

Open in app ↗



Search

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

Makine Öğrenmesi — Çoklu Doğrusal Regresyon

Samed Harman · [Follow](#)

4 min read · Sep 5, 2020



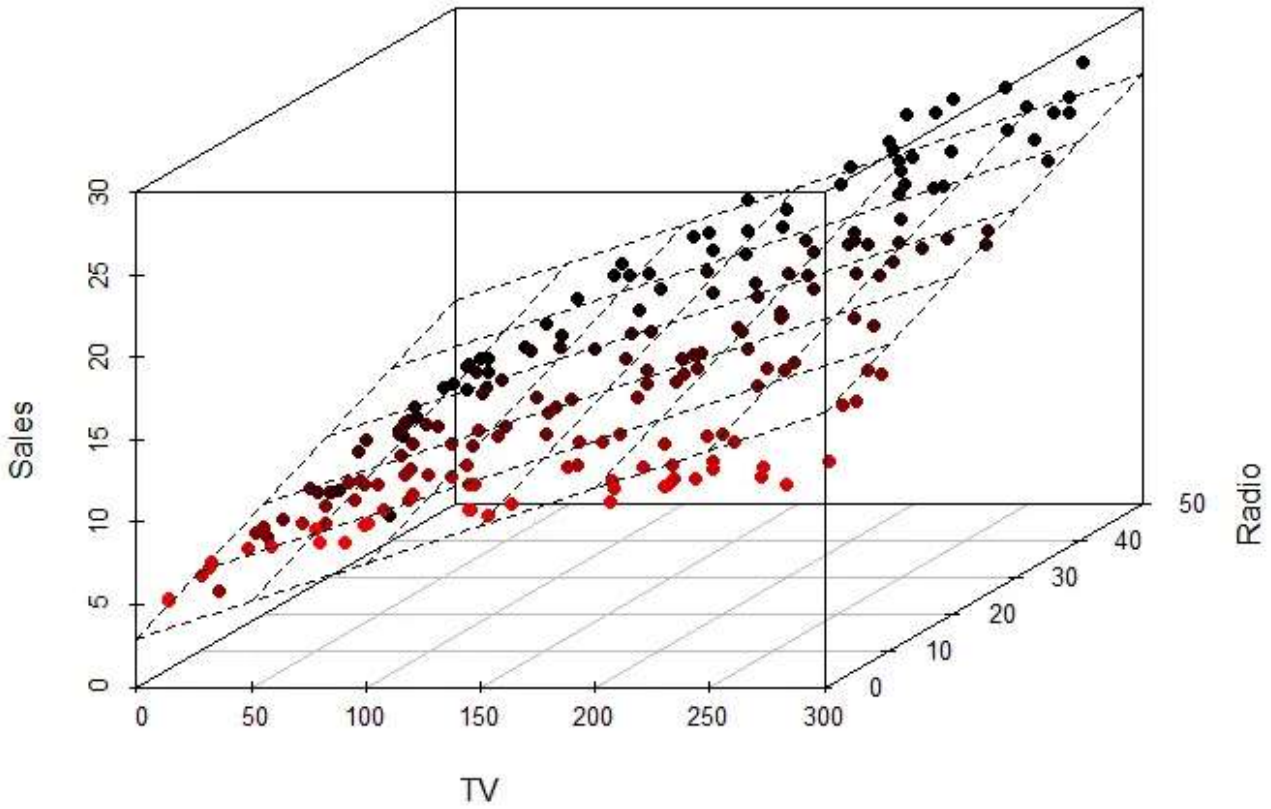
Share



More



Çoklu Doğrusal Regresyon — Multiple Linear Regression



Basit doğrusal regresyon modeli hatırlarsak iki boyutlu uzayda yer alan veri modelleri için uygun olan $y = ax + b$ doğrusunu çizmeye çalışan bir prediction (tahmin) algoritması idi. Bu doğrunun çizilmesindeki amaç ise tahmin edilmesi istenen değere (bağımlı değişken) en yakın sonuçları üretebilmektir.

