



UNIVERSITETI I EVROPËS JUGLINDORE
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА
SOUTH EAST EUROPEAN UNIVERSITY

ANALIZA E RRJETEVE SOCIALE DUKE PËRDORUR PLATFORMA TË HAPURA

Arben Tafa

Programi i Biznesit dhe Informatikës

Shkurt / 2022

Shkup



Programi i Informatikës dhe Biznesit

Punim Diplome
Viti akademik 2013-2015

Arben Tafa

ANALIZA E RRJETEVE SOCIALE DUKE PËRDORUR PLATFORMA TË HAPURA

Shkurt/ 2022



Me titullin “Mësues i letërsisë dhe gjuhës shqipe”, unë, Arif Hajredini, i diplomuar në Universitetin e Tiranës konfirmoj se dokumentacioni është i lektyruar duke u përmbajtur rregullave të letërsisë dhe gjuhës shqipe.

Përmbajtja

LISTA E FIGURAVE.....	6
1. HYRJA DHE MOTIVIMI	9
1.1 Hyrje	9
1.2 Motivimi	11
1.3 Përshkrimi i problemit dhe objektivat.....	12
1.4 Hipotezat dhe pyetjet hulumtuese.....	13
2. ZHVILLIMI I RRJETEVE SOCIALE	14
2.1 Koncepti WEB 2.0.....	14
2.2 Impakti në Internet pas WEB 2.0.....	15
2.3 Rritja eksponenciale e rrjeteve sociale.....	16
3. TË DHËNAT NGA RRJETET SOCIALE	18
3.1 Hyrje – rritja e të dhënave nga rrjetet sociale	18
3.2 Metodot e analizës së të dhënave	20
3.2.1 Text Sentiment Analysis - Analiza e ndishme e përmbajtjes së tekstit	22
3.2.2 Analiza e përmbajtjes së tekstit në rrjetin social Twitter.....	23
3.2.3 Shembull konkret i analizës së të dhënave nga Twitter.....	24
4. PLATFORMAT E HAPURA TE PËRDORURA	25
4.1 Apache Kafka.....	25
4.1.1 Komponentët e Kafka	27
4.2 Postgre SQL.....	28
4.3 ElasticSeach	28
4.4 Jupyter Notebook	28
5. REALIZIMI TEKNIK I PROCESIT	29
5.1 Twitter API	30
5.2 Konfigurimi i platformave të hapura - hyrje.....	32
5.2.1 Konfigurimi Apache Kafka dhe komponentëve të saj.....	33
5.2.2 Konfigurimi i Elasticsearch dhe Postgres SQL dhe komponentëve të tyre	35
5.2.3 Konfigurimi i konektorëve në Kafka	36
5.2.4 Twitter – Kafka Connector	37
5.2.5 Kafka – Elasticsearch Connector	40

5.2.6 Kafka – PostgreSQL	41
5.3 Leximi i shënimeve të ardhura nga Twitter	42
5.3.1 Pranimi i shënimeve në Kafka	42
5.3.2 Dërgimi i shënimeve në Postgre SQL dhe Elasticsearch.....	43
6. Rastet për studim (Case study)	46
6.1 Rasti për studim (I) – Të dhënat me fjalën kyçe “Kosovo”	46
6.2 Rasti për studim (II) – Të dhënat me fjalët kyçe “Ronaldo dhe Messi”	51
6.3 Rasti për studim (III) – Të dhënat me fjalët kyçe “Pepsi dhe Coca Cola”	54
7. KONKLUZIONI.....	58
8. RERERENCAT	62

LISTA E FIGURAVE

Fig.1 - Kalimi nga WEB.1.0 në web.2.0.....	11
Fig.2 – Platformat në WEB 2.0.....	13
Fig.3 - Rritja eksponenciale e shfrytëzuesve në rrjete sociale 2004-2019.....	14
Fig.4 - Numri i përdoruesve nga krijimi i rrjeteve sociale.....	15
Fig.5 - Topologjia e Kafka Ekosistemit.....	22
Fig.6 - Topologjia e Kafka Ekosistemit.....	23
Fig.7 - Krijimi i Twitter API.....	26
Fig.8 - Tokenat e nevojshme nga profili i API në Twitter.....	27
Fig.9 - docker-compose konfigurimi.....	27
Fig.10 - Kafka cluster brenda Docker.....	30
Fig.12 -Konfigurimi i Elasticsearch përmes docker-compose.....	31
Fig.13 - Konfigurimi i Postgre SQL përmes docker-compose.....	31
Fig.14 - Interface-i i konektorëve në Kafka.....	32
Fig.15 - Konfigurimi i Twitter Connector.....	35
Fig.16 – Qasja në Twitter Connector.....	35
Fig.17 - Konfigurimi i ElasticSearch Connector.....	36
Fig.18 – Qasja në ElasticSearch Connector.....	36
Fig.19 - Konfigurimi i PostgreSQL Connector.....	37
Fig.20 – Test postim në Twitter për koneksion.....	38
Fig.21 - Pranimi i Tweet-t në Kafka.....	38
Fig.22 - Paraqitja e të dhënave në Kafka-UI.....	38
Fig.23 – Shënime të paraqitura përmes UI – Dejavu.....	39
Fig.24 - SQL Query përmes UI – Dejavu.....	40
Fig.25- Krijimi i koneksionit në PostgreSQL.....	41
Fig.26 - Të dhënat në Postgre SQL.....	41
Fig. 27-Paraqitja e shënimeve dhe importimi i paketave përmes Python.....	43
Fig.28 - Listimi i kolonave.....	44
Fig.29- Filtrimi i të dhënave.....	44
Fig.30,31 - Vizualizimi i shënimeve përmes Jupyter Notebook.....	45
Fig.32 - Analiza e polarizimit të tekstit përmes sentiment.polarity.....	48
Fig.33 - Analiza e polarizimit të tekstit përmes sentiment.polarity -2.....	51
Fig.34 - Paraqitja e teksteve në bazë të polarizimit.....	52

ABSTRAKTI

Gjatë këtij procesi punimi është paraqitur dhe përfshirë i gjithë procesi i konfigurimit, analizës, mbledhjes, ruajtjes së shënimeve nga rrjeti social **Twitter** duke përdorur platformat e hapura të cilat janë pa pagesë.

Në këtë punim do të shihni pjesën praktike dhe teorike si dhe procesin e realizimit teknik në analizën e të dhënave. Te pjesa e teorisë është diskutuar për analizat nga publikime nga artikuj të cilat kanë lidhje me të dhëna nga rrjetet sociale. Njëkohësisht janë shqyrtuar edhe metodat e ndryshme të analizës së përmbajtjes së tekstit. Në pjesën tjetër, atij teknik është paraqitur detajisht realizimi përmes platformave të hapura (Apache Kafka, PostgreSQL, ElasticSearch), mbledhja dhe ruajtja e të dhënave në kohë reale nga rrjeti social Twitter. Në bazë së këtyre të dhënave (Jupyter Notebook – Python) kemi arritur deri te rezultatet konkrete nga rastet e studimit, që janë marrë nga të dhënat e rrjetit Twitter.

Në shembujt e marra për të realizuar këtë punim i bazuar nga rrjeti Twitter duke i filtruar ato pikërisht me fjalët kyçe si “Messi”, “Ronaldo”, “Kosovo”, dhe “Pepsi dhe CocaCola”. Rezultatet kanë qënë inkurajuese mbas përfundimit të procesit duke i përdorur fjalët e lartpërmendura. Qëllimi kryesor ka qënë analiza e përmbajtjes së tekstit, specifikisht në pjesën negative dhe pozitive të përmbajtjes së tekstit.

ABSTRACT

This paper presents and discusses the entire process of configuring, collecting, storing and analyzing data from the social network Twitter, using all open source platforms - free of charge.

The paper includes the theoretical part and the practical part or the technical summary of the complete process for data retrieve, storage and data analysis (Data Science). The theoretical part elaborates the findings and analysis from various publications and articles related to data from social networks, and various methods of text content analysis. In the part of technical realization is presented in detail how it can be technically realized through open-source platforms (ElasticSearch, PostgreSQL, Apache Kafka) storage and collection of data in real time from Twitter, also Jupyter Notebook – Python as part of the analysis for concrete results from some case studies that are taken from the data of the Twitter network.

In the case studies, elaborated in this paper, the data were taken from the Twitter social network by filtering them in advance with keywords such as “Messi”, “Ronaldo”, “Kosovo” and “Pepsi and Coca Cola”. Based on these data, various analyzes have been made and encouraging results have been obtained. The main focus has been on text content analysis, specifically on the negative and positive polarization of text content.

1. HYRJA DHE MOTIVIMI

1.1 Hyrje

Me zhvillimin e rrjeteve sociale këto të dhëna janë shumëfishuar dhe kjo është një mundësi për të gjithë, që ato t'i shfrytëzojnë në mënyrën më të mirë. Në pjesën fillestare të këtij punimi është paraqitur një historiat i lëvizjes të rrjeteve sociale ndër vite që nga fillimi i WEB 2.0 dhe ndikimit të tyre në botën e internetit dhe në jetën tonë. Është përmbledhur rëndësia e të dhënave duke u bazuar në disa studime shkencore të publikuara dhe theks i veçantë në këtë pjesë i është vënë metodave të analizës së shënimeve nga rrjetet sociale, veçanërisht të dhënave nga Twitter. Nga disa metoda të analizës së përmbajtjes të të dhënave, për këtë punim metoda kryesore është ajo e analizës së përmbajtjes së tekstit, e cila është përdorur edhe për të analizuar të dhënat që janë mbledhur nga Twitter.

Pjesa kryesore e këtij punimi është pjesa e realizimit teknik të analizës, ruajtjes dhe mbledhjes të të dhënave të marra nga Twitter, të gjitha këto duke përdorur platforma të hapura (pa pagesë). Në këtë pjesë qëllimi ka qënë që në mënyrë më të thjeshtë dhe të kuptueshme të krijoj një përmbledhje sepse janë zgjedhur platforma të hapura si **Kafka**, **PostgreSQL** dhe **ElasticSearch**, duke u bazuar në për parësitë, mundësitë dhe fleksibilitetin e tyre. Konfigurimi i duhur i këtyre platformave ka sjellur një rezultat mjaft inkurajues dhe të kënaqshëm, duke na mundësuar që të dhënat t'i marrim, ruajmë dhe pastaj nga to të nxjerrim analiza sa më konkrete. Analiza e të dhënave është mundësuar gjuha programuese **Python**, me klasat e gatshme si "**Panda**", "**TextBlob**" dhe "**Heapq**".

Bazuar në analizën e të dhënave jam fokusuar të paraqes disa raste studimi, që në mënyrë sa më të thjeshtë të sqaroj rëndësinë e të dhënave nga rrjetet sociale dhe rezultatet konkrete nga analiza e këtyre të dhënave. Bazuar ne rrethanat në vendin tonë

rrjeti social Twitter nuk është shumë i përdorur dhe poashtu janë marrë tri raste studimi me të dhëna nga rrjetet sociale, në mënyrë që të jenë sa më të qarta rezultatet.

Të dhënat janë marrë duke u fokusuar me fjalë si:

- **“Messi / Ronaldo”**- Duke pasur parasysh se këta dy futbollistë janë padyshim më të mirët në botë vitet e fundit, dilema se cili lojtar është më i famshëm, apo të themi më i miri në botë, më ka shtyer që si rast studimi në këtë punim të përdor të dhënat nga Twitter që përmbajnë emrat e tyre. Nga këto informacione kemi ardhur në disa rezultate mjaft të mira lidhur me popullaritetin e tyre, opinionin e shfrytëzuesve për këta futbollistë dhe analizën e polarizimit negativ apo pozitiv të përmbajtjes së postimeve nga shfrytëzuesit në Twitter.
- **“Kosovo”** - Në këtë mënyrë janë filtruar të gjitha postimet në Twitter që kanë përmbajtur këtë fjalë. Arsyeja pse është përzgjedhur kjo fjalë është shkaku se në periudhën e fundit ka qenë aktuale marrëveshja në mes të shtetit të Kosovës dhe Serbisë. Ideja e këtij hulumtimi ka qenë që të mirren shumë të dhëna në lidhje me shembujt e lartpërmendura dhe më pas të analizohen. Si rezultat i këtyre postimeve, mund të themi se kemi ardhur në disa përfundime se si njëri nga kandidatët për president të SHBA-ve ka arritur që këtë marrëveshje ta përdorë për fushatën e tij politike ose në çfarë kontesti (pozitiv apo negativ) është përmendur emri Kosovë nga shfrytëzues vendor dhe ndërkombëtar.
- **“Pepsi / CocaCola”** – Ky shembull është marrë me qëllim të krahasimit të dy brendeve botërore dhe fama e tyre në rrjetet sociale. Meqë këto dy kompani janë konkurrent gjigand në tregjet botërore, na është mundësuar që të dhënat nga Twitter për këto kompani t’i përdorim për analizë në këtë punim.

Punimi është i ndarë në gjashtë pjesë. Pjesa e parë përmban një përshkrim të shkurtër të punimit, motivimin dhe pyetjet hulumtuese e hipotezat e ngritura. Në pjesën e dytë është paraqitur një historik i shkurtër i zhvillimit të rrjeteve sociale, ku është përfshirë koncepti

dhe ndikimi i Web 2.0 në internet. Në pjesën e tretë është përmbledhur rëndësia e të dhënave nga rrjetet sociale dhe janë shpjeguar metodat e analizës së të dhënave. Në pjesën e katër dhe pestë janë paraqitur platformat e përzgjedhura për realizimin dhe hapat e ndjekur gjatë realizimit teknik të gjithë procesit. Pjesa e fundit ka të bëjë me procesin e të dhënave që janë marrë nga Twitter duke përdorur gjuhën programuese Python.

1.2 Motivimi

Duke pasur parasysh faktin që të dhënat sot janë të çmueshme (disa mendojnë se vlera e të dhënave është ekuivalente me vlerën e naftës në bursa ndërkombëtare) kam vendosur të përzgjedh këtë temë, me qëllim që të jap kontribut sado modest drejt kuptimit se si të dhënat mund të merren dhe përpunohen nga rrjetet sociale dhe mënyrës së si nga ato mund të nxirren rezultate konkrete, të gjitha duke përdorur platforma të hapura.

Bazuar në njohuritë time në sektorin e IT-së, specifikisht në baza të të dhënave dhe analizë të shënimeve, e kam konsideruar të arsyeshme që këtë tematikë ta përmbledh në mënyrë sa më të thjeshtë dhe të kuptueshme dhe njëkohësisht të jap një mesazh që të dhënat nga rrjetet sociale janë shumë të dobishme për zhvillim dhe marrjen e vendimeve, opinioneve dhe veprimeve konkrete.

1.3 Përshkrimi i problemit dhe objektivat

Duke qenë se mbledhja e të dhënave nga rrjetet sociale tek ne në nuk praktikohet apo ndoshta nuk i kushtohet mjaft rëndësi rezultateve që këto të dhëna mund të ofrojnë, gjatë punimit do të mundohem ta sqaroj kryesisht rëndësinë e këtyre të dhënave dhe realizimin teknik që mundëson marrjen, ruajtjen dhe përpunimin e tyre. Në tregun e sotëm ka shumë mekanizma dhe softuere të cilët janë të ndërtuar dhe të gatshëm për të dhënat nga rrjetet sociale, por ata janë edhe të kushtueshëm, kështu tratativa kryesore është realizimi duke përdorur platforma të hapura Open-Source, të cilat janë pa pagesë.

Online ka shumë komunitete që zhvillohen në platforma të hapura dhe nëse publikohen, mund të shfrytëzohen nga kushdo. Ato platforma janë zakonisht të gatshme të publikohen në forume ose WEB faqe (p.sh **github.com**), ku të gjithë kanë qasje në to për t'i shfrytëzuar, njëkohësisht qëllimi i publikimit të tyre është që edhe të tjerët të kontribuojnë në zhvillimin e tyre të mëtejshëm, përmirsimin, sygjerimin apo çfarë do tjetër kontributi, i cili do të përmirsonte performancën e tyre.

Si objektiva dhe rezultate nga ky punim mund të theksohen:

- Njohja me mënyrën e analizës së të dhënave nga rrjetet sociale (shembull: Analiza e tekstit).
- Njohja me platforma të hapura dhe konfigurimi i tyre.
- Marrja e të dhënave nga **Twitter** në kohë reale (real time).
- Regjistrimi i të dhënave në platforma të hapura (**Elastic Search** dhe **RDBMS¹ : Postgre SQL**).
- Përpunimi dhe analiza e të dhënave (Python – Jupyter Notebook)/
- Të gjetura konkrete nga rastet e studimit.

¹ RDBMS - Relational DataBase Management System – Bazat e të dhënave relacionale

1.4 Hipotezat dhe pyetjet hulumtuese

Për të pasur sa më lehtë përpunimin e këtij hulumtimi dhe për të realizuar një analizë sa më konkrete të të dhënave, kemi ngritur disa hipoteza dhe pyetje hulumtuese.

