Final Projesi

Öğretim Üyesi: Prof.Dr.MEHMET ERDAL BALABAN

Dersin Adı ve Kodu: VERİ ANALİTİĞİ (YBS317)

Proje Üyeleri: Furkan Kütük

Halil İbrahim Altuntaş

Kaan Hacıömeroğlu

Proje Konu Başlığı: Meme Kanserinin İyi Huylumu Kötümü Olduğunu Tahmin Etmek

Veri Kaynağı ve Adı: Breast Cancer Wisconsin (Diagnostic) Data Set | Kaggle

Özet

Verideki özellikler, bir göğüs kitlesinin ince iğne aspirasyonunun (FNA) dijitalleştirilmiş bir görüntüsünden hesaplanmış verilerden oluşmaktadır.**(Biyopsiden alınan parçanın sayısal verilere dönüştürülmesiyle)** Meme kanserin iyi huylu veya kötü huylu olduğunu tahmin etme projesidir. Projede makina ögrenmesi modelleriyle tahmin edip modellerden hangisinin daha performanslı olduğunu tespit edip o modelli kullanarak test verisinden tahmin yapar.

Giriş

Meme kanseri, meme dokusunu oluşturan hücre gruplarından birinin değişime uğraması ve kontrolsüz olarak çoğalması nedeniyle oluşan tümör sonucu ortaya çıkan bir hastalıktır.

Memede kitle görülmesi sık rastlanılan bir bulgudur. Tespit edilen her kitle kanser anlamına gelmemektedir. Kitleler iyi huylu (benign) ve kötü huylu (malign) olmak üzere ikiye ayrılmaktadır. Memede kitle saptandığında mutlaka araştırılması gerekir. Bu araştırma yöntemi ise İnce iğne aspirasyonu (FNA) ile bulunabilir. İnce iğne aspirasyonu (FNA), bir kist, enfeksiyon, iyi huylu bir tümör veya kanser olup olmadığını belirlemek için bir göğüs yumrağından bir hücre örneği almak için kullanılan bir prosedürdür.

Meme kanseri çevresel ve genetik faktörlerin birleşimi ile oluşur. Genetik faktörler (%10)arasında ailede meme kanseri olması, çevresel faktörler (%90) arasında da birçok etken bulunmaktadır.

Meme kanseri genellikler kadınlara özgü bir hastalık olarak görülse de erkeklerde de %1 oranında görülmektedir. Her yıl yaklaşık 300 erkeğe meme kanseri tanısı konulmaktadır.

Problemi Tanımlama ve Anlama

Meme kanserine ilişkin veri seti üzerinde öğrenme algoritmaları kullanarak bir kişinin meme kanserinde iyi huylu ve kötü huylu olup olmadığını öngörebilmek bu çalışmanın ana amacını oluşturmaktadır

Breast Cancer Wisconsin (Diagnostic) Data Set | Kaggle bu kaynaktan aldığımız veri setine göre iyi huylu ve kötü huylu olma oranları ;

İyi Huylu %62.7 Kötü Huylu % 37.3 'dir.

Veri Anlama ve Özetleme

	Tahmin İçin Kullanılan Nitelikler		
	Nitelik	Türkçesi	Veri Tipi
1	Diagnisis	Tanı	Character
2	Radius Mean	Yarı çap	Numeric
3	Texture Mean	Doku	Numeric
4	Perimeter Mean	Çevre	Numeric
5	Area Mean	Alan	Numeric

Veri setinde kullandığımız verilerin öz nitelikleri bunlardır.Bunun dışında kullandığımız verilerde aynı şekilde numeric verilerdir.

	Hedef Nitelik		
1	Breast Cancer	Meme kanseri	ikili 1:M 2:B

###VERİYİ ANLAMA Kod Satırları

```
> library(readr)
> prc <- read.csv("kanser.csv", stringsAsFactors = FALSE)
> View(prc)
> stringsAsFactors = FALSE    #Stringi factore çeviriyor
> str(prc)          #Verinin yapılandırılmış olup olmadığını kontrol etmek için
'data.frame':   569 obs. of  33 variables:

> prc <- prc [-1]       #gereksiz sütunları kaldırma
> prc <- prc [-1]
> table(prc$diagnosis)     # iyi huylu ve kötü huyludan kaç tane oldugunu sorduk

  B   M
357 212
> prc$diagnosis <- factor(prc$diagnosis, levels = c("B","M"), labels = c("Iyi Huylu","Kotu Huylu"))
> round(prop.table(table(prc$diagnosis)) * 100, digits = 1)

 Iyi Huylu Kotu Huylu
      62.7       37.3
```

Veri Hazırlama ve Ön İşlemler

Eksik Değer Sorgulama

```
> is.na(prc$diagnosis)
  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [27] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [53] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [79] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[105] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[131] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[183] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[209] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[235] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[261] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[287] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[339] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[365] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[391] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[417] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[443] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[469] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[495] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[521] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[547] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

Diagnosis tablosundaki eksik veriler olup olmadığını kontrol ediyoruz bu verileri TRUE ve FALSE olarak yazdırıyoruz

TRUE= Null değer var FALSE = Null değer yok

Dummy

```
> table(prc$diagnosis )

 Iyi Huylu Kotu Huylu
       357         212
> dummy_tani <- as.data.frame(model.matrix(~ 0 + diagnosis , data = prc))
> dummy <- cbind(prc$diagnosis , dummy_tani)
> head(dummy)    #Dummy verilerimizdeki İyi huylu ve Kötü huylu değişkenlerini 0 , 1 olarak yazdırır
  prc$diagnosis diagnosisIyi Huylu diagnosisKotu Huylu
1    Kotu Huylu                  0                   1
2    Kotu Huylu                  0                   1
3    Kotu Huylu                  0                   1
4    Kotu Huylu                  0                   1
5    Kotu Huylu                  0                   1
6    Kotu Huylu                  0                   1
```

Normalize

