju. Prema DataFlairu, [27] Tableau nudi brzi postupak analize podataka, a rezultiralo je vizualizacijama u obliku interaktivnih nadzornih ploča i radnih listova. tablo Djeluje u sinkronizaciji s drugim Big Data alatima kao što je Hadoop. Tableau je ponudio mogućnosti kombiniranja podataka koje su najbolje na tržištu. Omogućuje učinkovitu analizu u stvarnom vremenu. Tableau nije vezan samo za tehnološku industriju, već je presudan dio i nekih drugih industrija. Za rad ovog softvera nisu potrebne tehničke ili programske vještine.

[27] https://data-flair.training/blogs/top-big-data-tools/ DataFlair: Big Data tools

Slika5. : Logo Tableau

Izvor: Proximous https://www.proximous.com/introduction-to-tableau-prep/

2.4.7. RapidMiner

RapidMiner je alat na više platformi koji nudi integrirano okruženje za nauku o podacima, strojno učenje i prediktivnu analizu. Nalazi se pod raznim licencama koje nude mala, srednja i velika vlasnička izdanja, kao i besplatno izdanje koje omogućuje 1 logički procesor i do 10.000 redaka podataka.

RapidMiner je alat otvorenog koda koji piše u javi. RapidMiner nudi visoku učinkovitost čak i kada je integriran s API-jevima i cloud uslugama. Pruža neke robusne alate i algoritme podataka.

Slika 6. : Logo RapidMiner

Izvor: itecor https://itecor.com/partners/rapidminer-logo/

2.4.8. R Program

R je programski jezik i programsko okruženje za statističku analizu, grafičku reprezentaciju i izvještavanje.

Neke od najvažnijih značajki R-a:

- R je dobro razvijen, jednostavan i učinkovit programski jezik koji uključuje uvjetovanja, petlje, rekurzivne funkcije definirane od strane korisnika i mogućnosti ulaza i izlaza. R ima učinkovito postrojenje za rukovanje i pohranu podataka,
- R pruža skup operatora za proračun na nizovima, popisima, vektorima i matricama. R nudi veliku, koherentnu i integriranu zbirku alata za analizu podataka.
- R nudi grafičke mogućnosti za analizu podataka i prikaz ili izravno na računalu ili ispis na papirima. Zaključno,
- R je svjetski najkorišteniji programski jezik statistike. To je prvi izbor znanstvenika s podacima i podupire ga živopisna i talentirana zajednica suradnika.
- R se podučava na sveučilištima i raspoređuje u kritičnim poslovnim aplikacijama. Ovaj će vas udžbenik naučiti programiranja
- R uz odgovarajuće primjere u jednostavnim i jednostavnim koracima.

Aplikativna primjena ovog programskog jezika će se prikazati u narednim poglavljima.

Slika 7. : Logo R programskog jezika

Izvor: Tie Talent https://tietalent.com/jobs/r-programming/

3. Primjena Big Data tehnologije u suvremenom bankarstvu

Podaci igraju ogromnu ulogu u razumijevanju vrijednih uvida u ciljanu demografiju i preferencije korisnika. Iz svake interakcije s tehnologijom, bez obzira na to je li aktivna ili pasivna, stvaramo nove podatke koji nas mogu opisati. Budući da se podaci prikupljaju putem proizvoda, video kamera, kreditnih kartica, mobitela i drugih dodirnih točaka, naš profil podataka eksponencijalno raste. Prema Linly Ku, [28] ako se pravilno analiziraju, ove podatkovne točke mogu nam puno objasniti o našem ponašanju, osobnostima i životnim događajima. Tvrtke mogu iskoristiti te spoznaje za poboljšanja proizvoda, poslovnu strategiju i marketinške kampanje kako bi se zadovoljile ciljanim kupcima.

Koncept velikih podataka i njegova važnost postoje već godinama, ali tek je nedavno tehnologija omogućila brzinu i učinkovitost pomoću koje se mogu analizirati veliki skupovi podataka. Kako će podaci - i strukturirani i nestrukturirani - u narednim godinama značajno narasti, prikupit će se i ispitati kako bi se otkrili neočekivani uvidi i čak pomogli predvidjeti budućnost. Veliki podaci promijenit će način na koji čak i najmanje tvrtke posluju jer prikupljanje i interpretacija podataka postanu dostupniji. Nove, inovativne i ekonomične tehnologije neprestano se pojavljuju i poboljšavaju što bilo kojoj organizaciji čini nevjerojatno lako da jednostavno implementira velika podatkovna rješenja.

3.1. Primjeri korištenja Big Date u velikim kompanijama

[28] https://www.plugandplaytechcenter.com/resources/impact-big-data-business/

3.1.1. Ikea

Ovaj švedski prodavač namještaja konstantno unapređuje svoje poslovanje kako bi bili u korak s vremenom. Ističu se jedinstvenim načinom proizvodnje, pakiranja, isporuke i recikliranjem. Svojim poslovanjem postavljaju standarde drugim kompanijama iz njihovog područja djelovanja.

Kako u navedenim aktivnostima Ikea svakodnevno inovira i mijenja pristupe poslovanju, tako i u primjeni alata poslovne inteligencije u svom poslovanju. Biti će naveden primjer korištenja Big Date tehnologije.

Još 2013. godine, IKEA je u svojoj aplikaciji pokrenula značajku za prepoznavanje slike i proširenu stvarnost. Kupci skeniraju predmete koje lajkaju svojim telefonom, izravno iz IKEA kataloga ili iz same trgovine. [29]Napredna analitika omogućuje im da virtualno postave namještaj koji im se svidio u njihove vlastite domove i vide kako to izgleda, kao i da promijene boju, veličinu i modele.

Druga upotreba velikih podataka odnosi se na preporuke za stavke. Slično kao u Amazonovom motoru, ovaj prodavač sugerira srodne ili komplementarne proizvode stavkama koje su klijenti pregledavali na web stranici, dodali u košaricu ili istaknuli kao favorite. Uz takvu personalizacijsku značajku, Ikea je omogućena za pružanje visoke razine zadovoljstva korisnika i ne samo u trgovini, već i u njihovoj web trgovini i aplikaciji. Nadalje, upotreba velikih podataka za proširenu stvarnost i naprednu analitiku doprinosi naporima tvrtke u održivosti smanjujući ljude koji zapravo voze do lokacije trgovine da bi nešto kupili. Na ovaj način, organizacija je odgovorna za otpremu velikog dijela narudžbi i ima slobodu optimizirati njihov prijevoz organizirajući ga na najmanje skup i najučinkovitiji način kako bi umanjila svoj utjecaj na okoliš.

3.1.2. Amazon

Neki će kupci možda smatrati neobičnim kada prodavaonica zna mnogo o njima jednostavno po kupljenim proizvodima. Amazon.com, Inc. (AMZN) vodeća je u prikupljanju, pohranjivanju, obradi i analiziranju osobnih podataka od vas i svih ostalih kupaca kao načina utvrđivanja načina na koji kupci troše svoj novac. Tvrtka koristi prediktivnu analitiku za

[29] https://blog.datumize.com/how-do-global-retail-leaders-use-big-data-5-real-life-examples#smooth-scroll-top

ciljani marketing kako bi povećala zadovoljstvo kupaca i izgradila lojalnost tvrtke. Način na koji Amazon koristi velike podatke pomogao je da se marka razvije u gigant među mrežnim prodavaonicama. [30]Jenifer Wills navodi načine na koje Amazon prati i prikuplja podatke o korisnicima te ih naknadno koristi za lakši proces kupovine.

Personalizirani sustav preporuka, Amazon je lider u korištenju sveobuhvatnog kolaboracijskog motora za filtriranje (CFE). Analizira koje ste artikle prethodno kupili, što se nalazi u vašoj internetskoj košarici ili na vašem popisu želja, koje ste proizvode pregledali i ocijenili te koje stavke najviše tražite. Te se informacije koriste za preporuku dodatnih proizvoda koje su kupili ostali kupci prilikom kupovine istih predmeta.

Preporuke za knjige iz Kindle Highlighting. Nakon što je 2013. kupio Goodreads, Amazon je integrirao uslugu društvenih mreža od oko 25 milijuna korisnika u neke od Kindle funkcija. Kao rezultat, Kindle čitatelji mogu istaknuti riječi i bilješke i podijeliti ih s drugima kao sredstvo za raspravu o knjizi. Amazon redovito pregledava riječi istaknute u vašem Kindleu kako bi utvrdio o čemu vas zanima. Tvrtka vam tada može poslati dodatne preporuke e-knjiga.

Naručivanje jednim klikom. Budući da veliki podaci pokazuju da kupujete negdje drugdje, osim ako vam proizvodi nisu brzo isporučeni, Amazon je kreirao naručivanje jednim klikom. Jednim klikom je patentirana značajka koja se automatski uključuje prilikom prve narudžbe i unosa adrese za dostavu i načina plaćanja. Kad odaberete narudžbu jednim klikom, imate 30 minuta tijekom kojih se možete predomisliti u vezi s kupnjom. Nakon toga proizvod se automatski naplaćuje putem vašeg načina plaćanja i dostavlja na vašu adresu.

Optimizacija cijena, veliki podaci koriste se i za upravljanje cijenama Amazona kako bi privukli više kupaca i povećali profit u prosjeku 25% godišnje.[31] Cijene se postavljaju prema vašoj aktivnosti na web mjestu, cijenama konkurenata, dostupnosti proizvoda, postavkama artikala, povijesti narudžbe, očekivanoj marži profita i drugim čimbenicima. Cijene proizvoda se obično mijenjaju svakih 10 minuta kako se veliki podaci ažuriraju i analiziraju. Kao rezultat toga, Amazon obično nudi popuste na najprodavanije artikle i ostvaruje veću zaradu od manje popularnih predmeta. Na primjer, cijena romana na popisu Best prodavača New

[30] https://www.investopedia.com/articles/insights/090716/7-ways-amazon-uses-big-data-stalk-you-amzn.asp
[31] https://www.investopedia.com/articles/insights/090716/7-ways-amazon-uses-big-data-stalk-you-amzn.asp

York Timesa može biti 25% manja od maloprodajne cijene, dok roman koji nije na popisu košta 10% više od iste knjige koju je prodao konkurent.

3.1.3. Starbucks

Starbucks je jedan od najpoznatijih brandova, vezanih za kavu, u svijetu već dugi niz godina. Posluju od 1971. te predstavljaju jedan od najprepoznatljivijih lanaca, općenito, u većini velikih gradova na svijetu.

Prema istraživanju poznatog časopisa Forbes, [32] Starbucksovi prihodi, u zadnje tri godine su rasli čak 26 %. Jedan od razloga takvom rastu je i big data tehnologija.

Brend koristi podatke u pogledu lokacije, demografije, ponašanja kod kupovine, trendova kupaca i drugih kako bi predvidjela uspjeh i buduće poslovanje svojih novih trgovina, koje će se otvoriti u različitim dijelovima svijeta. Na ovaj način organizacija uspijeva umanjiti rizik otvaranja prodavaonice na neprofitabilnoj lokaciji i na kraju spriječiti bilo kakav bankrot trgovine.

Štoviše, Starbucks koristi podatke o klijentima koje generira za marketinške poticaje i usluge njege za stalno komuniciranje s klijentima, čak i kad nisu u nekom od njihovih lokala. A to ima dvostruki učinak: pružiti personalizirane proizvode i postići višu razinu zadovoljstva kupaca.

[32] https://www.forbes.com/sites/greatspeculations/2019/09/26/starbucks-top-line-to-grow-by-10-in-fy-2019/#4fe578da494f

Slika 8. : Logo Starbucksa

Izvor: Indeed https://www.indeed.com/cmp/Starbucks

3.1.4. Costco

Kako Zornitsa Stoycheva navodi,[33] u slučaju Costca, big data je doslovno spasila živote korisnika.

Naime, Costco je američka multinacionalna korporacija na veliko, specijalizirana za maloprodaju svih vrsta prehrambenih proizvoda, kao i osobnih predmeta i predmeta za kućanstvo. Svrha u koju ovaj poslovni subjekt koristi velike podatke je prilično impresivna: Costco detaljno prati svaku narudžbu. To uključuje tko je poslao narudžbu (i kontaktne podatke), kada je kupnja izvršena i točno koji je artikl otpremljen kupcu. Možda vam ovo ne zvuči kao impresivno, ali slijedeći primjer će vas pokrenuti.

Costco je 2019. godine kupio i prodao šaržu voća, za koju se pokazalo da je potencijalno zagađena listerijama. Integriranje podataka korporaciji je dalo prednost da identificira svakog klijenta koji je kupio voće iz ove određene serije i upozori ih na moguću prijetnju. Oni ne samo da su to učinili, već su alarmirali kupce koristeći dva različita načina komunikacije: najprije telefonom, a potom i dopisom.

3.1.5. Asos

ASOS je ogromni britanski modni prodavač koji je promijenio igru kupovine odjeće i dodataka. Brend je u svoju aplikaciju uveo značajku prepoznavanja slike kako bi svojim kupcima omogućio skeniranje bilo kojeg odjevnog predmeta koji im se sviđa. Na temelju

[33] https://blog.datumize.com/how-do-global-retail-leaders-use-big-data-5-real-life-examples#smooth-scroll-top

svojih karakteristika, motor generira popis sličnih proizvoda: na primjer, ako vidite jaknu koja vam se sviđa, možete je slikati, a ASOS će vam pokazati svoje jakne sa sličnim karakteristikama, kao što su model, boja, pribor itd.

Druga upotreba te napredne analitike je značajka "Stil i podudaranje", gdje vam aplikacija prikazuje prijedloge kako upariti odjeću koju ste skenirali i kako poboljšati njihov modni izgled. Uz takvu upotrebu velikih podataka, ASOS postaje preferirana marka za kupovinu i modne savjete. Nadalje, te su karakteristike izvrstan marketinški alat za popularizaciju proizvoda i pozicioniranje tvrtke u umu svojih klijenata.

3.2. Primjeri upotrebe Big Data tehnologije u bankarstvu

DataFlair tim navodi [34] kako Big Data obnavlja svijet, a niti jednu industriju nije ostavila netaknutom svojim ogromnim prednostima. Nastao je kao spasilac za bankarsku industriju. Big data je dosad uštedio puno prihoda od bankarskih tvrtki i ima još mnogo toga za ponuditi u narednim godinama. To im daje uzdah olakšanja jer vođenje bankarske tvrtke nije tako lako kao što izgleda. Veliki podaci u bankarskoj industriji pomažu bankama u upravljanju rizikom, otkrivanju prijevara i u zadovoljstvu klijenata.

Navode nekoliko ključnih područja primjene tehnologije velikih podataka u bankarstvu. Niže će biti objašnjeni konkretni primjeri u tri navedena područja:

- Upravljanje rizicima
- Otkrivanju prijevara
- Zadovoljstvu kupaca

[34] https://data-flair.training/blogs/big-data-in-banking/

Slika 9. : Big Data u bankarstvu

Izvor: Dana Flair https://data-flair.training/blogs/big-data-in-banking/

3.2.1. Upravljanje rizicima (Risk Managment)

Uspostavljanje snažnog sustava upravljanja rizikom od najveće je važnosti za bankarske organizacije ili će u suprotnom morati pretrpjeti velike gubitke prihoda. Da bi ostali živjeti u konkurentnom svijetu i povećali profit koliko mogu, organizacije moraju stalno izmišljati nove stvari. Analizom velikih podataka tvrtke mogu u stvarnom vremenu otkriti rizik i očito spasiti kupca od potencijalne prijevare.

Slika10: Big Data upravljanje rizicima

Izvor: Data Flair https://data-flair.training/blogs/big-data-in-banking/

Data Flair tim navodi United Overseas Bank (UOB) Singapore case[35] kao primjer korištenje Big Data tehnologije na području upravljanja rizicima. UOB je treća najveća banka u jugoistočnoj Aziji. Odlučili su iskoristiti Big Datu kako bi riješili jedan od ključnih problema u upravljanju poslovanja svake banke, upravljanje rizicima.

