

Open in app ↗



Search

Get unlimited access to the best of Medium for less than \$1/week. [Become a member](#)

PCA (Principal Component Analysis) Temel Bileşenler Analizi

Gülcan Öğündür · [Follow](#)

2 min read · Jan 14, 2020



Share



More

Türkçesi “Temel Bileşenler Analizi” olan PCA tanıma, sınıflandırma, görüntü sıkıştırma alanlarında kullanılan yararlı bir istatistiksel tekniktir. Temel amacı yüksek boyutlu verilerde en yüksek varyans ile veri setini tutmak ancak bunu yaparken boyut indirgemeyi sağlamak olan bir tekniktir. Fazla boyutlu verilerdeki genel özellikleri bularak boyut sayısının azaltılmasını, verinin sıkıştırılmasını sağlar. Boyut azalmasıyla bazı özelliklerin kaybedileceği kesindir; fakat amaçlanan, bu kaybolan özelliklerin popülasyon hakkında çok az bilgi içeriyor olmasıdır. Bu yöntem, yüksek korelasyonlu değişkenleri bir araya getirerek, verilerdeki en çok varyasyonu oluşturan “temel bileşenler” olarak adlandırılan daha az sayıda yapay değişken kümesi oluşturur.

PCA verideki gerekli bilgileri ortaya çıkarmada oldukça etkili bir yöntemdir. PCA’ın arkasında yatan temel mantık çok boyutlu bir veriyi, verideki temel özellikleri yakalayarak daha az sayıda değişkenle göstermektir.

Veri setimiz üzerinde PCA yapabilmek için sklearn kütüphanesinden PCA metotunu çağırmamız gereklidir.

Öncelikle iris veri setimizi sklearn’den indirelim. iris veri setimizde 3 farklı sınıfa ait 4 öznitelik bulunmaktadır.

[illegible]

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	sinif
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

Veri seti üzerinde yapmış olduğumuz PCA dönüşümü bize 2 bileşen seçmenin, verilerin toplam varyansının yaklaşık % 97.7'sini koruyabileceğimizi söylemektedir. Tüm bileşenleri kullanmak istemediğimiz sadece **ana bileşenleri** kullanmak istediğimiz için bu varyans oranı bizim için yeterlidir.

```
In [5]: ## PCA
from sklearn.decomposition import PCA
pca = PCA(n_components = 2, whiten= True ) # whiten = normalize
pca.fit(x)

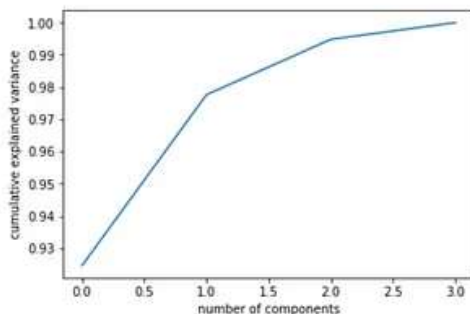
x_pca = pca.transform(x)

print("variance ratio: ", pca.explained_variance_ratio_)

print("sum: ",sum(pca.explained_variance_ratio_))

variance ratio: [ 0.92461621  0.05301557]
sum: 0.977631775025
```

```
In [6]: pca = PCA(whiten=True).fit(x)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance');
plt.show()
```



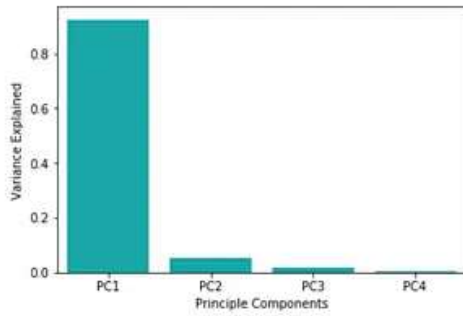
```
In [7]: df_sns = pd.DataFrame({'var':pca.explained_variance_ratio_,
                             'PC':['PC1','PC2','PC3','PC4']})

df_sns
```

```
Out[7]:
```

	PC	var
0	PC1	0.924616
1	PC2	0.053016
2	PC3	0.017185
3	PC4	0.005183

```
In [8]: sns.barplot(x='PC',y="var",data=df_sns, color="c")
plt.ylabel("Variance Explained")
plt.xlabel("Principle Components")
plt.show()
```



