ABSTRACT

İnsanların ne kadar yaşayacaklarını bilmeleri, hayatlarını daha iyi planlamalarına yardımcı olur Ekonomik ,sosyolojik ve Toplumsal dinamikler de tamamen bu konu cevresinde gelisir

Yasam beklentisi ; Devlet Politikalari ,Saglik ve Hastane Sektoru , **Emeklilik ve Sigortacılık Sektörü , Uzun Ölçekli Finans ve Bankacılık İşlemleri** icin hayati onem tasir.

Bu durumdan dolayi verisetimiz: Ulkelerin yil bazinda Hayat beklentisi uzerine olmustur.

Bu calismada Ulkelere gore yasam beklentisi’ni tahmin etmek icin makine ogrenimi algoritmalari uyguladik

Makine ogrenimi algoritmalarimizi secerken ,,,, dikkat ettik

Kullandiigmiz algoritmalar

Decision tree….

Random forest gibi etkili olan classifier supervised algoritmalari kullanildi

En iyi model 99% dogruluk ile Decision Tree algoritmasi elde edildi

Giris

Verisetini ilk elimize aldiigmzda

INTRODUCTION

Insanoglu dunyaya geldiklerinden beri yasami ,olumu kac yil yasayacagini ve nasil daha fazla yasayacagini arastirip ne kadar yasayacagi gibi bir belirsizligi ortadan kaldirmaya calismistir. Bunlara sifali icecekler, yabani otlar’dan baslayip penisilin ve suanki gunumuzde yapay zeka destekli dna gen dizilimine ozel ilaclara kadar gelmistir.

Bizim bu veriyi secmemizin asil ve insani tarafi bunlar iken gunumuzun sehir hayatinda ve modern insani ilgilendiren taraflari da basli basina cok buyuk bir muammadir. Ornegin :Devlet; kamu çalışan sayısı, emeklilik yaşı,emeklilik ikramiyesi ve primi gibi konularda ortalama hayat beklentisini göz önünde bulundururak politikalarını belirleyebilir.bu sayede sistem bozulmaz ve insanoglunun olusturdugu modern kapitalizm carki donmeye devam edebilir. Bu konuya iliskin benzer konular **Sağlık ve Hastane Sektörü Bankacılık İşlemleri iliskilendirilebilir.**

“Bu calismanin amaci makine ogrenmesi modelleri kullanarak kullandigimiz algoritmalarin yasam beklentisi tahmin performansini karsilastirmak ve gelistirmektir.

Hayat beklentisi etkileyen faktorler

“İnsan yaşam beklentisini etkileyen birçok faktör üzerine çeşitli çalışmalar yapılmıştır. Yaşam süresiyle ilişkili olduğu belirlenen bazı unsurlar, çevresel, sosyoekonomik davranislaridir Monsef ve Mehrjardi(2015) yaşam süresini sosyal, ekonomik ve çevresel faktörler açısından incelemiştir. Araştırmaları, işsizlik ve enflasyonun yaşam süresi üzerinde olumsuz etkisi olan temel ekonomik faktörlerin ne olduğudur,Daha iyi bir maddi durumun daha yüksek yaşam kalitesi sağladığı düşünülmekte ve bu nedenle finansal durum ve eğitim, yaşam süresini olumlu yönde etkileyen faktörler arasındadir.Pek cok ulkede, yapılan araştırmalar, kadınların erkeklerden daha uzun yaşadığını ortaya koymuştur

### YASAM BEKLENTISI ICIN VERI ONISLEME

|  |  |
| --- | --- |
| Country | Country |
| Year | Year |
| **Status** | Developed or Developing or Undeveloped |
| Infant deaths | umber of Infant Deaths per 1000 population |
| Alcohol | Number of Infant Deaths per 1000 population |
| Percentage expenditure |  |
| Hepatitis B | Hepatitis B (HepB) immunization coverage among 1-year olds (% |
| Measles | Measles - number of reported cases per 1000 population |
| BMI | Average Body Mass Index of the entire population |
| Under-five deaths | Number of under-five deaths per 1000 population |
| Polio | olio (Pol3) immunization coverage among 1-year-olds (%) |
| Total expenditure | eneral government expenditure on health as a percentage of total government expenditure (%) |
| Diphteria | iphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%) |
| HIV/AIDS | eaths per 1 000 live births HIV/AIDS (0-4 years) |
| GPD | Gross Domestic Product per capita (in USD) |
| Population | The population of the country |
| Thinness 5-9 years | Prevalence of thinness among children and adolescents for Age 10 to 19 (%) |
| Income composition of resources | Human Development Index in terms of income composition of resources (index ranging from 0 to 1) |
| Schooling | Number of years of Schooling(years) |
| **Life expectancy** | ife expectancy in age |