UOB je započeo kockanje koristeći sustav upravljanja rizikom koji se temelji na velikim podacima. Proračun vrijednosti rizika zahtijeva veliku količinu vremena, obično traje do 20 sati. Kroz svoj sustav upravljanja rizicima Big Data, UOB je sada mogao izvršiti isti zadatak u samo nekoliko minuta, a sa ciljem da ga ubrzo realizira u stvarnom vremenu.

Dodatne podatke i informacije o ovom caseu je relativno teško naći, ali dovoljna je informacija koju navodi Dana Flair tim kako bismo shvatili ogromnu prednost korištenja Big Date u procijeni rizika. Na ovaj način UOB je dobio bržu i točniju informaciju, zadovoljnog korisnika te je oslobodila resurse za druge radnje unutar organizacije.

3.2.2. Otkrivanje prijevara

Brzo rastući digitalni svijet pruža nam brojne pogodnosti, ali s druge strane, rađa i razne vrste prijevara. Naši osobni podaci sada su osjetljiviji na cyber napade nego ikad prije i to je najveći izazov s kojim se suočavaju bankarske organizacije. Upotrebljavajući analitiku velikih podataka s nekim algoritmima strojnog učenja, organizacije sada mogu otkriti prijevare prije nego što se mogu staviti. To se postiže prepoznavanjem nepoznatih obrazaca potrošnje korisnika, predviđanjem neobičnih aktivnosti korisnika itd.

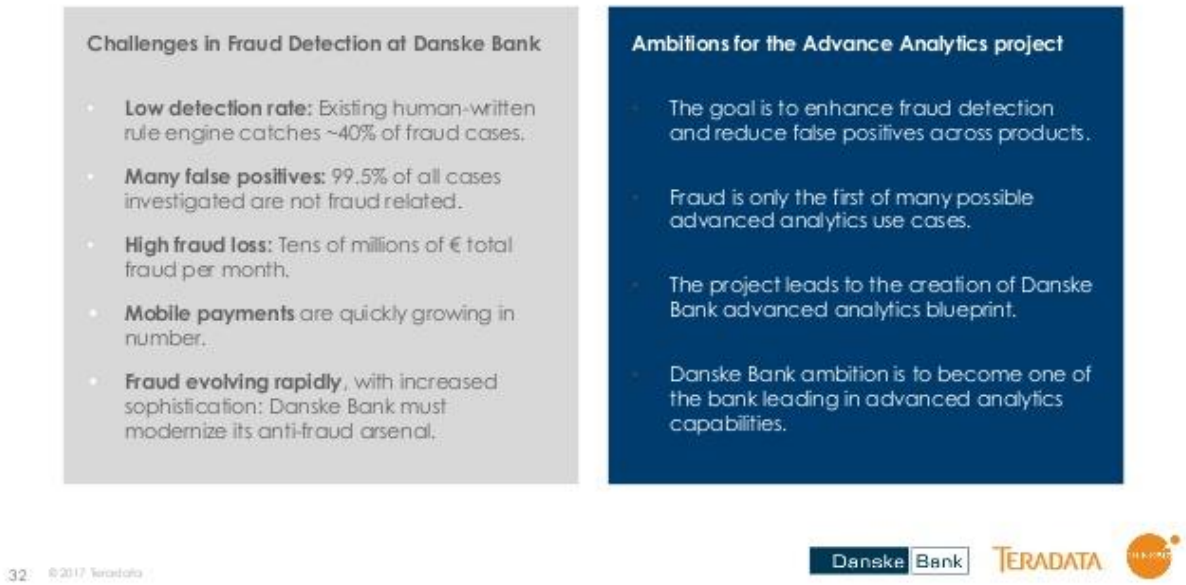
Katherine Knowles-Marchione i Mariah Kolpek[36] navode kako Danske Bank razlikuje dvije vrste prijevara, prevare s kupcima i „prevarante“. Kupac je u središtu prevare. Na primjer, kupac prima e-poštu od građanina u udaljenoj zemlji sa zahtjevom da mu pošalje novac kako bi pomogao ublažavanju teškoća. I onda se događa prava profesionalna prijevara kada „prevarant“ prati savršeno vrijeme za ozbiljnu štetu. To može uključivati zlonamjerni softver

[35] https://data-flair.training/blogs/big-data-in-banking/
[36] https://www.teradata.com/Blogs/Danske-Bank-Innovating-in-Artificial-Intelligence

koji je zaražen u banci ili na kojem se uzimaju osobne iskaznice i zlonamjerni softver se dodaje uređajima.

Navedeno tjera menadžment Danske Bank-a da zapošljava platformne, tehničke i podatkovne inženjere, znanstvenike o podacima, poslovne, pa čak i visoko osposobljene kriminalističke istražitelje Svi oni rade sa stručnjacima iz AI i Deep Learninga za uvođenje inovacija. Teradata je sustavu omogućio da donosi autonomne odluke u stvarnom vremenu koje su usklađene s visoko postavljenim sigurnosnim propisima. Rješenje pruža nove razine detalja, kao što je vrijeme serije i nizovi događaja kako bi se što bolje pomoglo banci u istragama prijevara. Cijelo rješenje bio je implementiran vrlo brzo - od starta do implementacije samo pet mjeseci.

Rezultati poslovanja nakon primjene Teradata rješenja su i više nego impresivni. Prije primjene AI i Deep Learninga, Danske Bank je imala 1200 lažnih pozitivnih rezultata dnevno. To su bili slučajevi koje su morali analizirati istražitelji Danske banke, ponekad čak i vanjske agencije poput Interpola. Sada je taj broj smanjen za 60%, čime su istražitelji banaka uštedjeli značajno vrijeme i omogućili im da istraže stvarne slučajeve prijevara. I to nije sve. Otkrivanje stvarnih pozitivnih veličina povećalo se na 50%. Timovi Danske bank vjeruju da je ovo tek početak.



Slika 11. : Izazovi u otkrivanju prijevara 'Danske Bank'Izvor: https://www.slideshare.net/RonBodkin/fighting-financial-fraud-at-danske-bank-with-artificial-intelligence

3.2.3. Zadovoljstvo kupaca

U ovoj novoj eri društvenih medija većina tvrtki se teško bori njihov način. Beyond The Arc[37], odlučio je istražiti, koristeći komentare na društvenim medijima, za pitanja površinske usluge koja mogu utjecati na zadržavanje korisnika. Za početak , analitičari podataka koristili su javno dostupne komentare na web stranicama društvenih medija, Facebook i Twitter, za prepoznavanje ključnih trendova u onome što ljudi govore o Banci Amerike.

Primjena podataka iz društvenih medija na Bank of America

U nastavku će biti preneseno analizu kompanije Byond The Arc [38]koja je provedena kroz tri koraka. Zašto Bank of America? Kao najveća banka u Americi po veličini imovine(ipreko 12% američkih depozita), poslovni analitičari bilježe da je banka maksimizirala udio depozita kroz akvizicije i da će uskoro trebati povećati stopu organskog rasta. To upućuje promjenu strategije poslovanja na način da će se morati fokusirati na postojeće kupce.

Korak 1- Prikupljanje podataka

U 15 dana, analitičari Beyond The Arca skupili su preko 41.000 komentara o Bank of America. Naglo su suzili 9.000 jedinstvenih komentara, od kojih je 88% bilo kratkih, tweet-ova osjećajno vođenih i 12% Facebook komentara koji su bili više narativne naravi.

Korak 2: Prepoznavanje teme putem analitike i poslovnog razumijevanja

Koristeći alate za prikupljanje podataka koji sortiraju i kategoriziraju velike količina podataka i korištenje pet plus godina analize podataka o korištenju financijskih usluga, analitičari otkrivaju primarne teme koje zabrinjavaju Bank of America, uključujući:

Mint.com - Problem veze između Mint.com i Bank of America izazvalo je veliko nezadovoljstvo mnogih kupaca, koje su izrazili putem Twittera i Facebooka. Neki su prijetili da će napustiti banku. Drugi su tvrdili da je Bank of America rekla im kako banka ima prekinute veze s agencijom Mint.com, pogrešna glasina koja se često ponavljala. Problem je

[37] https://beyondthearc.com/wp-content/media/cases/BTA-CaseStudy-BankofAmerica-SocialMediaAnalytics_10-5-11.pdf
[38] https://beyondthearc.com/wp-content/media/cases/BTA-CaseStudy-BankofAmerica-SocialMediaAnalytics_10-5-11.pdf

bio riješen tjedan dana kasnije, objava je stigla sa Mint.com. Bank of America nikada nije pokazala zabrinutost za navedene korisnike.

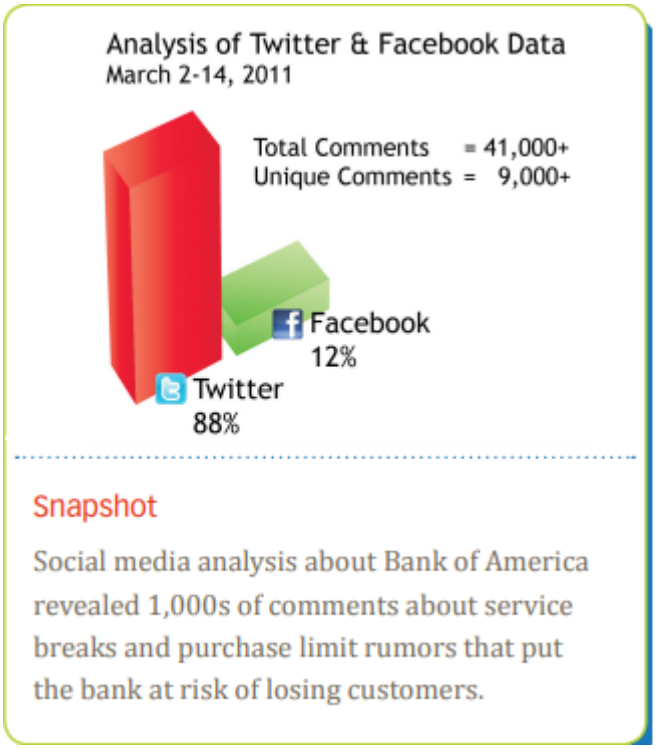
Prekid usluga - od 40.000 komentara identificirali su skoro 20 prekida u usluzi kao što su:

• Neuspjeh slanja novih debitnih kartica kupcima kojima su istekle kartice.

• Plaćanja koja nisu objavljena na vrijeme

• Porezni obrasci koji nisu stigli.

• Umnožavanje troškova na računima kupaca. Budući da ova situacija stvara frustrirane kupce i često je motivirajući faktor u prebacivanju banaka, ta su pitanja ključna područja koja zahtijevaju daljnju analizu i djelovanje.



Slika 12. : Analiza Twitter i Facebook podataka- Istraživanje Beyond the Arc

Izvor: file:///C:/Users/ipejic2/Desktop/BTA-CaseStudy-BankofAmerica-SocialMediaAnalytics_10-5-11.pdf Beyond The Arc

Korak 3 – Strategija i djelovanje

Procjena podataka s društvenih medija omogućuje bankama da identificiraju ključna pitanja koja nastaju i brzo reagiraju prije nego što se prošire izvan kontrole. U primjeru Mint.com, analizirajući komentare, Bank of America mogla se brzo otkriti područja za zabrinutost i proaktivno obavještavanje kupaca o situaciji, umjesto da dopuštaju širenje dezinformacija.

Zaključak - Korištenje podataka s društvenih medija za pokretanje poboljšanja usluga

Analitika na društvenim medijima moćan je alat koji pomaže tvrtkama koje pružaju financijske usluge da brzo utvrde probleme korisnika i pratiti raspoloženje javnosti o poslovanju. Kategoriziranjem velike količine komentara za utvrđivanje ključnih tema i trendova, analitika teksta može pretvoriti podatke društvenih medija u jasne pokazatelje za poboljšanja usluga, koja vam pomažu zadržati kupce i pružiti bolje korisničko iskustvo.

3.3 Husky Big Data Platforma

Kako ne bi svi primjeri ostali na razini velikih vanjskih financijskih institucija i kako se ne bi stekao dojam da kod nas Big Data nije stigla, navest ćemo primjer koji će možda i najbolje dočarati mogućnosti Big Date u suvremenom bankarstvu. Opisat ćemo rješenje tvrtke Combis koje je napravljeno za Erste Banku Hrvatska.

S razvojem inteligentnih uređaja raste i količina podataka koji oni generiraju. Iz ovog razloga razloga došlo je od razvoja Husky Big Data [39]platforme koja omogućuje prikupljanje, obradu, sigurnost i monetizaciju velikih količina podataka koje generiraju i pohranjuju mobilni operateri, banke i druge tvrtke i organizacije koje raspolažu velikom količinom informacija. Husky spaja podatkovni promet mobilnih mreža, transakcijski promet s Core banking sustava ili drugi vitalni sustav neke organizacije s Customer Relationship Management sustavom te kreira vrijedne izvještaje koji mogu omogućiti rast poslovanja ili usmjeravanje poslovanja u segmente koji donose snažniji profit. Prezentacija ovih informacija je u formi heat map-a, dashborda i izvještaja ili konkretnih akcija koje se pokreću automatski na temelju prikupljenih

[39] Husky. Combisovo (BIG) DATA rješenje. Brošura za financijski sektor. 2019. https://bigdata.combis.hr/wp-content/uploads/2017/08/Husky_za_financije.pdf

podataka. Također, u kontekstu podataka, moguć je pregled lokacija korisnika i njihovih aktivnosti po različitim vremenskim razdobljima – dnevno, tjedno ili mjesečno.

Kako je već napomenuto, Husky je jedinstvena platforma za monetiziranje podataka. Pristup primijenjen u procesu monetizacije podataka ide u dva smjera:

- interna monetizacija koja se postiže pametnom optimizacijom internih resursa, nakon promatranja ponašanja klijenata, tehnologije i usluga koje klijenti koriste. Odnosno, optimizacija poslovanja korištenjem podataka za „in-house" analize i informiranije donošenje odluka, kao i za poboljšanje prodajnih i marketinških aktivnosti u pravo vrijeme, i
- eksterna monetizacija, odnosno monetizacija u odnosu na vanjske klijente, pružajući im odgovarajuće podatke koje mogu koristiti za poboljšanje svog poslovanja.

Smjerovi monetizacije omogućeni su kroz slojeve aplikacije pomoću odvojenih modula:

1. Praćenje trendova i sociodemografskih analiza kretanja

Husky omogućuje analizu kretanja i najčešćih lokacija klijenata po određenim sociodemografskim parametrima. Konkretno, lokacija podrazumijeva točnu adresu na kojoj klijent banke koristi bankomat ili POS uređaj za kartičnu transakciju. Uz pomoć takvih podataka mogu se definirati trendovi u lokacijama korištenja bankomata i POS-uređaja i kretanjima po pojedinim grupama klijenata i sociodemografskim skupinama. To može biti vrlo korisno u smislu kreiranja personaliziranih ponuda čime se mogu unaprijediti marketinške i prodajne aktivnosti. Same lokacije mogu se vizualno prikazati u obliku heap map-a i drugih oblika izvještaja, što omogućuje bolje razumijevanje lokacija koje su najučestalije za klijente. Na temelju toga mogu se donijeti odluke o pozicioniranju bankomata, dodatne mreže POS uređaja, pa čak i poslovnica banke.

