# **Sentetik Veri: Tanımı, Türleri, Kalite Ölçümü ve Finans/Bankacılık Uygulamaları**

**Giriş**

Sentetik veri, gerçek dünya verilerinin özelliklerini ve kalıplarını taklit etmek amacıyla yapay olarak oluşturulmuş bilgilerdir 1. İnsan tarafından üretilmeyen bu veriler, bilgisayar simülasyonları veya algoritmalar aracılığıyla elde edilir ve gerçek dünya verilerine uygun maliyetli bir alternatif sunar 2. Sentetik verinin önemi günümüzde giderek artmaktadır çünkü gizliliği koruma, veri güvenliğini artırma ve hassas veya erişimi kısıtlı verilere ulaşma zorluklarını aşma gibi çeşitli avantajlar sunar 1. Özellikle makine öğrenimi, veri analizi ve yazılım testi gibi alanlarda yaygın olarak kullanılan sentetik veriler, kuruluşların büyük miktarda veriyi daha akıllıca kullanmasını sağlayarak rekabet avantajı elde etmelerine yardımcı olur 1. Ayrıca, veri gizliliği düzenlemelerine uyumu kolaylaştırır ve farklı endüstrilerde işbirliklerini teşvik eder 1.

**Sentetik Veri Türleri**

Sentetik veriler, oluşturulma yöntemlerine ve amaçlarına göre farklı türlere ayrılabilir. Genel olarak kabul gören sınıflandırmalar arasında tamamen sentetik veriler, kısmen sentetik veriler ve hibrit sentetik veriler yer alır 3.

Tamamen sentetik veriler, herhangi bir orijinal veri içermeyen ve tamamen yapay olarak oluşturulan verilerdir 3. Bu tür veriler genellikle yapay zeka modellerini eğitmek veya test etmek için kullanılır ve gerçek dünya verilerinin istatistiksel kalıplarını taklit eder 1. Örneğin, bilgisayar grafikleri veya simülasyonlar kullanılarak üretilen tamamen sentetik görüntüler buna örnek verilebilir 5.

Kısmen sentetik veriler ise, hassas bilgiler hariç tutularak orijinal verilerdeki tüm bilgileri korur 3. Gerçek verilerden türetilen bu tür veriler, özellikle gizliliği koruma ihtiyacının yüksek olduğu durumlarda tercih edilir 6. Örneğin, müşteri verilerindeki kişisel olarak tanımlanabilir bilgilerin (PII) sahte verilerle değiştirilmesi ancak diğer istatistiksel özelliklerin korunması kısmen sentetik veri oluşturma yöntemlerinden biridir 6.

Hibrit sentetik veriler, hem gerçek dünya verilerini hem de sentetik olarak üretilmiş verileri bir arada içerir 3. Bu yaklaşım, veri setini çeşitlendirmek, modelin genelleme yeteneğini artırmak veya sınırlı gerçek veri durumlarını simüle etmek amacıyla kullanılabilir 5.

Finans ve bankacılık sektöründe, bu sentetik veri türlerinin çeşitli uygulamaları bulunmaktadır. Örneğin, dolandırıcılık tespiti için tamamen sentetik işlem verileri oluşturulabilirken, müşteri davranışlarını analiz etmek için kısmen sentetik müşteri profilleri kullanılabilir 4. Zaman serisi verileri de finans sektöründe önemli bir yer tutar ve işlem verileri, piyasa verileri veya yatırım verileri gibi çeşitli alanlarda sentetik olarak üretilebilir 4.

**Sentetik Veri Kalitesinin Ölçümü**

Sentetik verinin kalitesini ölçmek, oluşturulan verilerin gerçek dünya verilerinin istatistiksel özelliklerini ne kadar doğru bir şekilde koruduğunu ve analiz amaçlarına ne kadar uygun olduğunu değerlendirmeyi içerir 9. Bu değerlendirme, sentetik verilerin güvenilir ve doğru olmasını sağlamak için kritik öneme sahiptir 9.

Sentetik veri kalitesini ölçmek için kullanılan temel boyutlar doğruluk (fidelity), yararlılık (utility) ve gizlilik (privacy) olarak sıralanabilir 10. Doğruluk, sentetik verinin orijinal verilerin istatistiksel özelliklerini ne kadar iyi yansıttığını ölçer 9. Yararlılık, sentetik verinin makine öğrenimi modellerini eğitmek veya veri analizi yapmak gibi gerçek dünya uygulamalarında ne kadar iyi performans gösterdiğini gösterir 9. Gizlilik ise, sentetik verinin hassas bilgileri ne kadar iyi koruduğunu ve ifşa etmediğini değerlendirir 9.

Doğruluğu değerlendirmek için kullanılan metrikler arasında istatistiksel benzerlik, Kolmogorov-Smirnov testi, kategori ve aralık tamlığı, sınır koruması, eksik veri benzerliği ve korelasyon metrikleri bulunur 10. Yararlılığı ölçmek için tahmin skoru, özellik önem skoru ve QScore gibi metrikler kullanılır 10. Gizliliği değerlendirmek için ise tam eşleşme skoru, satır yeniliği, doğru atıf olasılık katsayısı, çıkarım, tekilleştirme ve bağlantı kurulabilirlik gibi metrikler kullanılır 10.

Finans ve bankacılık sektöründe sentetik veri kalitesini ölçmek için de benzer metrikler ve yöntemler kullanılır. Özellikle, sentetik verinin dolandırıcılık tespiti veya risk yönetimi gibi spesifik finansal görevlerde ne kadar etkili olduğunu değerlendirmek için fayda testleri büyük önem taşır 9. Ayrıca, finansal verilerin hassasiyeti nedeniyle gizlilik koruma metrikleri de bu sektörde ayrı bir önem arz eder 9. Kuruluşlar, sentetik verilerini kullanmadan önce manuel kontrollerden geçirebilir veya otomatik araçlar kullanarak gerçek ve sentetik veri kümeleri arasındaki tutarsızlıkları kontrol edebilirler 6.

**Tabular Sentetik Veri Üretimi Yöntemleri ve Python Kütüphaneleri**

Tabular sentetik veri üretimi, satırlar ve sütunlar halinde düzenlenmiş yapılandırılmış verilerin yapay olarak oluşturulması sürecidir 1. Bu tür veriler genellikle veri tabanlarında ve elektronik tablolarda bulunur ve finans ile bankacılık gibi birçok sektör için kritik öneme sahiptir 12. Son beş yılda, tabular sentetik veri üretimi için çeşitli yöntemler ve Python kütüphaneleri geliştirilmiştir.

Derin öğrenme modelleri, tabular sentetik veri üretiminde önemli bir rol oynamaktadır. Üretken Çekişmeli Ağlar (GAN'lar), Varyasyonel Oto Kodlayıcılar (VAE'ler) ve özellikle tabular veriler için geliştirilmiş olan CTGAN (Conditional Tabular GAN) gibi modeller, gerçek veri setlerinin istatistiksel özelliklerini yakalayarak benzer yapıda yeni veri örnekleri üretebilir 5. GAN'lar, bir üretici ve bir ayırt edici ağdan oluşur; üretici gerçek verilere benzer veriler üretmeye çalışırken, ayırt edici üretilen verinin gerçek mi yoksa sentetik mi olduğunu belirlemeye çalışır 5. VAE'ler ise veriyi sıkıştırarak gizli bir uzaya dönüştürür ve bu gizli temsilden yeni veri örnekleri oluşturur 5. CTGAN, özellikle kategorik ve sürekli değişkenlerin bir arada bulunduğu tabular veri setlerinde karmaşık ilişkileri öğrenmek ve bu doğrultuda sentetik veri üretmek için tasarlanmıştır 12.

Python ekosisteminde tabular sentetik veri üretimi için çeşitli kütüphaneler bulunmaktadır. Bunlardan bazıları şunlardır:

* **SDV (Synthetic Data Vault):** Tabular, çok tablolu ve zaman serisi sentetik veri oluşturmak için çeşitli makine öğrenimi algoritmaları sunan kapsamlı bir Python kütüphanesidir 15. SDV, GaussianCopula'dan CTGAN'a kadar farklı modelleri destekler ve sentetik verinin kalitesini değerlendirme araçları da sunar 3.
* **YData Synthetic (şimdi YData SDK):** GAN'lar (CGAN, WGAN, CTGAN vb.) ve VAE'ler gibi çeşitli üretken modelleri içeren bir başka popüler Python kütüphanesidir 19. YData SDK, veri profesyonelleri için bağlantılar, metaveri yönetimi, veri kalitesi profilleme ve sentetik veri üretimi gibi özellikler sunar ve en iyi üretken modeli otomatik olarak seçip optimize eden bir API'ye sahiptir 19.
* **Gretel.AI:** Bulut tabanlı bir platform olmasının yanı sıra Python API'leri de sunan Gretel.AI, yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verilerden gerçekçi sentetik veriler üretebilir 16. Gretel, özellikle gizliliği koruma ve veri anonimleştirme konularına odaklanır ve büyük dil modellerini (LLM'ler) de destekler 20.
* **synthpop:** Özellikle gerçek dünya veri setlerinin sentetik versiyonlarını oluşturmak için tasarlanmış bir R paketidir 16. İstatistiksel modelleri kullanarak orijinal verilerdeki ilişkileri ve dağılımları taklit eder.
* **Faker:** Daha çok sahte metin verisi üretimiyle tanınsa da, tabular veri oluşturmak için de kullanılabilir 21. Özellikle test ve geliştirme amaçları için çeşitli veri türlerinde (isimler, adresler, finansal bilgiler vb.) sahte veriler üretebilir.

