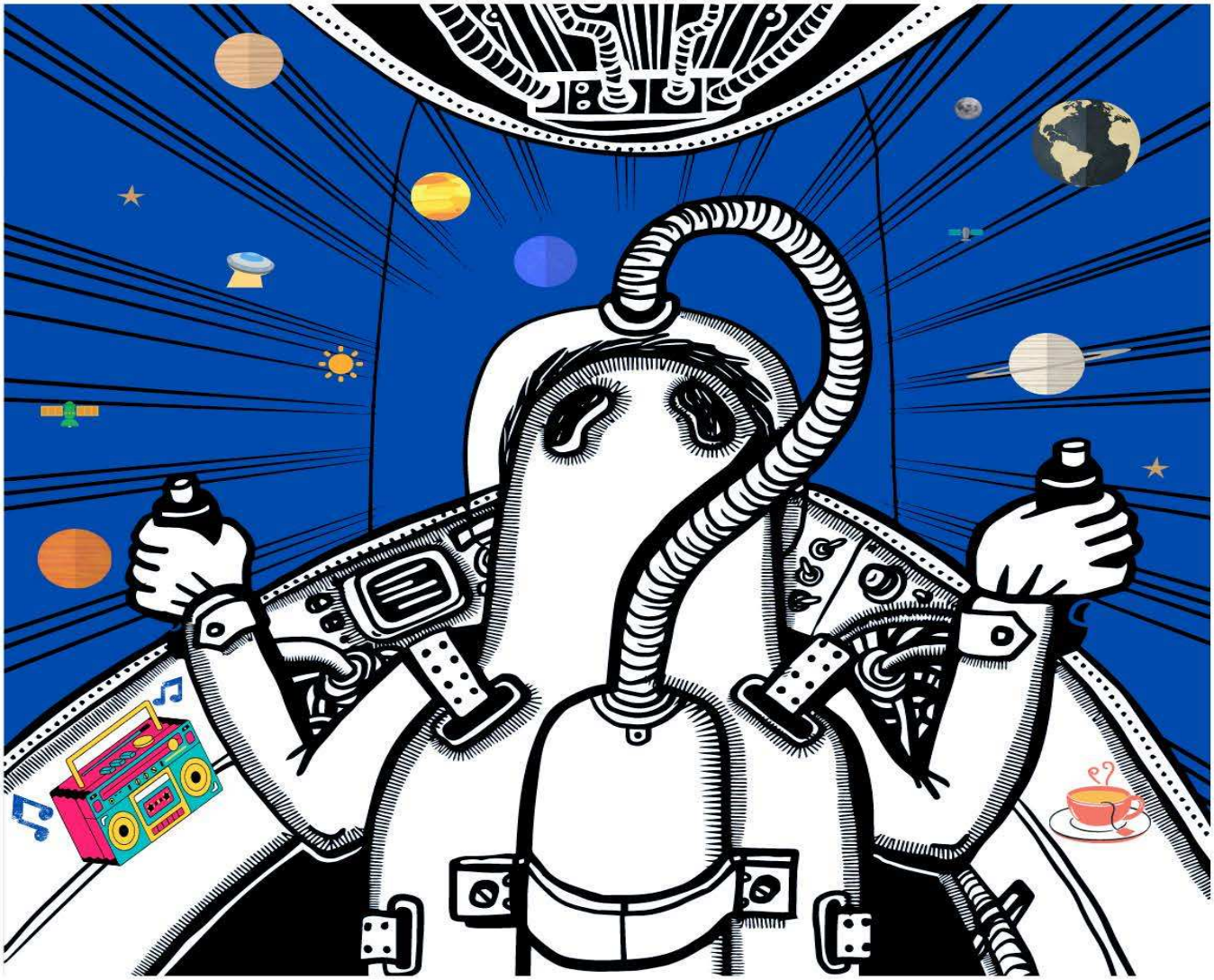
MAKİNE ÖĞRENMESİ ÖĞRETİCİSİ-PROJE 1
DÜNYA MUTLULUK RAPORU ANALİZİ
PYTHON İLE

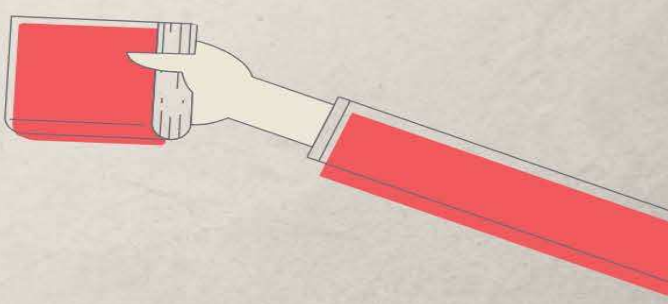
MAKİNE ÖĞRENMESİNE YOLCULUK



BERKANT ASLAN

ÖNSÖZ

Makine Öğrenmesini Python dili ve proje ile öğrenilmesini sağlamak için hazırlamış olduğum bu çalışmayı sizlerle paylaşıyorum. Sorularınız ve tavsiyeleriniz için berkantaslan@hotmail.com.tr olan mail adresime mail atabilirsiniz. Bu kitap üzerinde zamanla bazı güncellemeler olabilecek ve güncellemeler oldukça paylaşacağım. Haberdar olabilmek için beni takip edebilirsiniz.



SEVGİLİME, BABAMA, ANNEME VE KIZ KARDEŞİME...

Bu kitabın sizlere Makine Öğrenmesi konusunda pusula olması dileklerimle...

KİM BU BERKANT ASLAN YAA?

Selçuk Üniversitesi'nde Mekatronik Mühendisliği alanında Yapay Zeka konusunda yüksek lisans yapmakta; Yapay Zeka, Veri Bilimi, Görüntü İşleme, Büyük Veri, İş Zekası ve Siber Güvenlik ilgi alanları olan; insanları çok seven ve insanlığa kendisini adamış; Türkiye aşığı; saygılı ve sevgili; Isparta'nın gülü :) ve sevgilisiyle çok mutlu :)

berkantaslan@hotmail.com.tr

/berkantaslan

@berkantaslan

/berkantaslan

/berkantaslan

İÇİNDEKİLER

1. GİRİŞ

Dünya Mutluluk Raporu küresel mutluluğun durumunu gösteren bir ankettir. Yaklaşık olarak 155 ülkenin mutluluk seviyelerine göre sıralandığı bu rapor hükümetlerin nasıl politika izleyeceklerine karar vermesi; kuruluşların ve sivil toplumun bazı kararlar alabilmesi için bu raporlar paylaşılmaya ve takip edilmeye devam etmektedir. Alanında önde gelen uzmanlar ekonomi, psikoloji, anket analizi, ulusal istatistikler, sağlık, kamu politikası gibi konular üzerinde çalışarak refah ölçümlerinin ulusların ilerlemesini nasıl etkili bir şekilde kullanabileceklerini açıklar. Küresel Terörizm Veri Tabanı (GTD), 1970'ten 2017'ye kadar dünyanın dört bir yanındaki terörist saldırılara ilişkin bilgileri içeren açık kaynaklı bir veri tabanıdır. GTD, bu süre zarfında meydana gelen ve şimdi 180.000'den daha da fazla olan ulusal ve uluslararası terör olaylarına ilişkin sistematik verileri içermektedir.

2015, 2016 ve 2017 Dünya Mutluluk Raporları'nda Dünya Mutluluk Sıralaması'nı hangi faktörün ne kadar etkilediğini ve Küresel Terörizm Raporu'nu kullanarak bu yıllarda ki terör olaylarını veri setimize ekleyerek terör olaylarının etkisinin de sıralamaya etkisini izledik. 2018, 2019 ve 2020 Mutluluk Raporları'nda Dünya Mutluluk Sıralaması'nı hangi faktörün ne kadar etkilediğini ve bu analizimiz sonucunda bu 5 senelik rapor ile Makine Öğrenmesi yardımıyla mantıklı rastgele değerler ile mutluluk puanı tahmini yapıldı.

Herkes hayatta mutlu olmayı arzular ve ilginç bir şekilde mutlu olmanın gereklilikleri kişiden kişiye değişir. Hepimiz mutluluk kavramına farklı anlamlar yatırıyoruz. İnsanlar hayatımızın farklı noktalarında mutluluğa farklı değerler verir. Bununla birlikte, hayatta mutlu olmanın ana bileşenleri olarak kabul edilen bazı hayati faktörler vardır. Fiziksel ve psikolojik sağlamlık mutlu olmak için çok önemlidir ve insanlar hastalandıklarında bunu gerçekten anlayabilirler. Bence mutlu olmanın en önemli bir diğer faktörü de hayattaki ihtiyaçları karşılama becerisidir (Ekonomik özgürlük). Her zaman boş mideyle mutlu olmanın mümkün olmadığı söylenir. Bireysel doğum ve hak özgürlüğü de hayatta mutlu olmak için büyük bir etki olarak kabul edilir.

Ana analiz korelasyonları (regresyon; çıktı = mutluluk, değişkenler = ekonomi, sağlık, özgürlük vb.) içerecek ve bir Makine Öğrenimi algoritması oluşturacak, ülkeleri mutluluk derecelerine ve puanlarına göre sıralandıracaktır. Örneğin mutluluk puanının 7,00'den fazla olduğu bir ülke gelişmiş bir ülke, mutluluk puanının 5,00 ile 7,00 arasında olduğu bir ülke gelişmekte olan bir ülke ve mutluluk puanı 5,00'ün altında olan bir ülke ise gelişmemiş ülkedir.

Bu analiz, insanların mutluluğu ve ülkeleri arasındaki bağlantıyı değerlendirecek. Bence yaşadığımız yer mutluluk oranımızın büyük bir faktörü. Birleşmiş Milletler tarafından hazırlanan Dünya Mutluluk Raporu, yaklaşık 155 ülkeyi vatandaşlarının kendilerini ne kadar mutlu gördüklerine göre sıralamaktadır. Ekonomik refah, ortalama yaşam süresi, sosyal destek, yaşam seçimleri yapma özgürlüğü ve hükümet yolsuzluğu seviyeleri gibi faktörlere dayanır. Mutluluk puanları ve sıralamaları Gallup Dünya Anketinden alınan verileri kullanır. Puanlar, ankette sorulan temel yaşam değerlendirme sorusuna verilen yanıtlara dayanmaktadır.

