	# - Number	- Legal Fees (USD): Dava süresince oluşan toplam hukuk maliyetleri (USD cinsinden) - Number of Evidence Items: Davada kullanılan delil sayısı - Severity: Davanın ciddiyet düzeyi (1: Düşük, 2: Orta, 3: Yüksek) - Outcome: Davanın sonucu (0: Aleyhte, 1: Lehinde)													
		Özellik Seçimi ve Ölçeklendirme													
In [128	features	Ozellik Seçimi ve Olçeklendirme features = ['Case Duration (Days)', 'Number of Witnesses', 'Legal Fees (USD)', 'Number of Evidence Items', 'Severity']													
		<pre>[features] StandardScaler()</pre>													
	X_scaled	= scaler.fit_tran	sform(X)												
	Gözlem: • Seçilen değişkenler, davaların özelliklerini ve ciddiyetini temsil ediyor.														
	• Ölçeklendirme ile tüm değişkenler aynı ölçeğe getirildi, böylece K-Means algoritması için uygun hale geldi.														
	Yorum: Kümeleme	ie analizi, davaların sü	üresi, maliyeti ve cid	diyeti gibi özellikle	rin benzerliklerine	e göre gruplandırma y	apacak.								
	Elbow Y	⁄öntemi ile Optim	nal Küme Sayısın	ı Belirleme											
In [134	inertia =														
	<pre>for k in kmear</pre>	<pre>= range(1, 11) k_range: ns = KMeans(n_clus</pre>	sters=k, random_st	cate=42)											
		<pre>ns.fit(X_scaled) tia.append(kmeans.</pre>	inertia_)												
	plt.plot	<pre>re(figsize=(8,5)) (k_range, inertia, el('Küme Sayısı (k</pre>	·												
	plt.ylabe	el('Inertia') e('Elbow Yöntemi i		Sayısı')											
		silas\anaconda3\Li	·	_		UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak o	n Windows with M	(L, when there	are less chunks	than available	threads.
	warnings	void it by setting s.warn(silas\anaconda3\Li			_	UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak c	n Windows with M	(L, when there	are less chunks	than available	threads.
	You can avoid it by setting the environment variable OMP_NUM_THREADS=1. warnings.warn(C:\Usens\silas\anaconda3\Lib\site_nackages\sklearn\cluster\ kmeans py:1429: UsenWanning: KMeans is known to have a memony leak on Windows with MKL, when there are less chunks than available threads.														
	You can av	C:\Users\silas\anaconda3\Lib\site-packages\sklearn\cluster_kmeans.py:1429: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OMP_NUM_THREADS=1. warnings.warn(C:\Users\silas\anaconda3\Lib\site-packages\sklearn\cluster_kmeans.py:1429: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads.													
	You can av	void it by setting s.warn(the environment	variable OMP_NUM	_THREADS=1.										
		void it by setting	·	_		UserWarning: KMear	is is kno	wn to hav	ve a mem	ory leak d	n Windows with Mk	(L, when there	are less chunks	than available	threads.
		void it by setting		_		UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak o	n Windows with Mk	(L, when there	are less chunks	than available	threads.
	C:\Users\s	silas\anaconda3\Li void it by setting		_		UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak c	n Windows with MH	(L, when there	are less chunks	than available	threads.
	C:\Users\s You can av	silas\anaconda3\Li void it by setting		_		UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak o	n Windows with MH	(L, when there	are less chunks	than available	threads.
		•		_	• •	UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak c	n Windows with M	(L, when there	are less chunks	than available	threads.
		•		_	• •	UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak c	n Windows with MI	(L, when there	are less chunks	than available	threads.
	warnings	,		ni ile Optimal											
	500 -	•													
	450 -														
	400 -														
	Ē 350 -														
	lner														
	300 -														
	250 -														
	200 -														
	L	2	4	6 Küme Sayısı (k)	8	10									
	Gözlem:														
	Elbow gra	afiğinde inertia değer	rleri küme sayısı artt	ıkça azalıyor. Grafik	cte belirgin bir "di	rsek" noktası k=3 civa	ırında göz	emleniyo	r. Bu, veri	seti için op	timal küme sayısını	n yaklaşık 3 oldu	ğunu gösteriyor.		
	Yorum: Kümeleme	ıe için üç farklı grup c	oluşturmak mantıklı	görünüyor. Elbow y	yöntemi, inertia'n	ın azalmaya devam et	mesine ra	ğmen dirs	ek nokta:	sından sonr	a düşüşün yavaşlad	ığını gösterir; bu,	daha fazla küme	eklemenin veri iç	; i
						hem de model sadeli <u>č</u>	,								
	K-Mean	ns ile Kümeleme													
In [138		<pre>k = 3 KMeans(n_clusters = kmeans.fit_pred</pre>		om_state=42)											
	data['Cludata.head	<pre>uster'] = clusters d()</pre>													
	You can av	void it by setting		_		UserWarning: KMear	ns is kno	wn to hav	ve a mem	ory leak o	n Windows with M	(L, when there	are less chunks	than available	threads.
Out[138	warnings Unnan	·	on (Days) Number	of Witnesses Lega	al Fees (USD) Nu	ımber of Evidence Iten	ns Severi	ty Outco	ome Clu	ster					
	0	0	300	8 12	35206 46106		49 46	2	0	2					
	2	2	136	2	16411	:	31	1	1	1					
	3	3 4	101 218	1 16	7911 6734		35	1	0	0					
	Gözlem:														
	• Veri s	setine K-Means küme		her satır için bir Clu	uster etiketi oluştı	uruldu.									
	 3 farklı küme (0, 1, 2) tespit edilmiş. Bazı davalar yüksek Case Duration ve Legal Fees değerlerine sahipken bazıları düşük değerlerde kümelenmiş. 														
	• Cluster değerleri, verideki sayısal özelliklerin birleşik dağılımını yansıtıyor. Yorum:														
	• Kümeleme, davaları benzer özellikler üzerinden gruplandırmış: \														
			·	•	•	luster 1 ise kısa süreli	ve düşük	maliyetli c	lavalar içi	n.					
		ayede, dava türlerine	•		on analiz ve sınıfla	ndırma yapılabilir.									
In [150		me Sonuçlarının re(figsize=(8,6))	Görselleştirilme	Sİ											
111 [130	<pre>plt.scatter(data['Case Duration (Days)'], data['Legal Fees (USD)'], c=data['Cluster'], cmap='viridis') plt.xlabel('Case Duration (Gün)')</pre>														
	-	el('Legal Fees') e('K-Means Kümelem ()	ne Sonuçları')												
	<figure si<="" td=""><td>ize 800x600 with 0</td><td></td><td>ns Kümeleme</td><td>Sonucları</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></figure>	ize 800x600 with 0		ns Kümeleme	Sonucları										
	50000	-	13 14 EQ	•	•	•									
			•		•	•									
	40000	• •	•			•									
		•	•	•••		•									
	9 30000	-	•		•	•									
	Legal	•	•			•									