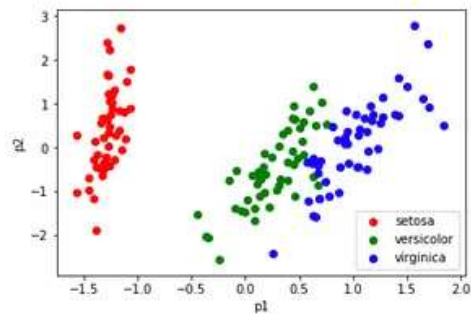
PCA dönüşümü işlemini tamamladıktan sonra sınıfların iki boyutlu vektörlerini aşağıdaki grafikte görebiliriz.

```
In [9]: df["p1"] = x_pca[:,0]
df["p2"] = x_pca[:,1]

color = ["red","green","blue"]
```

```
In [10]: for each in range(3):
    plt.scatter(df.p1[df.sinif == each],df.p2[df.sinif == each],color = color[each],label = iris.target_names[each])

plt.legend()
plt.xlabel("p1")
plt.ylabel("p2")
plt.show()
```



Principal Component

Pca

Feature Extraction



Follow

Written by Gülcan Öğündür

685 Followers

Business Intelligence Specialist at [sahibinden.com](https://www.sahibinden.com) in Istanbul. MS, Big Data and Business Analytics.
[linkedin.com/in/gulcanogundur/](https://www.linkedin.com/in/gulcanogundur/)

More from Gülcan Öğündür

Tahminlenen (Predicted)

(Actual)	Tahminlenen (Predicted)	
	True Positives (TP)	False Negatives (FN)
	False Positives (FP)	True Negatives (TN)



Gülcan Öğündür

Doğruluk (Accuracy) , Kesinlik(Precision) , Duyarlılık(Recall) ya da F1 Score ?

Veri bilimi projelerinde en doğru modelin hangisi olması gerektiğine karar vermek için iş birimlerinden gelen talepleri iyi...

3 min read · Nov 9, 2019



436



4



$$= \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$



Gülcan Öğündür

Mean (Ortalama), Mode(Mod), Median (Ortanca Değer), Standard Deviation (Standart Sapm

Bir verinin nasıl dağıldığını görebilmek için Histogram , box plot vb grafiklerden yararlanabilir ve görselleştirebiliriz. Ancak Mean...