Louise Millard 2011'de Makine Öğrenimi Yöntemleri ile Veri Madenciliği ve Küresel Mutluluk analizi yapıyor. Bu rapor, mutluluk değerleri olan 123 ülkeyi kapsıyor. Bu raporda ekonomi, sağlık, iklim verileri ele alınmaktadır. [1]

Natasha Jaques ve arkadaşları, 2015'te Makine Öğrenimi Yöntemleriyle Öğrencilerin Mutluluğunu Fizyoloji, Telefon, Hareketlilik ve Davranışsal Verilerden Tahmin Etme analizini yapıyor. Bu rapor, elektro termal aktivite (fizyolojik stresin bir ölçüsü) ve 3 eksenli ivmeölçer (adımlar ve fiziksel aktivite ölçüsü); akademik faaliyet, uyku, uyuşturucu ve alkol kullanımı ve egzersizle ilgili sorular içeren anket verileri; Telefon araması, SMS ve kullanım modelleriyle birlikte telefon verileri; gün boyunca kaydedilen koordinatlarla konum verileri. Bu makalede, mutlu ve mutsuz üniversite öğrencilerini ayırt etmek için bir makine öğrenimi algoritmasını analiz ediyorlar, hangi önlemlerin mutluluk hakkında en fazla bilgiyi sağladığını değerlendiriyorlar ve mutluluk, sağlık, enerji, uyanıklık ve stres dahil olmak üzere refahın farklı bileşenleri arasındaki ilişkiyi değerlendiriyorlar. [2]

2. VERİ HAKKINDA

Dünya Mutluluk Raporu olan ilk kullanılan veri, küresel mutluluk durumuna ilişkin dönüm noktası niteliğindeki bir araştırmadır. 155 ülkeyi mutluluk seviyelerine göre sıralayan rapor, hükümetler, kuruluşlar ve sivil toplum politika belirleme kararlarını bilgilendirmek için mutluluk göstergelerini giderek daha fazla kullandıkça küresel kabul görmeye devam ediyor. Ekonomi, psikoloji, anket analizi, ulusal istatistikler, sağlık, kamu politikası vb. alanlarda önde gelen uzmanlar, ulusların ilerlemesini değerlendirmek için refah ölçümlerinin nasıl etkili bir şekilde kullanılabileceğini açıklamaktadır.

Küresel Terörizm Veri Tabanı (GTD) olan ikinci kullanılan veri, 1970'den 2017'ye kadar dünyanın dört bir yanındaki terör saldırıları hakkındaki bilgileri içeren açık kaynaklı bir veri tabanıdır. GTD, bu sırada meydana gelen yerel ve uluslararası terör olaylarına ilişkin sistematik verileri ve 180.000'den fazla terör olayı içermektedir.

3. PROJE TANIMI

Çalışma, Mutluluk Puanının etkilerini incelemektedir. İnsanların mutluluk düzeyi bazı durumlardan etkilenebilir. CRISP-DM yöntemiyle Makine Öğrenimi, en azından en fazla etkiyi ölçer. Değerler mutluluk için en önemli şeyleri söyler. Projenin amacı, hükümetlere, kuruluşlara ve sivil topluma püf noktaları sağlar. Kurmak için Çoklu Doğrusal Regresyon modeli seçilir.

Öte yandan bu veri setleri ve tüm çıktılar küresel terör saldırıları ile tekrar analiz edilecektir. Planım, "Dünya Mutluluk Raporlarında olmayan yeni bir bağımsız değişken yaratmak. Bu amacın temel amacı, coğrafyanın terör saldırılarının tetikleyicisi olup olmadığını ve bu saldırıların vatandaşların mutluluğu üzerinde bir rolü olup olmadığını görmektir.

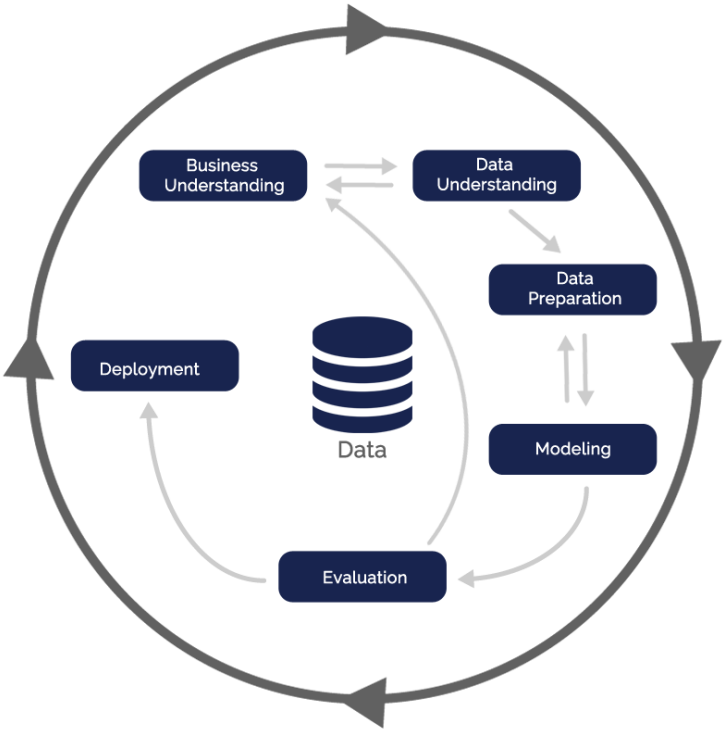
Bu projenin temel amacı, sosyoloji ve siyasete makine öğrenimi ve istatistik ile yaklaşmaktır. Çok ilginç olacak çünkü bu, veri biliminin hayatımızdaki yararını anlamak için güzel bir örnek. Veriler sayısal veya sözel değerler içerse de daha iyi bir dünya inşa etmeye yardımcı olacak analizler ve çıkarımlar yapabiliriz.

4. METODOLOJİ

Veri analizi için Python programlama dili ve Çoklu Doğrusal Regresyon Makine Öğrenme Metodu kullanıyoruz. Küresel Terörizm Raporu'ndan ülke, silah saldırısı vb. Olarak uygun verileri alıyoruz ve bu sütunlar Dünya Mutluluk Raporumuza ekleniyor. Bu veri çerçevesi, test (1/3) ve tren (2/3) olarak rastgele bölünür. Ve bu veri çerçevesi üzerinde çoklu doğrusal regresyon yöntemi çalışır. Ve hangi özelliğin mutluluk puanını diğerlerinden daha fazla etkilediğini buluyoruz. Yöntem için herhangi bir sorun varsa, diğer yöntemlere geçeceğiz ve çıktı değerlerini karşılaştıracağız.

4.1. CRISP-DM Metodu

Veri madenciliği projelerinin daha etkili, daha hızlı, daha güvenli, daha az maliyetli hale gelmesi için veri madenciliği için standart süreçleri tanımlayan bir model olan CRISP-DM (Cross Industry Standard Process for Data Mining) yöntemi. Bu yöntem, problemi ve ne yapmamız gerektiğini anlamakla başlar ve Makine Öğrenimi için veri hazırlama, modelleme ve değerlendirme ile devam eder.



Görsel-1: CRISP-DM Metodu Yapısı

4.2. Veriyi Anlamak

Dünya Mutluluk Raporu mutluluk puanları, ekonomik durumu (Kişi Başına GSYİH), sağlık durumu (Yaşam Beklentisi), özgürlük durumu, güven durumu (Devlet Yolsuzlukları), cömertlik durumu olan ülkeleri içerir. Mutluluk puanlarının bu değerlerle ilişkisini bulmayı anlıyoruz.

Küresel Terörizm Veri tabanı raporu, terör olaylarının görüldüğü ülkeleri, saldırı türlerini, silah türlerini vb. İçeren yılları içerir. Verimize yeni özellikler almanın ve eklemenin yeterli olduğunu anlıyoruz. Verimize bu veriden yıllar ve ülkeleri almak yeterlidir.

4.3. Veri Ön İşleme

Veri ön işleme kısmı, verilerin işlenmesi için önemlidir. Bu bölüm kütüphaneyi ve veri içe aktarmayı, veri temizlemeyi, kullanılacak uygun veri düzenlemeyi, Makine Öğrenimi bölümüne geçmek için veri çerçevesine veri eklemeyi içerir. Makine Öğrenimi bölümü için ön işlemlere ihtiyacımız var.

4.3.1. Kütüphane ve Veri İçe Aktarma

Süreçlere göre hangi kütüphanelerin kullanılacağına karar veriyoruz. Python'da veri ve veri çerçevesi işlemleri için Pandas kütüphanesi, hesaplamalar için NumPy kütüphanesi, çizim için Matplotlib ve

Seaborn kütüphanesi, "import" komutu ile Makine Öğrenimi işlemi için Sci-Kit Learn kütüphanesi kullanılır. Veri kullanımı csv (virgülle ayrılmış değerler), excel, html vb. Olabilir ve "read_csv / excel (" dosya_adı ")" komutu alınır. Veri dosyası çalışma dizini olmalıdır. Çalışma dizininde değilse, dosya_adı için dosya dizinini yazabilirsiniz. Ve bu herhangi bir değişken olarak tanımlanır.

```
import numpy as np
import pandas as pd

data1 = pd.read_csv('../input/world-happiness-report-with-terrorism/WorldHappinessReportwithTerrorism-2015.csv')
data2 = pd.read_csv('../input/world-happiness-report-with-terrorism/WorldHappinessReportwithTerrorism-2016.csv')
data3 = pd.read_csv('../input/world-happiness-report-with-terrorism/WorldHappinessReportwithTerrorism-2017.csv')
data4 = pd.read_csv('../input/world-happiness/2018.csv')
data5 = pd.read_csv('../input/world-happiness/2019.csv')
data6 = pd.read_csv('../input/world-happiness-report/2020.csv')
data7 = pd.read_csv('../input/global-terrorism-report-for-world-happiness-report/GlobalTerrorismReport-2015.csv')
data8 = pd.read_csv('../input/global-terrorism-report-for-world-happiness-report/GlobalTerrorismReport-2016.csv')
data9 = pd.read_csv('../input/global-terrorism-report-for-world-happiness-report/GlobalTerrorismReport-2017.csv')
```

4.3.2. Veri Temizleme

Bazı Makine Öğrenimi algoritmaları, eksik değerlere sahip verileri çalıştıramaz. İşlemler için düzeltilmesi gerekir. Eksik değerler NaN (sayı değil), "?", Boşluk vb. Görülebilir. Eksik sayısal değerler sabit istatistiksel yöntemler (ortalama vb.), belirli sayılar koyarak veya satırları silerek görülebilir. 2018 yılında sadece bir tane ve sütunların ortalamasına değiştirildi. Onun dışında çalışmamız için veriler eksik değere sahip olmadığı için başka bir işlem uygulanmamıştır. Büyük-küçük harf veya kelimeler veya farklı diller arasındaki boşluk gibi sütun adı tutarsızlığı olduğundan, keşfetmeden önce verileri teşhis etmemiz gerekir. Büyük-küçük harf veya boşluk sorunu düzeltildi. Verilerimizin sütun adlarında büyük-küçük harf veya boşluk varsa, sütun adlarını değiştirmeliyiz. IDE'de kodlayarak veya Excel'de yazarak sütun adlarını değiştirebilirsiniz.