```
> normalize <- function(x) { return ((x - min(x)) / (max(x) - min(x))) }
> prc_n <- as.data.frame(lapply(prc[2:31], normalize))
> summary(prc_n$radius_mean)       ##RADIUS_MEAN tablosunun özetini alır
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.2233  0.3024  0.3382  0.4164  1.0000
> summary(prc_n)       ##normalize değerlerinin özetini verir
  radius_mean      texture_mean    perimeter_mean     area_mean      smoothness_mean  compactness_mean concavity_mean    concave.points_mean symmetry_mean
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000      Min.   :0.0000
 1st Qu.:0.2233   1st Qu.:0.2185   1st Qu.:0.2168   1st Qu.:0.1174   1st Qu.:0.3046   1st Qu.:0.1397   1st Qu.:0.06926   1st Qu.:0.1009      1st Qu.:0.2823
 Median :0.3024   Median :0.3088   Median :0.2933   Median :0.1729   Median :0.3904   Median :0.2247   Median :0.14419   Median :0.1665      Median :0.3697
 Mean   :0.3382   Mean   :0.3240   Mean   :0.3329   Mean   :0.2169   Mean   :0.3948   Mean   :0.2606   Mean   :0.20806   Mean   :0.2431      Mean   :0.3796
 3rd Qu.:0.4164   3rd Qu.:0.4089   3rd Qu.:0.4168   3rd Qu.:0.2711   3rd Qu.:0.4755   3rd Qu.:0.3405   3rd Qu.:0.30623   3rd Qu.:0.3678      3rd Qu.:0.4530
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000      Max.   :1.0000
 fractal_dimension_mean   radius_se         texture_se      perimeter_se       area_se         smoothness_se     compactness_se     concavity_se
 Min.   :0.0000         Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
 1st Qu.:0.1630         1st Qu.:0.04378   1st Qu.:0.1047   1st Qu.:0.04000   1st Qu.:0.02064   1st Qu.:0.1175   1st Qu.:0.08132   1st Qu.:0.03811
 Median :0.2439         Median :0.07702   Median :0.1653   Median :0.07209   Median :0.03311   Median :0.1586   Median :0.13667   Median :0.06538
 Mean   :0.2704         Mean   :0.10635   Mean   :0.1893   Mean   :0.09938   Mean   :0.06264   Mean   :0.1811   Mean   :0.17444   Mean   :0.08054
 3rd Qu.:0.3404         3rd Qu.:0.13304   3rd Qu.:0.2462   3rd Qu.:0.12251   3rd Qu.:0.07170   3rd Qu.:0.2187   3rd Qu.:0.22680   3rd Qu.:0.10619
 Max.   :1.0000         Max.   :1.00000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
 concave.points_se  symmetry_se     fractal_dimension_se  radius_worst    texture_worst    perimeter_worst    area_worst      smoothness_worst compactness_worst
 Min.   :0.0000     Min.   :0.0000   Min.   :0.00000       Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:0.1447     1st Qu.:0.1024   1st Qu.:0.04675       1st Qu.:0.1807   1st Qu.:0.2415   1st Qu.:0.1678   1st Qu.:0.08113   1st Qu.:0.3000   1st Qu.:0.1163
 Median :0.2070     Median :0.1526   Median :0.07919       Median :0.2504   Median :0.3569   Median :0.2353   Median :0.12321   Median :0.3971   Median :0.1791
 Mean   :0.2235     Mean   :0.1781   Mean   :0.10019       Mean   :0.2967   Mean   :0.3640   Mean   :0.2831   Mean   :0.17091   Mean   :0.4041   Mean   :0.2202
 3rd Qu.:0.2787     3rd Qu.:0.2195   3rd Qu.:0.12656       3rd Qu.:0.3863   3rd Qu.:0.4717   3rd Qu.:0.3735   3rd Qu.:0.22090   3rd Qu.:0.4942   3rd Qu.:0.3025
 Max.   :1.0000     Max.   :1.0000   Max.   :1.00000       Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
 concavity_worst    concave.points_worst symmetry_worst   fractal_dimension_worst
 Min.   :0.00000    Min.   :0.0000       Min.   :0.0000   Min.   :0.0000
 1st Qu.:0.09145    1st Qu.:0.2231       1st Qu.:0.1851   1st Qu.:0.1077
 Median :0.18107    Median :0.3434       Median :0.2478   Median :0.1640
 Mean   :0.21740    Mean   :0.3938       Mean   :0.2633   Mean   :0.1896
 3rd Qu.:0.30583    3rd Qu.:0.5546       3rd Qu.:0.3182   3rd Qu.:0.2429
 Max.   :1.00000    Max.   :1.0000       Max.   :1.0000   Max.   :1.0000
> prc_train <- prc_n[1:350,]            ## knn modeli için test tablosu ve öğrenme tablosu oluşturma
> prc_test <- prc_n[351:569,]
> prc_train_labels <- prc[1:350,1]
> prc_test_labels <- prc[351:569,1]
```

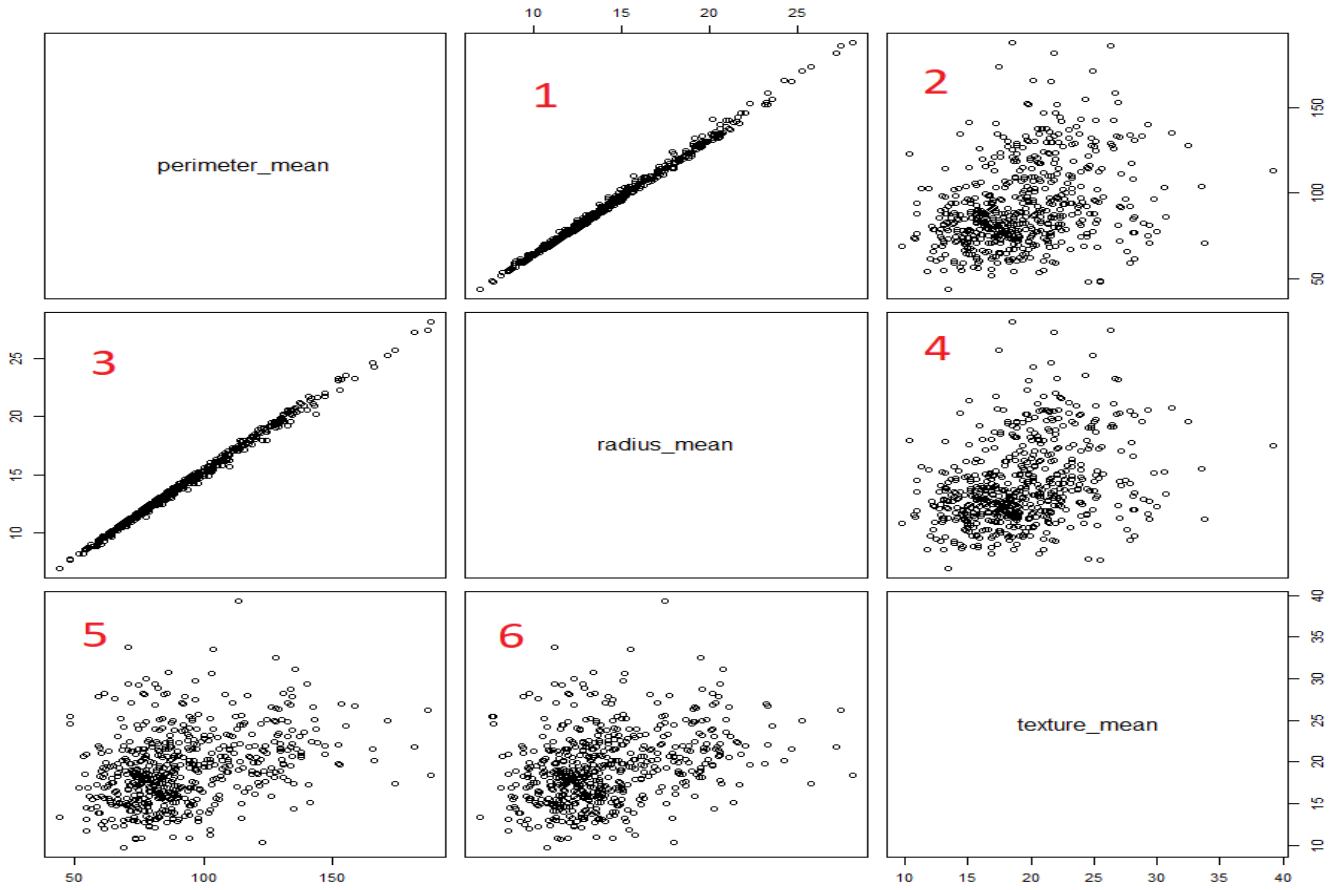
En Çok Tekrar Eden Veri

```
> prc2 <- table(prc$diagnosis)
> encok_deger <- names(prc2[which.max(prc2)])
> encok_deger
[1] "Iyi Huylu"
```

GÖRSELLEŞTİRME

Üç Niteliğin Karşılaştırılması

```
> pairs(~perimeter_mean + radius_mean + texture_mean, data=prc )
```



1 Numaralı Görsel : perimeter_mean ve radius_mean in birleşimidir. Perimeter(Y) arttıkça Radius'ta(X) artıyor.

2 Numaralı Görsel: perimeter_mean ve texture_mean in birleşimidir. Perimeter(Y) 50-100 Texture(X) 10-20 arasında yoğunluk gösteriyor.

3 Numaralı Görsel: radius_mean ve perimeter_mean in birleşimidir Radius(Y) arttıkça Perimeter(X) artıyor.

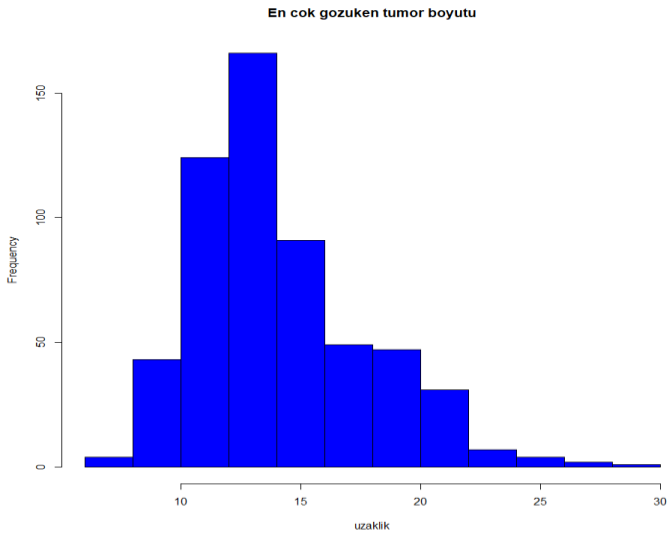
4 Numaralı Görsel: radius_mean ve texture_mean in birleşimidir. Radius(Y) 10-15 Texture(X) 10-20 arasında yoğunluk gösteriyor.

5 Numaralı Görsel: perimeter_mean ve texture_mean in birleşimidir. Texture(Y) 10-25 Perimeter(X) 50-100 arasında yoğunluk gösteriyor.

6 Numaralı Görsel: radius_mean ve texture_mean in birleşimidir. Texture(Y) 10-25 Radius(X) 10-15 arasında yoğunluk gösteriyor.

Histogram

