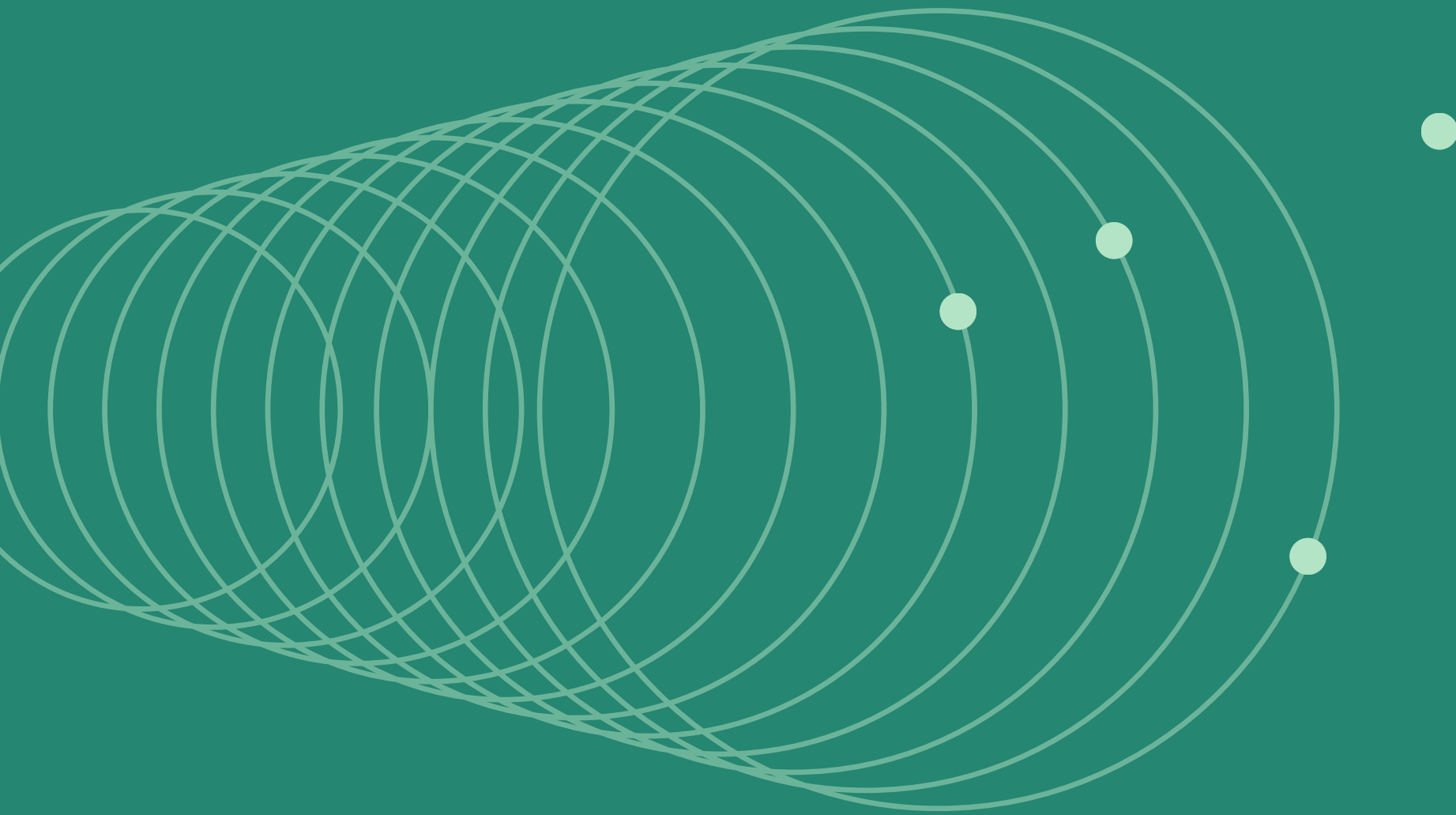




Breast Cancer K Nearest Neighbors

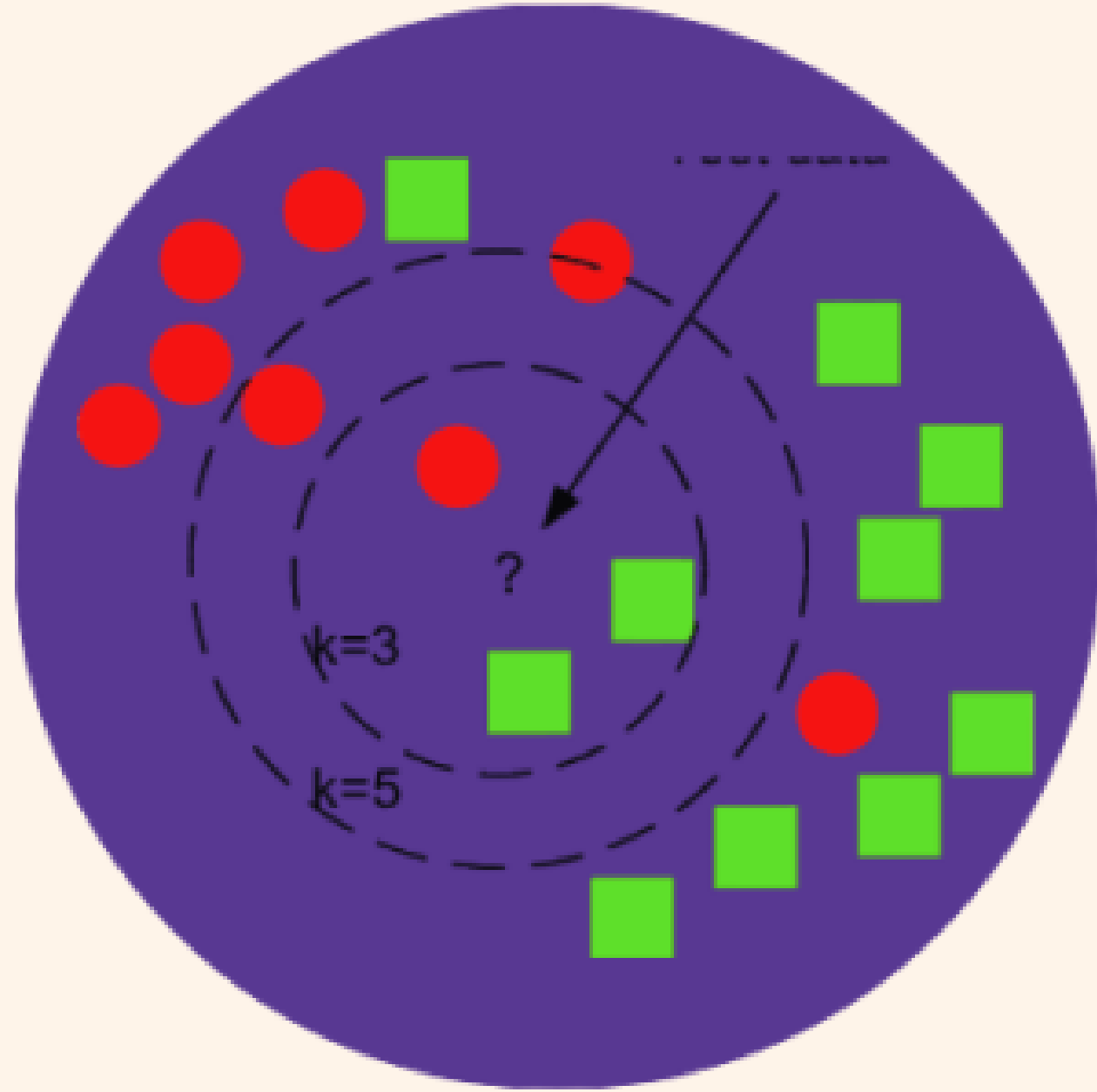
Uğur Can Akçay
18360859017

Giriş



Sınıflandırma Yöntemleri

- Decision Tree based Methods
- Rule-based Methods
- Nearest-neighbor
- Naïve Bayes and Bayesian Belief Networks
- Nearest-neighbor
- Neural Networks
- Deep Learning
- Support Vector Machines



K-En Yakın Komşu (KNN)

KNN, makine öğrenimi ve veri madenciliği alanında sıkça kullanılan bir sınıflandırma ve regresyon algoritmasıdır. Temel prensibi, benzerlik ölçütleri kullanarak yeni verileri sınıflandırmak veya tahmin etmektir.

