

T.C
SAKARYA ÜNİVERSİTESİ
İŞLETME FAKÜLTESİ

İSTANBUL 3. HAVALİMANI:
TWITTER VERİLERİ İLE METİN MADENCİLİĞİ
VE DUYGU ANALİZİ

BİTİRME ÇALIŞMASI
Metehan DURUTUNA

Bölüm: Yönetim Bilişim Sistemleri

Bu çalışma .../.../20... tarihinde aşağıdaki jüri tarafından oybirliği/oyçokluğu ile kabul edilmiştir.

Jüri Başkanı	Jüri Üyesi	Jüri Üyesi
<input type="checkbox"/> Kabul	<input type="checkbox"/> Kabul	<input type="checkbox"/> Kabul
<input type="checkbox"/> Red	<input type="checkbox"/> Red	<input type="checkbox"/> Red
<input type="checkbox"/> Düzeltme	<input type="checkbox"/> Düzeltme	<input type="checkbox"/> Düzeltme

İÇİNDEKİLER

1. GİRİŞ	1
1.1. Çalışmanın Amacı ve Önemi	1
1.2. Çalışmanın Yöntemi	1
1.3. Çalışmanın Kapsamı	2
2. LİTERATÜR İNCELEMESİ	2
2.1. İstanbul Havalimanı Literatür İncelemesi	2
2.2. Metin Madenciliği ve Duygu Analizi Üzerine Literatür İncelemesi	4
2.3. Makine Öğrenmesi ile Metin Sınıflandırma Üzerine Literatür Çalışması	5
3. İSTANBUL HAVALİMANI HAKKINDA	6
3.1. İstanbul Havalimanı Çevresel Faktörleri	7
3.2. İstanbul Havalimanı Yapısı	8
3.3. İstanbul Havalimanı Yolcu İstatistikleri	8
4. YÖNTEM	10
4.1. Veri Setinin Elde Edilmesi	11
4.1.1. Veri Seti Hakkında	13
4.2. Veri Ön İşleme	14
4.2.1. Metinlerin Küçük Harfe Çevrilmesi	14
4.2.2. Metinlerdeki Gereksiz Karakterlerin Temizlenmesi	15
4.2.3. Metinlerdeki Noktalama İşaretlerinin ve Sayıların Temizlenmesi	15
4.2.4. Metinlerdeki Stop Words'ün Temizlenmesi	16
4.2.5. Az Geçen Kelimelerin ve Boşlukların Kaldırılması	17
4.3. Makine Öğrenmesi Veri Seti Hazırlama	18
4.4. Makine Öğrenmesi İçin Kullanılacak Algoritmalar	20
4.5. Modelin Eğitilmesi ve Uygulanması	22
4.6. Duygu Analizinin Uygulanması	27
5. BULGULAR	29
5.1. Tüm Verilere Ait Duygu Skorları	29
5.2. Tüm Verilere Ait Tahminler	30
5.3. Sonuçların Görselleştirilmesi	31
SONUÇ	36
KAYNAKÇA	37

1. GİRİŞ

Şu an içinde bulunduğumuz dünyada teknoloji her geçen gün gelişmeye devam ediyor. Teknolojinin bu kadar gelişmesiyle birlikte insanların internet dünyasında geçirdiği vakit ve internete verdiği değerde artıyor. İnsanların interneti bu kadar çok kullanması sosyal medya platformlarının da gelişmesini sağlıyor. Milyonlarca insanın her gün çok fazla vakit geçirdiği sosyal medya platformlarında ürettikleri verilerde çok yüksek boyutlara ulaşıyor. Bu veriler ışığında pek çok konu hakkında analiz yapılmasını kolaylaştırıyor ve bu verilere ulaşmakta oldukça kolay oluyor. Bu çalışmamızda Türkiye’de vatandaşların İstanbul 3. Havalimanı hakkında düşüncelerini analiz etmeye yönelik bir çalışma yapıldı. Sosyal medya platformu olan Twitter’den elde edilen veri seti kullanıldı. Bu verilere, çeşitli yöntemler uygulanarak atılan tweetlerin duyguları analiz edildi.

1.1. Çalışmanın Amacı ve Önemi

İstanbul’da iki havalimanı olmasına rağmen üçüncü bir havalimanı yapma fikri ilk olarak 2007 yılında kulislere konuşulmaya başlandı. 2013 yılında proje için somut adımlar atıldı. Bu adımla birlikte İstanbul’a yeni bir havalimanı yapma fikri, projenin yapılacağı bölge ve dönemin ekonomik şartlarından dolayı vatandaşlar tarafından iyi ve kötü eleştirilere maruz kaldı.

Bu çalışmanın amacı, Twitter’da insanların 3. Havalimanı için atmış olduğu tweetler üzerinden olumlu veya olumsuz düşünceleri belirlemeye yönelik bir çalışmadır.

1.2. Çalışmanın Yöntemi

Bu çalışmada, yöntem olarak anlaşılır olabilmesi ve doğru sonuçlara ulaşabilmek adına 5 adımı bir yol izlenmiştir. Bunlar sırasıyla, Python programlama dili kullanarak veri toplama, veri ön işleme, makine öğrenmesi seti oluşturarak kategori etiketleme, duygu analizi ve sonuçların görselleştirilmesidir. Bu adımlar veriyi anlamlı ve okunur bir hale getirerek sonuçlar üzerinden doğru bir şekilde çıkarım yapmamıza ve o konu hakkında fikir yürütebilmemize olanak sağlamaktadır. Aynı zamanda analizler sonucunda elde edilen çıktılar, Microsoft Excel kullanılarak kaydedilmiştir.

1.3. Çalışmanın Kapsamı

Çalışmada kullanılan veriler, sosyal medya platformu olan Twitter'dan elde edilmiştir. Verileri elde etmek için 'İstanbul Havalimanı', '3. Havalimanı' ve 'İstanbul 3. Havalimanı' gibi arama terimleri aynı anda kullanılmıştır. Aynı zamanda tweetler çekilirken 2014 yılının 1 Ocak tarihinden günümüze kadar olan tweetler çekilmiştir. Çalışmada, toplamda 103.108 adet tweet kullanılmıştır. 103.108 adet veriyi ekonomi, siyasi ve çevresel yönden değerlendirebilmek zor olacağı için her kategoride 250'şer veri örneklem olarak seçilip bir makine öğrenmesi veri seti hazırlanmıştır. Hazırlanan veri seti eğitilip tüm veride bir kategori etiketlemesi yapılmıştır.

2. LİTERATÜR İNCELEMESİ

Bu çalışmada, doğru bir şekilde ilerleyebilmek adına, İstanbul Havalimanı, Makine Öğrenmesi ile Metin Sınıflandırma, Metin Madenciliği ve Duygu Analizi ile ilgili daha önceden yapılan çalışmalar incelenmiştir. Literatür incelemesini yapmamızın asıl nedeni çalışmamıza ışık tutabilmesi, bu çalışmayı yaparken bize yol göstermesi içindir.

2.1. İstanbul Havalimanı Literatür İncelemesi

İstanbul 3. Havalimanı Avrupa Yakası'nda bulunan Arnavutköy ilçesinin sınırları içerisinde yer alan Yeniköy, Tayakadın ve İmrahor'da bulunuyor. İstanbul'a yeni bir havalimanının yapılıyor olması beraberinde tartışmalara yol açtı. Bu tartışmaların başında ise İstanbul Havalimanı'nın çevreye vereceği olumlu veya olumsuz etkileriydi.

Şekil 1. İstanbul Havalimanı bölgesinin önceki ve sonraki hali



Kaynak: (Köker, 2018)

Şekil 1’de görüldüğü gibi İstanbul Havalimanı bölgesinin yapılmadan önceki ve yapıldıktan sonraki değişimi gözlemlemekteyiz. Tartışmalara neden olan çevredeki değişimin, doğal yaşam alanına olan müdahalesiydi. Bu müdahaleye rağmen bazı vatandaşlar bu projenin ekonomik anlamda kazanç getireceğini, bazı vatandaşlar ise bu projenin rant için yapıldığını, bölgeye olan zararını ve ekonomik anlamda sadece bazı insanlar için kazanç getireceğini savundu.

Özalp ve Sönmez tarafından 2022 yılında yapılan bir çalışmada İstanbul’a yapılan 3. Havalimanı’nın bölgedeki doğal tabiata olan olumlu veya olumsuz etkilerinin yanında bölgenin bu denli bir büyük projeye dahil olmasının aynı zamanda çevresinde meydana getireceği değişime de değinmiştir. Bölgenin ormanlık bir alan olması, etrafında çok fazla yerleşimin olmaması, zamanla değişeceği anlamına geliyor. Çünkü ekonomiye çok büyük olumlu veya olumsuz katkısı olacak olan bu projenin, çevresinin her geçen gün gelişecek olmasını gösteriyor. Çalışmada değinilen bir başka konu ise proje başlanmadan önce bölgenin çevresindeki arazilerin satın alınmasına yönelik eleştirilerdi. Bunun nedeninin ise rant olduğu vurgulanmıştır. Ayrıca bölgede çok hızlı bir şekilde konut yapımına başlanmış olduğuna ve bölgedeki doğal yaşam alanının ihlal edildiğine değinilmiştir. (Özalp ve Sönmez, 2022)

Kılıç ve Turgut tarafından 2021 yılında yapılan bir çalışma, İstanbul Havalimanı’nın ve Atatürk Havalimanı’nın bulundukları konuma göre lojistik ve ulaşım yönünden olumlu ve olumsuz yönlerine değinmiştir. Atatürk Havalimanı’nın şehrin merkezinde olmasından dolayı şehir içerisinde birtakım karışıklıklara neden olacağı yönünde sonuçlara ulaşılmıştır. Bunun nedeni İstanbul şehrinde çok fazla araç trafiğinin olması, olası havayolu yolculukları için gecikmelere neden olacağı vurgulanmıştır. Aynı zamanda trafikte çok fazla zaman geçirmenin, araçların sürekli dur kalk yaparak ilerlemesinin de araçlarının yakmış olduğu yakıtların havayolu kirliliğine ve çevre kirliliğine neden olacağı vurgulanmıştır. Bu nedenden dolayı İstanbul Havalimanının şehrin dışında olması ulaşım sırasında herhangi gecikmenin çok fazla olmayacağına değinmiştir. İstanbul Havalimanı’na insanların ulaşımı kolaylaştırmak adına birden fazla ulaşım alternatifinin hayata geçirilmesini gerektiği belirtilmiştir. Ayrıca Atatürk Havalimanı şehrin ortasında yer almasından dolayı kargo şirketleri açısından avantajlı olduğu savunuluyor.

Avantajlı olmasının sebebi, ulařtırma aısından civar ilelere olan yakınlık kargo řirketlerinin ulařım maliyetlerini azaltıyor. Ama İstanbul Havalimanı řehrin dıřında olması ve merkeze uzak olması, kargo řirketlerinin maliyetini artırdığı ve müşteri memnuniyetini azalttığı yönünde konulara değinilmiştir. (Kılı ve Turgut, 2021)

2.2. Metin Madenciliğı ve Duygu Analizi Üzerine Literatür İncelemesi

Günümüzde sosyal medya kullanımının sürekli artması, insanların her gün farkında olup ya da olmadan veriler paylaşmasına neden oluyor. Sosyal medyada paylaşılan herhangi bir yazı, fotoğraf veya yorum aslında insanların verileridir. Bu veriler her geen gün katlanarak yüksek boyutlu verilere dönüşüyor. İşte tam bu noktada insanların verilerinden anlamlı bir veri elde edilmek için Metin Madenciliğı yapılıyor. Metin madenciliğı yapılırken metinde geen tüm kelimelerin küçük harflere çevrilmesi, gereksiz kelime ve simgelerin kaldırılması, sayıların ve noktalama işaretlerinin temizlenmesi gibi işlemleri kapsamaktadır. Verilere bu yöntem uygulandığında verinin okunması ve üzerinde yapılacak olan analizlerde daha doğru sonuçlar elde edilmesine olanak sağlıyor. (Beřkirli ve diğ., 2021)

Sosyal medyanın beraberinde getirdiğı devasa büyüklükteki verilerin metin madenciliğı ile gereksiz kelimelerin, simgelerin kaldırılması, noktalama işaretlerinin temizlenmesi, sayıların kaldırılması gibi ön işleme aşamalarının ardından veriden değeri, ulařılması gereken bilginin alınması için duygu analizi yöntemi kullanılır. Bunlar metinlerin duygusuna yönelik olumlu veya olumsuz řekilde ayrılması yöntemidir. Bazı durumlarda ise okunan metinde tam olarak duygu belirlenemediğı zamanlarda nötr duygu olarak ayrıca karřımıza çıkıyor. Duygu analizi, bir kiřiye, ürüne veya topluluğı karřı olan duygularını anlamaya yardımcı olur. (Akgül ve diğ., 2016)

Koca tarafından 2021 yılında yapılan bir alıřmada Twitter'dan 5.000 adet tweet verisi elde edilmiştir. Elde edilen tweetlerin Bitcoin ile ilgili bir duygu analizi alıřması yapılmıştır. Bitcoin ile ilgili tweetler incelendiğinde çoğı tweetin aslında duygu patlaması sırasında atıldığına dair bir ıkarım yapılmıştır. Örneğın Bitcoin'in anlık olarak fiyatının artması veya ařağı yönde akılması kimi kullanıcılar tarafından mutlulukla karřılanırken kimi kullanıcılar tarafından ise olumsuz bir řekilde karřıldığı gözlenmiştir. Bunun nedenin ise tamamen o anki duygudan kaynaklandığı belirtilmiştir.