Çoklu doğrusal regresyon modeli ise sistemde birden fazla bağımsız değişkenin sonucu etkilediği durumlarda imdadımıza yetişmektedir. Aslında kullanımı ve mantığı basit doğrusal regresyon ile benzerdir. Sadece birden fazla değişken için içinde olduğundan dolayı doğru denklemimiz şu şekilde temsil edilir:

$$y = b_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n$$

Buradaki denklem her bir bağımsız değişkene ait katsayı değeri ve d hata miktarı içermektedir. Örneğin doğrusal regresyon modeli için:

$$\text{Satış} = a + b_1(\text{TV}) + b_2(\text{Radio}) + e$$

denklemini a sabit sayısı ve Ay bağımsız değişkeni ile beraber e hata miktarını içerirken;

$$\text{Satış} = a + b(\text{Ay}) + c(\text{Miktar}) + d(\text{Reklam}) + e$$

denklemini a sabit sayısı ve Ay, Miktar, Reklam bağımsız değişkenleri ile e hata miktarını içermektedir.

Değişkenlerin Sisteme Etkilerinin Ölçümü

Denkleminde görüldüğü üzere çoklu doğrusal regresyonda sistemde birden fazla bağımsız değişken bulunmaktadır. Bu sebeple değişkenlerin sisteme olan etkileri birbirinden farklı olabilir ve bu etkiler olumlu veya olumsuz sonuçlanabilir.

Değişkenlerin sisteme olan etkilerini ölçebilmek için birkaç yöntem mevcuttur:

Tüm değişkenlerin dahil edilmesi: Sezgisel olarak kullanabilecek bir yöntem olup kriterlerin bizim tarafımızdan belirlendiği yöntemdir. Aslında diğer yöntemleri kullanmadan önce bir keşif yapmak maksadıyla kullanılır.

Geriye doğru eleme (Backward Elimination): Adım adım karşılaştırma yöntemlerinden birisidir. Öncelikle bir önem değeri (significance value) belirlenir. O anki en yüksek p-value değerine (probability value) sahip değişken belirlenir ve şayet $P > SL$ ise değişken sistemden kaldırılır ve model yeniden oluşturulur, ardından bu adım tekrarlanır. Tüm değişkenler için $P < SL$ olduğunda ise eleme sonlandırılır.

İleriye doğru eleme (Forward Elimination): Geriye doğru eleme yöntemi ile aynı adımları içerir. Fark olarak ise burada algoritma başlangıçta geriye doğru elemenin aksine tek bir değişken içerir ve her adımda değişken sayısı artarak ilerler.

Çift Yönlü eleme (Bidirection Elimination): Bu yöntemde ise bir değişkenin sisteme eklenme ve sistemden elenme kriteri ayrı ayrı bulunur, yani iki adet SL değeri bulunabilir. Sistem öncelikle en düşük p-value değerine sahip değişkeni ele alır. Ardından bu değişkeni sabit tutarak diğer tüm değişkenler sisteme dahil edilir ve tekrar en düşük p-value değerine sahip değişken tespit edilerek o değişkenin sistemde kalması sağlanır.

Skor Karşılaştırması (Score Comparison): Yine sezgisel olarak yaklaşılabilecek yöntemlerden birisidir. Bir başarı kriteri belirlenir ve olası tüm regresyon modelleri

inşa edilir. Ardından belirlenen kritere en uygun model yöntem olarak seçilir.

P-Value dedin durdun o neymiş?

P-Value için şöyle bir tanım yazılmış:

Belirli bir istatistiksel model altında, verilerin istatistiksel toplamının gözlemlenen değere eşit veya daha yüksek olma olasılığıdır.

Peki nedir bunun meali?. Örneğin kavunların olgunlaşmasına etki eden faktörler ile olgun olan kavunları tahmin etmeye çalıştığımız bir modelimiz olsun. Bu modeldeki hipotezimiz ise bitki köküne doğru çapası yapılmayan kavunların çürüyeceği yönünde olsun. Çıktı sonuçlarımıza bakarak elde ettiğimiz bir örnek ise bu hipotezimizi çürütür nitelikte olsun yani örnekte kavunun dibi yanlış çapalandığı halde kavun olgunlaşmış olsun. Biz sadece bu örneğe bakarak ifade ettiğimiz bu hipotez (**H0** hipotezi olarak geçer) doğrudur veya yanlıştır diyemeyiz. Burada belirtilen p-value değeri ise bize aslında kaç örnek sonucunda bu hipotezi çürütebilirizi söylemektedir. Genellikle bu p-value değeri 0.05 gibi bir yüzdelik olarak alınır ve p-value küçüldükçe **H0** hatalı olma ihtimali artar. Artıkça da bu hipotezin tersi olan hipotezin (**H1**) hatalı olma ihtimali artar. Eleme yöntemleri kısmında da bahsettiğim üzere biz bu değeri değişkenlerimizin sisteme etkisi kısmında bir kriter olarak kullanmaktayız.