Finans sektöründe, bu kütüphaneler müşteri profillerini temel alarak sentetik veriler oluşturmak, dolandırıcılık senaryolarını simüle etmek veya risk yönetimi modellerini test etmek gibi çeşitli amaçlarla kullanılabilir 12. Örneğin, bankacılık sektöründe müşteri profillerine dayalı sentetik veri üretmek isteyen bir şirket, GAN'ları kullanarak gerçek müşteri bilgilerini temel alarak yeni ancak benzer özelliklere sahip müşteri profilleri oluşturabilir 12.

**Metin Tabanlı Sentetik Veri Üretimi Yöntemleri ve Python Kütüphaneleri**

Metin tabanlı sentetik veri üretimi, doğal dil işleme (NLP) ve metinle ilgili diğer görevlerde kullanılmak üzere yapay olarak metin verileri oluşturma sürecidir 5. Bu tür veriler genellikle gerçek verinin mevcut olmadığı veya hassas bilgiler içerdiği durumlarda kullanılır. Son üç yılda, metin tabanlı sentetik veri üretimi yöntemleri ve Python kütüphanelerinde önemli gelişmeler yaşanmıştır.

Büyük Dil Modelleri (LLM'ler), metin tabanlı sentetik veri üretiminde en yaygın kullanılan yöntemlerden biridir 14. GPT, BERT, Llama gibi modeller, geniş metin veri kümeleri üzerinde eğitilerek tutarlı ve bağlamsal olarak alakalı metinler üretebilirler 28. Bu modeller, sentetik müşteri yorumları, finansal haberler, sohbet kayıtları veya diğer metin tabanlı verileri oluşturmak için idealdir 14.

Makine öğrenimi tabanlı diğer yöntemler arasında Üretken Çekişmeli Ağlar (GAN'lar) ve Varyasyonel Oto Kodlayıcılar (VAE'ler) da yer alır 5. Bu modeller, karmaşık veri yapılarını ve dağılımlarını modelleyerek yüksek kaliteli sentetik metinler üretebilirler. Kural tabanlı yöntemler ise, belirli iş mantığını ve kalıplarını izleyerek metin verileri oluşturur 1.

Python ekosisteminde metin tabanlı sentetik veri üretimi için çeşitli kütüphaneler bulunmaktadır:

* **Faker:** Çeşitli dillerde ve formatlarda gerçekçi sahte metin verileri (isimler, adresler, e-postalar, telefon numaraları, finansal bilgiler, metin cümleleri vb.) üretmek için popüler bir Python kütüphanesidir 21. Finans sektöründe sahte kredi kartı numaraları, banka hesap numaraları veya para birimi isimleri gibi veriler oluşturmak için kullanılabilir 22.
* **Gretel.AI:** Yapılandırılmış ve yapılandırılmamış verilerden sentetik veri üretebilen Gretel.AI, metin verileri için de yetenekler sunar. Özellikle büyük dil modelleriyle entegrasyonu sayesinde yüksek kaliteli sentetik metinler oluşturulabilir 16.
* **SDV (Synthetic Data Vault):** Tabular verilerin yanı sıra zaman serisi ve metin verileri için de sentetik veri üretme yetenekleri sunar 16.
* **Mimesis:** Faker'a benzer şekilde çeşitli dillerde ve formatlarda sahte metin verileri üretebilen bir başka Python kütüphanesidir 21.

Finans ve bankacılık sektöründe metin tabanlı sentetik veriler, müşteri hizmetleri kayıtlarını simüle etmek, finansal haberler oluşturmak veya dolandırıcılıkla ilgili metinleri analiz etmek gibi çeşitli uygulamalarda kullanılabilir 5. Örneğin, bankalar müşteri destek sistemlerini eğitmek için sentetik sohbet kayıtları oluşturabilir veya finansal analistler piyasa trendlerini incelemek için sentetik finansal haberler üretebilirler 16.

**Sentetik Verinin Finans/Bankacılık Uygulamaları (Son 3 Yıl)**

Son üç yılda, sentetik verinin finans ve bankacılık sektöründeki uygulamaları önemli ölçüde artmıştır 14. Bu artışta, yapay zeka ve makine öğrenimi teknolojilerindeki gelişmelerin yanı sıra veri gizliliği ve güvenliği konusundaki artan farkındalık da etkili olmuştur.

Sentetik verinin finans ve bankacılıkta öne çıkan bazı uygulama alanları şunlardır:

* **Dolandırıcılık Tespiti ve Önleme:** Sentetik veri, çeşitli sahte işlemleri simüle ederek bankaların makine öğrenimi modellerini eğitmesine olanak tanır 27. Nadir görülen dolandırıcılık senaryolarını sentetik olarak oluşturmak, modellerin bu tür olayları daha iyi tespit etmesine yardımcı olur 14. American Express ve J.P. Morgan gibi finans kuruluşları, dolandırıcılık tespit yeteneklerini artırmak için tabular sentetik veriyi kullanmaktadır 7. *Medium* platformunda yer alan makaleler de sentetik verinin kara para aklamayı önleme faaliyetlerinde nasıl kullanılabileceğini göstermektedir 33. Sentetik verinin çok çeşitli dolandırıcılık senaryoları, hatta uç durumları oluşturabilme yeteneği, dolandırıcılık tespit sistemlerinin tehditlere karşı dayanıklılığını ve uyum yeteneğini artıran güçlü bir araçtır 27.
* **Risk Yönetimi ve Stres Testi:** Sentetik veri, aşırı piyasa koşullarını ve çeşitli risk senaryolarını simüle ederek finansal modellerin stres testinde önemli bir rol oynar 27. Bankalar, nadir görülen "Kara Kuğu" olaylarını sentetik veri kullanarak simüle edebilir ve modellerini iklim değişikliği gibi yeni dışsal gelişmeler üzerinde eğitebilir 14. *Medium* platformunda yer alan makaleler, yapay zeka tarafından üretilen sentetik verinin finans sektöründe stres testlerini nasıl dönüştürdüğünü ele almaktadır 34. Aşırı ve benzeri görülmemiş piyasa koşullarının simülasyonunu mümkün kılarak, sentetik veri finansal kuruluşların potansiyel riskleri proaktif bir şekilde değerlendirmesine ve yönetmesine olanak tanır, bu da finansal sistemin genel dayanıklılığını artırır 27.
* **Müşteri Davranışı Analizi:** Sentetik veri, müşteri davranışlarını ve işlem modellerini taklit ederek müşteri tercihleri ve eğilimlerine dair içgörüler sağlar 27. Bu içgörüler, bankaların kişiselleştirilmiş finansal hizmetler ve ürünler geliştirmesine yardımcı olabilir.
* **Algoritmik Ticaret:** Sentetik piyasa verileri, ticaret algoritmalarını ve stratejilerini geriye dönük test etmek için kullanılır 7. Bu, gerçek piyasa koşullarında risk almadan stratejilerin performansını değerlendirmeyi sağlar.
* **Kredi Değerlendirme:** Sentetik veri, kredi puanlama modellerini geliştirmede ve test etmede kullanılır, bu modellerin adil ve tarafsız olmasını sağlar 27. Sınırlı kredi geçmişi olan küçük işletmeler için kredi risk modellerini eğitmek amacıyla sentetik finansal veriler kullanılabilir 35.
* **Mevzuata Uygunluk:** Sentetik veri, çeşitli uyumluluk senaryolarını simüle ederek finans kurumlarının tüm düzenleyici gereklilikleri karşılamasını sağlar 27.
* **Ürün Geliştirme:** Sentetik veri, yeni finansal ürünlerin piyasaya sürülmeden önce test edilmesine ve iyileştirilmesine olanak tanır 4.
* **Veri Paylaşımı ve İşbirliği:** Sentetik veri, veri gizliliğinden ödün vermeden üçüncü taraf satıcılar, ortaklar ve araştırmacılarla güvenli veri paylaşımını sağlayarak işbirliğini kolaylaştırır 3. Bu, finansal kuruluşların inovasyonu teşvik etmek ve yeni çözümler geliştirmek için diğer kuruluşlarla daha kolay işbirliği yapmasına olanak tanır.
* **Eğitim:** Gerçek müşteri verilerini ifşa etmeden yeni çalışanları eğitmek için sentetik veriler kullanılabilir 27.

Son 3 yıldaki önemli gelişmeler arasında Gartner'ın 2026 yılına kadar işletmelerin %75'inin sentetik müşteri verisi oluşturmak için Üretken Yapay Zeka kullanacağını öngörmesi yer almaktadır 14. Dijital Dubai'nin Ekim 2022'de sentetik veri kullanımı için bir çerçeve başlatması da bu alandaki artan ilgiyi göstermektedir 14. *Medium* platformunda yayınlanan makaleler, yapısal veri türleri için yapay zeka tabanlı sentetik veri üretimini ve bankacılık sektöründeki örnek uygulamalarını (2024) 12, yapay zekanın finans sektöründe verimliliği ve büyümeyi nasıl artırdığını anlatırken sentetik verinin risk yönetimi alanındaki potansiyeline değinmeyi (2024) 36 ve yapay zekanın finans alanındaki uygulamalarını genel olarak ele alırken, sentetik verinin dolandırıcılık tespiti ve risk yönetimi gibi alanlardaki önemine işaret etmeyi (2023) 37 içermektedir. Üretken Yapay Zeka'nın giderek artan şekilde benimsenmesi, finans ve bankacılıkta sentetik veri oluşturma ve kullanma süreçlerinde önemli bir hızlanmaya işaret etmekte, bu da daha karmaşık uygulamalara ve potansiyel olarak yeni iş modellerine yol açmaktadır 14.