2. Analiza socijalnih mreža (ASM)

Ova analiza obuvaća praćenje povezanosti klijenata banke kroz njihove međusobne transakcije u svrhu detektiranja međusobnih odnosa i praćenja ponašanja klijenata. Uočavaju se transakcije koje su učestale između pojedinih klijenata banke, te se u tom slučaju zaključuje na koji način su oni povezani, kao supružnici, poslovni partneri, obitelj, prijatelji, odnosno osobe koje imaju određeni međusobni odnos. Kroz analizu prikupljenih podataka omogućuje se detektiranje lidera i sljedbenika (koji klijenti prvi usvajaju neke proizvode i

usluge, a koji ih slijede), te pomaže u detektiranju i prevenciji odljeva klijenata. Primjerice, ukoliko banka ima više klijenata iz jednog kućanstva (npr. 4) i iz nekog razloga dva klijenta iz tog kućanstva zatvore tekuće račune, veća je vjerojatnost da će i ostali članovi tog kućanstva zatvoriti svoje tekuće račune. Kad se ovakva situacija detektira na vrijeme, moguće je ponuditi dodatne pogodnosti članovima kućanstva koji su još uvijek klijenti banke, kako bi se spriječio i njihov odljev. Dodatno, kroz analizu socijalnih mreža može se pratiti i širenje pojedinih proizvoda i/ili usluga kroz mrežu klijenata, što može unaprijediti sami proizvod ili uslugu ali i olakšati definiranje skupina i/ili pojedinaca kojima se treba pristupiti s ponudom za konkretni proizvod ili uslugu.

3. Praćenje aktivnosti na web-u

Ovom metodom prati se interes klijenta na određenim web servisima, kao i podnošenje zahtjeva i druge aktivnosti klijenta na web servisima (npr. internet bankarstvo). Kroz praćenje aktivnosti na određenim internet lokacijama, omogućuje se kvalitetniji prodajni i marketinški nastup, s personaliziranim i individualno prilagođenim ponudama i drugim aktivnostima, što povećava interes ali i konverziju od strane klijenta. Ovo se postiže na način da se aktivnosti klijenata nakon praćenja prepoznaju, te se na temelju tih aktivnosti personalizirane ponude sastavljaju upravo kako bi odgovarale aktivnostima koje je klijent poduzeo. Primjerice ako je klijent istraživao uvjete za stambeni kredit ali nije kontaktirao banku, već je samostalno istraživao o tome na internet stranicama banke (kreditni kalkulator, kamatne stope, itd.), može mu se ponuditi personalizirana ponuda za kredit, prilagođena njegovim trenutnim primanjima i situaciji. Dodatno, ovim putem mogu se definirati mjesta na internet lokacijama na kojima klijenti nailaze na probleme prilikom korištenja internet usluga (korak na kojem se neki online zahtjev prekine i ne završi, ne dolazi do konverzije prilikom npr. online zahtjeva za gotovinskim kreditom), te se isti mogu ukloniti, čime se poboljšava kvaliteta usluge i percepcija klijenta o istoj.

4. Sentiment analiza ciljanog dijela web-a

Kroz sentiment analizu ciljanog dijela web-a može se postići bolje vrednovanje pojedinih proizvoda ili usluga, te se može postići i unaprjeđenje proizvoda na temelju obavljene analize prikupljenih podataka. Konkretno, može se pratiti što korisnici društvenih mreža i drugih javno dostupnih izvora podataka pišu o pojedinoj usluzi, proizvodu ili instituciji općenito. Definiranjem određenih ključnih riječi koje sustav traži na takvim mjestima može se zaključiti radi li se o generalno pozitivnom ili negativnom stavu i na temelju toga zaključiti u kojem

smjeru usmjeriti daljnji razvoj proizvoda ili usluge. Dodatno, uz integraciju s reklamacijskim centrom i uz ASM (Analiza socijalnih mreža), nudi se široki uvid u mogući odljev klijenata i prevencija istog, te daje mogućnost generiranja personaliziranih, pravovremenih ponuda prema pojedinim klijentima ili grupama klijenata obzirom na rezultate provedene sentiment analize. Primjerice sentiment analiza može dovesti do zaključka da gotovinske kredite više koriste osobe srednje životne dobi, iz određenih regija. Sukladno tim informacijama upravo se taj proizvod može još više kanalizirati prema upravo tim grupama ljudi, a kako bi se postigao maksimalni tržišni učinak u smislu prihvaćanja tog proizvoda od strane klijenata.

5. Praćenje aktivnosti u stvarnom vremenu ili u duljem vremenskom periodu

Praćenje aktivnosti klijenata u stvarnom vremenu omogućuje kreiranje ponuda temeljenih na informacijama dostupnim u realnom vremenu, koje bi mogle biti od značaja za klijenta upravo u trenutku u kojem se dogodio okidač za generiranje ponude. Konkretno, klijent u određenom trenutku u stvarnom vremenu učini određenu aktivnost koja predstavlja okidač koji pokreće predefinirani set aktivnosti koje dovode do generiranja ponude za tog klijenta. Na primjer, ukoliko se primijeti povećana potražnja za gotovinskim kreditima od strane mlađe populacije sa srednjom razinom prihoda, može se targetirati čitavu tu grupu klijenata s agresivnijom kampanjom koja promovira upravo taj tip proizvoda, kako bi se maksimalno iskoristio trenutni trend koji je prisutan i prepoznat na tržištu.

6. Ciljana komunikacija na temelju trenutne lokacije

Kroz praćenje aktivnosti klijenta u stvarnom vremenu, moguće je odrediti točnu trenutnu lokaciju pojedinog klijenta (na primjer podizanje gotovine s bankomata ili provlačenje kartice kroz pojedini PoS uređaj), te u tom trenutku pokrenuti predefiniranu i automatiziranu aktivnost kroz neki od komunikacijskih kanala (ATM, SMS, Email) u svrhu promotivnih, marketinških ili prodajnih aktivnosti. Konkretno, ukoliko klijent pokaže aktivnost na bankomatu u određenom trgovačkom centru, a u tom istom centru se nalazi maloprodavač koji u tom trenutku ima promotivnu akciju na svoje proizvode ukoliko se za proizvode plati karticom banke klijenta, tad klijent dobije obavijest o toj akciji, čime ga se pokušava potaknuti na kupnju.

7. Praćenje aktivnosti u sigurnosne svrhe

U ovom kontekstu prate se aktivnosti klijenta te se uočavaju određeni uzorci ponašanja. Ukoliko se na bilo kojem promatranom sustavu (ATM, PoS, mobilno bankarstvo, Internet

bankarstvo) dogodi odstupanje od utvrđenih standarda i uzoraka ponašanja klijenta, aktivira se upozorenje te se pokreću određene predefinirane preventivne i sigurnosne aktivnosti, ovisno o tipu odstupanja. Primjerice, prema utvrđenom uzorku klijent dosad niti jednom nije krivo unio PIN na bankomatu ili u mBanking aplikaciji. Ukoliko se dogodi da se za to korisničko ime ili tu karticu klijenta dogodi situacija da je PIN krivo unešen, pokreće se predefinirani set sigurnosnih aktivnosti. U ovom kontekstu prate se i lokacije klijenta na kojima se događaju aktivnosti, te se i one definiraju na standardne i nestandardne. Ponovno, u slučaju pojave određene aktivnosti od strane klijenta na nestandardnoj lokaciji aktiviraju se određene predefinirane aktivnosti, u svrhu validacije aktivnosti od strane klijenta.

8. Praćenje transakcija

Ova metoda omogućava praćenje transakcija klijenta te uočavanje određenih nepravilnosti na temelju predefiniranih pravila koja banka odredi prije početka promatranja aktivnosti svojih klijenata. Pravovremeno uočavanje ovog tipa nepravilnosti uvelike pomaže u borbi protiv tzv. „first party fraud-a“, koje može nanijeti veliku materijalnu štetu financijskoj instituciji. Ipak, kroz pravilno definiranje ovog tipa nepravilnosti, i to u stvarnom vremenu, uvelike se smanjuje mogućnost da se opisana materijalna šteta i ostvari.

3.4 Prijetnje funkcioniranju Big Data tehnologiji

Prema ENISA Threat Landscape 2013[40], prijetnja je "netko ili nešto sa sposobnostima, jasnom namjerom iskazati prijetnju i poviješću činjenja istog“. Big Data vlasnici podataka moraju biti svjesni koje prijetnje dolaze iz koje skupine prijetnji. U ovom poglavlju neće se otkrivati izvori prijetnje već će se navesti definirani izvori prijetnje kroz publikacije ENISA Threat Landscape 2013.

3.4.1. Izvori prijetnje

Korporacije

[40] https://www.enisa.europa.eu/publications/enisa-threat-landscape-2013-overview-of-current-and-emerging-cyber-threats

Odnosi se na organizacije, poduzeća koja usvajaju i / ili se bave neprimjerenom napadačkom taktikom. U ovom se kontekstu korporacije smatraju neprijateljskim izvorima prijetnji i njihova motivacija je izgraditi konkurentsku prednost u odnosu na konkurente koji im predstavljaju glavnu metu. Ovisno o njihovoj veličini, korporacije obično posjeduju značajne mogućnosti, u rasponu od tehnologije do ljudske inženjerske inteligencije, posebno u njihovom području djelovanja.

Cyber kriminalci

Oni su po prirodi neprijateljski raspoloženi. Štoviše, njihova motivacija obično je financijska korist i njihova razina vještina danas je prilično visoka. Internetski kriminalci mogu se organizirati na lokalnoj, nacionalnoj razini ili čak međunarodnoj razini.

Cyber teroristi

Proširili su svoje aktivnosti i uključili se u cyber-napade. Njihova motivacija može biti politička ili vjerska, a njihova sposobnost varira od nižih do izuzetno visokih. Preferirane mete cyber terorista uglavnom su kritična infrastruktura (npr. javno zdravstvo, proizvodnja energije, telekomunikacije) jer njihovo ne funkcioniranje uzrokuju ozbiljan utjecaj na društvo i vlast. Valja napomenuti da u javnim publikacijama, pojam cyber terorista i dalje je nedefiniran.

Script kiddies

Oni su nekvalificirani pojedinci koji koriste skripte ili programe koje su drugi razvili za napad računalnim sustavima i mrežama te web-lokacijama.

Internetski društveni hakeri (hacktivisti)

Oni su politički i društveno motivirani pojedinci koje koriste računalne sustave kako bi prosvjedovali i promicali svoje stavove. Njihovi su tipični ciljevi web stranice visokog profila, korporacije, obavještajne agencije i vojne institucije.

Zaposlenici

Odnose se na zaposlenike, operativno osoblje ili sigurnosne stručnjake unutar kompanije. Oni mogu imaju insajderski pristup resursima tvrtke. Najčešće su to rastreseni i nezadovoljni zaposlenici. Ovakva vrsta prijetnji posjeduje značajnu količinu znanja koja im omogućuje učinkovito izvršenje napada na imovinu koja pripada organizaciji.

Vlasti

Imaju mogućnost za cyber napade visoke razine i koriste ih na svojim protivnicima. U posljednje se vrijeme sve češće spominju cyber napadi raznih vladi. Razina sofisticiranosti zlonamjernih softvera potvrđuje da vlasti imaju veliku količinu resursa i visoku razinu vještina i znanja.

4. Primjena alata Big Date u procesu segmentacije kupaca

U svrhu prikazivanja važnosti velikih podataka i procesa strojnog učenja u današnjem poslovnom svijetu provest ćemo bazično empirijsko istraživanje koristeći upravo alate strojnog učenja i velikih podataka. Koristeći programski jezik R obradit ćemo prikupljene podatke i kroz njihovu obradu prikazati jednu od esencijalnih primjena strojnog učenja u poslovnoj sferi - segmentaciju kupaca. [41]Bit segmentacije kupaca je razdioba baze podataka o kupcima u razlučive i homogene grupe s ciljem razvijanja diferenciranih poslovnih strategija baziranih na karakteristikama novodobivenih grupa. Da bi došli do takvih homogenih grupa potrebno je izvršiti određeni broj koraka koji se kreću od prikupljanja podataka, preko programiranja sve do tumačenja dobivenih grafičkih i numeričkih pokazatelja.

Prikupljanje informacija u današnjem svijetu velikih podataka, uz poznavanje osnovnih načela potrebnih programskih jezika, jednostavno je i brzo. Ono se proteže od prikupljanja podatak iz "open-source" softwarea kao što je Hadoop sve do direktnog pristupa podatcima putem API-a (aplikacijskog programskog sučelja) ili pristupa podatcima kroz SQL baze. Sami podatci iz različitih izvora mogu imati i različite oblike poput csv, json, xml, html. Zbog nemogućnosti pristupa privatnim bazama poslovnih subjekata koji bi se uklapali u okvire ovog rada za bazu podataka ćemo koristiti ograničene podatke istraživanje koje je proveo odjel razvoja OTP banke dopustili su nam korištenje korigirane baze u svrhu ovog rada. Najbrži ali i najteži put prikupljanja podataka za ovakvo istraživanje je putem ankete. Valja napomenuti da iole ozbiljni poslovni subjekti podatke skupljaju svakodnevno kroz same procese svog poslovanja a nikako samo putem provođenja anketa.

4.1. Analogija provođenja procesa segmentacije kupaca

Osim prikupljanja podataka kompletan proces segmentacije kupaca odvijat će se sljedećim mehanizmom:

- prikupljanje podataka
- uvoz podataka u R
- programiranje
- vizualizacija podataka po spolu

[41] Tsiptsis, K. K., & Chorianopoulos, A. (2011). Data mining techniques in CRM: inside customer segmentation. John Wiley & Sons. stranica 4.

- vizualizacija podataka po godinama starosti
- analiza godišnjeg prihoda kupaca
- analiza "Spending scorea" kupaca
- određivanje optimalnog broja klastera
- vizualizacija rezultata klasteriranja u procesu segmentacije
- zaključak segmentacije

Proces prikupljanja podataka u našem slučaju bit će simulacija cijelog procesa prikupljanja i kreiranja baze podataka. Naime, proces prikupljanja podataka odvija se kroz svakodnevno prikupljanje sveukupnog obujma raznih vrsta podataka stvorenih kroz same poslovne procese unutar tvrtke. Pod tim podrazumjevamo podatke iz prodajnog prometa, podatke o kupcima prikupljene putem kyc-a (know your client), podatke koje potrošači dobrovoljno predaju putem ispunjenih anketa, podatke internog kartičnog prometa kao što je npr. K plus card, podatke o stanju broja posjetitelja fizičkih trgovina dobivenih iz rfid tagova i slično. Kako u našem slučaju nije moguće ostvariti pristup takvim podatcima iskoristiti ćemo bazu istraživanje OTP banke na uzorku od 200 ispitanika. Filtriranu bazu podatak iz ovog ispitivanja su nam dali djelatnici banke sa odjela razvoja, uz dopuštenje za korištenje u ovom radu. Podatci će biti raspoređeni unutar pet sljedećih kategorija: Potrošačev ID, Spol, Godine, Godišnji prihod, „Spending score“ (1-100). Baza podataka sastojat će se od CSV (comma-separated values) podataka.

Sljedeći korak je uvoz podataka u odabrani software za programiranje podataka. Kako smo za bazu podataka stvorili novonastali skup od 100 subjekata i pridali im pet kategorija potrebno je odabrati software za obradu istih. Za tu svrhu odabrali smo R, programski jezik nastao na temeljima S jezika razvijenog 1976. godine [42] s ciljem stvaranja novog programerskog jezika za statistička istraživanja. Glavna prednost R-a pred drugim jezicima je visoka razina prihvaćenosti programa kao open source alata koji nudi veliki broj ekstenzija tj. „paketa“, posebno onih iz domene statističkih istraživanja.

S obzirom da se naša baza podataka sastoji od csv (comma separated values) vrijednost koje su visoko kompatibilne s R jezikom sam proces uvoza podataka je jako jednostavan. Jedan od najjednostavnijih načina je unošenjem naredbe read.csv(„“)u konzolu.

[42] Peng, R. D. (2016). R programming for data science. Leanpub.

