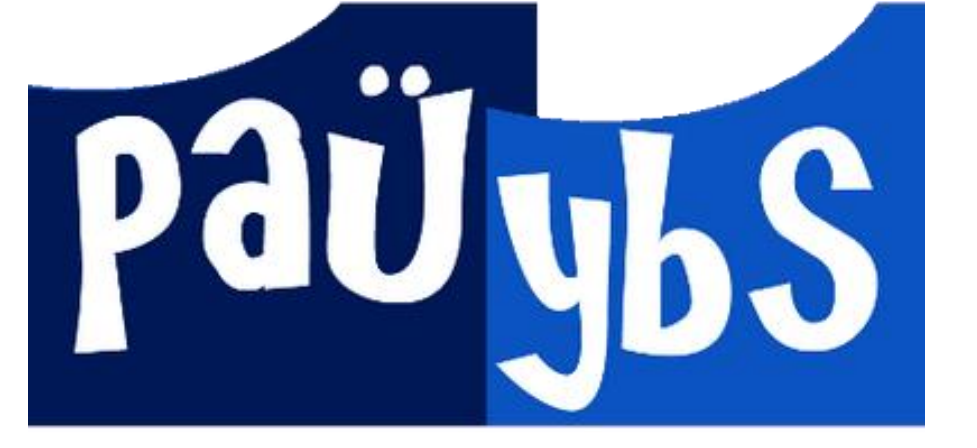




# ÇEŞİTLİ KANSER TÜRLERİNDE SINIFLANDIRMA ALGORİTMALARI KULLANILARAK VERİ MADENCİLİĞİ

Muhammed CEYLAN  
Fatih KARTAL  
Levent KARABIYIK

Prof. Dr. Selçuk Burak HAŞILOĞLU  
Öğr. Gör. Hamid YEŞİLYAYLA



Yönetim Bilişim Sistemleri Bölümü , Pamukkale Üniversitesi, Denizli

## Giriş

Çalışmada PAU Hastaneleri onkoloji bölümü hastalarına uygulanan 2012-2018 yılları arası tümör markerleri sonuçları kullanılmıştır. Belirlenen tümör markerleri ve kanser türleri sonucunda toplam 142 kanser tanısı konmuş hasta veri setini oluşturmaktadır. Veri üzerinde yapılan işlemler sonucunda 3 kanser türü belirlenmiştir. Belirlenen kanser türleri akciğer, karaciğer ve kolon kanseridir. Veri madenciliği işlemleri için veri setini 69 hasta oluşturmaktadır. Lisans tez çalışmasının amacı; onkoloji bölümü hastalarının verileri üzerinden kanser tanıları tahmin ederek kullanılan parametrelerin sonuçlara nasıl etki ettiğine dikkat çekmek ve doktorlara hasta teşhisinde bulunurken öneri sunmaktır.