Aynı zamanda bazı Bitcoin yatırımcıları tarafından beklenmedik alımlar veya satışlar esnasında kullanıcıların tepkisine neden olduğu yönünde bir çıkarım yapılmıştır. Sonuç olarak tweetler incelendiğinde atılan tweetlerin anlık olarak o anki Bitcoin durumuyla alakalı olduğu belirlenmiştir. (Koca, 2021)

2.3. Makine Öğrenmesi ile Metin Sınıflandırma Üzerine Literatür Çalışması

Günümüz çağında internetin çok sık bir şekilde yaşam alanımıza girmesi beraberinde kullanılan sosyal medya araçlarının yaygınlaşmasına neden olmuştur. Bununla beraber her geçen gün internet ortamında bulunan veriler devasa bir şekilde artmaya devam ediyor. İnternet ortamından çekilen, temizlenmemiş veriler içerisinden bir anlam çıkarabilmek epey zahmetli bir işlem olarak karşımıza çıkıyor. İşte bu noktada devreye makine öğrenmesi algoritmaları giriyor. Makine öğrenmesi algoritmalarının mantığı ise yüksek boyutlu veri setlerinin sınıflandırılmasının zor olduğu anlarda bir örneklem belirleyerek aynı veri setinde daha az veriyle etiketleme yapılarak bir kategori oluşturulur ve böylece bu kategori sayesinde veriler bir sınıflandırılmaya tabii tutulur. (Toçoğlu ve diğ., 2019)

Erdi ve diğerleri tarafından 2021 yılında yapılan çalışmada, Twitter’da insanları sürekli rahatsız edip, kötü deneyim yaşamlarına neden olan Trol hesaplarla ilgili bir çalışma gerçekleştirmişlerdir. Yapılan bu çalışmada 100 adet Trol hesap incelenmiş ve 238.925 adet tweet elde edilmiştir. Çalışmanın mantığı trol mü yoksa gerçek bir hesap mı diye ayırım yapılmasıdır. Bunu anlamak için kullanıcıların, kullanmış olduğu kelime sıklığına bakılmıştır. Böylece hesabın gerçek veya trol olduğu saptanmıştır. Bu kelimeler ise çalışmada incelenmesi gereken 2.102 adet özellik ortaya çıkarmıştır. Bu özelliklerin hangi hesaplar tarafından kullanıldığı tespit edilmeye çalışılmıştır. Örneğin 100 kullanıcı hangi özelliği kullandıysa bu 1 olarak alındı eğer o kullanıcı o özelliği kullanmadıysa 0 olarak kaydedilerek bir sınıflandırma oluşturulmuştur. Daha sonra bu bilgiler için diğer bir öznitelik olarak hesaplar gerçek ise sütuna 0, trol ise 1 olarak eklendi. Makine öğrenmesi algoritmaları yöntemlerinden olan Support Vector Machine, Random Forest, Logistic Regression yöntemleri bu bilgiler dahilinde test edildi. LR yöntemi sıkça kullanılan 50 özellik içerisinde 0-1 olarak en iyi sonucu (%93.93) verdiği gözlemlenmiştir. Tüm özelliklerin olduğu veri de ise en yüksek başarımlı SVM algoritması (%87,57) ile elde edilmiştir. (Erdi ve diğ., 2021)

3. İSTANBUL HAVALİMANI HAKKINDA

İstanbul'a yeni bir havalimanı yapılabilmesi için ilk ihale 2013 yılında yapıldı. İhaleyi kazanan şirketler ise İGA adı altında bir şirket kurarak havayolu işletmeciliğini 25 yıllığına devraldı. Planlanan proje ise Cumhuriyet tarihinin en büyük projesi olarak adlandırılıyordu. İnşaat, 4 fazda yapılacak olup, yapımı tamamlandığında ise dünyanın en büyük havalimanı olma özelliğini taşıyacaktır. Yeni Havalimanı'nın 29 Ekim 2018'de açılışı yapıldı. Açılıştan sonra ise 'İstanbul Havalimanı' adını aldı. İstanbul Havalimanı'nda ilk uçuş Ankara'ya gerçekleştirildi. (Şantiye Dergisi, 2020)

İstanbul Havalimanı, dünyanın en büyük havalimanlarından biridir. Ayrıca bulunduğu konum itibari ile havayolu taşımacılığında merkezi bir öneme sahiptir. Dünya'da havacılık sektörü adına popüler olan yayınlar arasında yerini alan 'Air Transport Awards' ödüllerinde 2021 ve 2022 yıllarında iki kez üst üste dünyada 'Yılın Havalimanı' seçildi. (Boon, 2022)

İstanbul Havalimanı, 'Travel and Leisure Dergisi' kullanıcıları tarafından dünyanın en iyi ikinci havalimanı seçildi. Okurlar oylarını belirlerken, seyahatte yaşadıkları iyi veya kötü anları, havayolu şirketi tarafından yapılan hizmeti, tesisler ve ulaşım gibi tamamen yolcuların refahına yönelik olan kaliteli hizmetleri değerlendirdi. İstanbul Havalimanı sırasıyla Osaka(Japonya), Zürih(İsviçre), Narita(Japonya), Hong Kong, Haneda(Japonya Hamad(Katar), Dubai, Incheon(Güney Kore) havalimanlarını gerisinde bırakarak Singapur Havalimanından sonra 'Dünyanın En iyi Uluslararası Havalimanları' oylamasında 2021 yılında 90,17 puan alarak ikinci oldu. (Brady, 2021)

Şekil 2. İstanbul Havalimanı



Kaynak: (İGA, 2018)

3.1. İstanbul Havalimanı Çevresel Faktörleri

Günümüzde yaşanan ve tüm dünyayı etkileyen olaylarda büyük kentler bundan daha çok etkilenmektedir. Olası göç vb. sorunlarda, barınma olanakları genişletebilmek için hiçbir plan yapmadan doğayı ve var olan yapıyı bozarak konut inşa ediliyor. İşte tam bu noktada 3. Havalimanı'nın yapılması ekonomik anlamda getireceklerinin yanında, çevreye vereceği olumsuz etkiler konuşulmaktadır. 3. Havalimanının yapılacağı bölgeye yakın olan Terkos gölünün inşaat süresi boyunca yapılacak olan çalışmalardan dolayı olumsuz yönde etkilendiği belirtilmiştir. Aynı zamanda Yeniköy civarına yapılacak olan havalimanının, Terkos gölünü besleyen Yeniköy deresi, Üstülük deresi ve Ceko deresinde, inşaat nedeni ile doğal yapısının bozulacağı söylenmiştir. Projenin yapıldığı bölge denize kıyısı olduğu için meteorolojiden alınan bilgilerden yola çıkarak denizden gelecek olan rüzgarlar nedeniyle iniş yapacak olan uçaklar için riskli olacağı belirtilmiştir. Proje yapılmadan önce bölge henüz doğal tabiatın bozulmadığı bir bölgeydi. Ayrıca doğal yaşam için de oldukça önemli bir noktaydı. İstanbul'u oluşturan ormanlık alanlarının %2,5'i İstanbul Havalimanı'nın yapıldığı bölgeyi kapsamaktaydı. Bu nedenden dolayı ormanlık alan tahrip edildiği için siyasi, ekonomik ve çevresel yönden tepkilere neden oldu. Bölgede 17 endemik bitki türünün olduğu ve bu oran İstanbul ile karşılaştırıldığı zaman %41'ine denk geldiği söylenmektedir. Ayrıca bölgede sürüngen, amfibi, kuş ve memeli türlerinin yaşadığı bilinmektedir. İstanbul havalimanı inşaatı sonrası Flora(Bitki Türü) ve Fauna'yı(Vahşi Yaşam) olumsuz yönde etkileyeceği raporlanmıştır. Rapordaki bir diğer detay ise havalimanının olduğu bölgenin doğal yaşamın içerisinde olduğundan kuşların göç yolu üzerinde yapıldığı raporlanmıştır. Göç yolu üzerinde olması uçaklar için büyük bir tehlike arz ettiği söylenmektedir. (TMMOB, 2014)

Şekil 3. İstanbul Havalimanı bölgesi inşaat öncesi



Kaynak: (Şimşek, 2018)

3.2. İstanbul Havalimanı Yapısı

İstanbul Havalimanı, tam anlamıyla hizmete başladığında 200 milyon yolcuyu ağırlamaya yetecek kapasiteye sahip olacak. İstanbul Havalimanı, terminal binası ise 1,3 milyon m²'lik büyüklüğü sahip bu anlamda dünyanın en büyük terminal binası olarak kayıtlara geçti. İstanbul Havalimanı'nın bünyesinde dünyanın en büyük uçakları arasında yer alan Airbus A380 ve Boeing 747-800 gibi uçakların iniş yapabileceği 4 km uzunluğunda 5 pist ve Karadeniz kıyısına paralel olarak uzanan 1 pist ile toplamda 6 pisti bünyesinde barındırıyor. Aynı zamanda İstanbul Havalimanı'nda iki taksi yolu olarak kullandığı ve bazı durumlarda acil pist olarak kullanabileceği pistleri bünyesinde barındırıyor. Ayrıca ödül alan lale tasarımlı 90 metre uzunluğunda 17 katlı bir hava trafik kontrol kulesi, İstanbul Havalimanı yapılarında yer alıyor. (TRT Haber, 2021)

İstanbul Havalimanı'nda 40.000 aracın rahat bir şekilde araçlarını park edebileceği otopark alanı mevcuttur. (igaaairport, 2019)

İstanbul Havalimanı'nda tüm dünyayı etkisi altına alan Covid-19 salgını nedeniyle 5.000 m² alana kurulan Covid-19 Test merkezi bulunmaktadır. (Bozdağ, 2021)

3.3. İstanbul Havalimanı Yolcu İstatistikleri

İstanbul Havalimanı, açılışını yaptığı 29 Ekim 2018 tarihinden itibaren pandemi yaşanmasına rağmen toplamda 112,5 milyonu aşkın yolcuya hizmet verdi. (Aslanhan, 2021)

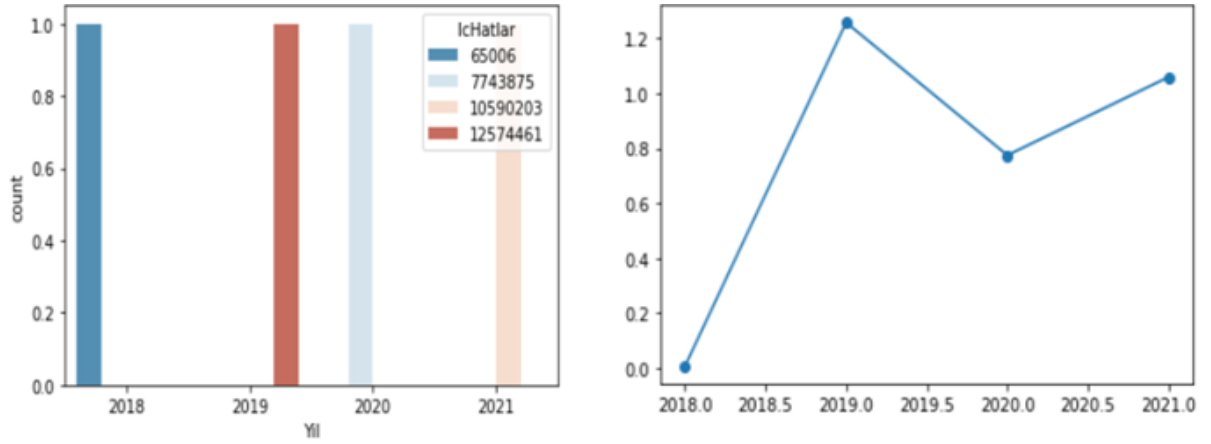
Şekil 4. İstanbul Havalimanı iç ve dış hatlar yolcu istatistikleri

Yıl	İç Hatlar	Dış Hatlar	Toplam
2018	65.006	30.199	95.205
2019	12.574.461	39.434.579	52.009.220
2020	7.473.875	15.936.505	23.410.380
2021	10.590.203	26.586.306	37.176.509

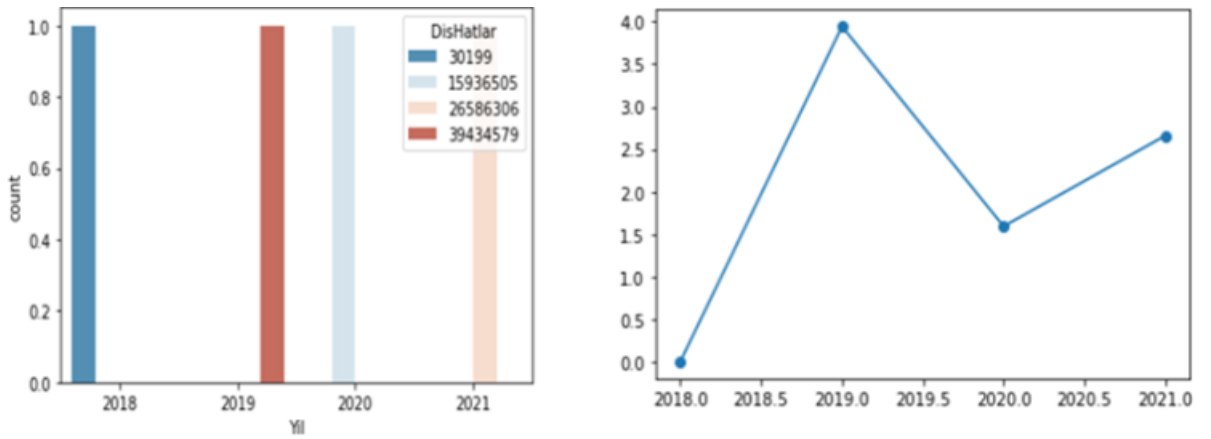
Kaynak: (DHMI, 2021)

Şekil 4'te gösterilen yıllara göre yolcu sayılarının İç Hatlar, Dış Hatlar ve Toplam Yolcu sayısını gösteren grafikler aşağıda yer almaktadır. Covid 19 salgınının şiddetli bir şekilde yaşandığı 2019-2020 yıllarında neredeyse tüm dünyada durdurulan uçuşlar nedeniyle yolcu sayılarında bir düşüş yaşanmıştır.