K-Nearest Neighbors

Çalışma Mantığı

- Veri noktalarını temsil etme: Veri setindeki her bir örneği bir nokta olarak düşünürüz. Bu noktalar, veri özelliklerine göre konumlandırılır.
- Sınıflandırma işlemi: Yeni bir veri noktası geldiğinde, KNN algoritması bu noktanın en yakın komşularını bulur. Komşuların sınıflarına bakarak, yeni veri noktasını ilgili sınıfa atar.
- Sınıflandırma için en yaygın yöntem, en yakın K komşunun çoğunluğunun sınıfını tahmin etmektir.

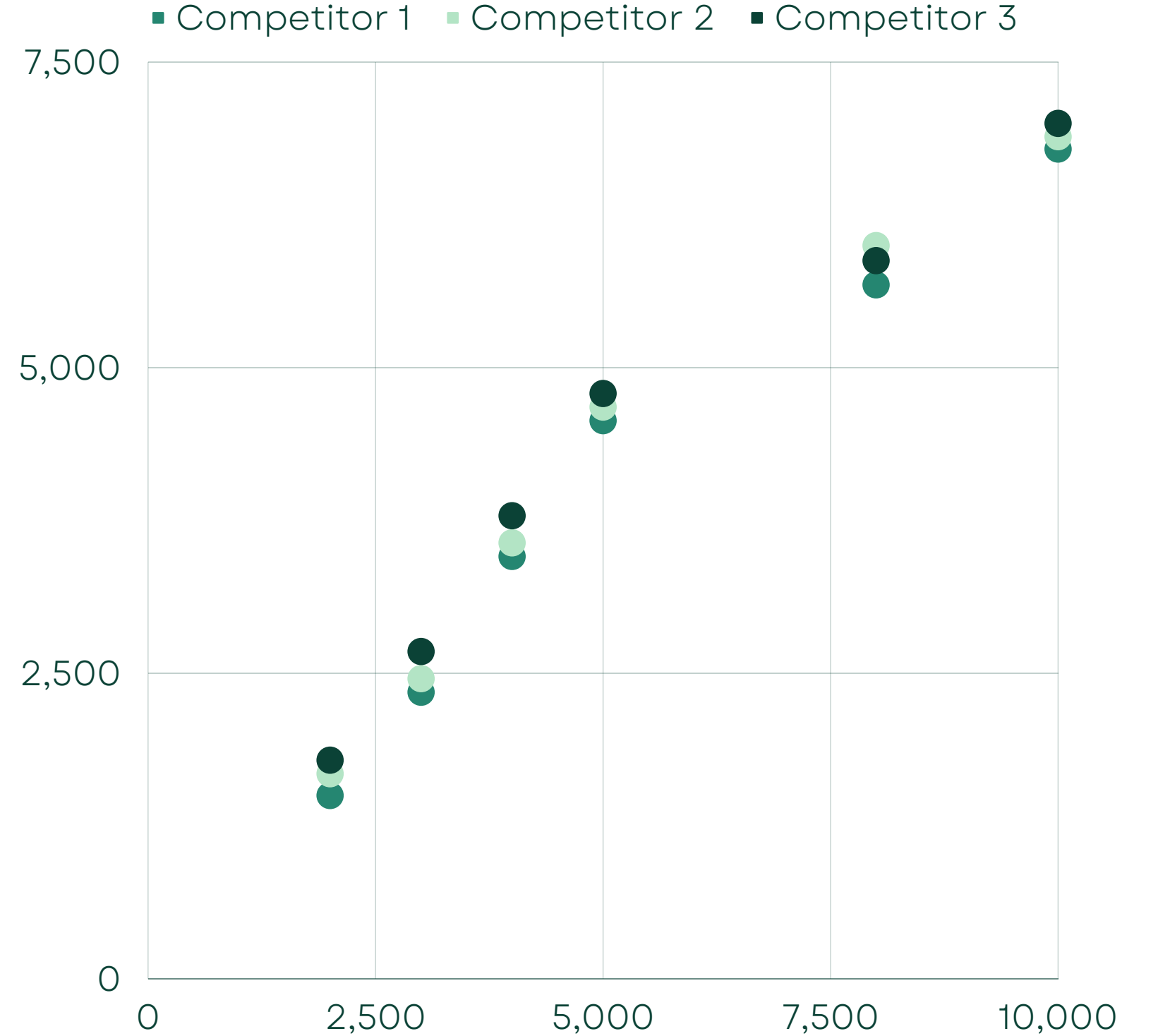
Çalışma Adımları

Veri Analizi	Veri Ön İşleme	Modelleme	Karşılaştırma
<ul style="list-style-type: none">Veri setinin yapısını ve içeriğini anlamak için veri setini inceleyin.Veri setindeki değişkenlerin türünü (sayısal veya kategorik) belirleyin.	<ul style="list-style-type: none">Eksik verileri ele alın. Eksik verileri doldurma veya eksik veri içeren örnekleri çıkarma gibi bir strateji uygulayın.Kategorik değişkenleri sayısal değerlere dönüştürün veya kategorik kodlamalar yapın.	<ul style="list-style-type: none">KNN algoritmasını kullanarak sınıflandırma veya regresyon modelini oluşturun.KNN modelinin hiperparametrelerini belirleyin, örneğin, K değeri, uzaklık metriği gibi.	<ul style="list-style-type: none">Karşılaştırma adımı, KNN modelinin performansını değerlendirmek için daha önce yapılmış benzer bir çalışma ile sonuçları karşılaştırmayı içerir.

Veri Setinin İncelenmesi

Veri seti, meme kanseri teşhisinde kullanılan özelliklerin ölçümlerini içerir.