```
> {uzaklik <- prc$radius_mean
+ hist(uzaklik, col="blue",main = "En cok gozuken tumor boyutu")} #veri setimizde en çok gözüken tümör boyutu 10 ila 15 arasıdır
```



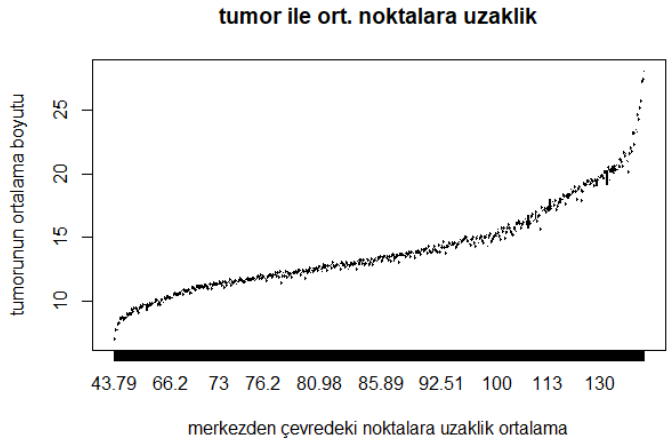
Veri setimizde en çok gözüken tümör boyutu 10 ila 15 arasıdır.

Veri setimizde en az gözüken tümör boyutu 20 ila 30 arasıdır.

Histagram numerik değerler için kullanılmıştır.

Boxplot

```
> boxplot(radius_mean~perimeter_mean,data=prc,xlab="merkezden çevredeki noktalara uzaklik ortalama",  #Tümörlerin hangi uzaklıkta ortalama ne boyutta olduğunu gösterir
+         ylab = "tumorunun ortalama boyutu" , main="tumor ile ort. noktalara uzaklik")    #veri setimize göre merkezden uzakta olan tümörün boyutu daha büyüktür
```



Tümörlerin hangi uzaklıkta ortalama ne boyutta olduğunu gösterir.

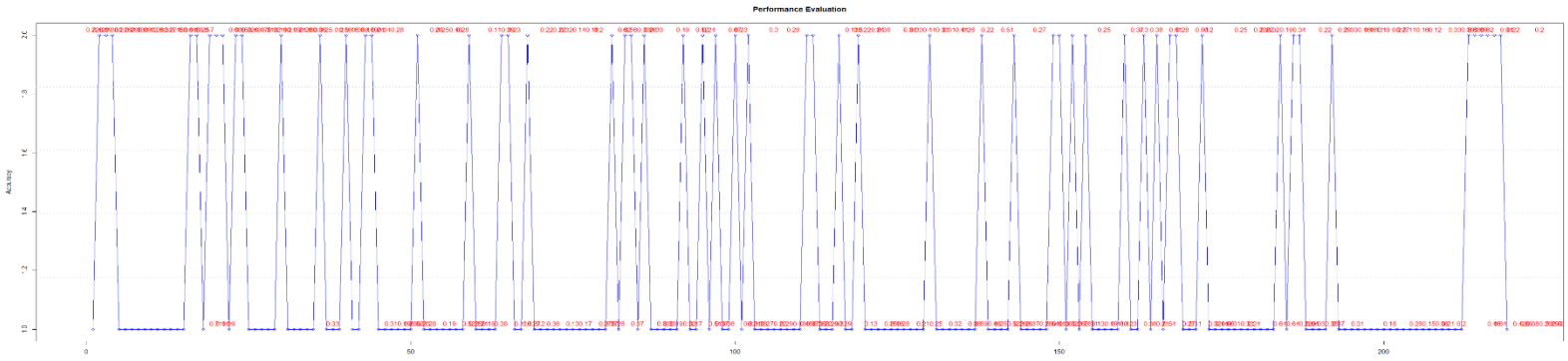
Veri setimize göre merkezden uzakta olan tümörün boyutu daha büyüktür.

Algoritma Seçimi ve Modelleme

1. Modelleme : KNN algoritması

```
> #install.packages("class")
> library(class)
> set.seed(1)
> prc_test_pred <- knn(train = prc_train, test = prc_test,cl = prc_train_labels, k=24)  # √569 = 23,85 ≈ 24
> head(prc_test_pred)                                                                   # k değeri veri setindeki gözlem sayısının kare kökü olarak belirledik
[1] Iyi Huylu  Kotu Huylu Kotu Huylu Kotu Huylu Iyi Huylu  Iyi Huylu
Levels: Iyi Huylu Kotu Huylu
> head(prc_test)
    radius_mean texture_mean perimeter_mean area_mean smoothness_mean compactness_mean concavity_mean concave.points_mean symmetry_mean fractal_dimension_mean
351   0.2214492    0.2489009      0.2066892 0.1177094       0.2074569        0.05189866     0.01946111          0.05775348     0.3085859              0.1548020
352   0.4150220    0.3216097      0.4374957 0.2609120       0.6470163        0.66566468     0.68275539          0.61729622     0.6641414              0.5490733
353   0.8873586    0.2620900      0.9011817 0.7917285       0.5621558        0.66535795     0.78912840          0.95079523     0.4525253              0.2369419
354   0.3833120    0.5421035      0.3746113 0.2430965       0.4493094        0.24019999     0.28936270          0.32569583     0.2964646              0.3091828
355   0.1968385    0.1474467      0.1896897 0.1022694       0.1815474        0.12655665     0.10555295          0.07311133     0.3181818              0.2289385
356   0.2640447    0.3165370      0.2634925 0.1451962       0.3156992        0.25894117     0.24133083          0.21824056     0.2388889              0.2502106
    radius_se texture_se perimeter_se    area_se smoothness_se compactness_se concavity_se concave.points_se symmetry_se fractal_dimension_se radius_worst
351 0.08759732 0.06899752   0.06917024 0.03591347      0.1655505      0.03559197   0.01502273         0.1192650  0.2009062             0.06126059   0.1903237
352 0.14807170 0.21300389   0.12816284 0.08296258      0.2588979      0.47569622   0.25133838         0.4324683  0.6690494             0.22232356   0.3354678
353 0.31986239 0.10844678   0.30462234 0.27325093      0.1582758      0.30175444   0.10772727         0.2856602  0.2176507             0.08603154   0.8964781
354 0.19623393 0.25322666   0.16100457 0.10565598      0.2993847      0.16566527   0.12404040         0.3307445  0.1873980             0.13723173   0.3763785
355 0.11251132 0.09923091   0.12123639 0.04116190      0.1301288      0.23746508   0.11376263         0.2786513  0.3255755             0.13598800   0.1490573
356 0.09005975 0.24703854   0.11567639 0.03864041      0.2767107      0.30115360   0.15835859         0.3724190  0.2604266             0.11436093   0.1935254
    texture_worst perimeter_worst area_worst smoothness_worst compactness_worst concavity_worst concave.points_worst symmetry_worst fractal_dimension_worst
351     0.2057569       0.1653469 0.08781459        0.1876114        0.03635358     0.02432907          0.1464605     0.2298443              0.08664568
352     0.3238273       0.3435928 0.17943866        0.5535891        0.46308855     0.54888179          0.7336770     0.5282870              0.32769251
353     0.3081023       0.8909308 0.74931184        0.5403817        0.54953382     0.51525559          0.9470790     0.4188843              0.21717172
354     0.5650320       0.3525574 0.21254424        0.6262299        0.20210340     0.32180511          0.5243986     0.2146659              0.25803489
355     0.1012793       0.1454754 0.06594082        0.1153668        0.09538085     0.09592652          0.1347766     0.1992904              0.09930474
356     0.2774520       0.1922905 0.08901888        0.2537806        0.16775815     0.19073482          0.3183849     0.1095998              0.11045520
```