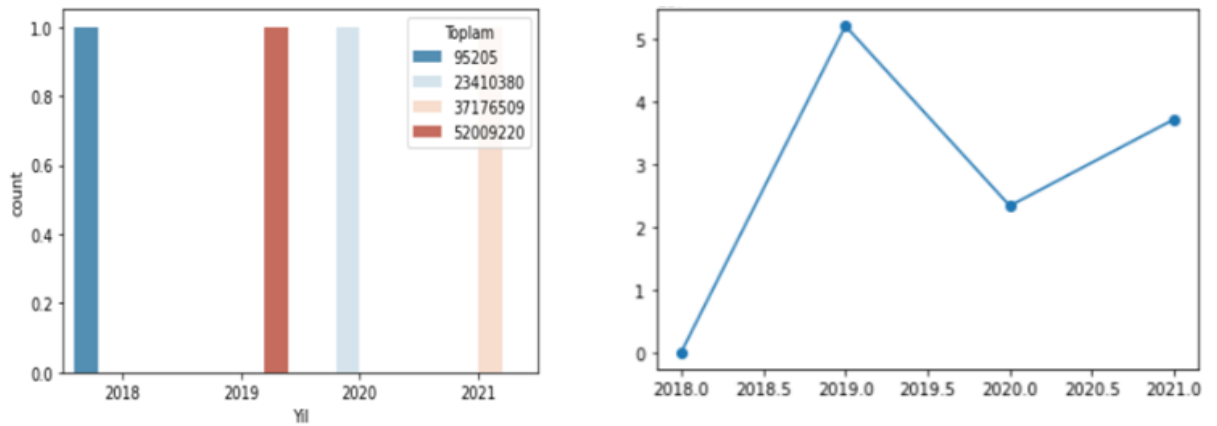
Şekil 5. İstanbul Havalimanı iç hatlar yolcu sayısı grafikleri



Şekil 6. İstanbul Havalimanı dış hatlar yolcu sayısı grafikleri



Şekil 7. İstanbul Havalimanı toplam yolcu sayısı grafikleri



4. YÖNTEM

İstanbul Havalimanı hakkında Twitter verileri ile duygu analizinin yapılması için veri seti elde edildikten sonra, veri ön işleme adımları uygulanıyor. Daha sonra yüzbinlerce olan veri setinde bir örneklem oluşturabilmek için makine öğrenmesi seti oluşturularak topikler belirlenip kategori etiketlemesi yapılıyor. Makine öğrenmesi setinde model oluşturulup, eğitilerek tüm verilere odaklı bir tahmin yapılıyor. Tüm veri setini olumlu ve olumsuz yönden gruplamak için duygu analizi uygulanıyor. Analizler sonrası elde edilen sonuçlar görselleştiriliyor.

Şekil 8.Yöntem diyagramı



4.1. Veri Setinin Elde Edilmesi

Twitter kullanıcılarının attığı tweetleri çekebilmek için sınırsız bir şekilde bizlere tweet çekebilme imkanı sunan ‘snsrcape’ kütüphanesi kullanarak ‘Google Colab’ üzerinden veri çekme işlemi gerçekleştirildi.

Şekil 9. Kütüphanelerin, tarihin ve arama terimlerinin yazdırılması

Metin Verilerinin Çekilmesi & İşlenmesi

```
[ ] #Snsrcape yükleme
!pip install -q snsrcape==0.3.4

[ ] #Gerekli olan kütüphanelerin import edildi.
import os
import pandas as pd
from datetime import date
import string
import time
import typing
import urllib.parse

[ ] #Geçerli yerel tarihi döndürür.
today = date.today()
end_date = today

[ ] #İstanbul 3. Havalimanı ile ilgili
#2014 yılının 1 ocak tarihinden itibaren atılan tweetlerini çekme işlemi gerçekleştirilecek.
search_term = '3.havalimanı OR istanbulhavalimanı OR İstanbul3.Havalimanı'
from_date = '2014-01-01'
```

Şekil 9’da görüldüğü gibi ilk olarak snsrcape 3.4 versiyonu ve beraberinde kullanacağımız kütüphaneler import edildi. İport ettiğimiz ‘time’ kütüphanesi bir sonraki adımda geçerli olan yerel saatin dönebilmesi için gerekli olan komut yazıldı. Daha sonra ‘3.Havalimanı’, ‘istanbulhavalimanı’ ve ‘İstanbul3.Havalimanı’ arama terimlerini yazarak 2014 yılının 1 Ocak tarihinden itibaren günümüze kadar atılan tweetleri çekebilmesi için gerekli olan işlem tamamlandı.

Şekil 10. Tweetlerin çekilmesi ve yazdırılması işlemi

Arama Terimleri İçin Toplam Tweet Sayısı

```
[ ] os.system(f"snsrape --since {from_date} twitter-search '{search_term} until:{end_date}' > result-tweets.txt")
    if os.stat("result-tweets.txt").st_size == 0:
        counter = 0
    else:
        df = pd.read_csv('result-tweets.txt', names=['link'])
        counter = df.size

    print('Number Of Tweets : ' + str(counter))
```

Number Of Tweets : 105260

Çekilen Tweetler İçerisinden 105260 adet Tweet'in Çekilmesi

```
[ ] max_results = 105260
```

Kullanıcılardan Tweet Çıkarma

```
[ ] extracted_tweets = "snsrape --format '{content!r}'+ f" --max-results {max_results}
    --since {from_date} twitter-search '{search_term} until:{end_date}' > extracted-tweets.txt"
    os.system(extracted_tweets)
    if os.stat("extracted-tweets.txt").st_size == 0:
        print('No Tweets found')
    else:
        df = pd.read_csv('extracted-tweets.txt', names=['content'])
        for row in df['content'].iteritems():
            print(row)

[ ] #Kullanacağım veri setini Csv formatında indirdim.
    df.to_csv("havalimanıtweets.csv")
```

Şekil 10'da ise ilk adımda yazmış olduğumuz arama terimlerini sadece içerik olarak tek bir sütun şeklinde çekme işleme gerçekleştirildi. Bir sonraki adımda çekilen tweetlerin ne kadarının sonuç olarak yazdırılması gerektiğiyle ilgili karar verdik. Çekmeye karar verdiğimiz maksimum sonuç, ilk olarak Şekil 9'da görülen 'string' kütüphanesi import ettiğimiz için Google Colab üzerinde 'txt' olarak yazdırıldı. Yazdırılan txt dosyasının ise 'csv' formatında çıktısı alındı.

4.1.1. Veri Seti Hakkında

Şekil 11’de görüldüğü gibi ‘İstanbul Havalimanı’ hakkında 2014-01-01 tarihinden itibaren günümüze kadar atılan tweetlerden, toplamda 105.200 adet tweet yazdırılmıştır. Veri setinde sadece 105.200 adet satır ve ‘content’ isimli içerisinde metinlerin yer aldığı bir öznitelik bulunmaktadır. Çekilen tweetleri ‘csv’ formatında alabilmek için ‘pandas’ kütüphanesi import edildi ve ardından ‘to_csv’ komutu ile ‘havalimanıtweets’ isimli veri setini kaydettik. Kaydedilen veri seti data isimli değişkene atayıp yazdırıldı.

Şekil 11. Veri seti

```
#csv dosyasını data isimli bir değişkene tanımladık.
data = pd.read_csv('/content/havalimanıtweets.csv')

data
```

	content
0	'İstanbul Havalimanı dünyanın en yoğun ikinci ...
1	'@tokcem - 3.havalimanı etrafına yapılaşma var...
2	'Gayrettepe İstanbul (3. Havalimanı)\nHalkalı ...
3	'@_mucahide007_ @Mrt_Ongun Bundan başka argüma...
4	'Atatürk Havalimanı yıkımı yargılanabilir!\n- ...
...	...
105195	'3.Havalimanı
105196	3.Havalimanı, 3.Köprü, Marmaray, Gebze Köprüsü...
105197	'@barisckmk @turkceder size rağmen 3. havalima...
105198	'@nihatkasikara @erdalddoglu 3.havalimani yapı...
105199	'@EnesBars sen 25 milyar avro 3.havalimanı iha...

105200 rows × 1 columns

```
data.shape
```

```
(105200, 1)
```

4.2. Veri Ön İşleme

Twitter'dan elde edilen veriler genellikle karmaşık bir yapıda oluyor. Bunun nedeni kullanıcıların büyük küçük harf uyumuna uymaması, kelime tekrar etmesi, gereksiz kelime veya simgeler kullanması gibi sorunlarla karşı karşıya kalınıyor. Ayrıca verinin analiz sürecinde noktalama işaretlerinin ve sayıların bulunması, verinin kirli ve aykırı değerlere sahip olmasına neden olur. Bu da analiz sonucuna etki edebiliyor. Tam bu noktada verinin temiz bir şekilde analiz edilebilmesi için 'pandas' kütüphanesinden yararlanarak veri, birtakım veri ön işleme adımlarından geçiyor.

4.2.1. Metinlerin Küçük Harfe Çevrilmesi

Veri ön işleme aşamasında, veri setinde normalleştirme yapmamız gerekiyor. Çünkü string(str) değerlerde 'istanbul' veya 'İSTANBUL' kelimesi iki farklı değer olarak algılanır. Bunun önüne geçebilmemiz için metinler 'lower' komutu ile büyük harften küçük harfe çevrilir.

Şekil 12. Tweetlerin küçük harfe çevrilmesi

Tweet'lerin küçük harflere çevrilmesi

```
[ ] data['content'] = data['content'].apply(lambda x: " ".join(x.lower() for x in x.split()))
```

```
[ ] data.head()
```

	content
0	'istanbul havalimanı dünyanın en yoğun ikinci...
1	'@tokcem - 3.havalimanı etrafına yapılaşma var...
2	'gayrettepe istanbul (3. havalimanı)\nhalkalı...
3	'@_mucahide007_ @mrt_ongun bundan başka argüma...
4	'atatürk havalimanı yıkımı yargılanabilir!\n- ...

4.2.2. Metinlerdeki Gereksiz Karakterlerin Temizlenmesi

Veri seti içerisindeki 'http', 'hashtag(#)', '@' işareti ve 'gt' gibi o an içerisinde yazılan gereksiz karakterler, veri seti içerisinde silinerek temizle işlemi gerçekleştirildi.

Şekil 13. Gereksiz karakterlerin silinmesi

```
[ ] data['content'] = data['content'].str.replace(r'http\S+', '')

[ ] data['content'] = data['content'].str.replace(r'@\S+', '')

[ ] data['content'] = data['content'].str.replace(r'#\S+', '')

[ ] data['content'] = data['content'].str.replace('gt', '')

[ ] data['content'].head()

0    'istanbul havalimanı dünyanın en yoğun ikinci...
1    ' - 3.havalimanı etrafına yapılaşma var mı ?'
2    'gayrettepe istanbul (3. havalimanı)\nhalkalı...
3    ' bundan başka argümanınız var mı öncesi ve s...
4    'atatürk havalimanı yıkımı yargılanabilir!\n- ...
Name: content, dtype: object
```

4.2.3. Metinlerdeki Noktalama İşaretlerinin ve Sayıların Temizlenmesi

Veri setini, noktalama işaretlerinden temizlememiz gerekmektedir. Aynı zamanda string ifadelerde tarih, saat veya herhangi bir sayının bulunması gereksizdir. Bu nedenle veri setinde bulunan sayılarda temizlenmiştir.

