

# Veliki podaci u turizmu

---

**Abramović, Luka**

**Undergraduate thesis / Završni rad**

**2023**

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

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:137:631872>

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

*Download date / Datum preuzimanja:* **2024-07-29**



*Repository / Repozitorij:*

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



Sveučilište Jurja Dobrile u Puli  
Fakultet informatike u Puli

**LUKA ABRAMOVIĆ**

**VELIKI PODACI U TURIZMU**  
Završni rad

Pula, rujan, 2023. godine

Sveučilište Jurja Dobrile u Puli

Fakultet Informatike u Puli

**LUKA ABRAMOVIĆ**

**VELIKI PODACI U TURIZMU**

Završni rad

**JMBAG:** 0303085591, redoviti student

**Studijski smjer:** Informatika

**Predmet:** Informacijska tehnologija i društvo

**Znanstveno područje:** društvene znanosti

**Znanstveno polje:** informacijske i komunikacijske znanosti

**Mentor:** izvr. prof. doc. dr. sc. Snježana Babić

Pula, rujan, 2023. godine



## IZJAVA O AKADEMSKOJ ČESTITOSTI

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

Student

U Puli, \_\_\_\_\_rujan\_\_\_\_, \_\_\_\_\_2023.\_\_\_\_ godine



## IZJAVA

o korištenju autorskog djela

Ja, Luka Abramović dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom „Veliki podaci u turizmu“ koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, rujan, 2023. godine (datum)

Student

## Sadržaj

<b>1. Uvod</b> .....	1
<b>2. Veliki podaci</b> .....	2
2.1 Opći pojam i karakteristike.....	2
2.2 Tehnologija obrade velikih podataka.....	5
2.3 Rudarenje velikih podataka.....	8
<b>3. Veliki podaci u turizmu</b> .....	13
3.1 Turizam.....	16
3.2 Vrste podataka u turizmu .....	16
<b>4. Primjeri primjene velikih podataka u turizmu</b> .....	21
<b>5. Prednosti i nedostaci korištenja velikih podataka u turizmu</b> .....	36
<b>6. Budući trendovi razvoja velikih podataka u turizmu</b> .....	38
<b>7. Zaključak</b> .....	40
<b>Literatura</b> .....	41
<b>Popis slika</b> .....	47
<b>Popis tablica</b> .....	48
<b>Sažetak</b> .....	49

## 1. Uvod

U današnje vrijeme turizam je jedna od najvećih grana poslovanja, međutim veliki dio ne posluje uz pomoć tehnologije, a tehnologija sve više napreduje. Jedna od tehnologija koja se koristi u turizmu su veliki podaci. Među najvažnijim korištenjem velikih podataka bi služilo kod upravljanja prihoda, gdje bi se kombinacijom prikupljenih informacija potpunosti hotela te vanjskih događaja omogućio profitabilni financijski izvještaj. Cilj ovog rada je kako bi se pobliže objasnila primjena velikih podataka u turizmu. U prvome poglavlju dana je definicija velikih podataka, prikazat će se opći pojmovi te karakteristike velikih podataka kako bi se steklo osnovno razumijevanje prije daljnjeg istraživanja. U drugom poglavlju prikazat će se vrste velikih podataka koji se upotrebljavaju u turizmu. Unutar trećeg poglavlja prikazat će se primjeri primjene velikih podataka u turizmu. Okvir četvrtog poglavlja prikazat će prednosti i nedostatke velikih podataka te peto i posljednje poglavlje koje služi za prikazivanje daljnjih trendova.

## 2. Veliki podaci

U okviru ovog poglavlja dan je opći pojam velikih podataka te će se prikazati vrste velikih podataka. Obradit će se karakteristike velikih podataka, te kojima se oni najčešće opisuju. Spomenuta je tehnologija koja se koristi kako bi se moglo upravljati velikim podacima, te načini prikupljanja velikih podataka.

### 2.1 Opći pojam i karakteristike

Za definiranje velikih podataka unutar literature postoji više različitih definicija:

- „Big data je skup podataka koji ima mogućnost spremanja, upravljanja i obradu podataka u proteklom vremenu“ (Rupali, Pratibha, 2018).
- „Često bi mnogi profesionalni analitičari podataka implicirali proces ekstrakcije, transformacije i opterećenje (ETL) za velike skupove podataka kao konotaciju velikih podataka“ (Wu i sur. 2016).
- „Veliki podaci bave se prilikama, izazovima, analizama i trendovima povezanim s velikim količinama podataka uzrokovanih sve većom i raznolikom zbirkom entiteta koja se kontinuirano stvara svijet od strane ljudi i oko ljudi, osobito putem pametnih senzora i uređaja“ (Wu et al., 2014; DeMauro, 2014.)

Veliki podaci su tehnologija kojom pomoću koje može se upravljati s velikim grupama podataka. Te vrste podataka su veoma korisne različitim organizacijama koje raspolažu s tim podacima na načine kojima su njima bitni. Volumen velikih podataka mjeri se u bitovima. Kako tehnologija sve više napreduje, veličina podataka koja se prije smatrala velikom tu današnjem svijetu se više ne smatra velikom količinom. Trenutno veličina velikih podataka se mjeri u petabiteovima te čak i exabiteovima (Clissa, 2022).

Veliki podaci obuhvaćaju veliki obujam pojmova. Razni podaci poput strukturiranih, polustrukturiranih, nestrukturiranih se implementiraju pomoću naprednih aplikacija, te



primjenjuju u svakodnevnom životu, te raznim aspektima poput obrazovanja ili zdravstva (Mahmoud, 2020).

Budući da su to velike količine podataka, one uglavnom neće biti obrađene unutar memorije, već će se obrađivati na mjestu kojem su prikupljene (Ormandjieva, Omidbakhsh, Trudel, 2020). Takve grupe podataka mogu biti:

- Strukturirane – U bazama podataka, strukturirane podatke smatramo one koji su uređeni. Primjeri takvih podataka su datum, mjesto, autor. Podaci koji pružaju organizirane informacije koje se spremaju u sustav upravljanja relacijskih baza podataka. Unutar tog sustava zadaća će biti shvatiti njihovu povezanost. Takvi podaci su uglavnom jednostavni za pretražiti jer se sastoje od jasnog definiranog tipa i njihovog uzorka, najčešće ti podaci će biti smješteni u retke i stupce (Ormandjieva i sur. 2020).
- Polustrukturirane – Polustrukturirani podaci oblik su strukturiranih podataka, podaci koji sadrže oznake kako bi se podijelili semantički elementi te ispunila ljestvica važnosti zapisa i polja podataka. Podaci koji ne prikazuju točnu i egzaktnu shemu, već sadrže oznake kojima se utvrđuju dijelovi tj. elementi podataka. Entiteti koji u grupama vjerojatno neće imati iste karakteristike, te podaci neće biti do kraja strukturirani (Ormandjieva i sur. 2020).
- Nestrukturirane – Pod nestrukturirane podatke smatraju se podaci koji nemaju poznati, stabilan oblik te strukturu. Vrste tih podataka spadaju pod medije poput audio, videa, te fotografija. Tekstualni podaci poput tablica, prezentacija te medijski podaci koji čine podaci s Facebook-a, Twitter-a, LinkedIn-a. U suštini su suprotno od strukturiranih podataka. Takvi podaci nemaju korektnu strukturu informacija, podaci zahtijevaju veću količinu spremnika te je puno kompleksnije analizirate takve podatke jer nisu strukturno uređeni (Ormandjieva i sur. 2020).

Glavne karakteristike velikih podataka bi mogli sažeti pojmom koji se naziva 3V-a. Pod tim pojmom se podrazumijeva da veliki podaci opisuju se:

- Volumen – Količina podataka mjeri koliko podataka organizacija ima koji nisu nužno njezini, sve dok im ima pristup. Povećanje obujma podataka rezultirat će u smanjivanju vrijednosti, zbog starosti, vrste, količine (Samir, 2016).

Pod volumen mogu se svrstati podaci koji imaju veliki kapacitet, uglavnom mogu se svrstati podaci prikupljeni od povezanih uređaja ili preko interneta ili različitih digitalnih uređaja. Trenutno u svijetu za jeftinije skladištenje većih količina podataka se npr. koristi Hadoop.

- Varijantnost – Kako bi se podaci mogli analizirati u različite skupine podataka, tu nam koristi raznolikost koja određuje vrstu, izvor podataka (Rupali, Pratibha, 2018). Pod varijantnost se referira na podatke koji su prikupljeni od različitih izvora i u različitim formatima. Ti podaci mogu dolaziti od senzora, pametnih telefona ili socijalnih mreža. Oblici u kojima ti podaci pristižu mogu biti audio, video, slika ili neke poruke.
- Velocitet – Brzina podataka mjeri brzinu kojom se skupovi podataka stvaraju, distribuiraju i agregiraju (Samir, 2016).

Pod velocitet podrazumijevamo brzinu procesa i protok transfera podataka, ali pod to ne spada samo brzina, već i vrijeme kad su podaci primljeni, tj kada su podaci krenuli, kada su pristigli i da li su došli svi odjednom ili u više serija.

Navedene karakteristike velikih podataka te pojmovi koji se podrazumijevaju pod glavnim karakteristikama su prikazani u tablici 1 veliki podaci – 3V.

<b>VOLUMEN</b>	<b>VARIJANTNOST</b>	<b>VELOCITET</b>
PB	Audio	Stvarno vrijeme
TB	Video	
GB	Slika	Periodično
MB	Socijalne mreže	
KB	Poruke	Serijski

*Tablica 1 Veliki podaci -3V*

*Izvor: Obrada autora prema : Ormandjieva i sur. , 2020*

## 2.2 Tehnologija obrade velikih podataka

Kako bi se moglo upravljati velikim količinama podataka kompanije su uglavnom uvijek koristile super-računala i to pod velikim troškom kako bi mogli upravljati tim podacima. Međutim kako u današnjem svijetu računalna tehnologija sve više napreduje razvijaju se nove tehnologije za baratanjem velikih podataka. Najveći napredak je krenuo kada su kompanije poput Google-a krenule razvijati softverski okvir i programske modele poput MapReduce i Hadoop. Analitičari za analizu velikih podataka koriste jedan od najpopularnijih platformi analize velikih podataka, ta platforma je Hadoop (Rajkumar i sur. 2016). U nastavku rada naveden je Hadoop kao jedan od najznačajnijih tehnologija velikih podataka te njegove karakteristike.

### **Hadoop**

Kako bi se velike količine podataka obrađivale na klasterima hardvera, jedan od alata je Apache Hadoop (Vignesh, 2013).

Hadoop je veoma popularan software pomoću kojeg se može analizirati velike podatke. Hadoop software je napisan pomoću programskog jezika Java. Funkcionira preko jednostavnog hardware-a koji nije skup i lagano ga je nabaviti.

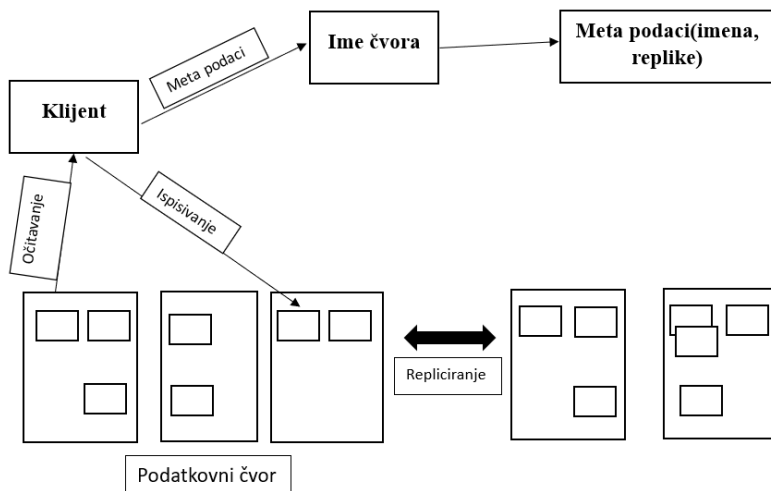
Sastoji se od 3 glavna dijela :

- HDFS – HDFS ili Distribuirani datotečni sustav Hadoop omogućava pohranu velikih količina informacija te obraditi bitne dijelove bez da se podaci izgube. Takav sustav se temelji na Google-om datotečnom sustav (Rupali, Pratibha, 2018).

Hadoop distribuirani datotečni sustav. HDFS je sustav koji koristi se i upotrebljava kako bi se mogli pohraniti podatci. Hadoop je razvio HDFS te on omogućuje distribuirati sustav datoteka i zbog toga je moguće obrađivati strukturirane i ne strukturirane podatke. Datoteke koje se nalaze unutar HDFS, su podijeljene na određenu veličinu. Te veličine su zadane na početnu veličinu od 64 megabita, ali ako je potrebno moguće je izmjenjivati određenu veličinu.

Namjenski poslužitelji služe kako bi se na njih pohranili aplikacijski podaci datotečnog sustava. Hadoop arhitektura ima dva kritična čvora, to su imenski i podatkovni čvor. Na podatkovnim čvorovima se replicira sadržaj datoteke.