```
> plot(x=1:length(prc_test_pred), y = prc_test_pred, type = "o", col = "blue", xlab = "k", ylab = "Accuracy",main = "Performance Evaluation")
> grid(NA, 5, lwd = 2)
> text(x=1:length(prc_train_labels), y = prc_train_labels,  round(prc_test$radius_mean,2), cex=1, pos=3,col="red")
```




```
> summary(prc_test$radius_mean)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0336  0.2179  0.2915  0.3238  0.3802  0.9673
```

#Uç değerleri buluyoruz

2. Modelleme : Navibayes

```
> #install.packages ("e1071")
> library (e1071)
> naiveB_model <- naiveBayes (prc [, 1: 2], prc [[1]])
> naiveB_model

Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = prc[, 1:2], y = prc[[1]])

A-priori probabilities:
prc[[1]]
 Iyi Huylu Kotu Huylu
 0.6274165  0.3725835

Conditional probabilities:
           diagnosis
prc[[1]]     Iyi Huylu Kotu Huylu
  Iyi Huylu         1          0
  Kotu Huylu        0          1

           radius_mean
prc[[1]]         [,1]     [,2]
  Iyi Huylu  12.14652 1.780512
  Kotu Huylu 17.46283 3.203971

> nb_predictions <- predict(naiveB_model, prc_test[,1:2])
> nb_probs <- predict(naiveB_model, prc_test[,1:2], "raw")
> results <- data.frame(prc_test[[1]], nb_predictions, nb_probs)
```

A-priorş probabilities kısmında verilerin yüzdelik değerlerini gösteriyoruz

Conditional probabilities kısmında eşleşen verileri 1 eşleşmeyen değerleri 0 gösteriyoruz

Radius_mean stünundaki iyi huylu ve kötü huylu değeri gösteriyor