Şekil 14. Noktalama işaretlerinin ve sayıların temizlenmesi

```
[ ] data['content'] = data['content'].str.replace('[^\w\s]', '')

[ ] data['content'] = data['content'].str.replace('\d', '')

[ ] data.head()
```

	content
0	istanbul havalimanı dünyanın en yoğun ikinci h...
1	havalimanı etrafına yapılaşma var mı
2	gayrettepe istanbul havalimanınınhalkalı istanb...
3	bundan başka argümanınız var mı öncesi ve so...
4	atatürk havalimanı yıkımı yargılanabilirn atat...

4.2.4. Metinlerdeki Stop Words'ün Temizlenmesi

Stop Words kelimeler, duraklama kelimeleri olarak da bilinmektedir. Veri ön işleme adımlarında, veri setinde geçen bu duraklama kelimeler örneğin, 'eğer' kelimesi iyi anlamda bir değer ifade etmediği için bu kelimenin daha çok negatif anlamda bir değeri vardır. İşte bu nedenden dolayı stop words kelimelerin veri setinden çıkarılması daha doğru bir işlem olur.

Şekil 15. Stop Words kaldırılması

Stop wordslerin çıkartılması

```
[ ] #NLTK kütüphanesi kurma işlemi
!pip install nltk

Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (3.2.5)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from nltk) (1.15.0)

[ ] #NLTK kütüphanesini içe aktarma
import nltk

[ ] #NLTK'dan durdurma sözcükleri indirme
nltk.download("stopwords")

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
True

[ ] #stopwords için gerekli olan kütüphane import edildi.
from nltk.corpus import stopwords

[ ] #Türkçe stopwords dolgu sözcüklerini sw adından bir değişkene tanımlıyoruz.
sw = stopwords.words("turkish")

[ ] sw

['acaba',
 'ama',
 'aslında',
 'az',
 'bazı',
 'belki',
 'biri',
 'birkaç',
 'birsey',
 'biz',
 'bu',
 'çok',
 'çünkü',
 'da',
 'daha',
 ...]

[ ] #stopwords sözcüklerini data isimli veri içerisinde ayırma işlemi yapıldı.
data['content'] = data['content'].apply(lambda x: " ".join(x for x in x.split() if x not in sw))
```

4.2.5. Az Geçen Kelimelerin ve Boşlukların Kaldırılması

Veri setinde, az geçen kelimelerin kaldırılma işlemi gerçekleştirildi. Bu kelimelerin az olması hiçbir sonuca etki etmeyeceği için performansta bir değişiklik olmayacaktır. İsteğe bağlı olarak kaldırılabilir. Tüm bu işlemler yapıldıktan sonra 'dropna(inplace=True)' komutu yazılarak veri seti içerisinde yer alan boş satırlar kaldırıldı.

Şekil 16. Az geçen kelimelerin ve boşlukların kaldırılması

En az geçen 10 kelimenin incelenmesi

```
[ ] kelime_sil = pd.Series(" ".join(data['content']).split()).value_counts()[-10:]
```

```
[ ] kelime_sil
```

ıstanbulla	1
edilmiş	1
akepenn	1
övgüsünü	1
zannedip	1
zahmetno	1
görürseniz	1
istemesem	1
edindik	1
ıslarına	1

dtype: int64

```
[ ] data['content'] = data['content'].apply(lambda x: " ".join(x for x in x.split() if x not in kelime_sil))
```

```
[ ] pd.Series(" ".join(data['content']).split()).value_counts()[-10:]
```

ediyormuşnyani	1
dıbinde	1
cehennn	1
gıt	1
havaalnını	1
kalıcazmis	1
beklemiyoruz	1
ndoğrudur	1
yaptınmade	1
bızım	1

dtype: int64

```
[ ] data.dropna(inplace=True)
```

4.3. Makine Öğrenmesi Veri Seti Hazırlama

Büyük veri setlerinde, metin sayısı artıkcça sınıflandırma yapabilmek zorlaşıyor. Bu çalışmadaki veri setinde ise bir makine öğrenmesi seti hazırlamamız gerekiyor. Bu seti hazırlarken belirlememiz gereken ilk şey İstanbul Havalimanı'nı hakkında atılan tweetlerin tam olarak hangi konuları içerdiğiyle ilgili bir karar vermek. 3 kategori belirlenip, her kategori için 250'şer tweet, toplamda 750 adet tweet ile kategori etiketlemesi yapılmıştır. Bu çalışmada da belirlenen kategoriler ise aşağıdaki gibidir.

- **Ekonomi:** İstanbul Havalimanı projesi için harcanan bütçe, bu projenin ülkemize getireceği ekonomik kazançları, kayıpları ve diğer ekonomik ifadeleri içeren tweetler alınmıştır.
- **Siyaset:** İstanbul Havalimanı projesi için siyasi söylem içeren tweetler alınmıştır.
- **Çevre:** İstanbul Havalimanı projesinin çevreye verdiği değişim, doğal yaşam alanı, projenin yapıldığı bölgenin iklim koşulları hakkında atılan tweetler alınmıştır.

Şekil 17. Makine Öğrenmesi seti

	A	B	C	D	E	F	G	H	I	J	K	L
1	topic	text										
2	ekonomi	istanbulHavalimanı yapılırken ne gereği Türkiye ekonomisine büyük bir yük yükleyorsunuz. Atıl bir yatırım										
3	ekonomi	Ülkenin sosyoekonomik durumu, insanı #istanbulhavalimanı 'nın gösterişli yapısı ve çöken çatısından farksız..										
4	ekonomi	yavo şu İstanbul havaalanını yolcuları karton dağıtıldı hani uzaya gidecektik hani dünyanın 5 ekonomisiz hani Avrupa bizi k										
5	ekonomi	@ekrem_imamoglu atatürk havalimanı 3. Havalimanı tartışmasının ekonomik temeli nedir? Ben 3. Havalimanının gerekliliğ										
6	ekonomi	Dünya'nın en iyi 3.Havalimanı seçilen ATATÜRK havalimanı #SuriyeliPolisOlamaz Sevda Demirel Metroda Nasuhi #dolar #b										
7	ekonomi	İstanbul Havalimanı dünyanın en yoğun ikinci havalimanı oldun https://t.co/dzH1AqdAor #istanbulhavalimanı https://t.c										
8	ekonomi	İstanbul havalimanlarından uçan yolcu sayısı 22 milyon arttı'nAvrupa'da geçen yıl hava yoluyla seyahat edenlerin yüzde 7'										
9	ekonomi	Avrupa ve ABD dız çöktürmüş dolar hamlesiyle zamların sebebi İmamoğlu ve kk 3. Havalimanı ile A10 ekonomisine zarar vı										
10	ekonomi	Tamam Engin Sarsmaz 2002'den sonra özelleştirmeler nere gitti ne olur örnek ver. Allah şahit görmek istiyorum. Hadi buy										
	A	B	C	D	E	F	G	H	I	J	K	L
252	siyaset	her zaman yaptığı aymazlığı yine yapıyor! Devasa Ak Parti projeleri üstüne heyula gibi çöküyor! Bay Ekrem kendine gel zeka										
253	siyaset	Ak parti hükümeti ve belediyesi zamanında yapıları açtınız.Ak parti hükümetine teşekkür yok mu?nKendinizi aldatmayın										
254	siyaset	gider ve belki başka bir parti adı altında o gelr'										
255	siyaset	Ak parti icraatlarına sahip çıkmalı										
256	siyaset	Ovhalde Ak Parti'nin yaptığı güzel işleri kiskanalar için de geçerli mi hocam bu. 3. Havalimanı 3. Köprü Çanakkale Köprüsü										
257	siyaset	Allah Ak Parti den razı olsun. IBB nin tüm işini o yapıyor resmen. \n#istanbulhavalimanı #thy https://t.co/9DemquTP6m '										
258	siyaset	Yazıklar olsun IBB hiçbir işe yaramıyorsunuz \nNe varsa AK Parti de var\n\n #istanbulhavalimanı #thy										
259	siyaset	İBB'den hayır yok yine AK Parti geçti işin başına #istanbulhavalimanı #thy https://t.co/dP7Ese9bd1										
260	siyaset	#istanbulhavalimanı nı açamadınız kalkmış İstanbul'u kurtarmak dan bahsediyorsunuz , #kgm ilk saatlerde kılını kımıldatma										
261	siyaset	CHP İzmir Milletvekili Tuncay Özkan, Ayşenur Ulus doğru diyor, başörtüsü sorununu CHP çözdü! Avrasya Tüneli CHP'nin prc										
	A	B	C	D	E	F	G	H	I	J	K	L
502	cevre	3. Havalimanı doğaya zarar vererek										
503	cevre	Doğa ile savaşılmaz.Mühendisler doğa ile anlaşıp.. YargılanacakSınız! #istanbulhavalimanı										
504	cevre	Doğayı mahvederek										
505	cevre	Atasözü düşmanlarına gelsin; "Doğa ile inatlaşan altına yapar!" #istanbulhavalimanı #istanbuldakar #istanbulairport										
506	cevre	Doğa tahribatı										
507	cevre	Kar berekettir hep olumsuzluklara bakmayın doğanın gücüne hayranlıkla bakın ..Zorlukların güzelliği getirdiğini hepimiz biliy										
508	cevre	Okumuş insanları dinlemezsek doğa son sözünü böyle söyler ve dinletir :) #kayağıyor #istanbulkar #istanbulhavalimanı ht										
509	cevre	Tarih 14 Eylül 2018: 3. Havalimanı şantiyesinde yer yerinden oynadı. O gün inşaat işçileri ..."niktidar-sermaye-siyaset çet										
510	cevre	Doğa ile inatlaşma olmaz. Ne önlem alırsanız alın										
511	cevre	atatürk Havalimanı'ndan kalkan uçak 3. Havalimanı'ndan neden kalkmıyor? Doğada nelere mal oldu? Yer seçimi neden i										

Şekil 18’de gösterilen makine öğrenmesi veri seti, Google Colab’a yazdırılma işleminde Türkçe karakter hatası verdiğiinden dolayı online belge dönüştürücü olan bir programda csv olarak kaydedilerek sorun halledildi. Ayrıca oluşturduğumuz bu veri setinde, veri ön işleme adımlarını uyguladık. Bunlar büyük harflerin küçük harflere çevrilmesi, http bağlantılarının, hashtag sembollerinin(#), @ işaretlerinin, noktalama işaretlerinin ve sayıların temizlenmesi gibi işlemler kapsar. Kullanılacak olan makine öğrenmesi setinin son hali ise Şekil 18’de görülmektedir.

Şekil 18. Makine Öğrenmesi seti son hali

	A	B	C	D	E	F	G	H	I	J	K	L
1	,topic;text											
2	0,ekonomi,istanbulhavalimanı yapılıırken gereği tırkıye ekonomisine bıkıyık bir yıkıklıyosunuz atı bir yatır											
3	1,ekonomi,İlkenin sosyoekonomik durumu insanın nın gısteriYli yapısız şİken şİatınsından farkı											
4	2,ekonomi,yavo istanbul havaalanı yolcuları karton daYııtlı hani uzaya gidecektik hani dınyanın ekonomisiz hani											
5	3,ekonomi,atatürk havalimanı havalimanı tartıYmasınn ekonomik temeli nedir ben havalimanın gerekliliYine dai											
6	4,ekonomi,dınyanın iyi havalimanı seİilen atatürk havalimanı sevda demirel metroda nasuhi erbakan yeni şİşanakkale k											
7	5,ekonomi,istanbul havalimanı dınyanın yoYun ikinci havalimanı oldun											
8	6,ekonomi,istanbul havalimanlarından uşan yolcu sayısız milyon arttınavrupada geşen yı hava yoluyla seyahat edenlerin											
9	7,ekonomi,avrupa abd dıı şİoktYrmıY dolar hamlesiyle zamları sebebi imamoYlu kk havalimanı a ekonomisine zarar ve											
10	8,ekonomi,tamam engin sarsmaz den sonra şİzelleYtirmeler nere gitti olur şİrnek ver allah İYahit gırmek istiyorum hadi buyur t											
11	9,ekonomi,kaan kimse İretimin karıYınsında deYil keYke dolarla mıYıYteri garantili yollar kıprıYler yerine verdiYimiz ver											
	A	B	C	D	E	F	G	H	I	J	K	L
252	250,siyaset,herzaman yaptıYı aymazıYı yine yapıyor devasa ak parti projeleri İstıne heyula şİİkıYor bay ekrem											
253	251,siyaset,ak parti hıYkmeti belediyesi zamanında yapılanı aşıtınsınzak parti hıYkmetine teYekkıYr yok munkendini											
254	252,siyaset,gider baYka bir parti adı altında gelr											
255	253,siyaset,ak parti sahip											
256	254,siyaset,ak partinin yaptıYı gızel iYleri kiskanalar geşerli mi hocam havalimanı kıprı şİşanakkale kıprııY											
257	255,siyaset,allah ak parti den razı olsun ibb nin iYini yapıyor resmen n											
258	256,siyaset,yazıklar olsun ibb hişbir iYi varsa ak parti varnn											
259	257,siyaset,yok yine ak parti geİti iYin baYına											
260	258,siyaset,nı kalkmıYı istanbulu kurtarmak dan bahsediyorsunuz ilk saatlerde kılını kımılatmadı sırf ibb zorda kal											
261	259,siyaset,chp izmir milletvekili tuncay İzkan ayıYenur ulus doYru diyor baYıİrtııY sorununu chp şİİııY avrasya tıne											
	A	B	C	D	E	F	G	H	I	J	K	L
502	500,cevre,havalimanı doYaya zarar vererek											
503	501,cevre,doYı savaYıımazmıhendisler doYı anıaıYıYargılanacaksın											
504	502,cevre,doYıYı mahvederek											
505	503,cevre,atasıY dıYımanlarına gelsin doYı inatlaYan altına yapar											
506	504,cevre,doYı tahribatı											
507	505,cevre,kar berekettir olumsuzluklara bakmayn doYıan gıYıne hayranla kla bakın zorlukları gızellik getirdiYini t											
508	506,cevre,okumuY insanlar dınelemesek doYı son sııZıYnı bıYle sıYler dinletir											
509	507,cevre,tarih eYıı havalimanı İYantiyesinde yer yerinden oynadı gıYn inıYaat iYİşileri niktidarsermayesiyaset şİsetelerini											
510	508,cevre,doYı inatlaYma olmaz İİnlem alırsanıı alı											

Şekil 19’da görüldüğü gibi oluşturduğumuz veri seti 750 satır ve 2 öznitelikten oluşmaktadır. Veri setinde grupta yaparak veri etiketlerinin 0, 1, 2 olarak sayısallaştırma işlemi gerçekleştirildi.