Korištenjem TCP protokola vrši se komunikacija između imenskog i podatkovnog čvora (Bagade i sur. 2019). Navedeni opis HDFS arhitekture je prikazan na slici 1 Hadoop – HDFS Arhitektura.

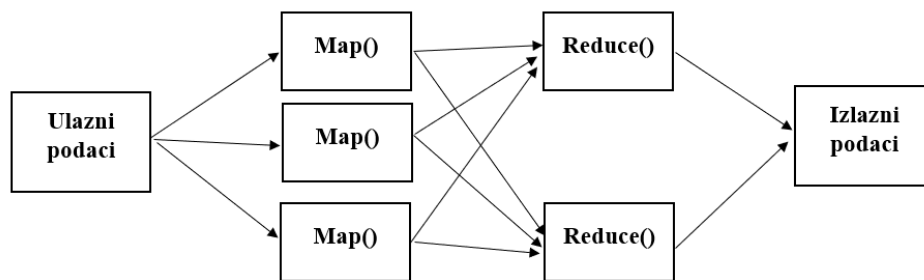


*Slika 1 Hadoop - HDFS Arhitektura*

*Izvor: Obrada autora prema Rupali, Pratibha, 2018*

- MapReduce –Map i Reduce skupa tvore MapReduce algoritam. U prvom dijelu Map podaci se pretvaraju u drugi skup podataka te se dijele, dok Reduce uzima izlazne podatke iz Map te rezultira ih u novi skup (Rupali, Pratibha, 2018). Navedeni proces je vidljivo prikazan na slici 2. Hadoop- MapReduce arhitektura

Hadoop se koristi MapReduce algoritmom, on je dizajniran kako bi bio u mogućnosti paralelno obrađivati podatke preko više različitih računalnih čvorova. Tim procesom bi se omogućilo ubrzano izračunavanje te skratilo se vrijeme kašnjenja podataka. MapReduce je moguće implementirati u različite programske jezike, najviše zastupljeniji programski jezik korištenja je Java. Sastoji se od dvije funkcije. Prva funkcija je Map funkcija, a druga je Reduce funkcija. Map funkcija je zaslužna za raspodjeljivanje ulaznih podataka u parove vrijednosti, nadalje ti parove se zasebno procesuiraju te ih preuzima Reduce funkcija koja ih povezuje i pretvara u izlazne podatke.



*Slika 2 Hadoop - MapReduce arhitektura*

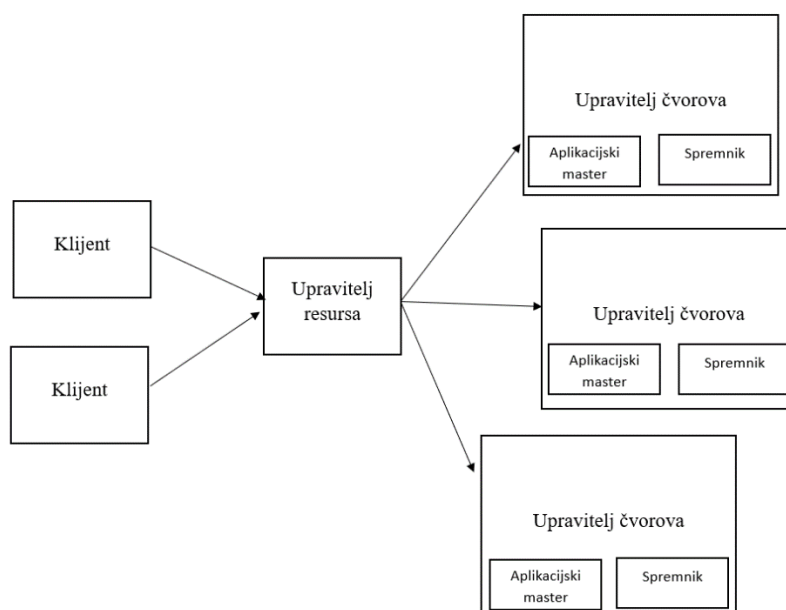
*Izvor: Obrada autora prema: Rupali, Pratibha, 2018*

- YARN – Yarn se smatra alatom za upravljanje resursima koji može dodijeliti resurse klastera za sve vrste okvira za obradu podataka implementiranih za Hadoop (dr. Yusuf i sur. 2017).

Pomoću kojeg može se upravljati razinama resursa. U usporedbi s MapReduce Yarn pruža bolju paralelnost obrađivanja podataka. Yarn je kompatibilan s API od MapReduce-a, prema ako korisnici trebaju ili žele koristiti dijelove od MapReduce-a preko Yarn-a sve što je potrebno je ponovno kompilirati. Yarn je bolji i kvalitetniji od MapReduce-a zbog toga što se dijeli na dvije komponente kako bi poboljšao efikasnost. Dvije glavne komponente na koje se dijeli Yarn su upravitelj resursa i upravitelj čvorova. Upravitelj resursa upotrebljava se za razmicanje ili slaganje resursa po aplikaciji. Unutar upravitelja čvorova imamo aplikacijskog mastera, njegova je uloga da održava komunikaciju s upraviteljom čvorova kako bi dijelovi procesa uspješno se završili. Unutar upravitelja čvorova također se nalazi i spremnik kojim je zadaća da brine o kvalitetnoj uporabi resursa koji su na raspolaganju.

U Yarn arhitekturi, nad glavnim čvorom klastera obrađuje upravitelj servisnih resursa. Yarn klijent nudi aplikaciju za upravljanje imovinom. Upravitelj aplikacija koji služi upravitelju resursa provjerava i prihvaća zahtjeve aplikacija od klijenta. Nakon ovog rasporeda, usluga upravitelja resursa dodjeljuje spremnik za glavnu aplikaciju na čvoru, a upravitelj čvorova na ovom čvoru koristi naredbu za pokretanje

glavne aplikacijske usluge (dr. Yusuf i sur. 2017). Navedeni proces Yarn arhitekture je prikazan na slici 3. Hadoop – Yarn arhitektura.



*Slika 3 Hadoop - YARN arhitektura*

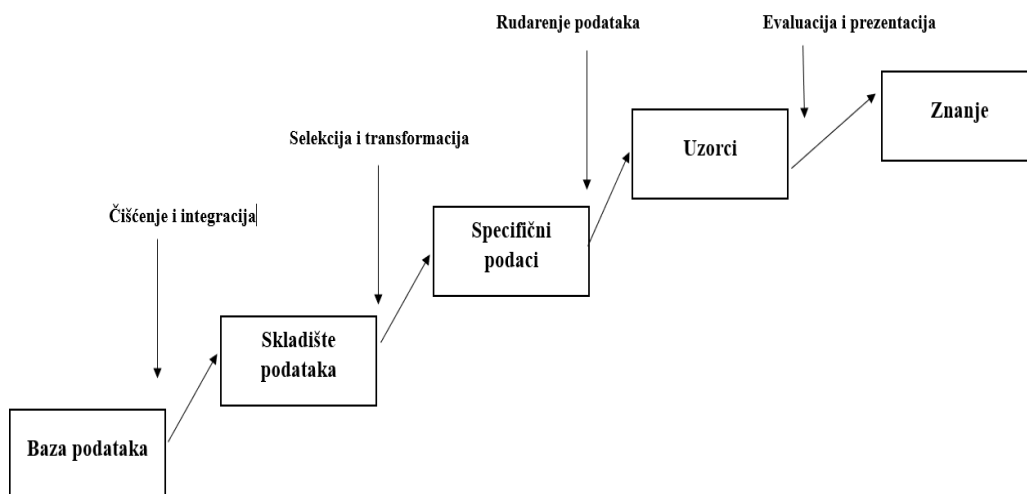
*Izvor: Obrada autora prema: dr. Yusuf i sur. 2017*

### 2.3 Rudarenje velikih podataka

Prilikom velikih količina podataka koji nisu precizirani, potpuni i korisni, koristi se proces rudarenja podataka kako bi se izvuklo znanje te korisne informacije. Prilikom početnog pokretanja sustava izvorni podaci nam služe kao temelj, takvi podaci mogu ne samo koristiti u znanju već i za zaštitu istih (Wanchun, 2017).

Napredak tehnologije je trenutno u svijetu doveo do toga da postoji jako puno podataka, podataka koji mogu biti veoma jeftini i lako dostupni. U tom slučaju dolazi do problema zbog težeg prepoznavanja kvalitetnih i efikasnih podataka. Pri tome nam pomaže rudarenje podataka. Rudarenje podataka je naime proces koji nam omogućava kako bi se moglo izvući korisne informacije i uzorke iz velikih podataka. Cilj je da se postigne pronaći neke nepoznate uzorke is podataka o kojima još nema znanja ili još nisu bili viđeni i upotrebljeni (Bharati, Ramageri, 2010). Generalni proces otkrivanja nekih novih znanja i vrijednih podataka bi se moglo opisati u 4 koraka:

- Čišćenje i integracija te spajanje podataka
- Izabir i transformiranje podataka
- Rudarenje podataka
- Procjena ili evaluacija i prezentacija znanja



*Slika 4 Generalni proces rudarenja podataka*

*Izvor: Obrad autora prema: Bharati, Ramageri, 2010*

Proces započinje s pročišćavanjem podataka i micanja nepotrebnih podataka koji nisu dosljedni. Nakon toga dolazi spajanje podataka iz različitih izvora koje je moguće kombinirati. U slučaju gdje se prvi korak provede kao korak preprocesiranja, podaci će biti spremljeni u skladište podataka. Nakon prvog koraka prelazi se na drugi korak gdje dolazi se do pažljivog biranja podataka iz skladišta podataka koji će nam biti potrebni za upotrebu unutar naše analize, te prelazi se na konverziju formatiranja podataka. Završetkom drugog koraka koristit će se različiti algoritmi te integrirani alati kako bi se uspješno mogli rudariti podaci. Tu se upotreba skladišta podataka i algoritama često kombinira jer su podaci u skladištu podataka uredno i strukturirano složeni te je veoma jednostavno i brzo za izvući podatke koji su nam potrebni. U posljednjem koraku procjenjuju se izvučene tehnike za koje se pretpostavlja da će biti korisne te prezentiranje procesa rudarenja korisnicima. Proces započinje oblikovanjem podataka u druge oblike. Nakon što su podaci obrađeni prelazi se na identificiranje uzorka. Odabrani uzorak se primjenjuje za daljnje predviđanje, te obrasci

se postavljaju za određeni ciljani rezultat (Bharati, Ramageri, 2010). Navedeni proces je prikazan pomoću slike 4. Generalni proces rudarenja podataka.

Prilikom rudarenja podataka razlikuju se dvije vrste načina rudarenja. Prva vrsta je prediktivna, a druga deskriptivna. Kako bi se otkrila svojstva ili komunikacija unutar podataka primjenjuje se deskriptivni model ( dr. Hayfron-Acquah i sur. 2016)

Metode rudarenja podataka mogu se podijeliti na dva dijela. Prva metoda rudarenja je deskriptivna analiza, dok je druga metoda prediktivna analiza. Glavna zadaća deskriptivne analize je opisivanje podataka i njihov veza, dok je glavna zadaća prediktivne analize predviđanje mogućih stavki. Prediktivna analiza se bazira na deskriptivnoj jer uzima njezine prikupljene informacije te pomoću njih kreira predviđanje.

Tehnike te algoritmi koji služe kako bi se otkrili dodatne informacije iz baze podataka su klasifikacija, grupiranje, regresija, umjetna Inteligencija, pravila pridruživanja, genetski algoritam, najbliži susjed i slični (Bharati, Ramageri, 2010).

Tu prikazuju se različite vrste metoda i tehnika koje se koriste.

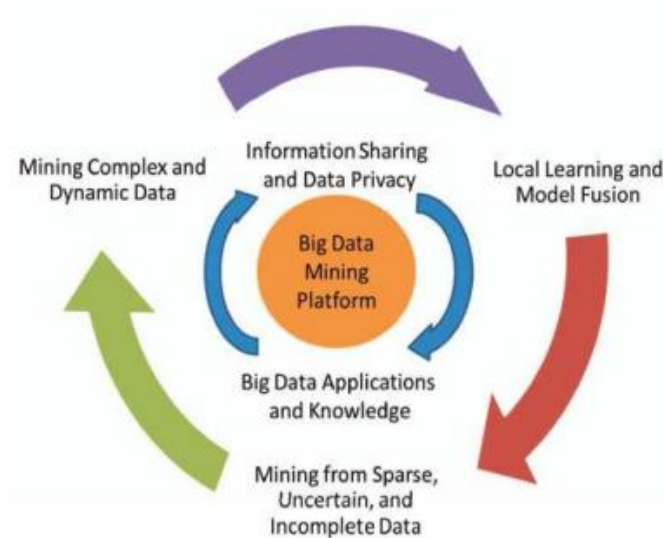
Neke od tih metoda su:

- Klasifikacija – Jedna od najzastupljenijih metoda ili tehnika rudarenja podataka je klasifikacija. Takva tehnika se primjenjuje nad skupom podataka koji je prethodno klasificiran (Bharati, Ramageri, 2010).  
Klasifikacija spada pod jednu od najčešćih metoda rudarenja podataka. Pod klasifikacijom najčešća metoda uporabe je metoda o izgradnji stabla odlučivanja. Unutar stabla odlučivanja primjenjuju se „IF-THEN“ pravila kako bi mogli predvidjeti rezultat podataka. Kako bi se ispitala točnost pravila kvalifikacije provodi se klasifikacijski test pomoću podataka. Ako se test prikaže točnim pravila se mogu primijeniti i na sljedeći dio podataka.
- Grupiranje – Grupiranje je metoda rudarenja podataka gdje se podaci koji imaju najviše zajedničkih karakteristika skupa spajaju u jednu grupu. Također omogućava i prikazuje grupiranjem pregled razlike između podataka, a ne samo zajednička svojstva.