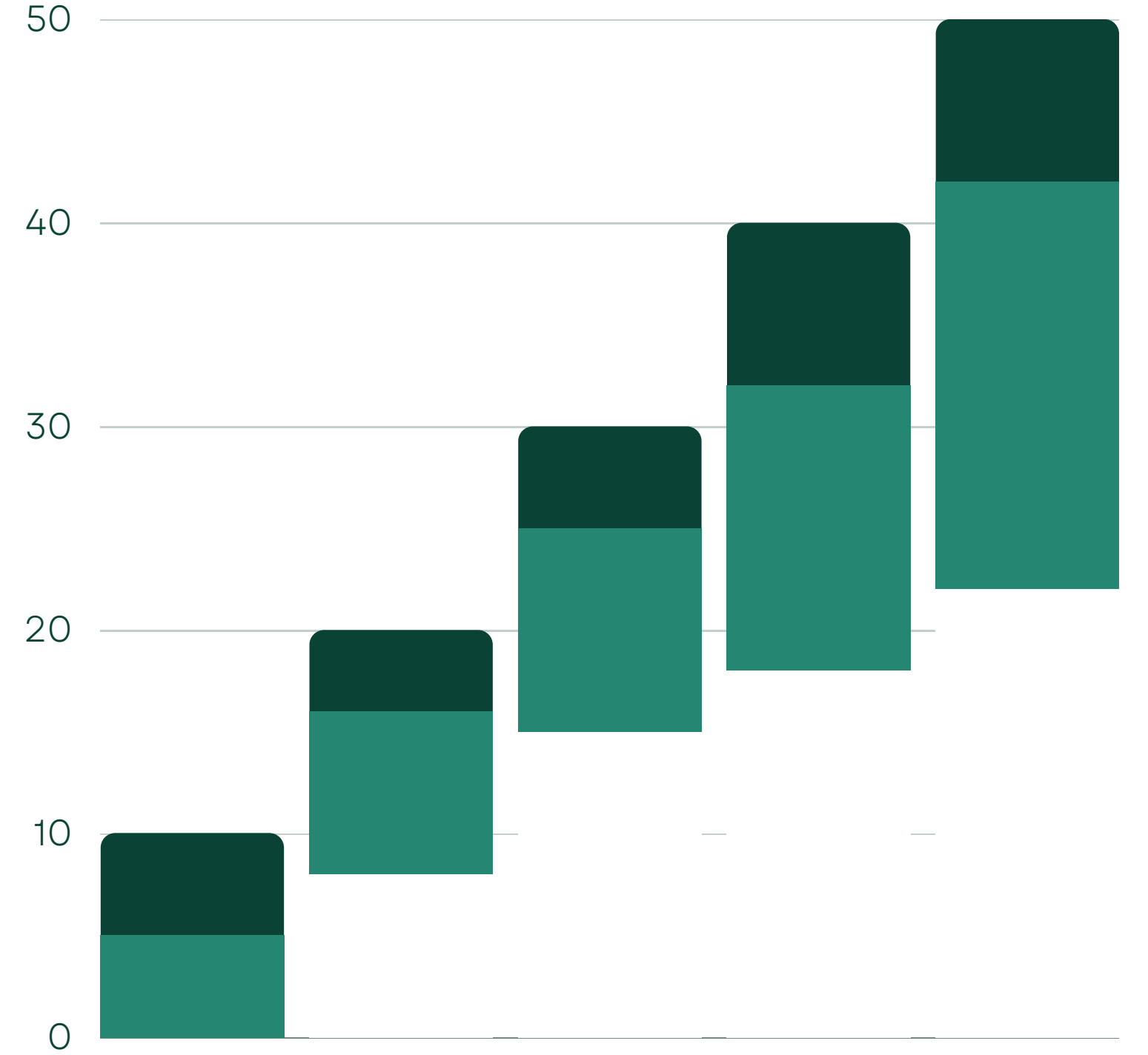
Veri seti, Wisconsin Üniversitesi Hastanesi tarafından toplanan ve daha sonra UCI Machine Learning Repository'de paylaşılan 699 hasta örneğini içerir. Her bir hastanın 10 farklı özelliğe ait sayısal değerlerden oluşan bir vektörü bulunmaktadır.



Veri Ön İşleme

Veri Ön İşleme aşaması, veri setinin analiz ve modelleme için uygun hale getirilmesini sağlar.

Veri Ön İşleme aşamasında, eksik değerlerin bulunduğu gözlemler çıkarılabilir veya eksik değerler yerine ortalama, medyan veya en yakın komşu değerleriyle doldurulabilir



Modelleme

01

Veri seti bölünerek eğitim ve test kümeleri oluşturuldu.