**Sonuç**

Sentetik veri, finans ve bankacılık sektöründe büyük bir potansiyele sahiptir ve önümüzdeki yıllarda kullanımının önemli ölçüde artması beklenmektedir 4. Finans kuruluşlarının %77'sinin büyük veriden yararlanmadan rekabeti kaybetmekten korkması, sentetik verinin bu alandaki stratejik önemini vurgulamaktadır 4. Ancak, sentetik veri kullanımının getirebileceği bazı zorluklar ve etik hususlar da göz önünde bulundurulmalıdır. Orijinal verilerin karmaşıklığını tam olarak yansıtmak her zaman mümkün olmayabilir ve bu durum tutarsızlıklara yol açabilir 3. Ayrıca, sentetik veriler gerçek verilerin yerini tamamen alamaz ve sentetik veri oluşturma süreçlerinde önyargıların oluşma riski de bulunmaktadır 3. Ne kadar uzman olunursa olsun, gerçek dünyadaki verilerin mükemmel bir taklidini oluşturmak teknik olarak zorlayıcı olabilir 6. Bu nedenle, sentetik veri kullanırken veri kalitesini sürekli olarak değerlendirmek, potansiyel önyargıları dikkate almak ve etik kullanımını sağlayacak uygun yönergeleri ve düzenlemeleri geliştirmek büyük önem taşımaktadır.

**Kaynaklar**

* AWS. "Sentetik Veri Nedir? - AWS." Erişim tarihi: 26 Mart 20256.
* Syntho.ai. "Sentetik Veri Nedir?" Erişim tarihi: 26 Mart 20251.
* Syntho.ai. "Sentetik Veri Türleri Nelerdir?" Erişim tarihi: 26 Mart 20258.
* Syntho.ai. "Sentetik Veri Finans Alanında." Erişim tarihi: 26 Mart 20254.
* Medium. "Bilgi Çağında Yeni Bir Zirve: Gerçek ve Sentetik Veri Arasında Yolculuk." Erişim tarihi: 26 Mart 20255.
* Shaip.com. "Sentetik Veri Kullanım Alanları, Riskleri ve Uygulamaları." Erişim tarihi: 26 Mart 202532.
* QuestionPro. "Sentetik Veri Nedir? Türleri, Yöntemleri, Kullanımı." Erişim tarihi: 26 Mart 20253.
* QuestionPro. "Sentetik Test Verileri Nedir? Nasıl Oluşturulur? Kullanım Örnekleri." Erişim tarihi: 26 Mart 202540.
* Syntho.ai. "Sentetik Veri Üretimi Kılavuzu: Tanımı, Türleri ve Uygulamaları." Erişim tarihi: 26 Mart 202538.
* Syntho.ai. "Syntho'nun Kalite Güvence Raporu." Erişim tarihi: 26 Mart 20259.
* GLEIF.org. "Veri Kalitesi Denetimleri." Erişim tarihi: 26 Mart 202541.
* Syntheticus.ai. "Sentetik Veri Kalitesi Nasıl Değerlendirilir." Erişim tarihi: 26 Mart 202510.
* Gretel.ai. "Sentetik Kalite ve Gizlilik Raporunu Değerlendirin." Erişim tarihi: 26 Mart 202542.
* Quantmetry.com. "Üretken Yapay Zeka: Sentetik Tabular Verileri Değerlendirme Rehberi." Erişim tarihi: 26 Mart 202511.
* SDV.dev. "SDMetrics." Erişim tarihi: 26 Mart 202543.
* Gretel.ai. "Değerlendirme." Erişim tarihi: 26 Mart 202544.
* Medium. "@denizkilinc. "Yapısal Veri Türleri İçin Yapay Zeka Tabanlı Sentetik Veri Üretimi." Erişim tarihi: 26 Mart 202512.
* K2view.com. "Sentetik Veri Üretimi Nedir?" Erişim tarihi: 26 Mart 202513.
* K2view.com. "Sentetik Veri Örnekleri." Erişim tarihi: 26 Mart 20257.
* Turing.com. "Sentetik Veri Üretim Teknikleri." Erişim tarihi: 26 Mart 202545.
* Developer.gs.com. "Sentetik Veri Üreticisi." Erişim tarihi: 26 Mart 202546.
* Betterdata.ai. "Finans ve Bankacılıkta Sentetik Veri İçin 10 Kullanım Durumu." Erişim tarihi: 26 Mart 202527.
* Unite.ai. "Veri Bilimi İçin En İyi 10 Python Kitaplığı." Erişim tarihi: 26 Mart 202547.
* Journo.com.tr. "En İyi Python Kütüphaneleri - Veri Analizi." Erişim tarihi: 26 Mart 202548.
* Github.com. "kadirturan0/python-veri-bilimi-kutuphaneleri." Erişim tarihi: 26 Mart 202549.
* Medium. "Kader Miyanyedi. "Veri Bilimi İçin Temel Python Kütüphaneleri: Numpy ve Pandas." Erişim tarihi: 26 Mart 202550.
* Talentgrid.io. "Python Kütüphaneleri." Erişim tarihi: 26 Mart 202551.
* Datacamp.com. "Sentetik Veri Üretimi." Erişim tarihi: 26 Mart 202516.
* Github.com. "sdv-dev/SDV." Erişim tarihi: 26 Mart 202517.
* Github.com. "ydataai/ydata-synthetic." Erişim tarihi: 26 Mart 202519.
* Docs.sdv.dev. "SDV." Erişim tarihi: 26 Mart 202518.
* Medium. "@ashish28082002.ak. "Tabular Sentetik Veri Üreticisi (Koşullu GAN'lar Kullanılarak)." Erişim tarihi: 26 Mart 202515.
* Savasanadam.com. "Sentetik Veri Türleri ve Oluşturma Yöntemleri." Erişim tarihi: 26 Mart 202552.
* Insights.daffodilsw.com. "2025'te Sentetik Veriler Hakkında Bilmeniz Gereken Her Şey." Erişim tarihi: 26 Mart 202529.
* Deepchecks.com. "Sentetik Veri Üretimi BDM'ler Hakkında Bilinmesi Gerekenler." Erişim tarihi: 26 Mart 202528.
* Intelligentcio.com. "Finans Kurumları Sentetik Veri Üretimi ile Tahmini Risk Modellerini Nasıl Oluşturabilir?" Erişim tarihi: 26 Mart 202514.
* Patika.dev. "Popüler Python Kütüphaneleri: Mutlaka Bilmeniz Gerekenler." Erişim tarihi: 26 Mart 202553.
* Coderspace.io. "En Popüler Python Kütüphaneleri." Erişim tarihi: 26 Mart 202554.
* Tirendazakademi.medium.com. "Veri Bilimi İçin 10 Python Kütüphanesi." Erişim tarihi: 26 Mart 202555.
* Aitude.com. "Sentetik Veri Oluşturmak İçin En İyi Python Paketleri." Erişim tarihi: 26 Mart 202521.
* Medium. "@sangitapokhrel911. "Python'da Faker ve DrawData ile Sahte Veri Üretme - Sentetik Veri Üretimi." Erişim tarihi: 26 Mart 202522.
* Medium. "@melihciray. "Python'da Faker ile Sahte Veri Üretme: Test ve Geliştirme İçin Bir Rehber." Erişim tarihi: 26 Mart 202523.
* Kaggle.com. "Adnan Anam. "Python Kullanarak Sınırsız Sahte Veri Oluşturun." Erişim tarihi: 26 Mart 202524.
* Stringfestanalytics.com. "Excel'de Python: Faker ile Sahte Veri Nasıl Oluşturulur." Erişim tarihi: 26 Mart 202525.
* Github.com. "joke2k/faker." Erişim tarihi: 26 Mart 202526.
* Cloud.google.com. "Gretel ve BigQuery DataFrames ile Sentetik Veri Üretimi." Erişim tarihi: 26 Mart 202520.
* Gretel.ai. "Akıllı AI Geliştirme İçin Gretel Navigator Veri Tasarımcısı: Sentetik Veri Oluşturmada Devrim." Erişim tarihi: 26 Mart 202530.
* Docs.gretel.ai. "Örnekler." Erişim tarihi: 26 Mart 202531.
* Huggingface.co. "gretelai/gretel-text-to-python-fintech-en-v1." Erişim tarihi: 26 Mart 202556.
* Kaggle.com. "ealaxi/paysim1." Erişim tarihi: 26 Mart 202557.
* Medium. "@a.skabar\_60534. "Gerçekçi Sentetik Finansal Zaman Serileri Üretme." Erişim tarihi: 26 Mart 202558.
* FCA.org.uk. "Finansal Hizmetlerde Sentetik Veri Kullanımı Raporu." Yayın tarihi: Mart 202459.
* Betterdata.ai. "Sentetik Veri Kullanarak Finans ve Bankacılık Sektöründe Dijital Dönüşüm." Yayın tarihi: 24 Haziran 202435.
* Ydata.ai. "Sentetik Veri Kalitenizi Görsel Olarak Nasıl Değerlendirirsiniz?" Yayın tarihi: 21 Kasım 202360.
* Keymakr.com. "Sentetik Verilerde Kalite ve Gerçekçiliği Sağlama." Yayın tarihi: 16 Mayıs 202461.
* Financialresearch.gov. "Finansal Makine Öğrenimi Dayanıklılığı." Yayın tarihi: Mayıs 202462.
* Medium. "@seaflux. "Finansta Yapay Zeka: Verimliliği ve Büyümeyi Artırma." Yayın tarihi: 16 Şubat 202436.
* Medium. "@aimonks. "Finansta Yapay Zeka Uygulamaları." Yayın tarihi: 16 Şubat 202437.
* Investing.com. "NMR/USD Sentetik Geçmiş Verileri." Erişim tarihi: 26 Mart 202563.
* Dergipark.org.tr. "Yapay Zeka Bankacılık Uygulamaları." Yayın tarihi: 7 Aralık 202464.
* [Ismailsen.com](https://www.google.com/search?q=Ismailsen.com).tr. "Finansta Makine Öğrenme Yöntemleri: Son Uygulamalar ve Gelecek Beklentileri." Erişim tarihi: 26 Mart 202565.
* Papel.com.tr. "Fintek Dünyasında 2024 Yılı Nasıl Geçti? 2025'te Fintek Sektöründe Neler Bekliyor?" Erişim tarihi: 26 Mart 202566.
* Srp-legal.com. "07.03.2025 Tarihinde Türkiye Bankalar Birliği İnternet Sitesinde Aralık 2024 Türkiye Finansal Sektör Ödeme Sistemleri Raporu Yayınlanmıştır." Yayın tarihi: 14 Mart 202567.
* Pwc.com.tr. "2025 ve Sonrasında Perakende Bankacılık." Erişim tarihi: 26 Mart 202568.
* Uptech.team. "Bankacılıkta Yapay Zeka Trendleri." Erişim tarihi: 26 Mart 202569.
* Ibm.com. "2025 Bankacılık ve Finansal Piyasalar Görünümü." Erişim tarihi: 26 Mart 202570.
* Mx.com. "2025 Tahminleri." Erişim tarihi: 26 Mart 202571.
* Tv.redhat.com. "Finansal Hizmetler 2025 İçin En İyi Trendler: Yapay Zeka ve Sentetik Veri." Erişim tarihi: 26 Mart 202572.
* Sbs-software.com. "Varlık Finansmanı 2025 İçin Trendler." Erişim tarihi: 26 Mart 202573.
* Finextra.com. "Riski Yeniden Tanımlamak: Yapay Zeka Tarafından Üretilen Sentetik Veri, Finansta Stres Testlerini Nasıl Dönüştürüyor?" Yayın tarihi: 15 Mayıs 202434.
* Cc-bei.news. "Sentetik Veri: Veriye Dayalı Finansal Hizmetlerin Geleceği." Yayın tarihi: 11 Aralık 202333.