- Pravilo asocijacije – Kako bi se prepoznala podudarnost između poznatih podataka i glavnog, primjenjuje se pravilo asocijacije. Pomoću te metode pronalazi se korelacija te odnos između složenih podataka (Wanchun 2017).  
Glavna zadaća asocijacije je da otkrije ako postoje veze povezanosti između podataka. ako se unutar baza podataka određeni podaci često sinkrono pojavljuju, to možemo nazvati uzorkom te pretpostavljamo da postoji određena veza povezanosti između tih podataka. Naime unutar pružene grupe podataka pravila za asocijaciju ima jako puno te nema svako pravilo vrijednost kojim bi se moglo pojačati uspješnost ili pronaći vrijednost.
- Neuronske mreže – Velika količina podataka se obrađuje pomoću matematičkog modela koji je u mogućnosti napraviti kompliciranu analizu podataka, takav model je neuronska mreža (Wanchun 2017).

Prilikom rudarenja podatka kod svakodnevnih sustava koji su namijenjeni za rudarenje podataka potrebne su računalne jedinice za analizu te usporedbu. Računalna platforma zahtijeva imati dostupan djelotvoran pristup podacima te računalnim procesorima, te za jednostavni rudarenje podataka potreban je običan PC s tvrdim diskom te CPU procesorom, međutim kod rudarenja velikih podataka to nije slučaj. Za rudarenje velikih podataka potrebno je računalo visokih performansi računalne platforme. Ta platforma će se postaviti pomoću alata paralelnog programiranja poput MapReduce (Wu i sur. 2014).



*Slika 4 Proces obrade velikih podataka*

*Izvor: Jadhav, 2013*

Za veoma napredan sustav koji je sposoban upravljanjem velikih podataka, veoma je potrebno da količina podataka bude u iznimno velikoj količini. Konceptualni prikaz toga prikazan je pomoću slike 4. Proces obrade velikih podataka, gdje su prikazana 3 sloja pristupa podataka, privatnosti podataka te algoritmi rudarenja velikih podataka. Na trećem sloju rudarenja velikih podataka cilja se na algoritme koji su zaslužni za rješavanje poteškoća koje se javljaju kod velikih količina podataka te različitih kompliciranih karakteristika podataka. Prvi dio trećeg sloja predstavlja različite podatke koji su obrađeni tehnikama fuzije podataka. Nakon toga podaci se pretražuju nakon procesiranja te će se bitni podaci vratiti na fazu prije procesiranja te raspodijeliti na odgovarajuće mjesto, dok se prvi sloj fokusira na određeno pristupanje podacima te drugi sloj na različite aplikacije primjene rudarenja podataka (Jadhav, 2013).

### 3. Veliki podaci u turizmu

U prošlom poglavlju obrađen je uvod u velike podatke. Unutar ovog poglavlja obradit će se vrste podataka koji se koriste te su najviše zastupljeniji unutar turizma. Trenutno se unutar turizma koriste veoma raznoliki podaci koji se mogu koristiti na različite načine. Veliki podaci trenutno unutar poslovanja poduzeća imaju veliki potencijal. Razna istraživanja te alati doprinose kako bi se povećala količina implementacije velikih podataka (Hela, 2020).

Takve podatke koji su najviše zastupljeniji u turizmu mogu se podijeliti na tri glavne kategorije. Najveća razlika između kategorija je u suštini izvor od kuda su prikupljeni podaci te lakoća prikupa određenih podataka (Peters, Keller, 2022). Te kategorije su:

- korisnički generirani sadržaj
- podaci uređaja
- transakcijski podaci

Četiri kategorije koje služe kako bi se podaci najbolje predstavili unutar turizma su podaci tražilice, web stranice, korisnički generirani sadržaj, podaci uređaja. Te vrste podataka uvelike omogućavaju lakše shvaćanje podataka u turizmu.

#### **Korisnički generirani sadržaj (UGC)**

Korisnički generirani sadržaj (UGC), sastoji se od podataka sadržaja koji je generiran od strane korisnika s društvenih mreža. Takvih podataka ima jako puno, ali nisu svi uvijek dostupni. UGC podaci su jedno od najčešćih, najzastupljenijih i najtraženijih podataka u turizmu. Izvor od kuda se takvi podaci najviše prikupljaju jesu socijalne mreže. Podaci prikupljeni sa socijalnih mreža poput Instagram, su veoma korisni za buduća predviđanja turizma (Peters, Keller, 2022). Korisnički generirani sadržaj ili UGC, mogu se podijeliti na dvije grupe :

- Online tekstualni podaci
- Online fotografski podaci

## Online fotografski podaci

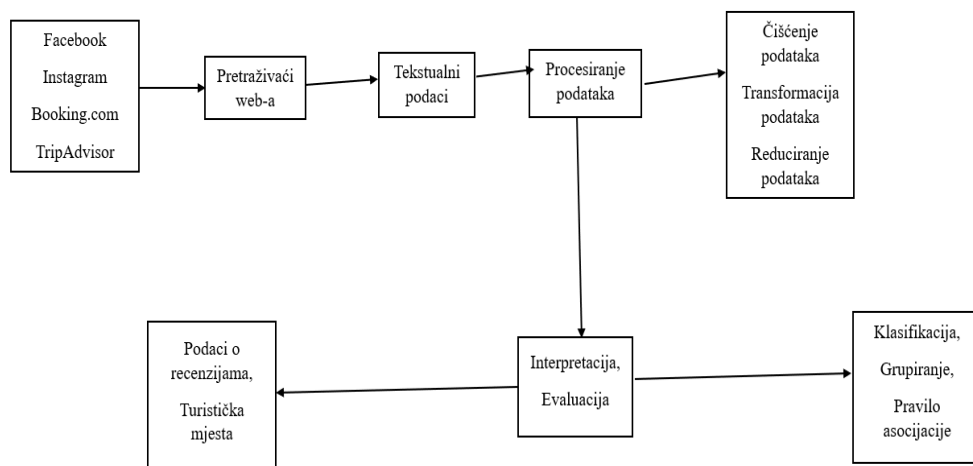
Jedna vrsta podataka koja je pružila veliki utjecaj su fotografski podaci, tj podaci fotografija na web stranica kako bi se najlakše prepoznale turističke destinacije. Fotografije kao podaci unutar turizma se najviše koriste pri marketingu kako bi se povećala pozornost turistima te sve ih se više privuklo (Hela, 2020).

Primjena podataka od fotografija s mreža je postalo veoma korisno kako bi se mogli predvidjeti potencijalni turisti, te predložiti buduće lokacije (Nitin, dr. Yadav, 2018).

Kao izvor prikupljanja fotografskih podataka najjači je Flickr zbog toga što omogućava pristup raznim API-evima koji su nam korisni i koji omogućavaju da pristup podacima bude veoma lakši. Prilikom prikupljanja podataka fotografija postoje zastupljene kategorije koje su od velike važnosti. Kategorije koje su zastupljene i korisne nam u prikupljanju u vrijeme kad je slika napravljena te kada je objavljena na internet, također podaci korisnika koji je objavio sliku te dodatne tekstualne informacije koje je korisnik uz tu sliku priložio.

## Online tekstualni podaci

Druga kategorija korisničkog generiranog sadržaja su tekstualni podaci. Budući da turizam i tehnologija veoma brzo napreduju, razvijaju se aplikacije koje omogućavaju turistima da iznesu svoje kriterije preko raznovrsnih platformi. Tekstualni podaci koji se prikupljaju mogu biti preuzeti s raznih izvora, pomoću pretraživača web-a koji će nam omogućiti da preuzmemo te podatke s izvora poput Facebook, Instagram, TripAdvisor, Booking.com (Peters, Keller, 2022). Proces započinje prikupljanjem tekstualnih podataka s raznih stranica. Drugi dio procesa spada analiza tih informacija koje su se prikupile, te to se dijeli na dvije pod faze. Prva faza je obrada podataka dok je druga procjena interpretacije (Nitin, dr. Yadav, 2018). Navedeni proces je prikazan pomoću slike 5. Procesiranje tekstualnih podataka.



*Slika 5 Procesiranje tekstualnih podataka*

*Izvor: Obrad autora prema: Nitin, Dr. K.P. Yadav*

## Podaci uređaja

Kako bi se prikupili podaci uređaja, o lokaciji i mobilnosti koristi se snimka sustava globalnog pozicioniranja ili GPS. Kako bi se uspješno povećao obujam prikupa podataka za turističku industriju, primjenjuje se prikupljanje podataka od različitih izvora koji god to mogu omogućiti. Trenutno od svih GPS spada pod najkorišteniji podataka u primjeni unutar turizma, a kako sve više napreduje smatra se kako će se kvaliteta podataka samo pojačavati. Vrste podataka koji se primaju od uređaja su uglavnom strukturirani i ne strukturirani podaci. Postoje razni načini kako bi se mogli prikupiti podaci, poput praćenja iznajmljenih bicikala, praćenje iznajmljenih vozila te iznajmljenih brodova (Peters, Keller, 2022).

Podaci uređaja su primarna kategorija prikupljanja podataka a pod nju mogu spadati GPS, mobilni podaci, Bluetooth (Hela, 2020).

## Transakcijski podaci

Transakcijski podaci veoma mogu pomoći za lakše shvaćanje ukupne potrošnje turista, vrijeme kada kupuje te učestalost i vrste kupnji. Takve informacije su veoma bitne kada je u

pitanju ciljanje određenih skupina turista i razvijanje turizma koji će se moći pružiti turistima ono što im je potrebno.

Primjeri transakcija mogu biti izvučeni iz različitih izvora. Evidencija tih transakcija može proizaći iz kupnje hrane, zrakoplovnih karti, smještaja (Hela, 2020).

Kada se promatra kroz nekoliko raznih kategorija potrošnje kako bi se najbolje analiziralo podatke, uz korištenje transakcijskih podataka najbolje je primijeniti prediktivnu analizu podataka.

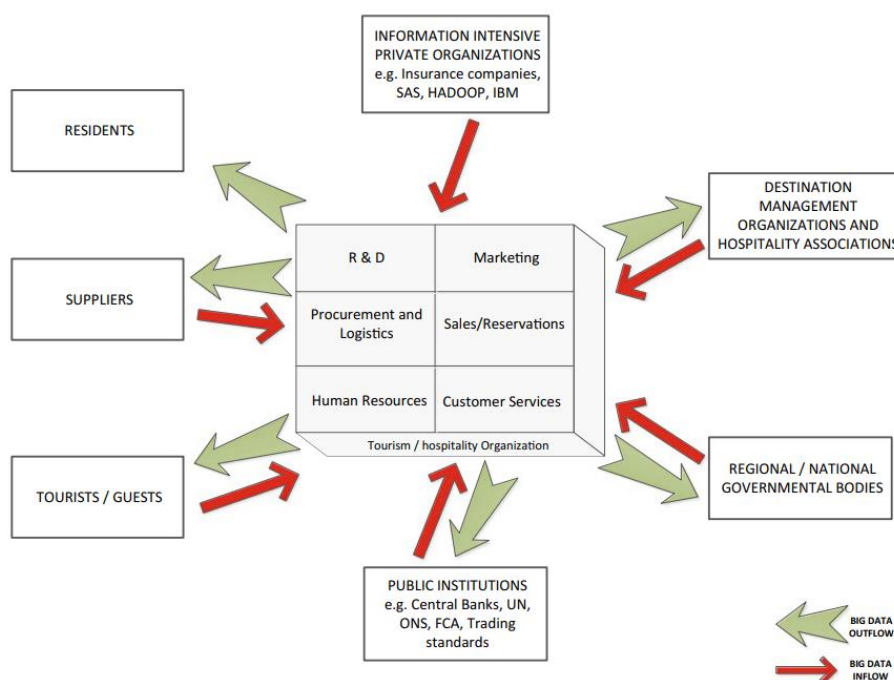
### 3.1 Turizam

„Turizam (engl. tourism), ukupnost odnosa i pojava koji proizlaze iz putovanja i boravka posjetitelja nekog mjesta, ako je takvo putovanje poduzeto radi odmora i uživanja te se njime ne zasniva stalno prebivalište i ne poduzima se neka gospodarska djelatnost“ (Hrvatska enciklopedija, 2021). Turizam je u ljudskoj povijesti uvijek postojao, ali sa sve većim napretkom zrakoplovne industrije putovanja su sve češća. Takav napredak veoma utječe na razna turistička središta pozitivno, ali i negativno (Theobald, 2004). Turizam više ne čine samo arheološka mjesta, poznati spomenici i razni prirodni fenomeni, već napretkom društva kako bi se privukli turisti sve više se razvijaju razni hoteli, restorani. Turizam pridonosi društvu u raznim aspektima, poput ekonomske profitabilnosti te socijalne profitabilnosti. Turisti koji posjećuju mjesto vrlo je vjerojatno da će potrošiti novac na lokalne proizvode ili usluge, te time pridonijeti ekonomskom razvitku tog mjesta. Kako bi se turisti lakše kretali vrlo je vjerojatno da će lokalne vlasti osigurati razne načine prijevoza te će se time i poboljšati promet mjesta. Također putovanjem turista ne samo da se razmjenjuje novac već se i razmjenjuju razne kulture ljudi koji se susreću. To može rezultirati prikazivanju ne toliko popularnih kultura koje će sada izaći na vidjelo te zainteresirati razne ljude za daljnje istraživanje (Revfine, 2023).