Pyetjet hulumtuese:

P-1: Si krijohet marrja dhe ruajtja e shënimeve në kohë reale duke përdorur platforma të hapura?

P-2: A është e mundur të caktohet mendimi i shfrytëzuesit në Twitter duke përdorur gjuhën programuese Python?

Hipotezat:

H-0: Apache Kafka dhe PostgreSQL / Elasticsearch si platforma të hapura janë të përshtatshme për marrjen dhe ruajtjen e shënimeve ngal Twitter.

H-1: Bazuar në analizën nga Twitter me fjalën “Kosovo”, ish kandidati për president të SHBA-ve Donald Trump e ka cekur marrëveshjen mes Kosovës dhe Sërbisë për fushatë zgjedhore.

H-2 Nga të marrat në Twitter, futbollisti Lionel Messi është lojtari më i popullarizuar dhe lojtari më i mirë i kohës ne favor me futbollistin Cristiano Ronaldo.

H-3:Në bazë të analizës së Twitter mund të konstatojme se cili nga brendet botërore **Pepsi** dhe **Coca Cola** ka reputacion më të madh dhe cila nga këto dy brende ka opinion pozitiv/negativ nga shfrytëzuesit.

2. ZHVILLIMI I RRJETEVE SOCIALE

2.1 Koncepti WEB 2.0

Koncepti WEB 2.0 ka zanafillën në një konferencë bashkëpunuese në mes të O'Reilly dhe Media Live International në vitin 2001. Në atë kohë web faqet vetëm kishin filluar të marrin tatëpjetën dhe shumë kompani kishin falementuar. Kjo konsiderohet si pikë startuese për konceptin WEB 2.0, në të cilin u vërejtën shumë mundësi që diçka të ndryshoj në përmajtjen në internet. Lindja e platformave të reja që kanë pasur përmbajtje interaktive krijoi mundësi dhe ide që të bëhet diçka më ndryshe dhe si rezultat i kësaj filloi një epokë e re.

Disa ekspertë kanë qenë skeptik lidhur me këtë koncept duke e gjykuar vetëm si emërtim komercial, përderisa shumë te tjerë e kanë parë si pikë kthese në zhvillimin e internetit në përgjithësi.

Me zhvillimin e konceptit e WEB 2.0 ka filluar një epokë e re në komunikimin nëpërmjet Internetit. Kjo ka krijuar një urë të re lidhëse në mes të përdoruesve duke shkëmbyer material, programe, ideja si dhe më e rëndësishmja, ka krijuar mundësi të reja të bërit biznes përmes marketingut në internet.

Si pasojë e ndryshimit nga koncepti 1.0 në 2.0 është paraqitur një tabelë, ku krahasohen faqet me përmbajtje të ndryshme komerciale dhe jo komerciale:

<i>Web 1.0</i>	<i>Web 2.0</i>
DoubleClick	--> Google AdSense
Ofoto	--> Flickr
Akamai	--> BitTorrent
mp3.com	--> Napster
Britannica Online	--> Wikipedia
personal websites	--> blogging
evite	--> upcoming.org and EVDB
domain name speculation	--> search engine optimization
page views	--> cost per click
screen scraping	--> web services
publishing	--> participation
content management systems	--> wikis
directories (taxonomy)	--> tagging ("folksonomy")
stickiness	--> syndication

Fig.1 - Ndryshimi nga WEB.1.0 në WEB.2.0[1]

2.2 Impakti në Internet pas WEB 2.0

WEB 2.0 nuk është vetëm ndryshim i versionit nga ai 1.0 si teori, por është edhe një koncept krejtësisht tjetër në shumë fusha, si:

- WEB dizajn fleksibil, azhurim dhe ripërdorim kreativ
- mundëson inventimin e app-eve të reja duke i ripërdorur dhe kombinuar aplikacionet ekzistuese në WEB dhe duke i kombinuar të dhënat nga burime të ndryshme.
- ofron mundësinë e formimit të rrjeteve sociale për shfrytëzuesit më interesim të përbashkët dhe
- përkrah bashkëpunimin dhe mundësinë e shkëmbimit të informatave në mes të shfrytëzuesve.

Më të njohurat në mesin e këtyre platformave WEB 2.0 janë rrjetet sociale. Platformat si këto janë urëlidhësë mes kompanive me klientet e tyre dhe të dhënat nga këto platforma sociale mund të shërbejnë si burim për njohuri ndaj konsumatorin, preferencat e tij, pëlqimet, mospëlqimet, trendet, sjelljet, mendimet etj.

Një trend tjetër në WEB 2.0 është rritja e tregtisë elektronike (E-Commerce). Pothuajse të gjithë po prezantojnë bizneset e tyre online për të përfituar nga rritja e konsumatorëve në internet. Arsyeja është që ata jo vetëm se mund të tregtojnë produktet e tyre në mënyrë efektive në internet, por gjithashtu mund të gjurmojnë performancën e tyre, shitjet si dhe të masin KPI² për kompaninë e tyre.

Të dhënat nga këto platforma WEB 2.0 janë të mahnitëse, edhe pse sjanë të përsosura sa duhet për të mbështetur një vendim nëse analizohet veç, pra shihet që të zhvillohen platforma që këto të dhëna ti përpunojnë dhe të sjellin një rezultat konkret. Burimet nga WEB 2.0 jo vetëm që mund të nxjerrin informacion të dobishëm, të kërkuar për mbështetjen e vendimit, por gjithashtu mund të gjurmojnë performancën e pas-vendimit. Ky hulumtim bazohet në gjetjen e një metode efektive për të mbledhur të dhëna në WEB 2.0, për t'i integruar dhe analizuar ato duke përdorur mjete efektive të disponueshme për analizën e të dhënave në internet.

² KPI -Key Performance Indicator – Vlere e matshme për arritjet e objektivave te kompanisë



Fig.2 - Platformat nëWEB 2.0[21]

2.3 Rritja e menjëhershme e rrjeteve sociale

Me përparimin e konceptit WEB 2.0 është krijuar lidhja shfrytëzuese mes veti duke shkëmber mendime,informata dhe publikime apo informacione të ndryshme përmes internetit. Rol të rëndësishëm në këtë ndërkomunikim ka pasur zhvillimi i rrjeteve sociale ku shfrytëzuesit krijojnë një profil në rrjet të caktuar dhe i ftojnë të tjerët për tu bashkangjitur atij rrjeti dhe pastaj bëjnë shkëmbimin e të dhënave në formate të ndryshme (Foto, Video, text e-mail etj) dhe bëjnë interaksion si shkëmbim të mendimeve,kritikave,publikimet personale shkencore etj.

Nga publikimi i Michael Haenlein dhe Andreas Kaplan 2010 “*The Challenges and Opportunities of Social Media*” [8] - rrjetet sociale janë një grup i aplikacioneve të bazuaranë Internet, të ndërtuara në themelet ideologjike dhe teknologjike të konceptit WEB2.0 dhe që lejojnë krijimin dhe transferin e teksteve të gjeneruara nga shfrytëzuesit.

Rrjetet sociale kanë ndryshuar botën në mënyrë pozitive nga aspekti komercial dhe interaktiv, por edhe kanë pasur ndikim negativ tek shumë njerëz. Përvetësimi i shpejtë dhe i gjerë i këtyre teknologjive po ndryshon mënyrën se si ne gjejmë partnerë, si kemi qasje në informacion nga lajmet dhe si organizohemi për të kërkuar ndryshime politike apo sociale. Rrjetet sociale kanë ndikuar edhe negativisht tek çdo person duke krijuar varësi të madhe tek shumë njerëz që i kalojnë të gjithë kufijtë në përdorimin e tyre.

Që nga viti 2004 rrjetet sociale kanë pasur një rritje marramendëse si platforma në internet. Sipas publikimit nga Esteban Ortiz-Ospina – “ *The rise of social Media* ” [22] si platformë e parë në vitin 2004 ka qënë rrjeti myspace, e cila në vitin 2006 arriti të tejkalojë edhe Google me vizita ditore në SHBA. Ajo kohë shënon fillimin e rritjes së madhe të platformave në internet. Në figurën më poshtë është paraqitur një grafikun që tregon rritjen e shfrytëzuesve të rrjeteve sociale nga viti 2004.

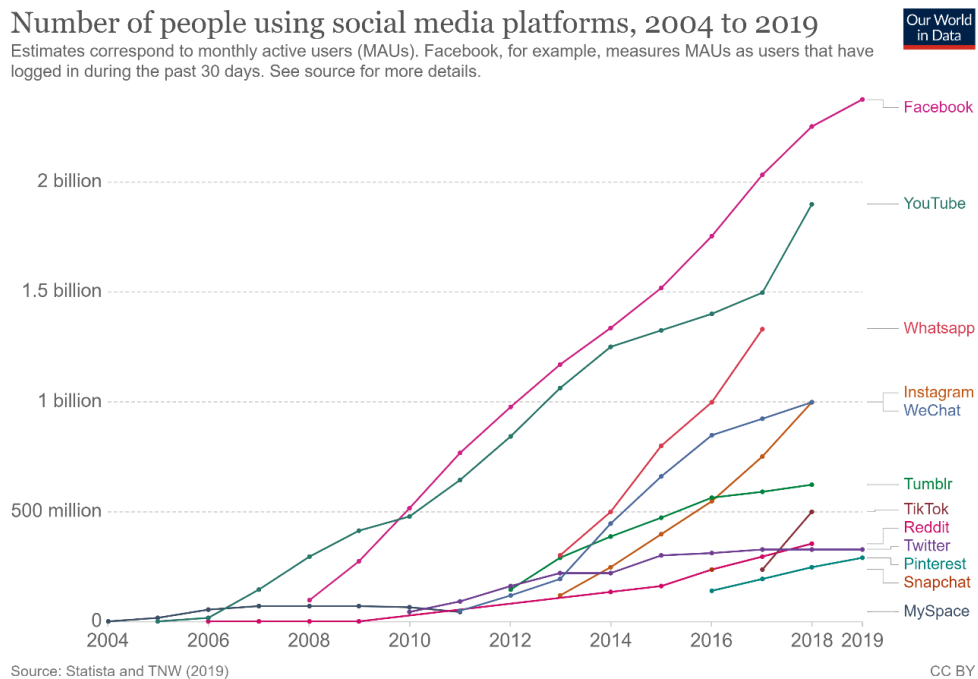


Fig.3 - Rritja eksponenciale e shfrytëzuesve në rrjete sociale 2004-2019[17]

Në bazë të kësaj figure, mund të konkludojmë se disa rrjete sociale për këto 15 vite të fundit kanë pasur rënje nënumrin e shfrytëzuesve të tyre. Nvarësisht nga konkurenca e atëhershme platformat si Hi5, MySpace janë shuar dhe platforma tjera të reja kanë hasur rritje shumë të shpejtë. Facebook, si rrjeti social më i madh në treg me mbi **2.4 miliardë** shfrytëzues aktiv është shndërruar në lider.

3. TË DHËNAT NGA RRJETET SOCIALE

Nga fillimi i këtij shekulli kompanitë e ndryshme i kanë përdorur WEB faqet për të krijuar ndërlidhje me konsumatorët e tyre. Kjo ka qenë e vetmja mënyrë që kompanitë të kuptojnë kërkesat, ankesat, popullaritetin apo ndjenjën e konsumatorëve të tyre. Sidoqoftë kjo nuk ka mjaftuar që të ketë një pasqyre reale lidhur me performancën e kompanisë dhe reflektimin e saj në publikun e gjërë. Me zhvillimin e rrjeteve sociale, ky trend bashkë me pjesën e marketingut tek kompanitë kanë ndryshuar rrënjësisht.

Në bazë të statistikave të publikuara në internet rreth 80 % e shfrytëzuesve përdorin rrjetet sociale dhe kjo paraqet një mundësi të madhe që këto të dhëna të përdoren për qëllime të ndryshme nga individë apo kompani.

3.1 Hyrje – rritja e të dhënave nga rrjetet sociale

Fillimi i përdorimit të rrjeteve sociale ka mundësuar fillimin e konceptit Big Data në mbarë botën. Vëllimi i madh i të dhënave nga rrjetat sociale ka ofruar një hap tjetër në aspektin e mbledhjes/përpunimit të të dhënave. Shumë rrjeteve sociale iu është dashur që në mënyrë konstante të ndryshojnë përmbajtjet dhe mundësitë që i kanë ofruar për përdoruesit e tyre, në mënyrë që të jenë në hap me trendin e rrjeteve sociale të kohës dhe konkurrencës. Twitter si rrjet social deri në vitin 2011 nuk ka lejuar që shfrytëzuesit të postojnë fotografi apo video. Nga viti 2011 kjo është e lejuar dhe tani sipas analizave [17] rreth 50% e përmbajtjes së të dhënave në Twitter janë fotografi ose video.

Bazuar në publikimet nga Esteban Ortiz-Ospina – “*The rise of social Media*” [22] mund të vërehet një rritje eksponenciale e shfrytëzuesve të rrjeteve sociale ndër vite. Facebook është lider në rritjen e shfrytëzuesve të tij për 10 vitet fundit, ku tani janë rreth 2,38 miliardë shfrytëzues me një rritje prej **2,275%** në krahasim me vitin 2008.

Facebook	100.00 million in 2008	2.38 billion in 2019	+2.27 billion	+2,275%
Flickr	3.68 million in 2004	80.00 million in 2013	+76.32 million	+2,077%
Friendster	5.97 million in 2004	28.80 million in 2009	+22.83 million	+382%
Google Buzz	166.03 million in 2010	170.00 million in 2015	+3.97 million	+2%
Google+	107.32 million in 2012	430.00 million in 2018	+322.68 million	+301%
Hi5	9.73 million in 2005	46.61 million in 2011	+36.88 million	+379%
Instagram	117.50 million in 2013	1.00 billion in 2018	+882.50 million	+751%
MySpace	980,036.00 in 2004	46.00 million in 2011	+45.02 million	+4,594%
Orkut	4.90 million in 2004	45.07 million in 2012	+40.17 million	+820%
Pinterest	143.25 million in 2016	291.00 million in 2019	+147.75 million	+103%
Reddit	248,309.00 in 2006	355.00 million in 2018	+354.75 million	+142,867%
Snapchat	238.65 million in 2016	238.65 million in 2016	+0.00	+0%
TikTok	239.14 million in 2017	500.00 million in 2018	+260.86 million	+109%
Tumblr	146.89 million in 2012	624.00 million in 2018	+477.11 million	+325%
Twitter	43.25 million in 2010	330.00 million in 2019	+286.75 million	+663%

Fig.4 - Numri i përdoruesve të rrjeteve sociale[22]

Rrjetet sociale janë pëlqyer dhe kanë filluar të përdoren nga shumë biznese. Gjithnjë e më shumë kompani po përdorin mjete të mediave sociale si Facebook dhe Twitter për të ofruar shërbime të ndryshme dhe për të bashkëvepruar me klientët. Si rezultat, një sasi e madhe e përmbajtjes së gjeneruar nga shfrytëzuesit është në dispozicion për përdorim në faqet e mediave sociale. Për të rritur epërsinë konkurruese dhe për të vlerësuar në mënyrë efektive mjedisin konkurrues të bizneseve, kompanitë duhet të monitorojnë dhe analizojnë jo vetëm përmbajtjen e gjeneruar nga klientët në faqet e tyre të mediave sociale, por edhe informacionin me tekst në faqet e mediave sociale të konkurrentëve të tyre. Në një përpjekje për të ndihmuar kompanitë të kuptojnë se si të kryejnë një analizë konkurruese të mediave sociale dhe të transformojnë të dhënat e mediave sociale në njohuri për vendimmarrësit dhe tregtarët elektronikë.

3.2 Metodatat e analizës së të dhënave

Në bazë të pohimeve të lartpërmendura, mund të vijmë në konkluzion se të dhënat nga rrjetet sociale mund të janë të dobishme nëse ato përdoren në mënyrë të duhur duke përdorur teknika të duhura. Pra, kushdo që dëshiron të merrë me analizën e shënimeve nga rrjetet sociale së pari duhet të përcaktoj qëllimin e analizës dhe gjithësesi të zgjedhë metodën më të përshtatshme për të ardhur deri tek një rezultat i dëshiruar dhe i suksesshëm.

Të gjithë metodat janë shpjeguar në mënyrë të shkëlqyer në **Ani Saponoutzi dhe Kostas Psanis** - *“Social Networking Data Analysis Tools & Challenges”* [9].

- **“Influence Analysis”**- kjo metodë e analizës parasheh identifikimin e shfrytëzuesve të besueshëm dhe me experiencë, ku të dhënat e tyre mund të shfrytëzohen për analizë sepse janë më kredibile, poashtu këta shfrytëzues mund të shpërndajnë lajmin tek shfrytëzuesit tjerë.
- **Link Analysis** - parasheh analizën e shfrytëzuesve që kanë interesim të përbashkët. Google si një ndër kompanitë më të mëdha përdor këtë metodë për identifikimin

e faqeve më të vizituara nga një grup i caktuar shfrytëzuesish duke u bazuar në algoritme.