#### Alıntılanan çalışmalar

1. Sentetik veri nedir? | Syntho bunu anlamanıza yardımcı olacak, erişim tarihi Mart 27, 2025, <https://www.syntho.ai/tr/what-is-synthetic-data/>
2. Sentetik Veri Nedir? - OpenZeka Blog, erişim tarihi Mart 27, 2025, <https://blog.openzeka.com/ai/sentetik-veri-nedir/>
3. Sentetik Veri: Nedir, Türleri, Yöntemleri + Kullanımı - QuestionPro, erişim tarihi Mart 27, 2025, <https://www.questionpro.com/blog/tr/sentetik-veri-nedir-turleri-yontemleri-kullanimi/>
4. Finansta Sentetik Veriler | Syntho, erişim tarihi Mart 27, 2025, <https://www.syntho.ai/tr/synthetic-data-in-finance/>
5. Bilgi Çağında Yeni Bir Zirve: Gerçek ve Sentetik Veri Arasında Yolculuk - Medium, erişim tarihi Mart 27, 2025, <https://medium.com/academy-team/bilgi-%C3%A7a%C4%9F%C4%B1nda-yeni-bir-zirve-ger%C3%A7ek-ve-sentetik-veri-aras%C4%B1nda-yolculuk-09acd7bfa78b>
6. Sentetik Veri nedir? - AWS, erişim tarihi Mart 27, 2025, <https://aws.amazon.com/tr/what-is/synthetic-data/>
7. Synthetic Data Examples that'll Knock Your SOX Off - K2view, erişim tarihi Mart 27, 2025, <https://www.k2view.com/blog/synthetic-data-examples/>
8. Zaman Serisi Sentetik Veri - Syntho, erişim tarihi Mart 27, 2025, <https://www.syntho.ai/tr/time-series-synthetic-data/>
9. Sentetik veri kalitesi | Syntho'nun kalite güvence raporu, erişim tarihi Mart 27, 2025, <https://www.syntho.ai/tr/synthos-quality-assurance-report/>
10. How to evaluate synthetic data quality - Syntheticus, erişim tarihi Mart 27, 2025, <https://syntheticus.ai/blog/how-to-evaluate-synthetic-data-quality>
11. Generative AI: a guide to evaluate synthetic tabular data - Quantmetry, erişim tarihi Mart 27, 2025, <https://www.quantmetry.com/blog/generative-ai-a-guide-to-evaluate-synthetic-tabular-data/>
12. Yapısal Veri Türleri İçin Yapay Zeka Tabanlı Sentetik Veri Üretimi | by Prof. Dr. Deniz Kılınç, erişim tarihi Mart 27, 2025, <https://medium.com/@denizkilinc/yap%C4%B1sal-veri-t%C3%BCrleri-i%CC%87%C3%A7in-yapay-zeka-tabanl%C4%B1-sentetik-veri-%C3%BCretimi-1176d32b3c6e>
13. What is Synthetic Data Generation? A Practical Guide - K2view, erişim tarihi Mart 27, 2025, <https://www.k2view.com/what-is-synthetic-data-generation/>
14. How financial institutions can build predictive risk models with Synthetic Data Generation, erişim tarihi Mart 27, 2025, <https://www.intelligentcio.com/me/2024/09/19/how-financial-institutions-can-build-predictive-risk-models-with-synthetic-data-generation/>
15. Synthetic Tabular Data Generator | by Ashish Kumar - Medium, erişim tarihi Mart 27, 2025, <https://medium.com/@ashish28082002.ak/tabular-synthetic-data-generator-using-conditional-gans-d98fcd974148>
16. Synthetic Data Generation: A Hands-On Guide in Python - DataCamp, erişim tarihi Mart 27, 2025, <https://www.datacamp.com/tutorial/synthetic-data-generation>
17. sdv-dev/SDV: Synthetic data generation for tabular data - GitHub, erişim tarihi Mart 27, 2025, <https://github.com/sdv-dev/SDV>
18. Welcome to the SDV! - Synthetic Data Vault, erişim tarihi Mart 27, 2025, <https://docs.sdv.dev/sdv>
19. ydataai/ydata-synthetic: Synthetic data generators for tabular and time-series data - GitHub, erişim tarihi Mart 27, 2025, <https://github.com/ydataai/ydata-synthetic>
20. Synthetic data generation with Gretel and BigQuery DataFrames | Google Cloud Blog, erişim tarihi Mart 27, 2025, <https://cloud.google.com/blog/products/data-analytics/synthetic-data-generation-with-gretel-and-bigquery-dataframes>
21. Top Python Packages to Generate Synthetic Data - AITUDE, erişim tarihi Mart 27, 2025, <https://www.aitude.com/top-python-packages-to-generate-synthetic-data/>
22. Generating Fake Data in Python with Faker and DrawData - Medium, erişim tarihi Mart 27, 2025, <https://medium.com/@sangitapokhrel911/generating-fake-data-in-python-with-faker-and-drawdata-synthetic-data-generation-dc77d73e8521>
23. Generating Fake Data in Python with Faker: A Guide for Testing and Development - Medium, erişim tarihi Mart 27, 2025, <https://medium.com/@melihciray/generating-fake-data-in-python-with-faker-a-guide-for-testing-and-development-8f47e9237c9c>
24. Generate Unlimited Fake Data Using Python - Kaggle, erişim tarihi Mart 27, 2025, <https://www.kaggle.com/code/adnananam/generate-unlimited-fake-data-using-python>
25. Python in Excel: How to generate fake data with Faker - Stringfest Analytics, erişim tarihi Mart 27, 2025, <https://stringfestanalytics.com/python-in-excel-how-to-generate-fake-data-with-faker/>
26. joke2k/faker: Faker is a Python package that generates fake data for you. - GitHub, erişim tarihi Mart 27, 2025, <https://github.com/joke2k/faker>
27. 10 Use Cases for Synthetic Data in Finance and Banking - Betterdata, erişim tarihi Mart 27, 2025, <https://www.betterdata.ai/blogs/10-use-cases-for-synthetic-data-in-finance-and-banking>
28. Synthetic Data Generation with LLMs: What You Need to Know, erişim tarihi Mart 27, 2025, <https://www.deepchecks.com/what-to-know-synthetic-data-generation-llms/>
29. Everything You Should Know About Synthetic Data in 2025 - Daffodil Software, erişim tarihi Mart 27, 2025, <https://insights.daffodilsw.com/blog/everything-you-should-know-about-synthetic-data-in-2025>
30. Build high-quality datasets for AI using Gretel Navigator Data Designer, erişim tarihi Mart 27, 2025, <https://gretel.ai/blog/build-high-quality-datasets-for-ai-using-gretel-navigator>
31. Use Case Examples - Gretel.ai, erişim tarihi Mart 27, 2025, <https://docs.gretel.ai/gretel-basics/getting-started/examples>
32. Sentetik Veri Kılavuzu – Kullanımlar, Yararlar, Riskler ve Uygulamalar | Shaip, erişim tarihi Mart 27, 2025, <https://tr.shaip.com/blog/synthetic-data-uses-risk-applications/>
33. Synthetic Data – The Future of Data-Driven Financial Services? | cc-bei.news, erişim tarihi Mart 27, 2025, <https://cc-bei.news/en/synthetic-data-the-future-of-data-driven-financial-services/>
34. Reinventing Risk: How AI-Generated Synthetic Data is Transforming Stress Testing in Finance: By Shailendra Prajapati - Finextra Research, erişim tarihi Mart 27, 2025, <https://www.finextra.com/blogposting/28072/reinventing-risk-how-ai-generated-synthetic-data-is-transforming-stress-testing-in-finance>
35. Digital Transformation in the Finance and Banking Industry using Synthetic Data - Betterdata, erişim tarihi Mart 27, 2025, <https://www.betterdata.ai/blogs/digital-transformation-in-the-finance-and-banking-industry-using-synthetic-data>
36. Artificial Intelligence in Finance Boosting Efficiency and Growth | by Jay | Medium, erişim tarihi Mart 27, 2025, <https://medium.com/@seaflux/artificial-intelligence-in-finance-boosting-efficiency-and-growth-69fbfe933c45>
37. Applications of AI in Finance - Medium, erişim tarihi Mart 27, 2025, <https://medium.com/aimonks/applications-of-ai-in-finance-b9fe7b323521>