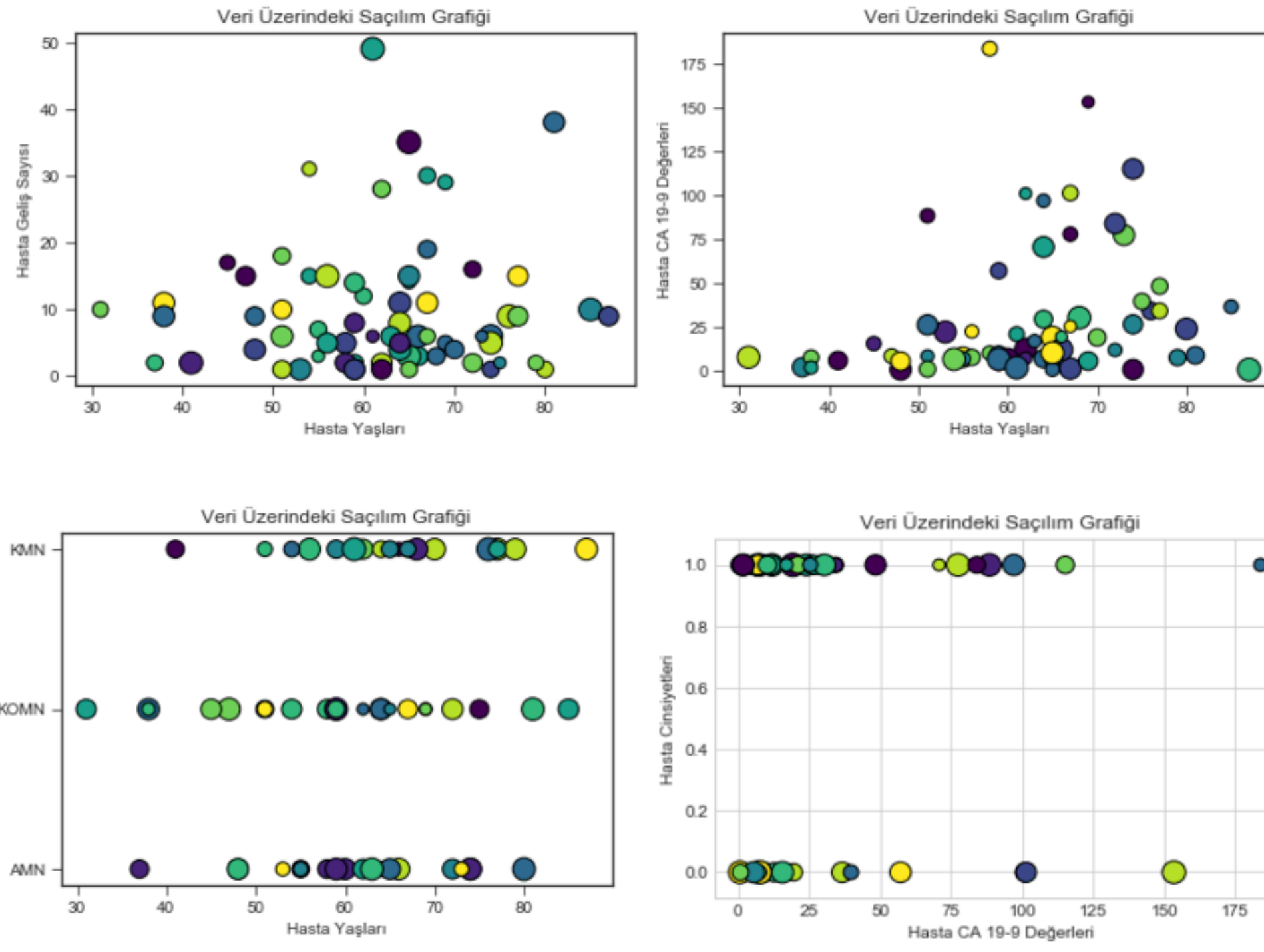
## Yöntemler

Bu çalışma Python dili kullanılarak gerçekleştirilmiştir. Python diline ait veri madenciliği alanında kullanılan Numpy, Pandas, Matplotlib, Sklearn kütüphaneleri ve bu kütüphanelere ait metrikler olan accuracy, recall, precision, sensitivity, specificity, F1 score ve support parametreler kullanılmıştır. Veri seti üzerinde Support Vector Machine, K nearest neighborhood ,Logistic Regression, Random Forest, Decision Tree, Naive Bayes ve Linear Discriminant Analysis sınıflandırma algoritmaları kullanılmıştır.