- **Community Detection**– kjo metodë parasheh analizen e një grupi të shfrytëzuesve në bazë të aktivitetit të tyre në rrjete sociale. Kjo metodë prezanton tri forma të analizës
 - Analiza statike “Numri I shfrytëzuesve në kohë të caktuar?”
 - Analiza e përkohëshme “Si është formuar një grup shfrytëzuesish”
 - Analiza e parashikimit “Si një grup mund të rritet-zgjerohet”

- **Sentiment Analysis** - Shqyrton analizën e detajuar të përmbajtjes së tekstit të në rrjete sociale. Kjo metodë do të jetë si fokus në shtjellimin e mëtejshëm në këtë punim.

Publikimi i *Bogdan Batrinca dhe Philip C. Treleaven* – “*A survey of tools, techniques and platforms-Social media analytics* ” [12] paraqiten dhe përshkruhen disa metoda dhe teknika për analizën dhe përdorimin e të dhënave nga rrjetet sociale

- **News analytics (Analiza e lajmeve)** – Parasheh mbledhjen dhe ruajtjen e lajmeve të ndryshme nga rrjetet sociale duke bërë analiza nga këto të dhëna.
- **Opinion mining** - parasheh analizën e të dhënave dhe opinionin e shfrytëzuesve për temë të caktuar. Këtu hyn dhe niveli i knaqësisë së klientëve për një gjë të caktuar, grupacion interesi, aktualitet apo qështje politike.
- **Sentiment text analysis** – Merret me analizën e tekstit të shkëmbyer në rrjetet sociale. Fokusi i kësaj metode është përmbledhjen e mëtejshme në këtë punim.

3.2.1 Analysis of text sentiment - Analiza e ndishme e përmbajtjes së tekstit

Rrjetet sociale ofrojnë shkëmbim të informatave në mes të shfrytëzuesve si komentetet, pëlqimet apo shpërndarjet të artikujve dhe përmbajtjeve të ndryshme. Nga këto të dhëna mund të dalin analiza dhe rezultate shumë të dobishme nëse ato përpunohen dhe analizohen në mënyrë të duhur dhe efikase. Opsionet që ofrohen nga rrjetet sociale për analiza nuk mund të nxjerrin rezultate të detajuara. Si mundësi shumë e mirë për nxjerrjen e informatave sa më konkrete është analiza e përmbajtjes së tekstit, ku si rezultat mund të arrijmë në shumë konkluzione si: opinionet e shfrytëzuesve, përcaktimi i polarizimit të shprehjeve nga shfrytëzuesi, shprehja e mendimeve, ndjenjave etj.

Artikuli i publikuar nga **Svetlana Kiritchenko dhe XiaodanZhu** “Sentiment Analysis of Short Informal Texts“[\[9\]](#) mund ti gjejnë taktikat dhe metodat për nxjerrjen e informave të rëndësishme nga teksti. Theks të veçantë autorët vënë në disa aspekte të analizës së përmbajtjes së tekstit si:

- zbulimi i fjalive objektive apo subjektive,
- përcaktimi i fjalës/fjalisë si pozitive, negative apo neutrale ndaj një teme apo subjekti të caktuar,
- zbulimi dhe qëllimi i shprehjes së shfrytëzuesit
- shprehja e emocionit (frikës, kënaqësisë/pakënaqësisë apo irritimit)

Sipas këtij punimi, përdorimi i këtyre sistemeve të analizave gjatë dekadës së fundit ka rezultuar me informata të duhura në të gjitha aspektet e biznesit, mjekësisë, gazetarisë, kujdesit ndaj klientëve, fatkeqësive natyrore etj. Këto sisteme kanë mundësuar publikime rreth metodave për analizë për përmbajtjen e tekstit në Twitter. Këto publikime do të diskutohen në pjesët e mëtejshme të punimit.

3.2.2 Përmbajtja e tekstit dhe analiza në rrjetin social Twitter

Siç e kemi theksuar në fillim të këtij punimi, pika kryesore e mbedhjes së informatave do të jetë në Twitter. Në lidhje me këtë platformë për analizën e teksteve eksistojnë shumë punime shkencore dhe publikime. Që të arrijmë deri te rezultati kërkuar, na duhet të kemi të qartë se çfarë kërkojmë si rezultat, qëllimi i cili tek ne është rritja e kënaqësi së klientëve, përmirsimi shërbimit, marrja e vendimeve dhe pas të shndërrohen në vendimmarrje.

Duke u bazuar në punimet e publikuara online [12],[11], (ku fokusi dhe tema e këtyre punimeve është kryesisht në mënyrën e nxjerrjes dhe përpunimit të informatave nga Twitter) si teknika të analizës së informatave nga Twitter zakonisht përdoren këto metoda:

- **Keyword** – Filtrimi i të dhënave në bazë të një apo më shumë fjalëve kyçe.
- **Përmbajtja e Tweetit dhe analiza** - Shprehja e emocioneve të fytyrës përmes simboleve (Emotions) , fjalët kyçe e të tjera...
- **Targeti (Caku)TAG** - i shprehjes përmes tekstit apo fjalisë, duke u bazuar në karakterin @ ku përmes këtij karakteri definohet (citohet) subjekti i caktuar (@Struga , @BankaXYZ , @CaffeXYZ etj)
- **Hashtag (#)** – Ky simbol përdoret më së shumti në rrjetin social Twitter. Më këtë simbol zakonisht theksohet (nënvizohet) një temë e caktuar apo pjesë e rëndësishme e fjalisë së shfrytëzuesit . Qëllimi i përdorimit të këtij simboli është për të bërë atë pjesë të tekstit më të dukshme dhe të lexueshme për shfrytëzuesit tjerë.

3.2.3 Shembull konkret i analizës së të dhënave nga Twitter

Në publikimin nga **Weiguo Fan** – “*The power of social media analytics*”[13] mund të shohim disa të dhëna nga analiza e rezultateve nga ompani të njohura lidhur me impaktin në rrjetet sociale. Në theks vendoset kompania e Hoteleve “**Accor**” e cila pas një periudhë të gjatë me reagime (Feedback) dhe komente të këqija angazhon një kompani për analizën e të dhënave të klientëve të saj për të përmirësuar gjendjen.

Qëllimi ishte të gjinden shkaqet se pse ka rënë aq shumë pakënaqësia e klienteve. Kompania e angazhuar kishte mbledhur të dhënat e klientëve të hoteleve “Accor” dhe të dhënat e konkurrentëve të tjerë, në mënyrë që të kenë një pasqyrë më të qartë dhe për të krahasuar të dhënat me të tjerët. Poashtu kompania “**Ancor**” që të përdoren rezervimet online kishte nis kampanja stimuluese.

Mbas analizave të bëra konkludohet se klientët janë të pakënaqur me hapësirën e çelësit të dhomave. Duke u fokusuar në këto analiza dhe marrjes së masave kompania ka pasur feedback për 55% më të mirë mbas 1 viti dhe numri i komenteve negative është reduktuar.

4. PLATFORMAT E HAPURA TE PËRDORURA

Platformat e hapura ose ndryshe Software-et e hapur janë programe kompjuterike të modifikuara nga grup njerëzish të ndryshëm kodi i te cilit është i hapur për zhvilluesit e tjerë që ta modifikojnë, përmisojnë apo shtojnë funksionalitete në të. Faktikisht çdo zhvillues i një platforme ka të drejtë tju japë lisencë zhvilluesve të tjerë në atë kod..

Në rastin tonë, për të realizuar hulumtimin kemi përdorur këto platforma:

- **Apache Kafka** – Për marrjen e shënimeve nga Twitter në kohë reale.
 - **Apache Zookeeper** – Shërbim lidhur me konfigurimin e proceseve.
- **PostgreSQL** – databazë relacionale për ruajtjen e shënimeve.
- **Elasticsearch** – për ruajtjen dhe analizën e shënimeve.
- **Jupyter Notebook** – për Data Science dhe raportime.
 - **Python 3** – gjuhë programuese për Data Science.

4.1 Apache Kafka

Apache Kafka si platformë komponentë bazë i ka Java dhe Scala ku dhe është ndërtuar nga kompania e njohur LinkedIn dhe për herë të parë është leshuar në përdorim në vitin 2011. Është një nga platformat që vitet e fundita morri ngritjen më të madhe e cila përdoret nga mijëra kompani ditët e sotme.

Gjithashtu në lidhje me këtë hulumim konsiderohet si platforma kryesore për marrjen e të dhënave nga Twitter. Kafka është sistem i bazuar në menaxhimin e rrjedhjes së informatave në kohë reale (Data flow Streaming) në mes të sistemeve të ndryshme dhe poashtu mundëson ruajtjen e atyre të dhënave nëpër Cluster të data serverave. Si në fillim është zhvilluar si platformë për përcjelljen e mesazhave si ndërmjetsues

(Middleware) midis sistemeve të ndryshme dhe është krijuar me kohë për Data Flow Streaming³.

Në librin “**Kafka: The Definitive Guide**” [2] e themeluar nga autorët e kësaj platforme, janë të përshkruara të gjithë funksionalitetet në mënyrë të hollësishme. Ky libër rekomandohet të lexohet nëse keni punë me Kafkën.

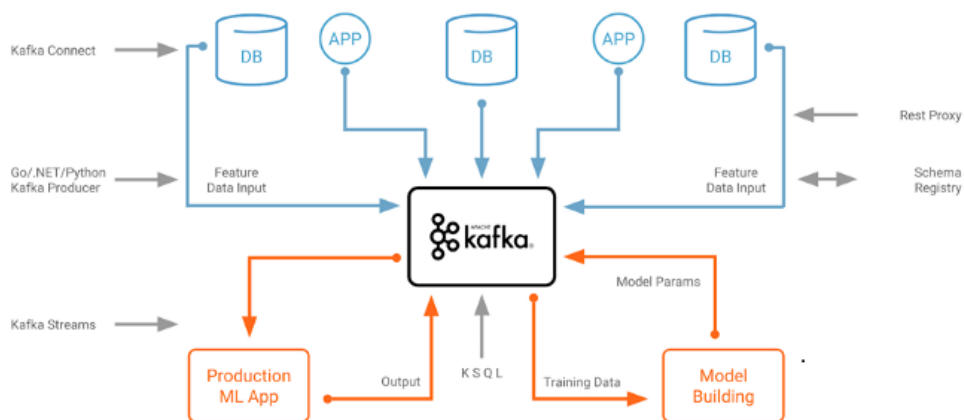


Fig. 5 - Topologjia e Kafka Ekosistemit[26]

³ Data Flow Streaming – Sistem për rrjedhjen e të dhënave në kohë reale ndërmjet sistemeve të ndryshme

4.1.1 Komponentët e Kafka

Bazuar në arkitekturën e kësaj platforme,egzistojnë 4 komponentë kryesorë:

- **Kafka Broker:**Merret me kërkesat e klientëve (prodhimin konsumimin dhe metadata) dhe ka qasje me Cluster-in e serverit ku është Kafka e ngritur. Egziston mundësia të jenë më shumë se një Broker në Cluster.
- **Zookeeper:**Ruan gjendjen e Culsterit ku është ngritur Kafka (Broker,topic,shfrytëzuesit).
- **Producer:** Dërgon dhe ruan shënimet në Brkoker (Topic).
- **Consumer:** I konsumon (i merr) shënimet nga Broker për aplikacionin.
- **Kafka-Topic:** Shënimet që vijnë në Kafka aktualisht janë të organizuara nepër Kafka Topic. Në Kafka topic shkuan Kafka producer nga ku shfrytëzuesi i lexon nga i njëjti topic.

Kafka memoron të dhënat në regjister dhe shfrytëzuesi është përgjegjës për hulumimin e tyre, i njohur si “OFFSET”. Në mënyrë tipike, një konsumator avancon kompensimin në një mënyrë lineare kur lexohen mesazhet. Sidoqoftë, pozicioni në të vërtetë kontrollohet nga konsumatori, i cili mund të konsumojë mesazhe në çdo mënyrë. Shënimet në Kafka janë të ndara në grupe,pra një temë mund të ketë më shumë nëngrupe.

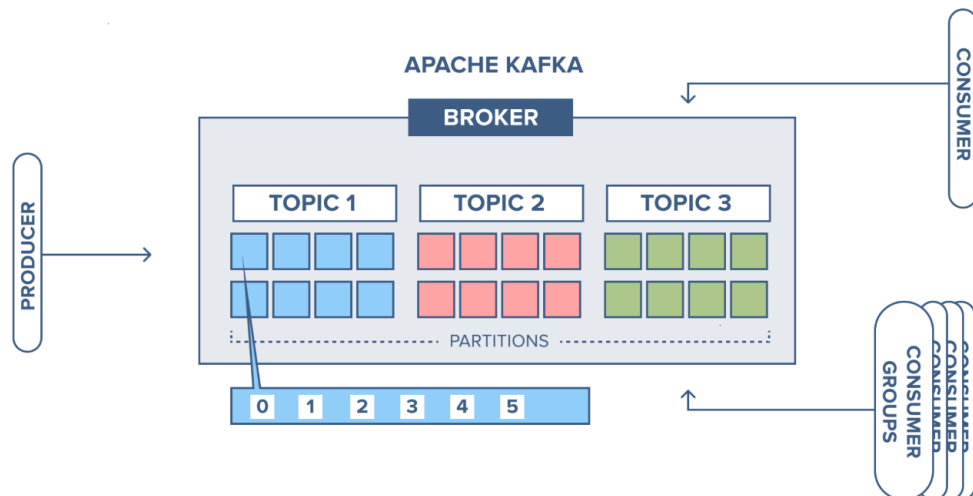


Fig. 6 - Topologjia e Kafka Ekosistemit[25]

4.2 Postgre SQL

Pjesë e projektit është edhe ruajtja e të dhënave dhe si platformë e cila është shumë e fuqishme për momentin është zgjedhur PostgreSQL. Kjo bazë e të dhënave është shumë stabile dhe ka një përdorim të gjërë nga kompanitë e ndryshme.

PostgreSQL mbështet si Querrie SQL (relacionale) ashtu edhe JSON (jo relacionale).

4.3 ElasticSeach

ElasticSearch është motor kërkimi e bazar në bibliotekën Lucene. Është shumë e avansuar e cila mundëson kërkim të plotë teksti, një ose më shumë Interface, HTTP dhe dokumente tjera. Elasticsearch është zhvilluar në Java e cila është shndërruar në model biznesi duke ofruar licenca për zhvillim dhe përdorim.

Ka një përdorim të madh për pjesën analitike dhe kërkim të shënimeve dhe gjen përdorim shumë në platformat në Cloud si p.sh AWS (Amazon WEB Services). Është shumë fleksibile dhe pranon gjuhën SQL si standard të kërkimit.

4.4 Jupyter Notebook

Platformë tjetër e hapur që kemi përdor është dhe Jupyter Notebook e cila mundëson përpunimin më të hollësishëm të shënimeve. Si burim i të dhënave mund të jenë të dhënat relacionale ose Json. Si gjuhë programuese e përdor Python, dhe ne në këtë rast e përdorim Python 3.

Jupyter është përzgjedhur më saktësisht për analizën e të dhënave të marra nga Tweeter deri te rezultatet e nevojshme. Si paketë ose librari të Python do të përdorim Pandas, textBlob dhe heapQ, të cilat do të na ndihmojnë në realizimin sa më të lehtë të procesit.

[26]

5. REALIZIMI TEKNIK I PROCESIT

Qëllimi kryesor në këtë hulumtim ka qënë realizimi i marrjes të të dhënave nga Tweeter dhe analiza e tyre duke përdorur platforma tjera që të arrijme te rezultati i dëshiruar.

Shikuar teknikisht, ka shumë mundësi se si të bëhet komplet ky proces, por ne kemi përzgjedhur një mënyrë bashkëkohore dhe lehtë të lexueshme dhe kuptueshme nga ndonjë palë tjetër që mund të shfaqë interesim për të punuar me këto platforma.

E gjithë ecuria e realizimit të procesit mund të ndahet në dy pjesë:

- **Pjesa inxhinierike**- mbledhja dhe ruajtja e të dhënave duke i konfiguruar me Tweeter API.
 - Konfigurimi i Twitter API.
 - Konfigurimi i platformave në Docker (Kafka ,Postgres ,Elasticsearch).
 - Krijimi i Kafka Konektorve – Source Connector (Twitter> Kafka).
 - Krijimi i Kafka Konektorve – Sink Connector (Kafka > Postgres, Kafka > Elasticsearch).
- **Pjesa analitike (Data Science)** Pas marrjes së analizave të vijmë deri te arritja e rezultateve të nevojshme.