Dummy Variable Trap (Kukla Değişken Tuzağı) Nedir?

Bu konu aslında makine öğrenmesinin genelinde karşımıza çıkabilecek ve veri ön işleme aşamasını ilgilendiren bir konu olsa da bu kısımda açıklamak istedim. Dummy variable diye tabir ettiğimiz kavram aslında bir değişkeni temsil eden başka bir değişken olarak düşünülebilir. Dummy variable olan değişkenin de sistemde olması bazı makine öğrenme algoritmalarının hatalı çalışabilmesine ve yanlış sonuçlar üretebilmesine sebep olabilir çünkü sistemde bulunan her bir değişkenin sonuca bir etkisi vardır. Kavramı bir örnekle açıklamaya çalışalım:

MEVSİM isimli İlkbahar,Yaz,Sonbahar,Kış değerlerinden oluşan bir kolon için encoding işlemi yapıldığını farz edelim. Oluşan kolonlar ise şu şekilde olsun:

$M1=1, M2=0, M3=0, M4=0$

$M1 = 0, M2=1, M3=0, M4=0$

$M1=0, M2=0, M3=1, M4=0$

$M1 =0, M2=0, M3=0, M4=1$

Burada yer alan Kış değerini temsil eden M4 kolonu aslında bir kukla değişken tuzağı içerir. Çünkü biz M1, M2 ve M3 kolonlarını 0 ile kodladığımızda zaten Kış değerini elde etmiş oluyoruz. Ayrıca bir M4 kolonuna ihtiyacımız olmuyor. Kısaca kukla değişken tuzağından korunmak için birbirinin yerine geçebilen diğer bir tabirle temsil edebilen değişkenlerden sadece bir tanesi sisteme alınmalıdır.

Bu yazımda çoklu doğrusal regresyon modelinden , değişkenler için eleme yöntemlerinden , p olasılık değerinden ve kukla değişken tuzağı kavramından bahsetmeye çalıştım. Hatalı gördüğünüz kısımları veya önerilerinizi bana bildirmekten çekinmeyin. Teşekkürler :).

Samed Harman - Necmettin Erbakan University - Ankara, Turkey | LinkedIn

View Samed Harman's profile on LinkedIn, the world's largest professional community. Samed has 1 job listed on their...