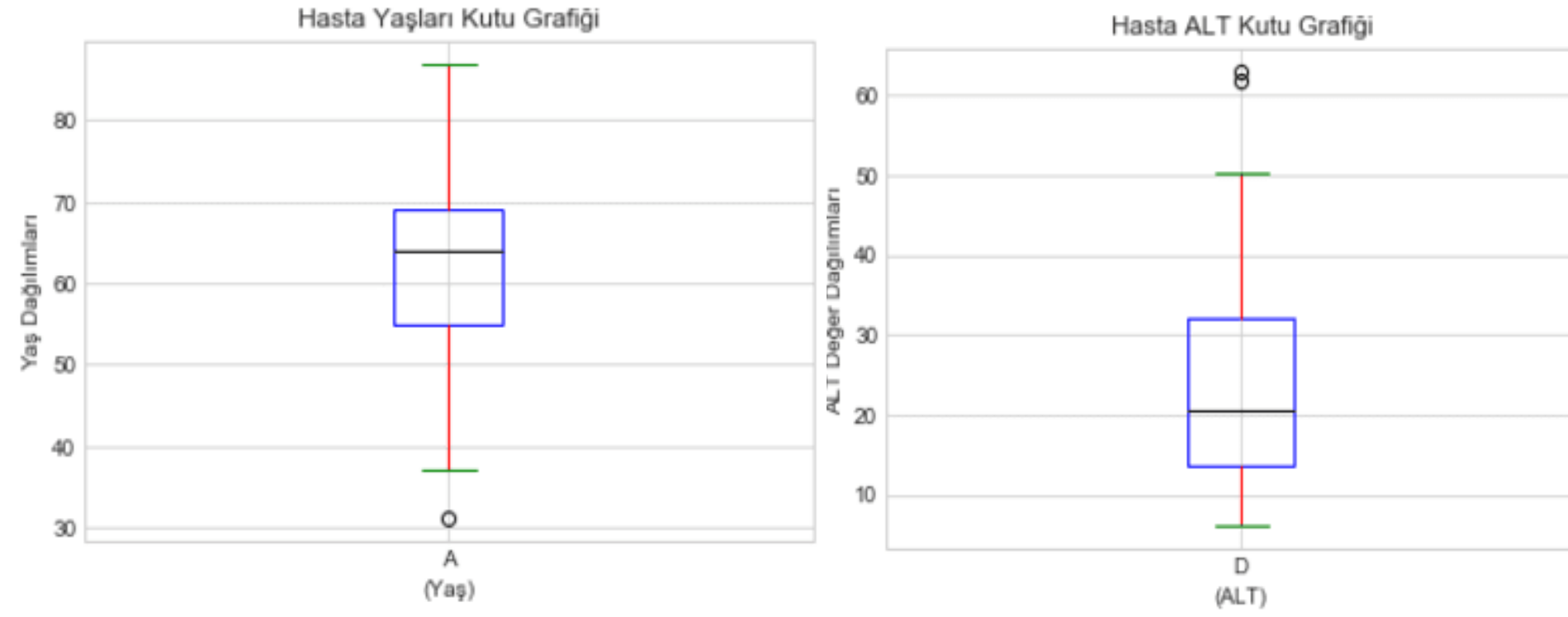
## Veri Seti

Pamukkale Üniversitesi hastaneleri onkoloji bölümü 2012-2018 yılları arası hasta verileri kullanılmıştır. İlk alınan veriler üzerinde yapılan literatür taraması sonucunda belirlenen parametrelere sahip hastalar seçilmiştir. Belirlenen parametreler Karbonhidrat antijeni 19-9 (CA 19-9), Alfa-feto protein (AFP), Alanin aminotransferaz (ALT), Karsinoembriyonik antijen (CEA) ve C-Reaktif protein (CRP) parametreleri kullanılmıştır. Veri setinde 36 akciğer kanseri, 39 karaciğer kanseri ve 67 kolon kanseri olmak üzere ilk aşamada 142 hasta bulunmaktadır. Bu hastalara ait parametre sonuçlarının çeyrek açıklık yöntemi kullanılarak aykırı değerleri tespit edilip verinin ikinci kısmı tamamlanmıştır. Uç değerlerin ortadan kaldırılması sonucunda 23 akciğer kanseri , 32 karaciğer kanseri ve 38 kolon kanseri olmak üzere verinin ikinci aşamasında toplam 93 hasta bulunmaktadır. Son olarak sınıflandırma algoritmalarında daha iyi sonuçlar elde etmek için hasta sayısı tüm kanser türlerinde 23 olarak belirlenmiştir. Veri seti toplamda 69 hastadan oluşmaktadır.