Elimizdeki verileri incelerken yapmamiz gereken on isleme islemleri vardi bunlar gerektigi anlarda : data cleaning, data transform , data drop ,,,, gibi islemler

Yukaridaki tabloda raw verilerimiz (5) ve ne olduklari gosterilmistir. Kalin ile yazili olan kolonlar bizim makine ogrenimi modellerini kullanarak sonuc almayi bekledigimiz y ciktilarimizdir. Alti cizili ile gosterilen features verilerimizde sapma miktarini minimuma indirmek adina belirli matematiksel modeler kullanarak olusturuldu

## Makine Ogrenimi Algoritmalari

Verilerimizi makine ogrenimi algoritmalarina sokmadan once sayisal veriler oldugundan emin olmaliyiz.string degerler birakmamaliyiz ve bazi kolon ikililerinin birbirine tam olarak dogrusal veya tersine olarak carpismasi yani korelasyon carpismasi gibi bir durumun yaratacagi overfitting ve underfitting olaylarini engellememiz gerekli

Burada kullandigimiz bazi modellerin nasil calistigi basitce modellenmistir(2)

A diagram of a triangle

Description automatically generated A graph with blue dots and a red line

Description automatically generated

### A1-Decision Tree

Bir karar ağacı, kararların bir ağacını oluşturan basit ve faydalı bir algoritmadır. Hem sınıflandırma hem de regresyon analizinde kullanılır. Bu yöntem, bir veri kümesini kökten yaprak düğümlere doğru büyüyen bir ağaç yapısına bölmeyi içerir. Her yaprak düğümünün durumu kök düğümüne bağlıdır ve yaprak düğüm koşulu sağlandığında büyür. Bu algoritmanın amacı, verilen veri kümesinden karar kurallarını öğrenerek gelecekteki değerleri tahmin etmektir. Sınıflandırmada, bir karar ağacı entropi ve bilgi kazancı kullanılarak oluşturulur. Entropi, ayrıca Shannon entropisi olarak da bilinir ve sonlu bir S kümesi için H(S) ile gösterilir. Verinin homojenliğini (belirsizlik veya rastgelelik) ölçer. Entropi çıkış değeri 0 ile 1 arasında değişir, burada 0 tamamen homojen veriyi temsil eder ve 1 eşit olarak bölünmüş veriyi temsil eder.

### A2-Random forest,

birçok karar ağacının bir araya gelerek bir model oluşturduğu bir ensemble (bir araya getirme) öğrenme tekniğidir. Her bir karar ağacı, rastgele özelliklerin bir alt kümesini kullanarak eğitilir ve ardından bu ağaçlar bir araya gelerek tahminler yapar. Bu yöntem, overfitting'i azaltır ve genel olarak daha iyi bir tahmin yapma yeteneğine sahiptir. Random forest, sınıflandırma ve regresyon problemleri için yaygın olarak kullanılır

### B1-Linear Regression

Doğrusal regresyon, bağımlı ve bağımsız değişkenler arasındaki istatistiksel ilişkiyi bulmak için kullanılır. Bağımsız değişken tahmin edici, bağımlı değişken ise yanıt değişkenidir ve hata, düzlemden uzaklık olarak ifade edilir. Amacı, verileri en iyi şekilde temsil eden bir çizgi çizmektir. Bu, bir değişkenin diğer bir değişkenle tam olarak ifade edilemediği durumlarda kullanılır, örneğin boy ve sıcaklık arasındaki ilişki gibi.

