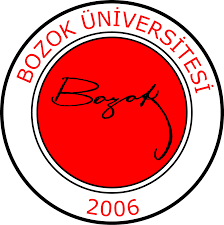
**T. C.**

**BOZOK ÜNİVERSİTESİ**

**MÜHENDİSLİK MİMARLIK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

****

**TWITTER ÜZERİNDEN ELDE EDİLEN VERİLER KULLANILARAK, TWEET DUYGU ANALİZİNİN YAPILMASI**

**Engin Karataş**

**16008118040**

**BİLGİSAYAR MÜHENDİSLİĞİ BİTİRME TEZİ**

**YOZGAT 2022**

**BOZOK ÜNİVERSİTESİ**

**MÜHENDİSLİK MİMARLIK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**TWITTER ÜZERİNDEN ELDE EDİLEN VERİLER KULLANILARAK, TWEET DUYGU ANALİZİNİN YAPILMASI**

**Engin Karataş**

Danışman: Doç.Dr. Mehmet BAKIR

Bölüm Başkanı: Doç.Dr. Mehmet BAKIR

**YOZGAT 2022**

**TWITTER ÜZERİNDEN ELDE EDİLEN VERİLER KULLANILARAK, TWEET DUYGU ANALİZİNİN YAPILMASI**

## ÖZET

Bir konu hakkında Twitter üzerinde insanların genel kanısı bilimsel yaklaşımlarla araştırmak istenildiğinde, ilgili konunun hastag(#) ilgili Tweet`leri okumak ve araştırma yapan araştırmacının bireysel fikri ile değerlendirmek gerekmektedir. Her bir tweet paylaşımını teker teker okuyup değerlendirmek bilimsel bir analiz niteliğine sahip olmayacaktır. Kullanıcıların duygularını nicel bir bakış ile değerlendirilmesini zorlaştıracaktır. Elastic Search kullanarak Twitter tweet analizi projesi, kullanıcıların yaptığı yorumlar bilgisayar tarafından analiz edilip konu bazında insanların genel kanılarının objektif bir şekilde değerlendirilip snıflandırma işlemlerini standartlar çerçevesinde değerlendirilmesini amaçlar. Sınıflandırılan bu veriler üzerinde filtrelemeler yaparak **eş zamanlı izlenebilen grafikler** oluşturulmuştur. Bu sayede bu grafikler üzerinde duygu verilerinin yorumlanması sağlanmıştır.

Bilgisayar bilimleri, insan diline en yakın yeni nesil 5. Nesil programlama dillerinin gelişimi ile gelen yeni teknolojiler ile arama, veri kaydetme, görselleştirme işlemleri, dağıtık sistemler üzerinde etkili şekilde çalıştırılabilir duruma gelmiştir. Yeni teknolojiler üzerinde çalışan projede, insan faktörünü en aza indirilmiş, yapılacak çalışmanın analiz edilmesini ve nicel sonuçlar elde edilip değerlendirilmesi kolaylaştırmıştır.

Tez özelinde Twitter üzerindeki hastag(topic)`lerden iki tanesi “Ukrayna savaşı ve Dolar Kuru” incelnmiştir ve sonuç çıkarılmıştır.

**Anahtar Kelimeler:** Twitter, Duygu Analizi, Elastic Search, Kibana, AWS, Python, Harita Görselleştirme

**PERFORMING A TWEET EMOTION ANALYSIS USING THE DATA OBTAINED THROUGH TWITTER**

## ABSTRACT

When it is desired to research the general opinion of people on Twitter with scientific approaches, it is necessary to read the tweets related to the hash(#) of the relevant topic and evaluate it with the individual opinion of the researcher. Reading and evaluating each tweet share one by one will not have the quality of a scientific analysis. It will make it difficult for users to evaluate their emotions from a quantitative perspective. The Twitter tweet analysis project using Elastic Search aims to analyze the comments made by the users by the computer, to evaluate the general opinions of the people on the basis of the subject in an objective way and to evaluate the classification processes within the framework of standards. Graphs that can be watched simultaneously were created by filtering on these classified data. In this way, the interpretation of emotion data on these graphs is provided.

With the development of new generation 5th generation programming languages ​​closest to human language, computer science has become able to be operated effectively on search, data recording, visualization processes and distributed systems with the development of new technologies. In the project working on new technologies, the human factor was minimized, making it easier to analyze the work to be done and to obtain and evaluate quantitative results.

**Keywords:** Twitter, Sentiment Analysis, Elastic Search, Kibana, AWS, Python, World Map Visualization

## TEŞEKKÜR

Yapmış olduğum proje esnasında, proje ilerleyişini takip eden, teknik konularda desteğini esirgemeyen değerli hocam Doç.Dr. Mehmet BAKIR`a teşekkürlerimi sunarım.

Python ile projenin temeli olan duygu analizi konusunda veri madenciliği alanında yardımlarını esirgemeyen değerli hocam Öğr. Gör. Ömer DAĞİSTANLI`ya teşekkür ederim.

Engin Karataş

Yozgat 2022

### İÇİNDEKİLER

[ÖZET i](#_Toc106314818)

[ABSTRACT ii](#_Toc106314819)

[TEŞEKKÜR iii](#_Toc106314820)

[İÇİNDEKİLER iv](#_Toc106314821)

[TABLOLAR DİZİNİ vi](#_Toc106314822)

[ŞEKİLLER DİZİNİ vii](#_Toc106314823)

[SİMGELER VE KISALTMALAR viii](#_Toc106314824)

[1. GİRİŞ 1](#_Toc106314825)

[1.1. Giriş **Error! Bookmark not defined.**](#_Toc106314826)

[1.2. Python Programlama 3](#_Toc106314827)

[1.2.1. Python Kütüphaneleri 3](#_Toc106314828)

[1.3. AWS (Amazon Web Service) Nedir? 6](#_Toc106314829)

[1.3.1. AWS`nin Sunduğu Önemli Hizmetler: 6](#_Toc106314830)

[1.4. ELK Stack Nedir? 7](#_Toc106314831)

[1.4.1. Elastic Search Nedir? 8](#_Toc106314832)

[1.4.1.1. Elastic Search Yararları 9](#_Toc106314833)

[1.4.2. Kibana Nedir? 11](#_Toc106314834)

[1.4.3. Logstash Nedir? 12](#_Toc106314835)

[1.4.4. Beats Nedir? 12](#_Toc106314836)

[1.5. ELK Stack ile Neler Yapılabilir? 13](#_Toc106314837)

[1.6. Elastic.co Elastic Search Nedir? 13](#_Toc106314838)

[1.6.1. Elastic.co Üzerinden Elastic Search Kullanımının Avantajları 14](#_Toc106314839)

[1.6.2. Elastic Search Araçları 14](#_Toc106314840)

[Elastic Search veritabanına girilen veriler üzerinde yönetim, görselleştirme sağlanması için bünyesinde birçok araç bulundurur. Bu araçlardan proje özelinde kullanılanlar aşağıdaki gibidir. 15](#_Toc106314841)

[1.7. Twitter Developer Account(Twitter Geliştirici Hesabı) Nedir? 15](#_Toc106314842)

[1.7.1. Twitter Geliştirici Hesap Türleri 16](#_Toc106314843)

[1.7.1.1. Bilimsel Makale İçin Geliştirici Hesabı 16](#_Toc106314844)

[1.7.2. Hobi Amaçlı Geliştirici Hesabı 16](#_Toc106314845)

[1.8. SSH Protokolü Nedir? 17](#_Toc106314846)

[2. YAPILAN ÇALIŞMALAR 17](#_Toc106314847)

[2.1. Twitter Developer Hesabı API Başvuru Süreçleri 17](#_Toc106314848)

[2.1.1. Projede Twitter Geliştirici Hesabı 17](#_Toc106314849)

[2.2. Bulut Sunucu Hizmetine SSH Protokolü ile Erişim 18](#_Toc106314850)

[2.3. AWS EC2 Sanal Makine Kullanımı 18](#_Toc106314851)

[2.3.1. AWS Sanal Makine Seçimi 18](#_Toc106314852)

[2.4. Python ile Sentiment(Duygu) Analizi 20](#_Toc106314853)

[2.4.1. Python ile duygu analizi 20](#_Toc106314854)

[2.5. Elastic Search Veri Tabanı İşlemleri 21](#_Toc106314855)

[2.5.1. CRUD Nedir? 21](#_Toc106314856)

[2.5.2. Elastic Search Python ile Ekleme İşlemi 22](#_Toc106314857)

[2.5.3. Elastic Search ile Okuma İşlemi 22](#_Toc106314858)

[2.5.4. Elastic Search ile Silme İşlemi 23](#_Toc106314859)

[2.5.5. Elastic Search ile Güncelleme İşlemi 23](#_Toc106314860)

[2.6. Kibana Üzerinden Monitoring(İzleme), Analytics(Mantıksal Çözümleme) ve Görselleştirme 24](#_Toc106314861)

[2.7. Dünya haritasında lokasyona göre duygu analizi 27](#_Toc106314862)

[2.7.1. Dünya haritası yazılımının geliştirilmesi 27](#_Toc106314863)

[3. SONUÇ 30](#_Toc106314864)

[KAYNAKÇA 31](#_Toc106314865)

[ÖZGEÇMİŞ 32](#_Toc106314866)

### TABLOLAR DİZİNİ

**SAYFA**

[Tablo 1 vi](#_Toc93435944)

Tablo 1

### **ŞEKİLLER DİZİNİ**

**SAYFA**

[**Şekil 1**. Projenin amacını gösteren diyagram 1](#_Toc93839668)

[**Şekil 2.** ELK Stack Nedir 8](#_Toc93839669)

[**Şekil 3.** Elastic Search full-text search çalışma prensibi 11](#_Toc93839670)

[**Şekil 4.** ELK Stack çalışma adımları 13](#_Toc93839671)

[**Şekil 5.** Kibana gösterge paneli, yeni görselleştirme ekle(create visualization) 25](#_Toc93839672)

[**Şekil 6.** Bir e-ticaret sisteminin Kibana ile eş zamanlı izlenmesi 26](#_Toc93839673)