Terörizm raporu, raporumuz için pek çok gereksiz bilgiye sahiptir. Dolayısıyla bu raporu raporumuza göre gereksiz sütunları silecek şekilde düzenliyoruz. Ve 2015, 2016 ve 2017'de sadece terörist saldırılarını filtreledik.

4.3.3. Kullanıma Uygun Olarak Veriyi Düzenleme

Veriler kategorik (nominal veya sıralı) veya sayısal (oran veya aralık) olabilir. Yalnızca sayısal verilere sahipsek, uygun veriler için düzenlememiz gerekmez. Yapmadıysak, verileri düzenlemeliyiz. Örneğin raporumuzda 2 kategorik verimiz var. Birincisi endeks numarası, ikincisi ülkeler. Bunlar mutluluk puanını etkiliyorsa, bu verileri sayı olarak kategorik olarak değiştirirken düzenlememiz gerekir. Ancak endeks numarası ve ülkeler mutluluk puanını etkilemez. Dolayısıyla, Makine Öğrenimi süreci için bunları göz ardı ediyoruz.

Terörün dünya mutluluk raporuna etkisini görmek istiyoruz. O halde Terörizm Raporundan uygun verileri almamız gerekiyor. Uygun veriler için sadece aynı yıl içinde kaç terör saldırısı olan ülkeleri alıyoruz. Dünya Mutluluk Raporu'nda olmayan diğer özellikleri ve ülkeleri görmezden geliyoruz. Mutluluk raporu için sadece etkilenmeye ihtiyacımız var. Terörizm olay sayılarını yalnızca yıllarla ve ülkelerle yıl olarak ayrı ayrı düzenlenmiş Küresel Terörizm Raporundan alıyoruz.

4.3.4. Veri Setine Veri Eklemek

Şimdi 2 veri çerçevemiz var: Dünya Mutluluğu veri çerçevesi ve sadece ülkeler ve sayıları içeren Terörizm veri çerçevesi. Dünya Mutluluk Veri Çerçevesine terörizm özelliği eklemeliyiz. Bu değerleri aynı yıl ve aynı ülkelerde ekliyoruz. Dünya Mutluluk Raporumuza terörizm olay sayılarını ekliyoruz.

4.4. Keşif Analizi

Verilerimizi anlamak ve üzerinde düşünmek istiyoruz. İlk olarak, istatistiksel bilgiler, veri çerçevelerinin kaç satır ve sütun olduğu vb. verilerden özet ve bilgi alıyoruz. Ve verileri teşhis etmek için keşif analizine ihtiyaç vardır.

4.5. Çoklu Doğrusal Regresyon Yöntemine Karar Verme

Çoklu Doğrusal Regresyon, problemdeki birden çok bağımsız değerden etkilenen çıktıyı kullanmaktır. Mutluluk puanları birden çok bağımsız değerden etkilenir. Ben de bu problem için bu yöntemi denemeye karar verdim.

$$y = \beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \epsilon$$

Formül-1: Çoklu Doğrusal Regresyon Yönteminin Formülü

Formülde, β0 yanlılık olarak bilinen sabit sayıdır, β1,2,3,4… katsayı olarak bilinen değişkenlerin çarpanı ve ϵ hata oranıdır. Bu formülden, doğrudan problem için inşa etmeyi anlıyoruz. Ve bu, ϵ (hata) olabileceği ve hesaplama eklenmesi gerektiği anlamına gelir. Çoklu Doğrusal Regresyon, uzayda 3 boyuttan fazla olabilir. B0 için her veride yalnızca 1 olan bir sütun ekliyoruz.

Y ve düz y (γi) değerleri arasındaki fark, artık olarak bilinir. Düz, minimum kare hatası (MSE) üzerine inşa edilmelidir. "N" örnek sayıdır.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \gamma i)^2$$

Formül-2: Minimum Kare Hata Formülü

4.6. Test ve Eğitim Kümesi Olarak Veriyi Bölme

Başarıyı ölçmek için test ve train kümesi olarak veri bölme yapılır. Makine Öğrenimi Algoritmasında amaç değerlendirme geliştirmektir. Genel olarak 1/3 testi ve 2/3 train için rastgele kullanılır, yüzdelik bölünme olarak bilinir. Makine öğrenimi algoritması train verileri üzerinde çalışır ve test verileri üzerinde tahmin yapar, tahmin ve test verilerinin mutluluk puanı karşılaştırılır. Ve başarı görülür ve Makine Öğrenimi Yönteminin iyi olup olmadığına karar verebiliriz. Başarıya göre kullanılan Makine Öğrenimi Yöntemi değiştirebiliriz.

4.7. Özellik Ölçekleme

Özellik Ölçekleme süreci Makine Öğrenimi için önemlidir. Farklı sütunların farklı veri hareketleri ve istatistiksel özellikleri (ortalama, minimum / maksimum değerler, standart sapma vb.) vardır. Yani sütunların etkileri farklıdır ve bu bir problemdir. Özellik ölçeklendirmeyi kullanarak bu sorunu çözmeliyiz. Özellik Ölçeklendirme iki klasik yöntemle yapılır: Standardizasyon ve Normalleştirme. Normalleştirme yöntemini kullanıyoruz. Sayısal veriler 0 ile 1 arasında ölçeklenir.

$$z = \frac{x - \mu}{\sigma} \qquad z = \frac{x - \min(x)}{[\max(x) - \min(x)]}$$

Formül-3: Standardizasyon ve Normalleştirme Formülleri

4.8. Sahte Değişken Tuzağına Dikkat Edin

Kukla değişken, veri çerçevesinin değişkenlerinin aynı değerlere sahip olduğu durumdur. Bazı Makine Öğrenimi Algoritmaları kukla değişkenden etkilenebilir. Veri çerçevesinde herhangi bir durum varsa, birini iptal etmek zorundayız. Örneğin, dize değerini sayısal değere dönüştürmek için erkek ve kadın değerleri 0 ve 1 olarak kodlanabilir. Ve bu iki kodlanmış sütunu kullanırsak sorun yaşayabiliriz. Çıktı, erkek veya kadın olabilen durumdan iki kez etkilenir. Sadece bir sütun yeterlidir ve ikincisi iptal edilir. Veri çerçevemiz için gerekli değildir.

4.9. Fine Tunning

Verilerin ve seçilen Makine Öğrenimi yönteminin başarılarına bakıyoruz. Herhangi bir sorun yaşarsak yöntemi değiştiririz. Veri başarısı için P-Değerlerine bakarız ve yöntemin başarısı için R^2 ve Düzeltilmiş R^2'ye bakarız.

4.9.1. P Değerlerine Bakmak

Bir p değeri, sonucun tesadüfen oluşma olasılığını temsil eden bir sonucun önemini belirtir. Düşük bir değer, sonucun rastgele oluşma ihtimalinin düşük olduğu ve dolayısıyla istatistiksel olarak anlamlı olduğu anlamına gelir. Yaygın olarak %5 ve %1 eşik değerleri kullanılır, bunun altında bir sonuç anlamlı olarak belirtilebilir. İstatistiksel testler, veri kümesinin boyutu gibi test parametrelerine göre olabilir ve bir p değeri, verilerin bu tür yönlerini hesaba katan karşılaştırılabilir bir değer sağlar. [1]

4.9.2. Yöntem Başarısının R^2 ve Düzeltilmiş R^2 ile Karşılaştırılması

R^2 aynı zamanda belirleme katsayısı olarak da bilinir ve R değerinin bir uzantısıdır. Regresyon modeli ile açıklanabilen etiketteki çeşitlilik oranını temsil eder. Bununla birlikte, R modelde kullanılan değişkenlerin kullanımı farklı olduğunda karşılaştırılamaz. Düzeltilmiş R^2, regresyonda kullanılan değişkenlerin sayısını dikkate alır. [1]

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \gamma i)^2}{\sum_{i=1}^{n}(yi - y\ average)^2} \quad \text{Adjusted } R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Formül-4: R^2ve Düzeltilmiş R^2'nin Formülleri

R^2 ve Düzeltilmiş R^2 değerleri 1'e ne kadar yaklaşıyorsa, seçilen yöntem için çok iyi bir durumdur. Çoklu Doğrusal Regresyon Yöntemi analizimiz için iyidir. Mutluluk puanlarının etkilerini analiz ediyoruz. Veri setinde bulunan tüm değerler önemlidir çünkü tüm değerler mutluluk puanlarını etkiler. Bu nedenle, hiçbir özelliği iptal etmiyoruz veya yok saymıyoruz.

5. VERİ ANALİZİ

World Happiness Rank 2015

RangeIndex: 158 entries, 0 to 157	
Data columns (total 12 columns):	
country	158 non-null object
region	158 non-null object
happinessrank	158 non-null int64
happinessscore	158 non-null float64
standarderror	158 non-null float64
economysituation	158 non-null float64
family	158 non-null float64
healthlifeexpectancy	158 non-null float64
freedom	158 non-null float64
governmentcorruption	158 non-null float64
generosity	158 non-null float64
dystopiaresidual	158 non-null float64
terrorismevent	158 non-null int64

Tablo-1: World Happiness Rank 2015 Veri Özeti

2015 verilerimizin 158 ülke, boş olmayan değerlere sahip 12 özellik içerdiğini Tablo-1'den anlıyoruz.

Mutluluk puanlarına göre ülkelerin isimlerini kelime torbası olarak görmek için İş Zekası araçlarına sahip Tableau programını kullanıyoruz.