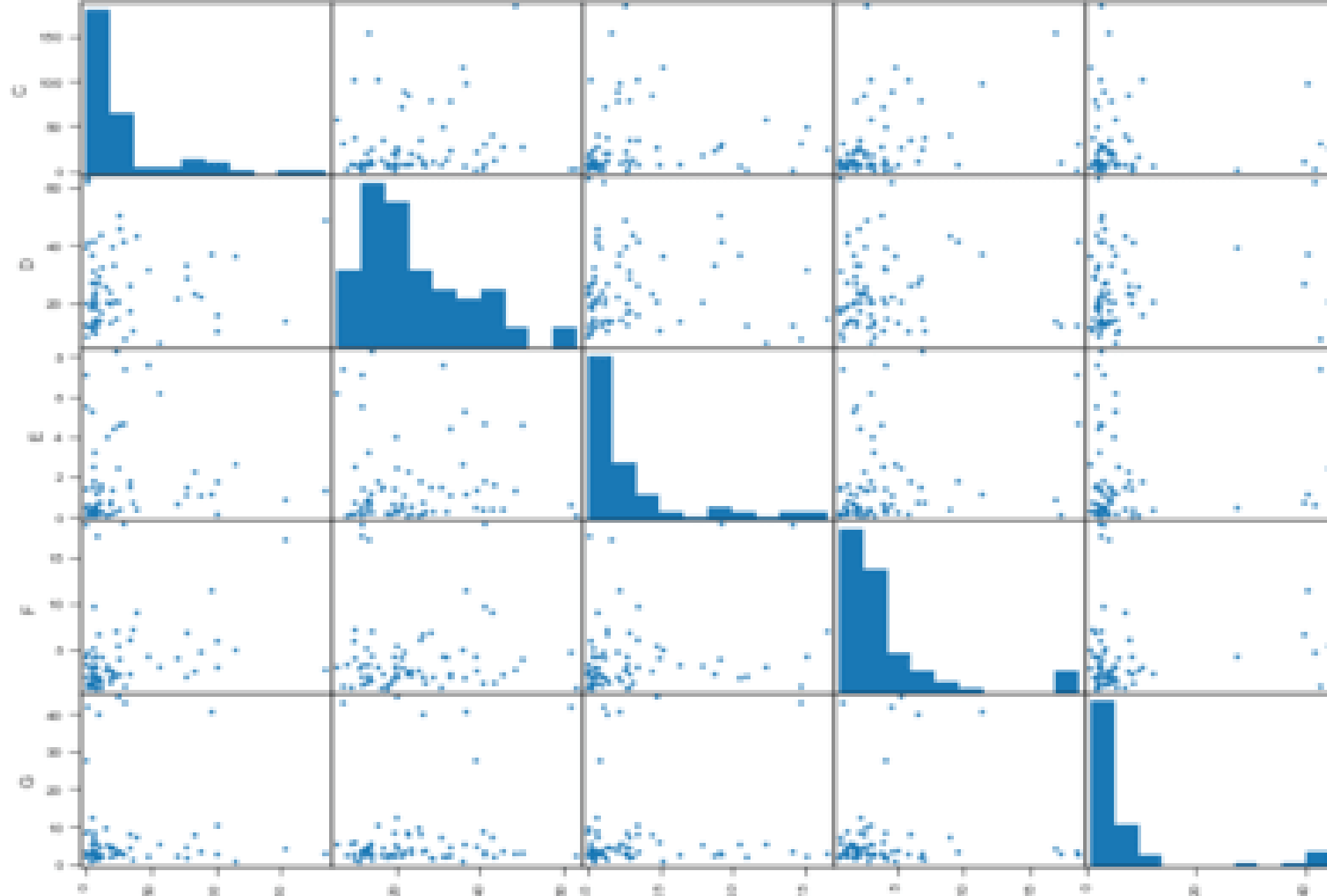
## Analiz



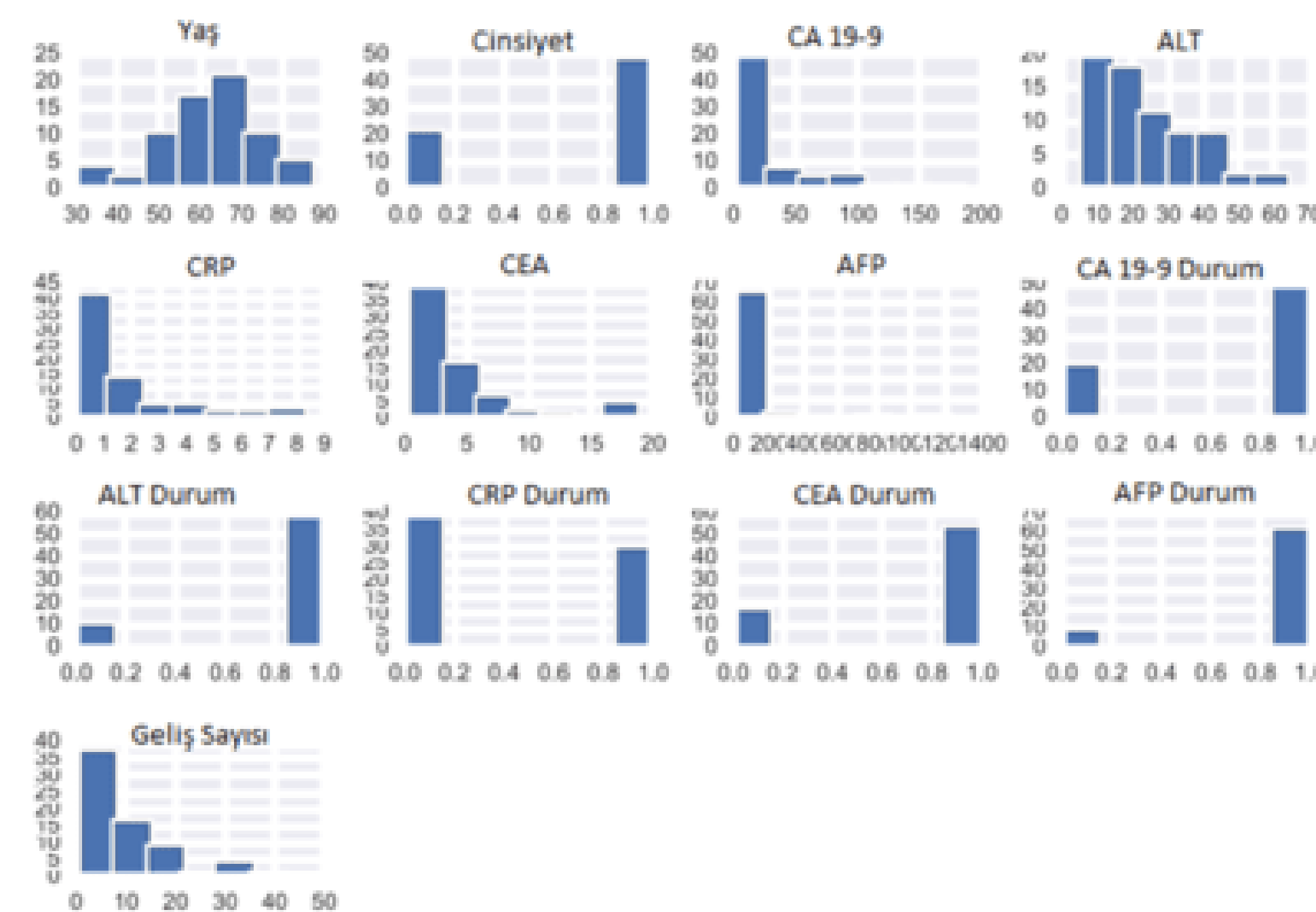
Şekil 1. Scatter (Saçılım) Grafikleri



Şekil 2. Boxplot (Kutu) Grafikleri



Şekil 3. Scatter Matris Grafiği

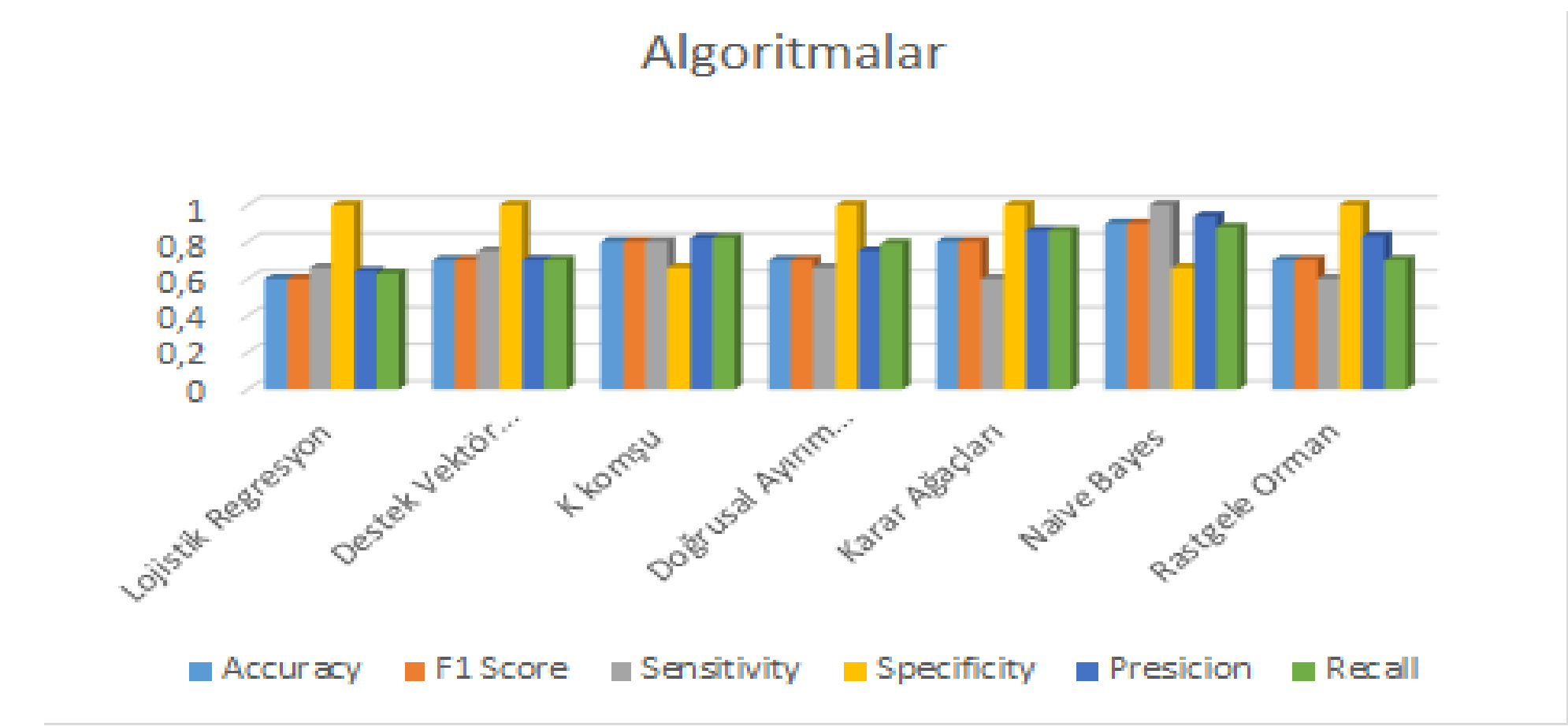


Şekil 4. Histogram Matris Grafiği

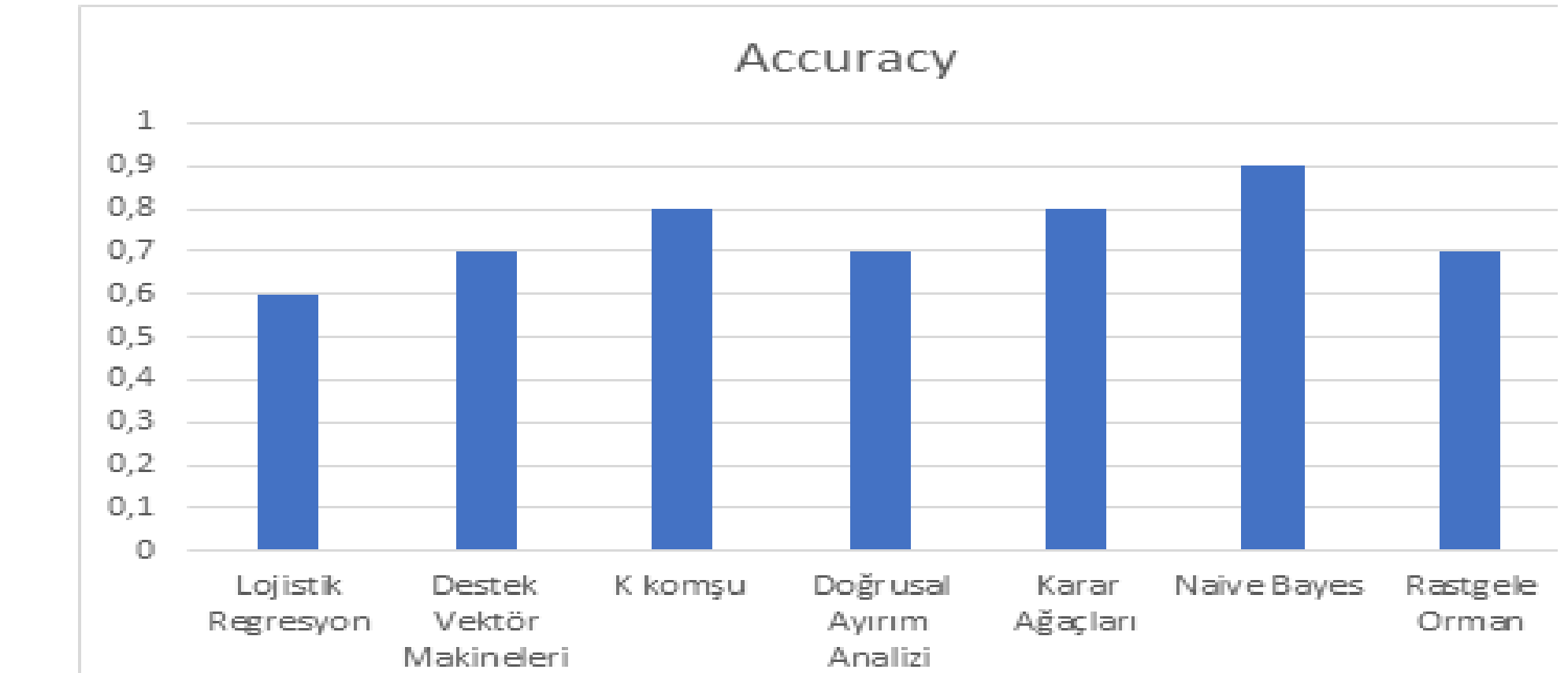
## Sonuçlar

Algoritmalar	Accuracy	F1 Score	Sensitivity	Specificity	Presicion	Recall
Lojistik Regresyon	0,6	0,6	0,66	1	0,64	0,63
Destek Vektör Makineleri	0,7	0,7	0,75	1	0,7	0,7
K komşu	0,8	0,8	0,8	0,66	0,82	0,82