5.1 Twitter API

Twitter API⁴ është një aplikacion i lansuar nga Twitter që mundëson lidhjen e rrjetit social Twitter me aplikacione tjera. Twitter në krahasim me rrjete tjera sociale si (Facebook , Instagram) e ka më të lehtë qasjen në API të vet dhe për shërbime të marrjes së të dhënave, si në rastin tonë, nuk ka nevojë për ndonjë pagesë.

Për të pasur mundësi të qasemi në të dhënat e këtij rrjeti social, nevojitet që të kemi një llogari të hapur në Twitter. Në rastin tonë të studimit kam përdorur llogarinë time private.

Hapi i parë që duhet të ndërmarim është të krijojmë qasje në API duke hyrë në <https://developer.twitter.com/en/application/>.

Në këtë seksion duhet të zgjedhim njerën nga arsyjet pse do të përdorim API dhe poashtu duke i plotësuar fushat e nevojshme që kërkohen aty.

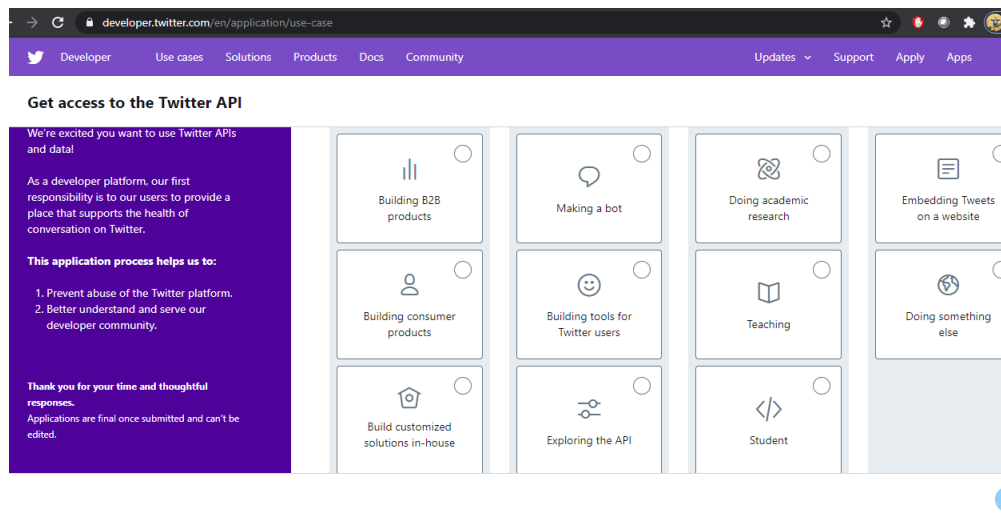


Fig. 7 - Krijimi i Twitter API

Pas përfundimit të procedurës pritet konfirmimi nga Tweeter e cila realizohet për një periudhë deri në 2 orë.

⁴ API – Application Program Interface

Pasi të kemi marrë përgjigje, gjenerohen **API_KEY** dhe **API_SECRET** dhe poashtu **ACCESS_TOKEN** dhe **ACCESS_SECRET**.

Paramentrat do të na mundësojnë më vonë lidhjen me Tweeter dhe Kafka.

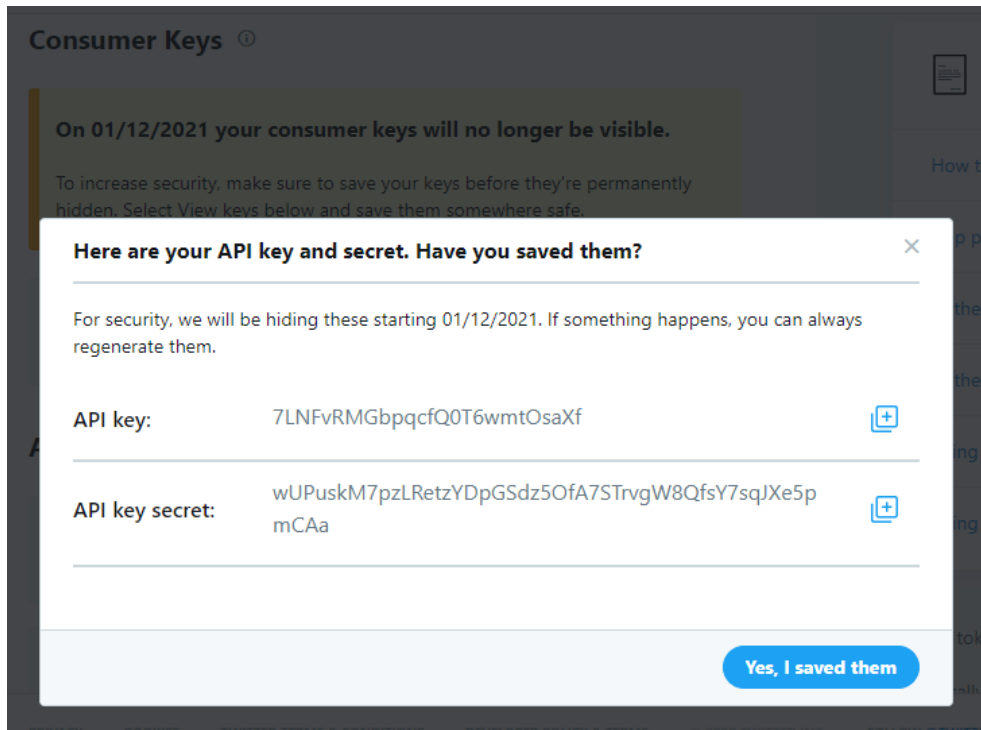


Fig. 8 -Tokenat e nevojshëm nga profili i API nëTwitter.

5.2 Platformat e hapura dhe konfigurimi i tyre - hyrje

Fillimisht, për konfigurimi e platformave të cekura më lartë na duhet një server ose paisje që operon në sistemin operativ Linux. Pasi që kjo është bërë për demonstrim dhe numri i të dhënave nuk është aq i madh, ne do të instalojmë të gjitha aplikacionet në kompjuter personal, duke përdorur **Docker**.

Docker është një aplikacion Open Source funksioni i të cilit është të bëjë paisje virtuale brenda sistemit operativ Windows. Pastaj përmes Docker mund të instalojmë aplikacionet që punojnë në sistemin operativ Linux.

Shkarkimi i aplikacionit Docker <https://docs.docker.com/docker-for-windows/install/>

Për të mundësuar instalimin e gjithë këtyre, na nevojitet që kompjuteri personal të ketë performacë të mirë. Rekomandohet që të ketë **16GB RAM** memorie dhe procesor **i7**.

Për instalimin e aplikacioneve që na nevojiten, do të përdorim imazhet e publikuara Open Source në **github.com**:

- Kafka dhe komponentët e saj - **landoop/fast-data-dev** [30]
- Elasticsearch - **itzg/elasticsearch:2.4.3** [31]
- Postgre SQL - **postgres:9.5-alpine** [32]

Në lidhje me Data Science do përdorim Anaconda e cila përmban Python dhe Jupyter.

Applikacioni mund të shkarkohet në linkun:

<https://www.anaconda.com/products/individual>

5.2.1 Komponentët e Apache Kafka dhe konfigurimi i saj

Në rastin tonë është përdorur imazhi i gatshëm **landoop/fast-data-dev**. Konfigurimi bëhet përmes aplikacionit Docker duke e lexuar file-in `docker-compose.yml`, si më poshtë. File përmban parametra konfiguruse për tre sisteme që do ti kemi të nevojshme në projekt.

```
1  version: "2"
2  services:
3  kafka-cluster:
4    image: landoop/fast-data-dev:latest
5    environment:
6      ADV_HOST: 127.0.0.1
7      RUNTESTS: 0
8      SAMPLEDATA: 0
9    ports:
10     - 2181:2181           # Zookeeper
11     - 3030:3030          # User Interface
12     - 8081-8083:8081-8083 # Rest proxu , Schema Registry , Kafka Connect
13     - 9581-9585:9581-9585 # JMX
14     - 9092:9092          # Kafka Broker
15  elasticsearch:
16    image: itzg/elasticsearch:2.4.3
17    environment:
18      PLUGINS: appbaseio/dejavu
19      OPTS: -Dindex.number_of_shards=1 -Dindex.number_of_replicas=0
20    ports:
21     - "9200:9200"
22  postgres:
23    image: postgres:9.5-alpine
24    environment:
25      POSTGRES_USER: postgres
26      POSTGRES_PASSWORD: postgres
27      POSTGRES_DB: postgres
28    ports:
29     - 5432:5432
```

Fig. 9 - docker-compose konfigurimi

Për të instaluar Kafka-n si dhe aplikacionet tjera duhet ndjekur hapat e mëposhtme:

1. E hapim Comand Promtnë PC lokale.
2. Ndryshohet direktiva ku docker-compose është ruajtur
3. Ekzekutojmë komandën: **docker-compose up kafka-cluster**.

Pas ekzekutimit, fillon të shkarkohet fotografija direkt nga interneti.

Kur përfundon procesi, Kafka Cluster është i gatshëm. Këtë mund ta shohim në Docker-Dashboard si më poshtë:

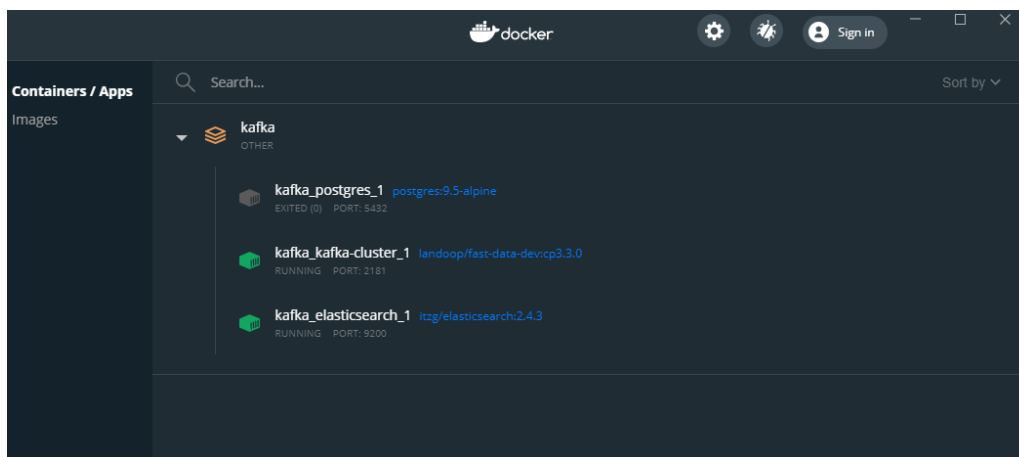
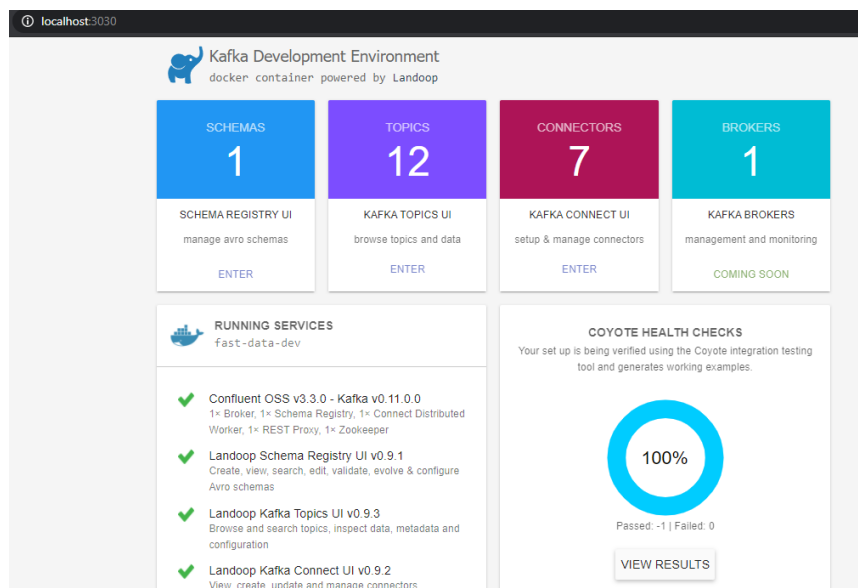


Fig.10 - Kafka cluster brenda Docker

Pasi Kafka tani është e konfiguruar dhe gati për përdorim, mund të hyjmë në User Interface, ku në bazë të YML file që është cekur më herët, mund të qasemi si WEB faqe. Në ndonjërin prej shfletuesve të internetit (Chrome , Mozilla, Microsoft Edge) mund të hyjmë në <http://127.0.0.1:3030> ose <http://localhost:3030> dhe do të paraqitet faqja si më poshtë.

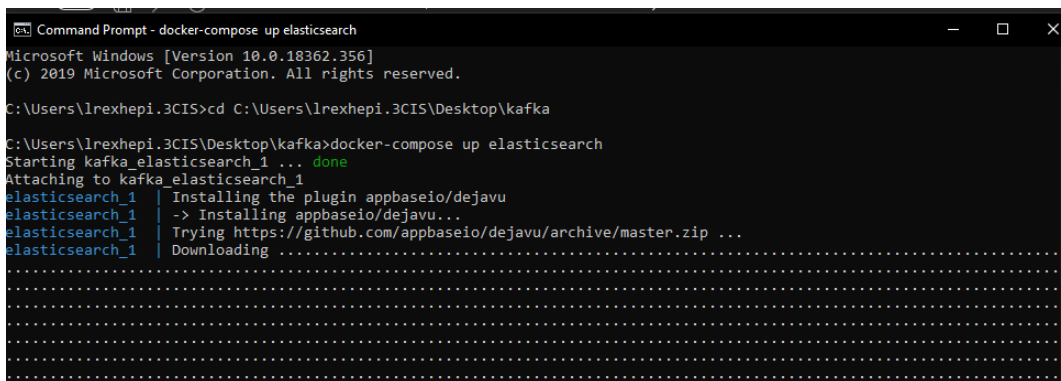
Mund të shihen pjesët e Kafka-s, cersionet si dhe mund të shfletohen apo të krijohet ndonjë topic, connector , gjëra që do të sqarojmë në vazhdimin e këtij punimi.



5.2.2 Komponentët e Elasticsearch dhe Postgres SQL dhe konfigurimi i tyre

Lartë u sqarua detajisht mënyra e instalimit të Kafkës përmes komandës docker-compose. Ngaqë file përmban parametra edhe të aplikacioneve tjera, në të njëjtën mënyrë mund të zbriten edhe Elasticsearch dhe Postgres duke përdorur komandën në Comand Prompt:

- **docker-compose up elasticseach**



```
Command Prompt - docker-compose up elasticsearch
Microsoft Windows [Version 10.0.18362.356]
(c) 2019 Microsoft Corporation. All rights reserved.

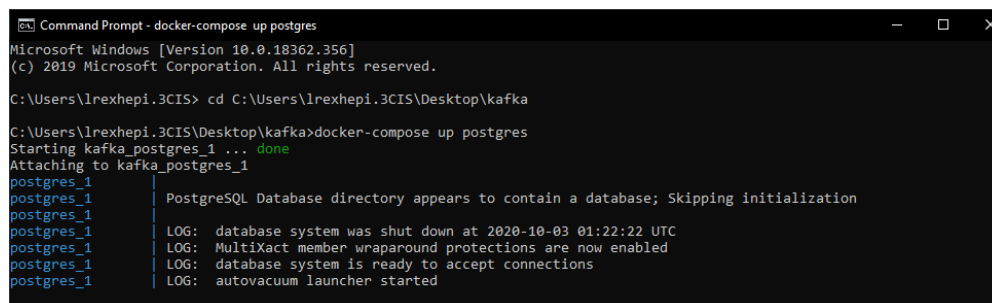
C:\Users\lrexhepi.3CIS>cd C:\Users\lrexhepi.3CIS\Desktop\kafka

C:\Users\lrexhepi.3CIS\Desktop\kafka>docker-compose up elasticsearch
Starting kafka_elasticsearch_1 ... done
Attaching to kafka_elasticsearch_1
kafka_elasticsearch_1 | Installing the plugin appbaseio/dejavu
kafka_elasticsearch_1 | -> Installing appbaseio/dejavu...
kafka_elasticsearch_1 | Trying https://github.com/appbaseio/dejavu/archive/master.zip ...
kafka_elasticsearch_1 | Downloading .....
```

Fig. 12- Konfigurimi i Elasticsearch përmes docker-compose

Linku <http://localhost:9200> shërben për të verifikuar se a është instaluar në rregull dhe a është aktive Elasticsearch. File docker-compose.yml përmban edhe një komponent shtesë e cila na mundëson qasjen përmes faqes në Elasticsearch. Nëse hyjmë në <http://localhost:9200/plugin/dejavu/> do të hapet faqja përmes të cilës mund të lexohen të dhënat, që do t'i sqarojmë më vonë gjatë këtij punimi.