### SİMGELER VE KISALTMALAR

# hastag

SQL Structered Query Language

ELK Elastic, Logstash, Kibana

GUI Graphical User Interface

JSON

API

CRUD

ID

HTTP

CURL

1. GİRİŞ

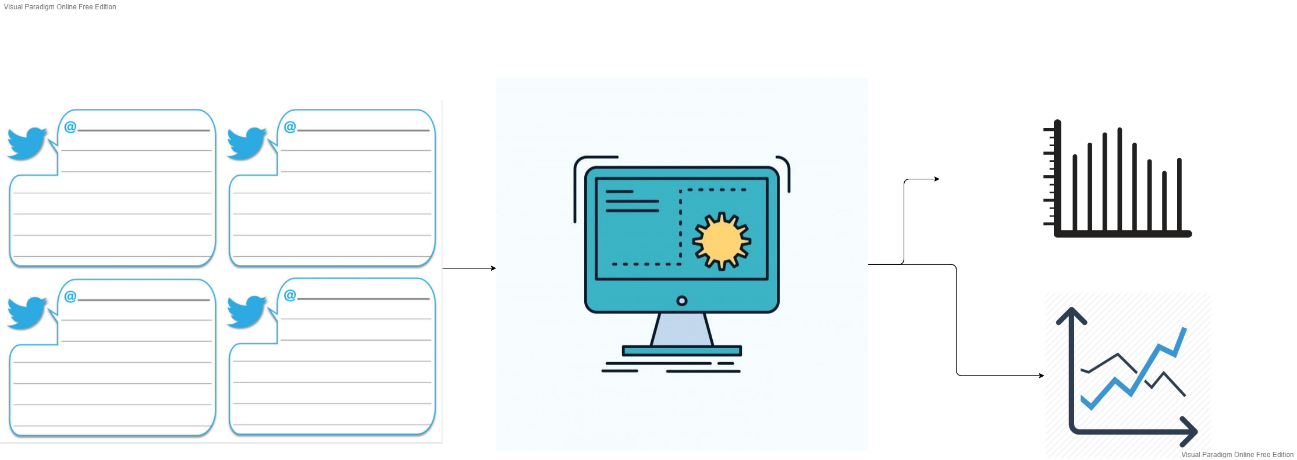
Twitter, günümüzde kullanılan, web 2.0 teknolojisi üzerinde çalışan bir sosyal medya, içerik oluşturma ve yayınlama sistemidir. Twitter`ın dünya genelinde kullanımı oldukça yaygın olduğundan, gerçek veriler için oldukça zengin bir kaynak sunmaktadır. Bu nedenle insan merkezli araştırmalarda, duygu analizlerinde araştırmacılar tarafından başvurulur. Kullanıcılar, konu bazında diğer insanların fikirlerini okuyabilir ve kendi fikirlerini de aynı konu üzerinden yayınlayabilir.

Twitter`da geçtiğimiz 2021 yılı içerisinde dünya çapında 322 milyon kullanıcı hesabı olduğu tahmin edilmektedir (1). Twitter etkileşimli bir platformdur. Kullanıcı, diğer kullanıcıların Tweet`lerini okuyup beğenebilir ve yorum yapabilir. Kullanıcılar, kendi profilleri üzerinden kendi düşüncelerini yazı formatında paylaşabildiği gibi, fotoğraf ve video eklemeleri de yapabilirler.

Konu bazında içerik paylaşılmasına hastag denir. Konu bazında içerik paylaşımları sekmesinden hastag ile diğer kullanıcıların yorumlarını incelenebilir.

Araştırmacılar, Veri Bilimciler, Büyük Veri mühendisleri veya herhangi sebepten dolayı kullanıcı veya kullanıcılar üzerinde çeşitli analizler yapmak isteyen kurum/kuruluşlar çeşitli sınıflandırmalar ve analizler yaparlar.

Analiz ve sınıflandırmalar, WEB tarayıcı üzerinde gezinme yoluyla okunup yorumlanabilir. Fakat araştırmanın sonucunda bilgisayarlı sistemlere göre, çok daha az başarı ve veri elde edilecektir.



**Şekil 1**. Projenin amacını gösteren diyagram

Şekil 1 `de otomatizme edilmiş bir sistem üzerinden elde edilen bilgilerin grafikler üzerinde gösterildiği genel olarak gösterilmiştir.

Tez için yapılan çalışmada Twitter`dan elde edilen bilgileri, bilgisayar bilimlerinin bir konusu olan veri madenciliği işlemlerinden geçirilmiştir. Veriler işledikten sonra kayıtların yazılabildiği veri tabanına veri olabilecek en anlamlı şekile getirilip kaydedilmeye çalışılmış ve sınıflandırılan bu anlamlı sonuçlar, en kolay şekilde incelenebilinmesi ve analiz edilebilmesi için görselleştirilmiştir.

**Tezin Amacı**

Duygu analizi seçim zamanlarında politika belirlemekte, yatırım dönemlerinde yükseliş ve azalış durumlarını tahmin etmede, bir ürünün üzerindeki memnunluğu ölçümlemek için kullanılabilir. Örneğin bir televizyon programının izleyiciler üzerindeki etkisini, beğenisi ölçülebilmesi için, izleyicilerin duygularının analiz edilmesi gerekir. Duyguların değerlendirilip sınıflandırılmasından sonra, verileri anlamlandırmak için grafiksel ara yüz oluşturulması projenin temel amacı olmasıyla birlikte, proje kazanımları aşağıdaki gibidir

* Duygu analiz yapmak
* Çok sayıda insana ait veriyi Twitter aracığıyla elde etmek
* Verileri veri madenciliği ile anlamlandırmak ve sonuçları sınıflandırmak
* Elde edilen sonuçları anlamlandıracak grafikler oluşturmak
* Anlamlı sonuçların bir kısmını Twitter üzerinde otomatik bir şekilde paylaşmak

Twitter`ın Kullanımı oldukça yaygın olduğundan, içerik bakımından oldukça zengin bir kaynak sunmaktadır. Kullanıcılar, konu bazında diğer kullanıcıların fikirlerini, konu için belirlenmiş hastag `ler ile oluşturulmuş konu sayfalarından okuyabilir ve kendi fikirlerini de aynı konu üzerinden yayınlayabilir. Kullanıcılar, kendi profilleri üzerinden kendi düşüncelerini yazı formatında paylaşabildiği gibi, fotoğraf ve video eklemeleri de yapabilirler. Twitter etkileşimli bir platform olduğundan, kullanıcı, diğer kullanıcıların Tweet`lerini okuyup beğenebilir ve yorum yapabilir.

Projede twitter üzerinden alınan veriler, Python programlama dili içerisine getirildikten sonra sınıflandırılmış ve Elastic Search veri tabanına yazılmıştır. Yazılan bu veriler, Elastic Search ile uyumlu çalışan görselleştirme, analiz aracı Kibana ile çalışılarak görselleştirilmiştir. Projede python programlama dili, duygu analizi yapılması ve sınıflandırılması için kullanılmıştır.

#### Python Programlama

Python programlama dili 5. Seviye yeni nesil bir programlama dilidir. Syntax(sözdizim) kuralları basit olduğundan ve hala geliştirilmeye devam edilen bir dil olduğundan dolayı Python oldukça popülerdir. Python programlama dilinin kullanım alanları aşağıdaki gibidir.

**Python programlama dilinin kullanım alanları:**

* Metin madenciliği
* Makine öğrenmesi
* Veri analizi
* Oyun geliştirme
* Web geliştirme
* Yapay zekâ
* Robotik uygulamaları
  + 1. Python Kütüphaneleri

Kütüphane kullanımı sayesinde geliştirici her ayrıntıyı, teorik ve matematiksel kısımları bilgisayara anlatmak için veri yapıları seviyesinde düşünerek çözüm getirmeye gerek kalmadan geliştirme yapabilir. Python programlama dili birçok kütüphaneye sahiptir.

Python programlama dili içerisinde çeşitli amaçlara yönelik birçok kod geliştirilmiş ve diğer geliştiricilerin hizmetine bir çok kütüphane sunulmuştur. Bu kütüphaneler amaca yönelik algoritmaları ve geliştirme aşamasında ekstra yük olacak kod bloklarını ve mantık süreçlerini azaltmaya yönelik kullanılan yardımcılardır. Python diline ait proje özelinde geliştirilmiş kütüphaneler aşağıda tanımlanmış ve projedeki kullanım alanları açıklanmıştır.

* + - 1. Tweepy

Tweepy adındaki Python kütüphanesi, Twitter API`yi kullanabilmek için geliştirilmiş ve hala geliştirilmeye devam eden açık kaynak bir kütüphanedir. Tweepy, tüketici ve erişim anahtarlarını kullanarak, Twitter API için tanımlanmış hedefe istekte bulunmak için Twitter API için standartlaştırılmış isteklerini gönderebilmek için geliştirilmiştir. API bilgileri kendi kütüphanesinde düzenlediği gibi, geliştiriciye sadece ulaşılmak istenilen hizmetler için genellikle tek satırda yazılabilen fonksiyonlar ile API isteği yürütülmesini sağlamıştır. Twitter API ile bağlantı sağlanırken consumer key, consumer secret, access token, acces token secret adında Twitter API oturumundan, geliştirici hesabı özelinde verilen anahtarlara ihtiyaç duyulur. Python komutları Tweepy kütüphanesi tarafından çalıştırılırken bu 4 anahtar kullanılır.

* + - 1. TextBlob

TextBlob, metin verilerini işlemek için kullanılan bir kütüphanedir. Konuşma bölümü etiketleme, duygu analizi, sınıflandırma, çeviri ve daha fazlası gibi yaygın doğal dil işleme işlemlerini yapmak için basit bir kullanım sağlayan Python kütüphanesidir. Türkçe metinleri TextBlob kütüphanesi aracılılığıyla herhangi bir yabancı dile çevirip, TextBlob kütüphanesiyle duygu analizini yapmak mümkün olmaktadır. TextBlob kütüphanesiyle sınıflandırılan duygular, word map(kelime bulutu), heat map(ısı haritası) gibi çeşitli yöntemler ile görselleştirilerek yorumlanması ve analiz edilmesi kolaylaştırılabilir.

* + - 1. **Elastic Search Python Kütüphanesi**

Elastic Search veri tabanı üzerindeki her türlü veri tabanı işlemini gerçekleştirmek için kullanılır. Elastic Search komutlarının geliştiriciye en kolay şekilde kullanması için geliştirilmiş bir kütüphanedir. Bu kolaylığının yanında hantal kalmaktadır. Elastic Search veri tabanı işlemleri bu kütüphane üzerinden yürütülmektedir. Proje özelinde Kısım 1.3`te bahsedilen ELK Stack Elastic.co bulut hizmeti üzerindeki veri tabanına Cloud ID adındaki özel bir anahtar ve deployment(dağıtım)`ın kullanıcı adı ve şifresi girilerek bağlantı sağlanıp yönetilmiştir. Yazılan kodlar, ELK bulut hizmeti üzerindeki veri tabanını doğrudan etkilemekte ve değiştirmektedir.