38. Sentetik Veri Üretimi Kılavuzu | Sentez - Syntho, erişim tarihi Mart 27, 2025, <https://www.syntho.ai/tr/guide-to-synthetic-data-generation-definition-types-applications/>
39. aws.amazon.com, erişim tarihi Mart 27, 2025, <https://aws.amazon.com/tr/what-is/synthetic-data/#:~:text=Sentetik%20veriler%2C%20ger%C3%A7ek%20d%C3%BCnya%20verilerini,eden%2C%20insan%20taraf%C4%B1ndan%20olu%C5%9Fturulmam%C4%B1%C5%9F%20verilerdir.>
40. Sentetik Test Verileri: Nedir, Nasıl Oluşturulur + Kullanım Örnekleri - QuestionPro, erişim tarihi Mart 27, 2025, <https://www.questionpro.com/blog/tr/sentetik-test-verileri-nedir-nasil-olusturulur-kullanim-ornekleri/>
41. Veri Kalitesi Denetimleri - GLEIF Veri Kalitesi Yönetimi - LEI Verisi, erişim tarihi Mart 27, 2025, <https://www.gleif.org/tr/lei-data/gleif-data-quality-management/data-quality-checks>
42. Synthetic Quality & Privacy Report - Gretel!, erişim tarihi Mart 27, 2025, <https://docs.gretel.ai/optimize-synthetic-data/evaluate/synthetic-quality-privacy-report>
43. SDMetrics - Synthetic Data Vault, erişim tarihi Mart 27, 2025, <https://docs.sdv.dev/sdmetrics>
44. Synthetic data evaluation for the metrics you care about - Gretel.ai, erişim tarihi Mart 27, 2025, <https://gretel.ai/evaluate>
45. Synthetic Data Generation: Definition, Types, Techniques, & Tools - Turing, erişim tarihi Mart 27, 2025, <https://www.turing.com/kb/synthetic-data-generation-techniques>
46. Synthetic Data Generator for Financial Contracts - Goldman Sachs Developer, erişim tarihi Mart 27, 2025, <https://developer.gs.com/blog/posts/synthetic-data-generator>
47. Veri Bilimi için En İyi 10 Python Kitaplığı (2025) - Unite.AI, erişim tarihi Mart 27, 2025, <https://www.unite.ai/tr/Veri-bilimi-i%C3%A7in-en-iyi-10-python-kitapl%C4%B1%C4%9F%C4%B1/>
48. Veri analizi ve görselleştirme için en iyi 10 Python kütüphanesi - Journo, erişim tarihi Mart 27, 2025, <https://journo.com.tr/en-iyi-python-kutuphaneleri-veri-analiz>
49. kadirturan0/python-veri-bilimi-kutuphaneleri: Veri Bilimi İçin En İyi 20 Python Kütüphanesi, erişim tarihi Mart 27, 2025, <https://github.com/kadirturan0/python-veri-bilimi-kutuphaneleri>
50. Veri Bilimi İçin Temel Python Kütüphaneleri: Numpy ve Pandas - Kader Miyanyedi - Medium, erişim tarihi Mart 27, 2025, <https://kadermiyanyedi.medium.com/veri-bilimi-i%CC%87%C3%A7in-temel-python-k%C3%BCt%C3%BCphaneleri-numpy-ve-pandas-97050f690247>
51. Bu 10 Python Kütüphanesi ile İleri Seviye Programlama - TalentGrid, erişim tarihi Mart 27, 2025, <https://talentgrid.io/tr/python-kutuphaneleri/>
52. Sentetik Veri Türleri Ve Oluşturma Yöntemleri - Savaşan Adam, erişim tarihi Mart 27, 2025, <https://www.savasanadam.com/teknoloji/sentetik-veri-turleri-ve-olusturma-yontemleri/>
53. Popüler Python Kütüphaneleri: Mutlaka Bilmeniz Gerekenler - Patika.dev, erişim tarihi Mart 27, 2025, <https://www.patika.dev/blog/populer-python-kutuphaneleri-mutlaka-bilmeniz-gerekenler>
54. En Popüler Python Kütüphaneleri | Coderspace Blog, erişim tarihi Mart 27, 2025, <https://coderspace.io/blog/en-populer-python-kutuphaneleri/>
55. Veri Bilimi için 10 Python Kütüphanesi - Tirendaz Akademi - Medium, erişim tarihi Mart 27, 2025, <https://tirendazakademi.medium.com/veri-bilimi-i%C3%A7in-10-python-k%C3%BCt%C3%BCphanesi-bacf0600d298>
56. gretelai/gretel-text-to-python-fintech-en-v1 · Datasets at Hugging Face, erişim tarihi Mart 27, 2025, <https://huggingface.co/datasets/gretelai/gretel-text-to-python-fintech-en-v1>
57. Synthetic Financial Datasets For Fraud Detection - Kaggle, erişim tarihi Mart 27, 2025, <https://www.kaggle.com/datasets/ealaxi/paysim1>
58. Generating Realistic Synthetic Financial Time Series - Medium, erişim tarihi Mart 27, 2025, <https://medium.com/@a.skabar_60534/generating-realistic-synthetic-financial-time-series-0ad6c1b9bc18>
59. Report: Using Synthetic Data in Financial Services, erişim tarihi Mart 27, 2025, <https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf>
60. How to Visually Evaluate Your Synthetic Data Quality? - Ydata.ai, erişim tarihi Mart 27, 2025, <https://ydata.ai/resources/how-to-visually-evaluate-your-synthetic-data-quality>
61. Ensuring Quality and Realism in Synthetic Data - Keymakr, erişim tarihi Mart 27, 2025, <https://keymakr.com/blog/ensuring-quality-and-realism-in-synthetic-data/>
62. Improving the resilience of machine learning in financial systems through synthetic data, erişim tarihi Mart 27, 2025, <https://www.financialresearch.gov/partnerships/files/financial-machine-learning-resilience.pdf>
63. NMR USD Sentetik Veri Geçmişi - Investing.com, erişim tarihi Mart 27, 2025, <https://tr.investing.com/crypto/numeraire/nmr-usd-historical-data>
64. Bankacılıkta Yapay Zeka Uygulamaları\* - DergiPark, erişim tarihi Mart 27, 2025, <https://dergipark.org.tr/en/download/article-file/4205893>
65. Finansta Makine Öğrenme Yöntemleri: Son Uygulamalar ve Gelecek Beklentileri, erişim tarihi Mart 27, 2025, <https://ismailsen.com.tr/finansta-makine-ogrenme-yontemleri-son-uygulamalar-ve-gelecek-beklentileri/>
66. Fintek Dünyasında 2024 Yılı Nasıl Geçti? 2025'te Fintek Sektörünü Neler Bekliyor?, erişim tarihi Mart 27, 2025, <https://papel.com.tr/blog/fintek-dunyasinda-2024-yili-nasil-gecti-2025-te-fintek-sektorunde-neler-bekliyor>
67. 07.03.2025 tarihinde Türkiye Bankalar Birliği İnternet Sitesinde “Aralık 2024- Türkiye Finansal Sektör Ödeme Sistemleri Raporu” yayınlanmıştır. - Srp-Legal, erişim tarihi Mart 27, 2025, <http://www.srp-legal.com/tr/2025/03/14/07-03-2025-tarihinde-turkiye-bankalar-birligi-internet-sitesinde-aralik-2024-turkiye-finansal-sektor-odeme-sistemleri-raporu-yayinlanmistir/>
68. 2025 ve Sonrasında Perakende Bankacılık - PwC Türkiye, erişim tarihi Mart 27, 2025, <https://www.pwc.com.tr/2025-ve-sonrasinda-perakende-bankacilik>
69. Top 15 AI Trends in Banking for 2025 - UPTech Team, erişim tarihi Mart 27, 2025, <https://www.uptech.team/blog/ai-trends-in-banking>
70. 2025 Global Outlook for Banking and Financial Markets - IBM, erişim tarihi Mart 27, 2025, <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/2025-banking-financial-markets-outlook>
71. Using Data to Grow Deposits and Engagement in 2025 - MX Technologies, erişim tarihi Mart 27, 2025, <https://www.mx.com/whitepapers/2025-predictions/?utm_lp=homepage+banner>
72. Detail - Top Trends for Financial Services 2025: AI and Synthetic Data - Red Hat TV, erişim tarihi Mart 27, 2025, <https://tv.redhat.com/detail/6365850671112/top-trends-for-financial-services-2025-ai-and-synthetic-data>
73. Emerging trends for asset finance in 2025 - Sopra Banking Software, erişim tarihi Mart 27, 2025, <https://sbs-software.com/insights/trends-for-asset-finance-2025/>