www.linkedin.com

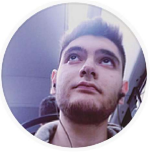
Machine Learning

Linear Regression

Makine Öğrenmesi

Artificial Intelligence

Yapay Zeka



Follow


Written by Samed Harman

65 Followers

Flutter Mobile App Developer at DeepSport | Computer Engineer

More from Samed Harman



 Samed Harman

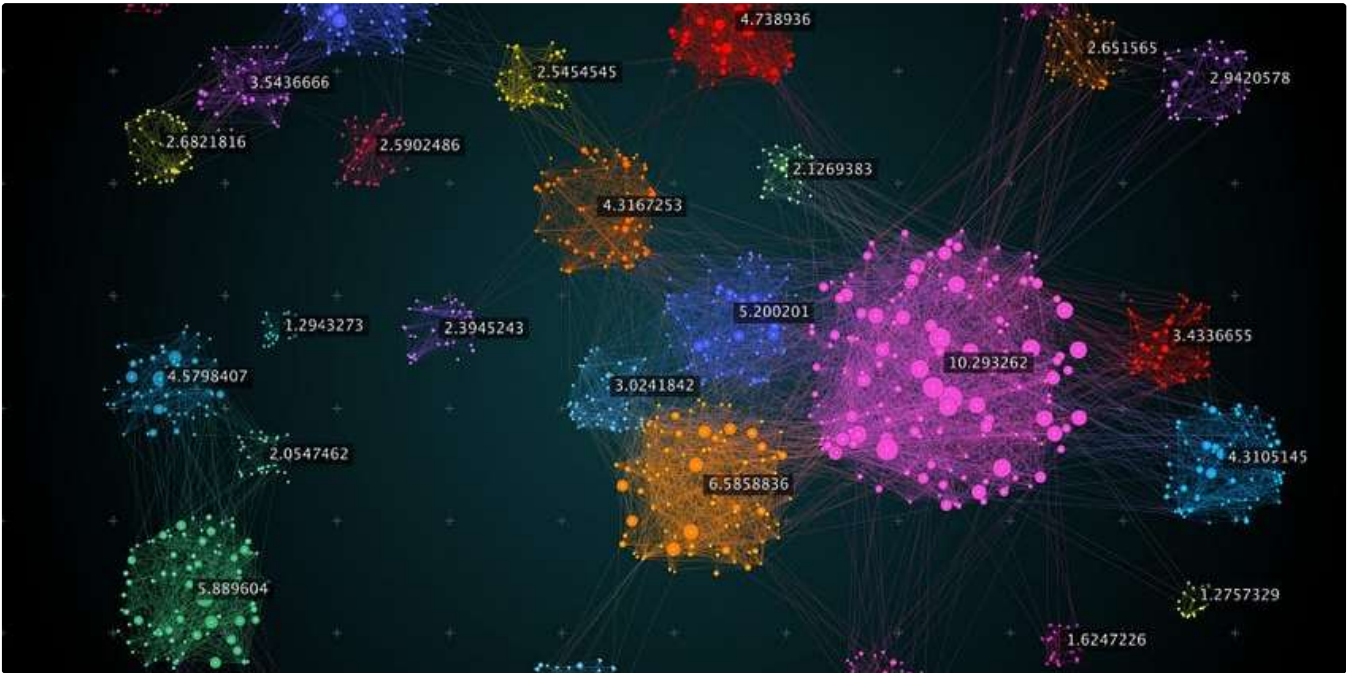
Flutter | Widget Communication With Callbacks (Flutter 3 , Dart 2.17)

Table of contents:

3 min read · Jun 17, 2022



5



 Samed Harman

Makine Öğrenmesi | Clustering (Kümeleme) Teknikleri

Kümeleme (Clustering) , makine öğrenmesi konseptlerinden biri olan Denetimsiz Öğrenme (Unsupervised Learning) için önemli bir kavramdır...

5 min read · Sep 25, 2020

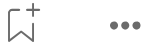


Samed Harman

Flutter | Navigator İşlemleri

Uygulamalarımızda genellikle birden fazla sayfaya ve bu sayfalar arasında geçişlerde bilgi aktarmaya ihtiyaç duyarız. Örneğin bir öğrenci...

5 min read · Sep 8, 2020





Samed Harman

Flutter | Executing Dart Code in Background

Hi all 📍, in this article I'm going to show you how do you execute your dart code if app state is background. For this, we are going to...

