

## VeriMadenciliği – Data Mining

### HW – 4

1. Bir A ülkesinde XX hastalığının görülme yaygınlığı %25 imiş. Bu hastalığı teşhiste kullanılan bir YY yöntemi var ama mükemmel çalışmıyor. Hasta birisi için hasta diye yaptığı teşhis %99 başarı ile doğru oluyor. Hasta olmayan birisi için hasta değil diye yaptığı teşhis ise %93 ile doğru oluyor. Buna göre test yaptıran birisi hastasının sonucunu aldıysa gerçekten hasta olma olasılığı nedir?
2. Asağıda verilen tabloyu kullanarak,  $x=\{E, 3, 3\}$  (cinsiyet, boy,kilo) olarak verilen verinin hangi beden olduğunu Bayesian sınıflandırma yaparak bulun.  
Tabloda boy değişkeni 1den 4e kadar 4 gruba, kilo değişkeni ise 1den 5e kadar 5 gruba ayrılmıştır.

Gender/cinsiyet	Length/boy	Weight/kilo	Size/beden
K	1	2	Small
K	2	2	Medium
K	2	3	Medium
K	1	1	Small
E	3	3	Medium
E	3	4	Large
E	1	2	Small
E	3	5	Large
E	2	3	medium
E	3	3	Large

3. Asağıdaki vize ve final sınav notlarından oluşan öğrenme kumesini kullanarak, vize notu verildiğinde final notunu tahmit edecek bir denklemleri linear regression yöntemi ile bulun.

Midterm (x)	Final (y)
72	84
50	63
81	78
74	78
94	90
86	78
59	49
83	88
65	66