Görsel-2: 2015 Dünya Mutluluk Sıralamasında Mutluluk Puanlarına Göre Ülkeler

```
print(data1.columns)

print(data1.info())

print(data1.describe())

df1 = data1["country"]
dff1 = df1.value_counts()
```

2015 verilerimizin 158 ülke, boş olmayan değerlere sahip 12 özellik içerdiğini ve diğer tüm istatistiksel bilgilere kod çıktısından ulaşıyoruz.

```
x1 = data1.iloc[:,5:].values
y1 = data1.iloc[:,3:4].values
```

Kod çıktısı olarak x mutluluk puanını etkileyen girdiler ve y sonuç olan mutluluk puanı sütunu olarak x ve y değişkenine atanırlar.

```
from sklearn.model_selection import train_test_split
x_train1, x_test1, y_train1, y_test1 = train_test_split(x1, y1, test_size=0.33, random_state=0)
```

Kod çıktısı olarak test ve eğitim kümeleri oluşturulur.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train1 = sc.fit_transform(x_train1)
X_test1 = sc.fit_transform(x_test1)
Y_train1 = sc.fit_transform(y_train1)
Y_test1 = sc.fit_transform(y_test1)
```

Kod çıktısı olarak normalizasyon işlemi yapılır.

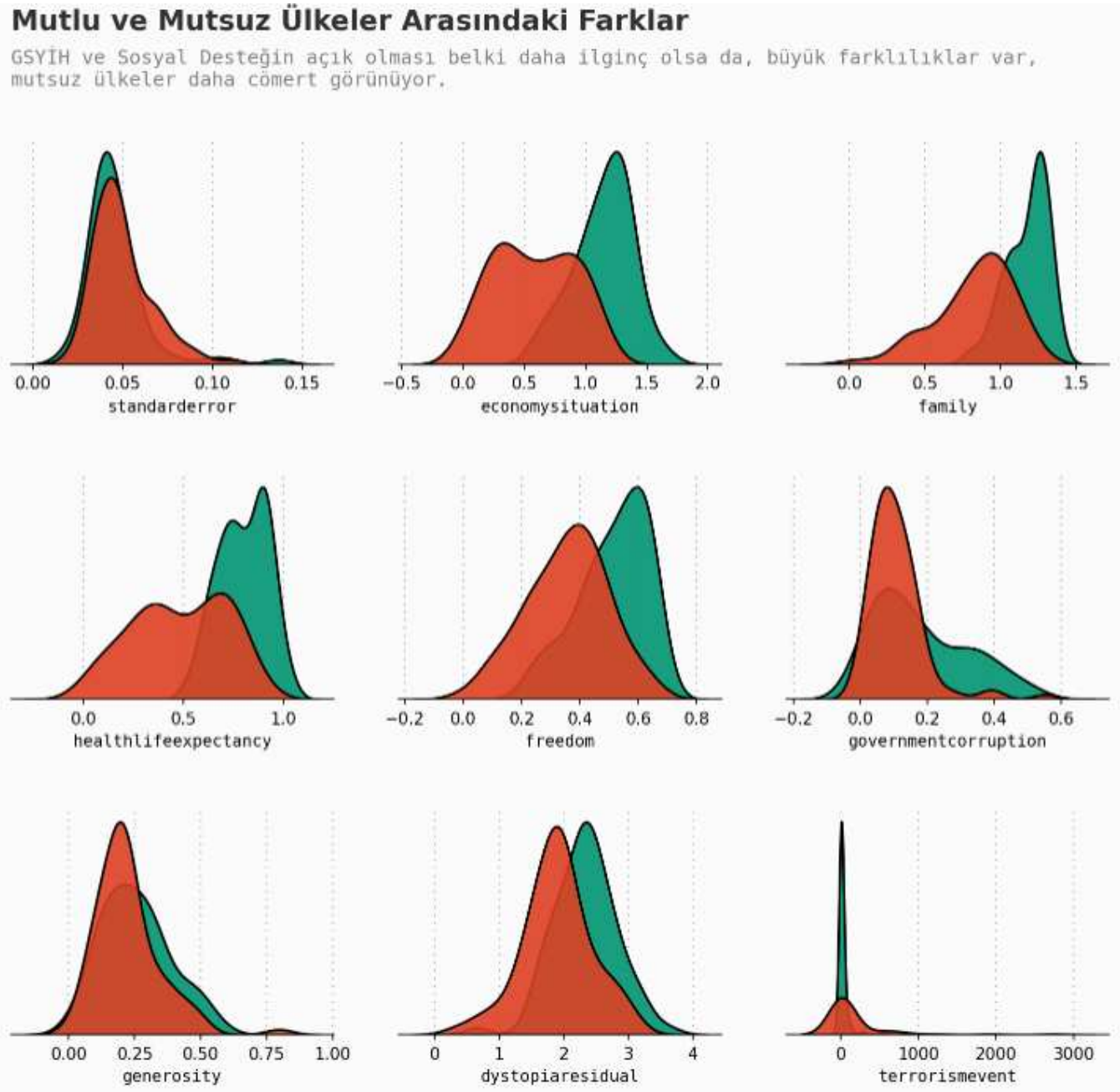
```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train1, y_train1)
print("b0: ", lr.intercept_)
print("other b: ", lr.coef_)
```

Kod çıktısı olarak makine öğrenmesi modeli oluşturulur ve modele göre ağırlık çarpan değerleri hesaplanır.

```
y_pred1 = lr.predict(x_test1)
prediction1 = lr.predict(np.array([[1.16492,0.87717,0.64718,0.23889,0.12348,0.04707,2.29074,542]]))
print("Prediction is ", prediction1)
```

```
y_pred1 = lr.predict(x_test1)
prediction1 = lr.predict(np.array([[1.198274,1.337753,0.637606,0.300741,0.099672,0.046693,1.879278,181]]))
print("Prediction is ", prediction1)
```

Türkiye için 2015 raporunda gerçek değerin 5.332 mutluluk puanı olacağını rapordan görüyoruz, modelimiz 2016 raporundan 5.38965305 ve 2017 raporundan 5.50012402 olarak tahmin ediyor. 2017 değerinin ve diğerlerinin farklılıklarının nedeninin özelliklerin ve terörizmin etkisinin değiştiğini tahmin ediyoruz.



```
low_c = '#dd4124'
high_c = '#009473'
background_color = '#fbfbfb'
fig = plt.figure(figsize=(12, 10), dpi=150,facecolor=background_color)
gs = fig.add_gridspec(3, 3)
gs.update(wspace=0.2, hspace=0.5)

newdata1 = data1.iloc[:,4:]
categorical = [var for var in newdata1.columns if newdata1[var].dtype=='O']
continuous = [var for var in newdata1.columns if newdata1[var].dtype!='O']

happiness_mean = data1['happinessscore'].mean()

data1['lower_happy'] = data1['happinessscore'].apply(lambda x: 0 if x < happiness_mean else 1)

plot = 0
for row in range(0, 3):
    for col in range(0, 3):
        locals()["ax"+str(plot)] = fig.add_subplot(gs[row, col])
        locals()["ax"+str(plot)].set_facecolor(background_color)
        locals()["ax"+str(plot)].tick_params(axis='y', left=False)
        locals()["ax"+str(plot)].get_yaxis().set_visible(False)
        locals()["ax"+str(plot)].set_axisbelow(True)
        for s in ["top","right","left"]:
            locals()["ax"+str(plot)].spines[s].set_visible(False)
        plot += 1

plot = 0

Yes = data1[data1['lower_happy'] == 1]
No = data1[data1['lower_happy'] == 0]

for variable in continuous:
        sns.kdeplot(Yes[variable],ax=locals()["ax"+str(plot)], color=high_c,ec='black', shade=True, linewidth=1.5,
alpha=0.9, zorder=3, legend=False)
        sns.kdeplot(No[variable],ax=locals()["ax"+str(plot)], color=low_c, shade=True, ec='black',linewidth=1.5, a
lpha=0.9, zorder=3, legend=False)
        locals()["ax"+str(plot)].grid(which='major', axis='x', zorder=0, color='gray', linestyle=':', dashes=(1,5)
)
        locals()["ax"+str(plot)].set_xlabel(variable, fontfamily='monospace')
        plot += 1

Xstart, Xend = ax0.get_xlim()
Ystart, Yend = ax0.get_ylim()

ax0.text(Xstart, Yend+(Yend*0.5), 'Differences Between Happy & Unhappy Countries', fontsize=15, fontweight='bold',
fontfamily='sansserif',color='#323232')
ax0.text(Xstart, Yend+(Yend*0.25), 'There are large differences, with GDP & Social Support being clear perhaps mor
e interesting though,unhappy\ncountries appear to be more generous.', fontsize=10, fontweight='light', fontfamily=
'monospace',color='gray')

plt.show()
```

```
import statsmodels.regression.linear_model as sm
X1 = np.append(arr = np.ones((158,1)).astype(int), values=x1, axis=1)
r_ols1 = sm.OLS(endog = y1, exog = X1)
r1 = r_ols1.fit()
print(r1.summary())
```

R squared:	1.000
Adj. R squared:	1.000

Tablo-2: Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2015 Verilerindeki Başarısı

Tablo-2'den modelimizin çok başarılı olduğunu anlıyoruz. Kod çıktısı olarak R-squared ve Adj. R-squared değerleri (1.000 ve 1.000) Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2015 Verilerindeki Başarısı'nı göstermektedir. Bu değerlerin 1 veya 1'e yakın olması modelimizin çok başarılı olduğunu anlıyoruz. Bu nedenle, seçili Makine Öğrenimi yöntemimizi değiştirmiyoruz.

Variables	**P-Value**
constant	0.586
economysituation	0.000
family	0.000
healthlifeexpectancy	0.000
freedom	0.000
governmentcorruption	0.000
generosity	0.000
dystopiaresidual	0.000
terrorismevent	0.846

Tablo-3: Değişkenlerin Dünya Mutluluk Sıralaması 2015 Verileri Üzerindeki Başarısı

P>|t| değeri Değişkenlerin Dünya Mutluluk Sıralaması 2015 Verileri Üzerindeki Başarısı'nı göstermektedir. 1 olarak eklediğimiz sabit değer ve eklenen terör olayı değerlerimizin p değerlerinin %5 veya %1 eşik değerlerinden fazla olduğunu Tablo-3'ten anlıyoruz. Bu, terörizm olaylarının değerlerinin modelimiz için uygun ve etkilenebilir olmadığı anlamına gelir.

World Happiness Rank 2016

RangeIndex: 157 entries, 0 to 156	
Data columns (total 13 columns):	
country	157 non-null object
region	157 non-null object
happinessrank	157 non-null int64
happinessscore	157 non-null float64
lowerconfidenceinterval	157 non-null float64
upperconfidenceinterval	157 non-null float64
economysituation	157 non-null float64
family	157 non-null float64
healthlifeexpectancy	157 non-null float64
freedom	157 non-null float64
governmentcorruption	157 non-null float64
generosity	157 non-null float64
dystopiaresidual	157 non-null float64
terrorismevent	157 non-null int64

Tablo-4: World Happiness Rank 2016 Veri Özeti

Tablo-4'ten 2016 verilerimizin 157 ülke, boş olmayan değerlere sahip 13 özellik olduğunu anlıyoruz.

Mutluluk puanlarına göre ülkelerin isimlerini kelime torbası olarak görmek için İş Zekası araçlarına sahip Tableau programını kullanıyoruz.

Görsel-3: 2016 Dünya Mutluluk Sıralamasında Mutluluk Puanlarına Göre Ülkeler

```
print(data2.columns)

print(data2.info())

print(data2.describe())

df2 = data2["country"]
dff2 = df2.value_counts()
```

2016 verilerimizin 157 ülke, boş olmayan değerlere sahip 13 özellik içerdiğini ve diğer tüm istatistiksel bilgilere kod çıktısından ulaşıyoruz.