### B2-Ridge Regression

Ridge regresyon, lineer regresyonun bir türevidir ve genellikle aşırı uyum (overfitting) problemlerini ele almak için kullanılır. Temel amaç, katsayıların büyüklüğünü sınırlamak ve bu şekilde modelin karmaşıklığını azaltmaktır. Ridge regresyonu, toplam kare hatasına bir ceza terimi ekleyerek çalışır. Bu ceza terimi, katsayıların karelerinin toplamıdır. Bu sayede, model aşırı uyum yapma eğiliminde olan katsayıları kontrol altına alır.

### B3-Lasso Regression

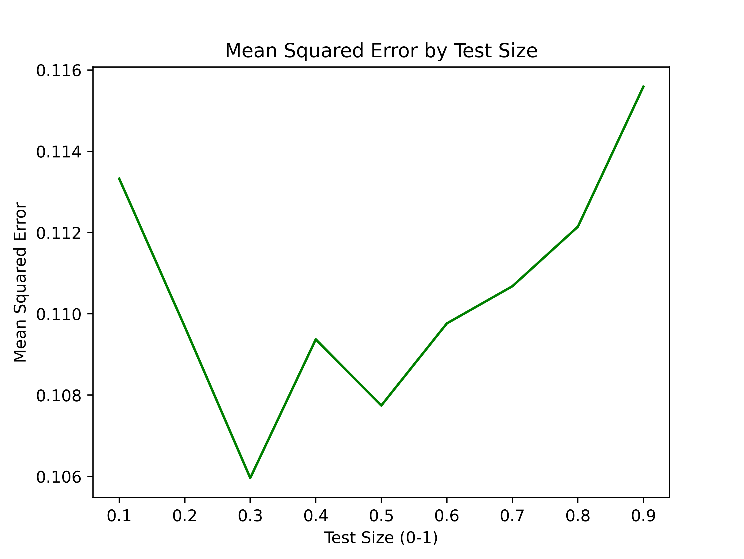
Lasso regresyonu, ridge regresyonuna benzer bir şekilde lineer regresyonun bir türevidir ve aşırı uyumu azaltmak için kullanılır. Ancak, lasso regresyonu, katsayıların toplam mutlak değerini sınırlayarak çalışır. Bu, gereksiz özellikleri belirleme ve modelin karmaşıklığını azaltma eğilimindedir. Lasso regresyonu, toplam kare hatasına bir ceza terimi ekleyerek çalışır, ancak bu ceza terimi katsayıların mutlak değerlerinin toplamı

## Makine ogrenimi algoritmalari ile tahmin

Makine Ogrenimi algoritmalari with supervising kategorisi icerisinde kendi icerisinde 2’ye ayrilir bunlar Regression ve Classificationdir. Bizim output almak istedigimiz bagimli degisken y “age” kolonu oldugu icin ve bu veri surekli (continuous) veri oldugu icin regression modelleri tercih etmek dogru olacaktir. 2.Senaryomuzda y outputumuz “status” kolonu sectigimizde ciktimizin 3 adet kategorik veri oldugunu bu yuzden de classification makine ogrenimi supervised modellerini kullanmamiz gerektigini goruruz.

Sirasiyla Linear regression,Ridge ve lasso,nun dogruluk degerlerini MSE ve R^2 metrigini baz alarak cizdiren fonksiyon gosterilmistir.