3 min read · Nov 6, 2019

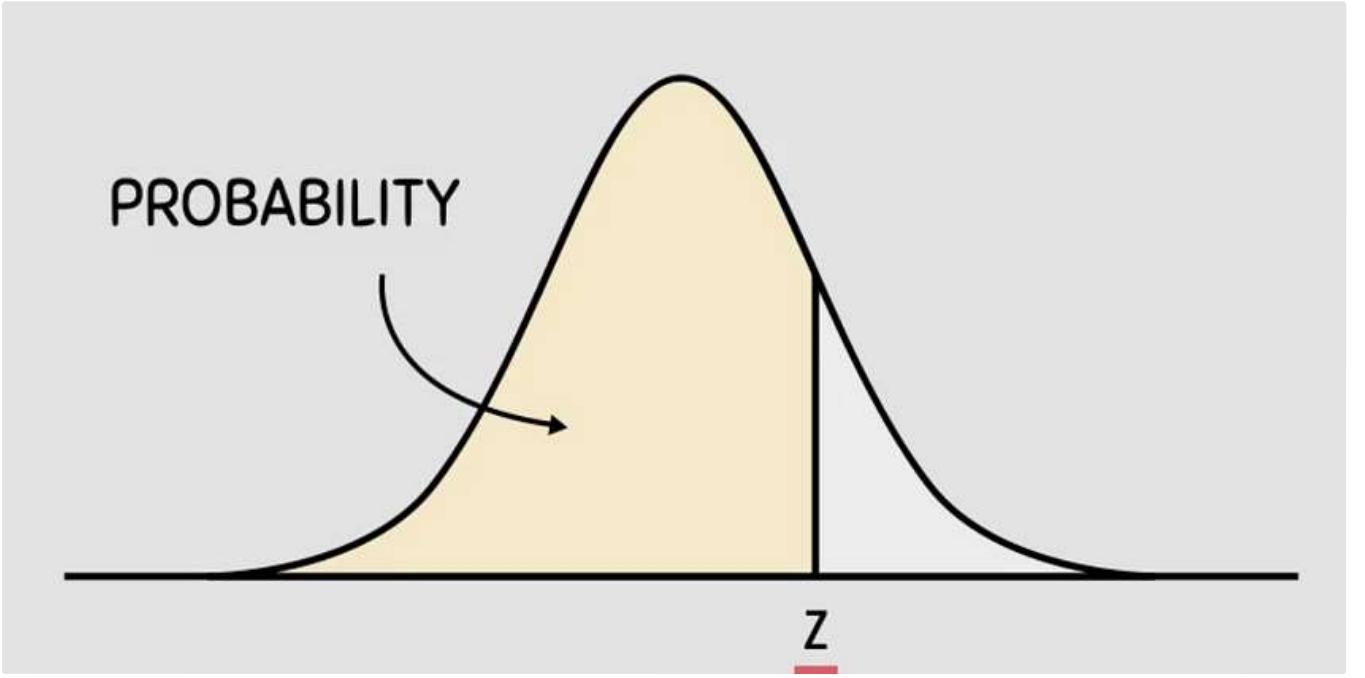


130



1





Gülcan Öğündür

Normal Dağılım , Z-Score ve Standardizasyon

Normal Dağılım çan şeklinde özel bir yoğunluk eğrisidir. Bu sebeple çoğu zaman çan eğrisi olarak da isimlendirilmektedir. Normal dağılım...