Şekil 19. Veri etiketinin sayısallaştırılması

```
[ ] #ml değişkeninin satır ve sütun sayısını göstermesi için ml.shape yaptık.  
ml.shape
```

```
(750, 2)
```

```
[ ] #ml verisinin içerisinde bulunan topic özniteliği için grupta yaptık.  
ml.groupby("topic").size()
```

```
topic  
cevre      250  
ekonomi    250  
siyaset    250  
dtype: int64
```

Veri Etiketlerinin Sayısallaştırılması

```
[ ] ml['labels'] = pd.factorize(ml.topic)[0]
```

```
[ ] #Topic özniteliğine 0,1,2 olarak etiket verdik.  
ml.groupby(["topic", "labels"]).size()
```

```
topic  labels  
cevre   2      250  
ekonomi 0      250  
siyaset 1      250  
dtype: int64
```

4.4. Makine Öğrenmesi İçin Kullanılacak Algoritmalar

Kategorilere ayırdığımız veri setinin başarımını ölçebilmek için 6 farklı makine öğrenmesi algoritması kullanılacaktır. Bu algoritmalar ise Navie Bayes, Logistic Regression, SVM, Random Forest, Gradient Boosting ve XGB algoritmalarıdır. Kullanılan bu algoritmalarla daha sonra tüm veriyi kapsayan kategori tahmini yapılacaktır.

- **Navie Bayes Algoritması:** Bu algoritma aslında tembel olarak bilinen bir sınıflandırma algoritmasıdır. Örneğin bir cismin rengine, boyuna veya çapına göre olan özelliklerini birbiriyle alakası olmasa bile bu özellikleri dikkate alarak işlem yapar. Çok karmaşık olan veri setlerinde elde ettiği yüksek başarımlar ile bilinen bir algoritmadır. (Hatipoğlu, 2018)
- **Logistic Regression:** Bir sınıflandırma algoritmasıdır. Bir görevi tahmin ederken olumlu veya olumsuz yönden tahmin eder. Tahminin 0-1 arasında olduğu durumlarda kullanılır. Örneğin, bizim yaptığımız çalışmada tweetlerin duygularını tahmin ederken olumlu veya olumsuz yönden kategorik bir şekilde tahmin yapar. (Agrawal, 2017)
- **Support Vector Machine:** SVM, veri seti içerisindeki her bir özelliği n boyutlu bir uzayda nokta çizer. N boyut diye tabir ettiğimiz ise veri setinde bulunun özelliklerdir. Bizim yaptığımız çalışmada duygu analizlerinin olumlu veya olumsuz yönden 2 parçaya ayırma işlemi yapılır. Olumlu veya olumsuz yönden bulunduğu doğrulara en yakın olanı bulmaya çalışır. (Pupalı, 2018)
- **Random Forest:** İçerisinde karar ağaçlarını barındıran bir tür sınıflandırma algoritmasıdır. Örneğin arkadaşlarınızla bir ders seçmeniz gerekiyor. Dersleri değerlendirirken akts, saat, ders tipi gibi özellikleri bakımından değerlendirerek bir seçim yapmanız gerekiyor. Herkes kendine göre uygunu seçtikten sonra en iyiyi bulabilmek için en çok tercih edilen dersi seçebilmektir. Rastgele ormanı anlatan en iyi yol budur. (Şimşek, 2018)
- **Gradient Boosting:** Çok fazla ve karmaşık verilerde oldukça güçlü tahminler yapmaya yarayan bir algoritmadır. Tek bir tahmindeki oranı artırmak için geçmişte yapılan tahminlerden yararlanır. (Kumar, 2020)
- **XGBoost:** XGB aslında Gradient Boosting algoritmasının daha da geliştirilmiş hali olan yeni nesil bir algoritmadır. Algoritmanın en büyük avantajlarından birisi Gradient Boosting'te olduğu gibi yüksek tahmin gücüne dayalıdır. Aynı zamanda overfitting olarak adlandırılan aşırı öğrenme durumunun da önüne geçer. (Martinez, 2019)

4.5. Modelin Eğitilmesi ve Uygulanması

Hazırlamış olduğumuz makine öğrenmesi veri setini, eğitilmesi için 'model_data' isimli bir değişkene atama işlemi gerçekleştirdik. Daha sonra 'sklearn' kütüphanesini import edildi. Hazırladığımız veri setinin %80'i eğitim, %20'si ise test verisi olarak ayrılmıştır.

Şekil 20. Eğitim ve Test verisi

```
Model ve Test Fonksiyonlarının Oluşturulması

[ ] #test için gerekli olan modeli oluşturmak için model_data isimli değişken atadık.
    model_data = ml[["text", "labels"]]

[ ] #test için gerekli olan kütüphaneler import edildi. Ayrıca test için %20 , eğitim verisi için %80 olarak ikiye ayrıldı.
    from sklearn.model_selection import train_test_split

    X_train, X_test, y_train, y_test = train_test_split(model_data["text"],
    model_data["labels"], test_size = 0.2, random_state = 4)
```

Hazırladığımız makine öğrenmesi veri setinde string değerler olduğu için veri seti sayısallaştırılmıştır.

Şekil 21. Metnin sayısallaştırılması

Metnin Sayısallaştırılması

```
#Metin sayısallaştırılması için gerekli olan kütüphane import edildi.
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer()

train_vectors = vectorizer.fit_transform(X_train)
test_vectors = vectorizer.transform(X_test)
print(train_vectors.shape, test_vectors.shape)

(600, 4801) (150, 4801)

print(train_vectors)

(0, 1793)      0.0970512964544771
(0, 2441)      0.39459836824281136
(0, 729)       0.32087648697541515
```


Oluşturduğumuz makine öğrenmesi veri seti için kullanılacak olan algoritmaların kütüphaneleri import edildi. Model oluşturup sınıflandırmaya tabii tutulan algoritmaların performanslarını değerlendirmek amacıyla Türkçe karşılığı ‘karışıklık matrisi’ olarak bilinen ‘confusion matrix’ ve algoritmaların doğruluk puanları alabilmek için ‘accuracy score’ metrikleri kullanıldı. Kullandığımız algoritmaların doğruluk puanları ekrana yazdırıldı.

Şekil 22. Algoritmaların doğruluk puanları

```
[ ] #Modellemeler için gerekli olan kütüphaneler import edildi.
    from sklearn.metrics import classification_report
    from sklearn.metrics import confusion_matrix
    from sklearn.metrics import accuracy_score
    from sklearn.linear_model import LogisticRegression
    from sklearn.svm import SVC
    from sklearn.naive_bayes import MultinomialNB
    from xgboost import XGBClassifier
    from sklearn.ensemble import RandomForestClassifier
    from sklearn.ensemble import GradientBoostingClassifier
    from sklearn.model_selection import cross_val_score
    import seaborn as sns

[ ] Accuracy for Naive Bayes: mean: 0.86 2sd: 0.03
    Scores:: [0.85833333 0.89166667 0.84166667 0.85833333 0.86666667]

    Accuracy for Logistic Regression: mean: 0.88 2sd: 0.02
    Scores:: [0.86666667 0.89166667 0.85833333 0.875      0.88333333]

    Accuracy for SVC: mean: 0.88 2sd: 0.03
    Scores:: [0.86666667 0.90833333 0.85833333 0.86666667 0.875      ]

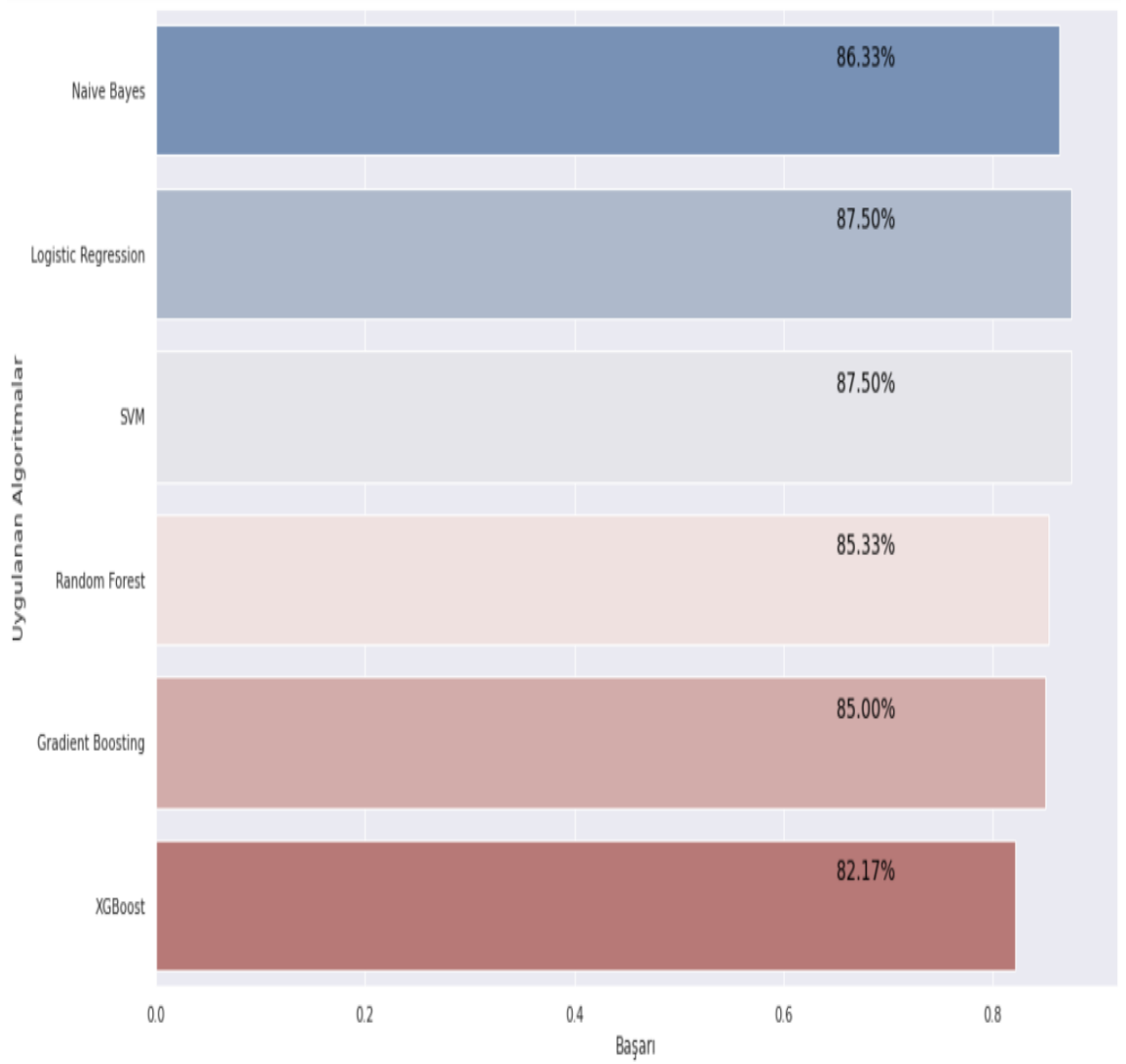
    Accuracy for Random Forest: mean: 0.85 2sd: 0.08
    Scores:: [0.80833333 0.90833333 0.88333333 0.86666667 0.8       ]

    Accuracy for Gradient Boosting: mean: 0.85 2sd: 0.03
    Scores:: [0.86666667 0.85833333 0.85833333 0.85      0.81666667]

    Accuracy for Xgboost: mean: 0.82 2sd: 0.03
    Scores:: [0.81666667 0.84166667 0.83333333 0.81666667 0.8       ]
```

Navie Bayes, Logistic Regression, SVM, Random Forest, Gardient Boosting ve XGBoost algoritmalarını öğretilip, yüksek başarımlı sağlamaya çalıştık. Genel anlamda iyi bir başarımlı sağladığımız söylenebilir. En iyi sonucu veren 3 algoritma sırasıyla Logistic Regression, %87.50 SVM, %87.50 Navie Bayes algoritması ise %86.33'lük bir başarımlı sağlamıştır.