* + - 1. **Geopy**

Geopy, Python geliştiricilerinin coğrafi lokasyonların ve diğer veri kaynaklarını kullanarak dünya çapındaki adreslerin, şehirlerin, ülkelerin ve yer işaretlerinin koordinatlarını kolayca bulmasını sağlar.

**Geopy Nominatim**

Nominatim, adınızı ve adresinizi verdikten sonra size veri sağlayan ve özel anahtar girilmesine gerek kalmayan API olarak da bilinen ücretsiz bir araçtır. Proje özelinde lokasyonların enlem ve boylamlarını almak için bu Class kullanılır.

* + - 1. **Pandas**

Pandas, veri işlemesi ve analizi için Python programlama dilinde yazılmış olan bir yazılım kütüphanesidir. Bu kütüphane temel olarak zaman etiketli serileri ve sayısal tabloları işlemek için bir veri yapısı oluşturur ve bu şekilde çeşitli işlemler bu veri yapısı üzerinde gerçekleştirilebilir olur. (2) Pandas içerisindeki tablo şablonlarından biri olan Data Frame Yapıları veri ön işleme adımlarında kolaylık sağlanması için kullanılmıştır.

* + - 1. **Re Kütüphanesi**

Re kütüphanesi, regular expression (düzenli ifadeler) işlemleri için oluşturulmuş bir kütüphanedir.

**Regular Expression nedir?**

* Yoğun veri yığınlarından gerekli bilgiler ayıklar.
* Genellikle form bilgisi gönderirken kullanıcı tarafından girilen girişi kontrol eder.
* Verileri kullanım amacına uygun bir formata çevirir. Örneğin email gönderirken hatalı bir format kabul edilemez.

#### AWS (Amazon Web Service) Nedir?

Amazon Web Service, bulut tabanlı, Web 2,0 üzerinde çalışan, istenilen miktardaki bulut işlem donanımları ve API`ları kiralanabilen bir platformdur. Amazon firmasının yan kuruluşu olan AWS, ödediğin kadar kullan mantığıyla web sunucuları, web siteleri için web servisleri sağlamaktadır. EC2 servisi, “cluster of computers” ile çalışır. Çalıştırılacak sistemler fiziksel olarak bir makineye ait değildir. Aynı küme içerisinde tanımlı donanımlar yazılım ile kontrol edilir. Sanal bir makine ile gerçek bir bilgisayarı ayırt edecek en büyük fark donanımsal özelliklerdir. EC2 servisi üzerinden satın alınan bir makinede sanallaştırma teknolojileri kullanılır. Tercih edilen işletim sistemi özellikleri ve donanım kapasiteleri, EC2 gösterge paneli üzerinden seçilir.

AWS hizmetlerine erişebilmek için öncelikle aws.amazon.com üzerinde kaydolmak gerekmektedir. Kaydolurken kredi kartı bilgileri girilmeli ve 1 USD($) ödeme yapılması gerekmektedir. Kaydolma işlemi tamamlandıktan sonra kullandığın kadar öde ilkesiyle AWS`nin sunduğu servisler kullanılabilir durumda olacaktır. AWS`nin sunduğu hizmetlerden bazıları aşağıdaki gibidir.

* + 1. AWS`nin Sunduğu Önemli Hizmetler:
* **Amazon EC2 (Elastic Compute Cloud):** EC2, bulutta bulunan işletim sistemlerinin kontrol edilebildiği ve oluşturulan bu işletim sistemine istenildiği zaman erişile bilinen bir sanal makine hizmetidir.
* **AWS Amazon S3 (Simple Storage Service):** Bu serviste; dosya, klasör, resim, belge gibi nesneler saklanabilir. Bu dosyalar, host üzerinde bulundurulmak yönetilebilirliği ve performansı düşürür. Özellikle resimler, web sitesi ilk açıldığında performansın büyük bir oranını etkilemektedir. Bu nedenle birçok büyük e-ticaret firması S3 gibi sistemler kullanır. Resimler, site yüklenirken S3 servisinden maliyeti düşük şekilde getirilir.
* **EKS (Kubernetes için Elastik Konteyner Hizmeti):** Bu araç ile Kubernetes kurulum yapılmadan Amazon bulut ortamında kullanılabilir.
* **Elasticache:** Elasticache, bulut bilişim ile verileri önbelleğe alma yeteneğini getirir. Çok kullanılan veriler önbelleğe alınır. Bu sayede ElasticSearch üzerindeki veri yük azaltılır
* **DynamoDB:** Yüksek performanslı bir NoSQL veri tabanıdır.
* **Storage Gateway:** Sunucu üzerinde kurulan bir sanal makinedir. Firma üzerindeki veriler AWS üzerinde yedeklenir ve verilerin korunumu artar
* **Glacier:** Bu arşivleme hizmeti, dosyaları uzun süreler boyunca ucuz bir şekilde saklayabilir.
* **IAM (Kimlik ve Erişim Yönetimi):** Kullanıcıların yönetilmesi ve birden çok gruplar oluşturulmasına olanak tanır.
* **Inspector:** Bir sanal makineye yüklenen ve tüm güvenlik açıklarını raporlayan bir güvenlik aracıdır.
* **Shield(Kalkan):** Web uygulamaları üzerinde giriş yaparken, DDOS saldırılarını engellemek için kullanılır.

#### SSH Protokolü Nedir?

Güvenli Kabuk (SSH), ağ hizmetlerinin güvenli olmayan bir ağ üzerinde güvenli şekilde çalıştırılması için kullanılan bir kriptolanmış ağ protokolüdür (3). Uzun yıllar boyunca kullanılan çeşitli bağlantı kurma protokolleri örneğin rsh, rexec, rlogin vardır. Fakat bu bağlantı yöntemleri özellikle şifreleri açık bir şekilde gönderdiğinden oldukça güvensizdir. SSH protokolü, güvenlik açığı bulunan bu protokollere göre daha güvenlidir. İki ağ arasında SSH protokolü ile iki yerel ağ arasında bağlantı sağlanabilir.

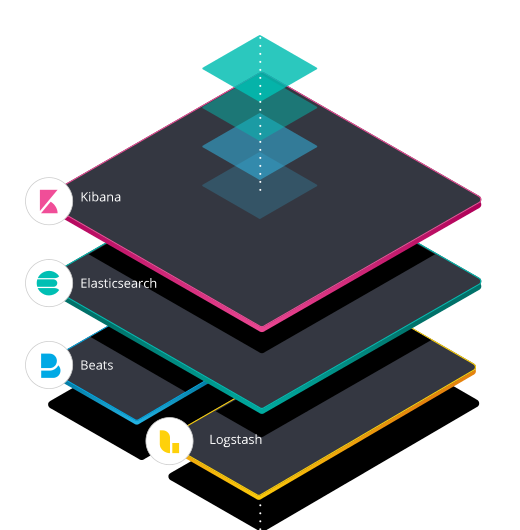
Uzaktaki bir bilgisayara bağlanmak için çeşitli şekillerde SSH protokolü çalıştırılabilir. SSH kullanım alanı genellikle Linux tabanlı işletim sistemlerine bağlanma şeklindedir. Bu bağlantı direkt işletim sistemine gönderilen bir komut satırı ile bağlantı sağlanması yöntemi şeklindedir.

SSH protokolü her ne kadar güvenli bir protokol olsa da Edward Snowden tarafından, Amerikan Ulusal Güvenlik Ajansı tarafından SSH oturum bilgilerinin çözülebildiğini, dinlenebildiğini ortaya çıkarmıştır (4).

AWS sanal makine üzerine bağlanırken SSH protokolü kullanılmıştır.

#### ELK Stack Nedir?

Şekil 2`de betimlenen Elastic Search, Kibana ve Logstash kelimelerinin baş harflerinden oluşan ELK Stack, beraber birçok işlemi bir araya getiren üç aracın kısaltımı için kullanılan bir kelimedir. Bu başlık altında ELK Stack içerisindeki ElasticSearch, Kibana ve Logstash teknolojileri anlatılacaktır.



**Şekil 2.** ELK Stack Nedir

* + 1. Elastic Search Nedir?

Elastic Search, Java dili ile yazılmış, Apache Lucene teknolojisi üzerinde çalışan, dağıtık sistemlerde çalışabilen, ilişkisel olmayan(unstructered) bir veri tabanı, analiz aracı ve Full-Text arama motorudur. Apache Lucene bir bilgisayar üzerinde arama yapabilir. Fakat dağıtık sistemlerde anlık gerçekleşecek aramalar söz konusu olduğunda Apache Lucene yetersiz kalmaktadır. Dağıtık sistemler üzerinde çalışabilmesi, Büyük Veri`ye sahip sistemlerde yüksek performans ile çalışması, esnek yapıya sahip olması gibi özellikleri Elastic Search`i güçlü kılan özelliklerindendir.

Veriler, Elastic Search üzerine veriler henüz kaydedilme aşamasında index`ler üzerinde ayrılır. Bu sayede aranan ifadeler kolaylıkla bulunabilir.

Elastic Search, açık kaynak bir sistem olmasından dolayı, Linux, MacOS, Windows üzerinde veya bu işletim sistemleri üzerinde oluşturulmuş konteynerler(sistemlerin dağıtılması ve hızlıca yönetilebilmesi için kullanılan bir sistem) üzerinde indirilip çalıştırılabilir.

Elastic Search sürüm kullanımı geriye uyumlu(backwards compatible)`dur. Bu nedenle geçmiş sürümlerde çalışan kodlar sonraki sürümlerde çalışmazken ileriki sürümlerde yazılan kodlar geçmiş sürümler ile çalışabilir.