### 3.2 Vrste podataka u turizmu

Prethodno su navedeni tipovi velikih podataka koji su najviše zastupljeniji u turizmu, te sada prikaz velikih podataka u turizmu. Kako svaka firma u turizmu koristi velike podatke, tako moraju imati i razvojni okvir tih podataka kako bi to bilo sve učinkovitije. Ovdje su

prikazana dva razvojna okvira. Prvi razvojni okvir je zaslužen za prikazivanje odnosa između podataka iz raznih izvora te turističke firme. Kako bi se primjenjivao pristup velikim podacima te mogućnost njihovog korištenja organizacijama su potrebni skupovi razvojni okvira pomoću kojih će se to omogućiti (Nikolaos, Jeremy 2019). Unutar centra prvog okvira prikazane su organizacije turizma. Prvo su prikazani međusobni odnosi unutar organizacija (Nikolaos, Jeremy 2019). Navedeno je prikazano pomoću slike 6. Protok velikih podataka u turizmu.



Slika 6 Protok velikih podataka u turizmu

Izvor: Nikolaos, Jeremy 2019.

Izlazni i ulazni podaci prikazani su strelicama, gdje zelena prikazuje izlazne podatke a crvena ulazne. Pod crvenim strelicama podrazumijeva se priljev velikih podataka od strane raznih subjekata koji doprinose poduzeću dostavljajući određene proizvode. Dok se pod crvenim strelicama mogu shvatiti podaci koju su obrađeni od strane raznih institucija, koji sadrže informacije poput koliko noćenja je provedeno unutar hotela ili koliko soba je rezervirano od strane turista (Nikolaos, Jeremy 2019).

Primjena velikih podataka može se usredotočiti na priljev velikih podataka koji je analiziran primjenom prediktivne analitike iz količina podataka na temelju ponavljajućih uzorka zaliha

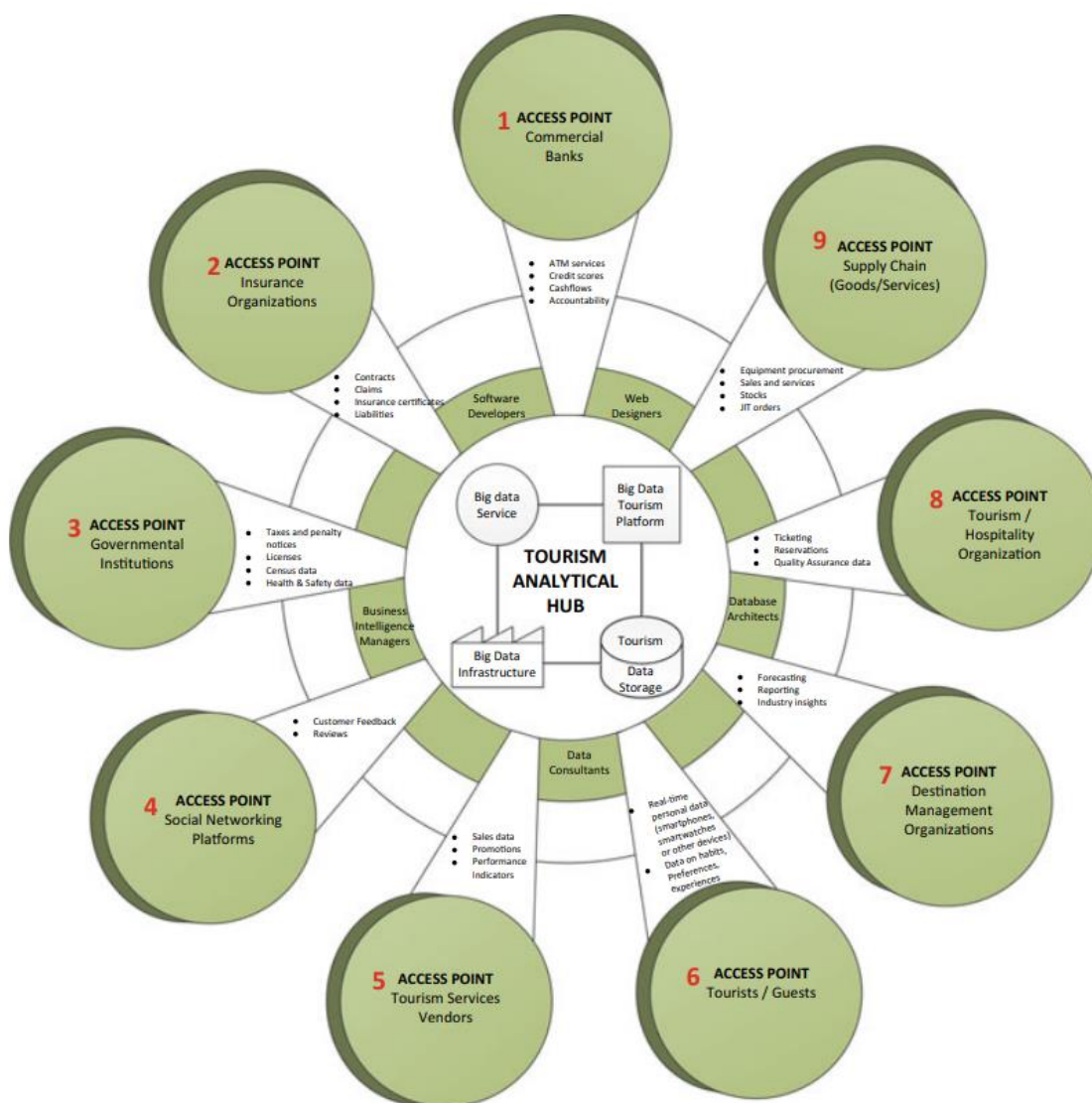
pružajući bogatije informacije za pružanje prilagođene usluge na temelju vrste gostiju ili kupaca u turističkoj ili ugostiteljskoj organizaciji (Nikolaos, Jeremy 2019). Implementacijom i dobrim kreiranjem razvojnih okvira, pomoću kojih će se upravljati podacima, pružit će se učinkovita sigurnost zaštite podataka (Anca, Hugues, 2020).

S obzirom na prikupljene podatke turista koji su prethodnih godina dolazili, predložit će im se nove opcije generirane preko tih podataka. Tu se kod izlaznih podataka prikazuje profit te kvaliteta uspješnosti. Dobiva se povratna informacija da li su se prikupljeni veliki podaci efektivno iskoristili. To ide u korist različitim partnerima koji su u tome sudjelovali.

Razvojni okviri koji su unutar turizma predstavljeni za upravljanje podacima, moraju proširiti svoje vidike s privatnosti i etičkih rješenja na pravednu i etičku razmjenu podataka i informacija (Anca, Hugues, 2020).



## Analitički razvojni okvir



Slika 7 Analitički razvojni okvir velikih podataka u turizmu

Izvor: Nikolaos, Jeremy 2019.

Prikazani analitički okvir je napravljen od 4 sloja, te središnjeg dijela gdje je glavna sredina obrade velikih podataka. Prvi sloj prikazuje 9 pristupnih točki od kuda se mogu prikupiti podaci. Svaki dio ima svoj vitalan element kojim pomaže u organizaciji. Kako bi se financijska inteligencija uspješno mogla obraditi te analizirati, banke pružaju podatke transakcije te elektronskih plaćanja (Nikolaos, Jeremy 2019). Drugi sloj je zaslužan za

prikazivanje za prikazivanje ključnog osoblja koji je osposobljen za analitičke sposobnosti pomoću kojih transformiraju podatke. Unutar tog sloja na osnovi transformiranih podataka donese se odluke koje utječu na daljnje poslovanje poduzeća. Treći sloj prikazuje vrste podataka koji se nalaze. Podaci koji se pružaju organizacijama interno i eksterno su veoma bitni kako turizam ne bi bio ograničen da bude učinkovit (Nikolas, Jeremy 2019). Zadnji tj. četvrti sloj prikazuje analitičko središte. To se bazira na potrebi za velikom količinom podataka kako bi se mogli analizirati te donijeti što bolja kvaliteta hotelijerstvu.

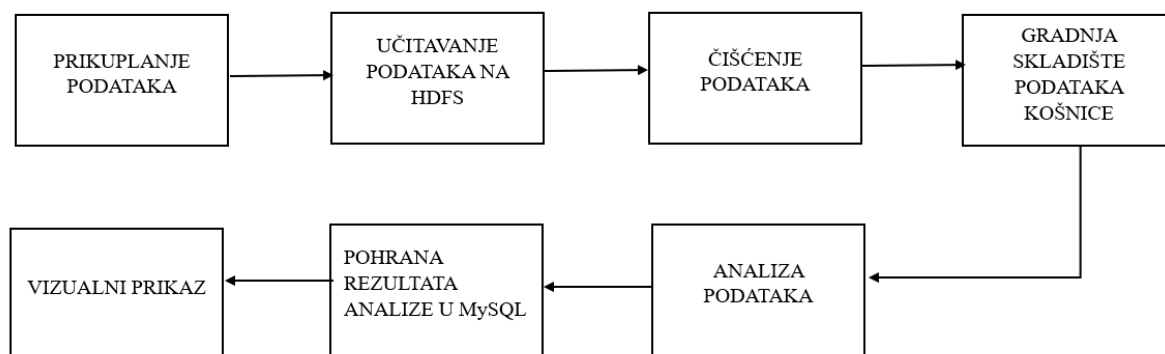
Turizam analitičko središte uključuje uslugu velikih podataka koja pruža podatke koje mogu prikupljati unutarnje ili vanjske organizacije; platforma velikih podataka kao što je integracijska platforma kao usluga (iPaaS) i blockchain ili drugi vlasnički softver koji podržava obradu i pristup velikim skupovima podataka bez potrebe za instaliranjem hardverskih uređaja ili međuprograma; i velika pohrana podataka kao što su skladišta podataka i poslužitelji (Nikolaos, Jeremy 2019). Navedeno je prikazano pomoću slike 7. Analitički razvojni okvir velikih podataka u turizmu.

## 4. Primjeri primjene velikih podataka u turizmu

### *Primjer 1: Analitička platforma hotelskih velikih podataka*

U ovom poglavlju obradit će se implementacija velikih podataka u turizmu koristeći neke tehnologije velikih podataka poput prethodno objašnjenog Hadoop-a, te Sparka, Amazon S3 usluga te klastera za prikupljanje podataka.

Kao što je spomenuto, razvojni okvir prve implementacije je pomoću Hadoop i Spark. U generalnoj obradi velikih podataka spomenute su komponente Hadoop-a koje se primjenjuju a to su HDFS i MapReduce. Sustav za analizu velikih podataka je prikazan procesom prikupljanja podata, pohranjivanja i čišćenja podataka, analiza i vizualizacija podataka (Dianwei, 2021). Prikazano na slici 8. Proces dizajna hotelske platforme velikih podataka.



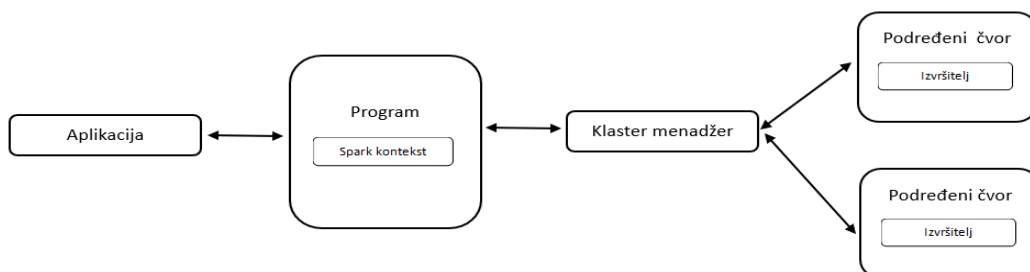
*Slika 8 Proces dizajna hotelske platforme velikih podataka*

*Izvor: Obrada autora prema: Dianwei, 2021.*

Dizajn cijelog procesa je zamišljen u dva dijela gdje je prvi dio zadatak da se dizajnira pretraživač za prikupljanje svi podataka od hotela, te da se podaci spremu u csv format. Drugi dio procesa je učitavanje podataka na Hadoop. Nakon što se podaci učitaju, koristeći hdfs i Python će se implementirati podaci. Kako bi se podaci filtrirali i očistili za analitičku platformu velikih podataka primijenit će se spark program (Dianwei, 2021).

