

## Veri Madenciliğinde Naive Bayes Algoritması Uygulamaları

Naive Bayes Sınıflandırıcı, örüntü tanıma problemine ilk bakışta oldukça kısıtlayıcı görülen bir önerme ile kullanılabilen olasılıksal bir yaklaşımdır. Bu önerme, örüntü tanımada kullanılacak her bir tanımlayıcı öznelik ya da parametrenin istatistik açıdan bağımsız olması gerekliliğidir. Naive Bayes, R Programlamada Denetimli Doğrusal Olmayan bir sınıflandırma algoritmasıdır. Naive Bayes algoritması, belirli bir özelliğin oluşumunun diğer özelliklerin oluşumundan bağımsız olduğu varsayımını yaptığı için “Naive” olarak adlandırılır.

**Teori:** Naive Bayes algoritması, Bayes teoremine dayanmaktadır. Bayes teoremi, başka bir X olayı meydana geldiğinde Y olayının koşullu olasılığını verir.

### Kullanılacak Paketler:

```
install.packages (naivebayes)
install.packages (dplyr)
install.packages (ggplot2)
install.packages (psych)
install.packages("fastNaiveBayes")

install.packages("e1071")

install.packages("caTools")
install.packages("caret")
```

### Verinin Tanıtılması:

```
data <- read.csv("D:/RStudio/NaiveClassifiaction/binary.csv", header = T)
head(data)
  Launch Thickness Appearance Spreading Rank
0      6          9          8         2
0      5          8          7         2
0      8          7          7         2
0      8          8          9         1
0      9          8          7         2
0      7          7          7         2
```

Tepki değişkeninin frekansını hesaplama.

```
xtabs(~Launch+Rank, data = data)
Rank
  Rank
Launch 1  2  3
0 12 21 13
1 21 15 13
```

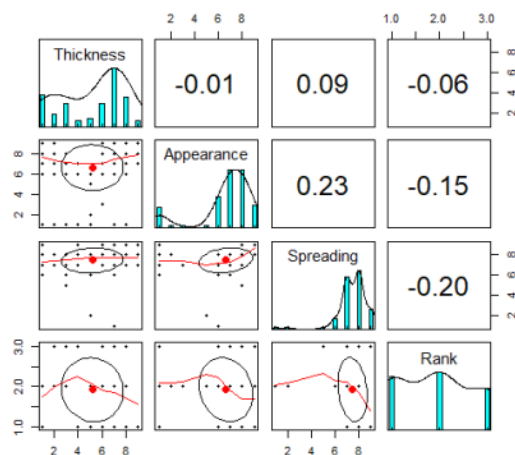
```
str(data)
data.frame': 95 obs. of 5 variables:
 $ Launch      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Thickness    : int  6 5 8 8 9 7 8 8 8 8 ...
 $ ColourAndAppearance: int  9 8 7 8 8 7 9 7 9 9 ...
 $ EaseOfSpreading : int  8 7 7 9 7 7 8 7 9 8 ...
 $ Rank        : int  2 2 2 1 2 2 2 2 1 2 ...
```

Launch ve Rank sütunları tamsayı değişkenleri olarak sınıflandırılmış. Tamsayı olarak görünen bu iki değişken Naive Bayes uygulayabilmemiz için faktör değişkenlerine dönüştürülmelidir.

```
data$Rank <- as.factor(data$Rank)
data$Launch <- as.factor(data$Launch)
```

### Verinin Görselleştirilmesi :

```
pairs.panels(data[-1])
```



Yorum: Bağımsız değişkenler arasında düşük korelasyon vardır.

### Naive Bayes Sınıflandırması :

```
model <- naive_bayes(Launch ~ ., data = train, usekernel = T)
model plot(model)
```

### Tahmin:

```
p <- predict(model, train, type = 'prob')
head(cbind(p, train))
```

	0	1	Launch	Thickness	Appearance	Spreading	Rank
1	0.9999637	3.629982e-05	0	1	9	8	2
2	0.9998770	1.229625e-04	0	1	8	7	1
3	0.9998804	1.196174e-04	0	1	7	7	1
4	0.9997236	2.764280e-04	0	1	8	9	1
6	0.9998804	1.196174e-04	0	1	7	7	1
7	0.9999637	3.629982e-05	0	1	9	8	2

Yorum: kalınlık düşük, yüksek görünüm, yayılma ve derece puanı 2 olduğundan, yani bu bize bu ürünün piyasaya çıkma şansının çok düşüktür olduğunu gösterir.

### Confusion Matrixleri:

#### Confusion Matrix – train data

```
p1 <- predict(model, train)
(tab1 <- table(p1, train$admit))
p1  0  1
0 30  5
1  5 34
1 - sum(diag(tab1)) / sum(tab1)
0.1351351
```

Yorum: Yanlış sınıflandırma %14 civarındadır. Eğitim modeli doğruluğu %86 civarında. Kötü bir model elde etmemiş oluruz.

#### Confusion Matrix – test data

```
p2 <- predict(model, test)
(tab2 <- table(p2, test$admit))
p2  0  1
0  8  0
1  3 10
1 - sum(diag(tab2)) / sum(tab2)
0.1428571
```

Yorum: R'deki Naive Bayes Sınıflandırmasına göre, test verilerinde yanlış sınıflandırma %14 civarındadır. Eğitim modelimizdeki doğruluğu arttırmak için veri ekleyebiliriz.

**KAYNAKÇA:**

<https://www.r-bloggers.com/2021/04/naive-bayes-classification-in-r/>

[https://scikit-learn.org/stable/modules/naive\\_bayes.html](https://scikit-learn.org/stable/modules/naive_bayes.html)

Yiğit Saçık 21936277