* + - 1. Elastic Search Yararları
* **Hızlıca arama yapılabilir:** Elastic Search`ın dağıtılmış altyapısı, büyük miktarda veriyi paralel olarak işler ve sorgular için en iyi eşleşmeleri hızla bulunmasını sağlar.
* **Elastic Search üzerine birçok yardımcı araç kurulabilmektedir:** Elastic Search görselleştirme için kibana ile uyumlu bir şekilde çalışır. Geliştirilen uygulama özelinde ihtiyaç duyulan birçok araç kurulabilir.
* **Veri tabanı olarak kullanılabilir:** Elastic Search verilerin kaydedilip erişilebildiği bir veri tabanı olarak kullanılabilir ve Metin verilerinde oldukça hızlı sonuçlar alınabilir.
* **Tamamen açık kaynaklıdır:** Proje üzerindeki popülerliğin artmasıyla birlikte topluluğun projeyi daha çok desteklemesi sağlanmıştır. Açık kaynaklı yazılımlar şeffaf ve güvenilirdir. Projenin özel olma, korunma maliyetlerinden de kurtulunmuş olunur.
* NoSQL veri tabanı gibi çalışır ve örneğin **MongoDB** gibi diğer NoSQL veri tabanlarından bilgileri kendi üzerine aktarabilir.
* **RestFul mimarisini üzerinde JSON objesi ile çalışır:** Elastic Search, basit bir REST tabanlı API, basit bir HTTP arabirimi ve şemasız JSON belgeleri sağlayarak hizmeti hızlı bir şekilde kullanmaya başlamanıza ve farklı kullanım durumları için uygulamalar tasarlamanıza olanak tanır.
* **Gerçek Zamanlı İşlemler Yapılabilir:** Gerçek zamanlıya yakın operasyonlar ile Veri okuma veya yazma gibi Elastic Search işlemleri genellikle bir saniyeden daha kısa sürede tamamlanır. Bu sayede uygulama üzerinde izleme ve anormallik tespiti gibi gerçek zamanlıya yakın kullanım amaçları için Elastic Search `ten faydalınabilir.
* **Kolay Uygulama Geliştirme Rahatlığı:** Elastic Search, Java, Python, PHP, JavaScript, Node.js, Ruby ve daha fazla dil için destek sağlar.
  + - 1. **Elastic Search Veri tabanı Terimleri ve İlişkisel Veri Tabanları Karşılaştırılması**
* **Type(Tür):** İlişkisel bir veri tabanındaki bir tablodur ve verileri mantıksal bölümlere ayırır. ElasticSearch, aynı dizinde birden çok Tür içerebilir.
* **Document(Belge):** Elastic Search ‘teki Tür(Type) yapısındaki bir satırı temsil eder. İlişkisel veri tabanlarında ise bir satırı temsil eder.
* **Mapping(Haritalama):** İndekslenen verinin veri tipinin tanımlanması işlemidir.
* **Fields(Alanlar):** Diğer veri tabanı türlerindeki sütunlara Elastic Search'te alanlar denir. Her belge birden fazla alan içerebilir.
* **Cluster(Küme):** Bir veya daha fazla düğümün(node) toplamıdır. Cluster(Küme) nedeniyle, tüm verilerle dizin oluşturma ve arama işlevleri oluşturulur.
* **Indice:** Herhangi arama motorunda bir veri tabanı işlemi aranırken, Elastic Search, aramaları Indice adlı bir dizin dosyasında gerçekleştirilir. Bir Elastic Search kümesi(cluster), birden çok index içerebilir.
  + - 1. **Elastic Search Full-Text Search Özelliği**

Elastic Search veri tabanı üzerinde yapılan aramalar çok hızlı bir şekilde gerçekleşir. Hızlı aramaları sağlayan Elastic Search özelliği, Full-Text Search özelliğidir. Yeni bir kayıt eklendiğinde metinleri index`lere ayırarak kaydeder. Yeni bir sorgu yazıldığında, bünyesine kaydettiği index`ler üzerinden arama yapar.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Şekil 3.** Elastic Search Full-Text Search çalışma prensibi

* + 1. Kibana Nedir?

Kibana, Elastic Search ile uyumlu bir görselleştirme platformudur. Elastic Search cluster(küme) içerisinde depolanan verileri aramak, görüntülemek ve analiz etmek için web tabanlı ara yüz sağlar. (4) Eğer Elastic Search üzerindeki veriler zaman belirtiyorsa, bir alan içerisinde zamanı belirten timestamp(zamanı belirten 13 haneli bir tamsayı) ile belirtilmelidir. Kibana üzerindeki görselleştirme(visualize) bölümü, görselleştirme eklentilerinden biri ile verileri görselleştirme olanağı sağlar. Bilgiler tablolar, çizelgeler, haritalar, histogramlar şeklinde gösterilebilir.

* + - 1. Kibana Avantajları
* **Etkileşimli Grafikler :** Kibana, Elastic Search veri tabanı ile eş zamanlı, etkileşimli olarak grafikler oluşturulmasını sağlayan raporlar sağlar. Zaman üzerinde dinamik arama yapılabilir. Belirli veri aralıklarına yakınlaştırıp uzaklaştırabilir.
* **Çizim desteği:** Kibana, coğrafi bilgileri sorunsuz bir şekilde veriler üzerine yerleştirile bilinen ve sonuçları bir harita üzerinde görselleştirebilen için güçlü bir coğrafi çizim yeteneğine sahiptir.
* **Önceden Oluşturulmuş Protokoller ve Filtreler:** Kibana'nın filtrelerini kullanarak, histogramlar, sorgular veya e-ticaret sisteminde en çok satılanlar gibi çeşitli analizler birkaç tıklamayla gerçekleştirebilir.
* **Kolay erişim paneli:** Panoları ve raporları kolayca oluşturabilir ve paylaşılabilir. Verileri görüntülemek için bir Chrome, Firefox gibi bir WEB tarayıcı yeterlidir.
  + 1. Logstash Nedir?

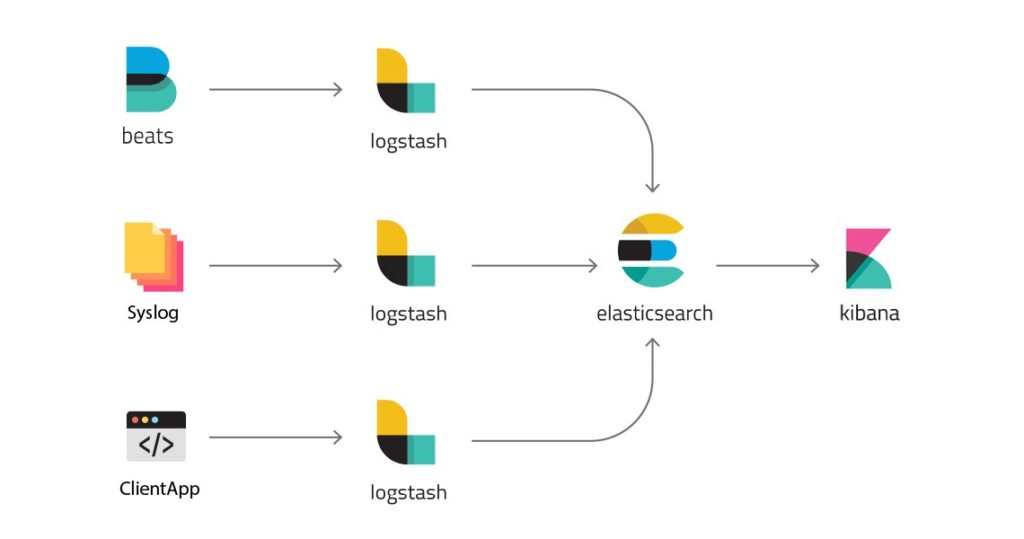
Logstash, çeşitli kaynaklardan veri toplanmasına, anında kayıt alınmasına ve istenilen hedefe göndermeye olanak tanıyan hafif, açık kaynaklı bir sunucu tarafı bir pipeline(veri işleme hattı)`dır. Elastic Search için bir veri hattı olarak kullanılır. Logstash, toplanan günlük verileri toplar. Topladığı verileri JSON formatına çevirir ve Elastic Search Cluster yapısında depolar.

* **Yapılandırılmamış Verileri Kolayca Yükler:** Sistem günlükleri(log), WEB sitesi günlükleri ve uygulama sunucusu günlükleri dahil olmak üzere çeşitli veri kaynaklarından yapılandırılmamış verileri kolayca ortaya çıkarılmasına olanak tanır.
* **Önceden Oluşturulmuş Filtreler Sağlar:** Logstash önceden oluşturulmuş filtreler sağlar, Elastic Search'te dizini üzerinde özel sorgular, modeller oluşturmak zorunda kalmadan sorgulama imkânı sunar.
* **Birçok Hazır Eklentiye Sahip Olma Özelliği:** Github'da halihazırda mevcut olan 200'den fazla eklentiyle, veri hattını(modeller) özelleştirmek için gereken eklentiler diğer geliştiriciler tarafından oluşturulmuştur. İhtiyaçlarınıza uygun bir eklenti yoksa, proje özelinde oluşturulabilir.
  + 1. Beats Nedir?

ELK Stack içinde tanımlı bir diğer teknoloji Beats`dır. Kelime anlamı atım olan Beats, tek amaçlı veri gönderimleri için ücretsiz ve açık bir platformdur. Yüzlerce veya binlerce makine ve sistemden Logstash veya Elasticsearch'e veri gönderir.

#### ELK Stack ile Neler Yapılabilir?

IOT, Web Servisleri, Veri tabanı Kayıtları, Cihaz durumları gibi sistemler ve uygulamalar üzerinde oluşan bilgi toplayıp, toplanılan bilgileri ve sistem altyapısı ve bilgilerini görselleştirme, analiz etmek için kullanılabilir. Bu sayede sistem üzerindeki yöneticinin toplanan bilgileri en hızlı şekilde analiz etmesine olanak tanır.



**Şekil 4.** ELK Stack çalışma adımları

#### Elastic.co Elastic Search Nedir?

En iyi şekilde performans elde edilebilmesi için donanımlar, Elastic Search kullanım yoğunluğuna ve birim zamanda kullanılan API isteklerinin sıklığına göre iyileştirilmelidir. Elastic Search ve Elastic Search üzerinde çalışan yardımcı araçları kurmak, konfügre etmek ve yedekleme işlemleri oldukça maliyetli bir süreçtir. Bu nedenle ELK ürünü ortaya çıkmıştır. Farklı bölgeler üzerindeki instance(örneklem)`ler üzerinde yönetimi ve performansı kolaylaştıracak konfigürasyonlar, kullanım kolaylığını arttıracak araçlar Elastic.co ürün sitesi üzerinde geliştiriciler için sunulmuştur.