Budući da se ovdje koristi Spark program za upravljanje velikim podacima, vanjske sustave skladištenja poput HDFS je važno držati blizu Spark sistema zbog toga što Spark sistemi očitavaju podatke iz istih. Spark program se sastoji od glavnog čvora čija je zadaća da se

poziva glavni dio aplikacije. Program se uglavnom sastoji od koda koji je napravljen od strane korisnika. Klaster menadžer te program skupa obavljaju proces unutar klastera, gdje će se nakon toga razdvojiti na podređene čvorove koji izvršuju zadatke koji je menadžer klastera zadao (Mohiuddin, Nahvi 2019).

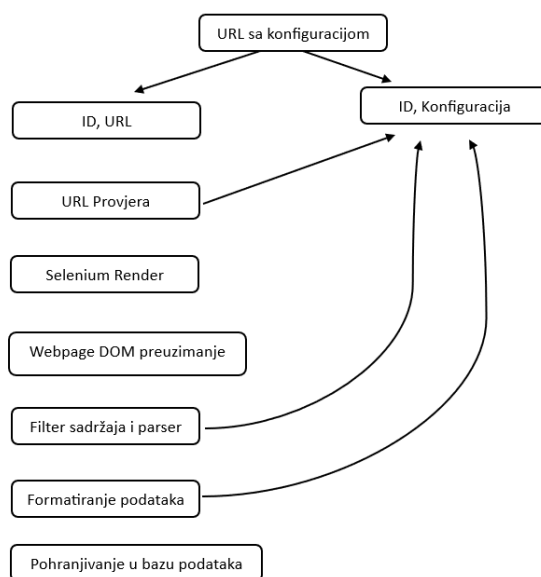


*Slika 9 Spark Arhitektura*

*Izvor: Obrada autora prema: Mohiuddin, Nahvi 2019*

Kako bi se izgradila i analizirala baza podataka moraju se kreirati košnice vanjskih tablica koje će se bazirati na putanjama u skladištu, koja imaju 2 dijela u hdfs. Prva tablica zastupa podatke komentiranja gostiju (hotel\_data), dok druga zastupa osnovne informacije o hotelu (hotel\_basic). U zadnjem dijelu obrađuje se vizualizacija podataka, prema kako bi se izgradio projekt koristi se SpringBoot, mybatis i Mysql te IDEA za razvojni alati. Ajax asinkronous se koristi kako bi se uspješno poboljšala kvaliteta stranice i korisnikovog iskustva (Dianwei, 2021).

Prilikom realizacije analitičke platforme velikih podataka primijenjen je modul pretraživača. Takav modul može se podijeliti na četiri dijela, gdje je prvi modul, koleksijski modul, takav modul je korišten da se analiziraju JSON podaci, ti podaci si prikupljeni od modula preuzimanja. Nakon što se podaci analiziraju preuzet će se odgovarajući podaci te izvori veza koji imaju konekciju. Drugi modul cjevovoda podataka služi kako bi se mogli formatirati podaci, za one koji su potrebni. Pod trećim modulom prelazi se na modul skladišta podataka, gdje se podaci koji su obrađeni pretraživačem spremljeni u određeno mjesto u skladištu podataka. Te zadnji modul, IP proxy modul, služi kako bi se moglo prikupiti IP proxye, da bi se omogućilo pretraživačima da ih ne blokiraju razne stranice. Nakon što je izgrađen klaster hadoop platforme, potrebno je instalirati hive i mysql. Potrebno je učitati csv datoteku s podacima koji su prikupljeni unutar hotela (Dianwei, 2021).



*Slika 10 Model pretraživača*

*Izvor: Obrada autora prema: Chaulagain i sur. 2017*

Kod modela pretraživača zahtjevom korisnika započinje pretraživanje priloženjem URL-a, gdje se konfiguracija sastoji od parser konfiguracije poput Xpath ili CSS izbornika te Regex-a. Unutar baze podataka se pohranjuju JSON podaci od url-ova te konfiguracije, te će iz tog JSON datoteke pretraživač pretražiti URL-ova i konfiguraciju. Prema selenium rendereru se prenosi ispravni URL, dok se neispravni prenose korisniku za informaciju. Koristeći EC2 mašinu MapReduce se može integrirati unutar pretraživača te se rezultat podataka može iskoristiti za prikaz vizualizacije (Chaulagain i sur. 2017).

Kako bi se izgradila košnica tablice osnovnih informacija o hotelu te podataka gostiju prilikom izgrađivanja skladišta podataka te analize podataka postoji određeni broj koraka. (Dianwei, 2021) Skripta pomoću koje se izgradi tablica podataka recenzija gostiju je:

„create vanjska tablica hotel\_podaci (korisnik\_ime string, hotel\_ime string, putovanje\_tip string, vrijeme1 string, vrijeme2 string, korisnik\_rezultat double) ROW FORMAT DELIMITED FIELDS TERMINATED BY',' LOCATION'/hotelpodaci' “ (Dianwei, 2021).

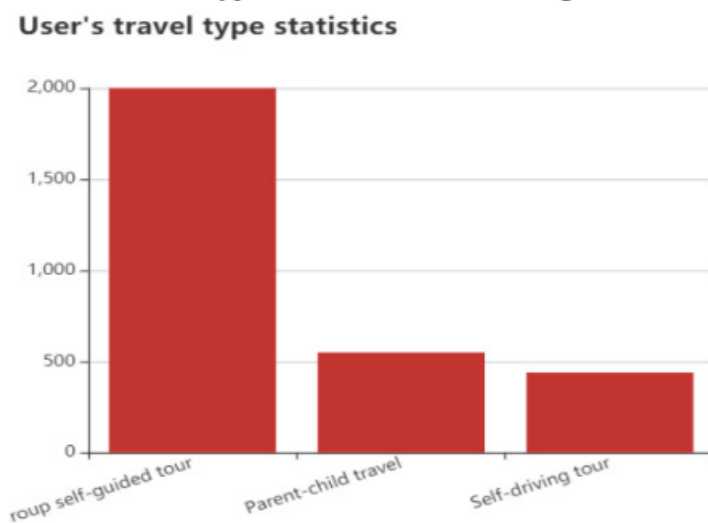
Nakon kreiranje tablica podataka korisnika, kreira se tablica o osnovnim podacima hotela (Dianwei, 2021). Skripta kojom se kreira tablica podataka hotela je:

„create vanjska tablica hotel\_osnove (id string, ime string, rezultat double, commentnum int, preporučeno int, adresa string, zvjezdice string, detalji zvjezdica string) )ROW FORMAT DELIMITED FIELDS TERMINATED BY',' LOCATION'/hotelosnove' “ (Dianwei, 2021).

Završetkom procesa, dolazi se do analiziranja svih podataka koji su prikupljeni te se podaci dostavljaju u mysql tablicu. Zadnji dio procesa je vizualizacija. Vizualizacija će se kreirati s obzirom na rezultate svih podataka koji su prikupljeni (Dianwei, 2021). Korišten je JavaEE razvojni okvir SpringBoot. Kod koji je korišten kako bi se asinkorono zatražio kod podataka je priloženi:

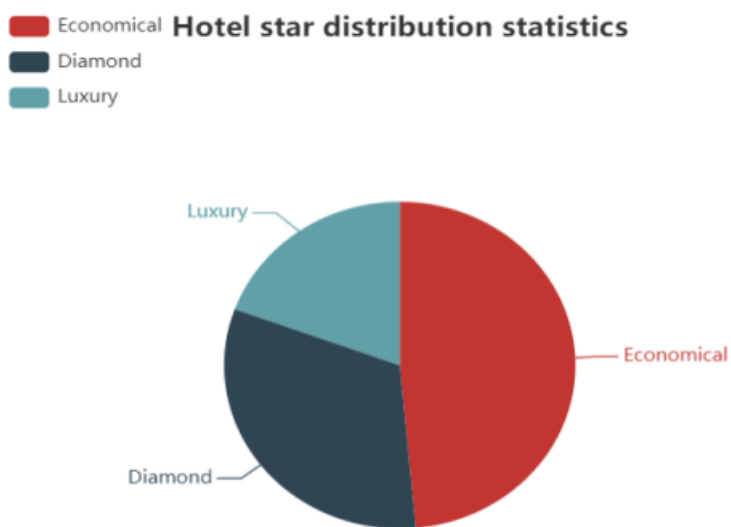
```
„ $.ajax({  
  type:"GET",  
  url:"/areastat",  
  dataType:"json",  
  async:false,  
  success:function (result) {  
    json = result;  
    for (var i = 0; i <result.length; i++){  
      var ob = {name:"",value:""};  
      ob.name = result[i].area_name;  
      ob.value = result[i].nums;  
      datatemp.push(ob);  
    }  
  },  
}); „ (Dianwei, 2021).
```

Postignuti rezultati su prikazani pomoću eecharts-a.



Slika 11 Statistika podataka korisničkih putovanja

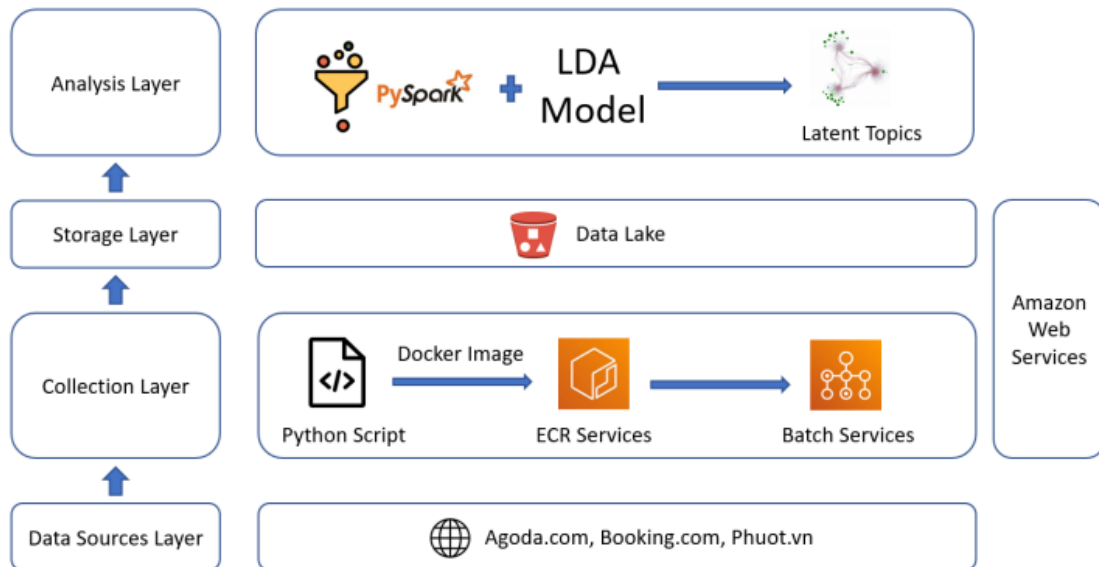
Izvor: Dianwei, 2021



Slika 12 Statistika hotelskih zvjezdica

Izvor: Dianwei, 2021

**Primjer 2: Model za obradu velikih podataka u turizmu prema recenzijama Booking-a i Agode**



*Slika 13 Modelni okvir za obradu velikih podataka*

*Izvor: Nguyen i sur.2023*

Prikazani model podijeljen je na 4 sloja. Prvi sloj su izvori od kuda su podaci prikupljeni. U drugom sloju prikazan je algoritam prema kojem su prikupljeni podatci s različitih platformi. Podaci koji su analizirani su prikupljeni bili s platforma Booking.com i Agoda. Na temelju Python koda provedeno je prikupljanje podataka. Budući da su te vrste podataka ne strukturirane, izabrano je u trećem sloju jezero podataka za pohranu podataka koristeći tehnologiju Amazon S3, te zadnji sloj gdje su vizualno prikazali rezultate. Prikazan je skup podataka koji je predstavljen sa PySpark tehnologijom (Nguyen i sur. 2023). Prikazano pomoću slike 14. Modelni okvir za obradu velikih podataka. Podatke koje su prikupljeni te iskorišteni su od strane korisnika preko recenzija koje su ostavili na platformama booking-a te agode. Za svaku platformu korišteni su određeni algoritmi, priložen je algoritam prema kojem su prikupljeni podatci sa agode.

„Algorithm 1: Data Collection Algorithm

$M = [ ]$

Initialize a list of 63 provinces and cities in Vietnam



P = ['an giang, 'ba ria – vung tau', 'bac lieu' ... 'ha noi', 'ho chi minh' ...]

for p in P do

(1) Information about the province/city you are looking for

H = Gather a list of hotels in the province/city p

for h in H do

(2) Hotel details

- Address

- Images

- Utilities

- ...