```
> results
    prc_test..1.. nb_predictions    Iyi.Huylu Kotu.Huylu
1      0.22144919      Kotu Huylu 0.0010678534  0.9989321
2      0.41502201      Kotu Huylu 0.0015905301  0.9984095
3      0.88735861      Kotu Huylu 0.0040582764  0.9959417
4      0.38331204      Kotu Huylu 0.0014908920  0.9985091
5      0.19683847      Kotu Huylu 0.0010145017  0.9989855
6      0.26404468      Kotu Huylu 0.0011665365  0.9988335
7      0.28723555      Kotu Huylu 0.0012238381  0.9987762
8      0.32604477      Kotu Huylu 0.0013257404  0.9986743
9      0.08978182      Kotu Huylu 0.0008104890  0.9991895
10     0.11619102      Kotu Huylu 0.0008568460  0.9991432
11     0.26309811      Kotu Huylu 0.0011642525  0.9988357
12     0.29906763      Kotu Huylu 0.0012540915  0.9987459
13     0.27351034      Kotu Huylu 0.0011896107  0.9988104
14     0.45051825      Kotu Huylu 0.0017095276  0.9982905
15     0.30380046      Kotu Huylu 0.0012663901  0.9987336
16     0.63699181      Kotu Huylu 0.0024857635  0.9975142
17     0.62563302      Kotu Huylu 0.0024302670  0.9975697
18     0.24747977      Kotu Huylu 0.0011271741  0.9988728
19     0.69709877      Kotu Huylu 0.0027998623  0.9972001
20     0.71129727      Kotu Huylu 0.0028793307  0.9971207
21     0.44341900      Kotu Huylu 0.0016850745  0.9983149
22     0.38851815      Kotu Huylu 0.0015068347  0.9984932
23     0.68100715      Kotu Huylu 0.0027122953  0.9972877
24     0.64645748      Kotu Huylu 0.0025329207  0.9974671
25     0.31752568      Kotu Huylu 0.0013027058  0.9986973
26     0.43489990      Kotu Huylu 0.0016561672  0.9983438
27     0.16986133      Kotu Huylu 0.0009589290  0.9990411
28     0.30664016      Kotu Huylu 0.0012738241  0.9987262
29     0.31610583      Kotu Huylu 0.0012989037  0.9987011
30     0.19399877      Kotu Huylu 0.0010085111  0.9989915
31     0.20299115      Kotu Huylu 0.0010275973  0.9989724
32     0.19210564      Kotu Huylu 0.0010045360  0.9989955
33     0.23990724      Kotu Huylu 0.0011096024  0.9988904
34     0.25599886      Kotu Huylu 0.0011472576  0.9988527
35     0.29812107      Kotu Huylu 0.0012516454  0.9987484
36     0.36059444      Kotu Huylu 0.0014231732  0.9985768
37     0.24747977      Kotu Huylu 0.0011271741  0.9988728
38     0.32651806      Kotu Huylu 0.0013270313  0.9986730
39     0.20299115      Kotu Huylu 0.0010275973  0.9989724
40     0.59486961      Kotu Huylu 0.0022857707  0.9977142
41     0.15518955      Kotu Huylu 0.0009299311  0.9990701
42     0.08296654      Kotu Huylu 0.0007989185  0.9992011
43     0.40271665      Kotu Huylu 0.0015511389  0.9984489
44     0.69236594      Kotu Huylu 0.0027738348  0.9972262
45     0.24227365      Kotu Huylu 0.0011150655  0.9988849
46     0.33503715      Kotu Huylu 0.0013504729  0.9986495
47     0.30900658      Kotu Huylu 0.0012800506  0.9987199
48     0.27540347      Kotu Huylu 0.0011942772  0.9988057
49     0.19305220      Kotu Huylu 0.0010065217  0.9989935
50     0.22807516      Kotu Huylu 0.0010826658  0.9989173
51     0.51725117      Kotu Huylu 0.0019563719  0.9980436
52     0.23422784      Kotu Huylu 0.0010965945  0.9989034
53     0.28297600      Kotu Huylu 0.0012131171  0.9987869
54     0.28202944      Kotu Huylu 0.0012107468  0.9987893
55     0.25363245      Kotu Huylu 0.0011416451  0.9988584
56     0.18737281      Kotu Huylu 0.0009946633  0.9990053
57     0.43348005      Kotu Huylu 0.0016513952  0.9983486
58     0.27776989      Kotu Huylu 0.0012001347  0.9987999
59     0.52103744      Kotu Huylu 0.0019713411  0.9980287
60     0.25031947      Kotu Huylu 0.0011338313  0.9988662
61     0.20725070      Kotu Huylu 0.0010367573  0.9989632
62     0.19210564      Kotu Huylu 0.0010045360  0.9989955
63     0.11434521      Kotu Huylu 0.0008535258  0.9991465
64     0.37905249      Kotu Huylu 0.0014779668  0.9985220
65     0.38567845      Kotu Huylu 0.0014981188  0.9985019
66     0.23233471      Kotu Huylu 0.0010922907  0.9989077
67     0.11472384      Kotu Huylu 0.0008542059  0.9991458
68     0.40318993      Kotu Huylu 0.0015526367  0.9984474
69     0.27067064      Kotu Huylu 0.0011826434  0.9988174
70     0.19778503      Kotu Huylu 0.0010165060  0.9989835
71     0.21718964      Kotu Huylu 0.0010584329  0.9989416
72     0.36485399      Kotu Huylu 0.0014356445  0.9985644
73     0.21908278      Kotu Huylu 0.0010626100  0.9989374
74     0.31610583      Kotu Huylu 0.0012989037  0.9987011
75     0.13067348      Kotu Huylu 0.0008833259  0.9991167
76     0.14430404      Kotu Huylu 0.0009089566  0.9990910
77     0.16560178      Kotu Huylu 0.0009504230  0.9990496
78     0.18074684      Kotu Huylu 0.0009809961  0.9990190
79     0.19636519      Kotu Huylu 0.0010135009  0.9989865
80     0.27161721      Kotu Huylu 0.0011849615  0.9988150
81     0.37479294      Kotu Huylu 0.0014651477  0.9985349
82     0.25647215      Kotu Huylu 0.0011483832  0.9988516
83     0.62468645      Kotu Huylu 0.0024256954  0.9975743
84     0.56031994      Kotu Huylu 0.0021331619  0.9978668
85     0.37289981      Kotu Huylu 0.0014594841  0.9985405
86     0.33125089      Kotu Huylu 0.0013400064  0.9986600
87     0.27871646      Kotu Huylu 0.0012024853  0.9987975
88     0.33409059      Kotu Huylu 0.0013478491  0.9986522
89     0.32509821      Kotu Huylu 0.0013231620  0.9986768
90     0.33314402      Kotu Huylu 0.0013452300  0.9986548
91     0.18879265      Kotu Huylu 0.0009976154  0.9990024
92     0.48696105      Kotu Huylu 0.0018404245  0.9981596
93     0.32178522      Kotu Huylu 0.0013141753  0.9986858
94     0.16986133      Kotu Huylu 0.0009589290  0.9990411
95     0.52293057      Kotu Huylu 0.0019788662  0.9980211
96     0.23706754      Kotu Huylu 0.0011030803  0.9988969
97     0.50967864      Kotu Huylu 0.0019267549  0.9980732
98     0.37006011      Kotu Huylu 0.0014510277  0.9985490
99     0.35728146      Kotu Huylu 0.0014135443  0.9985865
100    0.66822850      Kotu Huylu 0.0026446007  0.9973554
101    0.23138814      Kotu Huylu 0.0010901449  0.9989099
102    0.59676274      Kotu Huylu 0.0022944235  0.9977056
103    0.23754082      Kotu Huylu 0.0011041648  0.9988958
104    0.35728146      Kotu Huylu 0.0014135443  0.9985865
105    0.26688438      Kotu Huylu 0.0011734138  0.9988266
106    0.30285390      Kotu Huylu 0.0012639213  0.9987361
107    0.22002934      Kotu Huylu 0.0010647044  0.9989353
108    0.29480808      Kotu Huylu 0.0012431195  0.9987569
109    0.28486914      Kotu Huylu 0.0012178710  0.9987821
110    0.13128875      Kotu Huylu 0.0008844679  0.9991155
111    0.47796867      Kotu Huylu 0.0018072724  0.9981927
112    0.96734346      Kotu Huylu 0.0047317881  0.9952682
113    0.35112878      Kotu Huylu 0.0013958255  0.9986042
114    0.21860949      Kotu Huylu 0.0010615643  0.9989384
115    0.29291495      Kotu Huylu 0.0012382723  0.9987617
116    0.29622793      Kotu Huylu 0.0012467667  0.9987532
117    0.29149510      Kotu Huylu 0.0012346487  0.9987654
118    0.12717119      Kotu Huylu 0.0008768519  0.9991231
119    0.50257939      Kotu Huylu 0.0018993741  0.9981006
120    0.21955606      Kotu Huylu 0.0010636567  0.9989363
121    0.12712386      Kotu Huylu 0.0008767647  0.9991232
122    0.23943395      Kotu Huylu 0.0011085129  0.9988915
123    0.37573950      Kotu Huylu 0.0014679872  0.9985320
124    0.25031947      Kotu Huylu 0.0011338313  0.9988662
125    0.18453311      Kotu Huylu 0.0009887839  0.9990112
126    0.27682332      Kotu Huylu 0.0011977884  0.9988022
127    0.34166312      Kotu Huylu 0.0013689759  0.9986310
128    0.32746462      Kotu Huylu 0.0013296168  0.9986704
129    0.21340338      Kotu Huylu 0.0010501253  0.9989499
130    0.43868617      Kotu Huylu 0.0016689565  0.9983310
131    0.24511335      Kotu Huylu 0.0011216548  0.9988783
132    0.32746462      Kotu Huylu 0.0013296168  0.9986704
133    0.30711345      Kotu Huylu 0.0012750671  0.9987249
134    0.31799896      Kotu Huylu 0.0013039755  0.9986960
135    0.41407544      Kotu Huylu 0.0015874667  0.9984125
136    0.25883856      Kotu Huylu 0.0011540271  0.9988460
137    0.36248758      Kotu Huylu 0.0014287033  0.9985713
138    0.58966350      Kotu Huylu 0.0022621338  0.9977379
139    0.22239576      Kotu Huylu 0.0010699576  0.9989300
140    0.45951063      Kotu Huylu 0.0017409833  0.9982590
141    0.24937290      Kotu Huylu 0.0011316081  0.9988684
142    0.51441147      Kotu Huylu 0.0019452155  0.9980548
143    0.52198400      Kotu Huylu 0.0019751002  0.9980249
144    0.25931185      Kotu Huylu 0.0011551591  0.9988448
145    0.29244167      Kotu Huylu 0.0012370633  0.9987629
146    0.37337309      Kotu Huylu 0.0014608981  0.9985391
147    0.26830423      Kotu Huylu 0.0011768669  0.9988231
148    0.25978513      Kotu Huylu 0.0011562921  0.9988437
149    0.54470159      Kotu Huylu 0.0020673770  0.9979326
150    0.64409106      Kotu Huylu 0.0025210530  0.9974789
151    0.38141890      Kotu Huylu 0.0014851343  0.9985149
152    0.32367836      Kotu Huylu 0.0013193035  0.9986807
153    0.26309811      Kotu Huylu 0.0011642525  0.9988357
154    0.76241185      Kotu Huylu 0.0031833416  0.9968167
155    0.10823986      Kotu Huylu 0.0008426305  0.9991574
156    0.12754981      Kotu Huylu 0.0008775496  0.9991225
157    0.24793305      Kotu Huylu 0.0011282811  0.9988717
158    0.19305220      Kotu Huylu 0.0010065217  0.9989935
159    0.44105258      Kotu Huylu 0.0016769972  0.9983230
160    0.40129680      Kotu Huylu 0.0015466536  0.9984533
161    0.22523546      Kotu Huylu 0.0010762939  0.9989237
162    0.37053339      Kotu Huylu 0.0014524339  0.9985476
163    0.30380046      Kotu Huylu 0.0012663901  0.9987336
164    0.35964788      Kotu Huylu 0.0014204158  0.9985796
165    0.38189219      Kotu Huylu 0.0014865717  0.9985134
166    0.20630413      Kotu Huylu 0.0010347151  0.9989653
167    0.53618250      Kotu Huylu 0.0020323063  0.9979677
168    0.61096124      Kotu Huylu 0.0023603113  0.9976397
169    0.27918974      Kotu Huylu 0.0012036622  0.9987963
170    0.27303706      Kotu Huylu 0.0011884468  0.9988116
171    0.10951772      Kotu Huylu 0.0008448999  0.9991551
172    0.83529746      Kotu Huylu 0.0036693652  0.9963306
173    0.20251787      Kotu Huylu 0.0010265843  0.9989734
174    0.31847224      Kotu Huylu 0.0013052463  0.9986948
175    0.13564296      Kotu Huylu 0.0008925898  0.9991074
176    0.07525202      Kotu Huylu 0.0007860109  0.9992140
177    0.30664016      Kotu Huylu 0.0012738241  0.9987262
178    0.25363245      Kotu Huylu 0.0011416451  0.9988584
179    0.32935775      Kotu Huylu 0.0013348020  0.9986652
180    0.24085380      Kotu Huylu 0.0011117846  0.9988882
181    0.22570874      Kotu Huylu 0.0010773534  0.9989226
182    0.22192248      Kotu Huylu 0.0010689050  0.9989311
183    0.31705239      Kotu Huylu 0.0013014372  0.9986986
184    0.63841166      Kotu Huylu 0.0024927839  0.9975072
185    0.18831937      Kotu Huylu 0.0009966304  0.9990034
186    0.64219793      Kotu Huylu 0.0025115966  0.9974884
187    0.34497610      Kotu Huylu 0.0013783171  0.9986217
188    0.22286904      Kotu Huylu 0.0010710112  0.9989290
189    0.03540158      Kotu Huylu 0.0007224304  0.9992776
190    0.03360310      Kotu Huylu 0.0007196795  0.9992803
191    0.21576980      Kotu Huylu 0.0010553103  0.9989447
192    0.35444176      Kotu Huylu 0.0014053401  0.9985947
193    0.36722041      Kotu Huylu 0.0014426176  0.9985574
194    0.29480808      Kotu Huylu 0.0012431195  0.9987569
195    0.32604477      Kotu Huylu 0.0013257404  0.9986743
196    0.31421269      Kotu Huylu 0.0012938507  0.9987061
197    0.15802925      Kotu Huylu 0.0009354778  0.9990645
198    0.15518955      Kotu Huylu 0.0009299311  0.9990701
199    0.12788111      Kotu Huylu 0.0008781605  0.9991218
200    0.18169341      Kotu Huylu 0.0009829376  0.9990171
201    0.18358654      Kotu Huylu 0.0009868315  0.9990132
202    0.19636519      Kotu Huylu 0.0010135009  0.9989865
203    0.27398362      Kotu Huylu 0.0011907757  0.9988092
204    0.11131620      Kotu Huylu 0.0008481038  0.9991519
205    0.27918974      Kotu Huylu 0.0012036622  0.9987963
206    0.15660940      Kotu Huylu 0.0009327005  0.9990673
207    0.15045672      Kotu Huylu 0.0009207559  0.9990792
208    0.11557575      Kotu Huylu 0.0008557379  0.9991443
209    0.36012116      Kotu Huylu 0.0014217939  0.9985782
210    0.21434995      Kotu Huylu 0.0010521964  0.9989478
211    0.33456387      Kotu Huylu 0.0013491604  0.9986508
212    0.19967817      Kotu Huylu 0.0010205260  0.9989795
213    0.38993800      Kotu Huylu 0.0015112106  0.9984888
214    0.65970940      Kotu Huylu 0.0026003585  0.9973996
215    0.68999953      Kotu Huylu 0.0027609067  0.9972391
216    0.62232003      Kotu Huylu 0.0024143018  0.9975857
217    0.45525108      Kotu Huylu 0.0017260156  0.9982740
218    0.64456434      Kotu Huylu 0.0025234224  0.9974766
219    0.03686876      Kotu Huylu 0.0007246819  0.9992753
```

3.Modelleme : Karar Ağacı