Harika, sentetik veriye dair hem genel okuyucuya hem de akademik meraklılara hitap eden bir araştırma raporu hazırlayacağım. Finans ve bankacılık sektörüne odaklanarak, özellikle son 3 yıldaki gelişmeleri ele alacağım; tabular veri üretimi için daha eski kaynaklara da başvuracağım.

Rapor; sentetik verinin tanımı, türleri, kalite ölçüm yöntemleri, metin ve tablo tabanlı üretim teknikleri ile bu alanlarda kullanılan popüler kütüphaneleri kapsayacak. Medium gibi güncel blog ve yayınlardan da kaynaklara yer vereceğim. Hazır olduğunda seninle paylaşıyor olacağım.

# Finans Sektöründe Sentetik Veri: Tanım, Yöntemler ve Uygulamalar

## Giriş

Veri gizliliğinin kritik önem kazandığı günümüzde, gerçek veriye alternatif olarak **sentetik veri** kullanımına ilgi hızla artmaktadır. Gartner firmasının öngörüsüne göre 2030 yılı itibariyle yapay zeka eğitiminde sentetik verilerin kullanımının gerçek verileri aşacağı ve sentetik veri pazarının $2.3 milyar büyüklüğe ulaşacağı tahmin edilmektedir ([Synthetic Data: A Passing Trend or the Future of AI? | by Sciforce | Sciforce | Medium](https://medium.com/sciforce/synthetic-data-a-passing-trend-or-the-future-of-ai-65a64e3234b7#:~:text=With%20Gartner%20predicting%20it%20will,concept%20%E2%80%94%20it%E2%80%99s%20the%20future)). Sentetik veri; finans ve bankacılık başta olmak üzere sağlık, perakende gibi pek çok sektörde maliyetleri azaltma, veri paylaşımını kolaylaştırma ve inovasyonu hızlandırma potansiyeliyle dikkat çekmektedir ([Synthetic Data: A Passing Trend or the Future of AI? | by Sciforce | Sciforce | Medium](https://medium.com/sciforce/synthetic-data-a-passing-trend-or-the-future-of-ai-65a64e3234b7#:~:text=By%20providing%20privacy,synthetic%20data%20drives%20AI%20innovations)) ([Synthetic Data: A Passing Trend or the Future of AI? | by Sciforce | Sciforce | Medium](https://medium.com/sciforce/synthetic-data-a-passing-trend-or-the-future-of-ai-65a64e3234b7#:~:text=%2A%20Healthcare%3A%20Generates%20realistic%2C%20privacy,demand%2C%20and%20test%20new%20products)). Bu raporda sentetik verinin tanımı ve türleri, sentetik veri kalitesini değerlendirme yöntemleri, tabular ve metin bazlı sentetik veri üretiminde kullanılan modern yöntemler ve araçlar, finans/bankacılık sektöründeki kullanım örnekleri ve sektöre dair son yıllardaki güncel gelişmeler ele alınmıştır. Çalışma, hem genel okuyucu kitlesine hem de akademik ilgililere yönelik kapsamlı bir bakış sunmayı amaçlamaktadır.

## Sentetik Veri Nedir? Türleri Nelerdir?

**Sentetik veri**, gerçek dünyadaki olaylar sonucu toplanmamış, bunun yerine algoritmik olarak üretilmiş ve gerçek verinin istatistiksel özelliklerini taklit eden yapay verilerdir ([What is Synthetic Data? Examples, Use Cases and Benefits | TechTarget](https://www.techtarget.com/searchcio/definition/synthetic-data#:~:text=Synthetic%20data%20is%20information%20that%27s,and%20%2034%20models)) ([GitHub - ydataai/ydata-synthetic: Synthetic data generators for tabular and time-series data](https://github.com/ydataai/ydata-synthetic#:~:text=What%20is%20synthetic%20data%3F)). Başka bir deyişle, orijinal (gerçek) veri setindeki örüntü ve ilişkileri öğrenen bir model tarafından oluşturulan ve orijinal veriyle benzer analiz amaçlarıyla kullanılabilen verilerdir ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=This%20post%20talks%20about%20synthetic,same%20purpose%20as%20the%20original)). Sentetik veri üretiminde gerçek kayıtlar veya kişisel bilgiler birebir kullanılmaz; üretilen veriler istatistiksel olarak gerçeğe çok benzer olsa da herhangi bir sentetik kayıt ile gerçek bir birey arasında birebir eşleşme bulunmaz ([Synthetic Data Will Transform Fintech AI As We Know It](https://blog.mondato.com/synthetic-data-will-transform-fintech-ai-as-we-know-it/#:~:text=,Chief%20Trust%20Officer%2C%20Mostly%20AI)). Bu yönüyle sentetik veri, veri gizliliğini korurken orijinal verinin sağladığı içgörüleri sunabilen güçlü bir **Mahremiyet Güçlendirici Teknoloji (PET)** olarak değerlendirilmektedir ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=1,used%20to%20create%20advanced%20modelling)).

**Sentetik veri türleri** kullanılan veri formatına göre çeşitli kategorilere ayrılır ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=This%20post%20presents%20the%20different,and%20tabular%20synthetic%20data)):

* **Tabular (Yapısal) Veri:** Sütunlar ve satırlardan oluşan tablo biçimli veriler. Örneğin müşteri demografik bilgileri veya işlem kayıtları gibi yapısal veri setleri sentetik olarak üretilebilir. Tabular sentetik veriler, gerçek veriyle aynı istatistiksel dağılım ve ilişkilere sahip yapay kayıtlar içerir.
* **Metin Verisi:** Doğal dilde yazılmış metinlerin yapay olarak üretilmesidir. Örneğin bir banka çağrı merkezi konuşmaları, e-posta yazışmaları veya dolandırıcılık senaryoları içeren metinler büyük dil modelleri kullanılarak sentetik olarak oluşturulabilir.
* **Görsel ve İşitsel Veri:** Resim, video veya ses gibi medya verilerinin yapay üretimidir. Örneğin otonom araç geliştirmede sanal sürüş senaryoları için sentetik video verileri, veya finans sektöründe çek görüntülerinin sentetik üretimi mümkündür. Büyük teknoloji şirketleri görüntü ve ses tabanlı sentetik veriyi yoğun şekilde kullanmaktadır (ör. Google’ın otonom sürüş birimi Waymo, eğitim verisinin önemli bir kısmını sentetik sürüş sahnelerinden elde etmiştir ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=,medical%20data%20for%20clinical%20research))).
* **Zaman Serileri:** Finans alanında sıkça karşılaşılan zaman serisi veriler (örn. hisse senedi fiyatları, işlem hacimleri) de sentetik olarak üretilebilir. Bu veriler genellikle tabular formata yakın olmakla birlikte, zaman bağımlılığı içerdiğinden özel modeller gerektirebilir.

Sentetik verinin kullanılma motivasyonları arasında veri erişimindeki kısıtlar, gizlilik kaygıları ve nadir görülen durumların modellenmesi sayılabilir ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=First%2C%20it%20can%20be%20a,a%20cause%20of%20data%20unavailability)). Kurum içinde farklı birimlerin veya iş ortaklarının veri paylaşımı yasal engeller nedeniyle zorlaştığında, sentetik veri gerçek veriye kıyasla **GDPR** gibi düzenlemelere uyumlu bir alternatif sunar ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=Under%20GDPR%2C%20the%20possibility%20of,potential%20of%20data%20decreases%20significantly)). Ayrıca elde yeterli miktarda gerçek veri olmadığında (veya dengesiz dağılıma sahip olduğunda), sentetik verilerle veri seti genişletilip zenginleştirilerek makine öğrenimi modellerinin başarımı artırılabilir. Örneğin, sahte bankacılık işlemleri üretip dolandırıcılık tespit algoritmalarını eğitmek veya çok nadir gerçekleşen piyasa krizlerini simüle ederek risk yönetimi senaryoları oluşturmak sentetik verinin olanakları arasındadır.