R = Gather a list of hotel customer feedback

for r in R do

(3) Customer feedback r

- Room type

- Travel type

- Response date

- Feedback (General, Positive, Negative)

- Rating

- ....

M += {(1), (2), (3)}

end for

end for

end for „ (Nguyen i sur. 2023).

Prvi korak u procesu prikupljanja podataka je identifikacija ciljne web stranice i podataka koje prikuplja. Nakon određivanja cilja, web stranica se pretražuje i prepoznaje se njezina struktura. Sljedeći je korak odrediti metodu indeksiranja i konfigurirati okruženje za pregledavanje weba. To uključuje odabir preglednika, stvaranje HTTP upita i postavljanje ulaznih parametara za alat ili web pretraživanje. Koristeći upite HTTP zahtjeva omogućiti će se pristup web stranici. Kako bi se podaci analizirali, koristiti će se alati koji su korisni prilikom pretraživanja weba. Analizirani podaci se zatim pohranjuju u bazu podataka ili

datoteku za kasniju upotrebu. Naposljetku, kod je optimiziran za indeksiranje više web stranica i svih grešaka koje se dogode tijekom procesa struganja weba (Nguyen i sur. 2023).

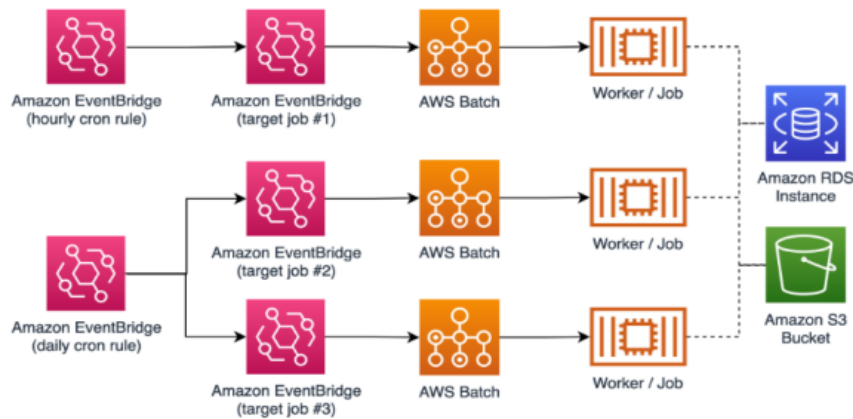
<b>BookingWebScraping</b>
- locationList : list - hotelList : list - language : string - location : string - directory : string - locationHeader : string - hotelHeader : string - source : string - inputFilename : string - country : string - countryName : string - web : string - cityKeyword : dict - hotelKeyword : dict - propertyTags : dict
+ getLocation(region : string, language : string) : list + getPropertiesReview() : list + addLocation(hotelID : int, hotelName : string, hotelURL ... + getPropertiesList(locationID : int, location : string) : list + getLocationID(location : string) : int + hotel2CSV(reviews : list) + location2CSV(location : string, reviews : list) + getLocationReviews(location : string) : list + getSource() : string + getLocationsByCountry() : list + getPropertySource() + getPropertiesByCountry() : list + getCountryReviews()

*Slika 14 Primjer web pretraživanja booking-a*

*Izvor: Mendoza i sur. 2021.*

Mogu se prikupljati podatci iz identificiranih izvora, koji mogu biti ručni, poluautomatski ili automatizirani. Podaci se mogu nalaziti, između ostalog, na papiru, u datotekama, skupovima podataka, ontologijama, informacijskim sustavima, društvenim mrežama, sensorima, mobilnim uređajima i internetu. Putem prilagođene aplikacije moguće je dohvatiti podatke automatski ili poluautomatski, u paketnom načinu rada ili u stvarnom vremenu (Mendoza i sur. 2021). Prikazano je pomoću slike 12. Primjer web pretraživanja booking-a.

Booking.com i Agoda se služe različitim API-evima, koji su Rest API te GraphQL. To je rezultiralo da se metode moraju kombinirati kako bi se mogli izvući podaci. Općenito, u svakom konkretnom slučaju istraživanje se provodi API metodom ili metodom analize HTML koda, koja kombinira tehnike povezane s JavaScriptom za izdvajanje podataka (Nguyen i sur. 2023).



Slika 15 AWS Usluge za skupljanje podataka

Izvor: Denot et al., 2021

Koristeći AWS Batch usluge zakazuje se određeno vrijeme za izvršenje određenih zadataka koji se trebaju izvršiti. Unutar modela kako bi se skripte izvukle primjenjuje se Python kod. Unutar Docker slike se pohranjuju rezultirane komponente te se Docker slika šalje na ECR. Kako bi se ta Docker slika s ECR-a izvršila za to će se pobrinuti Amazon S3 usluga (Nguyen i sur. 2023).

Spremljeni podaci su pohranjeni u jezero podataka jer se podaci sastoje od različitih tipova podataka. Za pohranu tih podataka iskorišten je Amazon S3 usluga. Amazon S3 usluga omogućuje spremanje bilo kakvih podataka na svom cloud-u. Amazon S3 usluga je prikazana pomoću slike 13.



Slika 16 Amazon S3 usluga

Izvor: <https://aws.amazon.com/s3/>

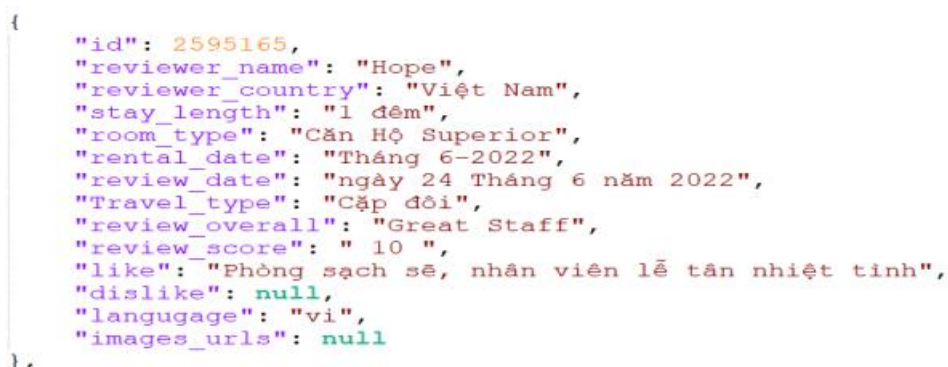
Prikazani rezultati su vrste podataka koji su izvučeni s platforme booking.com. Napravljena je struktura koja se sastoji od dva dijela, di je prvi dio koji pripada hotelu i njegovim podacima, a drugi dio spada pod podatke korisnika koje je on pustio unutar recenzije (Nguyen i sur. 2023).



Slika 17 Podaci hotela prikupljeni sa booking.com

Izvor: Nguyen i sur. 2023

Unutar podatkovnog jezera nalaze se podaci o hotelima, pod tim podacima spadaju naziv hotela, adresa hotela te recenzije i fotografije. Kako bi pohranio se put svake slike booking.com koristi domenu „<https://cf.bstatic.com>“ te se na to implementirao korijenski URL (Nguyen i sur. 2023). Navedeni podaci su prikazani pomoću slike 14. Podaci hotela prikupljeni sa booking.com.



Slika 18 Podaci korisnika prikupljeni preko recenzija

Izvor: Ho, Nguyen, Le, Le, Nguyen, Mai, Tran, & Trung 2023

Povratne informacije turista pohranjuje hotel. Informacije poput recenzija, datuma najma, vrste sobe, pozitivnih komentara te negativnih komentara. Ti podaci su pohranjeni u JSON

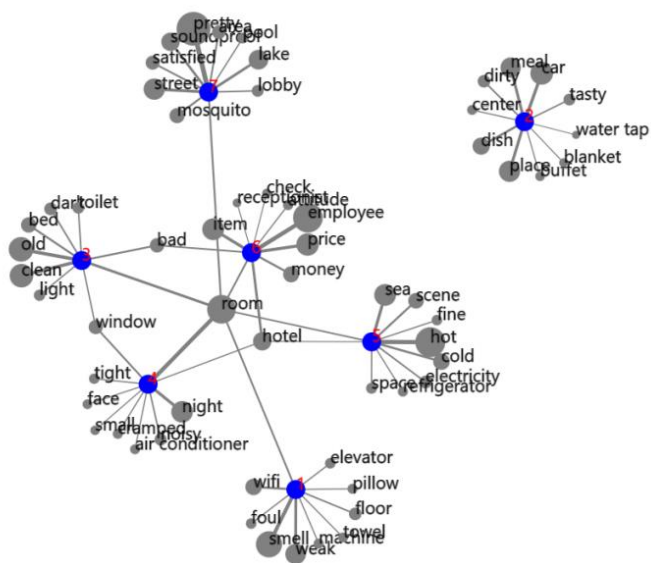
strukturi (Nguyen i sur. 2023). Navedeni podaci su prikazani pomoću slike 15. Podaci korisnika prikupljeni preko recenzija. Unutar istraživanja izdvojeno je 47 793 negativnih komentara s platformi. Ispitane su teme na koje se turisti žale za vrijeme njihovog boravka (Nguyen i sur. 2023). Kako bi procesirali podaci su primijenjena je PySpark tehnologiju, te primjenjuje se LDA model. Prikupljeno je više od milijun povratnih informacija kupaca s dvije platforme Agoda (763 911 komentara) i Booking (409 092 komentara). Također prikupljeno je za studiju 190.872 razmjene s web stranice Phuot.vn (Nguyen i sur. 2023). Navedeno je vidljivo pomoću priložene slike 16. Prikaz prikupljenih podataka.

Kriterij	Agoda.com	Booking.com	Phuot.vn
Vrijeme	7/2007 – 8/2022	6/2019 – 8/2022	5/2007 – 8/2022
Broj hotela(Popis članaka)	6,362	12,477	5,216
Broj komentara	763,911	409,092	190,872
Broj jezika	36	45	1
Broj komentara na Vijetnamskom/Engleskom/Nekategorizirani	93,942/ 433,600/ 4,039	89,925/ 55,640/ 191,115	190,872/ 0/ 0

*Tablica 2 Prikaz prikupljenih podataka*

*Izvor: Obrada autora prema: (Nguyen i sur.2023).*

Kako bi se prikazala povezanost između najčešćih riječi unutar istraživanja izgrađena je mreža grafikona za vizualizaciju (Nguyen i sur. 2023). Prikazano pomoću slike 17. Grafički prikaz najčešćih riječi.



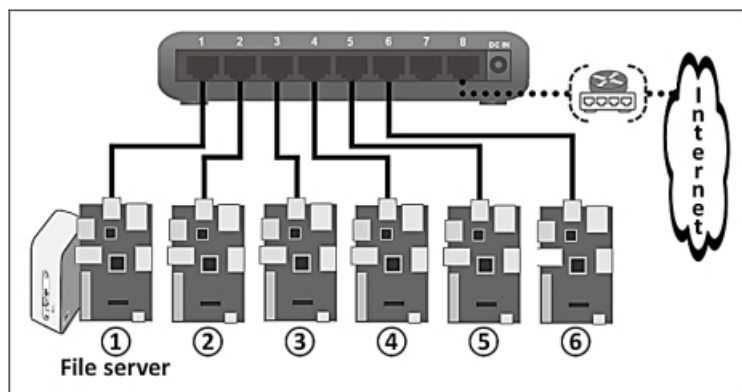
Slika 19 Grafički prikaz najčešćih riječi

Izvor: , Nguyen i sur. 2023

### **Primjer 3: Veliki podaci PI klaster**

U turizmu nemaju sve tvrtke sredstva da prikupljaju velike količine podataka te analiziraju ih za kvalitetna rješenja. Raspberry Pi, jeftino jednoplošno računalo male snage daje pristupačnu alternativu tradicionalnim radnim stanicama za zadatak koji zahtijeva malo računalne snage (d'Amore i sur. 2015). Sustav ima rješenje da radi u isto vrijeme paralelno s više adresa, te da se uspijeva riješiti problem koji imaju kompanije di prikupljanjem velikih količina podataka dođe do poteškoća s hardwareom. Oni su takav sustav opisali kao:

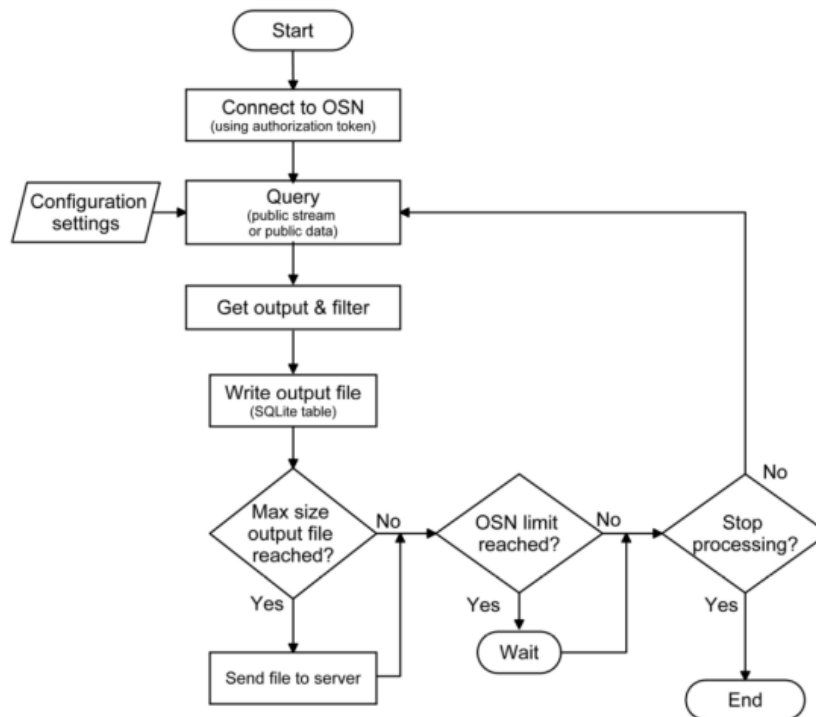
„Cilj je učiniti dostupnim jeftino rješenje koje je jednostavno i dovoljno fleksibilno i ne zahtijeva previše specijalizirane vještine, tako da mala organizacija ili tvrtka može postaviti personaliziranu analizu ili istraživački program prema autonomno odlučivanje koja će pitanja istražiti i koji podaci najbolje odgovaraju studiju ciljeva“ (d'Amore i sur. 2015).