- **docker-compose up postgres**



```
Command Prompt - docker-compose up postgres
Microsoft Windows [Version 10.0.18362.356]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\lrexhepi.3CIS> cd C:\Users\lrexhepi.3CIS\Desktop\kafka

C:\Users\lrexhepi.3CIS\Desktop\kafka>docker-compose up postgres
Starting kafka_postgres_1 ... done
Attaching to kafka_postgres_1
kafka_postgres_1 | PostgreSQL Database directory appears to contain a database; Skipping initialization
kafka_postgres_1 | LOG: database system was shut down at 2020-10-03 01:22:22 UTC
kafka_postgres_1 | LOG: MultiXact member wraparound protections are now enabled
kafka_postgres_1 | LOG: database system is ready to accept connections
kafka_postgres_1 | LOG: autovacuum launcher started
```

Fig. 13- Konfigurimi i Postgre SQL përmes docker-compose

Figura 13. tregon se baza e të dhënave është instaluar dhe është gati për të pranuar lidhjen. Është e nevojshme edhe një editor për ti menaxhuar të dhënat më lehtë, që të kemi qasje dhe ti menaxhojmë përmes user interface. Ne kemi përzgjedhur SQLelektron, një aplikacion Open Source, me mundësi shkarkimi <https://sqlectron.github.io/>. Mënyra e lidhjes me Bazë të dhënash dhe mundësitë e tjera që ofron ky aplikacion do t'i trajtojmë në vazhdim të këtij punimi.

5.2.3 Konfigurimi i konektorëve në Kafka

Siç është sqaruar në nën kapitullin 5.2.2. të këtij punimi, përmes web-it duhet të qasemi në <http://localhost:3030>. Kjo mundëson realizimin më të lehtë të konfigurimit. Gjithësesi, të gjitha konfigurimet mund të bëhen edhe me comanlline interface (CLI). Ky interface vetëm na lehtëson punën, që në mënyrë vizuale të kemi më të qartë procesin e konfigurimit.

Pasi që hyjmë tek WEB interface i Kafka-s duhet të shkojmë tek **Connectorse** pastaj **NEW** dhe do të hapet dritarja si më poshtë:

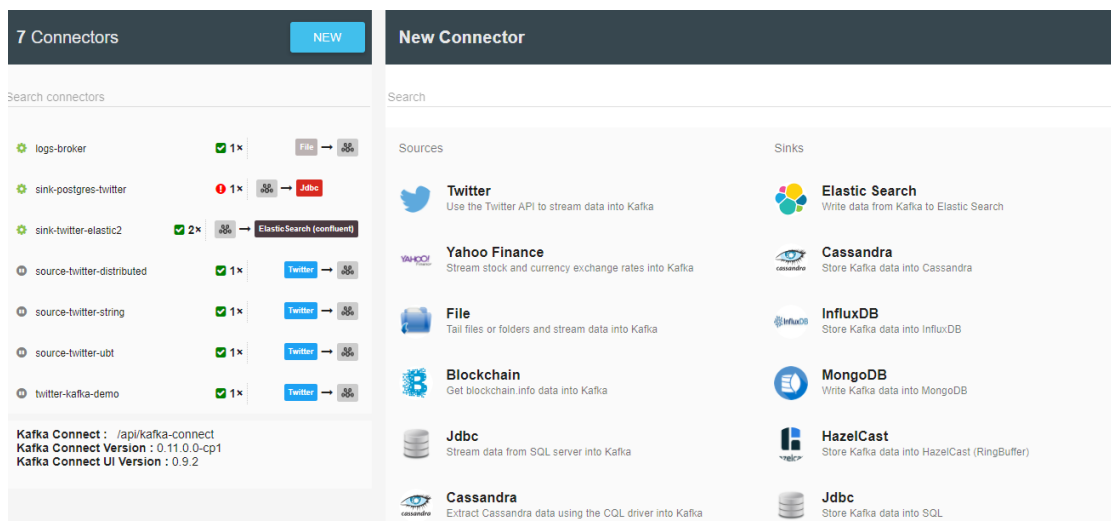


Fig. 14 - Interface-i i konektorëve në Kafka

Pjesa e djathtë shpjegohen koneksionet dhe çfare duam të zgjedhim për konfigurim. Konektorët në kafka ndahen nëdy pjesë:

- **Source Connectors** – Lidhja e konektorëve me aplikacionet, në rastin tonë Tweeter.
- **Sink Connectors** - Koneksioni i Kafka me platforma tjera.

Për realizim na duhen konfigurim i tre konektorev:

- **Twitter – Kafka** - për qasje direkte me Twitter API.
- **Kafka – Elasticsearch** – bartje nga Kafka-Topic në cluster-in e elasticsearch.
- **Kafka – PostgreSQL** – ruajtje e shënimeve në Postgres SQL.

5.2.4 Twitter – Kafka Connector

Më lartë kemi përmendur se Twitter API kërkon katër parametra, të cilët janë gjeneruar dhe përdoren për konfigurim. Këta katër parametra do të jenë si pjesë konfiguruese e koneksionit. Para koneksionit duhet krijuar Kafka-Topic, ku shënimet nga Twitter do të rruhen në Kafka-Cluster. Për të krijuar Kafka-Topic duhet të qasemi në Kafka përmes Comand Line Interface. Qasja me e shpejtë është përmes Docker, ku duke klikuar në butonin “>_”, Kafka paraqitet aktive.

Pasi të paraqitet dritarja e CLI , ekzekutojmë komandën:

kafka-topics --create --topic source-Twitter-ubt --partitions 2 --replication-factor 1 -zookeeper 127.0.0.1:2181

- “kafka-topic-ubt” – emri i topic-ut ku do të ruhen shënimet
- “partition 2” – përcaktojmë se sa particione për ruajtjen e shënimeve dëshirojmë t’i kemi (varësisht nga resurset e paisjes)
- “Replication-factor 1– përcaktojmë se sa kopje të këtij topic-ut do ti kemi
- “zookeeper 127.0.0.1:2181” – aktivizojmë zookeeper, i cili bën menaxhimin e proceseve në mes të Topic-ut dhe Kafka-Cluster.

Krijohet topicu pasi e ekzekutojmë komandën dhe **Kafka UI > Topics** na mundëson qasjen që ta shohim.

Meqë tash me topicu është krijuar mund të aktivizohet kafka-costumer përmes të cilit i monitorojmë shënimet që do të na vijnë nga Twitter përmes komandës:

kafka-console-consumer --topic demo-Twitter --bootstrap-server 127.0.0.1:9092

Tani infrastruktura është gati për të filluar me krijimin e Konektorit për Twitter. Kafka UI (Kafka WEB interface) na mundëson që shumë lehtë të krijojmë duke i përcjellur hapat: Connectors – New – zgjedhim Twitter si opcion dhe do të hapet dritarja si në figurën më poshtë. Për të krijuar koneksionin na nevojiten parametrat hyrës të cilët janë pri-konfiguruar më parë. Pasi të jenë plotësuar të gjithë parametrat konfigurues e klikojmë butonin “Create”.

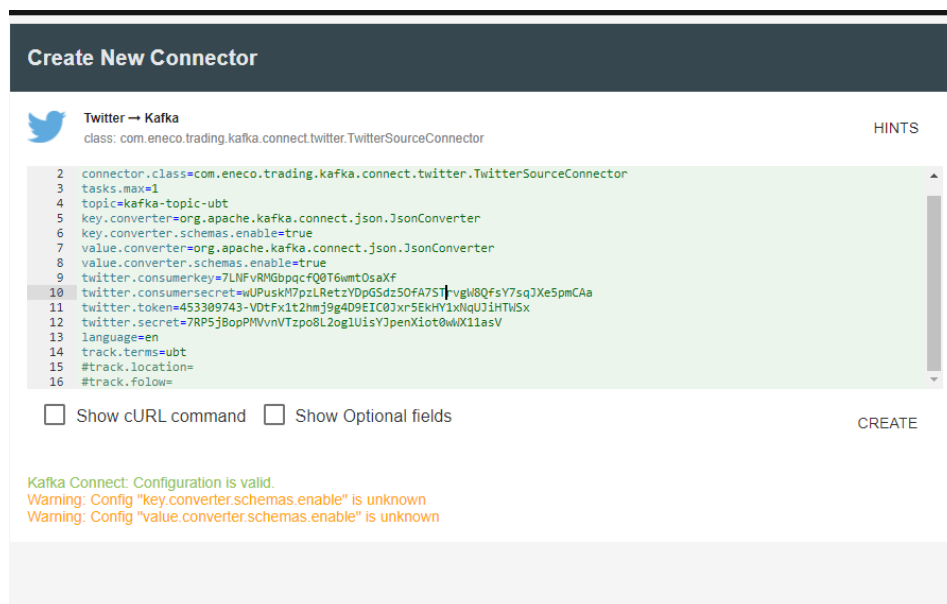


Fig. 15 - Konfigurimi i Twitter Connector

Duhet konsideruar se token-at dhe çelësat nga Twitter duhet të përdoren nga llogaria të cilën e kemi krijuar për Twitter API, të sqaruara me lartë.

Track.term sështë fjala kyçe / filter për shënimet nga Twitter. Në këtë rast ne kemi përdorur fjalen “**Kosovo**”. Mund të përdoren më shumë se një fjalë në filter si p.sh **track.terms=Kosovo** dhe në do ti pranojmë të gjitha shënimet që përmbajnë njërin nga këto fjalë.

Ky konektor e ka mundësinë që të bëjë filtrimin edhe në bazë të **lokacionit** (track.location) dhe në bazë të shfrytëzuesit të caktuar (track.follow). Në rastin tonë këto do t'i lemë si opcionale.

Pasi të krijohet konektori, tek lista e konektorëve krijohet Kafka UI e cila mund të modifikohet sipas nevojës.

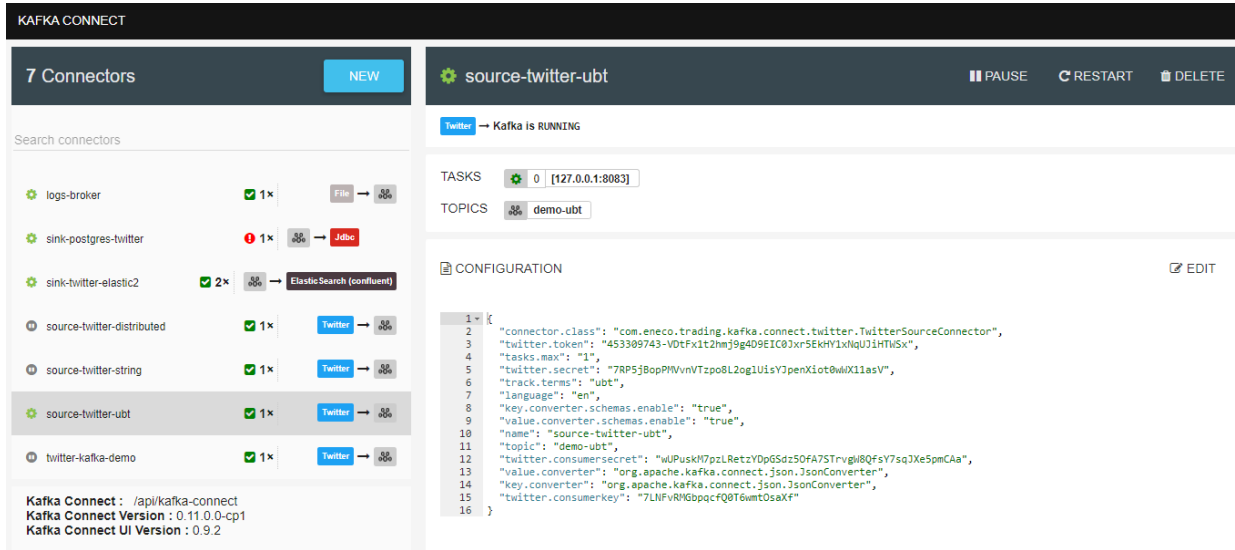


Fig. 16 – QasjanëTwitter Connector

5.2.5 Kafka – Elasticsearch Connector

Mënzra e këtij konektori nuk ka ndonjë ndryshim krijoet njëjtë si ato më lartë. Si solucion në këtë rast e zgjedhim **SINK Connector ElasticSearch (Confluent)**. Ky konektor do të lexoj shënimet nga Topic që e kemi kjuar më herët dhe do të shkuaj në ElasticSeach.**Topics** – duhet të ceket topic-u nga i cili do të lexojmë shënime, në rastin tonë është topic-u i krijuar nga ne “demo-twitter-ubt”.

Connection.url – është linku dhe porti ku është instaluar Elasticsearch, i cili në rastin tonë është instaluar lokalisht.

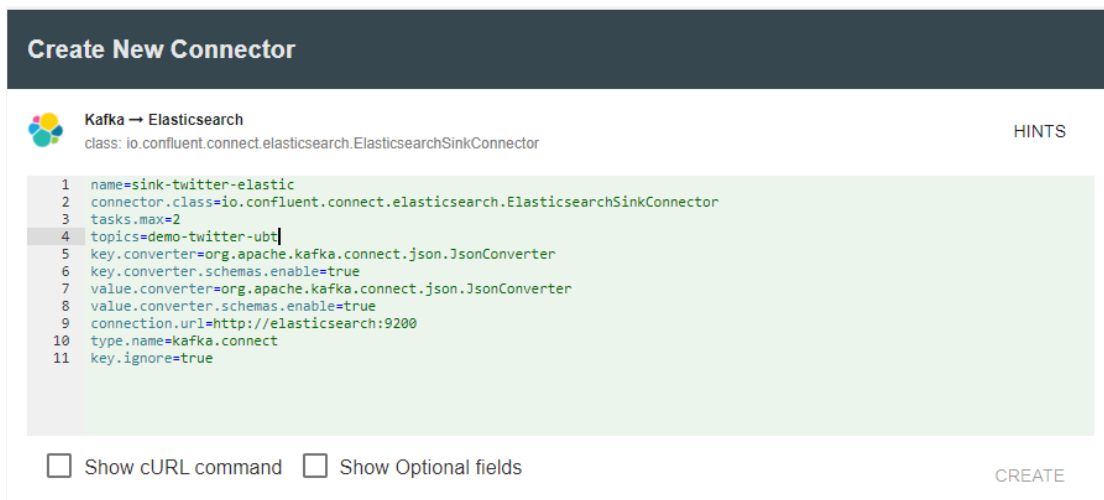


Fig. 17 - Konfigurimi i ElasticSearch Connector

Mbasi të krijohet koneksioni, konektori duhet të jetë “running” dhe atë mund ta modifikojmë, fshijmë ose pauzojmë sipas nevojës.