* + 1. Elastic.co Üzerinden Elastic Search Kullanımının Avantajları
* Farklı ülke ve bölgelerde farklı bulut servisleri üzerinden instance`leri otomatik oluşturur. Bu sayede sistem yoğunluğuna göre farklı bölgelerdeki servislerin yoğunluğuna göre gelen istekleri dağıtır.
* Kibana, Enterprise Search, APM, Fleet gibi Elastic Stack ile çalışan araçları içerisinde barındırır. Sürümlerden kaynaklanan sorunlar meydana gelmeden, optimize sürümler ile çalışır.
* Elastic Search veri tabanı, araçları arasında hızlı bir şekilde incelenebilmesi
* Servis örneklerinin yoğunluğunu ve sağlığını hızlı bir şekilde ölçümlenebilmesi
* Elastic Cloud üzerinde çalışan RestfulAPI ile, farklı sistemler üzerinden çalışan yazılım dillerinden Elastic Search operasyonlarının hızlı bir şekilde yapılabilmesi
* Farklı deployment(dağıtım)`lar üzerinden hızlıca geçiş yapılabilmesi
* Deneme sürümü ile kullanılabilmesi
* Günlük log(kayıt) alınabilme
* Veri analizi ve görselleştirme yapılabilmesi için araçları barındırması
* Eş zamanlı olarak verileri izleyebilmesi
* Elastic Search Veri tabanı üzerinde sorgular, Manage(Yönetim) sekmesinden ara yüz ile hızlıca gönderilebilir.
  + 1. Elastic Search No Sql Veritabanı ile Sql veritabanının karşılaştırılması.

SQL ilişkisel veritabanı üzerinde tanımlı tablolar, veritabanları Elastic Search veri tabanları arasında farklı ifadelere karşılık gelmektedir. Tez üzerinde ilişkisel ve ilişkisel olmayan veritabanlarının karşılaştırılması yapılmayacaktır. Fakat çoğu kitle tarafından bilinen çalışma alanı olan SQL`in Elastic Search ile farklılıkları aşağıdaki Tablo 1`de gösterilmesi uygun görülmüştür.

|  |  |
| --- | --- |
| **Elastic Search(No SQL)** | **SQL** |
| DATABASE | INDEX |
| TABLE | TYPE |
| ROW | DOCUMENT |
| COLUMN | FIELDS |
| SCHEMA | MAPPING |

Tablo 1 – Elastic Search ve SQL üzerindeki terimlerin karşılaştırılması

* + 1. Elastic Search Araçları

Elastic Search veritabanına girilen veriler üzerinde yönetim, görselleştirme sağlanması için bünyesinde birçok araç bulundurur. Bu araçlardan proje özelinde kullanılanlar aşağıdaki gibidir.

* + - 1. Developer Console(Geliştirici konsolu)

Developer konsol, ilgili deployment üzerinde ilgili index, type vb. üzerinde yönetim sağlamak için kullanılan bir I/O(Girdi/Çıktı) mantığıyla çalışan script(yazılım) çalıştırılabilecek bir arayüzdür. Elastic.co üzerinden Deployment>Manage>Dev Tool>Console yolu takip edilerek developer consol`a erişilebilir.

#### Twitter Developer Account(Twitter Geliştirici Hesabı) Nedir?

Twitter üzerinden verilerin dağıtımı Twitter Developer Platform aracılığıyla gerçekleşir. Tüm verilere ölçülebilir, izlenebilir bir ortam sağlar. API isteği gerçekleştirilen hesapları yetkinliklerine göre inceler ve yapılan isteklerin geliştirici hesabının yetkilerini, Tweet yazma ve elde etme sınırlarını bir yazılım vasıtasıyla kontrol eder.

Twitter üzerinden başvuru süreçleri başlatılması gerekmektedir. Niyeti belirten birçok paragraf, e-posta ile yazışma sonrasında ancak hesap twitter tarafından verilebilmektedir. Bu nedenle proje amacının henüz başvuru yapılmadan belirlenmesi gerekmektedir.

Twitter API, birçok twitter özelliklerine erişimi ve bunlarla etkileşim kurulması veya oluşturulmasını sağlar:

* Tweetler
* Kullanıcılar
* Spaces(twitter canlı yayın ortamı)
* Direkt Mesajlar
* Listeler
* Trendler
* Media
* Yerler
  + 1. Twitter Geliştirici Hesap Türleri

Twitter geliştirici hesabı üzerinden verilere erişim sağlanması için bir geliştirici hesabına ihtiyaç duyulur. Geliştirici hesabına, twitter hesabı üzerinden başvurulur. Twitter geliştirici hesabına başvuru yaparken, geliştirici hesabının türü ne olursa olsun aşağıdaki sorulara cevap verilmesi Twitter geliştirici hesabı alınabilmesi için gereklidir.

* Twitter verilerini ve/veya API'lerinin nasıl kullanılması planlanmaktadır?
* Twitter verilerini analiz edilecek mi?
* Uygulama Tweet, Retweet, Beğen, Takip Et veya Direkt Mesaj işlevlerini kullanacak mı?
* Twitter dışında Twitter içeriği hakkında Tweetler veya toplu veriler görüntülenmek planlanmaktamıdır?
* Tarafınızdan oluşturulan ürün, hizmet veya analiz, Twitter içeriğini veya türetilmiş bilgileri bir devlet kurumunun kullanımına sunacak mı?
  + - 1. Bilimsel Makale İçin Geliştirici Hesabı

Bilimsel araştırmalar, akademik projeler için twitter tarafından verilen bir hesap türüdür. Bu hesabın alınabilmesi için akademik bir makalede, başvurulan geliştirici hesabındaki isim sahibinin, ilgili makalede adının geçmesi veya daha önceden bu tür çalışmalar yapıldığını ispat edecek nitelikte bir akademik makale başvuru sürecinde belirtilmelidir.

* + 1. Hobi Amaçlı Geliştirici Hesabı

Hobi amaçlı geliştirici hesabı, akademik amaçlı başvurulan geliştirici hesaplarına göre daha kolaydır. Başvuru süreçlerinde akademik amaçlı geliştirici hesaplarında göre daha az bilgi istenir.

1. YAPILAN ÇALIŞMALAR

#### Twitter Developer Hesabı API Başvuru Süreçleri

Twitter üzerinden kullanıcıların bilgilerinin elde edilebilmesi, Twitter üzerinde bot hesabının paylaşım yetkisinin verilmesi ve diğer bütün Twitter üzerinden yapılan işlemlerin yapılabilmesi için Twitter geliştirici hesabına ihtiyaç vardır. Bu bağlamda, bu başlık altında Twitter API başvuru süreçleri anlatılmıştır.

* + 1. Projede Twitter Geliştirici Hesabı

Twitter`in verilerinin ölçülebilir şekilde kullanılabilmesi için geliştiriciler için oluşturduğu, API erişim süreçlerinin yapıldığı platform üzerinde hobi Twitter API erişimi isteği gerçekleştirildi. Twitter API`ye erişim sağlanması için gerçekleştirilen bu başvuru sürecinde, projenin detayları hakkında, hangi amaçla, nerede ve hangi verilerle çalışılacağı bilgileri form olarak gönderilmiş ve sonrasında e-posta yoluyla devam eden iletişim sonucunda hobi hesabı, Twitter tarafından proje için tarafımıza verilmiştir. Hobi hesabı kapsamında aylık olarak 2 milyon adet Tweet fetch(getirme) edilebilir durumdadır (5).

#### Bulut Sunucu Hizmetine SSH Protokolü ile Erişim

SSH bağlantısında bir köprü niteliği oluşturacak bağlantıyı sağlayan ve giriş bilgilerini tutup bağlantının gerçekleştirilmesini kolaylaştıran programlar vardır. Bu bağlamda AWS Sanal makine bağlantısı gerçekleştirmesi, SSH protokolünü destekleyen PUTTY programı ile sağlanmıştır.

AWS üzerinde makineye bağlantı gerçekleştirilebilmesi için makine örneği oluşturulurken, AWS tarafından verilen özel bir anahtara ihtiyaç duyulur. Bu anahtar AWS ara yüzü üzerinden sadece makine örneği oluşturulurken bir sefere mahsus AWS tarafından verilir ve kaydedilmesi veya indirilmesi gerekmektedir. Bu anahtarı belirli bir algoritmaya göre çözümleyip özel bağlantı anahtarı oluşturulmasını sağlayan başka bir programa daha ihtiyaç vardır. Bu bağlamda Putty programının anahtar oluşturucu programı PuttyGen ile AWS üzerinden alınan anahtarı, RSA algoritmasına göre oluşturulur ve Putty programında özel bağlantı anahtarı olarak kullanılır. Algoritma sonucunda elde edilen özel anahtar Putty programına eklenir. Oturum açma için gerekli bilgiler(ip adresi, port numarası vb.) de girildikten sonra ilgili session(oturumu) kaydedildikten sonra farklı zamanlarda hızlıca bağlantı sağlanabilir duruma gelir ve bağlantı sağlanır.

#### AWS EC2 Sanal Makine Kullanımı

Sanal makineler genellikle IoT, WebServer, Windows Server, Restful Servis gibi servisler oluşturmak için kurulan bilgisayarlardır. Farklı sistemlerin veya doğrudan bir projenin taleplerini karşılar. Sanal makine seçimi, kullanılacak sistemin yoğunluğuna göre seçilmelidir. Projede özelinde bulut sunucu üzerinde kurulmuş bir sanal makine vardır. Sanal makine kurulum ve kullanımı ile geliştiricinin bilgisayarının bozulma, elektrik, internet kesintileri gibi riskler en aza indirilir. Sanal makineler, AWS üzerinde çalışan paylaşımlı Cluster yapısı içerisinde olduğundan oluşacak tehditlerden minimum şekilde etkilenecektir.

* + 1. AWS Sanal Makine Seçimi

AWS EC2 sanal makine üzerinde seçim yaparken çalıştırılmak istenilen sistemin ne kadar performansa ihtiyaç duyduğu, hafıza ve belleğini ne ölçüde kullanacağı maliyet açısından oldukça önemli ve dikkate alınıp hesaplanması gereken bir konudur. Milyarlarca Tweet verisi saklanmak istenseydi terabaytlar seviyesinde bir donanıma ihtiyaç duyulabilirdi. Proje özelinde elde edilecek maksimum Tweet sayısı 2 milyon olduğundan, elde edilecek veri miktarı ve kapladıkları alan da düşünüldüğünde 10 GB veri depolama alanı proje özelinde gerekli ve yeterlidir. Ayrıca proje, data-intensity(veri-yoğun) bir uygulama olmadığından, tek çekirdek bir işlemci kullanımı yeterlidir. Projede Amazon Linux 2 AMI (HVM) - Kernel 5.10 64-bit ARM işlemciye sahip bir sanal makinesi kurulmuştur ve 10 GB depolama alanı, SSD disk yapısı içerisinde saklanmaktadır.