*Slika 20 Raspberry Pi Cluster*

*Izvor: d'Amore i sur. 2015*

„Klaster se sastoji od šest Raspberry Pi modela B strojeva povezanih putem Ethernet prekidača“ (d'Amore i sur. 2015). Prikazano pomoću slike 18. Raspberry Pi Cluster.



Slika 21 Raspberry Pi klaster dijagram

Izvor: d'Amore i sur., 2015

Kako bi se s OSN-ova mogli preuzeti podaci pomoću skripta koje se dizajnirane za to prikazano je prema Slici 22. Raspberry Pi klaster dijagram. Programi koji su bili namijenjeni za preuzimanje podataka napravljeni su pomoću programerskog jezika Python. Budući da je dostupna velika količina programa koji omogućavaju korištenje te implementaciju raznih API-eva, uvelike je olakšan pristup raznim platformama (d'Amore i sur. 2015).

Provedena je demonstracija clustera gdje je unutar grada Lugano izabrana lokacija od 5km radijusa te to područje testirano za geolociranje ljudi unutar te turističke destinacije. Cluster je bio zadužen kako bi prikupio podatke s društvenih poput Facebook-a, Twitter-a ( trenutnog X-a), Instagrama (d'Amore i sur. 2015).

Za svaki izvor odabrane su sve objave (tweetovi, izvješća, slike) s geokodiranim oznakama (u području interesa). Dobiveni objekti se zatim grupiraju u popis lokacija (geografska širina i dužina) i ponderiraju prema broju elemenata koji se odnose na svaku lokaciju. Taj se popis zatim koristi kao ulaz u skriptu Heatmap.js Patricka Wieda ([www.patrick-wied.at/static/heatmapjs](http://www.patrick-wied.at/static/heatmapjs)). Kako bi se kreirala vizualizacija geokodiranih podataka korištena je Heatmap.js Javascript biblioteka (d'Amore i sur. 2015). Rezultati su prikazani pomoću



tablice 3. Točke interesa (POI) i spomeni prikupljeni za različite OSN-ove (d'Amore, Baggio, Valdani, 2015).

<b>OSN</b>	<b>POIs</b>	<b>Mentions</b>
Facebook	291	12923
Foursquare	15	57
Instagram	1958	4739
Twitter	24	1659
Total	2288	19378

*Tablica 3 Točke interesa (POI) i spomeni prikupljeni za različite OSN-ove*

*Izvor: d'Amore i sur. 2015*

Provedena vježba može biti od koristi bilo kakvim sudionicima unutar turizma ili menadžerima. Konstantnim provođenjem vježbe prikazat će se sve veći kvalitetniji broj podataka na nekome području (d'Amore i sur. 2015).

## 5. Prednosti i nedostaci korištenja velikih podataka u turizmu

### Prednosti

Veliki podaci iako su veoma korisni i efektivni unutar turističke industrije, kao i svaka druga tehnologija oni imaju svoje prednosti i nedostatke. Primjer za prednost velikih podataka je da omogućava lakše donošenje odluka što se tiče upravljanja prihoda. Moguće je predvidjeti buduće prihode s obzirom na analizu podataka koji su prikupljeni. Turizam i ugostiteljstvo kreiraju iznimno velike količine podataka pružatelja usluga te kupaca. Takve vrste podataka su veoma korisne kako bi se odradila analiza podataka za poboljšanje donošenja odluka na svim segmentima (Mariani, 2019).

Neke od prednosti velikih podataka su što nisu financijski zahtjevna tehnologija. Cijena naravno ovisi o količini ljudskih resursa koji su potrebni za brigu o upravljanju velikih podataka, ovisno koliko su velike količine podataka koje su pohranjene. Trenutno u svijetu pohrana strukturiranih podataka koja se dostavlja pomoću tehnologije velikih podataka, je od strane raznih organizacija viđena kao financijski najisplativija opcija. Primjer usporedbe troškova tvrtke, procijenjeno je da je trošak skladištenja jednog terabajta tijekom godine bio 37.000 USD za tradicionalnu relacijsku bazu podataka, 5.000 USD za uređaj baze podataka i samo 2.000 USD za Hadoop cluster (Thomas, Jill, 2013). Analiziranjem podataka postoji mogućnost predviđanja budućih turističkih kretanja. S obzirom na prikupljene podataka zaključuje se koje stvari najviše privlače turiste te su najviše korisnije, te s obzirom na to moguće je turistima pružiti stvari koje su ciljane za njihovo zadovoljstvo, a ne trošenje na stvari koje ne pridonose prihod.

Pozitivan čimbenik u korištenju aplikacije za analizu velikih podataka je ponuda pretraživanja informacija, sustav preporuka, dinamičko određivanje cijena i korisnička služba koja omogućuje interakciju sa zajednicom članova. Prikupljanjem različitih podataka u eri velikih podataka, kao što su geografska distribucija, emocionalne preferencije, ponašanje kupaca pri kupnji te društvene veze i hobiji, tvrtke mogu ciljati na potražnju, preferencije, odnose i druge načine da zadovolje kupce (Thi Mai, Shu-Yi, 2017).

Velikih izvor prihoda su društvene mreže tako da s oglašavanjem i primanjem podataka s društvenih mreža može se vidjeti koji turisti iz određenih dijelovi svijeta su najviše zainteresirani za određenu lokaciju. Trenutno unutar turizma upravljanje prihodima poput

popunjenosti rezervacija, lokalnim događajima te praznicima, sve više se povezuje učinkovitost uporabe velikih podataka (Anca, Hugues, 2020).

## **Nedostaci**

Glavni rizik kod korištenja velikih podataka je sigurnost. Budući da se koriste iznimno velike količine podataka potrebna je zaštita tih podataka. Firme su konstantno pod napadima hakera, te konstantno su u potrazi za dobrim analitičarima velikih podataka kako bi im oni pružali pravilno održavanje. Unutar 21. stoljeća primjer jedne od najvećih povreda podataka zadesilo je turističku tvrtku Marriott International. Takav propust rezultirao je utjecajem na 500 milijuna kupaca (Anca, Hugues, 2020).

Prilikom davanja privatnih informacija poput emailova, osobnih iskaznica ili kreditnih kartica predaju se informacije tvrtkama na raspolaganje gdje oni imaju kontrolu nad pristupom tih informacija, te je važno pohraniti i osigurati te informacije. Kako bi se razriješile sve zabrinutosti oko sigurnosti korisničkih podataka moraju se postaviti dobri razvojni okviri kojima će se definirati upravljanje podacima kako bi se mogli pružiti najviše mogući kvaliteti podaci, taj dio je najvažniji prilikom odabira analitičkih rješenja (Anca, Hugues, 2020).

Jedan od najvećih problema u poslovanju je curenje podataka van okvira za koji su osmišljeni. Curenje podataka bi za poslovanje imalo ogromne posljedice zbog financijske nadoknade svojim korisnicima, te potencijalno gubljenje svojih korisnika koji koriste njihove usluge čime bi rezultiralo financijskom kolapsu. Veliki problemi dolazi kod organizacija koje nisu sposobne da riješe i upravljaju velikim podacima te to rezultira u curenju podataka što će rezultirati u tome da treće strane imaju pristup tim podacima (Thi Mai, Shu-Yi, 2017).

## 6. Budući trendovi razvoja velikih podataka u turizmu

Zadatak budućih istraživanja je usmjeriti svoje ciljeve prema boljem shvaćanju velikih podataka u turizmu te način na koji se sve mogu koristiti. Podatke treba analizirati te shvatiti koja vrsta podataka koja se primjenjuje u turizmu nije dovoljno istražena, koja se najviše koristi, a koja vrsta podataka se najmanje koristi. Podaci uređaja, podaci na web stranicama nisu dovoljno istraženi, to uglavnom rezultira zbog spriječenosti pristupa podacima (Petters, Keller, 2022). Budući da podaci mogu imati osjetljive informacije, cilj sljedećih istraživanja može biti osiguravanje i zaštita podataka za sigurno korištenje unutar etičkih granica. Budući da su u ovom radu prikazani model prikupljanja podataka te vizualizacije, sljedeća istraživanja bi se mogla usredotočiti na sustav za preporuku ponuda prema prikupljenim podacima i njihov budući razvoj. Turizam nije jedina industrija koja ima mnoštvo ljudi već i zrakoplovna industrija, istraživanja bi se mogla posvetiti ako postoji korelacije između tih industrija te zajedničkih podataka koji bi se mogli iskoristiti za daljnji napredak. Također buduća istraživanja mogu prikazati statistiku velikih podataka, način i kako se koriste ovisno o funkciji te cilju poduzeća. Neki od problema koji bi se mogli istraživati u budućnosti na temu velikih podataka u turizmu:

- „Kako digitalizacija i veliki podaci oblikuju strukturu turističke industrije i institucionalne snage na djelu?
- Koji su glavni tehnološki, organizacijski, strukturni i kulturni izazovi povezani s implementacijom i upravljanjem velikih podataka u turističkoj industriji?
- Koje su vještine i sposobnosti potrebne baviti velike podatke u turizmu u sljedećim godinama?
- Koji su kritični čimbenici uspjeha koji utječu na učinkovitost korištenja velikih podataka u turističkoj industriji?
- Koja su komplementarna ulaganja koja tvrtke u turističkoj industriji trebaju održati za iskorištavanje potencijala velikih podataka?
- Koji su mehanizmi koji stoje iza stvaranja strateške vrijednosti velikih podataka u turističkoj industriji? „ (Ardito, Raguseo, Cerchione, Del Vecchio, 2019).

Jedna od tehnologija koja veoma može poboljšati protok podataka te točnost recenzija unutar turizma je blockchain. Kombiniranje blockchain-a te umjetne inteligencije postiže se efikasno upotrebljavanje velikih podataka u financijske svrhe (Lana, Giungiato, Tricase, 2021).

Primjena umjetne inteligencije te velikih podataka unutar turističke industrije je veoma mala. Činjenica da je umjetna inteligencija veoma nova tehnologija, te unutar turizma postoji veoma puno prostora kako i gdje bi se primjena velike inteligencije te analiza velikih podataka mogla istraživati. Istraživanja bi mogla prikazati pomoću kojih aplikacija i kako umjetna inteligencija prikuplja podatke te kako analiza velikih podataka se kombinira sa prikupljenim rezultatima za stvaranje korisnog sadržaja (Stroumpoulis, Kopanaki, Varelas, 2022).

## 7. Zaključak

Veliki podaci su nova tehnologija koja se u posljednje vrijeme sve više koristi. Zbog napretka tehnologije veliki podaci se koriste u raznim sektorima. Trenutno sektori koji imaju najviše zastupljene velike podatke su financijski sektori, gdje velike firme koriste velike podatke za analizu tržišta. Upotreba velikih podataka se znatno povećala u turističkom sektoru, ali još nije dovoljno iskorišteno puni potencijal velikih podataka u turizmu, te smatram kako će se sve više upotrebljavati i veoma efikasno doprinijeti budućem razvoju. Smatram kako treba pridonijeti i pažnju načinu prikupljanja velikih podataka. Senzori su trenutno veoma efikasni te njihova integracija možemo veoma unaprijediti bilo kakvo prikupljanje podataka. U procesu analize veoma će pridonijeti Hadoop tehnologija. Hadoop je veoma koristan prilikom analiziranja podataka turista te rezervacija i recenzija. Također razvoj AI-a se ne smije zanemariti, sa sve većim napretkom smatram kako AI će imati veliku ulogu u analiziranju te upravljanju podataka. Svatko tko nije upoznat s velikim podacima, smatram kako će ovaj rad znatno olakšati put prema nadograđivanju znanja o velikim podacima, te da će se lakše shvatiti primjena i načini iskorištavanje velikih podataka u turizmu, ali gdje se i slični slučajevi mogu primijeniti na druge sektore koji imaju potencijal za iskorištavanjem velikih podataka.