## Sentetik Verinin Kalitesini Ölçme Yöntemleri

Gerçek veriyi taklit eden yapay verilerin **kalitesini değerlendirmek**, hem üretilen verinin kullanışlılığını (faydasını) hem de gizlilik açısından güvenliğini doğrulamak açısından kritik bir adımdır. Sentetik verinin kalitesini üç boyutta ele almak yaygın bir yaklaşımdır ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=The%20three%20dimensions%20of%20synthetic,data%20quality%20evaluation)):

1. **Aslına Uygunluk (Fidelity):** Sentetik verinin istatistiksel olarak ne derece gerçek veriye benzediğinin ölçüsüdür. Yüksek aslına uygunluk, sentetik verideki dağılımların ve ilişkilerin orijinal veriyle neredeyse aynı olduğunu gösterir. Bunu değerlendirmek için orijinal veri ile sentetik veri karşılaştırılarak ortalama, medyan, standart sapma, minimum-maksimum değerler, çeyrekler arası aralık gibi temel istatistikler incelenir ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=Exploratory%20statistical%20comparisons)). Her bir değişkenin histogramları üst üste konularak dağılımlar görsel ve sayısal olarak kıyaslanabilir; örneğin histogram örtüşme oranı veya Kolmogorov-Smirnov testi gibi metrikler kullanılabilir ([Hazy synthetic data quality metrics explained](https://hazy.com/resources/2021/01/15/hazy-synthetic-data-quality-metrics-explained#:~:text=Hazy%20synthetic%20data%20quality%20metrics,an%2080%20percent%20histogram%20overlap)). Ayrıca sentetik verideki kategorik değişkenlerin sınıf oranlarının gerçeğe yakınlığı da kontrol edilir. **Korelasyon analizi**, değişkenler arası ilişkilerin korunup korunmadığını anlamak için önemlidir: Orijinal veri setindeki korelasyon matrisi ile sentetik verininki karşılaştırılarak, ilişkilerin ne kadar iyi yakalandığı bir **korelasyon skoru** ile ölçülebilir (1’e yakın skor, korelasyonların birebir korunduğunu gösterir) ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=Correlation%20score)) ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=match%20at%20L276%20The%20correlation,correlations%20have%20been%20perfectly%20matched)).
2. **Fayda (Utility):** Sentetik verinin beklenen kullanım amacına hizmet etme derecesini ifade eder. Yani sentetik veri üzerinde çalıştırılan analizlerin veya eğitilen modellerin, gerçek veri kullanıldığında elde edilecek sonuçlara ne kadar yakın sonuçlar verdiği değerlendirilir. Örneğin, sentetik veriye dayanarak eğitilen bir makine öğrenimi modelinin, gerçek veriyle eğitilen modele yakın bir doğruluk veya F1 skoru elde etmesi, sentetik verinin yüksek faydaya sahip olduğunu gösterir ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=,to%20Synthesize%20Artificial%20Financial%20Datasets)). Sıklıkla uygulanan bir yöntem, *Train on Synthetic, Test on Real (TSTR)* şeklinde sentetik veride model eğitip gerçek veri üzerinde test etmek ve performansı ölçmektir. Benzer şekilde, gerçek veride önemli olan özniteliklerin (feature importance) sentetik veride eğitilen modelde de benzer sırada çıkması beklenir ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=match%20at%20L338%20to%20be,as%20the%20original%20real%20data)). Fayda metriğini ölçmek için QScore gibi özel bileşik metrikler de önerilmiştir ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=match%20at%20L304%20Using%20the,the%20original%20data%20has%20performed)).
3. **Gizlilik (Privacy):** Üretilen sentetik verinin orijinal veriden sızdırabileceği özel bilgilerin değerlendirmesidir. İyi bir sentetik veri, gerçek tekil kişi veya işlemlere dair hassas bilgileri ele vermeden genel örüntüleri yansıtmalıdır. Bu açıdan bakıldığında, sentetik veri ile orijinal veri arasında *bir-ya-da-sıfır* eşleşmelerin olmaması kritik bir kriterdir. Değerlendirme için sentetik kayıtlardan herhangi birinin gerçek veri kayıtlarına ne kadar benzediği ölçülebilir; örneğin her sentetik kaydın en yakın gerçek kayıtla mesafesinin hesaplanması ve bu mesafenin makul bir eşiğin üstünde olması beklenir (çok yakın veya birebir eşleşme varsa, bu bir gizlilik alarmıdır). Ayrıca **üyelik çıkarımı (membership inference)** testleri uygulanarak sentetik veri üretme modelinin, belirli bir gerçek verinin eğitim kümesinde olup olmadığını ele verip vermediği anlaşılabilir ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=With%20increasing%20legal%20and%20ethical,any%20of%20the%20original%20data)) ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=,inadvertently%20synthesized%20by%20our%20model)). Daha ileri seviyede, farklılaştırılmış gizlilik (differential privacy) garantileriyle çalışan modeller kullanılmışsa, buradaki ε (epsilon) değeri gizlilik seviyesini nicel olarak temsil edecektir. Özetle, gizlilik metriği sentetik verinin orijinaliyle ne kadar “az” benzeştiğini (doğrudan kopyalama yapmadığını) güvence altına alır.

Bu üç boyut bazen birbiriyle çelişebilir: **Aslına uygunluk** ve **fayda** genellikle benzer yönde hareket ederken, gizlilik ile tam uyum arasında bir denge gözetmek gerekir. Çok yüksek aslına uygunluk, modelin orijinal veriyi ezberleme riskini artırarak gizliliği zedeleyebilir; öte yandan gizliliği maksimize etmek adına fazla gürültü eklemek de verinin faydasını düşürebilir. Bu nedenle gerçek dünya uygulamalarında bu üç boyut arasında bir **ödünleşim (trade-off)** söz konusudur ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=match%20at%20L467%20Moreover%2C%20a,from%20the%20generated%20synthetic%20data)). Sentetik veri üreticileri ve kullanıcıları, kullanım amacına göre öncelikleri belirleyerek (örneğin finansal bir dolandırıcılık tespit yarışmasında fayda ön planda iken, müşteri verisi paylaşımında gizlilik ön planda tutulabilir) uygun dengeyi sağlamalıdır.

Akademik ve endüstriyel alanda sentetik veri değerlendirmesi için çeşitli metrik setleri ve araçlar geliştirilmiştir. Örneğin, MIT tarafından geliştirilen SDV (Synthetic Data Vault) projesinin bir parçası olan **SDMetrics** kütüphanesi, sentetik verinin gerçek veriye göre istatistiksel benzerliği ve aykırı değer içerip içermediği gibi pek çok metriği otomatik hesaplayabilmektedir. Benzer şekilde, bazı ticari platformlar (ör. Gretel.ai) her sentetik veri üretimi sonrası verinin kalite ve gizlilik skorlarını içeren raporlar sunmaktadır ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=Gretel%E2%80%99s%20Synthetic%20Data%20Quality%20Report,synthetic%20data%20generation%20process%20went)) ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=models%2C%20check%20out%20Gretel%E2%80%99s%20case,identifying%20and%20redacting%20datasets)).

## Tabular Sentetik Veri Üretim Yöntemleri ve Araçları

Finans ve bankacılıkta en yaygın veri tipi olan **tabular (tablo biçimli) veriler** için yıllar içinde çeşitli sentetik veri üretim yöntemleri geliştirilmiştir. Bu yöntemleri kabaca **istatistiksel yaklaşımlar** ve **modern yapay öğrenme yaklaşımları** olarak ikiye ayırmak mümkündür.

**İstatistiksel yöntemler**, her değişkenin dağılımını ve değişkenler arası ilişkileri klasik istatistiksel modellerle yakalayıp buradan yeni örnekler türetmeyi hedefler. Örneğin, bir veri kümesindeki her sütun için uygun olasılık dağılımları veya çok değişkenli ortak dağılımlar (ör. *copula* yöntemleri) tahmin edilerek bu dağılımlardan rastgele örneklem almak yoluyla sentetik veriler üretilebilir. Bir diğer yaklaşım olarak, karar ağaçları gibi yöntemlerle bir sütunun değerleri diğer sütunlara koşullu olarak örneklenebilir. R dilinde yaygın kullanılan **synthpop** paketi, hassas mikro verilerin sentetik versiyonlarını üretmek için bu tür istatistiksel eşleme ve sınıflandırma ağaçları tabanlı teknikler sunmaktadır ([synthpop - Generate synthetic data](https://www.synthpop.org.uk/#:~:text=Image%3A%20tree)). Synthpop’un amacı, orijinal verinin yapısını koruyan ancak gerçek bireyleri içermeyen veriler üreterek veri setinin istatistiksel özelliklerini araştırmacılara açarken gizliliği sağlamaktır ([CRAN: Package synthpop - R Project](https://cran.r-project.org/package=synthpop#:~:text=CRAN%3A%20Package%20synthpop%20,to%20users%20for%20exploratory%20analysis)). Bu yöntemler özellikle görece küçük ve yapısı iyi anlaşılan veri setlerinde başarılı sonuçlar verir ve açıklanmaları nispeten kolaydır.