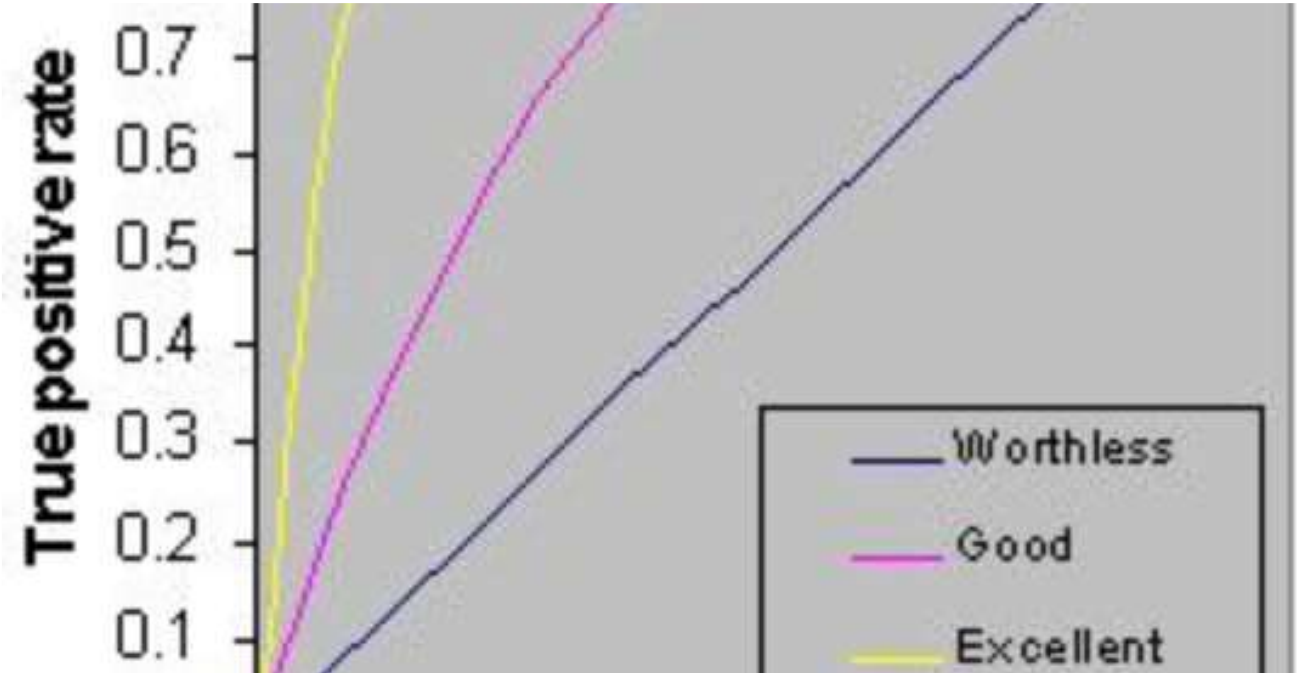
4 min read • Dec 25, 2019



155



1



Gülcan Öğündür

ROC ve AUC

ROC eğrisi sınıflandırma problemleri için çok önemli bir performans ölçümüdür. ROC bir olasılık eğrisidir ve altında kalan alan olan AUC...

2 min read · Jan 12, 2020



220



See all from Gülcan Öğündür

Recommended from Medium



Alexandru Lazar in ILLUMINATION

Ten Habits that will get you ahead of 99% of People

Improve your life and get ahead of your peers in 10 simple steps

9 min read · Nov 18, 2023

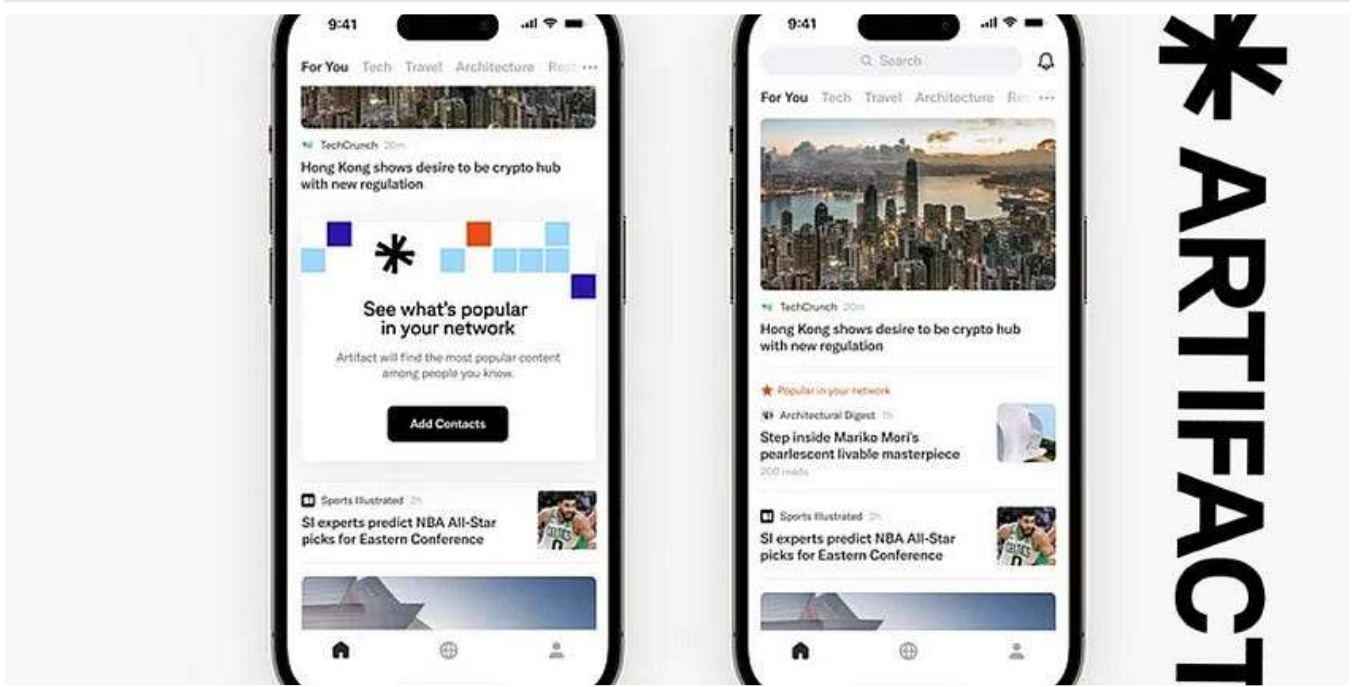


15.2K



273





 Gowtham Oleti

Apps I Use And Why You Should Too.

Let's skip past the usual suspects like YouTube, WhatsApp and Instagram. I want to share with you some less familiar apps that have become...