### 1-Linear regression



A red line graph with numbers

Description automatically generated

### 2-Ridge Regression

A green line graph with numbers and numbers

Description automatically generated

A graph with a red line

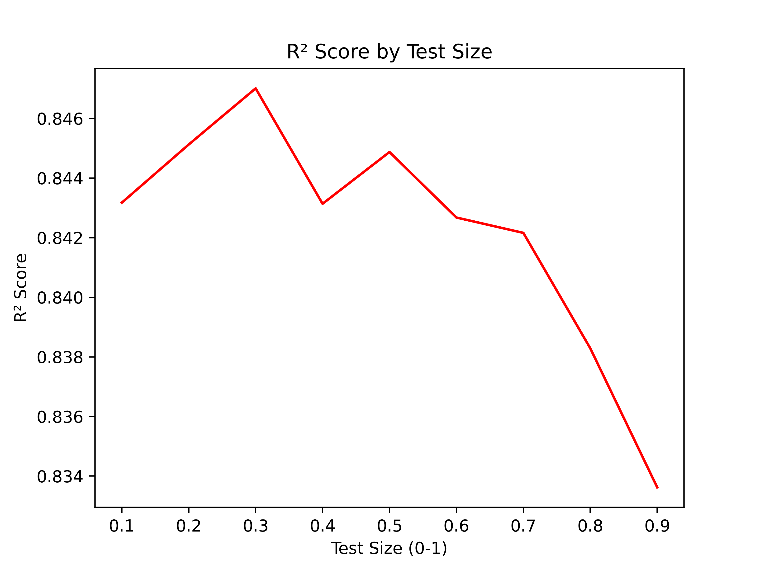
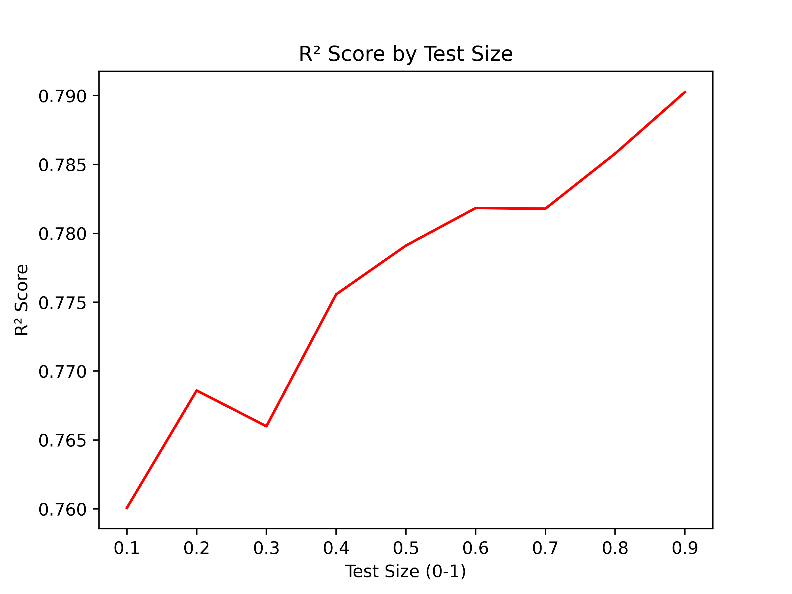
Description automatically generated

3-Lasso Regression

RESIM1 -RESIM2 YUKLENECEKTIR

### GELISTIRME

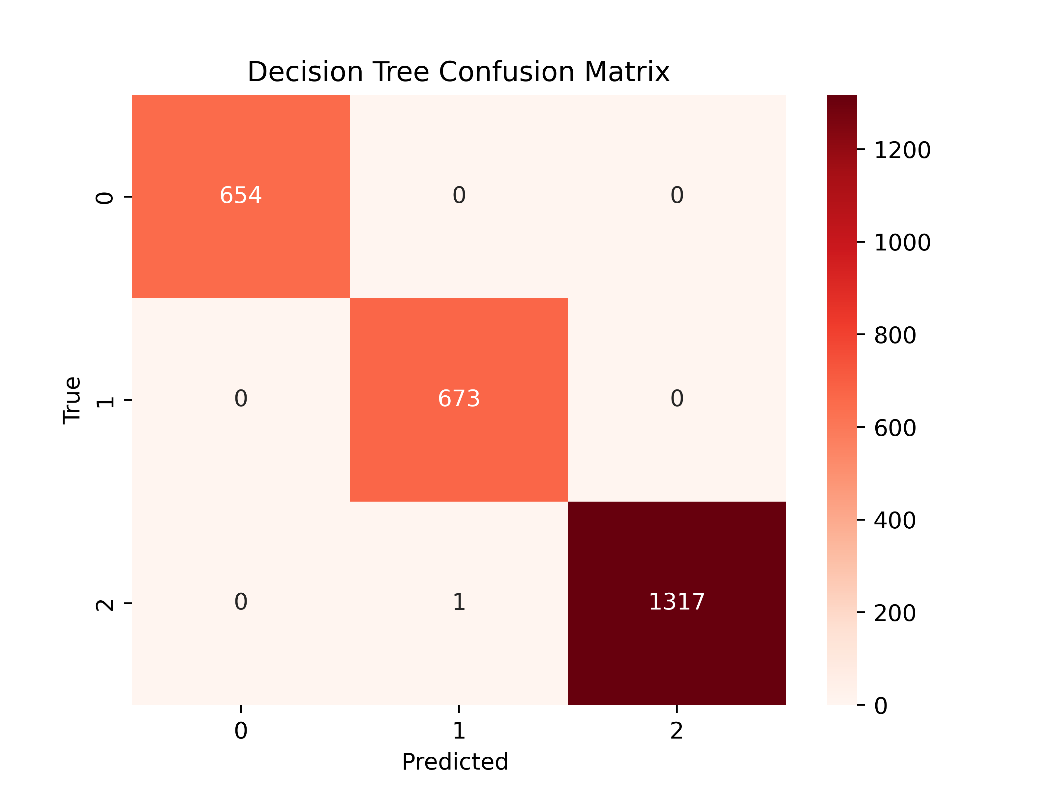
Verilerimizin egitirken farkettigimiz train\_test\_split methodundaki test size ve randomstate’I en optimal deger aldigimizda algoritmamizin gozle gorulur derecede (MSE :30->8) gibi dogruya yakinsidigini gorduk. En iyi parametreleri bulabilmek adina test\_size 0-1 araliginda 0.1 range degeri ile random state ise 1-ile 100 degerleri araliginda +8 range degeri ile en optimal parametreleri bularak sistemimizi yuzdelik olarak ortalama 35% oraninda guclendirmis olduk.Bu kodlama islemini tum modellerimiz icin (classificationve regression) uyguladiktan sonra en optimaller icin grafiklerini cikarttik(MSE R2)(5)