```
> #3.Modelleme
> #install.packages("caret")
> library(caret)
> set.seed(1)
> egitimIndisleri <- createDataPartition(y = prc$diagnosis, p = .70, list = FALSE)
> egitim_C45 <- prc[egitimIndisleri,]
> test_C45 <- prc[-egitimIndisleri,]
> numeric_coloumns<-c(2:32)
> for(i in 1:ncol(prc)){
+   if(i %in%numeric_coloumns)
+     prc[,i]<-as.numeric(prc[,i])
+   else
+     prc[,i]<- as.factor(prc[,i])
+ }
> summary(prc)
     diagnosis    radius_mean      texture_mean   perimeter_mean     area_mean      smoothness_mean   compactness_mean  concavity_mean    concave.points_mean
 Iyi Huylu :357   Min.   : 6.981   Min.   : 9.71   Min.   : 43.79   Min.   : 143.5   Min.   :0.05263   Min.   :0.01938   Min.   :0.00000   Min.   :0.00000
 Kotu Huylu:212   1st Qu.:11.700   1st Qu.:16.17   1st Qu.: 75.17   1st Qu.: 420.3   1st Qu.:0.08637   1st Qu.:0.06492   1st Qu.:0.02956   1st Qu.:0.02031
                  Median :13.370   Median :18.84   Median : 86.24   Median : 551.1   Median :0.09587   Median :0.09263   Median :0.06154   Median :0.03350
                  Mean   :14.127   Mean   :19.29   Mean   : 91.97   Mean   : 654.9   Mean   :0.09636   Mean   :0.10434   Mean   :0.08880   Mean   :0.04892
                  3rd Qu.:15.780   3rd Qu.:21.80   3rd Qu.:104.10   3rd Qu.: 782.7   3rd Qu.:0.10530   3rd Qu.:0.13040   3rd Qu.:0.13070   3rd Qu.:0.07400
                  Max.   :28.110   Max.   :39.28   Max.   :188.50   Max.   :2501.0   Max.   :0.16340   Max.   :0.34540   Max.   :0.42680   Max.   :0.20120
 symmetry_mean    fractal_dimension_mean   radius_se       texture_se      perimeter_se       area_se        smoothness_se      compactness_se      concavity_se
 Min.   :0.1060   Min.   :0.04996        Min.   :0.1115   Min.   :0.3602   Min.   : 0.757   Min.   :  6.802   Min.   :0.001713   Min.   :0.002252   Min.   :0.00000
 1st Qu.:0.1619   1st Qu.:0.05770        1st Qu.:0.2324   1st Qu.:0.8339   1st Qu.: 1.606   1st Qu.: 17.850   1st Qu.:0.005169   1st Qu.:0.013080   1st Qu.:0.01509
 Median :0.1792   Median :0.06154        Median :0.3242   Median :1.1080   Median : 2.287   Median : 24.530   Median :0.006380   Median :0.020450   Median :0.02589
 Mean   :0.1812   Mean   :0.06280        Mean   :0.4052   Mean   :1.2169   Mean   : 2.866   Mean   : 40.337   Mean   :0.007041   Mean   :0.025478   Mean   :0.03189
 3rd Qu.:0.1957   3rd Qu.:0.06612        3rd Qu.:0.4789   3rd Qu.:1.4740   3rd Qu.: 3.357   3rd Qu.: 45.190   3rd Qu.:0.008146   3rd Qu.:0.032450   3rd Qu.:0.04205
 Max.   :0.3040   Max.   :0.09744        Max.   :2.8730   Max.   :4.8850   Max.   :21.980   Max.   :542.200   Max.   :0.031130   Max.   :0.135400   Max.   :0.39600
 concave.points_se   symmetry_se       fractal_dimension_se  radius_worst   texture_worst   perimeter_worst    area_worst     smoothness_worst  compactness_worst
 Min.   :0.000000   Min.   :0.007882   Min.   :0.0008948    Min.   : 7.93   Min.   :12.02   Min.   : 50.41   Min.   : 185.2   Min.   :0.07117   Min.   :0.02729
 1st Qu.:0.007638   1st Qu.:0.015160   1st Qu.:0.0022480    1st Qu.:13.01   1st Qu.:21.08   1st Qu.: 84.11   1st Qu.: 515.3   1st Qu.:0.11660   1st Qu.:0.14720
 Median :0.010930   Median :0.018730   Median :0.0031870    Median :14.97   Median :25.41   Median : 97.66   Median : 686.5   Median :0.13130   Median :0.21190
 Mean   :0.011796   Mean   :0.020542   Mean   :0.0037949    Mean   :16.27   Mean   :25.68   Mean   :107.26   Mean   : 880.6   Mean   :0.13237   Mean   :0.25427
 3rd Qu.:0.014710   3rd Qu.:0.023480   3rd Qu.:0.0045580    3rd Qu.:18.79   3rd Qu.:29.72   3rd Qu.:125.40   3rd Qu.:1084.0   3rd Qu.:0.14600   3rd Qu.:0.33910
 Max.   :0.052790   Max.   :0.078950   Max.   :0.0298400    Max.   :36.04   Max.   :49.54   Max.   :251.20   Max.   :4254.0   Max.   :0.22260   Max.   :1.05800
 concavity_worst  concave.points_worst symmetry_worst   fractal_dimension_worst
 Min.   :0.0000   Min.   :0.00000      Min.   :0.1565   Min.   :0.05504
 1st Qu.:0.1145   1st Qu.:0.06493      1st Qu.:0.2504   1st Qu.:0.07146
 Median :0.2267   Median :0.09993      Median :0.2822   Median :0.08004
 Mean   :0.2722   Mean   :0.11461      Mean   :0.2901   Mean   :0.08395
 3rd Qu.:0.3829   3rd Qu.:0.16140      3rd Qu.:0.3179   3rd Qu.:0.09208
 Max.   :1.2520   Max.   :0.29100      Max.   :0.6638   Max.   :0.20750
```

Eğitim ve test veri setini oluşturmak için “caret” paketini yükledik

Prc tablosunun özetini çıkarttık