## Literatura

1. Mousavian, S., Miah, S.J. & Zhong, Y. (2023). A design concept of big data analytics model for managers in hospitality industries. *Pers Ubiquit Comput*
2. M. Mariani, G. Di Fatta and M. Di Felice, (2019)"Understanding Customer Satisfaction With Services by Leveraging Big Data: The Role of Services Attributes and Consumers' Cultural Background".
3. Noviyanti, Isnaeni et al. (2020) 'Towards Big Data as Official Statistics: Case Study of the Use of Mobile Positioning Data to Delineate Metropolitan Areas in Indonesia'. 1 Jan.
4. Jingjing Li, Lizhi Xu, Ling Tang, Shouyang Wang, Ling Li, (2018) Big data in tourism research: A literature review, *Tourism Management*, Volume 68, Pages 301-323.
5. Sharma, Nitin, and K. P. Yadav. (2018) "Big Data impacts in Tourism Research." *Think India Journal* 21.1. 41-53.
6. Becha, M., Riabi, O., Benmessaoud, Y., Masri, H. (2020). Applications of Big Data in Tourism: A Survey. In: Yang, X., Wang, CD., Islam, M.S., Zhang, Z. (eds) *Advanced Data Mining and Applications. ADMA 2020. Lecture Notes in Computer Science()*, vol 12447. Springer, Cham.
7. Peters, Scott, and Peter Keller (2022) "Applications and issues of big data in tourism research." .
8. Ho, Ping-Tsan (2022)"Smart tourism recommendation method in southeast Asia under big data and artificial intelligence algorithms." *Mobile Information Systems*.

9. ITF (2021), Big Data for Travel Demand Modelling: Summary and Conclusions, ITFRoundtable Reports, No. 186, OECD Publishing, Paris.
10. Kimball, R. and Ross, M. (2013) The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. 3rd Edition, John Wiley & Sons, Inc., Indianapolis.
11. Sinha, Sapna & Bhatnagar, Vishal & Bansal, Abhay. (2017). A Framework for Effective Data Analytics for Tourism Sector: Big Data Approach. International Journal of Grid and High Performance Computing.
12. K. A. Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi and A. Sabri, (2021) "Hybrid recommender system for tourism based on big data and AI: A conceptual framework," in *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 47-55
13. Sigala, Marianna & Rahimi, Roya & Thelwall, Mike. (2019). Big Data and Innovation in Tourism, Travel, and Hospitality Managerial Approaches, Techniques, and Applications: Managerial Approaches, Techniques, and Applications
14. Zhu, Wanchun (2017) "Application of Data Mining Technology in Tourism Information Recommendation." *2017 2nd International Conference on Materials Science, Machinery and Energy Engineering (MSMEE 2017)*. Atlantis Press.
15. Perwej, Yusuf, et al (2017) "An empirical exploration of the yarn in big data." *International Journal of Applied Information Systems (IJ AIS)* 12.9. 19-29.
16. Jagadale, Rupali, and Pratibha Adkar(2018) "A Review Paper on Big data Hadoop." *International Journal on Recent and Innovation Trends in Computing and Communication* 6.5. 131-135.



17. Beakta, Rahul (2015) "Big data and hadoop: A review paper." *International Journal of Computer Science & Information Technology* 2.2. 13-15.
18. Klievink, Bram, et al (2017) "Big data in the public sector: Uncertainties and readiness." *Information systems frontiers* 19.2. 267-283.
19. Al-Khasawneh, Mahmoud. (2020). Big Data Applications and Tools.
20. Banik, Abhinandan & Bandyopadhyay, Samir. (2016). Big Data-A Review on Analysing 3Vs.
21. Ormandjieva, Olga et al (2020) "Measuring the 3V's of Big Data: A Rigorous Approach." IWSM-Mensura.
22. An Overview of Data Mining Models (2016) (Descriptive and Predictive) Kwame Boakye Agyapong, Dr. J.B Hayfron-Acquah, Dr. Michael Asante, Volume 4 Issue 5.
23. Fouskas, Konstantinos & Misirlis, Nikolaos & Karanatsiou, Dimitra & Vlachopoulou, Maro. (2018). Big data analysis in Tourism and Hospitality: a mapping literature review.
24. A Quinquennium mapping literature review on big data analytics in tourism and hospitality,( 2022) Nikolaos Misirlis, Dirk van der Steenhoven.
25. X. Wu, X. Zhu, G. -Q. Wu and W. Ding (2014) "Data mining with big data," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97-107.
26. Ahmed Oussous, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, Samir Belfkih, (2018) Big Data technologies: A survey, *Journal of King Saud University - Computer and Information Sciences*, Volume 30, Issue 4, Pages 431-448.

27. Data Mining: Concepts and Techniques, 3rd ed. The Morgan Kaufmann Series in Data Management Systems, (2018) Morgan Kaufmann Publishers..
28. Bharati, M. & Ramageri, Bharati. (2010). Data mining techniques and applications. Indian Journal of Computer Science and Engineering. 1.
29. Buyya, Rajkumar, Rodrigo N. Calheiros, and Amir Vahid Dastjerdi (2016) Big Data: Principles and Paradigms.
30. Big Data Analytics with R and Hadoop, (2013) Vignesh Prajapati, Packt Publishing.
31. Mariani, M.M and Borghi, M. (2019), “Industry 4.0: a bibliometric review of its managerial intellectual structure and potential evolution in the service industries”, Technological Forecasting and Social Change, Vol. 149, 119752
32. Martin, Kirsten. (2015). Ethical Issues in Big Data Industry. MIS Quarterly Executive.
33. Yallop, Anca & Seraphin, Hugues. (2020). Big data and analytics in tourism and hospitality: opportunities and risks. Journal of Tourism Futures. ahead-of-print. 10.1108/JTF-10-2019-0108.
34. Wu X, Huang J. (2023) Construction of a rural tourism information service management system for multi-source heterogeneous data processing. PeerJ Comput Sci.
35. Mahmoud, Omar & Rady, Ahmed & Eljalil, Sabreen. (2021). Big Data Analytics’ Utilization in Egyptian Hotels’ Decision-Making. What are The Challenges of Applying BDA?. Journal of Association of Arab Universities for Tourism and Hospitality

36. Chen, Wei-Chih & Chen, Wen-Hui & Yang, Sheng-Yuan. (2018). A Big Data and Time Series Analysis Technology-Based Multi-Agent System for Smart Tourism. *Applied Sciences*
37. d'Amore, Mariano & Baggio, R. & Valdani, Enrico. (2015). A Practical Approach to Big Data in Tourism: A Low Cost Raspberry Pi Cluster
38. Le TM, Liaw S-Y. (2018) Effects of Pros and Cons of Applying Big Data Analytics to Consumers' Responses in an E-Commerce Context. *Sustainability*.
39. Thomas H. Davenport, J. D., 2013. *Big Data in Big Companies*, s.l.: International Institute for Analytics.
40. Sivarajah, Uthayasankar & Kamal, Muhammad & Irani, Zahir & Weerakkody, Vishanth. (2016). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*
41. Ho, T., Nguyen, V.-H., Le, T., Le, H.-S., Nguyen, T.-D., Mai, T. T.-C., Tran, T.-A., & Truong, H. P. (2023). A new model for collecting, storing, and analyzing big data on customer feedback in the tourism industry. *Interdisciplinary Journal of Information, Knowledge, and Management*, 18, 225-249.
42. Mendoza-Moreno, J.F.; Santamaria-Granados, L.; Fraga Vázquez, A.; Ramirez-Gonzalez, G. *OntoTouTra* (2021): Tourist Traceability Ontology Based on Big Data Analytics. *Appl. Sci.* 11, 11061.
43. Peters, Scott and Keller, Peter, "Applications and issues of big data in tourism research" (2022). *Travel and Tourism Research Association: Advancing Tourism Research Globally*. 18.
44. Ardito, Lorenzo & Cerchione, Roberto & Del Vecchio, Pasquale & Raguseo, Elisabetta. (2019). Big data in smart tourism: challenges, issues and opportunities. *Current Issues in Tourism*

45. turizam. Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2021.
46. Mr. Ketan Bagade | Mrs. Anjali Gharat | Mrs. Helina Tandel " (2019) A Review Paper on Big Data and Hadoop for Data Science" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-1, pp.1216-1221
47. Theobald, William F. Global Tourism / Edited by William F. Theobald. (2015) 3rd ed. Burlington, MA: Elsevier Butterworth-Heinemann.
48. L. Clissa, Survey of Big Data sizes in 2021 (2022), arXiv
49. X. Wu, X. Zhu, G. -Q. Wu and W. Ding, (2014) "Data mining with big data," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107.
50. Jadhav, Mrs. Deepali Kishor. (2013) "Big Data: The New Challenges in Data Mining".
51. Shaikh, Eman & Mohiuddin, Iman & Alufaisan, Yasmeen & Nahvi, Irum. (2019). Apache Spark: A Big Data Processing Engine. 1-6.
52. Chaulagain, Ram & Pandey, Santosh & Basnet, Sadhu & Shakya, Subarna. (2017). Cloud Based Web Scraping for Big Data Applications. 138-143.
53. Rana, R. L., Giungato, P., & Tricase, C. (2021). Implementation of Blockchain Technology in the Tourism Industry: A Systematic Literature Review. New Trends in Sustainable Business and Consumption, 1, 594-602.
54. STROUMPOULIS, ASTERIOS & Kopanaki, Evangelia & VARELAS, SOTIRIOS. (2022). ROLE OF ARTIFICIAL INTELLIGENCE AND BIG DATA ANALYTICS IN SMART TOURISM: A RESOURCE-BASED VIEW APPROACH. 99-108. 10.2495/ST220091.

## Popis slika

Slika 1 Hadoop - HDFS Arhitektura.....	6
Slika 2 Hadoop - MapReduce arhitektura.....	7
Slika 3 Hadoop - YARN arhitektura.....	8
Slika 4 Generalni proces rudarenja podataka.....	9
Slika 6 Procesiranje tekstualnih podataka.....	15
Slika 7 Protok velikih podataka u turizmu.....	17
Slika 8 Analitički razvojni okvir velikih podataka u turizmu.....	19
Slika 9 Proces dizajna hotelske platforme velikih podataka.....	21
Slika 10 Spark Arhitektura.....	22
Slika 11 Model pretraživača.....	23
Slika 12 Statistika podataka korisničkih putovanja.....	25
Slika 13 Statistika hotelskih zvjezdica.....	25
Slika 14 Modelni okvir za obradu velikih podataka.....	26
Slika 15 Primjer web pretraživanja booking-a.....	28
Slika 16 AWS Usluge za skupljanje podataka.....	29
Slika 17 Amazon S3 usluga.....	29
Slika 18 Podaci hotela prikupljeni sa booking.com.....	30
Slika 19 Podaci korisnika prikupljeni preko recenzija.....	30
Slika 20 Grafički prikaz najčešćih riječi.....	32
Slika 21 Raspberry Pi Cluster.....	33
Slika 22 Raspberry Pi klaster dijagram.....	34

## **Popis tablica**

Tablica 1 Veliki podaci -3V .....	4
Tablica 2 Prikaz prikupljenih podataka.....	31
Tablica 3 Točke interesa (POI) i spomeni prikupljeni za različite OSN-ove .....	35

## **Sažetak**

Veliki podaci u kombinaciji s turizmom imaju veliki potencijal u pretpostavljanju budućih turističkih kretnji, bazirajući se na analiziranju podataka koji su se prikupili iz prošlosti.

Cilj ovog završnog rada bio je upoznati se s velikim podacima, kako se i na koje načine oni mogu razvijati u turizmu te di se sve koristi i na koji način. Ovaj rad se sastoji od velikih podataka koji su objašnjeni, te kako se ti podaci koriste pomoću određenih aplikacija i kako možemo te podatke prikupiti. Objašnjeni su veliki podaci u turizmu koji su najviše zastupljeni i kako se unutar turizma ti podaci mogu razviti te implementirati unutar hotela ili poduzeća. Ovaj rad može pridonijeti bilo kome tko je tek ušao u svijet velikih podataka, te treba postaviti temelje za daljnjim usmjerenjem prema nadograđivanju znanja.

**Ključne riječi:** Veliki podaci, Hadoop, razvojni okvir podataka, implementacija podataka

## **Abstract**

Big data in combination with tourism has great potential in resetting future tourist movements, based on the analysis of data collected from the past.

The goal of this final paper was to become familiar with big data, how and in what ways it can be developed in tourism, and where it is used and in what way. This paper consists of big data explained, and how this data is used by certain applications and how we can collect this data. Big data in tourism, which are the most prevalent, are explained and how within tourism this data can be developed and implemented within hotels or companies. This work can contribute to anyone who has just entered the world of big data, and should lay the foundations for further direction towards upgrading knowledge.

**Keywords:** Big data, Hadoop, data development framework, data implementation