10 min read · Nov 14, 2023

 11.3K  199

Lists



Practical Guides to Machine Learning

10 stories · 860 saves



Unbecoming

10 Seconds That Ended My 20 Year Marriage

It's August in Northern Virginia, hot and humid. I still haven't showered from my morning trail run. I'm wearing my stay-at-home mom...

★ • 4 min read • Feb 16, 2022



72K



1030



Followers will receive notifications when you add a new product.

2112

Favorites

The number of times people added your products to their favorites.



\$413.01



Paul Rose

I Found 3 Passive Income Ideas Anyone Can Start

Build your First or Next Passive Income Stream Today

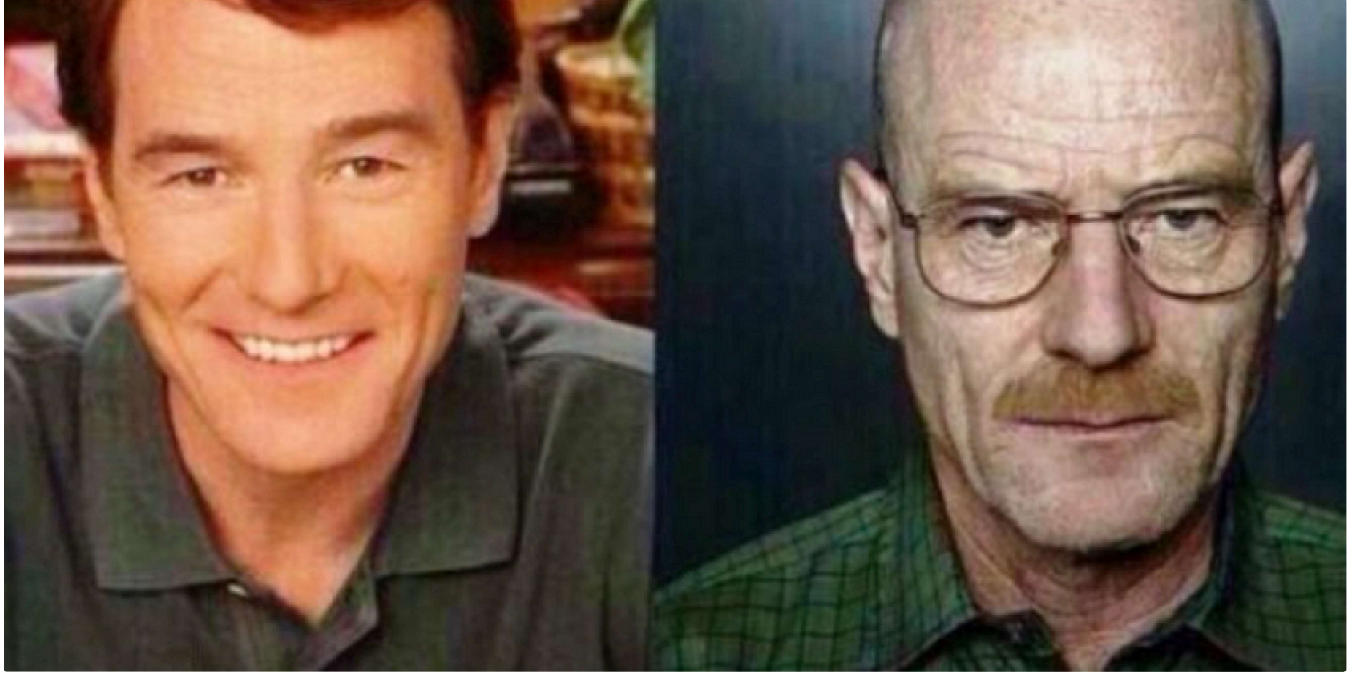
9 min read · Nov 3, 2023



9.1K



200



David Goudet

This is Why I Didn't Accept You as a Senior Software Engineer

An Alarming Trend in The Software Industry



· 5 min read · Jul 26, 2023



7.4K



75





Scott-Ryan Abt in Pitfall

Bye Bye, Spotify

And see ya later, all you subscription services in my little empire

🌟 · 4 min read · Aug 19, 2023

👏 18K

💬 425



See more recommendations