#### API Key Güvenliğinin sağlanması

Github nedir

Projede github kullanımı

#### Python ile Veri Ön İşleme

Ön işleme adımları, verileri temizleyen ve makine öğrenimi modellerinin verileri daha doğru kullanmasını sağlayan bir dizi yöntemdir. Gerçek hayatta üzerinde çalışılan veri bilimi projelerinde toplanılan veriler istenilen kadar temiz olmayacaktır. Ön işleme/ön eleme adımlarında tutarsız, eksik ve görültülü veriden gerçek veri temizlenilmeye çalışılır.

Python programlama dilinde ön işleme adımları aşağıdaki şekilde gerçekleştirilmiştir.

Aşağıdaki kod, Twitter`dan gelen tweet`leri okur ve bölüm 1.1.1.5`te anlatıldığı üzere bir dataframe`a dönüştürür. Tanımlanan search\_words fonksyonuna apply fonksyonu vasıtasıyla tweet parametre olarak gönderilir. Gönderilen parametreye göre bir regex işlemi uygulanır. Bu işlemin sonucunda kelimeler hariç tüm noktalama işaretleri, fazla boşluklar kaldırılmış olunur. Ve işlemin sonucunda bir dizi oluşturulur. Oluşuturulan dizi result isimli değikene atanır. Bir sonraki satırda Join() fonksyonu çalışır. Bu fonksyon verilen dizinin tüm elemanları arasına join fonksyonuna paremetre olarak gönderilen değeri ekler ve string olarak çıktı üretilir. search\_words fonksyonu sonucunda return edilen değer, aralarında boşluk bulunan kelimelerden oluşan bir string(metin) ifadesi olmuş olur.

df["only\_words"] ifadesi, dataframe içerisindeki only\_words sütununu ifade eder. Bu sütuna ön işleme yapılan veriler eklenmiş olunur. Son adımda, tweetText değişkenine, ilk sütunun içerisinde bulunan metne atama yapılır.

 df = pd.DataFrame({"tweetTextData": [dict\_data["text"]],})

 def search\_words(text):

     result = re.findall(r"\b[^\d\W]+\b", text)

     return " ".join(result)

df["only\_words"] = df["tweetTextData"].apply(lambda x: search\_words(x))

 tweettext = df["only\_words"].values[0]

#### Python ile Sentiment(Duygu) Analizi

Projenin temel amaçlarından olan Twitter üzerinden duygu analizi, Python programlama dili kullanılarak yapılmıştır. Twitter API bağlantısı için Tweepy kütüphanesi kullanılmıştır. Anlık olarak Tweet`ler Python programlama dilinde, Tweepy kütüphanesi üzerinde tanımlı bir kütüphane olan StreamListener kullanılarak dinlenmiştir. Tweepy kütüphanesi vasıtasıyla Tweet`ler, birer birer getirilir ve JSON formatında dönüştürülür.

Elastic Search veri tabanına Twitter`dan getirilen verilerin yazılması sağlanmıştır. Python ile ElasticSearch-Py veri tabanı arasındaki bağlantıları, ekleme, getirme gibi işlemler için Python programlama dili için yazılmış ElasticSearch-Py kütüphanesi kullanılmıştır. Olumlu ve olumsuz sınıflama TextBlob kütüphanesi ile gerçekleştirilmiştir. 1 ile -1 arasında olumlu ve olumsuz şeklinde sınıflandırıp, Elastic Search veri tabanına kaydedilmiştir. Kibana ile görselleştirme işleminin gerçekleştirilmesi için, Elastic.co üzerinden deneme hesabı üzerinde Elastic Search Kibana ile görselleştirme yapılacaktır.

* + 1. Python ile duygu analizi

Projenin mantıksal gerçekleşmesi gereken her bir işlem Python dili üzerinden gerçekleşmektedir. Python dili ayrıca çeşitli API`lere bilgi taşınması, getirilmesi ve aksiyon alınması hususunda stabil bağlantı koşullarını sağlayan teknolojilere sahiptir. Python programlama dili ile Twitter API`a bağlantı gerçekleştirilir. Bu bağlantı, Twitter API üzerinden geliştiricilere sağlanan oturum açma anahtarı(Authantication Key) ile sağlanır. Python programlama diliyle Twitter arasında bağlantı kurulmasını sağlayan Twitter API vasıtasıyla, Tweet`ler üzerinde izin verilen get(elde etme) işlem limitleri, proje özelinde elde edilen hobi hesabının limitleri kadardır.

#### Elastic Search API ile Veri Tabanı İşlemleri

Elastic.co üzerinde Elastic Search veri tabanına, Python programlama dili ile bağlantı kurulmuştur. Veri tabanı işlemleri, Python programlama dili ile, ElasticSearch-Py kütüphanesini kullanarak gerçekleştirilir.

Elastic Search Restful mimari üzerinde kurgulanmıştır bu nedenle erişimler RestAPI teknolojisi üzerinde çalışır. Bu nedenle diğer Restful teknolojiler ile tam uyumlu ve JSON formatında veriler ile haberleşmeye tam uyumludur.

Python programlama diliyle yazılmış ElasticSearch-Py kütüphanesinde gönderilen istekler, http request(http isteği) ile Elastic.co üzerindeki veri tabanına RestAPI ile gönderir. (6) Veri tabanı teknolojileri üzerinde en temel işlemler CRUD işlemleridir. Bir sonraki başlık altında CRUD işlemleri tanıtılacaktır.

* + 1. CRUD Nedir?

CRUD kelimesi, Create Read, Delete ve Update kelimelerinin baş harflerinden oluşan bir kısaltmadır. Veri tabanında, kullanıcı tarafından yapılacak temel işlemlerin kısaltmasıdır.

CRUD kelimesindeki harfler aşağıda açıklanmıştır:

* Create: Veri tabanı üzerinde kayıt eklemek için kullanılan terimdir
* Read: Veri tabanından veri getirmek için kullanılan bir terimdir
* Delete: Koşullu veya koşulsuz şekilde kayıtları silmek terimidir
* Update: Veri tabanı üzerinde genellikle bir koşul belirterek belirli bir kaydı güncellemek için kullanılan bir terimdir.
  + 1. Elastic Search Python ile Ekleme İşlemi

Ekleme yapma işlemi Elastic.co üzerindeki Elastic Search içerisindeki Index, Document ve Document ID belirtilerek yapılır. Eğer Document ID belirtilmez ise, ID otomatik artan şekilde değişir. Index üzerinde bir değişiklik yapıldığında, işlemin onaylanması için ilgili Index ismi yazılır ve refresh ile yenilenerek sonraki gelecek kod satırlarının güncel Index kullanılması sağlanır.

* + 1. Elasticsearch API isteklerinin çalıştırılması

Yapılan işlemler Geliştirici Konsolu vasıtasıyla yapılabilir ve yalnızca Elasticsearch API'lerini destekler. Kibana API'leri ile Konsol ile etkileşim kurulamaz ve bunun yerine curl veya başka bir HTTP aracı kullanılması gerekir. (7)

* + - 1. Elastic Search ile Ekleme İşlemi

Aşağıdaki kod, elastic.co üzerinde 1000 adında bir document`i dolar index`i üzerine ekler. (8)

PUT dolar/\_doc/1000

{

"account\_number": 1000,

"balance": 65536,

"firstname": "Engin",

"lastname": "Karatas",

"age": 23,

}

* + - 1. Elastic Search ile Okuma İşlemi

Python ElasticSearch-Py kütüphanesi, veri tabanı üzerinde hızlıca okuma yapılabilecek fonksiyona sahiptir. Arama işlemi gerçekleştirilirken, arama yapılmak istenilen Index ve koşul girilmelidir. İlişkisel veri tabanlarında tüm kayıtları getirmek için Select deyimi kullanılırken, Elastic Search üzerinde tüm verileri getirmek için Match\_All özelliği kullanılır. Match\_All özelliği, Query objesi içerisinde yazılan bir özelliktir. Sorgular, Query objesi içerisinde gönderilir.

* Aşağıdaki kod Elastic Searh üzerindeki ukraine\_crisis\_topic isimli index`teki dokümanların hepsini listeler. (8)

GET ukraine\_crisis\_topic/\_search

{

"query": {

"match\_all": {}

}

}

* Aşağıdaki kod Elastic Searh üzerindeki ukraine\_crisis\_topic isimli index`teki dokümanların JSON formatında görünümü iyileştirilmiş halinde çıktı üretir. Bu işlemler CURL ile yapılır. CURL kullanımı zor olduğundan developer console üzerinden istekler gerçekleştirilmesi önerilir
  + - 1. Elastic Search ile Silme İşlemi

Silme işleminde silinecek işlem bir Index ise Index adı parametre olarak gönderilir. Silinecek işlem bir Document ise, Index ismi ile ilgili Document Type ve Document ID belirtilmelidir. Verilen bilgiler sağlandığı durumda, veri tabanındaki ilgili kayıt silinecektir.

Aşağıdaki kod, Developer Console üzerinde index içerisinde girilen sayıdaki document`in silinmesi işlemini gerçekleştirir. (9)

DELETE dolar/\_doc/1

* + - 1. Elastic Search ile Güncelleme İşlemi

Güncellenecek kaydın ilk parametreleri genelden özele doğru gitmektedir. Index, Document Type ve Document ID belirtilir. Belirtilen kayıt bulunduktan sonra bir diğer parametreyle güncellenecek obje yerine değiştirilecek yeni obje belirtilir. Güncellenecek kayıt mevcut ise, Elastic Search koşulları sağlayan bu yeni objeye güncellenecek obje ile değiştirir.

#### Index Export CSV İşlemi

Elastic Search üzerinde oluşturulmuş index, içerisindeki tüm verileriyle birlikte export edilebilir. Bu çıktı sonucunda .csv adında standartlaştırılmış bir format ile çıktı üretilir. Bu çıktıyı elde etmek için

Yapılan işlemler sonucunda elde edilen .csv formatı Excell programında tüm sütunlar ve değerleri şeklinde sonuçlar görüntülenebilir.

* + 1. Export İşlemi İçin Rapor Oluşturulması

İndex`in export işlemi gerçekleştirilmesi için ilgili indexin rapor olarak oluşturulması gerekmektedir. **Analytics > Discover** yolu izlenir. Bu ekranda export edilecek fields(sütunlar) belirtilebilir, eğer belirtilmezse tüm sütunları kapsayan bir csv dosyası oluşturulabilir. Belirli bir filtre özelliğine göre veriler düzenlenebilir.