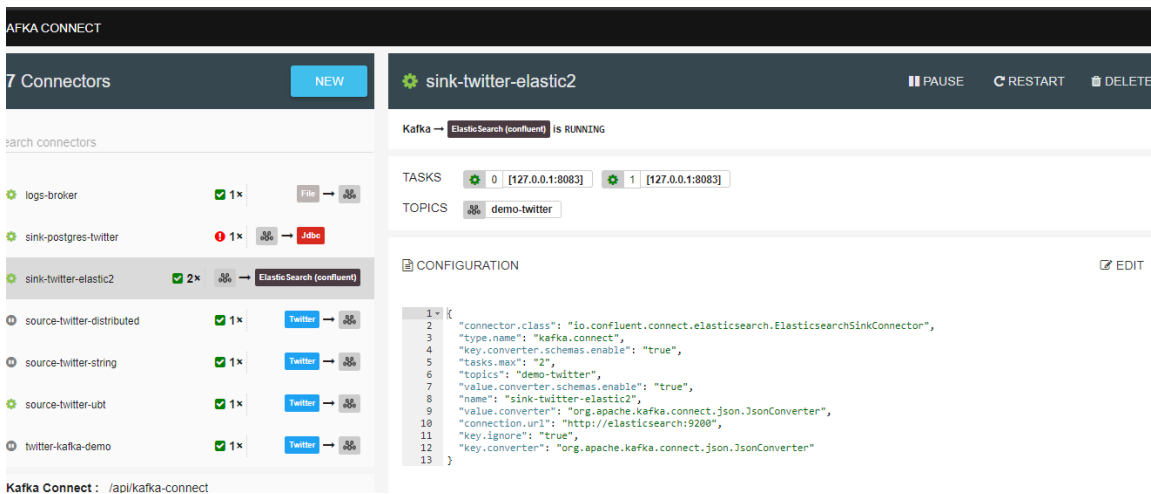


Fig. 18 – QasjanëElasticSearch Connector

5.2.6 Kafka – PostgreSQL

Në mënyrë të njejtë si më parë mund të krijojmë konektorin për t'u lidhur me Bazën e të dhënave dhe ruajtjen e shënimeve aty. Tek New Connector zgjedhim opcionin JDBC dhe pastaj i fusim parametrat e nevojshëm. Gjatë procesit të instalimit të PostgreSQL nëpërmjet docker-compose, janë përcaktuar do parametra te docker-compose.yml që shërben për krijimin e userit, fjalëkalimin ku skema vijon si:

environment:

POSTGRES_USER: **postgres**

POSTGRES_PASSWORD: **postgres**

POSTGRES_DB: **postgres**

Këto parametra mund t'i ndrojmë varësisht nga përzgjedhja jonë dhe ato do të na duhen të përdoren në konektorin si më poshtë:

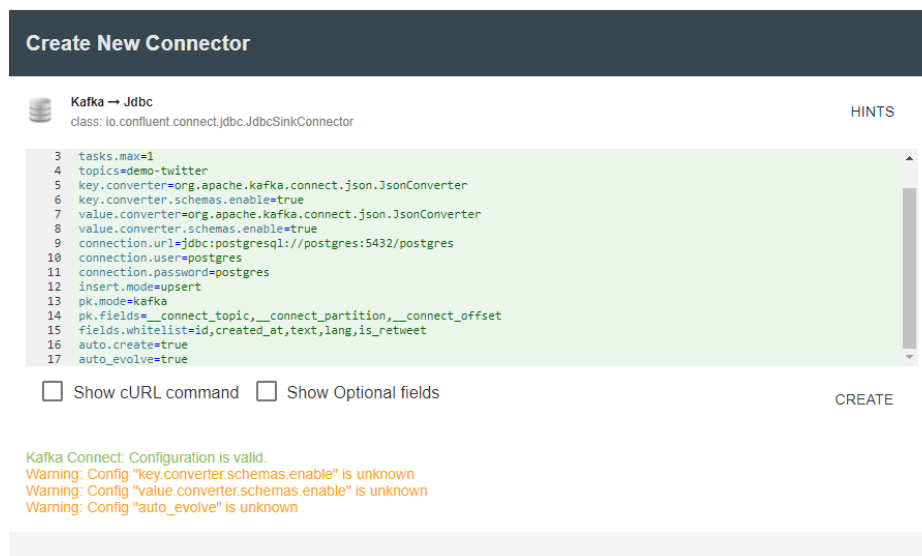


Fig. 19 - Konfigurimi i PostgreSQL Connector

Pasi të përfundojë procedura, klikohet Create dhe tek lista paraqitet konektori në Kafka UI si në rastet e lartpërmendura.

5.3 Leximi i të ardhurave nga Twitter dhe shënimi i tyre

Tani procesi për të pranuar shënime nga Tweeter është gati ku mund të shohim si paraqiten në Kafka-Topic, Elasticsearch dhe Bazë të dhënash.

5.3.1 Pranimi i shënimeve në Kafka

Që të testohet funksionaliteti duhet testuar në shënimet sipas konfigurimeve që kemi bërë.

Në faqen Tweeter kemi shkruar një tweet me fjalën “Kosovo”,ku Kafka që është konfiguruar më herët e pranon shënimin duke na njoftuar ne formen si mëposhtë



Fig. 20 – Test postim ne Twitter per koneskion.

```
"string","optional":true,"field":"display_url"},{"type":"string","optional":true,"field":"expanded_url"},{"type":"string","optional":true,"field":"url"}],{"optional":false,"name":"com.eneco.trading.kafka.connect.twitter.Url"},"optional":true,"field":"urls"},{"type":"array","items":{"type":"struct","fields":[{"type":"int64","optional":false,"field":"id"},{"type":"string","optional":true,"field":"name"},{"type":"string","optional":true,"field":"screen_name"}],"optional":false,"name":"com.eneco.trading.kafka.connect.twitter.UserMention"},"optional":true,"field":"user_mentions"},"optional":false,"name":"com.eneco.trading.kafka.connect.twitter.Entities","field":"entities"},"optional":false,"name":"com.eneco.trading.kafka.connect.twitter.Tweet"},"payload":{"id":1321116241331978240,"created_at":"2020-10-27T15:47:11.000+0000","user":{"id":453309743,"name":"Arben","screen_name":"@tafarben","location":"Struga","verified":false,"friends_count":12,"followers_count":3,"statuses_count":29},"text":"Kosovo is now independent country","lang":"en","is_retweet":false,"entities":{"hashtags":[],"media":[],"urls":[],"user_mentions":[]}}
```

Fig. 21 - Pranimi i Tweet-avenëkafka përmes Kafka-Consumer

```

▶ Key: { schema: null, payload: null }
▼ Value:
  ▶ schema:
    { type: struct, fields: [object Object],[object Object],[object Object],[object Object],[o
  ▼ payload:
    id: 1321116241331978200
    created_at: 2020-10-27T15:47:11.000+0000
    user:
      id: 453309743
      name: Arben
      screen_name: @tafarben
      location: Struga
      verified: false
      friends_count: 12
      followers_count: 3
      statuses_count: 29
      text: Kosovo is now independent country
      lang: en
      is_retweet: false
    entities: { hashtags: , media: , urls: , user_mentions: }

```

Fig. 22 – Paraqitja e të dhënave në Kafka-UI

5.3.2 Dërgimi i shënimeve në Postgre SQL dhe Elasticsearch

Gjëja më e rëndësishme deri tani është se pranimi i të dhënave në kohë reale nga Twitter në kafka është kryer me sukses. Pra, tani shënimet janë të ruajtura lokalisht në cluster-at e kafka-s dhe kjo ka qenë edhe një nga objektivat kryesore. Detyra jonë është që tani ato shënime t'i ruajmë në platforma, të cilat mundësojnë leximin dhe manipulimin e lehtë me to.

Elasticsearch:

Qasjen krijohet përmes UI ku instalimi është krijuar nepërmjet linkut më poshtë <http://localhost:9200/plugin/dejavu/> dhe në kohë reale i shohim shënimet për herë të parë. Këtu tani kemi një pasqyrë të rezultatit tonë të deritanishëm.

ElasticSearch ka disa mundësi për manipulim me të dhëna si:

- Filtrimi i të dhënave në UI
- Exportimi i të dhënave përmes UI në formati JSON⁵
- Ekzekutimi i queries (pytësorët) në SQL

⁵ JSON - JavaScript Object Notation

Poshtë janë postuar lista e figurave të dhënave në Elasticsearch përmes Dejavu shtojcës:

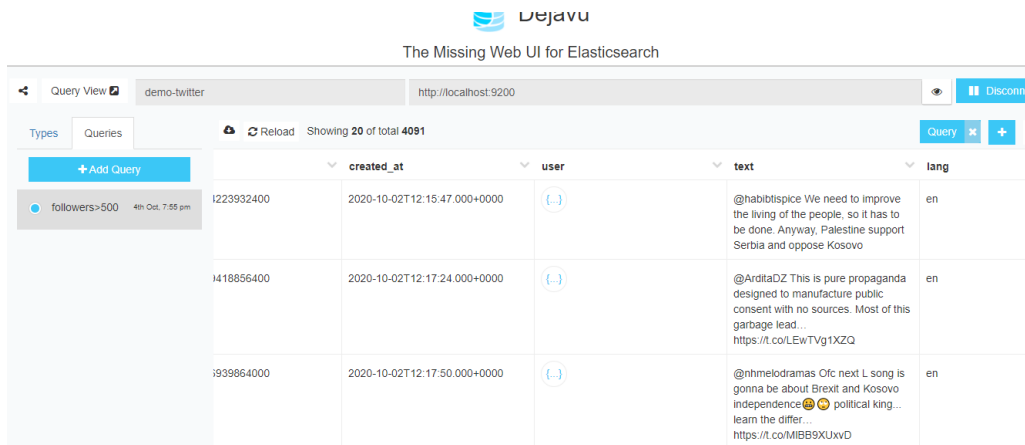


Fig. 23 – Shënimet dhe paraqitja përmes UI – Dejavu

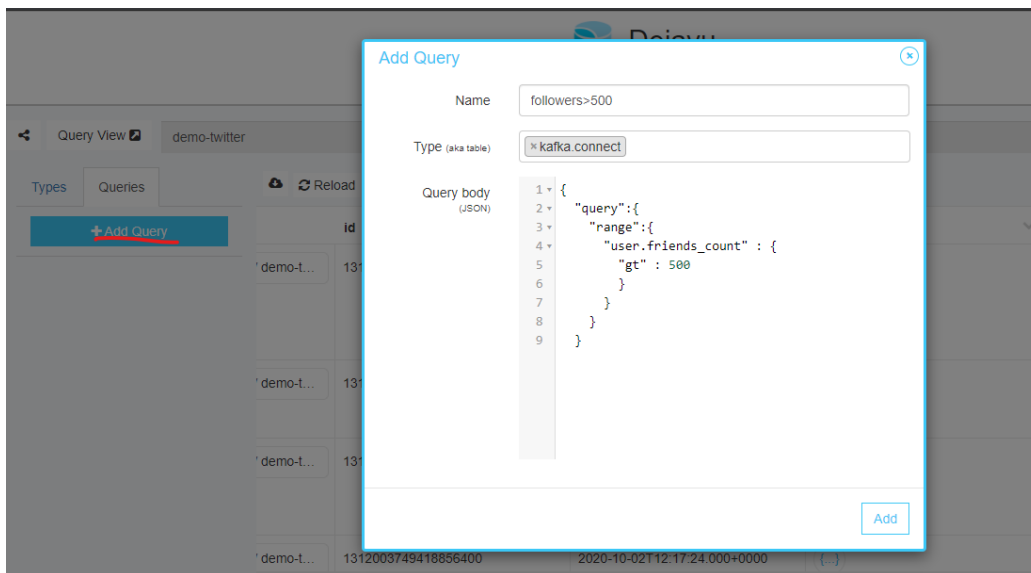


Fig. 24 - SQL Query përmes UI – Dejavu

PostgreSQL:

Për tu bërë leximi më i lehtëi të dhënave në databazë SqlElectron si opensource.Pasi të shkruhën në databazë racionale, fitohet kuptiumi plotë për procesin.

Tani shënimet i shohim në formatin më të zakonshëm dhe është e lehtë të kuptojmë dhe manipulojmë me të dhëna përmes SQL, si në çdo bazë të dhënash tjetër.

Së pari duhet të qasemi përmes editorit duke ja dhënë kredencialet që i kemi shkruar gjatë implementimit të platformave.

The screenshot shows a 'Server Information' configuration form. The 'Name' field is set to 'postgres'. The 'Database Type' is set to 'PostgreSQL'. The 'Server Address' is '127.0.0.1'. The 'User' is 'postgres'. The 'Password' is masked with dots. The 'Initial Database/Keyspace' is 'postgres' and the 'Initial Schema' is 'Schema'. There are checkboxes for 'SSH Tunnel' and 'Filter'. At the bottom, there are buttons for 'Test', 'Duplicate', 'Cancel', 'Save', and 'Remove'.

Fig. 25 - Krijimi i koneksionit në PostgreSQL

Pasi që qasja është kryer me sukses ne mund të shohim të dhënat tona të ruajtura në tabela dhe të manipulojmë me ato të dhëna, varësisht prej rastit të studimit.

The screenshot shows a PostgreSQL query editor. The query is: `SELECT * FROM 'public'.'demo-twitter' where text like ('Agreement%');`. The results table has columns: `__connect_to...`, `__connect_pa...`, `__connect_off...`, `created_at`, `id`, `text`, `lang`. The results show several rows of data from the 'demo-twitter' table.

__connect_to...	__connect_pa...	__connect_off...	created_at	id	text	lang
demo-twitter	1	630	2020-10-01T09:...	131160442759...	#OBC Washington agreement, high expectations in Kosovo ht...	en
demo-twitter	1	749	2020-10-01T11:...	131162614702...	Almost a month after the agreement signed by #Kosovo and #S...	en
demo-twitter	1	778	2020-10-01T11:...	131162943236...	What calculation is behind Trump's mediation in Serbia-Kosovo...	en
demo-twitter	1	813	2020-10-01T11:...	131163404909...	RT @LowyInstitute: What calculation is behind Trump's mediati...	en
demo-twitter	1	943	2020-10-01T13:...	131165996987...	Today the Swedish embassy signed a new agreement with @S...	en
demo-twitter	1	971	2020-10-01T13:...	131166566948...	Cost Norway from @MareFDW: Decision @LowyInstitute...	en

Fig. 26 - Të dhënat në PostgreSQL

6.Rastet për studim (Case study)

Për gjithë projektin kemi marrë tre shembuj që do përdorim që të arrijmë feedback ne fund edhe atë:

- “Ronaldo dhe Messi”
- “Kosovo”
- “Pepsi dhe CocaCola

Nga këto shënime jemi përpjekur të bëjmë një analizë dhe të nxjerrim disa rezultate relativisht të thjeshta për t’u analizuar, megjithatë shumë interesante dhe të rëndësishme. Për analizë është përdorur *Jupyter Notebook*, i cili përfshinë gjuhën programuese *Python*.

6.1 Rasti për studim (I) – Të dhënat me fjalën kyçe “Kosovo”

Në këtë rast studimi do të përdoret metoda e numërimit të fjalëve të caktuara në përmbajtjen e tekstit të postimeve me fjalën kyçe “Kosovo”. Thënë më saktë,do të marrim si shembull kandidaturën e zgjedhjeve të fundita presidenciale që u zhvilluan në SHBA dhe do të numërohen se sa herë në këto postime është përmendur kandidati Donald Trump në krahasim me kundër kandidatin demokrat Joe Biden. Duke pasur parasysh se marrëveshja në mes Kosovës dhe Serbisë ishte nënshkruar pak kohë para fushatës elektorale në SHBA, ne do të tentojmë të vërtetojmë hipotezën tonë, se a është përdorur kjo marrëveshje për fushatë, pra për pikë politike nga presidenti aktual Donald Trump.

Shënimet e eksportuara nga *Elasticsearch* (të sqaruara më lartë në punim) janë në formatin *JSON*. Duke përdorur *Jupyter Notebook* këtë file mund ta lexojmë duke shkruar kod në gjuhën programuese *Python*.Kjo analizë do na ndihmojë të paraqesim një të gjetur interesante dhe atraktive, që ka të bëjë me zgjedhjet presidenciale në SHBA. Pra pyetja hulumtuese këtu do ishte: cili nga kandidatët amerikan për president është përmendur në postimet që kanë të bëjnë me Kosovën dhe sa është shfrytëzuar tema e Kosovës për fushatë politike.

Python na mundëson leximin e shënimeve nga JSON file dhe të paraqesim në mënyrë vizuale, gjë që shton atraktivitetin e analizës. Në vijim do të shpjegojmë edhe mënyrat e ndryshme se si mund të manipulohet me të dhëna.

Poshtë janë shpjeguar hapat si kemi arritur deri aty.

```
In [1]: import json
import pandas as pd
from pandas.io.json import json_normalize #package for flattening json in pandas df
import seaborn as sns
```

```
In [2]: #Load json object
with open(r'C:\Users\RBK0\Desktop\data_twitter_final.json', encoding='UTF-8') as f:
    d = json.load(f)
```

```
In [3]: tweets = json_normalize(d)
tweets.head(10)
```

C:\Users\RBK0\anaconda3\lib\site-packages\ipykernel_launcher.py:4: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize instead after removing the cwd from sys.path.

```
Out[3]:
```

	_index	_type	_id	_score	_source.id	_source.created_at	_source.user.id	_source.user.name	_source.user.screen_nam
0	demo-twitter	kafka.connect	demo-twitter+1+3458	1	1312003199805649000	2020-10-02T12:15:13.000+0000	1273716694561837000	Serbian Ban-Evasion Enthusiast	Redacted_Shari
1	demo-twitter	kafka.connect	demo-twitter+1+3459	1	1312003273797316600	2020-10-02T12:15:30.000+0000	477138227	Jessica	xhxsik

Fig. 27 - Importimi i paketave / importimi dhe paraqitja e shënimeve përmes Python.

In [1]: Në këtë pjesë të kodit në Python janë importuar paketat si:

- **pandas** që bën të mundëshme manipulimin me të dhëna, dhe
- **Json** që bën të mundëshme leximin e file-it të formatit json.

In [2]: Lexim i *Json* file nga direktoriumi ku file-i është ruajtur lokalisht.