Nakon uvoza podataka u R software može se provesti obrada podataka putem pisanja valjanih naredbi. Naredbe se razvijaju u prozoru nazvanom „R script“ ili R skripta dok se njihovo izvršavanje odvija u „Consolu“ ili konzoli. Za razvijanje naredbi s ciljem provođenja segmentacije kupaca potrebno je određeno poznavanje razvijanja funkcija unutar R jezika ali i snažno poznavanje statističkih pojmova i teorije. Sam proces „programiranja“ podataka dovest će nas do analitičkih i vizualnih podataka potrebnih za vizualizaciju podataka po spolu, vizualizaciju podataka po godinama starosti, analizu godišnjeg prihoda kupaca, analizu „Spending scorea“ kupaca, određivanje optimalnog broja klastera i vizualizaciju rezultata klasteriranja u procesu segmentacije.

Vizualizacija podataka po spolu prikazat će se korz upotrebu „pita“ grafikona te jednostavnih analitičkih podataka koji nam govore koliko elemenata naše baze pripada ženskom a koliko muškom spolu.

Vizualizacija podataka po godinama starosti prikazat će se putem histograma u kojem je vidljivo koliko elemenata pripada kojem starosnom razredu.

Analiza godišnjeg prihoda vizualno će se prikazati također putem histograma. Distribucija podataka iz ove kategorije bit će prikazana Kernelovom krivuljom (dijagramom) gustoće.

Analiza „Spending scorea“ kupaca specifična je kategorija ovog istraživanja. Naime, kako su podatci prikupljeni anketnim putem oni nikako ne mogu sadržavati vjerodostojne podatke koji se odose na „koeficijent potrošnje“. Definiranje samog „Spending scorea“ ili „koeficijenta (ili procjene) potrošnje“ kao pojam nije naišlo na konsenzus u literaturi. Rezultat toga je različito definiranje te raziličiti načini računanja ovog pojma od strane različitih subjekata koji ga koriste u svojim izračunima. Jedna od najčešćih tehnika korištena pri izračunu „Spending Scorea“ je RFM [43] tehnika i njene postojeće varijacije. Kratica RFM-a dolazi od triju varijabli koje se uzimaju u obzir pri izračunavanju konačnog RFM koeficijenta, recency–frequency–monetary. Varijabla „Recency“ odnosi se na posljednji vremenski trag, pečat, u kojem je kupac obavljao kupnju, „Frequency“ prikazuje koliko često kupac obavlja kupnju te varijabla „Monetary“ koja prikazuje omjer najviše vrijednosti kupnje s omjerom kupnje svakog pojedinog subjekta. Iako ne postoji konsenzus u literaturi može se poopćiti i zaključiti da je „Spending score“ rezultat potrošačkih navika kupaca a definiran je odnosom najviše razine potrošnje i potrošnje kupca na kojeg se koeficijent odnosi. Problem ovog istraživanja je

[43] Hughes, A. M. (1994). Strategic database marketing. Chicago: Probus Publishing Company

nemogućnost vlastitog izračunavanja „Spending scorea“ Kako nismo u mogućnosti sami izračunati „spending score“ anketiranih osoba, iz razloga što ne poznajemo njihove potrošačke sklonosti niti općenito niti prema određenim subjektima, zamolili smo anketirane da sami sebi pridaju određeni koeficijent. Iz toga proizlazi da je „spending score“ iz naše baze podataka produkt iskrene procjene anketiranih o odnosu njihove osobne potrošnje u odnosu na najviše razine potrošnje subjekata koji ih svakodnevno okružuju.

Sljedeći korak analize je određivanje optimalnog broja klastera. Tijekom procesa klasteriranja potrebno je odrediti optimalan broj klastera za određenu bazu podataka. Optimalnim brojem klastera izvlačimo najvišu razinu „korisnosti“ istraživanja i postižemo više razine analitičke i grafičke pripadnosti elemenata (kupaca) određenom klasteru. Postoje tri popularne metode za određivanje optimalnog broja klastera

- Metoda lakta
- Metoda siluete
- „Gap“ statistika [44]

Nakon provedenog procesa klasteriranja podataka i određivanja optimalnog broja klastera potrebno je analizirati sam „karakter“ klastera. Za primjer, možemo pretpostaviti da će se određeni subjekti istraživanja (kupci) grupirati u klaster za koji su karakteristična visoka primanja ali i visoki godišnji troškovi (prikazani kroz „spending score“). Isto tako, možemo postaviti teorijsku pretpostavku da postoji vjerojatnost da se formira klaster za koji će biti karakteristična niska primanja ali visoki godišnji troškovi.

Upravo nam obrada podataka u R-u provedena kroz sve prethodne korake pruža empirijsku analizu „karaktera“ klastera ali i same preraspodjele kupaca iz naše baze podataka unutar određenih klastera. Na temelju te analize moguće je donijeti empirijski utemeljene zaključke o postojanju određenog broja značajnih klastera, o samom „karakteru“ klastera i o pripadnosti određenog subjekta (kupca) određenom klasteru. Sve te informacije pružaju jake temelje za razvoj visoko specijaliziranog i diversificiranog pojedinačnog pristupa svakom kupcu s ciljem minimiziranja troškova, maksimiziranja korisnosti i stvaranja što profitabilnijeg modela poslovanja.

[44] Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2), 411-423.

4.2. Provođenja procesa segmentacije kupaca

4.2.1. Prikupljanje podataka

Proces prikupljanja podataka u našem slučaju bit će simulacija cijelog procesa prikupljanja i kreiranja baze podataka. Sam proces prikupljanja podataka u radu tvrtki odvija se kroz svakodnevno prikupljanje sveukupnog obujma raznih vrsta podataka stvorenih kroz same poslovne procese unutar tvrtke. Pod tim podrazumijevamo podatke iz prodajnog prometa, podatke o kupcima prikupljene putem kyc-a (know your client), podatke koje potrošači dobrovoljno predaju putem ispunjenih anketa, podatke internog kartičnog prometa kao što je npr. K plus card, podatke o stanju broja posjetitelja fizičkih trgovina dobivenih iz rfid tagova i slično. Kako u našem slučaju nije moguće ostvariti pristup takvim podatcima kroz anketiranje 200 ispitanika stvorit ćemo bazu od 200 subjekata i pet kategorija. Podatci će biti raspoređeni unutar pet sljedećih kategorija: Potrošačev ID, Spol, Godine, Godišnji prihod, „Spending score“ (1-100). Baza podataka sastojat će se od CSV (comma-separated values) podataka.

4.2.2. Uvoz podataka u R software

Prvi korak u provođenju bilo kakve analize podataka u R programskom jeziku je učitavanje podataka te bazičan pregled strukture samih podataka. Najjednostavniji put za učitavanje podataka je putem naredbe read.csv data.

Naša baza podataka nalazi se na radnoj površini („desktop“) pod nazivom „Baza_podataka.csv“ i kao takvu možemo je uvesti u bazu R alata. Međutim, radi jednostavnijeg pregleda i manipuliranja naredbama i podatcima bazu ćemo u R uvesti kao zasebnu varijablu pod nazivom „customer_data“ i to sljedećom naredbom:

```
customer_data=read.csv(„/home/desktop/Baza_podataka.csv“)
```

Nakon učitavanja baze podataka u bazu R-a moguće je izvršiti pregled strukture podatak kao i pregled temeljnih statističkih pokazatelja. Provođenjem funkcije:

```
str(customer_data)
```

dobivamo sljedeći grafički prikaz.

```
## 'data.frame':    200 obs. of  5 variables:
##  $ CustomerID            : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Gender                : Factor w/ 2 levels "Female","Male": 2 2 1 1 1 1 1 1 2 1
...
##  $ Age                   : int  19 21 20 23 31 22 35 23 64 30 ...
##  $ Annual.Income..k..    : int  15 15 16 16 17 17 18 18 19 19 ...
##  $ Spending.Score..1.100.: int  39 81 6 77 40 76 6 94 3 72 ...
```

Slika 13. Pregled strukture uvezene baze podataka

Iz Slike 1. vidljivo je da se baza sastoji od 200 promatranja svake od 5 varijabli („Potrošačev ID“, „Spol“, „Godine“, „Godišnji prihod“, „Spending score“).

Pregled temeljnih statističkih pokazatelja možemo pregledati provođenjem naredbe summary. Valja napomenuti da provođenje ove operacije pri analizi varijabli kao što su „Potrošačev ID“ ili „Spol“ neće dovesti do nikakvih valjanih zaključaka. Razlog tomu je što kod prve varijable imamo niz brojeva od 1-200 a kod druge popis tvoren od dvije znamenke, 1 i 2, gdje 1 označava osobe koje su se definirale kao žene a broj 2 osobe koje su se definirale kao muškarci. Provođenje naredbe summary za varijablu „Godine“ provodimo na sljedeći način.

```
summary(customer_data$Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   28.75   36.00   38.85   49.00   70.00
```

Rezultati naredbe Summary vidljivi su na Slici 13. Vidimo da najmlađi ispitanik iz baze podataka ima 18 godina dok najstariji ima 70, medijan je vrijednost 36 dok srednja vrijednost iznosi 38,85.

Učitavanjem podataka u R dobili smo strukturiranu bazu podataka, analizirali njenu strukturu i očitali bazične statističke pokazatelje. Provođenjem ovih naredbi dobili smo statističke temelje i mogućnost provođenja daljnje analize.

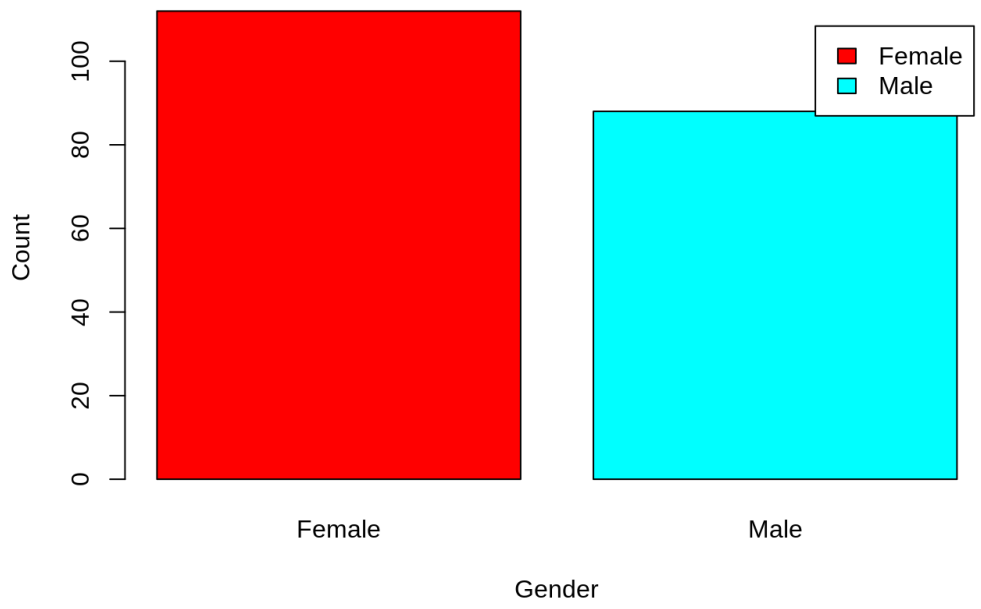
4.2.3. Vizualizacija podataka po spolu

Kako provođenjem naredbe „summary“ ne bi došli do zadovoljavajućih vizualnih i analitičkih podataka potrebno je provesti dodatnu analizu. Cilj ove analize je utvrđivanje postojanja „ravnopravne“ strukture unutar baze podataka a sve u svrhu dobivanja što korisnijih podataka za poboljšanje poslovanja. Za primjer, ako se poslovanje tvrtke „X“ bazira na strukturi klijenata od koji se njih 93% izjašnjava kao žene tada i sama baza podataka po svojoj strukturi mora sadržavati približno jednak omjer muških i ženskih klijenata tj. 7% i 93%. Ako, u ovom fiktivnom slučaju, baza podataka sadrži 85% muškaraca rezultati koji će se dobiti provedenim istraživanjima neće davati realnu sliku sveopćeg poslovanja tvrtke već će davati sliku klastera unutar dominantno muške strukture klijenata. Ovaj problem problem je stvaranja vjerodostojnog uzorka iz šire populacije prati sva statistička istraživanja. Upravo je to jedna od najvećih prednosti Big Date, analiziranje potpune populacije a ne samo određenog uzorka. S obzirom da je naš uzorak relativno mali i sadrži podatke „prilagođene“ (ili „izmišljene“) našem cilju istraživanja potrebno je stvoriti što ujednačeniju bazu ženskih i muških subjekata. Tu bazu vizualno ćemo predočiti kroz dva dijagrama putem dvije naredbe. Prije toga stvoriti ćemo i zasebnu varijablu „a“ s ciljem lakšeg programiranja podataka.

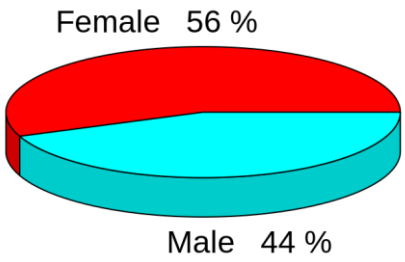
```
a=table(customer_data$Gender)
```

```
barplot(a,main="StupcaniGraf",
      ylab="Count",
      xlab="Gender",
      col=rainbow(2),
      legend=rownames(a))
```

```
pct=round(a/sum(a)*100)
lbs=paste(c("Female","Male")," ",pct,"%",sep=" ")
library(plotrix)
pie3D(a,labels=lbs,
  main="PitaGraf")
```



Slika 14. Stupčasti grafikon – apsolutni udio po spolovima



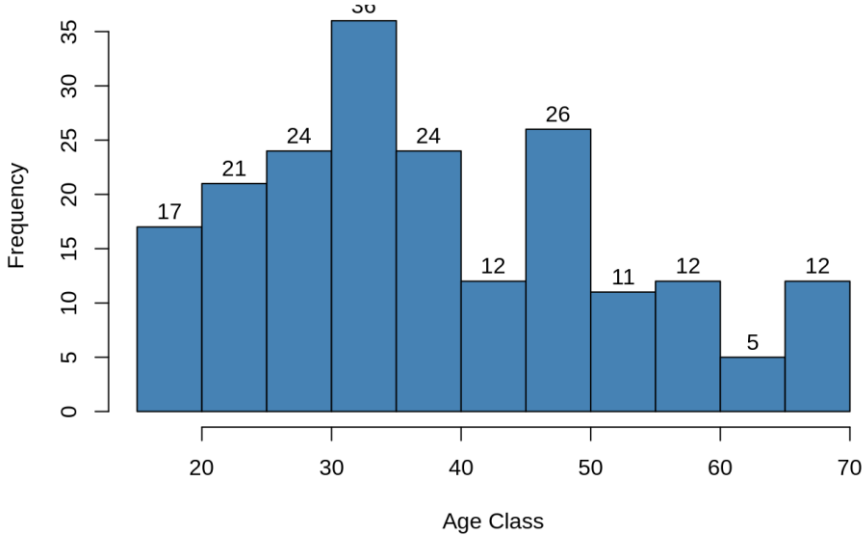
Slika 15. Pita grafikon – relativni udio po spolovima

Iz dobivenih grafova Slika 14. i Slika 15. vidljivo je da je struktura zadovoljavajuće ujednačena i da je udio žena 56% a muškaraca 44%.