Şekil 23. Algoritma sonuçlarının yazdırılması



En iyi sonuçları veren 3 farklı makine öğrenmesi algoritması ile tüm veri setini kapsayan bir kategori etiketlemesi yapılacaktır. Veri seti içerisindeki tweetler sayısallaştırılmıştır. Ardından ekonomi, siyaset ve çevre kategorileri için bir model oluşturulmuştur.

Şekil 24. Kategori Modelleme

```
[ ] test_vectors_ = vectorizer.transform(data["content"].astype('U').values)
    print(test_vectors_.shape)

(103108, 4801)
```

```
[ ] print(test_vectors_)

(0, 4361)      0.48723910240066226
(0, 2982)      0.5105529658788056
(0, 2074)      0.2564513339499233
```

```
[ ] predicted = clf.predict(test_vectors_)
    tahmin = pd.DataFrame(predicted)
    tahmin.rename(columns = {0:'tahmin'}, inplace = True)
    data["tahmin_naive_bayes"] = tahmin
```

```
[ ] data.loc[data['tahmin_naive_bayes'] == 0, ['tahmin_category_nb']] = 'ekonomi'
    data.loc[data['tahmin_naive_bayes'] == 1, ['tahmin_category_nb']] = 'siyaset'
    data.loc[data['tahmin_naive_bayes'] == 2, ['tahmin_category_nb']] = 'çevre'
```

```
[ ] predicted = LogicReg.predict(test_vectors_)
    tahmin = pd.DataFrame(predicted)
    tahmin.rename(columns = {0:'tahmin'}, inplace = True)
    data["tahmin_logistic"] = tahmin
```

```
[ ] data.loc[data['tahmin_logistic'] == 0, ['tahmin_category_logistic']] = 'ekonomi'
    data.loc[data['tahmin_logistic'] == 1, ['tahmin_category_logistic']] = 'siyaset'
    data.loc[data['tahmin_logistic'] == 2, ['tahmin_category_logistic']] = 'çevre'
```

```
[ ] predicted = svm.predict(test_vectors_)
    tahmin = pd.DataFrame(predicted)
    tahmin.rename(columns = {0:'tahmin'}, inplace = True)
    data["tahmin_svm"] = tahmin
```

```
[ ] data.loc[data['tahmin_svm'] == 0, ['tahmin_category_svm']] = 'ekonomi'
    data.loc[data['tahmin_svm'] == 1, ['tahmin_category_svm']] = 'siyaset'
    data.loc[data['tahmin_svm'] == 2, ['tahmin_category_svm']] = 'çevre'
```

Eğitilen makine öğrenmesi algoritmalarından en yüksek başarıyı sağlayan Navie Bayes, Logistic Regression, Support Vector Machine algoritmaları, veri ön işleme adımlarından sonra elde edilen 103.108 adet veriye de uygulanmıştır. Algoritmalar tarafından tweetlere, ekonomi, siyaset ve çevre yönünden etiketleme yapıldı. Yapılan etiketlemeler sonucunda, algoritma tahminleri karşılaştırılmıştır.

Veri görselleştirme bölümünde bu algoritmaların kategori etiketlemesi yaparken ekonomi, siyaset ve çevre yönünden nasıl bir etiketleme yaptığını, sayısal olarak da göreceğiz. Şekil 25'te ise tüm verilere uygulanan kategori etiketlemesinin ilk 15 verisi, ekrana yazdırıldı.

Şekil 25. Algoritma karşılaştırmaları

```
data.head(15)
```

	content	tahmin_naive_bayes	tahmin_category_nb	tahmin_logistic	tahmin_category_logistic	tahmin_svm	tahmin_category_svm
0	istanbul havalimanı dünyanın yoğun ikinci hava...	0	ekonomi	0	ekonomi	0	ekonomi
1	havalimanı etrafına yapılaşma var	2	çevre	2	çevre	2	çevre
2	gayrettepe istanbul havalimanınınhalkalı istanbu...	0	ekonomi	0	ekonomi	0	ekonomi
3	bundan başka argümanınız var öncesi sonrasıyla...	2	çevre	2	çevre	2	çevre
4	atatürk havalimanı yıkımı yargılanabilirn atat...	2	çevre	0	ekonomi	2	çevre
5	chp izmir milletvekili tuncay özkan ayşenur ul...	1	siyaset	1	siyaset	1	siyaset
6	istanbul havalimanından yolcularına yeni hizmet	0	ekonomi	0	ekonomi	0	ekonomi
7	değil yıl önce yaşam dolu bir kuzey ormanları ...	2	çevre	2	çevre	2	çevre
8	son dönemlerde reklam işine ağırlık verdi basi...	2	çevre	2	çevre	2	çevre
9	hele havalimanı muhteşemlik muazzam ellerine s...	1	siyaset	1	siyaset	1	siyaset
10	havalimanının olduğu bölgede köy hayvanlarına ...	2	çevre	2	çevre	2	çevre
11	uçak inemez dedikleri istanbul havalimanı yılında	2	çevre	2	çevre	2	çevre
12	gözüünün icine sokayım mi yavuz selim sultan k...	1	siyaset	1	siyaset	1	siyaset
13	mert bey havalimanı çevresinden mi ev almak ma...	1	siyaset	1	siyaset	1	siyaset
14	yılında avrupadaki havalimanları arasında yolc...	0	ekonomi	0	ekonomi	0	ekonomi

4.6. Duygu Analizinin Uygulanması

Duygu analizinin yapılması için, data isimli mevcut veri setimize 'data_new' ismini veriyoruz. Duygu analizi sonucu, data_new isimli değişken üzerine yazılacaktır. Sonraki adımda ise, duygu analizi için gerekli olan model kurgulandı. Transformers kütüphanesi import edildi.

Şekil 26. Duygu Analizi ilk adım

```
[ ] data_new = pd.DataFrame(data["content"])
```

```
data_new.shape
```

```
(103108, 1)
```

```
[ ] data_new["content"] = data_new["content"].apply(lambda r: str(r))
```

```
[ ] #Transformers yükleme  
!pip install transformers
```

```
Collecting transformers
```

```
Downloading transformers-4.18.0-py3-none-any.whl (4.0 MB)
```

```
|████████████████████████████████████████| 4.0 MB 7.5 MB/s
```

```
Collecting huggingface-hub<1.0,>=0.1.0
```

```
Downloading huggingface_hub-0.5.1-py3-none-any.whl (77 kB)
```

```
|████████████████████████████████████████| 77 kB 7.6 MB/s
```

```
[ ] #Gerekli olan kütüphane import edildi. olumlu olumsuz metinleri sınıflandırmak için bir model kuruldu.  
from transformers import AutoModelForSequenceClassification, AutoTokenizer, pipeline
```

```
model = AutoModelForSequenceClassification.from_pretrained("savasy/bert-base-turkish-sentiment-cased")
```

```
tokenizer = AutoTokenizer.from_pretrained("savasy/bert-base-turkish-sentiment-cased")
```

```
sa = pipeline("sentiment-analysis", tokenizer = tokenizer, model = model)
```

```
Downloading: 100% ██████████ 596/596 [00:00<00:00, 3.85kB/s]
```

```
Downloading: 100% ██████████ 422M/422M [00:18<00:00, 52.9MB/s]
```

```
Downloading: 100% ██████████ 39.0/39.0 [00:00<00:00, 1.08kB/s]
```

```
Downloading: 100% ██████████ 256k/256k [00:00<00:00, 633kB/s]
```

```
Downloading: 100% ██████████ 112/112 [00:00<00:00, 2.60kB/s]
```

Şekil 27’ de ‘sentiment_list’ adında bir liste oluşturulup tüm veri setinin, duygu skorlarını olumlu veya olumsuz yönden analizi gerçekleştirildi. Bu işlem yaklaşık 3 saat sürdü. Bir sonraki adımda 95.000’inci satırın duygu skoru ekrana yazdırıldı. Daha sonra csv olarak kaydedilmek için ‘spredict_list’ içerisinde çıktı sonuçlarını yazdırdık. Data_new veri setine, ‘label’ ve ‘score’ sonuçları birleştirildi. Bu işlemin sonucunda ‘duyguanalizicikti’ ismiyle csv çıktısı alınarak görselleştirilme aşamasına geçildi.

Şekil 27. Duygu Analizinin yapılması

```
[ ] #Sentiment_list olarak bir liste oluşturuldu.
#listenin içine olumlu olumsuz duyguları ve skorlarını yazdırma işlemi gerçekleştirildi.
sentiment_list = []
for i in data_new["content"]:
    sentiment_list.append(sa(i))

[ ] #sentiment listesini yazdırdık.
sentiment_list

[[{'label': 'positive', 'score': 0.9524326920509338},
 {'label': 'negative', 'score': 0.9971930384635925},
 {'label': 'positive', 'score': 0.513373613357544},
 {'label': 'negative', 'score': 0.9987950325012207},
 {'label': 'negative', 'score': 0.8606646060943604},
 {'label': 'negative', 'score': 0.9515908360481262},
 {'label': 'positive', 'score': 0.5678892135620117},
 {'label': 'positive', 'score': 0.6107219457626343},
 {'label': 'positive', 'score': 0.8064418407891833},
 {'label': 'negative', 'score': 0.9987950325012207},
 {'label': 'negative', 'score': 0.8606646060943604},
 {'label': 'negative', 'score': 0.9515908360481262},
 {'label': 'positive', 'score': 0.5678892135620117},
 {'label': 'negative', 'score': 0.9987950325012207},
 {'label': 'positive', 'score': 0.513373613357544},
 {'label': 'negative', 'score': 0.9971930384635925},
 {'label': 'positive', 'score': 0.9524326920509338}],

[ ] #sentiment_list'in 95.000'inci index'inde bulunan sonucu ekrana yazdırdık.
print(sentiment_list[95000][0])

{'label': 'positive', 'score': 0.8874590992927551}

[ ] spredict_list = []
for i in range(0, len(sentiment_list)):
    spredict_list.append(sentiment_list[i][0])

[ ] spredict_list

[{'label': 'positive', 'score': 0.9524326920509338},
 {'label': 'negative', 'score': 0.9971930384635925},
 {'label': 'positive', 'score': 0.513373613357544},
 {'label': 'negative', 'score': 0.9987950325012207},
 {'label': 'negative', 'score': 0.8606646060943604},

[ ] data_new["label"] = spredict_list["label"]
data_new["score"] = spredict_list["score"]

[ ] data_new.to_csv('duyguanalizicikti.csv')
```

5. BULGULAR

İlk olarak çekilen 105.200 adet veri setine, metin madenciliği uygulandı. Veri ön işleme adımlarından sonra temizlenen veriden 103.108 adet twitter verisi elde edildi. Hazırlanmış olduğumuz makine öğrenmesi seti, 6 farklı algoritma ile eğitildi ve başarımları alındı. Elde ettiğimiz 103.108 adet veri ile makine öğrenmesi algoritmaları, Navie Bayes, Logistic Regression ve Support Vector Machine algoritmaları ile metin sınıflandırma işlemleri uygulandı. Son olarak atılan tweetlerin, duygularını anlamaya yönelik olumlu ve olumsuz olarak ayırım yapıldı. Bir diğer sütuna ise atılan tweetlerin duygu skorları yazdırıldı. Bunun sonucunda 6 farklı makine öğrenmesi algoritmasından elde ettiğimiz başarı yüzdeleri, görselleştirme adımıyla sunulacaktır. Duygu skorları için 'data_new' değişkeni, tahmin sonuçları için ise 'data' isimli değişken kullanılmıştır.

5.1. Tüm Verilere Ait Duygu Skorları

Duygu skorlarının, tüm veri setinde yapılan sonucu 'data_new' olarak yazdırılmıştır. Data_new, veri setinde toplamda 103.108 adet satır ve 3 adet sütun bulunmaktadır.

Şekil 28. Duygu Analizi sonucu

data_new			
	content	label	score
0	istanbul havalimanı dünyanın yoğun ikinci hava...	positive	0.952433
1	havalimanı etrafına yapılaşma var	negative	0.997193
2	gayrettepe istanbul havalimanınınhalkalı istanbu...	positive	0.513374
3	bundan başka argümanınız var öncesi sonrasıyla...	negative	0.998795
4	atatürk havalimanı yıkımı yargılanabilirn atat...	negative	0.860665
...
103103	havalimanı	positive	0.871141
103104	havalimanı köprü marmaray gebze köprüsü izmir ...	positive	0.972389
103105	size rağmen havalimanı olacak kanal istanbul n...	negative	0.994233
103106	havalimani yapilirsa almanyaya insani yardım g...	positive	0.566309
103107	sen milyar avro havalimanı ihalesi verırsen ol...	negative	0.873342
103108 rows × 3 columns			

5.2. Tüm Verilere Ait Tahminler

3 farklı makine öğrenmesi algoritması NB, LR ve SVM ile tüm veri setinde, ekonomi, siyaset, çevre yönünden kategori etiketi yapıldı. Aynı zamanda ‘sentiment’ adı ile yeni bir öznitelik ekleyip kullanıcıların duyguları ayrı bir sütunda görüntülenebiliyor. Son olarak ‘data’ isimli değişkenin sonucunu ekrana yazdırma işlemi yapıldı. Data veri seti, toplamda 103.108 adet satır ve 8 adet sütundan oluşmaktadır.