Rapor oluşturmak için aynı sayfa üzerinde bulunan **Share** butonuna basılıp, **CSV Reports** seçeneği seçilir. Bu ekranda yapılan değişiklikler kaydedilebilir.

* + 1. Rapor Görüntüleme ve İndirme işlemi

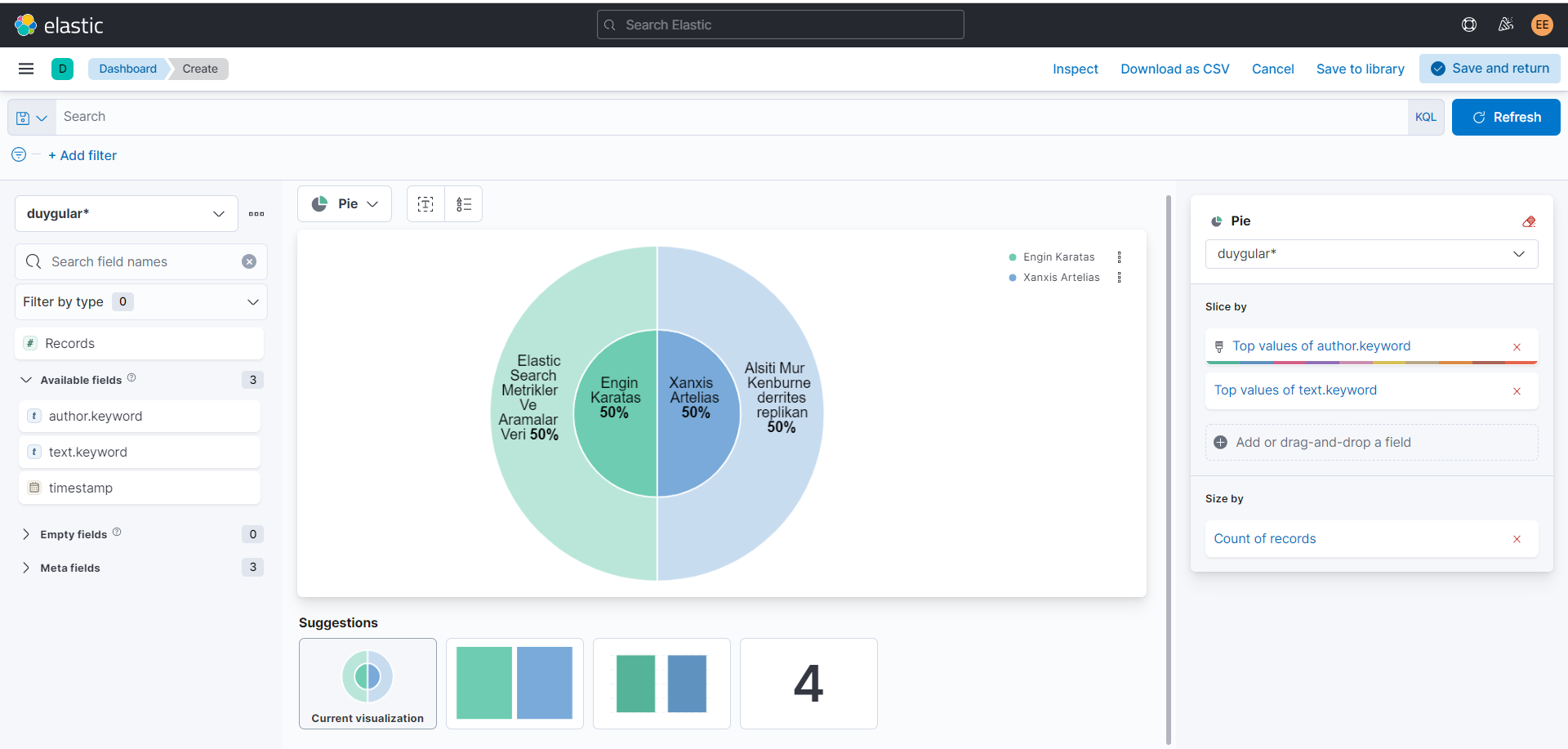
Raporlama özelliği, elastic.co üzerinde Stack Management başlığı altında tanımlanmıştır. **Management > Alerts and Insights > Reporting** yolu takip edilerek ulaşılabilir. Bu sayfa üzerinde oluşturulan tüm raporlar gözükür. Oluşturulan rapor indirilmek istenildiğinde, raporun en sağında bulunan indirme butonuna tıklandığında indirme işlemi gerçekleştirilmiş olunur.

#### Kibana Üzerinden Monitoring(İzleme), Analytics(Mantıksal Çözümleme) ve Görselleştirme

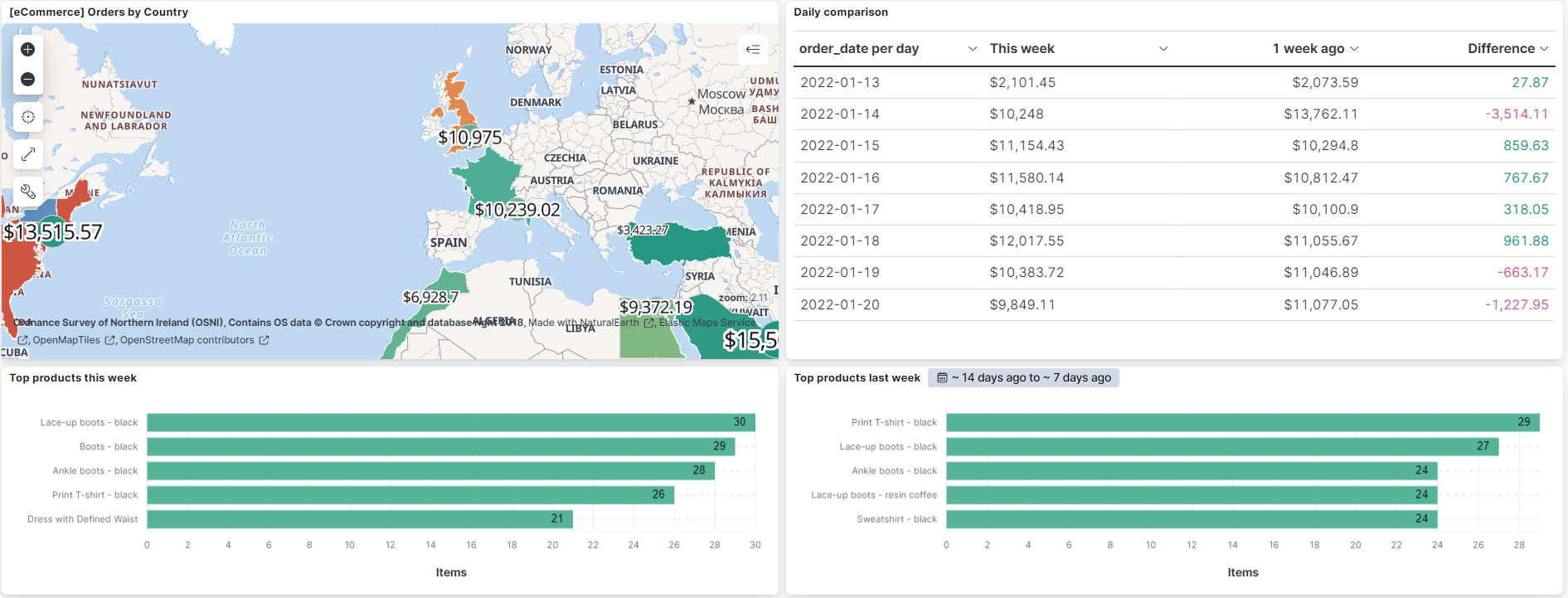
Kibana üzerinde görselleştirme yapılması için, ilgili indexlerin data view(veri görseli) olarak tanımlanması gerekmektedir. Bunun için ilgili deployment`ta ki index ,data view olarak tanımlanmalıdır. Kibana>Data View yolu içerisinde **Create Data View** butonuna tıklayarak eklenmek istenen data view seçilir ve eklenir. Bu aşamadan sonra **Analytics** içerisindeki Dashboard, Canvas, Maps ilgili index değerlerine göre oluşturulabilir.

Kibana, toplanan anlamlı verileri analiz ettikten sonra bir görselleştirme işlemi gerçekleştirir. Şekil 5`te görüldüğü üzere hacim grafiği, pasta grafiği, çubuk grafik, çizgi veya dağılım grafiği şeklinde kolayca gösterilebilir. Gösterge Paneli`ne geçiş yapıldığında Şekil 6`te görüldüğü üzere bir e-ticaret sisteminin ayrıntılı grafikleri çıkarılabilir. Aylık ülkeye göre satışlar, geçen haftaya göre satış oranlarının değişimi, bu haftanın ve geçen haftanın en çok satılan ürünlerinin grafiği vb. grafikler Şekil 6`te görülmektedir.

Kibana Gösterge Paneli(dashboard), kaydedilmiş birden çok görselleştirmeyi tek bir görünümde birleştirmeye olanak tanır. Gösterge panelindeki öğeler düzenlenebilir ve yeniden boyutlandırılabilir. Doğrudan Kibana GUI'den karmaşık grafikler oluşturmak mümkündür. Gösterge panelinde oluşturulan tablolar kolayca kaydedilebilir.



**Şekil 5.** Kibana gösterge paneli, yeni görselleştirme ekle(create visualization)



**Şekil 6.** Bir e-ticaret sisteminin Kibana ile eş zamanlı izlenmesi

Kibana üzerinde görselleştirme yapılabilmesi için, Elastic Search veri tabanı üzerinde doğru field`lere uygun verilerin yazılması gerekmektedir. Kibana, Elastic.co üzerine kaydedilen verileri arar. Bunun için Kibana kullanılırken, üzerinde arama yapılacak Index verilmelidir.

#### Ukrayna Rusya savaşnın izlenmesi

Ukrayna savaşının sonucunda belirli tarihler arasında çeşitli veriler incelenmiştir. Bu incelenen verilerin sonucu, tezin sonuç kısmında gösterilmektedir. İncelenen verilerin sayısı ve hangi tarihler arasındaki maddelerde gösterilmektedir.

* 10-11-12 Mayıs 2022 tarihleri arasında 32 bin veri incelenmiştir.

Chart, bar chart

Description automatically generated

* 3-4-5 Haziran tarihleri arasında 84.114 kayıt incelenmiştir.