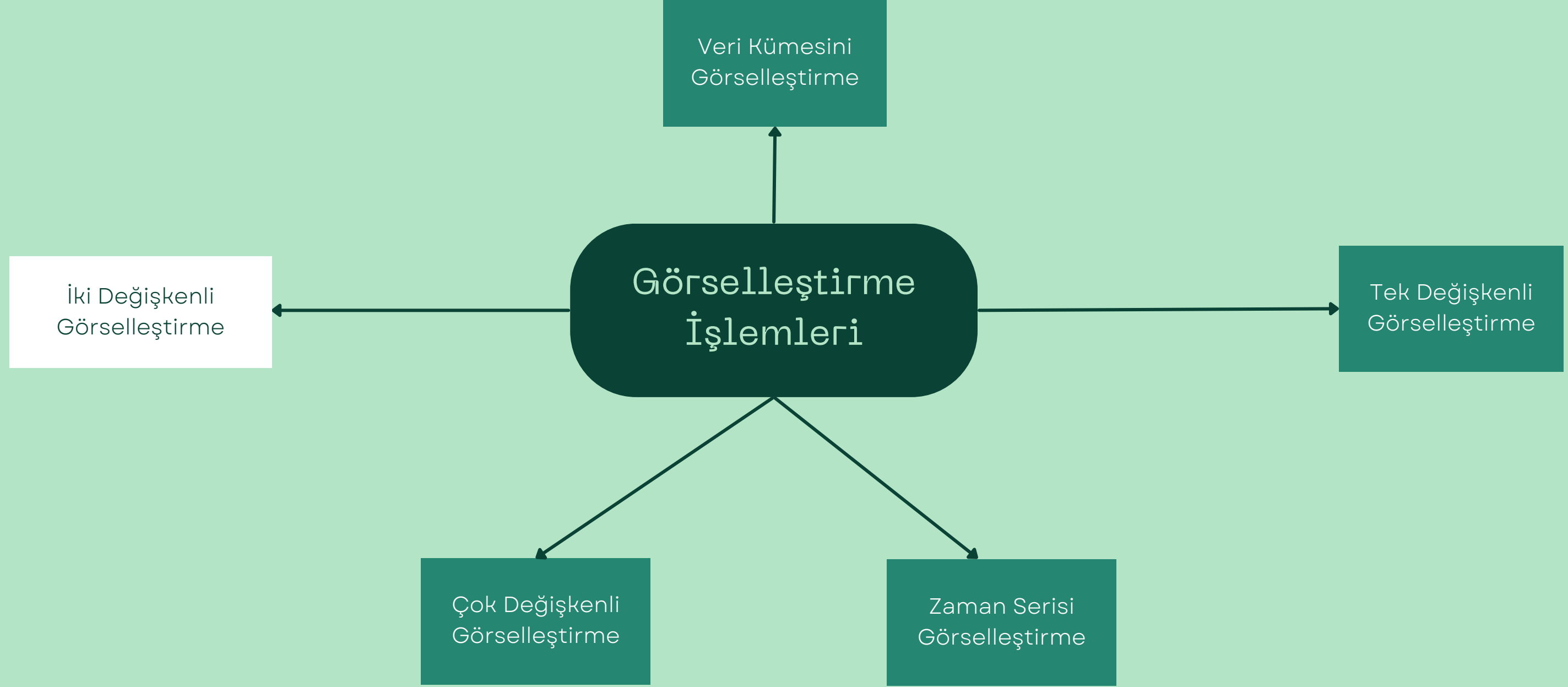
02

KNN modeli eğitim verileriyle eğitildi ve parametreleri ayarlandı

03

Modelin eğitim doğruluğu ve performansı değerlendirildi.

Görselleştirme



Bu başlıklar altında yer alan görselleştirme işlemleri, veri analizi sürecinde verileri anlamak, ilişkileri keşfetmek ve sonuçları görsel olarak sunmak için kullanılan araçlardır.

Sonuçların Karşılaştırılması



İkilinin
Accuracy değeri
0.96%

Ugur Can AKCAY
0.98%

	precision	recall	f1-score	support
0	0.96	0.98	0.97	121
1	0.97	0.93	0.95	67
accuracy			0.96	188
macro avg	0.96	0.95	0.96	188
weighted avg	0.96	0.96	0.96	188

...	precision	recall	f1-score	support
2	0.99	0.99	0.99	91
4	0.98	0.98	0.98	49
accuracy			0.99	140
macro avg	0.98	0.98	0.98	140
weighted avg	0.99	0.99	0.99	140

Accuracy	0.96	0.985
Sensitivity	1.0(Neighbors>320)	1.0
Specificity	0.98	0.978
Fmeasure	0.94	0.978



Marjan Ghazisaeedi ve Arash Roshanpoor The Performance of K-Nearest Neighbors on Malignant and Benign Classes: Sensitivity, Specificity, and Accuracy Analysis for Breast Cancer Diagnosis 2017'li yazısına göre karşılaştırmalar yapılmıştır.

Teşekkürler