3 min read · Jan 17

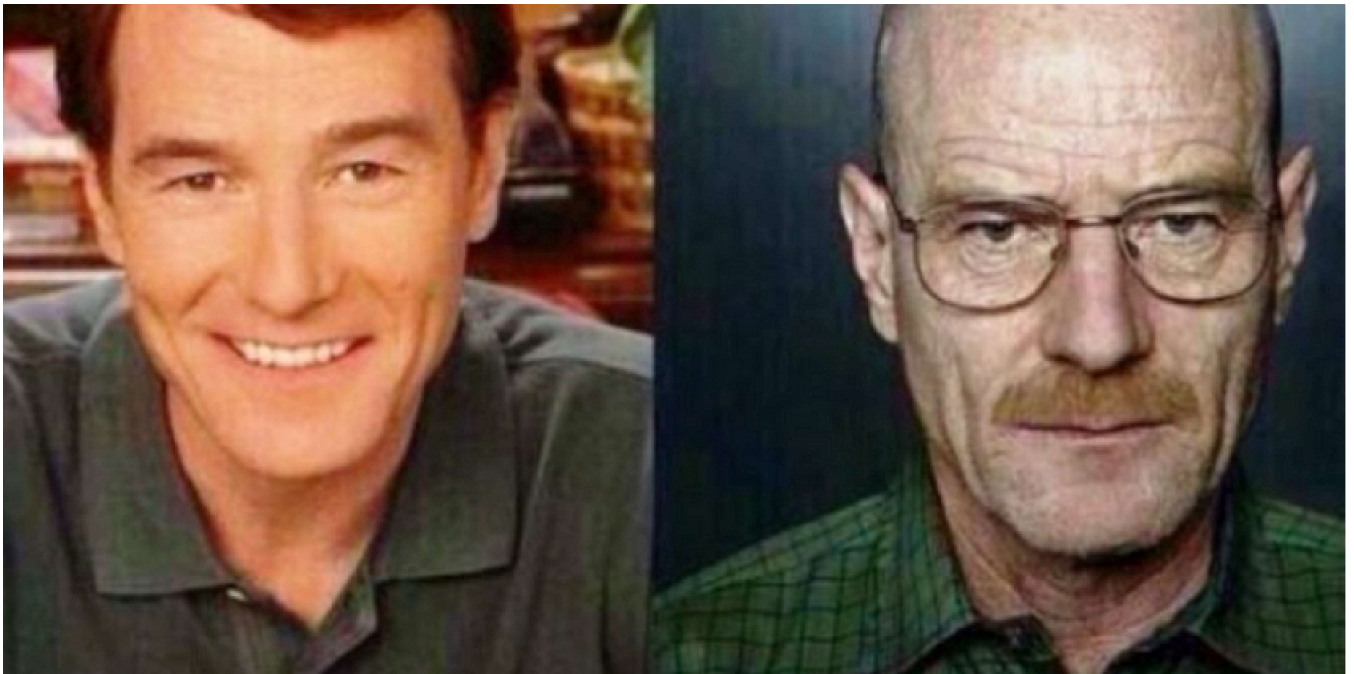



16



See all from Samed Harman

Recommended from Medium



 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

 6.9K  72



 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



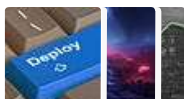
13.4K



234



Lists



Predictive Modeling w/ Python

20 stories · 719 saves



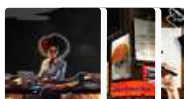
Practical Guides to Machine Learning

10 stories · 825 saves



AI Regulation

6 stories · 233 saves



Natural Language Processing

1027 stories · 502 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...