Chart, bar chart

Description automatically generated

#### Dünya haritasında lokasyona göre duygu analizi

Yapılan çalışmalar bölümü 2.4 kısmında bahsedilen duygu analizi işlemleri, dünya haritası üzerinde Şekil 7`de ki gibi gösterilmiştir.

* + 1. Dünya haritası yazılımının geliştirilmesi

Dünya haritası özelliği kullanılabilmesi için, elastic.co üzerinde ilgili index üzerinde veri tiplerinin doğru tanımlanması ve Python üzerinde verileri doğru formatta gönderilmesi gerekmektedir.

2.7.1.1 ve 2.7.1.2 kısımlarında Elastic Search ve Python üzerindeki yazılım geliştirme aşamalarından bahsedilecektir.

* + - 1. Dünya haritasının Elastic Search üzerindeki geliştirmeleri

Dünya haritası özelliği kullanılabilmesi için, elastic.co index üzerinde veri tiplerinin doğru tanımlanması gerekmektedir. **Elastic Search Developer Console** üzerinde ilgili index`in location isimli property tipini geo\_point olarak olarak değiştirilmelidir. Bu sayede Python tarafından eklenecek location isimli property`nin türü geo\_location olarak gösterilir. (8)

Bu bağlamda tipi değiştirilecek olan property`nin index`i yeni oluşturulmalıdır. Çünkü mevcut index üzerinde değiştirme yapmak hata meydana getirilmektedir.

Aşağıdaki kod bloğu index\_adi isimli yere index`in adı yazılır.

PUT index\_adi

{

"mappings": {

"properties": {

"location": {

"type": "geo\_point"

}

}

}

}

* + - 1. Dünya haritasının Python (10)üzerindeki geliştirmeleri

Python tarafından eklenecek veri içerisinde, geo\_location formatı olarak seçilen property için aşağıdaki şekilde girdi yapılmalıdır. Bu örnekte girilecek olan property, elastic.co üzerinde location olarak tanımlandığından, location adında bir property, index`e ekleme yapılan document içerisine aşağıdaki formatlardan herhangi biri girilmelidir şeklinde girilebilir. (8)

**Elastic Search Location Tip Değiştirme**

PUT index\_adi

{

"text": "Geopoint as an object",

"location": {

"lat": 41.12,

"lon": -71.34

}

}

**Veya:**

PUT index\_adi

{

"text": "Geopoint as a string",

"location": "41.12,-71.34"

}

Python üzerinden index üzerine ekleme yapılırken location bilgisi şekildeki gibi girilir. Bir başka deyişle Elastic Search restful format başlığı adındaki kodlar, Python ile eklenmek istediğinde aşağıdaki şekilde girilmelidir.

        es.index(

            index="ukraine\_crisis\_topic",

            doc\_type="\_doc",

            body={

                    "location":{

                        "lat":15,

                        "lon":27

                    }

            },

        )

Proje özelinde lokasyon property`si içerisinde girilen lat ve lon değerleri, tweet`ler üzerinden alınan lokasyonların Geopy API`si sayesinde elde edilen coğrafi konum verileri ile dinamik olarak verilmektedir.

Enlem ve boylam, bir coğrafi koordinat sisteminin düzlemindeki bir konumu tanımlamak için kullanılan bir çift sayıdır (koordinatlardır). Sayılar ondalık derece biçimindedir ve enlem için -90 ila 90 ve bu değerler, boylam için -180 ila 180 arasındadır. (11)

#### Ekstra çalışmalar

Tez çalışması esnasında, tezde beklenen konular dışında yapılan çalışmayı büyütmek için çeşitli denemeler yapılmıştır. Yapılan ve bahsedilmemiş işlemler aşağıdaki şekildedir.

* Elastic Search kurulumunu elastic.co ürünü haricinde manuel olarak linux sistem üzerine kurulmuştur.
* Linux Server üzerinde Python kodlarını daha iyi test edilebilmesi için Jupyter Notebook kurulmuştur.
* Farklı Elastic Search sürümleri ve migration(sürümler arası geçişler) yapılmıştır.
* Onlarca hatalar alınmıştır ve çözülmüştür.

1. SONUÇ

Twitter`ın açık kaynak API`ini kullanarak, Tweet analizi gerçekleştirilmiştir. Girdi olarak Tweet`ler Twitter platformunda, en fazla 280 karakter uzunluğunda metin verileri alınmıştır. Çıktı olarak insanların konu/konular hakkındaki düşünceleri analiz edilmiştir.

Sonuç kısmında ilişkisel olmayan Elastic Search veri tabanı üzerinde kayıtlı analiz sonucu verilerinin, Kibana üzerinde görselleştirilmiştir.

Analiz edilen verilerden elde edilen anlamlı veriler, Twitter API üzerinden Twitter üzerinde paylaşılmak üzere Twitter API`ye başvuru yapılan profil hesabından Public Account (herkese açık hesap) üzerinden paylaşılmıştır.

Elde edilen sonuçlarda ukryna savaşı ve dolar kuru incelenmiş ve çeşitli çıktılar elde edilmiştir.

#### Ukrayna Savaşı Analiz Sonuçları

#### Dolar İncelenmesi Sonuçları

### *KAYNAKÇA*

1. **Number of Twitter users worldwide from 2019 to 2024. *Statista.* [Çevrimiçi] [Alıntı Tarihi: 1 Ocak 2022.] statista.com/statistics/303681/twitter-users-worldwide/.**

**2. Pandas. *Vikipedi.* [Çevrimiçi] 2022 02 15. [Alıntı Tarihi: 2022 6 17.] https://tr.wikipedia.org/wiki/Pandas.**

**3. Güvenli Kabuk (SSH). *Vikipedi.* [Çevrimiçi] [Alıntı Tarihi: 2 Ocak 2022.] https://tr.wikipedia.org/wiki/G%C3%BCvenli\_kabuk.**

**4. Güvenli kabuk. *Wikipedia.com.* [Çevrimiçi] [Alıntı Tarihi: 20 Ocak 2022.] https://en.wikipedia.org/wiki/Secure\_Shell.**

**5. Docs. *Twitter Developer Platform.* [Çevrimiçi] https://developer.twitter.com/en/docs.**

**6. Accounts. *Twitter Developer Platform.* [Çevrimiçi] [Alıntı Tarihi: 20 Ocak 2022.] https://developer.twitter.com/en/docs/twitter-ads-api/campaign-management/api-reference/accounts.**

**7. Run Elasticsearch API requestsedit . [Çevrimiçi] [Alıntı Tarihi: 17 6 2022.] https://www.elastic.co/guide/en/kibana/8.2/console-kibana.html#console-kibana. ELK Version [8.2].**

**8. Search, Elastic. Elk Documentation. [Çevrimiçi] [Alıntı Tarihi: 13 6 2022.] https://www.elastic.co/guide/en/elasticsearch/reference/current/geo-point.html.**

**9. Execute CRUD Operations in Elasticsearch. *A Cloud Guru.* [Çevrimiçi] [Alıntı Tarihi: 13 6 2022.] https://acloudguru.com/hands-on-labs/execute-crud-operations-in-elasticsearch.**

**10. geeksforgeeks. *How to get Geolocation in Python?* [Çevrimiçi] [Alıntı Tarihi: 8 6 2022.] https://www.geeksforgeeks.org/how-to-get-geolocation-in-python/.**

**11. Mapbox Docs. [Çevrimiçi] [Alıntı Tarihi: 6 13 2022.] https://docs.mapbox.com/help/glossary/lat-lon/#:~:text=Latitude%20and%20longitude%20are%20a,180%20to%20180%20for%20longitude..**

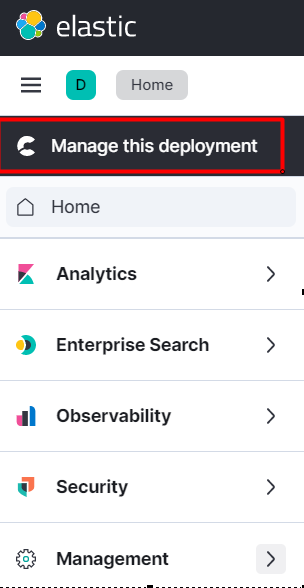
**12. *Building an IoT Data Hub with Elasticsearch, Logstash and Kibana.* Bajer, Marcin. Starowislna : IEEE, 2017.**

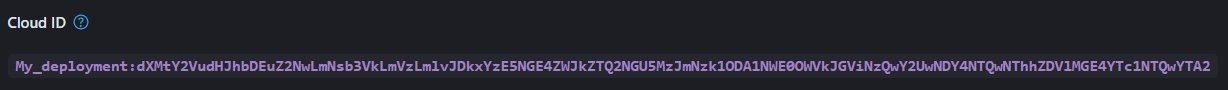
### ÖZGEÇMİŞ

Engin Karataş, Eylül 1999 doğumlu Kocaeli, Gebze yerlisidir . Yozgat Bozok Üniversitesi Bilgisayar Mühendisliği lisans öğrencisidir. Birçok kez onur öğrencisi olarak seçilmiştir. Birçok yazılım, donanım etkinliklerine katılmıştır. Kendini yazılım alanında geliştirmiş ve farklı kategorideki projelere sahiptir. Ağ üzerinden çok oyunculu ve tek oyunculu oynanabilen oyunlar yazmıştır. Yazılım geliştirme prensiplerine hâkim olmayı kendine prensip edinmiştir. Java, Angular, C#, Nodejs, Javascript teknolojileriyle projeler geliştirmiştir. Drone Kontrol istasyonu, e-ticaret yönetimi, firma içi kullanıcıların raporlarının takibi ve yönetimi, ağ elemanları ve çeşitli donanım teknolojileri üzerinde monitoring(izleme) vb. projelerde bireysel veya grup halinde çalışmıştır.

ELASTIC CLOUD API KEY erişimi

Elastic search cloud id alınabilmesi için elastic.co enterprise sitesi üzerinden şekil x`te görüldüğü üzere navbar üzerinde bulunan “Manage this deployment” yazısına tıklayarak ilgili sayfaya girilmelidir.



Bu sayfa üzerinde ilgili deploymente ait instance bilgileri, var olan aplication endpointleri, cluster id leri ve yazılım ile erişilip, işlemleri yazılımla yapabilmek için şekil y de görüldüğü üzere gerekli Cloud ID bulunmaktadır. 

Not: proje özelinde Cloud ID, Python programlama dili ile elastic.co enterprise üzerine bağlanılmak üzere kullanılmıştır.