```
> #install.packages("RWeka")
> library(RWeka)
> C45_modeli <- J48(diagnosis ~., data=egitim_C45)
> #C45_modeli <- J48(prc ~., data=egitim_C45, control = Weka_control(R = true, M = 6))  #Control etmek için
> print(summary(C45_modeli))

=== Summary ===

Correctly Classified Instances         396               99.2481 %
Kappa statistic                          0.9839
Mean absolute error                      0.0141
Root mean squared error                  0.084
Relative absolute error                  3.0154 %
Root relative squared error             17.3679 %
Total Number of Instances              399

=== Confusion Matrix ===

   a   b   <-- classified as
 250   0 |   a = Iyi Huylu
   3 146 |   b = Kotu Huylu
> print(C45_modeli)
J48 pruned tree
------------------

radius_worst <= 16.76
|   concave.points_worst <= 0.1357: Iyi Huylu (232.0/2.0)
|   concave.points_worst > 0.1357
|   |   texture_worst <= 27.37
|   |   |   area_se <= 21.19: Iyi Huylu (10.0)
|   |   |   area_se > 21.19
|   |   |   |   smoothness_worst <= 0.1349: Iyi Huylu (6.0/1.0)
|   |   |   |   smoothness_worst > 0.1349: Kotu Huylu (5.0)
|   |   texture_worst > 27.37: Kotu Huylu (12.0)
radius_worst > 16.76
|   concavity_worst <= 0.231
|   |   texture_mean <= 18.83: Iyi Huylu (5.0)
|   |   texture_mean > 18.83: Kotu Huylu (8.0)
|   concavity_worst > 0.231: Kotu Huylu (121.0)

Number of Leaves  :     8

Size of the tree :      15
```

C4.5 algoritmasıyla çalışabilmek için “RWeka” paketini kullandık

Summary kısmında modelin eğitim verisi üstündeki performansı %99.2481 değerinde çıkmıştır

Karar ağacını çalıştırmak için J48 fonksyonunu kullandık

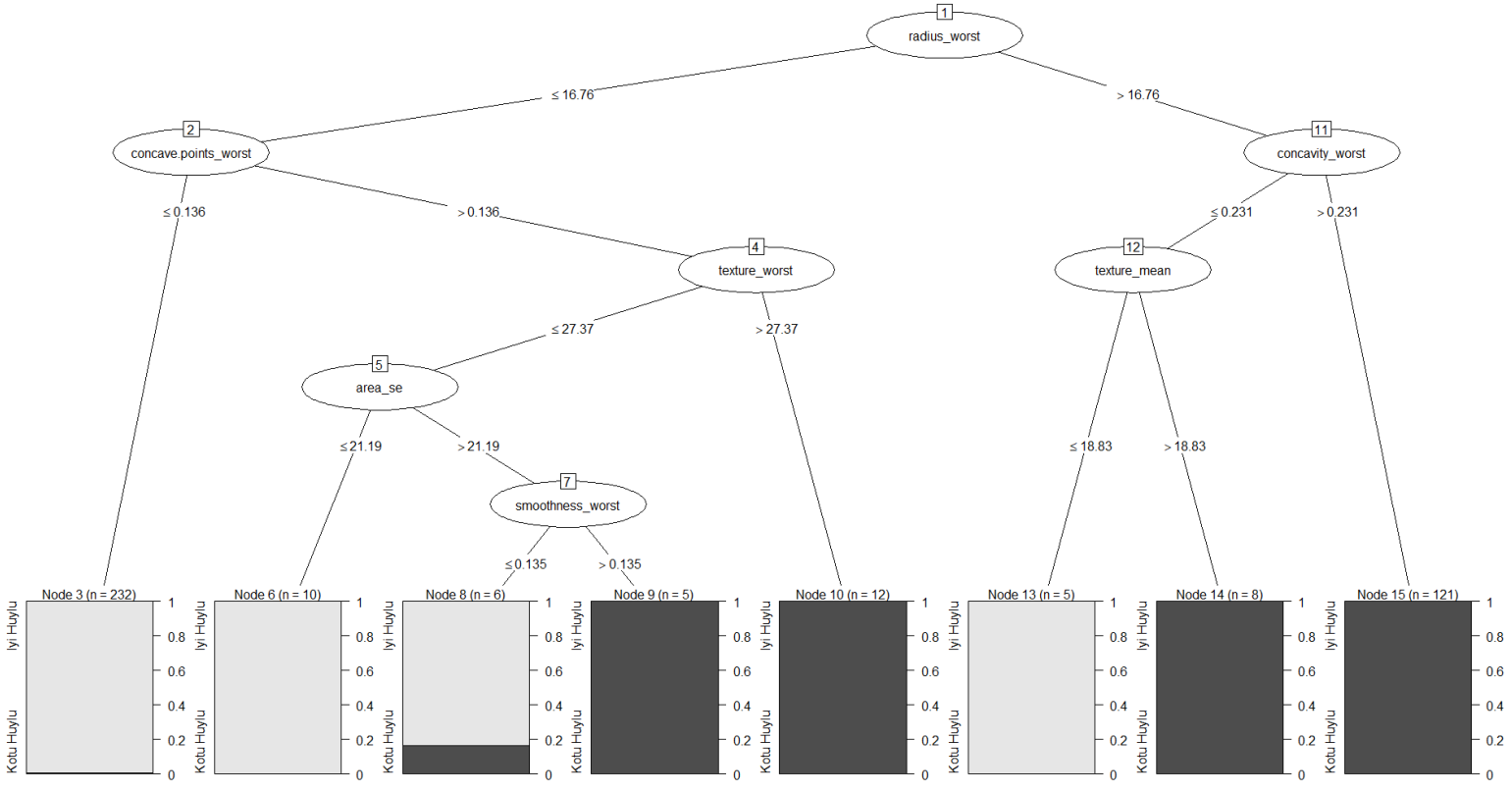
Niteliklerin önem derecesi

```
> #install.packages("FSelector")
> library(FSelector)
> information.gain(diagnosis ~., data=egitim_C45) > gain.ratio(diagnosis ~., data=egitim_C45)
                        attr_importance                          attr_importance
radius_mean                  0.40532046 radius_mean                0.33106893
texture_mean                 0.11841186 texture_mean               0.17241757
perimeter_mean               0.41374475 perimeter_mean             0.38142732
area_mean                    0.40763453 area_mean                  0.33243733
smoothness_mean              0.06473818 smoothness_mean            0.10116412
compactness_mean             0.20317190 compactness_mean           0.18938440
concavity_mean               0.34222256 concavity_mean             0.36441667
concave.points_mean          0.45033421 concave.points_mean        0.33163454
symmetry_mean                0.04516805 symmetry_mean              0.06628570
fractal_dimension_mean       0.00000000 fractal_dimension_mean     0.00000000
radius_se                    0.28267345 radius_se                  0.23658895
texture_se                   0.00000000 texture_se                 0.00000000
perimeter_se                 0.26697562 perimeter_se               0.20327509
area_se                      0.36561153 area_se                    0.40461326
smoothness_se                0.00000000 smoothness_se              0.00000000
compactness_se               0.06859988 compactness_se             0.09940197
concavity_se                 0.14956663 concavity_se               0.15876610
concave.points_se            0.13650831 concave.points_se          0.13050719
symmetry_se                  0.00000000 symmetry_se                0.00000000
fractal_dimension_se         0.00000000 fractal_dimension_se       0.00000000
radius_worst                 0.49289580 radius_worst               0.41494842
texture_worst                0.11191714 texture_worst              0.16483088
perimeter_worst              0.50737778 perimeter_worst            0.40271012
area_worst                   0.47330985 area_worst                 0.50515670
smoothness_worst             0.09259731 smoothness_worst           0.13620888
compactness_worst            0.20188952 compactness_worst          0.19484888
concavity_worst              0.33741993 concavity_worst            0.26897422
concave.points_worst         0.43778666 concave.points_worst       0.37256363
symmetry_worst               0.12016461 symmetry_worst             0.12975635
fractal_dimension_worst      0.04359890 fractal_dimension_worst    0.07259820
```

"FSelector" paketini yükledik

Belirli bir veri kümesinden ve bir hedef özelliğinden öznitelikleri seçmek için işlevler içeren information.gain ve gain.ratio filtrelemelerini kullanarak sütunların önem derecesini yazdırdık