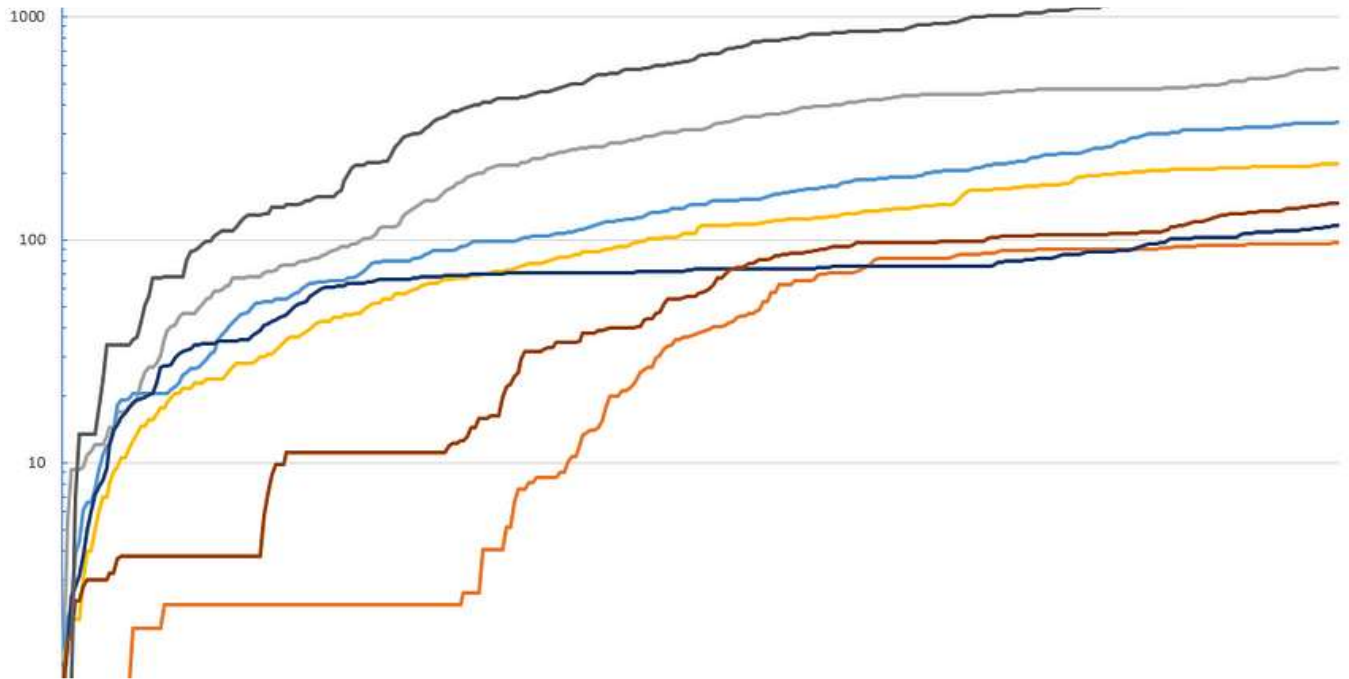
4 min read · Feb 16, 2022



71K



1024



Pau Blasco i Roca in Towards Data Science

My Life Stats: I Tracked My Habits for a Year, and This Is What I Learned

I measured the time I spent on my daily activities (studying, doing sports, socializing, sleeping...) for 332 days in a row.

12 min read · Nov 21

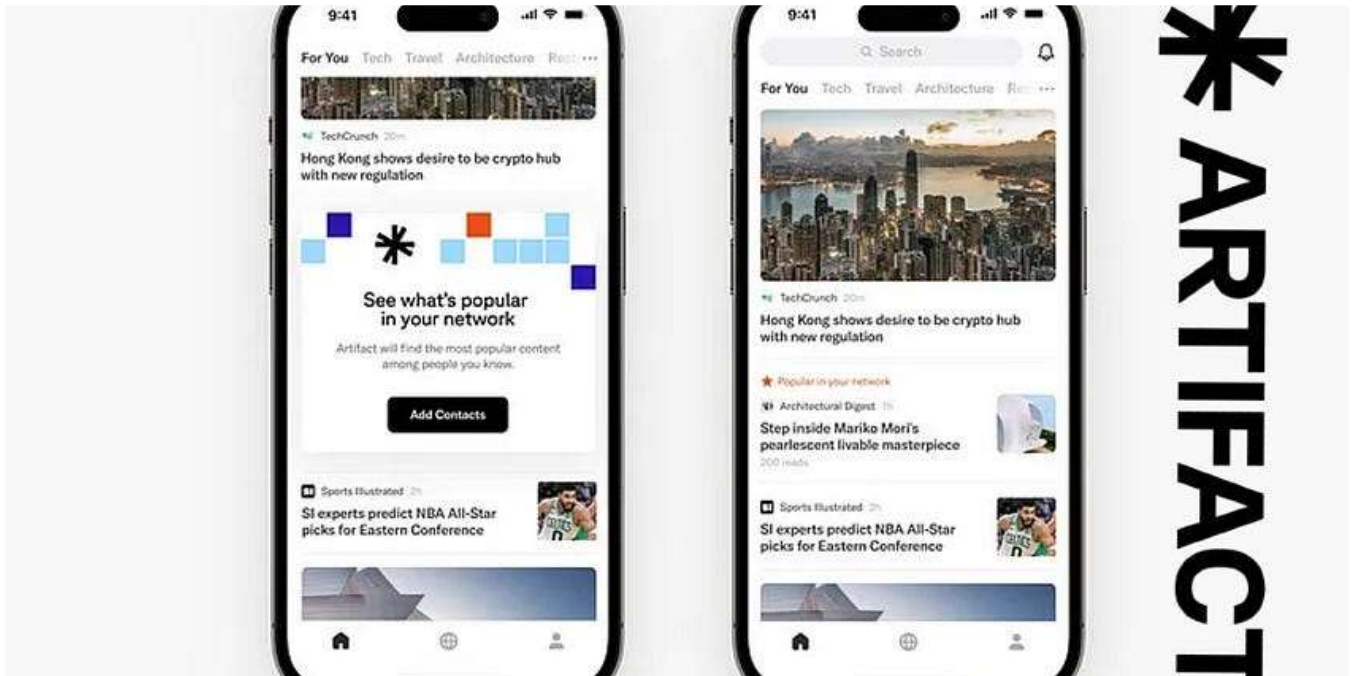


5K



88





 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



9.4K



169



 AL Anany 

The ChatGPT Hype Is Over—Now Watch How Google Will Kill ChatGPT.

It never happens instantly. The business game is longer than you know.

✦ • 6 min read • Sep 1



20K



635



See more recommendations