4.2.4. Vizualizacija podataka po godinama starosti

Udio ženskih ili muških kupaca temeljan je ali često manje bitan podatak u stvaranju poslovnih strategija. Neusporedivo veću korisnost od strukture po spolu daje struktura po godinama. Najkorisniji vizualni prikaz strukture podataka varijable „Godine“ je prikaz iste putem histograma.

```
hist(customer_data$Age,
   col="blue",
   main="Histogram Godine",
   xlab="Age Class",
   ylab="Frequency",
   labels=TRUE)
```

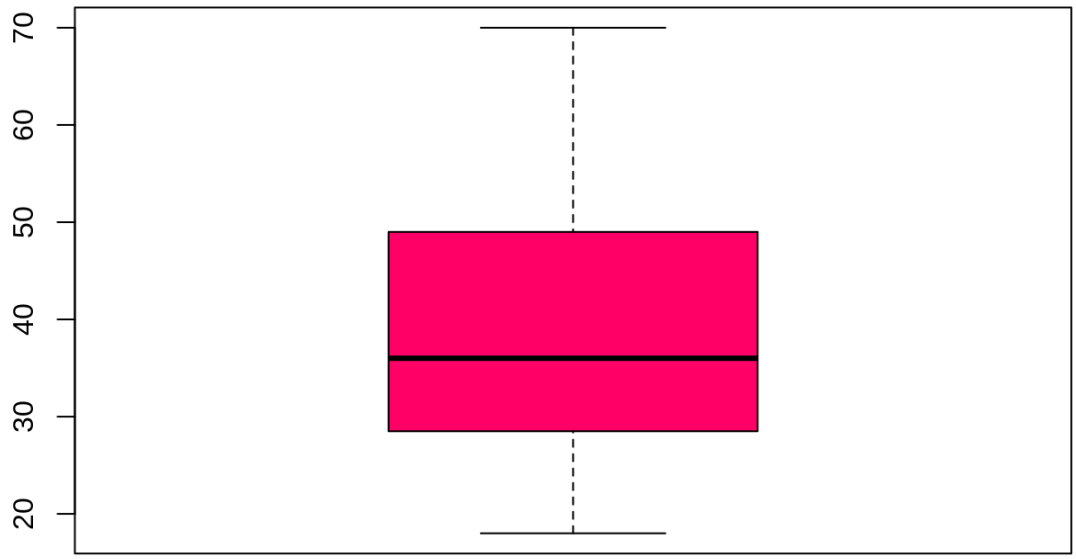


Slika 16. Histogram – varijabla „Godine“

Histogramom sa Slike 16. prikazana je jasna struktura ispitanika strukturirana po razredima u intervalu od 5 godina. Tako možemo uočiti da najveći broj ispitanika, njih 36, pripada intervalu od 30-35 godina. Najmanje ih pripada intervalu od 60-65 godina.

Već ranije smo prikazali određene statističke pokazatelje za varijablu godine. Ti pokazatelji daju nam sliku distribucije promatranih podataka. Iste podatke moguće je i vizualno prikazati pute box grafikona. On nam pokazuje minimalnu i maksimalnu vrijednost, prvi i treći kvartil te vrijednost medijana. Prikazujemo ga jednostavnom naredbom boxplot().

```
boxplot(customer_data$Age)
```



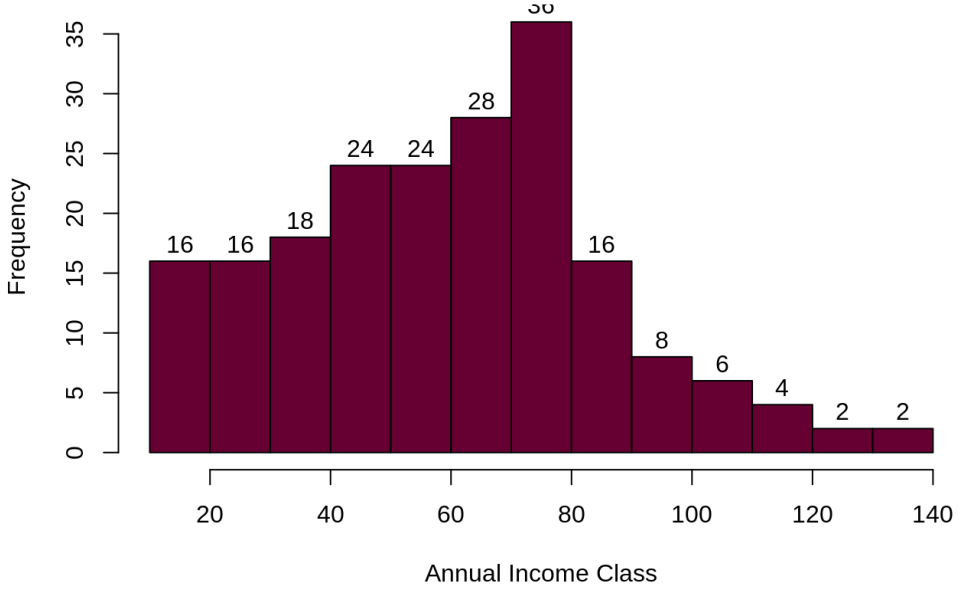
Slika 17. Box grafikon – varijabla „Godine“

4.2.5. Analiza godišnjeg prihoda kupaca

S gledišta marketinških analiza i kreiranja marketinških strategija vizualizacija podataka po starosti i spolu bitan je čimbenik u prilagođavanju prodaje proizvoda segmentu koji je kroz iste analize identificiran kao strateški najvažniji kontingent. Problem prodajnih strategija temeljenih samo na analizi spola i starosti kupaca je taj što nerijetko daju rezultate koji se ne podudaraju sa samim ciljem strategije. Izvor tog problema često se nalazi u tome što najzastupljeniji segment, npr. žene u dobi od 30-35 godina, nije segment koji donosi najviše prihode tvrtci koja provodi analize. S ciljem dobivanja potpune slike o aktivnosti, udjelu i potrošnji kupaca potrebno je provesti analizu prihoda kupaca te analizu njihovog „spending scorea“.

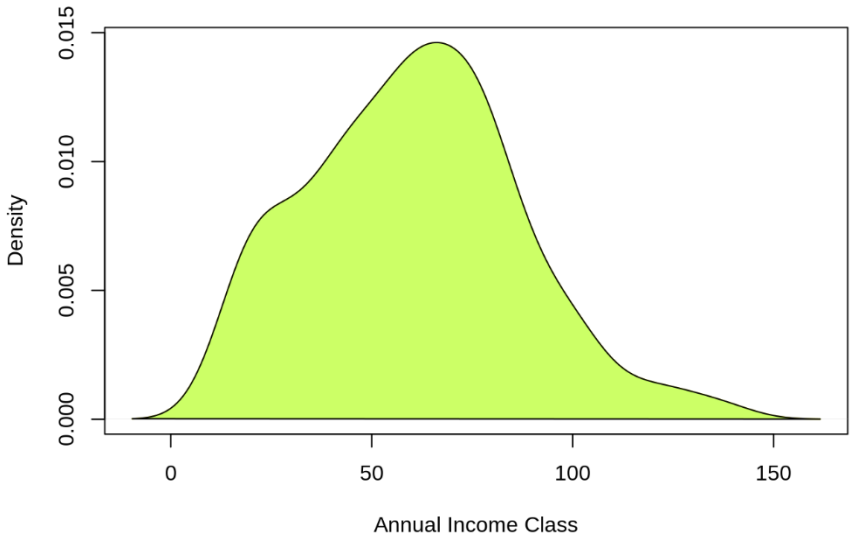
Distribuciju potošača po varijabli „Godišnji prihod“ grafički ćemo prikazati putem histograma.

```
plot(density(customer_data$Annual.Income..k..),
   col="yellow",
   xlab="Annual Income Class",
   ylab="Density")
polygon(density(customer_data$Annual.Income..k..),
      col="#ccff66")
```



Slika 18. Histogram – varijabla „Godišnji prihod“

Iz histograma je vidljivo da najveći broj ljudi zarađuje od 70-80 tisuća kuna godišnje a najmanji broj ih godišnje uprivredi 120-140 tisuća kuna. Analiziranjem dodatnih statističkih pokazatelja vidljivo je da u strukturi baze podataka najmanja godišnja plaća iznosi 15 tisuća kuna dok najviša iznosi 137 tisuća. Prosječna godišnja plaća cijele populacije je 60,56 tisuća kuna. Zanimljivo je i provesti analizu podataka u svrhu prikazivanja Kernelove krivulje gustoće s ciljem analize distribucije podataka o godišnjim prihodima. Iz Slike 19. na kojoj je prikazana Kernelova krivulja gustoće moguće je uočiti da podatci za „Godišnje prihode“ imaju oblik normalne distribucije.



Slika 19. - Kernelova krivulja gustoće – varijabla „Godišnji prihod“

4.2.6. Analiza „spending scorea“ kupaca

Provođenjem naredbe summary prikazat ćemo osnovne statističke pokazatelje za varijablu „Spending score“.

```
summary(customer_data$Spending.Score..1.100.)
```

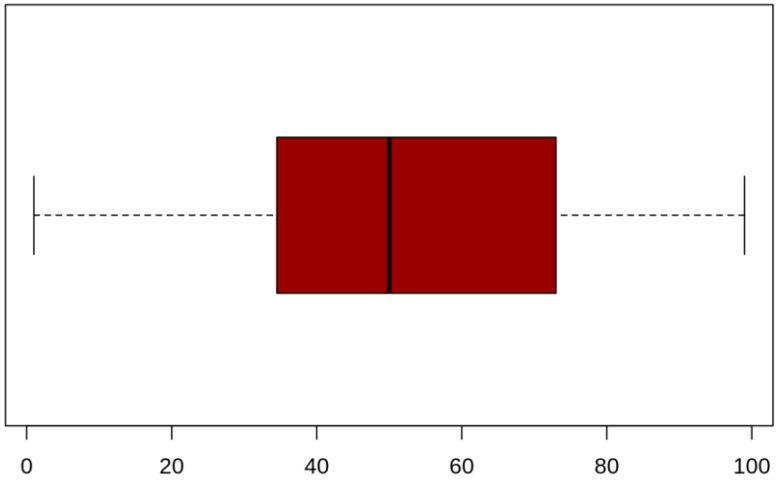
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   34.75   50.00   50.20   73.00   99.00
```

Slika 20. Statistički pokazatelji za varijablu „Spending score“

Iz dobivenih podataka vidljivo je da je minimalna vrijednost „Spending score“ varijable 1 a maksimalna 99. Srednja vrijednost varijable je 50.20 dok je medijan 50.

Iste podatke možemo pokazati grafički putem potonje (Slika 17.) spomenutog box grafikona.

```
boxplot(customer_data$Spending.Score..1.100.,
  horizontal=TRUE)
```

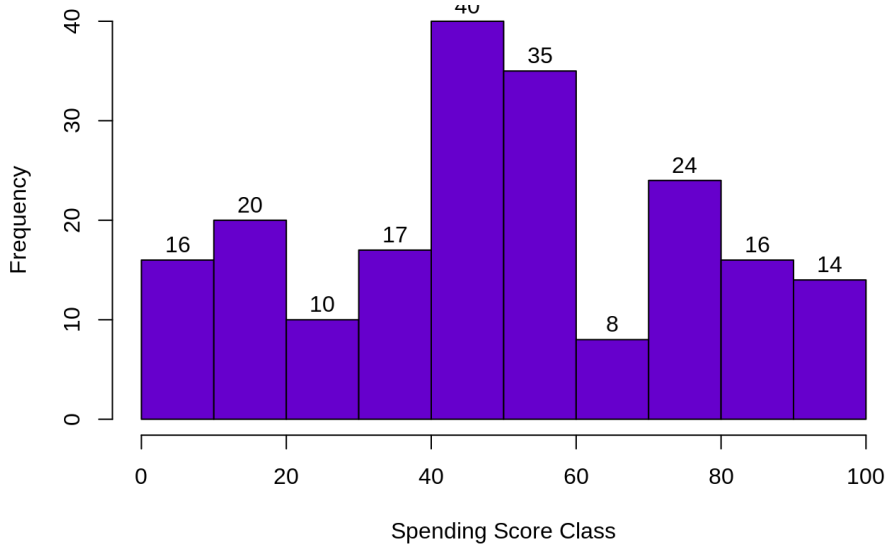


Slika 20. Box grafikon – varijabla „Spending score“

Iz slike 20. vidljiva je distibucija podataka od minimuma 1 do maksimuma 99, s prvim kvartalom u 34,75 te rećim kvartalom u 73.

Distribuciju varijable „Spending score“ prikazat ćemo i putem histograma.