In [3] : *json.normalize* e cila bën leximin e strukturës së file-it *JSON* dhe strukturon të dhënat që të jetë më i lehtë konfigurimi i tyre.

```

In [7]: for col in tweets.columns:
        print(col)

INDEX
TYPE
ID
SCORE
SOURCE_ID
SOURCE_CREAT
SOURCE_USER_ID
SOURCE_USER_NAME
SOURCE_USER_SCRNAME
SOURCE_USER_LOCATION
SOURCE_USER_VERIFIED
SOURCE_USER_FRIENDSCNT
SOURCE_USER_FOLLOWERSCNT
SOURCE_USER_STATUSES CNT
SOURCE_TEXT
SOURCE_LANG
SOURCE_ISRETWEET
SOURCE_ENTITIES_HASHTAGS
SOURCE_USER_MEDIA
SOURCE_USER_URLS
SOURCE_ENTITIES_USERMENTIONS

```

Fig. 28 - Listimi i kolonave

In [7] : Në këtë pjesë të kodit janë shtypur / listuar kolonat e tabelës së të dhënave. Kjo na mundëson qasje me të lehtë nëshënime, duke ditur strukturën e tabelës dhe të dhënat që i përmbanë ky set i shënimeve.

```

In [8]: a = tweets.SOURCE_TEXT.str.count("realDonaldTrump")

In [9]: tweets['TRUMP_MENTIONS'] = a

In [10]: b = tweets.SOURCE_TEXT.str.count("Biden")

In [11]: tweets['BIDEN_MENTIONS'] = b

In [12]: trump = len(tweets[(tweets['TRUMP_MENTIONS'] == 1) & (tweets['SOURCE_USER_FOLLOWERSCNT'] > 20)])

In [13]: biden = len(tweets[(tweets['BIDEN_MENTIONS'] == 1) & (tweets['SOURCE_USER_FOLLOWERSCNT'] > 20)])

In [14]: trump
Out[14]: 395

In [15]: biden
Out[15]: 144

```

Fig. 29- Filtrimi i të dhënave

In [8] dhe In [9] :Në variablen (**a**) e cila është e tipit “list” janë filtruar dhe ruajur shënimet që në kolonen (source_text) përmajnë fjalën “realDonaldTrump”.

In [10] , In [11] :Në variablen (**b**), e cila është e tipit “list”, janë filtruar dhe ruajur shënimet që në kolonen (source_text) përmajnë fjalën “Biden”.

In [12] , In [13]: Në variablen (**trump**), respektivisht në variablen (**biden**), janë ruajtur numri i tweet-ave që përmbajnë fjalën “realDonaldTrump” dhe “Biden” nga shfrytëzuesit që kanë më shumë se 20 ndjekës. Ky kusht është vendosur që të dhënat të janë sa ma kredibile në mënyrë që shfrytëzuesit, të cilët nuk janë të rregullt në Twitter, të mos hyjnë në kalkulim.

In [14] , In [15]: Në këtë pjesë është paraqitur në ekran rezultati nga kalkulimet më larta për dy variablat që i kanë kalkuluar/ ruajtur shënimet për kushtet që i kemi vendosur.

Duke shfrytëzuar paketën **matplotlib** Python përmes Jupyter Notebook, në menyrë vizuale mund ti paraqesim rezultatet e lartpërmendura. Kur shohim rezultatin në mënyrë vizuale mund të kemi një konkluzion të qartë për rezultatin e arritur. Të dhënat e vizualizuara janë paraqitur në figurat 28 dhe 29.

Rezultatet na sjellin në konkludim se është e logjikshme hipoteza e ngritur gjatë fushatës politike që kemi përmendur. Pra në pjesën dërmuese të postimeve në lidhje me shembullin tonë emri i Donald Trump është përdorur në krahasim me Joe Biden.

```

import matplotlib.pyplot as plt
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
persons = ['Donald Trump', 'Joe Biden']
data = [trump, biden]
ax.bar(persons,data)
plt.show()

```

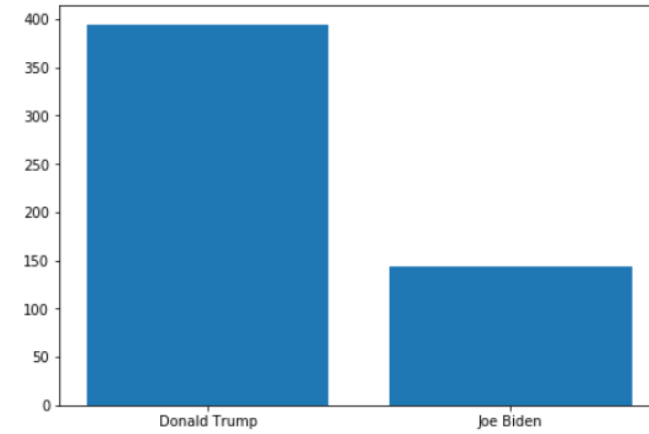
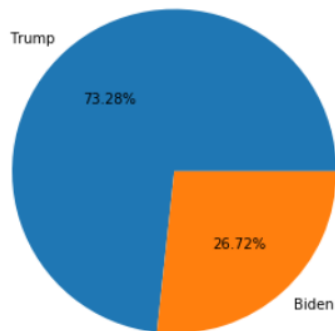


Fig. 30-31 Vizualizimi i shënimeve përmes Jupyter Notebook

```

from matplotlib import pyplot as plt
import numpy as np
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
ax.axis('equal')
CANDIDATES = ['Trump', 'Biden']
data = [trump,biden]
ax.pie(data, labels = CANDIDATES, autopct='%1.2f%%')
plt.show()

```



6.2 Rasti për studim (II) – Të dhënat me fjalët kyçe “Ronaldo dhe Messi”

Për këtë rast studimi, të dhënat nga Twitter i kemi marrë duke përdorur fjalët kyçe Messi dhe Ronaldo. Me qëllim të shtjellimit të analizës së përmbajtjes së tekstit kemi provuar që të marrim një shembull që mund të na sjellë shumë shënime, dhe si ide kemi përzgjedhur këta dy futbollist, për të cilët shfrytëzuesit në internet bëjnë mjaft shumë postime. Shënimet nga Twitter për këta dy futbollist janë marrë për periudhën dy ditore (48 orë). Qysh në shikim të parë, pas marrjes së të dhënave, mund të konkludojmë që futbollisti Messi është dukshëm më i popullarizuar në Twitter.

Për të arritur rezultat sa më të detajuar, kemi tentuar që këto shënime t'i analizojmë në mënyrë më të thellë. Ideja ka qenë që të bëjmë analizën e tekstit dhe të gjejmë feedback-un e tekstit dhe të arrijmë në konkludim se për cilin lojtar ka më shumë komente negative.

Përmes Python duke përdorur paketën “**textblob**”[33], kemi analizuar secilin tweet që kemi pranuar dhe kemi përcaktuar polarizimin dhe subjektivitetin e përmbajtjes së teksteve për 2 lojtarët në fjalë.

TextBlob është librari që përdoret për Python, e cila është ndërtuar të përdoret për procesimin e të dhënave tekstuale duke përdorur “Natural Language Processing (NLP)”. Një nga metodat e kësaj librarie, që do ta përdorim në këtë rast studimi, është metoda “**Sentiment.Polarity**” edhe “**Sentiment.Subjectivity**”.

Gjatë përdorimit të kësaj metode, ajo na kthen vlerat numerike nga -1.0 deri në 1.0 për polaritet (ku -1.0 është shumë negativ ndersa 1.0 është shumë pozitiv) dhe nga 0.0 deri në 1.0 për subjektivitet (ku 0.0 është shumë objektive ndersa 1.0 është shumë subjektive). Më poshtë përmes figurave është paraqitur mënyra e realizimit duke përdorur **textblob**si paketë e Python dhe analiza e përmbajtjes së tekstit.

```

In [2]: import json
        from textblob import TextBlob

In [3]: with open(r'C:\Users\lrexhepi.3CIS\Downloads\MessiRonaldo.json', encoding='UTF-8') as f:
        messi_ronaldo = json.load(f)

In [4]: blobs = [(
        tweet['_source']['text'],
        TextBlob(tweet['_source']['text']),
        ) for tweet in messi_ronaldo];

In [5]: #Messi Polarization
total_messi_tweets = [b for b in blobs if 'messi' in b[0].lower()]
messi_positive = [b for b in blobs if 'messi' in b[0].lower() and b[1].sentiment.polarity > 0]
messi_neutral = [b for b in blobs if 'messi' in b[0].lower() and b[1].sentiment.polarity == 0]
messi_negative = [b for b in blobs if 'messi' in b[0].lower() and b[1].sentiment.polarity < 0]
#Messi subjectivity
messi_objectivity = [b for b in blobs if 'messi' in b[0].lower() and b[1].sentiment.subjectivity <= 0.5]
messi_subjectivity = [b for b in blobs if 'messi' in b[0].lower() and b[1].sentiment.subjectivity > 0.5]
#Ronaldo Polarization
total_ronaldo_tweets = [b for b in blobs if 'ronaldo' in b[0].lower()]
ronaldo_positive = [b for b in blobs if 'ronaldo' in b[0].lower() and b[1].sentiment.polarity > 0]
ronaldo_neutral = [b for b in blobs if 'ronaldo' in b[0].lower() and b[1].sentiment.polarity == 0]
ronaldo_negative = [b for b in blobs if 'ronaldo' in b[0].lower() and b[1].sentiment.polarity < 0]
#Ronaldo subjectivity
ronaldo_objectivity = [b for b in blobs if 'ronaldo' in b[0].lower() and b[1].sentiment.subjectivity <= 0.5]
ronaldo_subjectivity = [b for b in blobs if 'ronaldo' in b[0].lower() and b[1].sentiment.subjectivity > 0.5]

In [9]: print(f'Total tweets for Messi: {len(total_messi_tweets)}')
        print('-----')
        print(f'Positive tweets about Messi: {len(messi_positive)}')
        print(f'Neutral tweets about Messi: {len(messi_neutral)}')
        print(f'Negative tweets about Messi: {len(messi_negative)}')
        print('-----')
        print(f'Messi objectivity tweets : {len(messi_objectivity)}')
        print(f'Messi subjectivity tweets: {len(messi_subjectivity)}')
        print('-----')
        print(f'Total tweets for Ronaldo: {len(total_ronaldo_tweets)}')
        print('-----')
        print(f'Positive tweets about Ronaldo: {len(ronaldo_positive)}')
        print(f'Neutral tweets about Ronaldo: {len(ronaldo_neutral)}')
        print(f'Negative tweets about Ronaldo: {len(ronaldo_negative)}')
        print('-----')
        print(f'Ronaldo objectivity tweets : {len(ronaldo_objectivity)}')
        print(f'Ronaldo subjectivity tweets: {len(ronaldo_subjectivity)}')

Total tweets for Messi: 29470
-----
Positive tweets about Messi: 13859
Neutral tweets about Messi: 12097
Negative tweets about Messi: 3514
-----
Messi objectivity tweets : 19357
Messi subjectivity tweets: 10113
-----
Total tweets for Ronaldo: 16374
-----
Positive tweets about Ronaldo: 7570
Neutral tweets about Ronaldo: 6393
Negative tweets about Ronaldo: 2411
-----
Ronaldo objectivity tweets : 10958
Ronaldo subjectivity tweets: 5416

```

Fig. 32 – Analiza e polarizimit të tekstit përmes sentiment.polarity

In [2]: Në këtë pjesë të kodit në Python janë importuar paketat si:

- **textblob** për analizën e tekstit
- **Json** që bënë të mundëshme leximin e file-it të formatit json.

In [3]: Lexim i Json file nga direktoriumi ku file-i është ruajtur lokalisht.

In [4] : Në këtë pjesë te kodit është krijuar array i cili mundëson ruajtjen e kolonës text nga shënimet që janë mbledhur në Json.

In [5] : Në këtë pjesë kodi ka filtruar tekstet nëpërmjet “**Sentiment.Polarity**” edhe “**Sentiment.Subjectivity**” të librarisë “textBlob”. Vlerat e kthyer (nga -1.0 deri në 1.0) janë kategorizuar u komentet negative janë konsideruar si vlera negative, vlerat “0” janë konsideruar si neutrale, ndërsa vlerat pozitive si komente pozitive. Në mënyrë të njëjtë është trajtuar edhe subjektiviteti i tekstit.

In [9] : Kjo pjesë e kodit është vetëm shtypje (paraqitje) e rezultateve të kalkuluara më lartë.

Nga këto vlera, që janë rezultat i kalkulimeve të sqaruara më lartë, mund të bëjmë një analizë të shpejtë lidhur me rastin tonë të studimit dhe hipotezën që e kemi ngritur për këtë rast studimi :

Total tweets for Messi: 29470 – from 45,844 (**64%**)

Positive tweets about Messi: 13859 – **47%**

Neutral tweets about Messi: 12097 – **41%**

Negative tweets about Messi: 3514 – **12%**

Messi objectivity tweets : 19357 – **66%**

Messi subjectivity tweets: 10113 – **34%**

Total tweets for Ronaldo: 16374 – from 45,844 (**36%**)

Positive tweets about Ronaldo: 7570 – **46%**

Neutral tweets about Ronaldo: 6393 – **29%**

Negative tweets about Ronaldo: 2411 – **15%**

Ronaldo objectivity tweets : 10958 – **67%**

Ronaldo subjectivity tweets: 5416 – **33%**

6.3 Rasti për studim (III) – Të dhënat me fjalëtkyçe “Pepsi dhe CocaCola”

Duke pasur në konsideratë se kompanitë vendore në vendin tonë nuk kanë mjaftueshëm postime në Twitter, për shtjellim të studimit tonë kemi marrë dy kompani të mëdha, të huaja, që në fakt janë konkurrent kryesor në botë për produktet e tyre (pjetet jo-alkoolike). Nga pikëpamja komerciale ky rast studimi për këto dy kompani të mëdha botërore mund të jetë shumë i dobishëm për kompanitë që dëshirojnë të bëjnë analiza nga rrjetet sociale për produktet e tyre, opinionin e shfrytëzuesve, gjetjen e kënaqësisë së klientave të tyre dhe shumë analiza tjera, varësisht nga veprimtaria e biznesit.

Te dhënat e marra nga Twitter janë filtruar me fjalët kyçe “Pepsi” dhe “CocaCola” (Coca Cola) për periudhën dy ditore (48 orë). Në krahasim me rastet e studimit më lartë ku rezultati i analizave është paraqitur vetëm në mënyrë numerike si statistikë, për këtë rast të studimit kemi tentuar që në bazë të analizave të japim një rezultat konkret, duke pasur parasysh anën komerciale, reputacionin dhe interesin ekonomik të këtyre kompanive.

Interesi i këtyre kompanive, por edhe i çdo kompanie tjetër, do të ishte që të kenë sa më të qartë vlerësimin dhe opinionin e shfrytëzuesve të tyre në internet dhe nga këto analiza të marrin vendime për përmirsim, ndryshim apo ndërrim të produkteve të tyre apo mënyrës së lansimit të produkteve në treg.

Duke përdorur Python dhe libraritë e saj si “**heapq**” dhe “**textBlob**” kemi paraqitur përmbajtjen e Tweet-ave më negative nga shfrytëzuesit në internet për këto dy kompani. Poashtu, si në rastin për studim II, do të krahasojmë këto 2 kompani në bazë të komenteve negative dhe pozitive.

Heapq (Heap queue algorithm) është një librari e Python ku ne kemi përdor metodat:

- **heapq.nlargest(n, iterable, key=None)**
- **heapq.nsmallest (n, iterable, key=None)**

Këto dy metoda kthejnë një listë me “n” elemente më të mëdhenj/ më të vegjël nga baza e të dhënave e përcaktuar nga një koeficient i definuar. Në rastin tonë koeficienti është vlera nga -1.0 derinë +1.0, nga metoda “sentimet.polarity” e librarisë textBlob.

Bazuar në këtë mund të shfaqim lirshëm komentet pozitive dhe negative nga të gjitha postimet në Twitter në lidhje me kompanitë. Në dy shembujt më poshtë është paraqitur mënyra se si kemi bërë krahasimin e dy kompanive në bazë të komenteve pozitive apo negative, si dhe shfaqjen e komenteve më pozitive dhe atyre më negative nga shfrytëzuesit në Twitter.

```
In [6]: stats = {
        'VERY POSITIVE': 0,
        'POSITIVE': 0,
        'NEUTRAL': 0,
        'NEGATIVE': 0,
        'VERY NEGATIVE': 0,
    }
    for text, blob in blobs:
        polarity = blob.sentiment.polarity
        if polarity < -0.5:
            stats['VERY NEGATIVE'] += 1
        elif polarity < 0:
            stats['NEGATIVE'] += 1
        elif polarity == 0:
            stats['NEUTRAL'] += 1
        elif polarity < 0.5:
            stats['POSITIVE'] += 1
        elif polarity <= 1:
            stats['VERY POSITIVE'] += 1
        else:
            raise RuntimeError(f'Unknown polarity: {polarity}')
    print(stats)

{'VERY POSITIVE': 780, 'POSITIVE': 1369, 'NEUTRAL': 3410, 'NEGATIVE': 1288, 'VERY NEGATIVE': 216}
```

Fig. 33 – Analiza e polarizimit të tekstit përmes `sentiment.polarity` - 2

In [6]: Kjo pjesë e kodit paraqet një mënyrë më ndryshe të përcaktimi dhe kategorizimit të vlerave të polarizimit të tekstit. Duke kaluar me “loop” nëpër secilin pjesë të kolones “text” bëjmë kalkulimin e secilës kategori në bazë të vlerave të përcaktuara.