```
x2 = data2.iloc[:,6:].values
y2 = data2.iloc[:,3:4].values
```

Kod çıktısı olarak x mutluluk puanını etkileyen girdiler ve y sonuç olan mutluluk puanı sütunu olarak x ve y değişkenine atanırlar.

```
from sklearn.model_selection import train_test_split
x_train2, x_test2, y_train2, y_test2 = train_test_split(x2, y2, test_size=0.33, random_state=0)
```

Kod çıktısı olarak test ve eğitim kümeleri oluşturulur.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train2 = sc.fit_transform(x_train2)
X_test2 = sc.fit_transform(x_test2)
Y_train2 = sc.fit_transform(y_train2)
Y_test2 = sc.fit_transform(y_test2)
```

Kod çıktısı olarak normalizasyon işlemi yapılır.

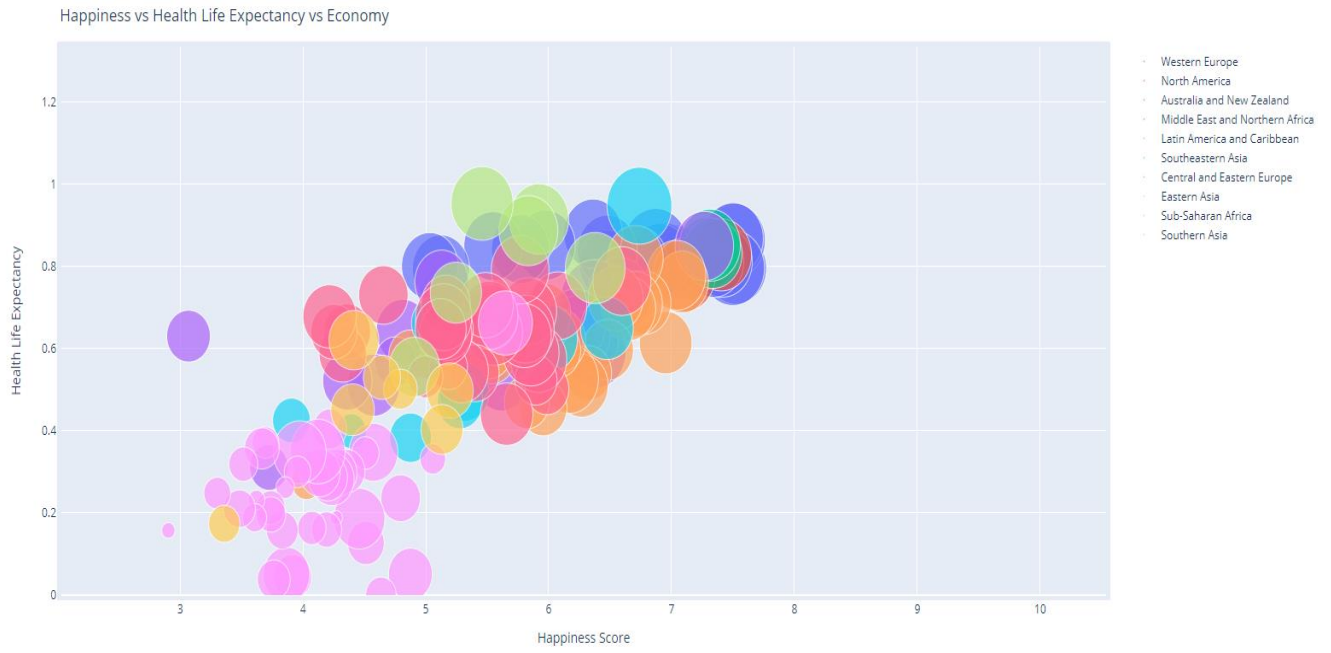
```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train2, y_train2)
print("b0: ", lr.intercept_)
print("other b: ", lr.coef_)
```

Kod çıktısı olarak makine öğrenmesi modeli oluşturulur ve modele göre ağırlık çarpan değerleri hesaplanır.

```
y_pred2 = lr.predict(x_test2)
prediction2 = lr.predict(np.array([[1.06098,0.94632,0.73172,0.22815,0.15746,0.12253,2.08528,422]]))
print("Prediction is ", prediction2)
```

```
y_pred2 = lr.predict(x_test2)
prediction2 = lr.predict(np.array([[1.198274,1.337753,0.637606,0.300741,0.099672,0.046693,1.879278,181]]))
print("Prediction is ", prediction2)
```

Türkiye için 2016 raporunda gerçek değerin 5.389 mutluluk puanı olacağını rapordan görüyoruz, modelimiz 2015 raporundan 5.33218602 ve 2017 raporundan 5.49986729 olarak tahmin ediyor. 2017 değerinin ve diğerlerinin farklılıklarının nedeninin özelliklerin ve terörizmin etkisinin değiştiğini tahmin ediyoruz.




```
figure = bubbleplot(dataset = data2, x_column = 'happinessscore', y_column = 'healthlifeexpectancy',
    bubble_column = 'country', size_column = 'economysituation', color_column = 'region',
    x_title = "Happiness Score", y_title = "Health Life Expectancy", title = 'Happiness vs Health Life Expectancy
vs Economy',
    x_logscale = False, scale_bubble = 1, height = 650)

po.iplot(figure)
```

```
import statsmodels.regression.linear_model as sm
X2 = np.append(arr = np.ones((157,1)).astype(int), values=x2, axis=1)
r_ols2 = sm.OLS(endog = y2, exog = X2)
r2 = r_ols2.fit()
print(r2.summary())
```

R squared:	1.000
Adj. R squared:	1.000

Tablo-5: Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2016 Verilerindeki Başarısı

Kod çıktısı olarak R-squared ve Adj. R-squared değerleri(1.000 ve 1.000) Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2016 Verilerindeki Başarısı'nı göstermektedir. Bu değerlerin 1 veya 1'e yakın olması modelimizin çok başarılı olduğunu anlıyoruz. Bu nedenle, seçili Makine Öğrenimi yöntemimizi değiştirmiyoruz.

Variables	**P-Value**
constant	0.281
economysituation	0.000
family	0.000
healthlifeexpectancy	0.000
freedom	0.000
governmentcorruption	0.000
generosity	0.000
dystopiaresidual	0.000
terrorismevent	0.619

Tablo-6: Değişkenlerin Dünya Mutluluk Sıralaması 2016 Verileri Üzerindeki Başarısı

P>|t| değeri Değişkenlerin Dünya Mutluluk Sıralaması 2016 Verileri Üzerindeki Başarısı'nı göstermektedir. Tablo-6'dan 1 olarak eklediğimiz sabit ve eklenen terör olayı değerlerimizin p değerlerinin %5 veya %1 eşik değerlerinden fazla olduğunu anlıyoruz. Bu, terörizm olaylarının değerlerinin modelimiz için uygun ve etkilenebilir olmadığı anlamına gelir. Ancak bu p değerlerinin azaldığını görüyoruz, bu modelimize olan etkilerinin artması anlamına geliyor.

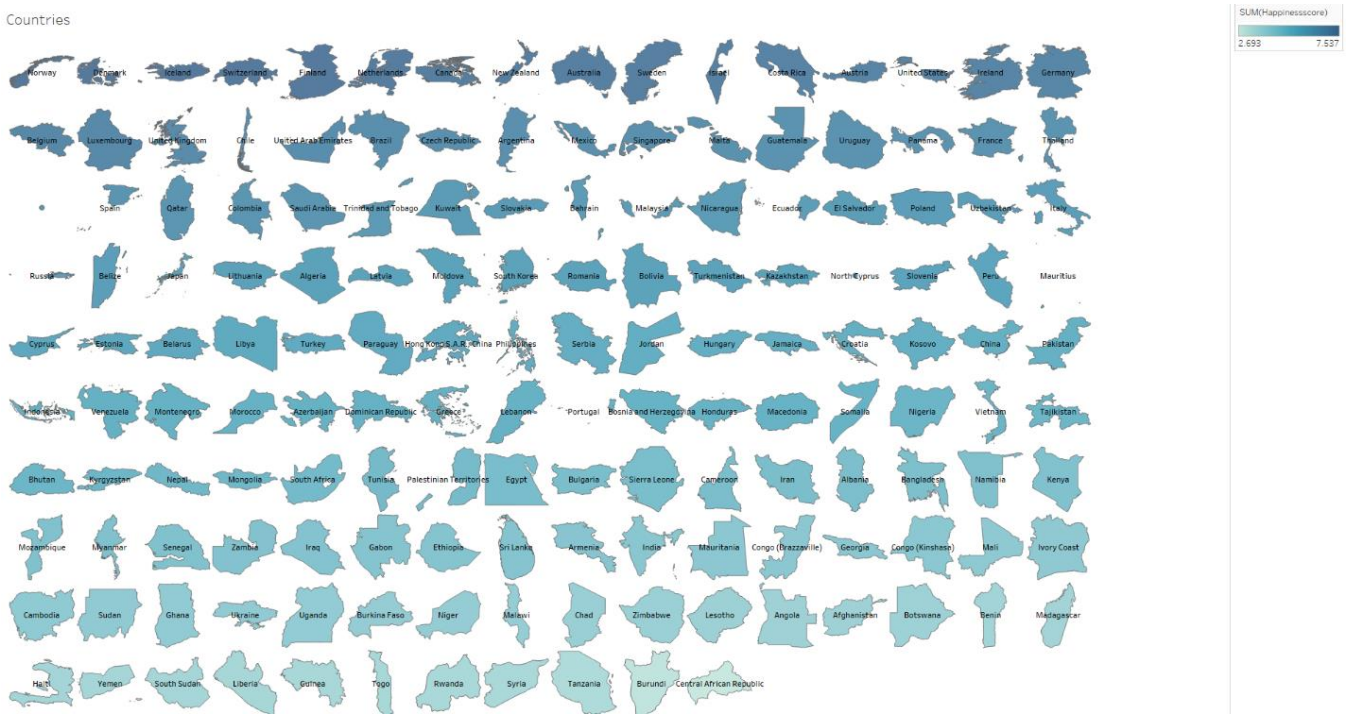
World Happiness Rank 2017

RangeIndex: 155 entries, 0 to 154	
Data columns (total 12 columns):	
country	155 non-null object
happinessrank	155 non-null int64
happinessscore	155 non-null float64
whiskerhigh	155 non-null float64
whiskerlow	155 non-null float64
economysituation	155 non-null float64
family	155 non-null float64
healthlifeexpectancy	155 non-null float64
freedom	155 non-null float64
generosity	155 non-null float64
governmentcorruption.	155 non-null float64
dystopiaresidual	155 non-null float64
terrorismevent	155 non-null int64

Tablo-7: World Happiness Rank 2017Veri Özeti

Tablo-7'den 2017 verilerimizin 155 ülke, boş olmayan değerlere sahip 12 özellik olduğunu anlıyoruz.