Şekil 29. Tahmin sonuçları

data								
	content	tahmin_naive_bayes	tahmin_category_nb	tahmin_logistic	tahmin_category_logistic	tahmin_svm	tahmin_category_svm	sentiment
0	istanbul havalimanı dünyanın yoğun ikinci hava...	0	ekonomi	0	ekonomi	0	ekonomi	positive
1	havalimanı etrafına yapılaşma var	2	çevre	2	çevre	2	çevre	negative
2	gayrettepe istanbul havalimanının halkalı istanbu...	0	ekonomi	0	ekonomi	0	ekonomi	positive
3	bundan başka argümanınız var öncesi sonrasıyla...	2	çevre	2	çevre	2	çevre	negative
4	atatürk havalimanı yıkımı yargılanabilir atat...	2	çevre	0	ekonomi	2	çevre	negative
...
103103	havalimanı	0	ekonomi	0	ekonomi	0	ekonomi	positive
103104	havalimanı köprü marmaray gebze köprüsü izmir ...	1	siyaset	1	siyaset	1	siyaset	positive
103105	size rağmen havalimanı olacak kanal istanbul n...	1	siyaset	1	siyaset	1	siyaset	negative
103106	havalimanı yapılırsa almanyaya insani yardım g...	1	siyaset	1	siyaset	1	siyaset	positive
103107	sen milyar avro havalimanı ihalesi venrsen ol...	0	ekonomi	0	ekonomi	0	ekonomi	negative

103108 rows x 8 columns

5.3. Sonuçların Görselleştirilmesi

Duygu analizi ve makine öğrenmesi algoritmaları ile metin sınıflandırma işlemleri sonucunda alınan çıktılar ile detaylı bir şekilde veri görselleştirme yapılmıştır. Kelime bulutunun oluşturulması, tüm verilere ait duygu skorları, duygu analizinin pasta grafik gösterimi, NB, LR ve SVM algoritmalarının tahmin sonuçlarının grafik ile gösterimleri sunulmuştur.

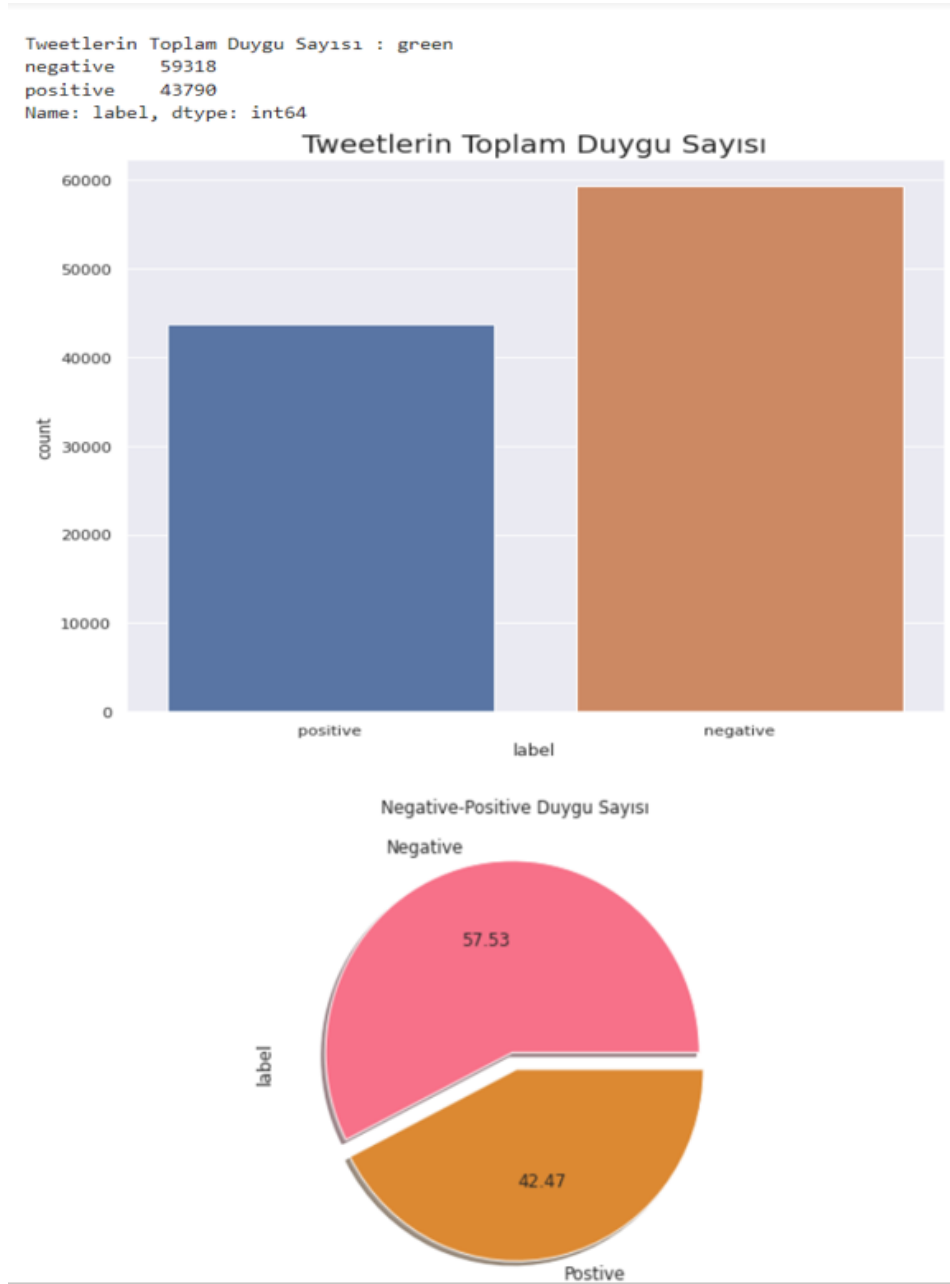
Şekil 30’da görüldüğü gibi metin madenciliği yapılarak veri ön işleme aşamasının ardından çıktısı alınan verinin, kelime bulutu oluşturuldu.

Şekil 30. Kelime bulutu



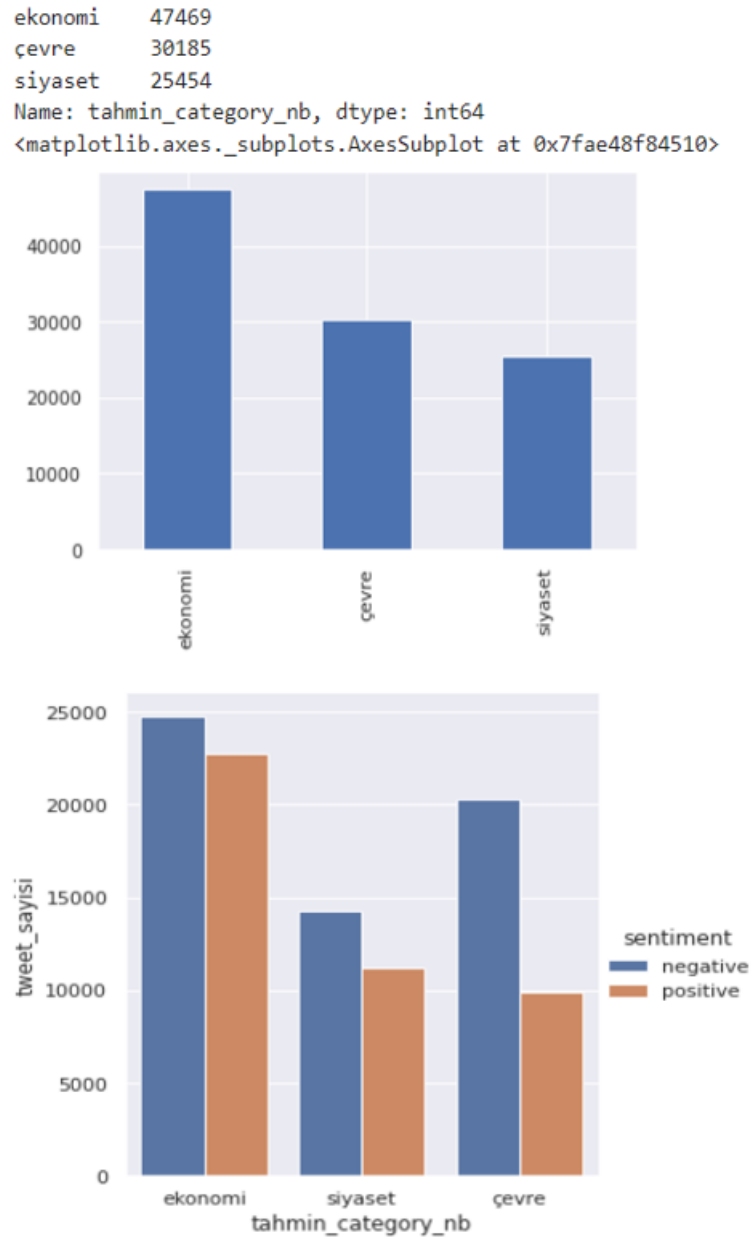
Şekil 31’de görüldüğü gibi yapılan duygu analizi sonucunda 59.318 tweetin duygusunun negatif olduğu, 43.790 adet tweetin duygusunun ise pozitif olduğu görülmektedir. Pasta grafikte ise duygu analizinin sonucunda tweetlerin duygu yüzdeleri, olumlu olarak %42.47, olumsuz olarak ise %57.53 olarak görülüyor.

Şekil 31. Duygu Analizi dağılımı



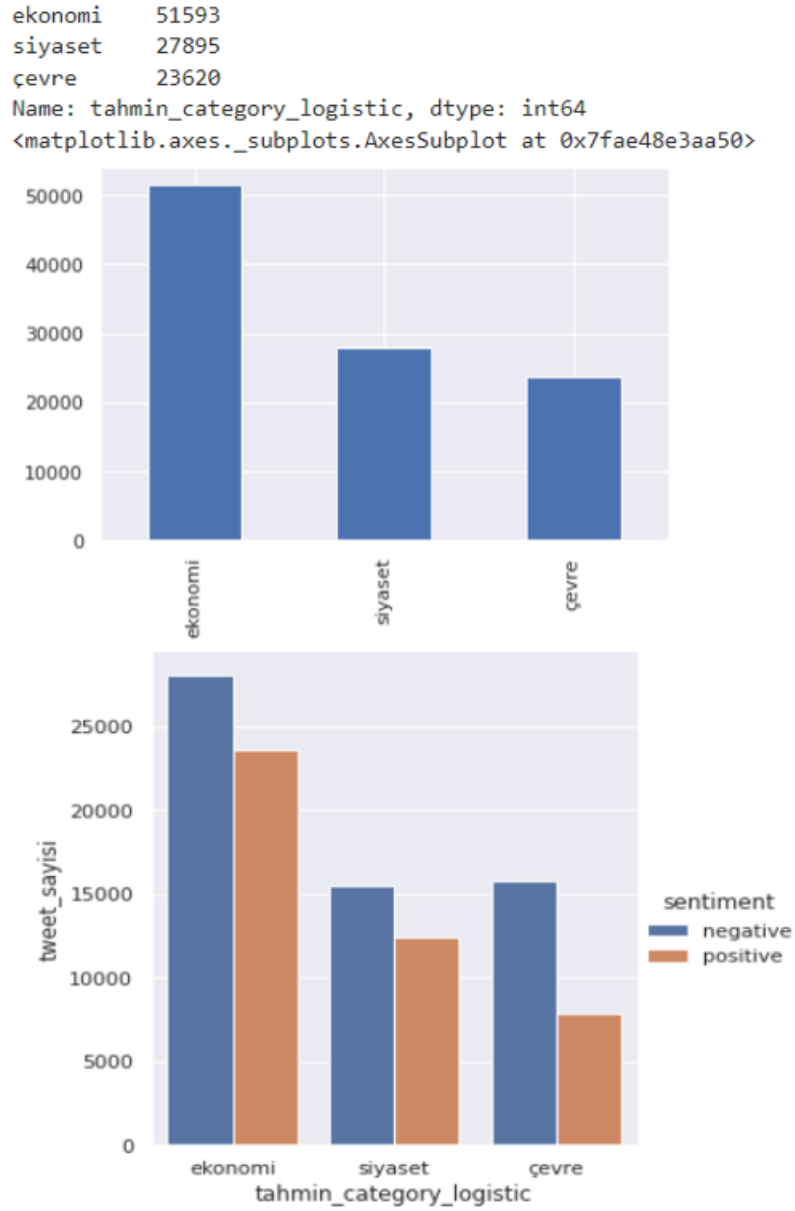
Şekil 32’de Navie Bayes algoritması kategori tahmin sonuçları görülmektedir. Navie Bayes algoritması %86.33’lük başarımları yakalamıştı. Bu başarımın sonucunda ise tweetlerin, 47.469’u ekonomi, 30.185’i çevre, 25.454 adet tweet ise siyaset olarak belirlenmiştir. Aynı zamanda olumlu ve olumsuz yönden ekonomi, siyaset ve çevre kategorilerinin oranları görselleştirilmiştir.