Nga analiza e të dhënave të mbledhura për dy kompanitë rezultatet e përmbajtjes së tekstit janë këto:

Pepsi: Total Tweets: **7063**

'VERY POSITIVE': **780**, 'POSITIVE': **1369**, 'NEUTRAL': **3410**, 'NEGATIVE': **1288**,
'VERY NEGATIVE': **216**

CocaCola: Total Tweets: **3537**

'VERY POSITIVE': **523**, 'POSITIVE': **817**, 'NEUTRAL': **1683**, 'NEGATIVE': **450**,
'VERY NEGATIVE': **64**

Nga këto të dhëna mund të konkludojmë se Pepsi është shumë më i popullarizuar në Tweet-er se sa CocaCola, në bazë të numrit të postimeve për periudhën e njëjtë kohore. Sa i përket komentet negative Coca Cola me 14% qëndron pak më mirë se Pepsi, i cili ka rreth 20% komente negative në Twitter.

Qëllimi jonë në këtë studim është që të mund të shohim përmbajtjen e komenteve negative dhe pozitive në lidhje me kompanitë që kemi marrur si shembull.

Në figurën më poshtë është paraqitur pjesa e kodit në Python përmes së cilës janë paraqitur 10 përmbajtjet më pozitive dhe 10 më negative.

```
In [9]: print('10 MOST POSITIVE TWEETS')
print('====')
for text, blob in heapq.nlargest(10, blobs, key=lambda tup: tup[1].sentiment.polarity):
    print(text)
    print('-----')

# 3 most negative tweets

print('10 MOST NEGATIVE TWEETS')
print('====')
for text, blob in heapq.nsmallest(10, blobs, key=lambda tup: tup[1].sentiment.polarity):
    print(text)
    print('-----')
```

Fig. 34 – Paraqitja e teksteve në bazë të polarizimit

Për ilustrim si rezultat do të paraqesim disa tekste më pozitive dhe negative të postuara në Twitter për kompaninë Coca Cola:

10 MOST POSITIVE TWEETS

====

CocaCola is the best!

Great job, Coca-Cola Consolidated! Community partnerships are so valuable.

<https://t.co/zHrPghMOse>

Awesome , Another week, another awesome SHI Upgrade! @CocaCola @graciesdream have a coke with me <3

RT @WVBeverageAssoc: Great job, Coca-Cola Consolidated! Community partnerships are so valuable.

<https://t.co/zHrPghMOse>

@bschieving @CocaCola The best virtual coke date 🥰

@Frankie10117 @CocaCola Plzzz!! Yesss!! I'll share one with you :D

@CocaCola @realUNOgame Thankyou for giving me beautiful memories with my family during the lockdown. I'd share one with you, my friend! ❤️

10 MOST NEGATIVE TWEETS

=====

Disgusting behavior by South Africans

RT @Unathi_Kwaza: Disgusting behavior by South Africans <https://t.co/HLYnIjk5fj>

Things are a bit grim these days. Here's something to brighten it a smidge @ccfreestyle @Spotify #ChristmasIsComing @CocaCola

@CocaCola Yes Danny, always amused me that it got passed the PC Police / Disgusted from Tonbridge Wells, but wait PC di... <https://t.co/u6o1ëkBQtb>

@patarimusic Worst management thats why #cocacola should always use youtube #worstmanagment #cokefest2020

Is this some god awful Coke advert song?
#TOTP

Idiotic!
<https://t.co/4VI2A98C3M> @CocaCola

i hate this i hate this i hate this @CocaCola

7.KONKLUZIONI

Duke u bazuar gjithmonë në rëndësinë që kanë të dhënat në ditët e sotme është shumë e rëndësishme që ato të dhëna t'i shfrytëzojmë sa më shumë që është e mundur, duke përfituar në mënyrë individuale apo shoqërore nga rezultatet e tyre. Në plotë kuptimin e fjalës mund të themi se të dhënat, konkretisht të dhënat nga rrjetet sociale që janë pa pagesë dhe duke përdorur platforma Open Source pa pagesë, mund të bëjnë ndryshime të mëdha në veprime dhe vendimmarje për secilin nga ne.

Edhe pse implementimi i tërë këtij procesi kërkon së pari përvojë dhe një njohuri të përgjithshme rreth disa fushave të teknologjisë informative, në këtë punim u vërtetua se kjo mund të realizohet me pak mund dhe përkushtim dhe në fund rezultati i fituar do dal mjaft interesant dhe inkurajues, që të punohet edhe më tutje në këtë drejtim, pra në zhvillimin e mekanizmave të tillë.

Në rastin tonë konkret e pamë se si nga të dhënat e marra nga Twitter mund të paraqesim disa të gjetura që mund të përdoren si mbështetje për veprime dhe vendimmarje të mëtejshme. Në bazë të pyetjeve hulumtuese dhe hipotezave që i kemi ngritur në fillim të këtij punimi, mund të nxjerrim konkluzionet duke u bazuar në rezultatet e arritura:

Pyetjet hulumtuese:

P-1: Si mundësohet marrja dhe ruajta e shënimeve duke përdorur platforma të hapura në kohë reale?

Konkluzioni – P1 : Në kapitullin 4 dhe 5 të këtij punimi u tregua se si është e mundur që shënimet nga Twitter t'i mbledhim në kohë reale dhe t'i përpunojmë e ruajmë ato duke i përdorur platformat e zgjedhura.

P-2: A përcakohet opinioni i klientit në Twitter duke përdorur gjuhën programuese Python?

Konkluzioni – P-2 : Lartë në këtë studim kemi analizuar 3 raste studimi përdorur gjuhën programuese Python dhe disa librari të saj. Duke i përdorur këto librari, përveç analizave

tjera, e pamë qënë mënyrë shumë të lehtë mund të bëjmë analizën dhe përcaktimin e polarizimit negativ dhe pozitiv të tekstit.

Hipotezat:

H-0: Si platforma të hapura Apache Kafka dhe PostgreSQL / Elasticsearch të cilat janë më të përshtatshme për marrjen dhe ruajtjen e shënimeve nga rrjeti social Twitter.

Konkluzioni H-0 :Gjatë këtij punimi, jam përpjekur që në mënyrë sa më të thjeshtë dhe të kuptueshme të krijoj një përmbledhje të procesit të mbledhjes, ruajtjes dhe përpunimit të të dhënave nga Twitter duke përdorur Apache Kafka dhe PostgreSQL / Elasticsearch. Këto platforma janë përzgjedhur për shkak të fleksibilitetit të tyre, mënyrës së lehtë të instalimit dhe konfigurimit. Këto platforma të përdorura janë mjaft të fuqishme dhe kanë përdorim shumë të madh në botën e teknologjisë informative.

Apache Kafka është një nga platformat më të përdorura lidhur me pranimin dhe dërgimin e mesazheve në mes të sistemeve të ndryshme, poashtu PostgreSQL është nga databazat Open-Source që përdoret më së shumti. Të gjitha këto veti kanë bërë të mundur që komplet procesi i shtjelluar në punim të bëhet shumë lehtë dhe kuptueshëm.

H-1: Në bazë të analizës së të dhënave nga Twitter me fjalën kyçe “Kosovo”, kandidati për president të SHBA-ve Donald Trump e ka përdorur marrëveshjen mes Kosovës dhe Serbisë për fushatë zgjedhore.

Konkluzioni H-1:Nga ky rast studimi, i analizuar në kapitullin 6.1 të këtij punimi, mund të konkludojmë se kandidati demokrat Donald Trump është favorizuar në postimet që kanë përmbajtur fjalën “Kosovo” në krahasim me kundër kandidatin Joe Biden.

Më poshtë po rishikojmë disa nga të gjeturat nga këto të dhëna:

- Nga **10,000** postime në Twitter kandidati **Trump** është përmendur në **391(73%)** postime, ndërsa kundërkandidati i tij **Joe Biden** është përmendur **141(27%)** herë.
- Vetëm rreth 14% e postimeve kanë qënë negative, ndërsa 86% e postimeve kanë qënë neutrale ose pozitive.

Shumë pozitive 920 9%

Pozitive	4602	46%
Neutrale	3120	31%
Negative	1156	12%
Shumë Negative	166	2%
Total	9964	

- Në 60% të postimeve është përmendur fjala “Agreement”.

H-2: Në bazë të të dhënave të marra Twitter, futbollisti Lionel Messi është lojtari më i popullarizuar dhe lojtari më i mirë i kohës, në krahasim me futbollistin Cristiano Ronaldo.

Konkluzioni H-2: Nga ky rast studimi, i analizuar në kapitullin 6.2, mund të konkludojmë se L.Messi është lojtar shumë më i popullarizuar në botë, në krahasim me futbollistin C.Ronaldo dhe po ashtu është edhe lojtari më i mirë në botë. Të gjeturat nga ky rast studimi mund të vërtetojnë këtë hipotezë:

- Për Lionel Messi janë postuar afër **30,000** postime apo **64%**, shifra të cilat janë dyfish më shumë se futbollisti Cristiano Ronaldo nga gjithsej **46,000** postime për **48** orë.
- Lionel Messi ka më shumë komente pozitive dhe neutrale, në krahasim me C.Ronaldo.
- Duke iu referuar fjalëve përcaktuese si (“greatest , goat , g.o.a.t , best) në përmbajtjen e postimeve (tekst) për këta dy lojtarë, ku këto fjalë janë përdorur për të përshkruar epitetin e lojtarit ma të mirë në botë, mund të konkludojmë se L.Messi është lojtari më i mirë në botë sepse prinë dukshëm në krahasim me C.Ronaldo. Nga **29470** postime për L.Messi , **5570** apo **19%** postime kanë përmbajtur fjalët përcaktuese. Nga **16374** postime për C Ronaldo , **1726** apo **11%** postime kanë përmbajtur fjalët përcaktuese.

H-3: Duke analizuar të dhënat në Twitter të arrijmë rezultat se cili nga këto brende gjigande si **Pepsi** dhe **CocaCola** ka reputacion më të madh dhe cila nga këto dy brende gëzon opinion pozitiv/negativ nga shfrytëzuesit.

Konkluzioni H-3 :Nga ky rast studimi, i analizuar në kapitullin 6.3 të këtij punimi, mund të vijmë në konkluzion se kompania Pepsi është më shumë e popullarizuar në Twitter se sa Coca Cola. Poashtu gjëja më e rëndësishme në këtë rast studimi është paraqitja e postimeve negative/pozitive të shfrytëzuesve. Kjo nënkupton se mund të analizojmë opinionin e shfrytëzuesve për kompanitë e cekura duke reaguar sidomos në raste kur komentet për produktet e tyre janë negative.

Të gjeteturat nga ky rast studimi:

- për 48 orë në Twitter janë bërë **10600** postime për kompanitë Pepsi dhe Coca Cola, ku **66%** kanë qenë për Pepsi dhe **34%** për Coca Cola. Nga këto rezultate mund të vërehet se Pepsi është shumë më i “përfolur” në rrjete sociale se sa konkurenti kryesor i tij.
- Coca Cola ka më pak postime negative se sa Pepsi.

Polarizimi i tekstit	Pepsi	Coca Cola	Pepsi	Coca Cola
Shumë pozitive	780	523	11%	7%
Pozitive	1369	817	19%	12%
Neutrale	3410	1683	48%	24%
Negative	1288	450	18%	6%
Shumë Negative	216	64	3%	1%
Total	7063	3537		

Duke i marë parasysh të gjitha këto rezultate, sado modeste që janë, mund të konkludojmë se të gjeturat nga shënimet nga rrjetet sociale mund të jenë shumë të dobishme dhe njëkohësisht shumë inkurajuese. Nëse e marrim parasysh faktin që këto shënime mund t’i marrim e përpunojmë nga platformat e hapura pa pagesë, kjo duhet të inkurajoj kompanitë, bizneset apo edhe individet që të marrin sa më shumë me këtë fushë, në mënyrë që analizat nga këto të dhëna t’i ndihmojnë ata për plasimin e produkteve, vendimarrjeve dhe mënyrës së bërit biznes përgjithësi.

8.RERERENCAT

- [1] - O'Reilly, Tim 2009. *“The Web 2.0.”* United States: O'Reilly Media.(12.12.2021)
- [2] - Narkhede, N., Shapira, G., Palino, T. 2017. *“Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale.”* United States: O'Reilly Media.(22.01.2022)
- [3] - Singh, Chanchal, Kumar, Manish. 2017. *“Building Data Streaming Applications with Apache Kafka.”* United Kingdom: Packt Publishing.(11.02.2022)
- [4] - Wise, Lyndsay. 2012. *“Using Open Source Platforms for Business Intelligence: Avoid Pitfalls and Maximize ROI.”* Netherlands: Elsevier Science.(26.01.2022)
- [5] - *“PostgreSQL: up and Running : A Practical Introduction to the Advanced Open Source Database”* - ISBN: 1449373194 (03.02.2022)
- [6] - Barry, Paul. 2016. *“Head First Python: A Brain-Friendly Guide. 2nd edition.”* Japan: O'Reilly Media.(14.01.2022)
- [7] - San Murugesan - *“Understanding WEB 2.0”* 30 October 2014.
DOI: 10.1109/MITP.2007.78(19.01.2022)
- [8] - Andreas Kaplan - *“Users of the World, Unite! The Challenges and Opportunities of Social Media”* February 2010. DOI: 10.1016/j.bushor.2009.09.003(25.01.2022)
- [9] - Kostas E. Psannis - *“Social Networking Data Analysis Tools & Challenges”*
October 2016. DOI:10.1016/j.future.2016.10.019(11.02.2022)
- [10] - Kostas E. Psannis - *“Twitter as a Corpus for Sentiment Analysis and Opinion Mining”* - January 2010. Conference: Proceedings of the International Conference on Language Resources and Evaluation (25.12.2021)
- [11] - Apoorv Agarwal, Boyi Xie, Ilia Vovsha - *“Sentiment Analysis of Twitter Data”*
Department of Computer Science Columbia University (04.02.2022)
- [12] -Bogdan Batrinca dhe Philip C. Treleven. *“: a survey of techniques, tools and platforms- Social media analytics”* (28.11.2021)

- [13] - Weiguo Fan June 2014. “*The Power of Social Media Analytics*” - Communications of the ACM 57(6):74-81 DOI: **10.1145/2602574**
- [14]- Axel Burns - “*User-Generated Content*”(2016),<http://snurb.info/files/2016/User-Generated%20Content.pdf> (13.01.2022)
- [15] Srijan Thapa,Geoff Skinner(2015),”*Review of Social Media Management Tools and Related Literature*”s(26.01.2022)
- [16] – Yuhuai Yu ,Hongfei Lin,Zhehuan Zhao(2016), ” *Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks*” (20.12.2021)
- [17]–Bogdan Batrinca,Philip C. Treleaven(2014),”*Social Media Analytics*” (22.09.2021)
- [18] – Xiaodan Zhu,(2013),<https://www.jair.org/index.php/jair/article/view/10896/25984> (20.11.2021)
- [20] – Minqing Hu,Bing Li,”*Mining and Summarizing Customer Reveiwes*”(25.12.2021)
- [21] - “*Kafka Docker for Development*”(2016), (23.01.2022)
- [22] – Raven Digimark,(2017),<https://www.ravendigimark.com/blog/what-is-web-2> (17.01.2022)
- [23] – Esteban Ortiz-Ospina,(2019), <https://ourworldindata.org/rise-of-social-media> (15.02.2022)
- [24] -Chandan Kumar,(2020), <https://medium.com/@chandank.kumar/ultimate-runbook-to-manage-apache-kafka-7e30eb6972ab> (20.01.2022)
- [25] – Lakshay Suri(2016), <https://lakshaysuri.wordpress.com/2016/10/02/apache-kafka/>(13.12.2021)
- [26] – Lovisa Johanson,(2019),<https://www.cloudkarafka.com/blog/2016-11-30-part1-kafka-for-beginners-what-is-apache-kafka.html> (13.02.2022)
- [27] – “*Elastic Cloud*”(2013),<https://www.elastic.co> (20.12.2021)

- [28] - Minqing Hu,Bing Li(2018),<https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf> (13.02.2022)
- [29] – Patrick Paroubek,Alexander Pak,(2010), "*Twitter as a Corpus for Sentiment Analysis and Opinion Mining*" (17.11.2022)
- [30] – Andry Marios,(2018),<https://github.com/lensesio/fast-data-dev> (19.02.2022)
- [31] – Alex Pak,(2017),<https://hub.docker.com/r/itzg/elasticsearch> (11.01.2022)
- [32] – Regis Belson(2019),https://hub.docker.com/_/postgres (17.02.2022)
- [33] –Steven Loria,(2020), <https://textblob.readthedocs.io/en/dev/> (25.01.2022)
- [34] – Andres Delfino,(2020),<https://docs.python.org/3/library/heapq.html> (14.01.2022)