Mutluluk puanlarına göre ülkelerin isimlerini ülkelerin şekli ile kelime torbası olarak görmek için İş Zekâsı araçlarının bulunduğu Tableau programını kullanıyoruz.



Görsel-4: 2017 Dünya Mutluluk Sıralamasında Mutluluk Puanlarına Göre Ülkeler

```
print(data3.columns)

print(data3.info())

print(data3.describe())

df3 = data3["country"]
dff3 = df3.value_counts()
```

2017 verilerimizin 155 ülke, boş olmayan değerlere sahip 12 özellik içerdiğini ve diğer tüm istatistiksel bilgilere kod çıktısından ulaşıyoruz.

```
x3 = data3.iloc[:,5:].values
y3 = data3.iloc[:,2:3].values
```

Kod çıktısı olarak x mutluluk puanını etkileyen girdiler ve y sonuç olan mutluluk puanı sütunu olarak x ve y değişkenine atanırlar.

```
from sklearn.model_selection import train_test_split
x_train3, x_test3, y_train3, y_test3 = train_test_split(x3, y3, test_size=0.33, random_state=0)
```

Kod çıktısı olarak test ve eğitim kümeleri oluşturulur.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train3 = sc.fit_transform(x_train3)
X_test3 = sc.fit_transform(x_test3)
Y_train3 = sc.fit_transform(y_train3)
Y_test3 = sc.fit_transform(y_test3)
```

Kod çıktısı olarak normalizasyon işlemi yapılır.

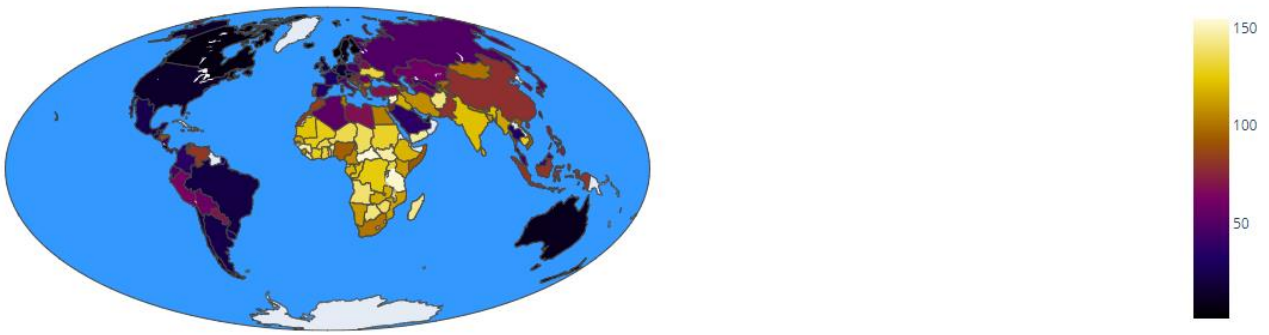
```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train3, y_train3)
print("b0: ", lr.intercept_)
print("other b: ", lr.coef_)
```

Kod çıktısı olarak makine öğrenmesi modeli oluşturulur ve modele göre ağırlık çarpan değerleri hesaplanır.

```
y_pred3 = lr.predict(x_test3)
prediction3 = lr.predict(np.array([[1.06098,0.94632,0.73172,0.22815,0.15746,0.12253,2.08528,422]]))
print("Prediction is ", prediction3)
```

```
y_pred3 = lr.predict(x_test3)
prediction3 = lr.predict(np.array([[1.16492,0.87717,0.64718,0.23889,0.12348,0.04707,2.29074,542]]))
print("Prediction is ", prediction3)
```

Kod çıktısı olarak Türkiye için 2017 raporunda gerçek değerin 5,5 mutluluk puanı olacağını rapordan görüyoruz, modelimiz 2015 raporundan 5.33249758 ve 2016 raporundan 5.38949664 olarak tahmin ediyor. Bu modeli kullanmak isterseniz, denkliği nedeniyle 2015 ve 2016 raporlarını kullanabilirsiniz.




```
trace1 = [go.Choropleth(
               colorscale = 'Electric',
               locationmode = 'country names',
               locations = data3['country'],
               text = data3['country'],
               z = data3['happinessrank'],
               )]

layout = dict(title = 'Happiness Rank',
                  geo = dict(
                      showframe = True,
                      showocean = True,
                      showlakes = True,
                      showcoastlines = True,
                      projection = dict(
                          type = 'hammer'
        )))

projections = [ "equirectangular", "mercator", "orthographic", "natural earth","kavrayskiy7",
               "miller", "robinson", "eckert4", "azimuthal equal area","azimuthal equidistant",
               "conic equal area", "conic conformal", "conic equidistant", "gnomonic", "stereographic",
               "mollweide", "hammer", "transverse mercator", "albers usa", "winkel tripel" ]

buttons = [dict(args = ['geo.projection.type', y],
           label = y, method = 'relayout') for y in projections]

annot = list([ dict( x=0.1, y=0.8, text='Projection', yanchor='bottom',
                    xref='paper', xanchor='right', showarrow=False )])

# Update Layout Object
layout[ 'updatemenus' ] = list([ dict( x=0.1, y=0.8, buttons=buttons, yanchor='top' )])
layout[ 'annotations' ] = annot