```
hist(customer_data$Spending.Score..1.100.,
   main="Histogram Spending score",
   xlab="Spending Score Class",
   ylab="Frequency",
   col="#6600cc",
   labels=TRUE)
```



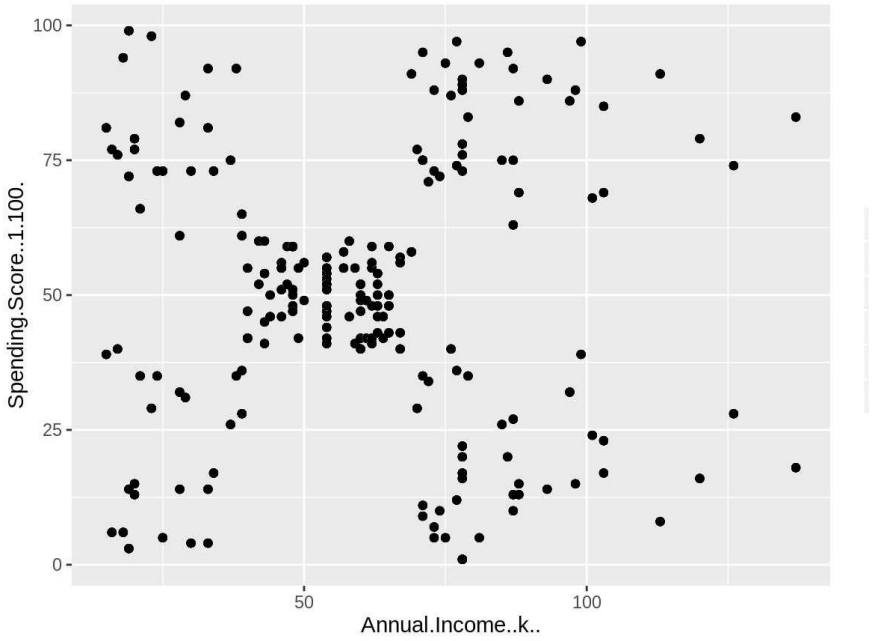
Slika 21. Histogram – varijabla „Spending score“

Iz histograma je vidljivo da najveći broj kupaca, njih 40, pripada grupi kojoj je dodijeljen „Spending score“ od 40 do 50. Čak 75 kupaca, 37,5%, ima dodijeljen „Spending score“ od 40 do 60.

Iz provedenih analiza već sada možemo uvidjeti koliki utjecaj strojno učenje i Big data mogu imati na poslovanje bilo kojeg poslovnog subjekta. Iz analiza pojedinačnih varijabli „Spol“, „Godine“, „Godišnji prihod“, „Spending score“ iz jednostavne CSV baze, s 200 promatranja za svaku varijablu, došli smo do ključnih podataka za poboljšanje poslovanja.

Udio žena u bazi podataka je 56% a muškaraca 44%. Najveći broj ispitanika, njih 36, pripada intervalu od 30-35 godina. Najmanje ih pripada intervalu od 60-65 godina. Najveći broj ljudi zarađuje od 70-80 tisuća kuna godišnje a najmanji broj ih godišnje uprivredi 120-140 tisuća kuna, najmanja godišnja plaća iznosi 15 tisuća kuna dok najviša iznosi 137 tisuća. Po „Spending scoreu“ 40 ispitanika pripada razredu sa spending scoreom od 40 do 50. Iako su

ovi podatci izuzetno korisni njihov najveći problem je njihova jednodimenzionalnost. R nam daje mogućnost da npr. za svakog subjekta iz „Spending score“ razreda od 40 do 50 provjerimo njegovu dob, spol i godišnje prihode ali takav proces provjere pojedinačnih podatak je prekompliciran i vremenski zahtjevan. Rješenje tom problemu je dodavanje dimenzija pri analizi podataka. Takva dvodimenzionalna analiza dala bi nam mnogo jasniji uvid u međuovisnost varijabli.



Slika 22. Međuovisnost varijabli „Godišnji prihod“ i „Spending score“.

Na slici 22. vidljiva je međuovisnost varijabli „Godišnji prihod“ i „Spending score“. Za svaku točku na predstavljenom grafikonu moguće je utvrditi spol i dob subjekta. Pažljivijim iščitavanjem grafikona mogu se uočiti određeni uzorci u grupiranju podataka. S ciljem povećanja dostupnosti proizvoda subjektima s višim „Spending scoreom“ potrebno je razlučiti uzorke, formirati ih u klastere i proučiti klastere koji su zanimljivi za daljnju analizu tj. klastere formirane u gornjem dijelu grafikona.

U ovom stadiju analize podataka potrebno je segmentirati subjekte u različite „pripadajuće grupe“. Sam proces segmentacije postaje relevantan za razumijevanje kupaca, za raspodjelu resursa i diversifikaciju proizvoda, kao i za razvoj novih pristupa tržištu[45]. Efektivna segmentacija kupaca postaje krucijalna za održivu strategiju proizvoda dok se menadžmentu

[45] (Palmer, R.A., Millier, P., 2004. Segmentation: identification, intuition, and implementation. Industrial Marketing Management 33, 779)

pruža slika je li njihov proizvod aktualan s potražnjom na tržištu ili ga je potrebno redizajnirati[46].

Određivanjem broja klastera omogućit će nam daljnje analize onih segmenata koji su bitni za formiranje poslovnih strategija.

4.2.7. Određivanje optimalnog broja klastera

Hijerarhijska analiza klastera je algoritam kojim se grupiraju što više slični uzorci u skupine zvane klasteri. Hijerarhijski klasteriranje se može izvesti na sirovim i strukturiranim podacima. Jednom kada se podaci uvedu u programe za obradu i učitaju se određene naredbe, računalo automatski izračunava matricu udaljenosti u pozadini. Proces klasteriranja podataka složen je i dugotrajan ali ga alati poput R-a čine dostupnim i manje kompliciranim.

U prvom koraku procesa odredi se početnih *k* točki ishodišta zamišljenih klastera te se nasumično i slučajnim odabirom odrede njihove koordinate na pripadajućem koordinatnom sustavu.

Na temelju nasumično poslaganih *k* točki formira se *k* klastera. Klasteri se formiraju na način da se svaka nova točka (promatranje) uvrsti (pripoji) u klaster čiji mu je centar najbliži. Na ovaj način se udaljenost koristi kao mjera sličnosti među objektima. Za pripajanje točki određenom *k* klasteru najčešće koristimo mjeru Euklidske udaljenosti ali mogu se koristiti i drugi tipovi udaljenosti kao što su Manhattan ili Minkowski. Ovim procesom formirali smo *k* klastera i njima pripadajuće točke. Međutim, kako su ishodišta klastera nasumično određena može se dogoditi scenarij u kojem su sve točke zbog svoje blizine pripojene jednom klasteru. [47]Ovaj problem rješava se neprestanim ponavljanjem radnje nasumičnog dodjeljivanja *k* točki ishodišta na koordinatni sustav sve dok se ne pojave dva slučaja u kojima isti klasteri sadrže iste točke.

Postoje i drugi načini formiranja klastera. Na primjer moguće je svakom formiranom klasteru, nakon prvog koraka, odrediti novi centar. Novi centar se određuje kao težište formiranog klastera. Njegove koordinate se dobivaju kao aritmetička sredina odgovarajućih koordinata članova klastera. Tada se ponavlja proces dodjele točaka novim *k* centrima. To dovodi do premještanja nekih objekata iz jednog klastera u drugi. Promjena sastava objekata u svakom

[46] Teichert, T., Shehu, E., & von Wartburg, I. (2008). Customer segmentation revisited: The case of the airline industry. Transportation Research Part A: Policy and Practice, 42(1), 227-242.
[47] Yadav, J., & Sharma, M. (2013). A Review of K-mean Algorithm. International journal of engineering trends and technology, 4(7), 2972-2976.

klasteru dovodi do promjene težišta, pa se ponovno ponavljamo cijeli opisani proces.[48] Koraci se ponavljaju sve dok se centri i klasteri ne pomaknu prema svojoj konačnoj poziciji.

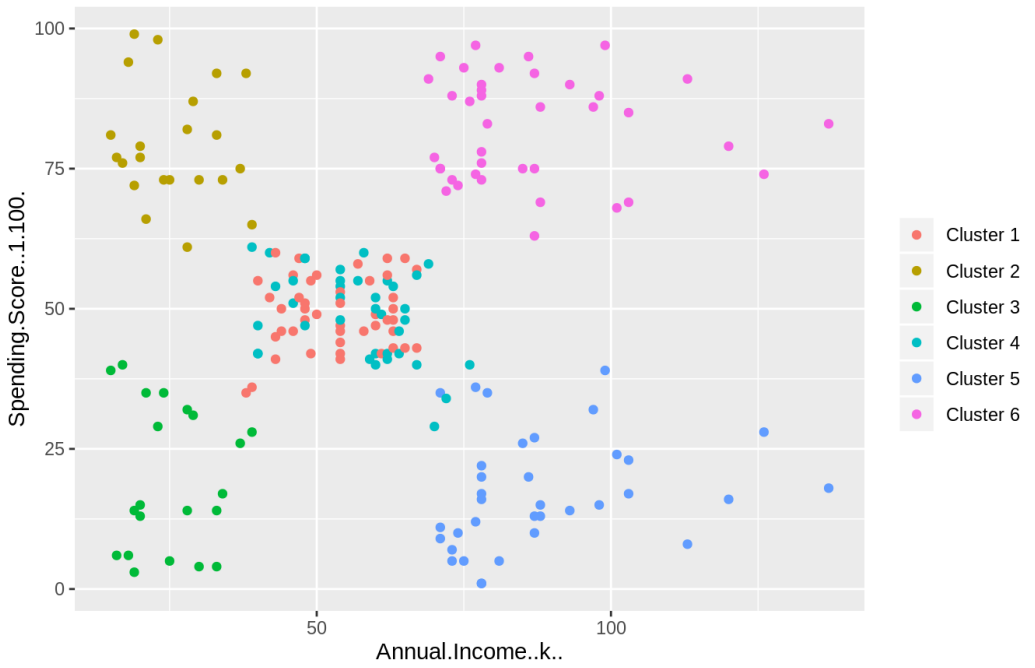
Prije nego što se krene u proces klasteriranja podataka potrebno je odrediti vrijednost *k* iz gore navedenog procesa. Naime, *k* predstavlja broj klastera koji su optimalni za određenu bazu podataka. Kako je vidljivo iz Slike 21, logičkim zaključkom možemo utvrditi da bi našoj bazi podatak optimalan broj klastera sezao od 4 do 6. Problem je kada baze podataka sadrže milione promatranja i veći broj varijabli, što nije rijedak slučaj. U svrhu određivanja broja klastera osmišljene su mnoge kompleksne metode odabira optimalnog broja *k*. Najčešće korištene metode su gap statistic metoda, metoda siluete (silhouette) te metoda lakta. Provođenjem gap statistic i silhouette metoda došli smo do optimalnog broja od 6 klastera za analizu potonje baze podataka. Kako su procesi provođenja svake od ovih metoda opširni i nisu prikladni za prikazivanje u ovom radu možemo i proizvoljno uzeti da $k = 6$ s ciljem lakše prezentacije utjecaja strojnog učenja i Big date na donošenje poslovnih odluka.

4.2.8. Vizualizacija rezultata klasteriranja u procesu segmentacije

Nakon što smo za optimalan broj klastera izabrali 6 možemo provesti naredbu za kreiranje vizualnog prikaza.