Doğrusal Ayırım Analizi	0,7	0,7	0,66	1	0,75	0,79
Karar Ağaçları	0,8	0,8	0,6	1	0,86	0,86
Naive Bayes	0,9	0,9	1	0,66	0,94	0,88
Rastgele Orman	0,7	0,7	0,6	1	0,83	0,7

Şekil 5. Algoritma metrik sonuçları



Şekil 6. Algoritmaların Metrik Sonuçları Grafiği



Şekil 7. Algoritmaların Accuracy Sonuçları Grafiği

Kullanılan sınıflandırma algoritmaları arasında en iyi sonucu Naive Bayes sınıflandırma algoritması vermektedir ve bu sınıflandırma algoritması kanser tanısı tahminlemede %90 oranında doğruluk seviyesine ulaşmıştır.

## Gelecekte Yapılacaklar

Sınıflandırma algoritmalarında elde edilen değerlerin daha iyi ve doğru sonuçlar vermesi adına Ensemble metotların kullanılması. Bu veriler üzerinden E-Sağlık alanında çalışmaların gerçekleştirilmesi. Pamukkale Üniversitesi' ne ait olan veriler ile ilerleyen aşamalarda bildiri ve makale çalışmalarını gerçekleştirmek.

## Kaynakça

- [1] Yayla N. (2014). Ratlarda Parasetamolle İndüklenen Akut Karaciğer Toksisitesi Üzerine Nigella Sativa L. Etanol Ekstresinin Etkilerinin Araştırılması, Atatürk Üniversitesi, Eczacılık Biyokimya Anabilim Dalı.
- [2] Kılıç L. Ve Aykan F. Onkoloji ve Tümör Belirteçleri, İstanbul Üniversitesi Onkoloji Enstitüsü.
- [3] Doğan R. (2009). Karaciğer Kist Hidatiği Hastalığında Kistin Safra Yolları ile İlişisini Belirlemede Tümör Markerleri ve Biyokimyasal Değerlerin Katkısı, Harran Üniversitesi.
- [4] Çolak T. (2007). İstatistiksel Süreç Kontrolü ve Uygulamalar, Çukurova Üniversitesi.
- [5] Kavzoğlu T. ve Çölkese İ. (2010, Temmuz). Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi, Harita, 144.