20000 -

10000

Gözlem:

Yorum:

Genel Analiz

Sonuç olarak:

100

150

ile ölçeklendirildi; bu, K-Means algoritmasının mesafe tabanlı mantığı için gerekliydi.

• Küme 0: Daha düşük maliyet ve kısa dava süresi

• Küme 2: Yüksek maliyet ve uzun süren davalar

• Küme 1: Orta düzey maliyet ve süre

200

Case Duration (Gün)

Farklı kümeler farklı renklerle görselleştirilmiş. Kümeler, davaların süresi ve maliyetine göre ayrılmış gibi görünüyor.

50

300

K-Means ile yapılan kümeleme, davaları maliyet ve süre özelliklerine göre anlamlı bir şekilde gruplandırıyor. Bu analiz, davaların hangi özelliklerde benzer olduğunu belirlemek için kullanılabilir.

350

Seçilen özellikler: Case Duration (Days), Number of Witnesses, Legal Fees (USD), Number of Evidence Items, Severity, dava süresi ve maliyetleri, delil sayısı gibi önemli faktörleri temsil ediyor. Sayısal değişkenler StandardScaler

Bu analiz, veri setindeki sınıf dengesizliğinin model performansını ve yorumlanabilirliğini doğrudan etkilediğini gösteriyor. Outcome değişkeninde farklı sınıflar olsaydı, karar ağacı görselleştirmesi ile hangi faktörlerin dava

sonucunu daha çok etkilediği net bir şekilde yorumlanabilirdi.Bu analiz, dava örneklerini maliyet, süre ve delil sayısı gibi faktörlere göre gruplandırmanın mümkün olduğunu gösteriyor. Kümeleme, dava yönetimi ve kaynak

250

Yapılan veri ön işleme, özellik seçimi ve K-Means kümeleme görselleştirmeleri, dava süreçlerindeki farklı örüntüleri ve gruplamaları anlamamıza yardımcı oldu.

Kümeleme sonuçları 3 küme olarak belirlendi. Grafikler, davaların farklı yoğunluk ve maliyet kombinasyonlarına göre ayrıldığını gösterdi. Örneğin:

Aykırı değerler bazı davalarda gözlemlendi (örneğin yüksek Legal Fees), fakat K-Means algoritması bu tür uç değerlere karşı nispeten dayanıklı.

planlaması açısından potansiyel olarak yol gösterici olabilir; örneğin yüksek maliyetli uzun süren davalar için farklı stratejiler geliştirilebilir.

In [110... # K-Means Kümeleme Ödevi — Veri Analizi Notebook

import matplotlib.pyplot as plt

data = pd.read_csv('dava.csv')

0

2

3

4

from sklearn.cluster import KMeans

import numpy as np

import pandas as pd

%matplotlib inline

data.head()

In [116... # Veri Seti Açıklaması:

3

In [112...

In [114...

Out[114...

Bu notebook .ipynb olarak çalıştırılmak üzere hazırlandı.

132

300

136

101

218

- Number of Witnesses: Dava boyunca dinlenen tanık sayısı

- Case Duration (Days): Davanın tamamlanması için geçen süre (gün cinsinden)

from sklearn.preprocessing import StandardScaler

İçinde hem kodlar hem de veriye dair analiz (gözlem ve yorum) hücreleri bulunuyor.

Unnamed: 0 Case Duration (Days) Number of Witnesses Legal Fees (USD) Number of Evidence Items Severity Outcome

8

12

1

16

35206

46106

16411

7911

6734

49

46

31

35

33

0

0

0