```
ggplot(customer_data, aes(x =Spending.Score..1.100., y =Age)) +
 geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
 scale_color_discrete(name=" ",
                      breaks=c("1", "2", "3", "4", "5","6"),
                      labels=c("Cluster 1", "Cluster 2", "Cluster 3", "Cluster 4", "Cluster 5","Cluster 6"))
```

[48] https://www.poslovnaucinkovitost.eu/kolumne/poslovanje/klasterizacija-k-means-algoritmom-u-excelu



Slika 23. Klasteri - Međuovisnost varijabli „Godišnji prihod“ i „Spending score“.

Iz Slike 23. jasno je vidljivo 6 formiranih klastera.

Klasteri 1 i 4 su klasteri koji obuhvaćaju isto područje koordinatnog sustava. U njima su sadržani potrošači „umjerenih godišnjih prihoda“ čiji je spending score također „umjeren“.

Klaster 2 sadrži potrošače niskog godišnjeg prihoda sklone visokoj potrošnji.

Klaster 3 sadrži potrošače niskog godišnjeg prihoda ali i niskog „Spending scorea“.

Klaster 5 formiran je od potrošača koji imaju niski „Spending score“ unatoč tome što imaju visoka godišnja primanja.

Zadnji je Klaster 6 koji obuhvaća sve potrošače visokih godišnjih primanja kojima je pridan visok „Spending score“.

Poslovni subjekt koji posjeduje ovako klasterizirane podatke može se odlučiti daljnje analize i kreiranje poslovnih strategija kreiranih na subjektima iz klastera 1, 2, 4 i 6.

Klasterizacijom podataka otvaraju nam se mogućnosti daljnjeg analiziranja podataka ali mogućnost provođenja analize glavnih komponenti i faktorske analize. [49]Ove analize od

[49] Dago, D. N., Fofana, I. J., Diarrassouba, N., Barro, M. L., Moroh, J. L., Dagnogo, O., ... & Giovanni, M. (2019). A Quick Computational Statistical Pipeline Developed in R Programing Environment for Agronomic Metric Data Analysis. American Journal of Bioinformatics Research, 9(1), 22-44.)

značajne su koristi jer svode brojne varijable na nekoliko faktora i tako omogućuju potpuniju sliku međuovisnosti varijabli.

4.2.9. Zaključak provođenja analize

Iako smo u kratkom vremenskom periodu proveli pojednostavljenu analizu segmentacije anketiranih osoba došli smo do vrijednih podataka kojima bez alata strojnog učenja ne bi imali tako lak pristup. Analiza istih podataka na gore predstavljen način iziskivala bi neusporedivo više vremena i ljudskog truda da se nismo služili strojnim učenjem. Također, neusporediva prednost analize velikih podataka putem strojnog učenja je mogućnost obrade podataka u realnom vremenu i mogućnost obrade prikupljenih podatak bez da isti zastare.

U ovoj kratkoj analizi saznali smo da je udio žena u korištenoj bazi podataka 56% a muškaraca 44%. Najveći broj ispitanika, njih 36, pripada intervalu od 30-35 godina. Najmanje ispitanika pripada intervalu od 60-65 godina. Najveći broj ispitanih zarađuje od 70-80 tisuća kuna godišnje a najmanji broj ih godišnje uprivredi 120-140 tisuća kuna, najmanja godišnja plaća iznosi 15 tisuća kuna dok najviša iznosi 137 tisuća. Po „Spending scoreu“ 40 ispitanika pripada razredu sa spending scoreom od 40 do 50. Istom analizom smo utvrdili da gledajući međuovisnost varijabli „Godišnji prihod“ i „Spending score“ ukupnu populaciju baze podataka možemo rasporediti u 6 klastera. Segmentiranjem 200 ispitanih u 6 grupa stvorili smo temelje za daljnje analize klijenata prema važnosti određenog klastera za poslovnu strategiju tvrtke i provođenje kompleksnijih analitičkih i grafičkih analiza kao što su analiza glavnih komponenti i faktorska analiza.

Istom metodologijom moguće je segmentirati korisnike bankarskih usluga prema njihovoj izloženosti kreditnom riziku ili prema bilo kojoj drugoj varijabli signifikantnoj u svakodnevnom bankarskom poslovanju. Nakon segmentacije i klasteriranja korisnika znatno lakše je upravljati rizicima plasmana bilo kojeg oblika kredita ili zaduženja.

4.3.0. Analiza hipoteze

Na temelju provedenog istraživanja te primjera navedenih u prošlom poglavlju, prihvaća se hipoteza H1: Big data tehnologiju moguće je koristiti u segmentaciji i profiliranju klijenata banke u svrhu upravljanja rizicima. Kroz treće poglavlje smo naveli primjere primjene Big Data tehnologije u različitim kompanijama. Najbolji i najmjerljiviji rezultati su kod caseova koji za cilj imaju konkretiziranje korisničkih podataka kroz alate Big Data tehnologije. Već nakon

navođenja tih primjera dalo se naslutiti kako je primjena moguća i u bankarstvu koje posjeduje enormnu količinu podataka za identificiranje slike potrošača. Navođenjem primjera u bankarstvu, pokazali smo stvarne primjere uspješne primjene ove tehnologije u svijetu. Dodatno, kroz analizu Huskya, Combisovo rješenja za Erste banku, obuhvatili smo prikaz najvažnijih primjena Big Date u poslovanju suvremenih banaka.

Osim teoretskog istraživanja te definiranja raznih područja primjene, u četvrtom poglavlju smo suzili područje istraživanja Big Data tehnologije na segmentaciju i profiliranje klijenata. Kroz jednostavni primjer prikazali smo kako pomoću alata Big Date (program R) radimo segmentiranje i profiliranje klijenta banke. Nakon segmentiranja i profiliranja, napravili smo klastere koji bankama omogućuju lakšu, bržu i točniju procjenu kreditnog rizika. Osim poboljšanja kvalitete i brzine procesa procjene rizika klijenta, Big Data omogućuje alokaciju većeg dijela resursa koje su do sada koristili za ovaj proces.

Iz svega navedenog, jasno je kako i zašto možemo prihvatiti hipotezu kao istinitu.

5. Svjesnost zaposlenika promatranih banaka o tehnologiji velikih podataka

Pojava bilo koje nove tehnologije uzrokuje promjenu radnih navika zaposlenika, unutar organizacije. Često menadžmentu poduzeća otpor promjenama, među zaposlenicima, predstavlja veliku prepreku pri primjeni novijih tehnoloških rješenja. Intenzitet otpora ovisi o nekoliko najčešćih faktora: djelatnost, dobna struktura, kompleksnost nove tehnologije, stupanj odbojnosti društva u cjelini na promjene, kvaliteta kadra i sl. Unatoč tome, benefiti novih tehnologija su toliko veliki, u ekonomskom i društvenom smislu, da u ovoj borbi sve češće i brže pobjeđuje tehnologija.

Kako do sada u kontekstu tehnoloških inovacija, tako i danas, u slučaju pojave Big Data tehnologije, postoje prepreke njenoj bržoj primjeni. Glavni razlog sporijem širenju primjene nije otpor zaposlenika već nekompetencija istih. Izvješće CapGeminij [50]a otkrilo je da 37% tvrtki ima problema s pronalaženjem odgovarajućih analitičara podataka koji bi mogli iskoristiti njihove podatke. Njihova najbolja varijanta je formirati zajednički tim za analizu podataka unutar kompanije, bilo putem prekvalifikacije postojećih radnika ili zapošljavanja novih radnika specijaliziranih za velike podatke.

Na tragu ovog problema, za potrebe ovog rada će se provesti analiza među zaposlenicima pojedini hrvatskih banaka. U analizi će se kroz upitnik doći do saznanja koliko su zaposlenici banaka, na odgovarajućim pozicijama, svjesni pojma i mogućnosti Big Data tehnologije.

5.1. Hipoteza istraživanja

Analiziranjem odgovarajuće literature i osobnog shvaćanja iste, postavljena je istraživačka hipoteza:

H1: Zaposlenici na relevantnim pozicijama u promatranim bankama, svjesni su prednosti Big data tehnologije u procesima poslovanja s klijentima

[50] https://www.piesync.com/blog/top-5-problems-with-big-data-and-how-to-solve-them/

5.2. Metodologija istraživanja

Primarni podatci su prikupljeni kroz jednokratno deskriptivno istraživanje na namjernom uzorku. Empirijsko istraživanje je provedeno metodom ispitivanja koristeći se anketnim upitnikom kao glavnim instrumentom istraživanja.

Cijela anketa je provedena "on line" odnosno putem interneta. Ovaj medij je izabran kako bi se lakše i brže pristupilo ispitanicima Anketiranje je provedeno putem elektronske pošte.

Analizirajući podatke, utvrđeno je da je anketi pristupilo 57 ispitanika od ukupno poslanih 112 upita što čini 50,89% cijelog namjernog uzorka ispitanika.

Svih 57 ispitanika su zaposlenici neke od banaka u Splitsko-dalmatinskoj županiji. Od 57 ispitanika, najviše ih ima 31-40 godina njih 57,9 %. 21,1% su ispitanici koji imaju između 19-30 godina, 17,5% su ispitanici između 41-50 godina te 3,5% njih imaju 51 i više godina.

Dodatno su raspoređeni po banci u kojoj rade. Tako imamo zaposlenike OTP banke, Wuestenrot, PBZ, Erste i Addiko banke.

U namjernom uzorku se nalaze relevantne pozicije za ispitivanje navedene hipoteze. Izabran je uzorak zaposlenika koji bi trebali biti upoznati s tehnologijom Big Date. Unutar uzorka se nalaze zaposlenici low, middle i top menadžmenta. Tako su ispitanici podijeljeni u 3 kategorije: a) Osobni bankar back office; b) Voditelj tima, odjela, poslovnice c) direktor, član uprave, predsjednik uprave. Ispitanici su trebali odabrati poziciju koja najbliže opisuje njihovu.

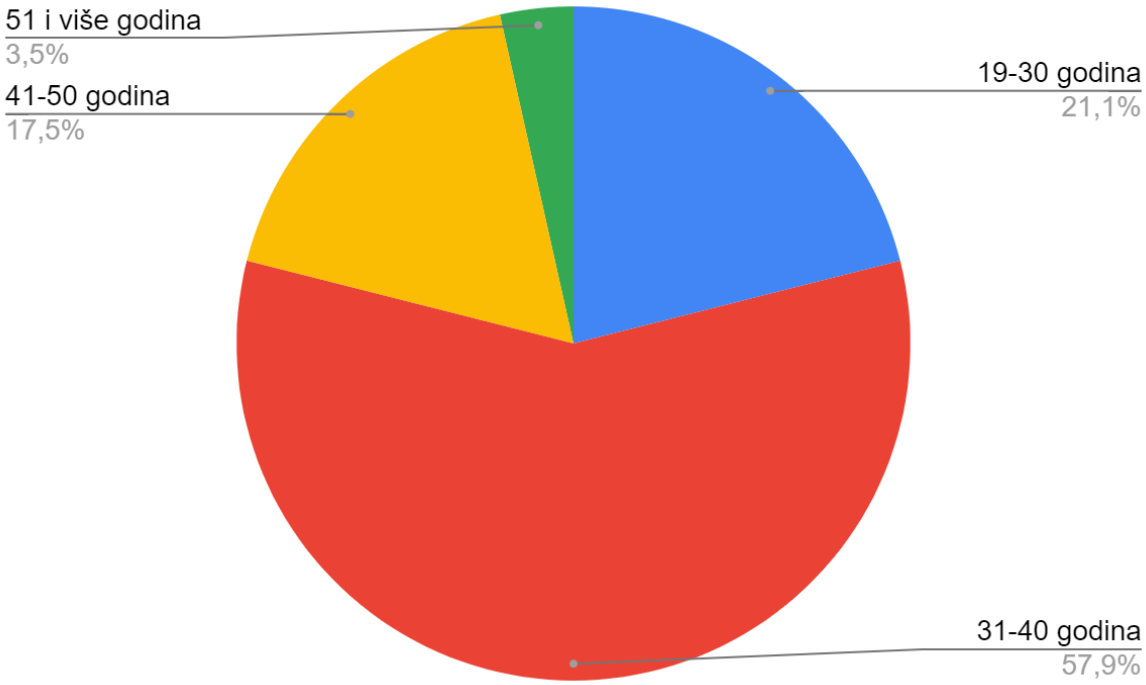
Anketni upitnik je sastavljen od 7 pitanja, kako bismo zadržali fokus ispitanika te kako bismo kroz malu količinu pitanja došli do informacija koje trebamo, kako bismo ispitali postavljenu hipotezu. Osim već navedenih segmentacijskih pitanja, dodana su 3 pitanja koja opisuju koliko su ispitanici svjesni Big Data tehnologije te koliko im pomaže u poslu.

5.3. Rezultati istraživanja

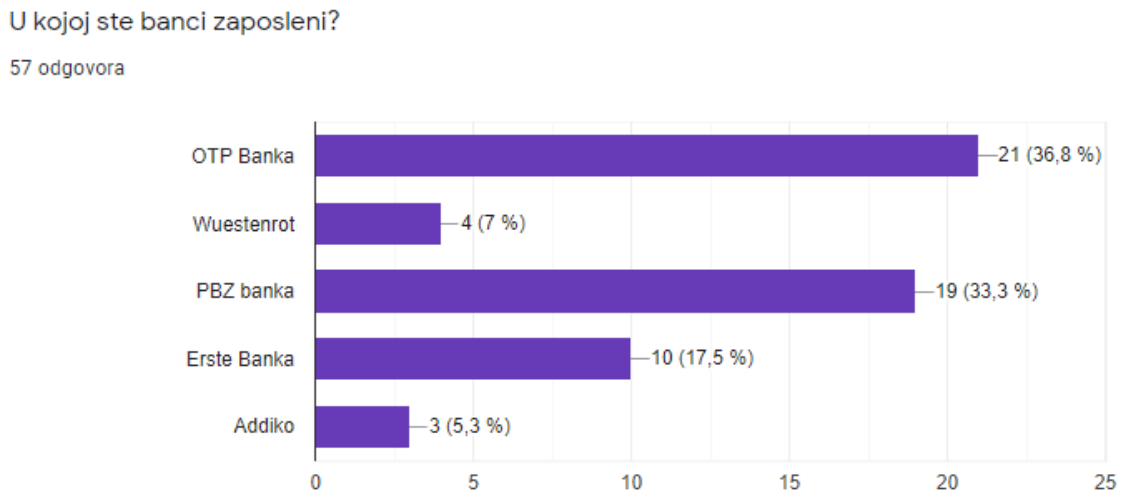
5.3.3. Deskriptivna analiza prikupljenih podataka

U ovom djelu će biti opisani detaljno rezultati koje smo dobili kroz anketni upitnik. Dobivene odgovori će biti prikazani i odgovarajućim grafičkim prikazom kako bi rezultat bio što jasniji.

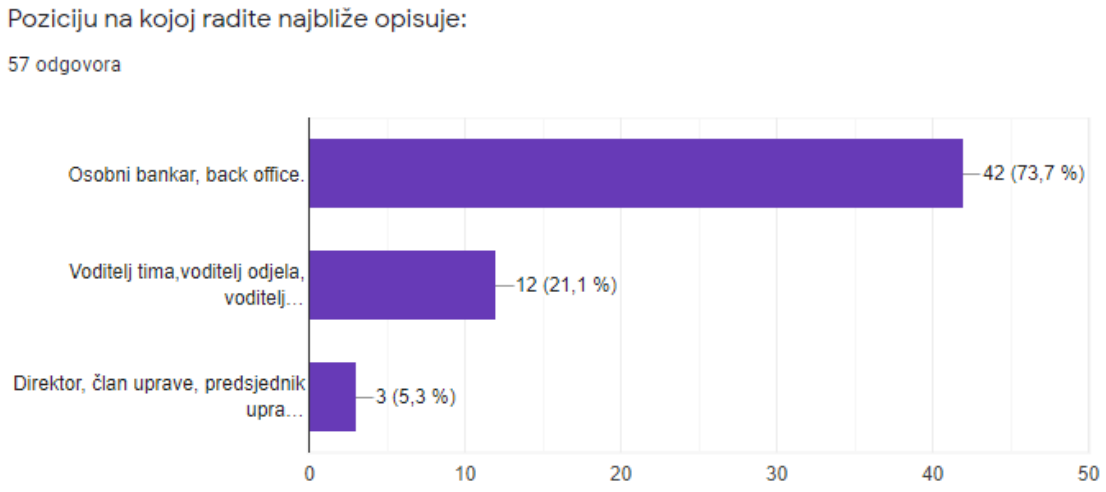
Kako smo naveli svi ispitanici su zaposlenici neke od banaka u Splitsko-dalmatinskoj županiji te njihova dobna struktura prikazana grafikonom izgleda ovako:



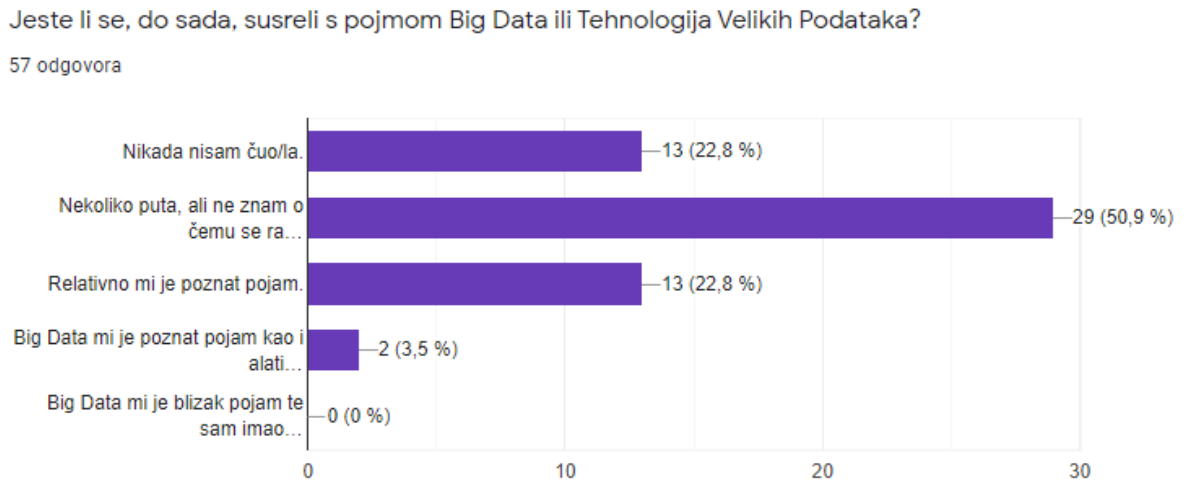
Jasno je vidljivo kako najviše ih ima 31-40 godina njih 57,9 %. 21,1% su ispitanici koji imaju između 19-30 godina, 17,5% su ispitanici između 41-50 godina te 3,5% njih imaju 51 i više godina.

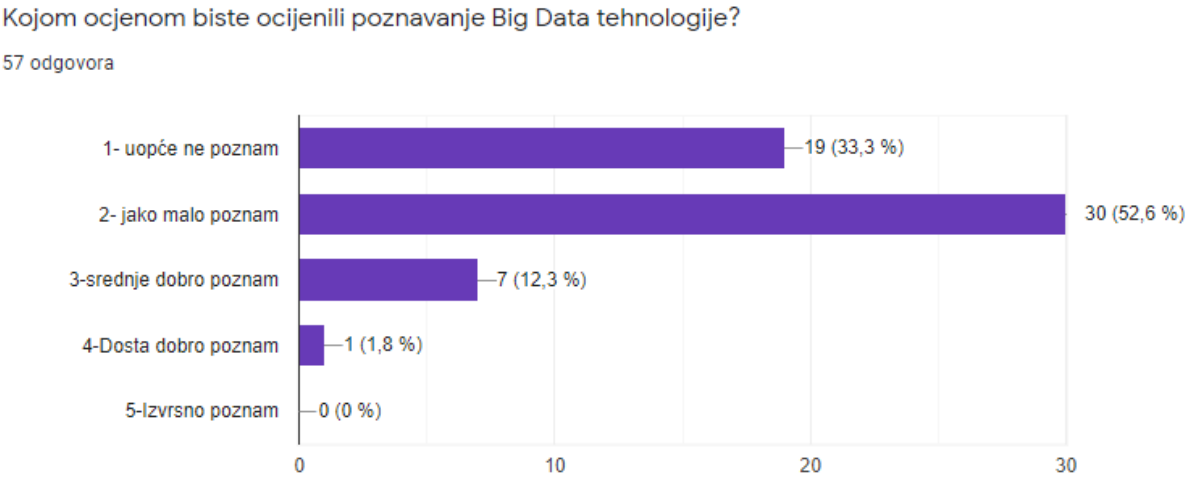


U ovom grafikonu će se prikazati struktura ispitanika po bankama u kojima su zaposleni. Primjećujemo kako 67,1%, preko dvije trećina ispitanika otpadaju na OTP i PBZ banku, 17,5 % na Erste Banku te manji dio, ukupno 12,3 % na Erste Addiko. Razlog ovakvoj strukturi leži u ukupnom broju zaposlenih, na relevantnim pozicijama, na području Splitkso-dalmatinske županije te suradnji predstavnika sindikata u navedenim bankama.



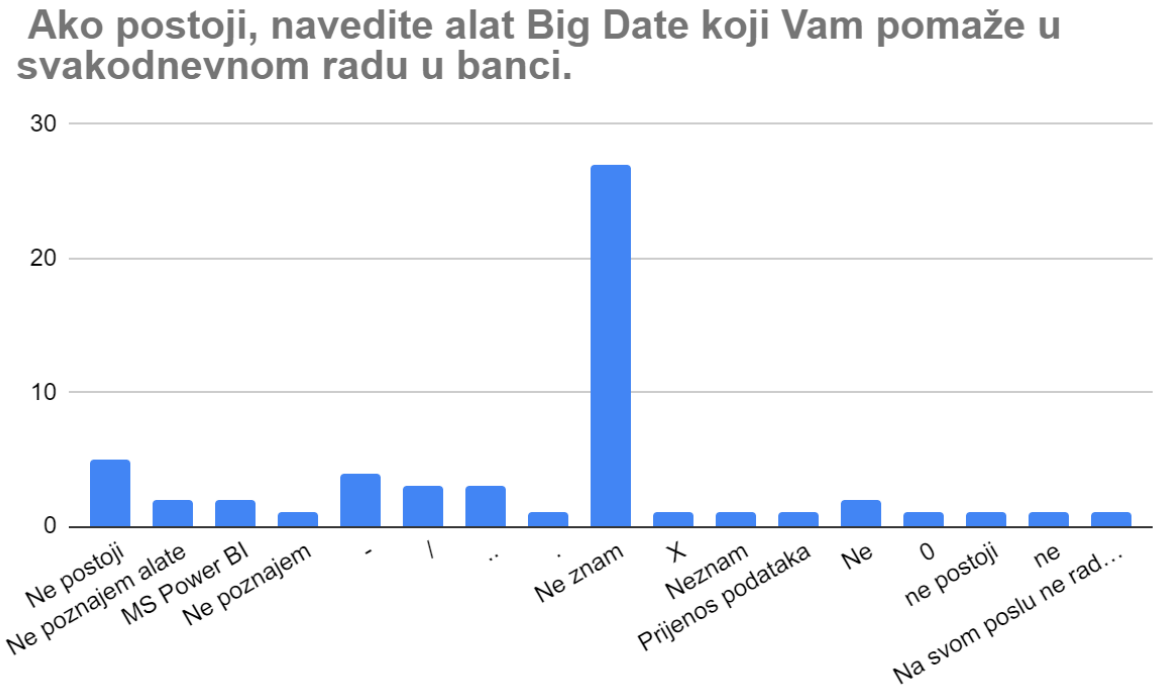
Vidljivo je kako je anketi pristupilo uvjerljivo najviše ispitanika koji svoju poziciju opisuju najbliže pozicij Osobnog bankara, back office, njih čak 73,7 %, Dok samo 5,3 % ispitanika se izjasnilo da su na poziciji direktor, člana uprave ili predsjednika uprave. Ovakav rezultat je logičan. Struktura u kompanijama je slična strukturi uzorka, top menadžment nema vremena za ispunjavanje anketa.





U gore navedena dva pitanja smo dobili rezultate mišljenja ispitanika o tome koliko su se često susreli s pojmom Big Data te koliko zapravo misle da poznaju Big Data tehnologiju. U prvom pitanju vidimo kako 50,9% ispitanika je nekoliko puta čula za pojam Big Datu ali ne znaju o čemu se točno radi. 22,8 % ispitanika tvrdi da im je Big Data relativno poznat pojam. Nikad nije čulo za Big Datu 22,8 % ispitanika te 3,5 % tvrdi kako pozna pojam i alate Big Data tehnologije. Nijedan ispitanik ne smatra big Datu kao blizak pojam te da je imao priliku koristiti alate Big Date.

U drugom pitanju je trend i struktura jasna te je vrlo slična prethodnom pitanju.



U ovom pitanju je ostavljeno ispitanicima opcija da napišu odgovor te im nije ponuđen nikakav prijedlog odgovora. Razlog tome je kako bismo provjerili postoji li eventualno nepodudaranje s odgovorima na prethodna pitanja. Imamo samo tri pozitivna odgovora, od kojih jedan nije prihvatljiv(prijenos podataka). Dva ispitanika su odgovorila MS Power BI koji se može uzeti kao prihvatljiv odgovor jer može poslužiti kao svojevrsna ekstenzija pojedinim Big Data alatima.

5.4. Zaključak istraživanja

Završetkom istraživanja koje smo proveli na namjernom uzorku od 57 zaposlenika 5 banka u Splitsko-dalmatinskoj županiji, dobili smo jasne rezultate o poznavanju Big Data tehnologije navedenog uzorka.

85,9 % ispitanika tvrdi kako ne pozna ili jako malo pozna Big Data tehnologiju, 12,3% srednje dobro pozna i 1,8 % odnosno jedan ispitanik tvrdi da dosta dobro pozna pojam Big Data tehnologije dok izvrsno poznavanje nitko nije naveo kao opciju. Posebno se ističe činjenica da od 57 ispitanika je samo troje (5,26%) dali ikakav odgovor na pitanje da navedu neki od alata Big Date koje koriste u poslovanju.

Ako promatramo rezultate istraživanja po hijerarhijskoj strukturi, vidimo da na zadnje pitanje od tri odgovora, dva su relevantna. Od dva relevantna jedan je iz grupe direktor, član ili predsjednik uprave. Dodatno, samo jedan ispitanik iz ove grupe je naveo bilo kakav odgovor. Ovaj podatak nam ukazuje na problem nepoznavanja tehnologije velikih podataka i na najvišim razinama menadžmenta banke.