fig = go.Figure(data = trace1, layout = layout)
po.iplot(fig)
```

R squared:	1.000
Adj. R squared:	1.000

Tablo-8: Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2017 Verilerindeki Başarısı

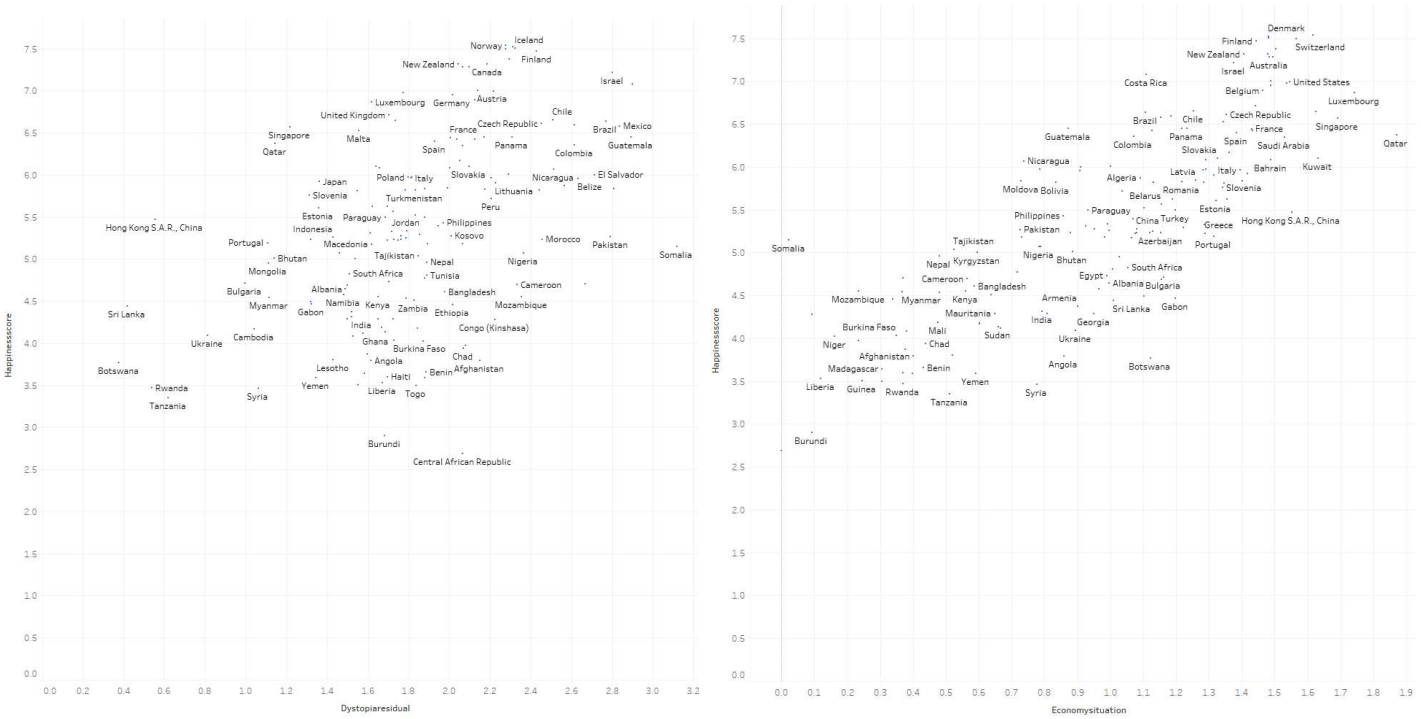
Tablo-8'den modelimizin çok başarılı olduğunu anlıyoruz. Kod çıktısı olarak R-squared ve Adj. R-squared değerleri(1.000 ve 1.000) Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2017 Verilerindeki Başarısı'nı göstermektedir. Bu değerlerin 1 veya 1'e yakın olması modelimizin çok başarılı olduğunu anlıyoruz. Bu nedenle, seçili Makine Öğrenimi yöntemimizi değiştirmiyoruz.

Variables	P-Value
constant	0.259
economysituation	0.000
family	0.000
healthlifeexpectancy	0.000
freedom	0.000
governmentcorruption	0.000
generosity	0.000
dystopiaresidual	0.000
terrorismevent	0.939

Tablo-9: Değişkenlerin Dünya Mutluluk Sıralaması 2017 Verileri Üzerindeki Başarısı

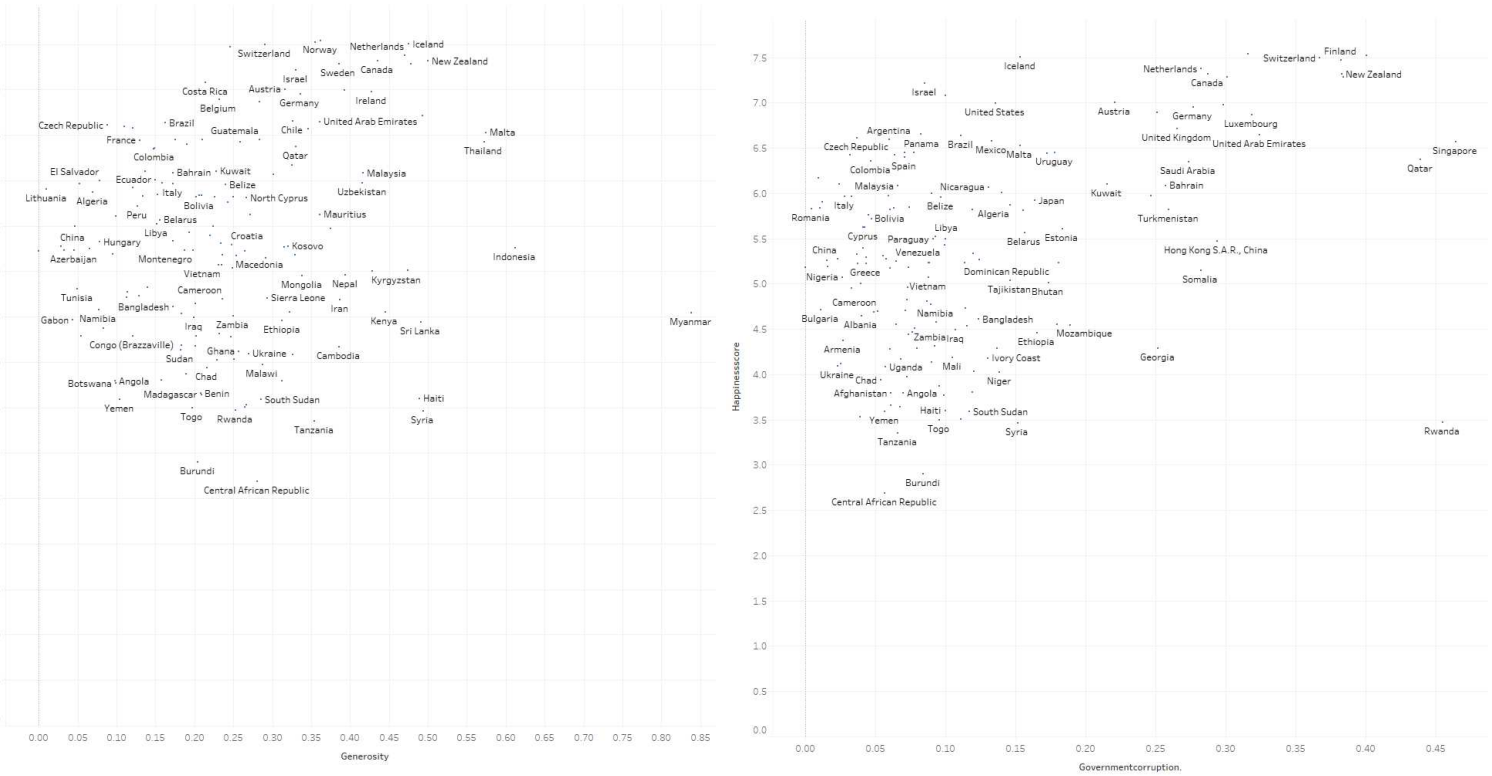
P>|t| değeri Değişkenlerin Dünya Mutluluk Sıralaması 2017 Verileri Üzerindeki Başarısı'nı göstermektedir. 1 olarak eklediğimiz sabit değer ve eklenen terör olayı değerlerimizin p değerlerinin %5 veya %1 eşik değerlerinden fazla olduğunu Tablo-9'dan anlıyoruz. Bu, terörizm olaylarının değerlerinin modelimiz için uygun ve etkilenebilir olmadığı anlamına gelir.

Yalnızca 2017 raporu için ülkelerin özellik değerlerinin mutluluk puanlarına ilişkin grafiklerini görmek için İş Zekâsı araçlarına sahip Tableau programını kullanıyoruz.

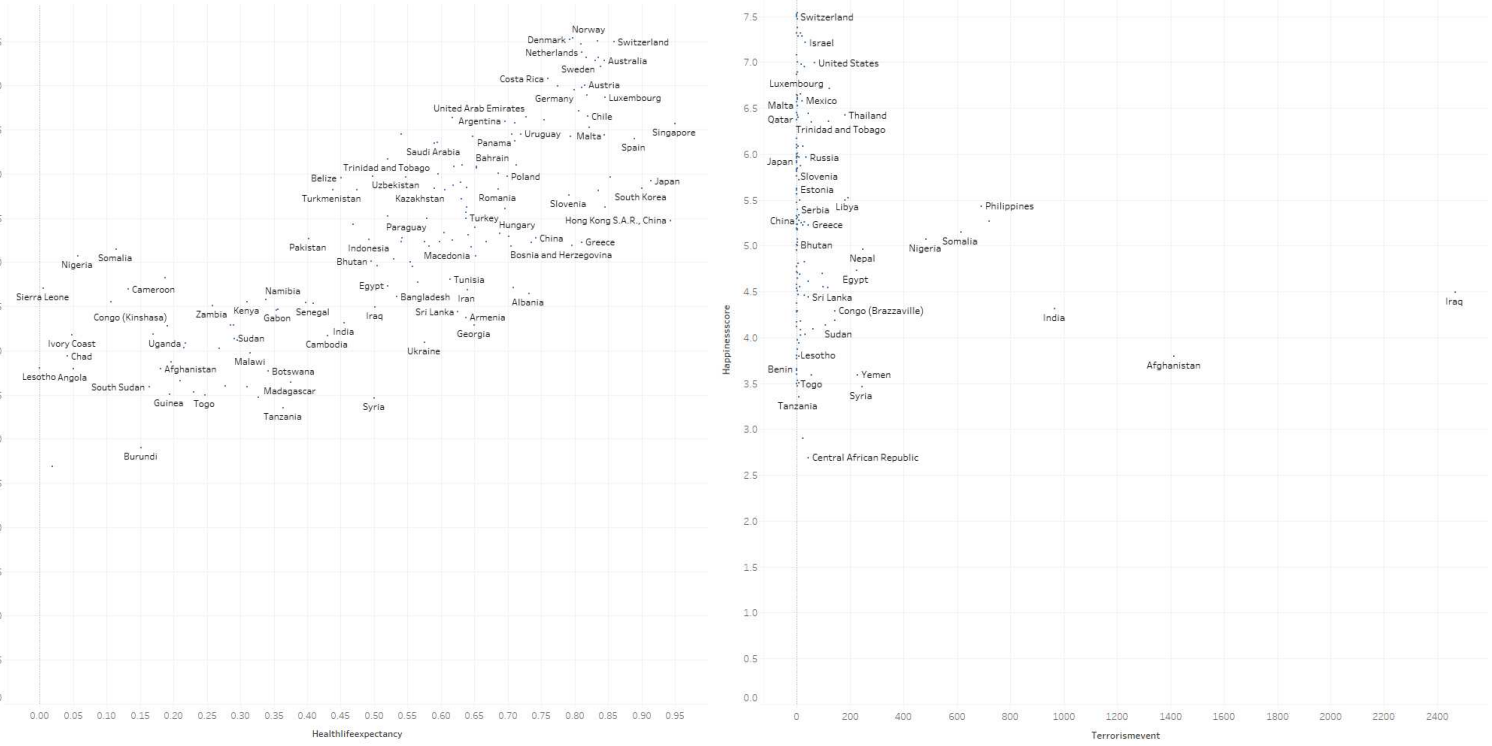


Grafik-1: Dystopia Residual and Economy to Happiness Skor Grafiği

Grafik-2: Family and Freedom to Happiness Skor Grafiği



Grafik-3: Generosity and Government Corruption to Happiness Skor Grafiği



Grafik-4: Health Life Expectancy and Terrorism Events to Happiness Skor Grafiği

World Happiness Rank 2018

```
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values=np.nan, strategy='constant',fill_value= 0)
impdata = data4.iloc[:,2:9].values
imputer = imputer.fit(impdata[:,2:9])
impdata[:,2:9] = imputer.transform(impdata[:,2:9])
data4.iloc[:,2:9] = impdata[:,:]
```

Kod çıktısı olarak eksik verilerin 0 olarak doldurulması sağlanmıştır.

Range Index: 156 entries, 0 to 155
Data columns (total 9 columns)
Country or region 156 non-null object
Score 156 non-null int64
GDP per capita 156 non-null float64
Social support 156 non-null float64
Healty life expectancy 156 non-null float64
Freedom to make life choices 156 non-null float64
Generosity 156 non-null float64
Perceptions of corruption 156 non-null float64

Tablo-10: World Happiness Rank 2018 Veri Özeti

```
print(data4.columns)

print(data4.info())

print(data4.describe())

df4 = data4["Country or region"]
dff4 = data4.value_counts()
```

2018 verilerimizin 156 ülke, boş olmayan değerlere sahip 7 özellik içerdiğini ve diğer tüm istatistiksel bilgilere kod çıktısından ulaşıyoruz.

```
x4 = data4.iloc[:,3:].values
y4 = data4.iloc[:,2:3].values
```

Kod çıktısı olarak x mutluluk puanını etkileyen girdiler ve y sonuç olan mutluluk puanı sütunu olarak x ve y değişkenine atanırlar.

```
from sklearn.model_selection import train_test_split
x_train4, x_test4, y_train4, y_test4 = train_test_split(x4, y4, test_size=0.33, random_state=0)
```

Kod çıktısı olarak test ve eğitim kümeleri oluşturulur.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train4 = sc.fit_transform(x_train4)
X_test4 = sc.fit_transform(x_test4)
Y_train4 = sc.fit_transform(y_train4)
Y_test4 = sc.fit_transform(y_test4)
```

Kod çıktısı olarak normalizasyon işlemi yapılır.

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train4, y_train4)
print("b0: ", lr.intercept_)
print("other b: ", lr.coef_)
```

Kod çıktısı olarak makine öğrenmesi modeli oluşturulur ve modele göre ağırlık çarpan değerleri hesaplanır.