### VERIMIZIN KONTROLU

Modelimizden guzel ve yuksek accuracy oranlari elde ettik. Bu durumda olasi bir overfitting soz konusu mu bunu gormek icin confusion matrix gibi yontemleri kullandik.Bu kontrolu classification targetimiz olan “status”icin baktik.

### Decision Tree Confusion Matrix



Random Forest Confusion Matrix

A graph with numbers and squares

Description automatically generated

Tahminlerimizde veri setimizi train\_test\_split fonksiyonu ile train ve test olarak belli oranda ayirdik(3) bu fonksiyondaki hyperparametreler : testsize ve traintestsplit default olarak verisetimize uyguladik.

A blue and yellow squares with black text

Description automatically generated

“age” kolonumuz icin farkli regresyon modellerinde birbirine yakin fakat farkli cevaplar aldik.(4)

Tablomuza baktigimizda ….. Goruldugu uzere …. Degerlerimizde overfitting vb bulunmamaktadir.

### SONUC

VERIMIZI GIRDIK SAYIALRI DENEDIK VE SU ULKENIN YASINI BILDI.

.

.

.

SONRASINDA TURKIYE ORNEGINDEN YAPTIK GDP VE YILI DEGISTRIDIK VE 2023U BILMESINI ISTEDIK. BIZE BIR DEGER VERDI 2023 YASAM BEKLENTISINE BAKTIGIMIZDA 75.5 OLDUGUNU GORDUK[2]

KULLANDIGIM KAYNAKLAR

[train\_test\_split — scikit-learn 1.5.0 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)

[GRADU-1498472565.pdf (tuni.fi)](https://trepo.tuni.fi/bitstream/handle/10024/101646/GRADU-1498472565.pdf?sequence=1&isAllowed=y)

[Stock-Price-Forecast-Using-Recurrent-Neural-Network.pdf (researchgate.net)](https://www.researchgate.net/profile/Noman-Javed-2/publication/333059580_Stock_Price_Forecast_Using_Recurrent_Neural_Network/links/5cda0b1d458515712ea94121/Stock-Price-Forecast-Using-Recurrent-Neural-Network.pdf#page=66)

[View of PREDICTION OF LIFE EXPECTANCY FOR ASIAN POPULATION USING MACHINE LEARNING ALGORITHMS (mohe.gov.my)](https://myjms.mohe.gov.my/index.php/mjoc/article/view/18218/10716)

HANGISINI NEREDEN ALDIGIMI SONRADAN EKLEYECEGIM.

EN IYI SONUC RESIMLERI KODDAN INDIREBILIRSIN

True false varsa tablo cikart

[1]

Pisal, Nurul Shahira, et al. "Prediction of life expectancy for Asian population using machine learning algorithms." *Malaysian Journal of Computing* 7.2 (2022): 1150-1161.