Iz ovih rezultata je vidljivo kako razina poznavanja tehnologije velikih podatak na odabranom uzorku je na poprilično niskoj razini. Usudio bih se reći i zabrinjavajućoj, s obzirom na mogućnosti primjene Big Data tehnologije u svakodnevnom poslovanju.

5.5.Analiza hipoteze

Prethodno istraživanje je rađeno kako bismo dobili informacije koje će nam pružiti uvid u razinu poznavanja Big Data tehnologije u određenim bankama Splitsko-dalmatinske županije. Informacije koje smo dobili nam ukazuju da hipotezu, H2: Zaposlenici na relevantnim pozicijama u promatranim bankama svjesni su prednosti Big Data tehnologije u procesima poslovanja s klijentima, ne prihvaćamo kao istinitu.

S obzirom na rezultate istraživanja ne treba se pretjerano truditi da bi objasnili razlog neprihvaćanja naveden hipoteze. Zaposlenici se nisu trudili prikriti nepoznavanje pojma Big Date. Razlog tome je to što misle da ni ne trebaju znati. Više od rezultata cijelog uzorka, zabrinjava razina (ne)poznavanja Big Data tehnologije zaposlenika na pozicijama direktora, člana uprave ili predsjednika uprave.

6. Zaključak

Big Data ili tehnologija velikih podataka nije budućnost, to je već nekoliko godina aktualna sadašnjost aktera na poslovnom tržištu. Gotovo sva velika poduzeća u svijetu, su prihvatili Big Datu kao tehnologiju od koje mogu imati velike koristi. Sukladno tome, velika većina njih je dio svojih ljudskih i financijskih resursa namijenila razvoju primjene Big Date u svom poslovanju. Paralelno s rastom primjene Big Date, raste i pojava kompanija koje su se specijalizirale za implementaciju Big Date u svakodnevno poslovanje. Kompanije, vlade, razne institucije i pojedinci koriste njihove usluge Pojedine kompanije, posebice velike korporacije, idu korak dalje i osnivaju vlastite odjele koji se bave Big Data tehnologijom.

Banke i financijske institucije ,također nije zaobišao trend povećanja važnosti Big Dana tehnologije. U radu su navedena ključna područja primjene ove tehnologije u poslovanju suvremene banke. Kroz pristup podacima potrošnje svojih klijenata alatima Big Data tehnologije mogu, u realnom vremenu, pratiti njihove navike, sklonosti, preferencije i sl. Na taj način mogu prilagoditi svoje aktivnosti različitim grupama klijenata. U istraživanju smo prikazali kako kroz alate Big Data tehnologije, banke mogu segmentiranjem i profiliranjem procijeniti kreditni rizik pojedine transakcije. Upravo sustavni i individualni rizik kroz kvalitetnu politiku upravljanja rizicima predstavlja jedan od najvažnijih faktora poslovanja suvremenih banaka i financijskih institucija. Osim valorizacije prošlih i trenutnih pojava u poslovanju, Big Data ima (može imati) još i veću važnost. Prediktivna analitika zasnovana na tehnologiji velikih podataka omogućuje poduzeću "predvidjeti budućnost". Naravno da postoje određena odstupanja, ali već sada stupanj preciznosti predviđanja budućih kretanja odabranih varijabli je zapanjujući. Sve veći broj banaka koriste ovaj alat u planiranju poslovnih aktivnosti.

Paradoksalno navedenom, dobili smo rezultate istraživanja o poznavanju Big Data tehnologije na uzorku zaposlenika banka. Rezultati ukazuju da unatoč važnosti tehnologije velikih podataka, svjesnost zaposlenika o istoj nije na zadovoljavajućoj razini. Svjesnost, poznavanje promatrane tehnologije se ne mijenja na većim hijerarhijskim razinama unutar poduzeća. Slijedom navedenog, da se zaključiti kako će proći još vremena do aktualiziranja ove teme u hrvatskim bankama.

S obzirom na potencijal mogućnosti primjene tehnologije velikih podataka, kao i stopu rasta podataka koji se generiraju svakodnevno, može se očekivati učestalija pojava i rast važnosti pojma Big Data u svakodnevnom poslovanju banka. Potpuno opravdano.

Popis literature:

Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu: Big data Challenge: a data management perspective, Key Laboratory of Data Engineering and Knowledge Engineering, School of Information, Renmin University of China, Beijing 100872, February 22, 2013, Front. Comput. Sci., 2013, 7

Russom, P. (2011). TDWI Best Practices Report – Big Data Analytics. TDWI Research.

Tsiptsis, K. K., & Chorianopoulos, A. (2011). Data mining techniques in CRM: inside customer segmentation. John Wiley & Sons. stranica 4.

Izvori s interneta:

1. J.P. Morgan: Solutions
   https://www.jpmorgan.com/global/technology/applied-AI-and-ML#key-initiatives
2. Data Flair: Big Data bank industry
   https://data-flair.training/blogs/big-data-in-banking/
3. Data Flair: Big Data bank industry
   https://data-flair.training/blogs/big-data-in-banking/
4. J.P. Morgan: About Us
   https://www.jpmorgan.com/global/technology/applied-AI-and-ML#key-initiatives
5. Kreditech: Case study Kreditech
   http://www.eurofinas.org/uploads/documents/Non-visible/Big%20data/Roundtable/Eurofinas_Kreditech.pdf
6. Martin Schoell(2015): Big Data Scoring for Consumer Lending
   https://www.hbs.edu/openforum/openforum.hbs.org/goto/challenge/understand-digital-transformation-of-business/kreditech-big-data-scoring-for-consumer-lending.html
7. Dr Ichak Adizes(2017)-O efektivnosti i efikasnosti i njihovim posljedicama EFOS

http://www.efos.unios.hr/poduzetnicke-strategije/wp-content/uploads/sites/231/2017/03/Efektivnost_efikasnost_Adizes.pdf

8. Sveučilište u Zadru(2018)- Metode Znanstvenih Istraživanja http://www.unizd.hr/portals/4/nastavni_mat/1_godina/metodologija/METODE_ZNANSTVENIH_ISTRAZIVANJA.pdf

8.Investopedia-Business Ess https://www.investopedia.com/terms/b/big-data.asp

9.Informacija, *Hrvatska enciklopedija, mrežno izdanje.* Leksikografski zavod Miroslav Krleža http://www.enciklopedija.hr/Natuknica.aspx?ID=27405

11. Informacija, *Hrvatska enciklopedija, mrežno izdanje.* Leksikografski zavod Miroslav Krleža(http://www.enciklopedija.hr/Natuknica.aspx?ID=27405)

12. J.P. Morgan: Solutions https://www.jpmorgan.com/global/technology/applied-AI-and-ML#key-initiatives

13. Data Flair: Big Data bank industry https://data-flair.training/blogs/big-data-in-banking/

14. J.P. Morgan: Key initiatives https://www.jpmorgan.com/global/technology/applied-AI-and-ML#key-initiatives

15.Case study: Kreditch: Big Dat In Online Consumer Finance http://www.eurofinas.org/uploads/documents/Non-visible/Big%20data/Roundtable/Eurofinas_Kreditech.pdf

16. Martin Schoell (2015): Big Data Scoring for Consumer Lending https://www.hbs.edu/openforum/openforum.hbs.org/goto/challenge/understand-digital-transformation-of-business/kreditech-big-data-scoring-for-consumer-lending.html

17. Dr Ichak Adizes(2017)-O efektivnosti i efikasnosti i njihovim posljedicama EFOShttp://www.efos.unios.hr/poduzetnicke-strategije/wp-content/uploads/sites/231/2017/03/Efektivnost_efikasnost_Adizes.pdf

18. UNIZD (2018): Metode Znanstvenih Istraživanja http://www.unizd.hr/portals/4/nastavni_mat/1_godina/metodologija/METODE_ZNANSTVENIH_ISTRAZIVANJA.pdf

19. Investopedia-Business Ess https://www.investopedia.com/terms/b/big-data.asp

20. Dumbill, E. (2012) What is big data?: An introductionto the big data landscape. O'Reilly Media Inc., http://www.springer.com/us/book/7983319106649

21. Julio, P. (2009) Big Data Analytics with Hadoop. LinkedIn Corporation. http://www.slideshare.net/PhilippeJulio/hadoop-architecture

22. Mitchell, R.L. (2014). 8 big trends in big data analytics. Computerworld, http://www.computerworld.com/article/2690856/8-big-trends-in-big-data-analytics.html

23. Economist Intelligence Unit (2015). The Deciding Factor. Big Data & Decision Making.Capgemini. http://capgemini.com/thought-leadership/the-deciding-factor-big-data-decision-making

24. Julio, P. (2009) Big Data Analytics with Hadoop. LinkedIn Corporation. http://www.slideshare.net/PhilippeJulio/hadoop-architecture

25. Lockwood, G. K. (2014). Conceptual Overview of Map-Reduce and Hadoop. http://www.glennklockwood.com/data-intensive/hadoop/overview.html

26. Schneider, R. D. (2013). Hadoop Buyer's Guide. Ubuntu. http://insights.ubuntu.com/wp-content/uploads/HadoopBuyersGuide_sm.pdf

27. Husky. Combisovo (BIG) DATA rješenje. Brošura za financijski sektor. 2019. https://bigdata.combis.hr/wp-content/uploads/2017/08/Husky_za_financije.pdf

28. David Reinsel – John Gantz – John Rydning (2018) The Digitization of the World From Edge to Core https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

29. Binu Mathew, Global Head of Development & Product Management, GE Oil & Gas Digital, 12/13/2016. How Big Data is reducing costs and improving performance in the upstream industry? https://www.worldoil.com/news/2016/12/13/how-big-data-is-reducing-costs-and-improving-performance-in-the-upstream-industry

30.Data Flair tim (2018): Top 10 Big Data Tools that you should know about https://data-flair.training/blogs/top-big-data-tools/

31.Cloudera official website: Front page. https://www.cloudera.com/products/open-source/apache-hadoop/apache-storm.html

32. Stephen Watts 2020 https://www.bmc.com/blogs/apache-cassandra-introduction/

33. Linly Ku 2019: The Impact of Big Data in Business

34. Zornitsa Stoycheva 2018: How do global retail leaders use Big data? https://blog.datumize.com/how-do-global-retail-leaders-use-big-data-5-real-life-examples#smooth-scroll-top

35. Jennifer Wills, 2020: 7 Ways Amazon Uses Big Data to Stalk You

https://www.investopedia.com/articles/insights/090716/7-ways-amazon-uses-big-data-stalk-you-amzn.asp

36. Forbes 2019 Starbucks Top Line To Grow By 10% in FY 2019 https://www.forbes.com/sites/greatspeculations/2019/09/26/starbucks-top-line-to-grow-by-10-in-fy-2019/#4fe578da494f

37. Katherine Knowles-Marchione and Mariah Kolpek (2017) Danske Bank: Innovating in Artificial Intelligence and Deep Learning to Detect Sophisticated Fraud

https://www.teradata.com/Blogs/Danske-Bank-Innovating-in-Artificial-Intelligence

38.Beyond The Arc (2011): BOA Case Study- Are You Missing Opportunities to Listen to Your Customers? https://beyondthearc.com/wp-content/media/cases/BTA-CaseStudy-BankofAmerica-SocialMediaAnalytics_10-5-11.pdf

39. Vanessa Rombaut 2020.: Top 5 problems with big data - and how to solve them

https://www.piesync.com/blog/top-5-problems-with-big-data-and-how-to-solve-them/

Popis slika i grafikona:

Sažetak

Banke u Hrvatskoj predstavljaju jedan od glavnih kotačića koji omogućuje kretanje i održavanje razine gospodarske aktivnosti na tržištu. Sukladno tome važnost banaka u privatnom i poslovnom sektoru je izuzetno velika.

Tendencija svakog poslovnog subjekta je imati zadovoljavajuću razinu efikasnosti poslovanja, u bankarskom sektoru posebna pažnja se pridaje spomenutom faktoru. Big Data tehnologija je postala jedan od najvažnijih alata u povećanju efikasnosti poslovanja suvremenih banaka. U radu je posebno analiziran pozitivan utjecaj tehnologije velikih podatka u procesu upravljanja rizicima. Ukazali smo na Big Datu kao alat koji omogućuje jednostavniju, točniju i bržu procjenu rizika kroz proces profiliranja i segmentiranja klastera korisnika.

Unatoč važnosti, pojam tehnologije velikih podatka kao i njeni alati, nije dovoljno poznat zaposlenicima banaka na promatranom uzorku.

U radu smo objasnili važnost primjene Big Data tehnologije u poslovanju suvremenih banaka. Dodatno, sukladno provedenom istraživanju, ukazali smo na problem nepoznavanje navedene tehnologije među promatranim zaposlenicima banaka.

Ključne riječi: Big Data, Suvremene Banke, Profiliranje i Segmentiranje korisnika,

Summary

Banks in Croatia are one of the main wheels that enable the movement and maintenance of the level of economic activity in the market. Accordingly, the importance of banks in the private and business sectors is extremely high.

Every business entity tends to have a satisfactory level of business efficiency, in the banking sector special attention is paid to this factor. Big Data technology has become one of the most important tools in increasing the efficiency of modern banks. The paper especially analyzes the positive impact of big data technology in the risk management process. We pointed to Big Date as a tool that allows for simpler, more accurate and faster risk assessment through the process of profiling and segmenting user clusters

Despite its importance, the concept of big data technology, as well as its tools, is not sufficiently known to bank employees in the observed sample

In this paper, we explain the importance of the application of Big Data technology in the modern banks. Also, by the conducted research, we pointed out the problem of ignorance of this technology among the observed employees of banks.

Key Words: Big Data, Modern Banks, Profiling and Segmenting User