```
import statsmodels.regression.linear_model as sm
X4 = np.append(arr = np.ones((156,1)).astype(int), values=x4, axis=1)
r_ols4 = sm.OLS(endog = y4, exog = X4)
r4 = r_ols4.fit()
print(r4.summary())
```

R squared:	0.789
Adj. R squared:	0.781

Tablo-11: Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2018 Verilerindeki Başarısı

Variables	P-Value
constant	0.000
GDP per capita	0.000
Social support	0.000
Healty life expectancy	0.015
Freedom to make life choices	0.000
Generosity	0.222
Perceptions of corruption	0.200

Tablo-12: Değişkenlerin Dünya Mutluluk Sıralaması 2018 Verileri Üzerindeki Başarısı

World Happiness Rank 2019

```
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values=np.nan, strategy='constant',fill_value= 0)
impdata = data5.iloc[:,3:9].values
imputer = imputer.fit(impdata[:,3:9])
impdata[:,3:9] = imputer.transform(impdata[:,3:9])
data5.iloc[:,3:9] = impdata[:,:]
```

Kod çıktısı olarak eksik verilerin 0 olarak doldurulması sağlanmıştır.

Range Index: 153 entries, 0 to 152
Data columns (total 20 columns)
Country name 153 non-null object
Regional indicator 153 non-null object
Ladder score 153 non-null float64
Standard error of ladder score 153 non-null float64
upperwhisker 153 non-null float64
lowerwhisker 153 non-null float64
Logged GDP per capita 153 non-null float64
Social support 153 non-null float64
Healthy life expectancy 153 non-null float64
Freedom to make life choices 153 non-null float64
Generosity 153 non-null float64
Perceptions of corruption 153 non-null float64
Ladder score in Dystopia 153 non-null float64
Explained by: Logged GDP per capita 153 non-null float64
Explained by: Social support 153 non-null float64
Explained by: Healthy life expectancy 153 non-null float64
Explained by: Freedom to make life choices 153 non-null float64
Explained by: Generosity 153 non-null float64
Explained by: Perceptions of corruption 153 non-null float64
Dystopia + residual 153 non-null float64

Tablo-13: World Happiness Rank 2019 Veri Özeti

```
print(data5.columns)

print(data5.info())

print(data5.describe())

df5 = data5["Country or region"]
dff5 = data5.value_counts()
```

2019 verilerimizin 156 ülke, boş olmayan değerlere sahip 7 özellik içerdiğini ve diğer tüm istatistiksel bilgilere kod çıktısından ulaşıyoruz.

```
x5 = data5.iloc[:,3:].values
y5 = data5.iloc[:,2:3].values
```

Kod çıktısı olarak x mutluluk puanını etkileyen girdiler ve y sonuç olan mutluluk puanı sütunu olarak x ve y değişkenine atanırlar.

```
from sklearn.model_selection import train_test_split
x_train5, x_test5, y_train5, y_test5 = train_test_split(x5, y5, test_size=0.33, random_state=0)
```

Kod çıktısı olarak test ve eğitim kümeleri oluşturulur.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train5 = sc.fit_transform(x_train5)
X_test5 = sc.fit_transform(x_test5)
Y_train5 = sc.fit_transform(y_train5)
Y_test5 = sc.fit_transform(y_test5)
```

Kod çıktısı olarak normalizasyon işlemi yapılır.

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train5, y_train5)
print("b0: ", lr.intercept_)
print("other b: ", lr.coef_)
```

Kod çıktısı olarak makine öğrenmesi modeli oluşturulur ve modele göre ağırlık çarpan değerleri hesaplanır.

```
import statsmodels.regression.linear_model as sm
X5 = np.append(arr = np.ones((156,1)).astype(int), values=x5, axis=1)
r_ols5 = sm.OLS(endog = y5, exog = X5)
r5 = r_ols5.fit()
print(r5.summary())
```

R squared:	1.000
Adj. R squared:	1.000

Tablo-14: Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2019 Verilerindeki Başarısı

Variables	P-Value
constant	0.194
Logged GDP per capita	0.081
Social support	0.524
Healty life expectancy	0.714
Freedom to make life choices	0.207
Generosity	0.181
Perceptions of corruption	0.172
Ladder score in Dystopia	0.194
Explained by: Logged GDP per capita	0.576
Explained by: Social support	0.000
Explained by: Healthy life expectancy	0.410
Explained by: Freedom to make life choices	0.605
Explained by: Generosity	0.996
Explained by: Perceptions of corruption	0.959
Dystopia + residual	0.000

Tablo-15: Değişkenlerin Dünya Mutluluk Sıralaması 2019 Verileri Üzerindeki Başarısı

World Happiness Rank 2020

```
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values=np.nan, strategy='constant',fill_value= 0)
impdata = data6.iloc[:,4:20].values
imputer = imputer.fit(impdata[:,4:20])
impdata[:,4:20] = imputer.transform(impdata[:,4:20])
data6.iloc[:,4:20] = impdata[:,:]
```

Kod çıktısı olarak eksik verilerin 0 olarak doldurulması sağlanmıştır.

Range Index: 156 entries, 0 to 155
Data columns (total 9 columns):
Country or region 156 non-null object
Score 156 non-null int64
GDP per capita 156 non-null float64
Social support 156 non-null float64
Healty life expectancy 156 non-null float64
Freedom to make life choices 156 non-null float64
Generosity 156 non-null float64
Perceptions of corruption 156 non-null float64

Tablo-16: World Happiness Rank 2020 Veri Özeti

```
print(data6.columns)

print(data6.info())

print(data6.describe())

df6 = data6["Country name"]
dff6 = data6.value_counts()
```

2020 verilerimizin 153 ülke, boş olmayan değerlere sahip 16 özellik içerdiğini ve diğer tüm istatistiksel bilgilere kod çıktısından ulaşıyoruz.

```
x6 = data6.iloc[:,4:].values
y6 = data6.iloc[:,2:3].values
```

Kod çıktısı olarak x mutluluk puanını etkileyen girdiler ve y sonuç olan mutluluk puanı sütunu olarak x ve y değişkenine atanırlar.

```
from sklearn.model_selection import train_test_split
x_train6, x_test6, y_train6, y_test6 = train_test_split(x6, y6, test_size=0.33, random_state=0)
```

Kod çıktısı olarak test ve eğitim kümeleri oluşturulur.

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train6 = sc.fit_transform(x_train6)
X_test6 = sc.fit_transform(x_test6)
Y_train6 = sc.fit_transform(y_train6)
Y_test6 = sc.fit_transform(y_test6)
```

Kod çıktısı olarak normalizasyon işlemi yapılır.

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train6, y_train6)
print("b0: ", lr.intercept_)
print("other b: ", lr.coef_)
```

Kod çıktısı olarak makine öğrenmesi modeli oluşturulur ve modele göre ağırlık çarpan değerleri hesaplanır.

```
import statsmodels.regression.linear_model as sm
X6 = np.append(arr = np.ones((153,1)).astype(int), values=x6, axis=1)
r_ols6 = sm.OLS(endog = y6, exog = X6)
r6 = r_ols6.fit()
print(r6.summary())
```

R squared:	0.779
Adj. R squared:	0.770

Tablo-17: Makine Öğrenimi Yönteminin Dünya Mutluluk Sıralaması 2020 Verilerindeki Başarısı

Variables	P-Value
constant	0.000
GDP per capita	0.001
Social support	0.000
Healty life expectancy	0.002
Freedom to make life choices	0.000
Generosity	0.327
Perceptions of corruption	0.075

Tablo-18: Değişkenlerin Dünya Mutluluk Sıralaması 2020 Verileri Üzerindeki Başarısı

Terrorism Report 2015

RangeIndex: 14963 entries, 0 to 14962
Data columns (total 2 columns):
year 14963 non-null int64
country 14963 non-null object

Tablo-19: Terörizm Raporu 2015 Verilerinin Özeti

```
print(data7.columns)

print(data7.info())

print(data7.describe())

df7 = data7["country"]
dff7 = df7.value_counts()
print(dff7)
```

Tablo-19'dan ve kod çıktısından, 2015 Terör Raporu verilerimizin yıl ve 14963 sıfır olmayan değeri olan ülkeler olduğunu anlıyoruz.

Country	Terrorism Counts
Iraq	2750
Afghanistan	1928
Pakistan	1243
India	884
Philippines	721

Tablo-20: Terörizm Raporu 2015 Verilerine İlişkin Terör Sayılarından Örnekler

Örnek olarak ülkeler terörizm sayılarına göre sıralanıyor. 2015 rapor modelimize tüm değerler yeni bir özellik olarak eklenmiştir.

Terrorism Report 2016

RangeIndex: 13587 entries, 0 to 13586
Data columns (total 2 columns):
year 13587 non-null int64
country 13587 non-null object

Tablo-21: Terörizm Raporu 2016 Verilerinin Özeti

```
print(data8.columns)

print(data8.info())

print(data8.describe())

df8 = data8["country"]
dff8 = df8.value_counts()
print(dff8)
```

Tablo-21'den ve kod çıktısından, 2016 Terör Raporu verilerimizin yıl ve 13587 sıfır olmayan değeri olan ülkeler olduğunu anlıyoruz.

Country	Terrorism Counts
Iraq	3360
Afghanistan	1617
India	1025
Pakistan	864
Philippines	632
Somalia	602
Turkey	542

Tablo-22: Terörizm Raporu 2016 Verilerine İlişkin Terör Sayılarından Örnekler

Örnek olarak ülkeler terör sayılarına göre sıralanıyor. 2016 rapor modelimize tüm değerler yeni bir özellik olarak eklenmiştir.

Terrorism Report 2017

RangeIndex: 10900 entries, 0 to 10899
Data columns (total 2 columns):
year 10900 non-null int64
country 10900 non-null object

Tablo-23: Terörizm Raporu 2017 Verilerinin Özeti

```
print(data9.columns)

print(data9.info())

print(data9.describe())

df9 = data9["country"]
dff9 = df9.value_counts()
print(dff9)
```

Tablo-23'ten ve kod çıktısından, Terör Raporu 2017 verilerimizin yıl ve 10900 sıfır olmayan değeri olan ülkeler olduğunu anlıyoruz.

Country	Terrorism Counts
Iraq	2466
Afghanistan	1414
India	966
Pakistan	719
Philippines	692
Somalia	614

Tablo-24: Terörizm Raporu 2017 Verilerine İlişkin Terör Sayılarından Örnekler

Örnek olarak ülkeler terör sayılarına göre sıralanıyor. Tüm değerler 2017 rapor modelimize yeni bir özellik olarak eklenmiştir.

6. SONUÇ

2015, 2016 ve 2017 Dünya Mutluluk Raporlarında, Küresel Terörizm Raporu'nu kullanarak Dünya Mutluluk Raporu'na hangi faktörün ne kadar etki ettiğini ve terör olaylarının bu sıralamaya etkisini takip ettik. Analizin sonunda, bu 3 raporda Makine Öğrenimini kullanarak mantıksal rastgele değerlerle mutluluk puanını tahmin ediyoruz ve bunların gerçekten işe yaradığını gördük. Modelimiz için Çok Doğrusal Regresyon Makine Öğrenme Metodunun iyi bir seçenek olduğunu gördük. Terör olaylarının değerlerinin mutluluk puanlarından etkilenmediğini, ancak etkisinin 2016'da arttığını gördük. 2018, 2019 ve 2020 verileri için Terörizm verileri olmadığından yalnızca kendi verileriyle analizi yapıldı.

7. KAYNAKLAR

[1] Louise Millard (2011) Data Mining and Analysis of Global Happiness: A Machine Learning Approach

[2] Natasha Jaques, Sara Taylor, Asaph Azaria, Asma Ghandeharioun, Akane Sano, Rosalind Picard (2015) Predicting Students' Happiness from Physiology, Phone, Mobility, and Behavioral Data