Şekil 32. Navie Bayes sonucu



Şekil 33'te görüldüğü gibi Logistic Regression tahmin sonuçları görülmektedir. Logistic Regression algoritması %87.50'lik başarımla en iyi sonuçlardan biridir. Bu başarımla tweetlerin, 51.593'ü ekonomi, 27.895'i siyaset, 23.620'si ise çevre kategorisini oluşturmaktadır. Ayrıca tweetlerin olumlu veya olumsuz yönden duyguları gösterilmiştir.

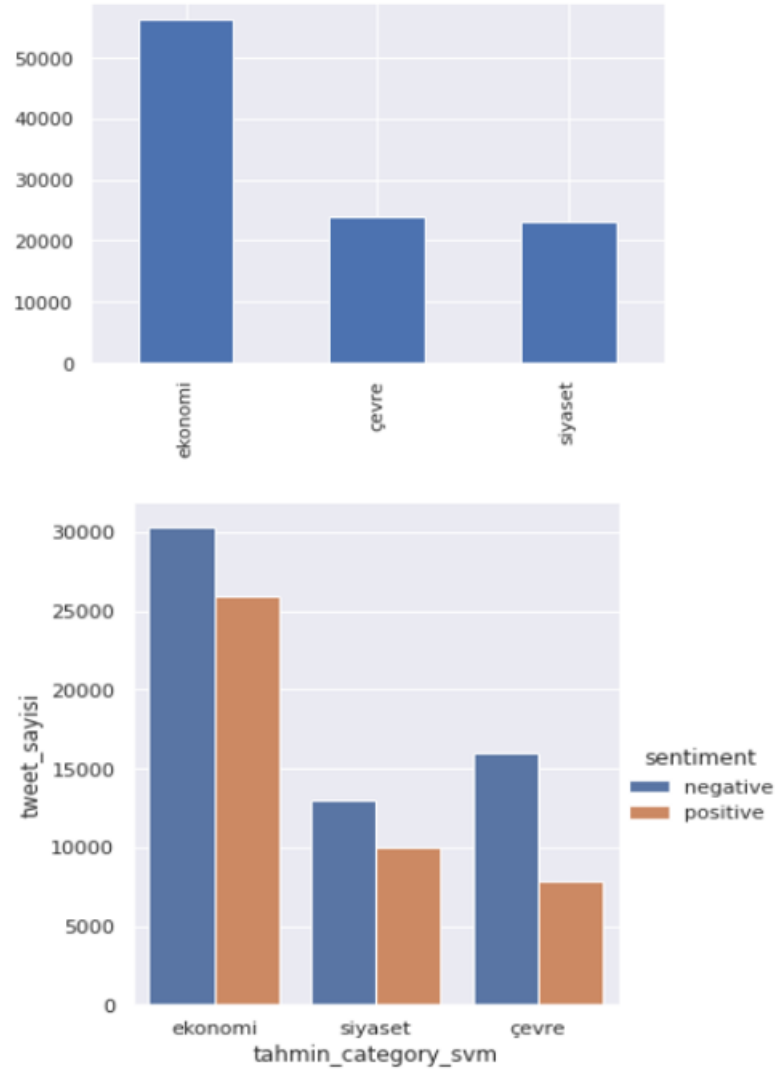
Şekil 33. Logistic Regression sonucu



Şekil 34'te görüldüğü gibi Support Vector Machine algoritmasının tahmin sonuçları görülmektedir. SVM algoritması en iyi sonuç veren algoritmalardan biri olup, %87.50'lik bir başarımlı sağlamıştır. SVM algoritmasının tahmin sonuçlarında, tweetlerin, 56.240'ı ekonomi, 23.876'sı çevre, 22.992'si ise siyaset tweetlerinden oluşmaktadır. Ekonomi, siyaset ve çevre kategorilerinin olumlu ve olumsuz karşılaştırmalarına yer verilmiştir.

Şekil 34. SVM sonucu

```
ekonomi    56240  
çevre      23876  
siyaset    22992  
Name: tahmin_category_svm, dtype: int64  
<matplotlib.axes._subplots.AxesSubplot at 0x7fae48d51590>
```



SONUÇ

Sosyal medya platformu olan Twitter'dan 'Python' ile 'Snsrape' kütüphanesinden yararlanılarak 'İstanbul Havalimanı' ile ilgili atılan tweetler çekilmiştir. Günümüzde şirketlerin bir ürüne, kişiye, topluluğa veya bizim çalışmamızdaki gibi bir proje olan tavrını görebilmek adına duygu analizi son zamanlarda aldığı hızlı ve etkili sonuçlarla pek çok çalışmada kullanılmaktadır. Aynı zamanda duygu analizi, çeşitli tekniklerle internet ortamında çok fazla zaman harcamadan ve diğer yöntemlere göre daha az maliyetli olmasından dolayı popülerliğini her geçen gün artırmaktadır.

Makine öğrenmesi seti oluşturarak çeşitli algoritmalar ile model eğitilmiştir. Eğitilen algoritmalar ile bir başarı oranı elde edilmiştir. En iyi sonuçlar Logistic Regression ile %87.50, SVM ile %87.50 ve Navie Bayes ile %86.33'lük bir başarı sağlanmıştır. Bu 3 algoritma, tüm verilere uygulanarak bir kategori etiketlemesi yapılmıştır. Yapılan işlem sonrasında en dikkat çeken detay ise 3 algortmada da en çok etiketlenen 'ekonomi' kategorisi olmuştur. Çalışmaya, duygu analizi adımlarını uygulayabilmek için Turkish Bert modeli ile birlikte Transformers kütüphanesinden yararlanarak olumlu ve olumsuz duyguları ve bu duyguların skorları elde edildi. Duygu Analizi sonucunda 59.318 tweet negatif, 43.790 tweet ise pozitif yönde bir analiz sonucu ortaya çıkmıştır. İstanbul Havalimanı bulunduğu konum itibariyle büyük bir öneme sahiptir. Projenin çevreye ve ekonomiye verdiği etkileri göz önünde bulundurduğumuzda insanların, İstanbul Havalimanı'nı iyi veya kötü yönde eleştirdiği görülmektedir.

Yaptığımız bu çalışmada, Twitter kullanıcıları bu projeyi daha çok ekonomik anlamda değerlendirmiştir. Projeye olan tavırları ise daha çok olumsuz anlamda olmuştur. Şekil 35'te duygu analizi sonunda elde edilen veri setinde negatif tweetlerin duygu skorlarının ortalaması 0.89, pozitif tweetlerin duygu skorlarının ortalaması ise 0,84 olarak görülmektedir.

Şekil 35. İstatistiksel sonuç

```
[ ] data_new.groupby(['label']).describe()['score']
```

	count	mean	std	min	25%	50%	75%	max
label								
negative	59318.0	0.890668	0.135228	0.500079	0.826305	0.958397	0.991764	0.999679
positive	43790.0	0.841975	0.140989	0.500020	0.746072	0.889909	0.956594	0.999828

KAYNAKÇA

Kaynakça

- Agrawal, Apoorva. (2017). <https://medium.com/data-science-group-iitr/logistic-regression-simplified-9b4efe801389>. medium: <https://medium.com>, Erişim tarihi: 08.05.2022
- Akgül, Eyüp S., Ertano, C. Diri, B. (2016). Twitter verileri ile duygu analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 22(2), 106-110.
- Aslanhan, Uğur. (2021). <https://www.aa.com.tr/tr/gundem/istanbul-havalimani-3-yilda-103-5-milyon-yolcuyu-agirladi/2407658>. Anadolu Ajansı: www.aa.com.tr, Erişim tarihi: 05.05.2022
- Beşkirli, Ayşe, Gülbandır, E. ve Dağ, İ. (2021). Metin Madenciliği Yöntemleri ile Twitter Verilerinden Bilgi Keşfi. *ESTUDAM Bilişim Dergisi*, 2(1), 21-25.
- Boon, Tom (2022). <https://simpleflying.com/istanbul-airport-of-the-year/>. simpleflying: www.simpleflying.com, Erişim tarihi: 04.05.2022
- Bozdağ, Abidin M. (2021). <https://www.aa.com.tr/tr/saglik/istanbul-havalimaninda-pcr-sonuclari-kisa-surede-yolculara-bildiriliyor/2260296>. Anadolu Ajansı: www.aa.com.tr, Erişim tarihi: 05.05.2022
- Brady, Paul. (2021). <https://www.travelandleisure.com/worlds-best/airports-international>. travelandleisure: www.travelandleisure.com, Erişim tarihi: 04.05.2022
- DHMI. (2021). <https://www.dhmi.gov.tr/Sayfalar/Istatistikler.aspx>. dhmi: <https://www.dhmi.gov.tr>, Erişim Tarihi: 05.05.2021
- Erdi, Bengisu, Şahin, E. A., Toydemir, M. S., ve Dökeroğlu, T. (2021). Makine Öğrenmesi Algoritmaları ile Trol Hesapların Tespiti. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9, 430-442.
- Hatipoğlu, Ekrem (2018). <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-naive-bayes-part-11-4a10cd3452b4>. medium: <https://medium.com>, Erişim tarihi: 08.05.2022
- İGA. (2018). <https://www.igairport.aero/tr/basin-merkezi/multimedya>. igairport.aero: <https://www.igairport.aero/tr>, Erişim tarihi: 04.05.2022

- igairport. (2019). <https://www.igairport.aero/tr/istanbul-havalimani/hizmetler/otopark.igairport.aero>: www.igairport.aero/tr, Eriřim tarihi: 05.05.2022
- Kılıç, Dođan ve Muhammed Turgut (2021). Kentsel Lojistik Açıısından İstanbul Havalimanı ve Atatürk Havalimanının Deđerlendirilmesi. *KOCATEPEİİBF Dergisi*, 21(2), 148-157.
- Koca, Gözde. (2021). Bitcoin Üzerine Twitter Verileri ile Duygu Analizi. *Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakülte Dergisi*, 22(4), 19-30.
- Köker, İrem (2018). <https://www.bbc.com/turkce/haberler-turkiye-45990453>. bbc news: <https://www.bbc.com>, Eriřim tarihi: 04.05.2022
- Kumar, Anjani (2020). <https://medium.com/analytics-vidhya/introduction-to-the-gradient-boosting-algorithm-c25c653f826b>. medium: <https://medium.com>, Eriřim tarihi: 08.05.2022
- Martinez, Viridiana R. (2019). <https://medium.datadriveninvestor.com/using-extreme-gradient-boosted-trees-in-machine-learning-classification-problems-a7bb04be759>. medium datadriveninvestor: <https://medium.datadriveninvestor.com>, Eriřim tarihi: 08.05.2022
- Özalp, Nur Y. ve Özdemir Sönmez (2022). Havalimanlarının Kent Makroformuna Etkisi: Yeni İstanbul Havalimanı. *Teknoloji ve Uygulamalı Bilimler Dergisi*, 4(2), 205-226.
- Pupalı, Rushikesh (2018). <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>. towardsdatascience: <https://towardsdatascience.com>, Eriřim tarihi: 08.05.2022
- Şantiye Dergisi. (2020). Limitlerin Zorlandığı İstanbul Havalimanı. *Şantiye Dergisi*, 383, 36-46. <https://www.santiye.com.tr/> Eriřim tarihi: 04.05.2022
- Şimsek, Zeynep (2018). <https://www.gazeteduvar.com.tr/gundem/2018/10/29/3-havalimaninin-3380-gunu>. gazeteduvar: www.gazeteduvar.com.tr Eriřim Tarihi: 05.05.2022
- Şimşek, Hakkı. K. (2018). <https://medium.com/data-science-tr/makine-%C3%B6%C4%9Frenmesi-dersleri-5-bagging-ve-random-forest-2f803cf21e07/data-science-tr/makine-%C3%B6%C4%9Frenmesi-dersleri-5-bagging-ve-random-forest-2f803cf21e07>. medium: <https://medium.com>, Eriřim tarihi: 08.05.2022

- TMMOB. (2014). 3.Havalimanı Teknik Raporu. Teknik Rapor, Türk Mühendis ve Mimar Odaları Birliği, https://www.tmmob.org.tr/sites/default/files/3.havalimani_ikk_rapor_20141208.pdf, Erişim Tarihi: 05.05.202
- Toçoğlu, Alp. M., Çekiten, A., Aygün, İ., ve Alpkoçak, A. (2019). Türkçe Metinlerde Duygu Analizi İçin Farklı Makine Öğrenmesi Yöntemlerinin Karşılaştırılması. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 21(63).
- TRT Haber. (2021). <https://www.trthaber.com/haber/ekonomi/istanbul-havalimani-dunyanin-en-iyileri-arasinda-601211.html>. trthaber: www.trthaber.com, Erişim tarihi: 05.05.2022