```
#install.packages("partykit")
library(partykit)
plot(C45_modeli)
```



Plot grafiğini kullanarak Karar Ağacını çıkardık

İyi huylu olanlar beyaz yapraklar Kötü huylu olanlar ise siyah yapraklar

Model Performans Değerlendirme

C4.5 Modelinin Performans Değerleri

```
> tahminler<-predict(C45_modeli, newdata = test_C45[,-1])
> confusionMatrix(tahminler, test_C45$diagnosis)
Confusion Matrix and Statistics

          Reference
Prediction   Iyi Huylu Kotu Huylu
  Iyi Huylu        101          5
  Kotu Huylu         6         58

               Accuracy : 0.9353
                 95% CI : (0.8872, 0.9673)
    No Information Rate : 0.6294
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.8617

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9439
            Specificity : 0.9206
         Pos Pred Value : 0.9528
         Neg Pred Value : 0.9062
             Prevalence : 0.6294
         Detection Rate : 0.5941
   Detection Prevalence : 0.6235
      Balanced Accuracy : 0.9323

       'Positive' Class : Iyi Huylu

> i <- 5
> (yeniveri_tahmini <- predict(C45_modeli, newdata = prc[i,-1]))
[1] Kotu Huylu
Levels: Iyi Huylu Kotu Huylu
> prc[i,1]
[1] Kotu Huylu
Levels: Iyi Huylu Kotu Huylu
```

Değerlerimizin doğruluk oranı %93 olduğu için güvenilirliği yüksektir.

İ=5 Verip tahmini değerimizi yazdırdık yazdırdığımız değere göre İyi Huylu veya Kötü Huylu olduğunu gösteriyor.

KNN Modelinin Performans Değerleri

```
> myConfMatrix <- table(veri2$diagnosis, prc$diagnosis, dnn = c("Model Tahminleri", "Gerçek/Referans"))
> myConfMatrix
                 Gerçek/Referans
Model Tahminleri Iyi Huylu Kotu Huylu
      Iyi Huylu        175         109
      Kotu Huylu       182         103
> dp <- myConfMatrix[1]
> dp
[1] 175
> yp <- myConfMatrix[3]
> yp
[1] 109
> yn<- myConfMatrix[2]
> yn
[1] 182
> dn <- myConfMatrix[4]
> dn
[1] 103
> dogruluk_orani <- (dp+dn)/(dp+yp+yn+dn)
> dogruluk_orani
[1] 0.4885764
> hata_orani <- 1 - dogruluk_orani
> hata_orani
[1] 0.5114236
> duyarlilik <- dp/(dp+yn)
> duyarlilik
[1] 0.4901961
> belirleyicilik <- dn/(dn+yp)
> belirleyicilik
[1] 0.4858491
> FNR <- 1 - duyarlilik
> FNR
[1] 0.5098039
> FPR <- 1 - belirleyicilik
> FPR
[1] 0.5141509
> PPV <- dp/(dp+yp)
> PPV
[1] 0.6161972
> NPV <- dn/(dn+yn)
> NPV
[1] 0.3614035
> FMeasure <- (2*duyarlilik*PPV)/(duyarlilik+PPV)
> FMeasure
[1] 0.5460218
> #caret::confusionMatrix(data = veri2$diagnosis, reference = prc$diagnosis)
```

```
> caret::confusionMatrix(data = veri2$diagnosis, reference = prc$diagnosis, mode = "everything")
Confusion Matrix and Statistics

          Reference
Prediction   Iyi Huylu Kotu Huylu
  Iyi Huylu        175        109
  Kotu Huylu       182        103

               Accuracy : 0.4886
                 95% CI : (0.4468, 0.5305)
    No Information Rate : 0.6274
    P-Value [Acc > NIR] : 1

                  Kappa : -0.0224

 Mcnemar's Test P-Value : 2.435e-05

            Sensitivity : 0.4902
            Specificity : 0.4858
         Pos Pred Value : 0.6162
         Neg Pred Value : 0.3614
              Precision : 0.6162
                 Recall : 0.4902
                     F1 : 0.5460
             Prevalence : 0.6274
         Detection Rate : 0.3076
   Detection Prevalence : 0.4991
      Balanced Accuracy : 0.4880

       'Positive' Class : Iyi Huylu
```

Gerçekte İyi Huylu olan hasta sayısı: 175

Gerçekte Kötü Huylu olan hasta sayısı : 103

Excel Tablo sonuçları:

		Gerçek Değerler		
		Pozitif	Negatif	Toplam
Tahmini Değerler	Pozitif	175	109	284
	Negatif	182	103	285
	Toplam	357	212	569

Doğruluk Oranı (ACC)	0,4885764
Hata Oranı (ERR)	0,5114236
Duyalılık (TPR)	0,4901961
Belirleyicilik (SPR)	0,4858491
Yanlış Negtaif Oranı (FNR)	0,5098039
Yanlış Pozitif Oranı (FPR)	0,5141509
Pozitif Öngörü Değeri (PPV)	0,6161972
Negatif Öngörü Değeri (NPV)	0,3614035
F Ölçütü	0,5460218

Üç Yollu Bölme Modelinin Performans Değerleri

```
> #install.packages("caret")
> library(caret)
> set.seed(1)
> my_indexes1 <- caret::createDataPartition(y = prc$diagnosis, times = 1, p = .80, list = F)
> training_ <- as.data.frame(prc[my_indexes1,])
> test <- as.data.frame(prc[-my_indexes1,])
> set.seed(1)
> my_indexes2 <- caret::createDataPartition(y = training_$radius_mean, times = 1, p = .80, list = F)
> training <- as.data.frame(training_[my_indexes2,])
> validation <- as.data.frame(training_[-my_indexes2,])
> table(prc$diagnosis)

 Iyi Huylu Kotu Huylu
       357        212
> table(training$diagnosis)

 Iyi Huylu Kotu Huylu
       231        137
> table(validation$diagnosis)

 Iyi Huylu Kotu Huylu
        55         33
> table(test$diagnosis)

 Iyi Huylu Kotu Huylu
        71         42
```

Bütün sorgulamalar sonucunda İyi Huylu olan hastalar çoğunlukta çıkmıştır.

Modellerin Genel Değerlendirme Tablosu

	Doğruluk	Hata	F-Ölçüsü	Tanısal Üstünlük Oranı
KNN	0,49	0,51	0,54	0,91
C4.5	0,93	0,07	0,95	131

Sırasıyla KNN algoritması ve C4.5 algoritması performansları karşılaştırılmıştır.

Tablodan hareketle en iyi performans C4.5 algoritması bulunmuştur.

Sonuç ve Bulgularınız (Tartışma)

Elde ettiğimiz bulgular sayesinde veri setimizde ki meme kanserinin İnce İğne Aspirasyonu (FNA) testi sayesinde İyi Huylu veya Kötü Huylu olma olasılığını test ettik. Sonuçlarımıza göre %62.7 İyi Huylu , %37,3 Kötü Huylu olduğunu sonucuna ulaştık.%93 oranında doğruluğu yakaladığı için test algoritmamız olarak C4.5'i seçtik.

İnce İğne Aspirasyonu (FNA) testi sonuçlarını C4.5 algoritmasını kullanarak yüksek doğruluk oranı sayesinde tümörlerin İyi Huylu veya Kötü Huylu olduğunu tahmin edilebilir. Bu test sayesinde tümörlerin İyi Huylu veya Kötü Huylu olma sonuçlarına daha hızlı ulaşabilir ve hastalığı kritik duruma gelmeden önlenebilir.

Kaynakça

https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29