**Modern yapay öğrenme tabanlı yöntemler** ise büyük ve karmaşık veri setlerinde daha üstün performans gösterebilmektedir. Bu alanda öne çıkan yaklaşımlar **Generative Adversarial Network (GAN)** ve **Variational Autoencoder (VAE)** gibi derin öğrenme teknikleridir. GAN tabanlı modeller, biri veri üreten (üretici) diğeri ise üretilen veriyi gerçek/veri ayrımı yaparak denetleyen (ayrımcı) iki sinir ağını karşı karşıya getirerek gerçekçi veri örnekleri üretir. Özellikle **CTGAN (Conditional Tabular GAN)** modeli, karma tipi verilerin (sürekli ve kategorik karışık) olduğu tablo verilerine özel olarak tasarlanmış bir GAN yaklaşımıdır ([CTGAN - YData-Synthetic](https://docs.synthetic.ydata.ai/2.0/synthetic_data/single_table/ctgan_example/#:~:text=CTGAN%20was%20specifically%20designed%20to,numeric%20and%20categorical%29%20data)). CTGAN, gerçek veri dağılımını koşullu modüllerle öğrenip, sentetik veride hem sayısal değişkenlerdeki dağılım biçimlerini hem de kategorik değişken oranlarını başarıyla taklit edebilir. 2019 yılında geliştirilen bu model, sentetik veri üretiminde bir benchmark (kıyas) haline gelmiş ve Python tabanlı açık kaynak **SDV** kütüphanesine entegre edilmiştir. GAN türevleri yanında, **VAE (Varyasyonel Otomatik Kodlayıcı)** tabanlı yöntemler de tabular veri sentezinde kullanılmaktadır. VAE’ler veri noktalarını bir latent (örtük) uzaya kodlayıp, bu uzayda örnekler alarak tekrar veri uzayına geri döndürür. Bu sayede orijinal veriye benzer ancak ondan farklı örnekler üretilebilir. SDV içerisinde **TVAE** adıyla tabular veriye uyarlanmış bir VAE modeli de bulunmaktadır. Derin öğrenme yaklaşımları, karmaşık çok boyutlu ilişkileri ve dağılımları öğrenmede başarılı olmakla birlikte, eğitim aşamasında büyük veri ve hesaplama gücü gerektirebilirler.

Yapay öğrenme yöntemlerinin yanı sıra, daha basit **kural tabanlı** veri üretim araçları da tabular verilerde kullanılmaktadır. Özellikle yazılım testleri veya hızlı prototipleme için, gerçekçi görünen rastgele veriler üretmek amacıyla Python dünyasında popüler olan **Faker** gibi kütüphaneler mevcuttur. Faker, isim, adres, telefon numarası, kredi kartı bilgisi gibi birçok alanda ön tanımlı veri üretme fonksiyonları ile veritabanlarını veya uygulamaları sahte verilerle doldurmayı sağlar ([GitHub - joke2k/faker: Faker is a Python package that generates fake data for you.](https://github.com/joke2k/faker#:~:text=Faker%20is%20a%20Python%20package,service%2C%20Faker%20is%20for%20you)). Ancak Faker gibi araçların ürettiği veriler *sentetik veri* kapsamına girmekle birlikte, bunlar genellikle gerçek bir veri setinin istatistiksel özelliklerini öğrenerek değil, önceden tanımlanmış örneklem havuzlarından rastgele seçerek veya kural bazlı oluşturulur. Dolayısıyla bu tür **"sahte veri" (dummy data)** araçları, istatistiksel yakınlık gerekmeyen senaryolarda (ör. arayüz testleri) kullanıma uygundur fakat gerçek verinin dağılımını korumak hedefiyle üretilen sentetik veri yaklaşımlarından ayrılır.

Açık kaynak topluluğu ve şirketler tarafından geliştirilmiş çeşitli sentetik veri araçları bulunmaktadır. Aşağıdaki tabloda önemli bazı yöntem ve araçlar özetlenmiştir:

| **Yöntem/Araç** | **Teknik Altyapı** | **Özellikler ve Kullanım Notları** |
| --- | --- | --- |
| **CTGAN** (YData) | Derin Öğrenme (GAN) | Tabular veriler için özel tasarlanmış koşullu GAN modeli. Sayısal ve kategorik verileri birlikte ele alarak gerçek veri dağılımlarını başarılı biçimde taklit eder ([CTGAN - YData-Synthetic](https://docs.synthetic.ydata.ai/2.0/synthetic_data/single_table/ctgan_example/#:~:text=CTGAN%20was%20specifically%20designed%20to,numeric%20and%20categorical%29%20data)). SDV kütüphanesinde uygulanabilir. |
| **TVAE** (SDV) | Derin Öğrenme (VAE) | Tabular VAE modeli. Veriyi latent alanda temsil ederek yeni örnekler üretir. Özellikle sürekli verilerde başarılıdır; SDV paketinde mevcuttur. |
| **Synthpop** (R) | İstatistiksel (Karar ağacı + Modelleme) | Hassas mikro verilerin sentetik kopyalarını üretmek için geliştirilmiş R paketi. Orijinal verinin yapısını karar ağaçları ve parametrik modeller ile yeniden oluşturarak, gizlilik korumalı sentetik veri setleri üretir ([synthpop - Generate synthetic data](https://www.synthpop.org.uk/#:~:text=Image%3A%20tree)). |
| **Faker** (Python) | Kural Tabanlı (Rastgele) | Test ve geliştirme amacıyla gerçekçi görünen rastgele veriler üretir. İsim, adres, finansal kimlik numarası gibi birçok kategori için hazır veri üreticileri bulunur ([GitHub - joke2k/faker: Faker is a Python package that generates fake data for you.](https://github.com/joke2k/faker#:~:text=Faker%20is%20a%20Python%20package,service%2C%20Faker%20is%20for%20you)). Gerçek veri dağılımını yansıtma iddiası olmaksızın, “dummy” veri sağlar. |
| **Gretel.ai** (Platform) | Derin Öğrenme (GAN, Transformer) | Bulut tabanlı bir sentetik veri platformu. Kullanıcı dostu arayüz ve API ile tabular ve metin verileri için sentetik veri üretir. GAN’ler ve diğer generatif modeller kullanarak yüksek kalite ve gizlilik ayarlı veriler sunar; üretim sonrası ayrıntılı kalite/gizlilik raporları verir. |
| **YData SDK** (Python) | Derin Öğrenme (Çoklu Model) | YData firmasının açık kaynak **ydata-synthetic** kütüphanesinin evrimi olan SDK, tabular ve zaman serisi veriler için birden fazla jeneratif modeli entegre şekilde sunar. Veri setine en uygun modeli otomatik seçip en iyi sonuca ulaşmayı hedefler ([GitHub - ydataai/ydata-synthetic: Synthetic data generators for tabular and time-series data](https://github.com/ydataai/ydata-synthetic#:~:text=With%20the%20update%20of%20%60ydata,specific%20dataset%20and%20use%20case)) ([GitHub - ydataai/ydata-synthetic: Synthetic data generators for tabular and time-series data](https://github.com/ydataai/ydata-synthetic#:~:text=,series%20data)). CTGAN, TimeGAN gibi modelleri arka planda barındırır. |
| **SDV (Synthetic Data Vault)** | Derin Öğrenme + İstatistiksel | MIT tarafından başlatılan açık kaynak proje. Tek tablo, çoklu tablo (ilişkisel veritabanı) ve zaman serileri için farklı sentetik veri modelleri içerir. CTGAN, TVAE, Copula modülleri gibi yöntemlerle kapsamlı bir sentetik veri üretim ve değerlendirme ekosistemi sağlar. Python diliyle kolay entegrasyon imkânı mevcuttur. |
| **GPT-3/4** (Büyük Dil Modeli) | Derin Öğrenme (Transformer) | Geniş ölçekli dil modelleri, özellikle *metin tabanlı* verilerin üretiminde yeni bir çığır açmıştır. GPT-3 ve GPT-4 modelleri, doğal dilde tutarlı ve gerçekçi metinler oluşturabilir. Finansal uygulamalarda, müşteri diyalogları, e-posta metinleri veya raporlar gibi serbest metinlerin sentetik üretiminde kullanılır. Ayrıca uygun yönlendirmeyle JSON/CSV formatında yapılandırılmış veri de üretebildikleri için, tabular veri üretiminde de yaratıcı bir yöntem olarak değerlendirilebilir. |

Yukarıda listelenenlerin yanı sıra **Mostly AI, Syntho, Hazy** gibi çeşitli şirketler de finans sektörüne yönelik sentetik veri yazılımları geliştirmektedir. Özellikle Mostly AI platformu, bankacılık verilerinin istatistiksel örüntülerini öğrenip *müşteri düzeyinde* sentetik veriler oluşturma konusunda uzmanlaşmıştır ([Synthetic Data Will Transform Fintech AI As We Know It](https://blog.mondato.com/synthetic-data-will-transform-fintech-ai-as-we-know-it/#:~:text=,Chief%20Trust%20Officer%2C%20Mostly%20AI)). Özetle, sentetik veri üretimi için hem açık kaynak hem de ticarî birçok seçenek bulunmaktadır ve kuruluşlar ihtiyaçlarına göre bu araçlardan faydalanmaktadır.

## Metin Tabanlı Sentetik Veri Üretimi

Doğal dildeki metinlerin sentetik olarak üretilmesi, son yıllarda **Büyük Dil Modelleri (Large Language Models, LLM)** sayesinde oldukça gerçekçi hale gelmiştir. Özellikle OpenAI tarafından geliştirilen GPT-3 ve GPT-4 gibi transformer tabanlı LLM'ler, insan dilini son derece tutarlı biçimde taklit edebilmekte ve istenen herhangi bir konuda metin üretebilmektedir ([Synthetic Data: A Passing Trend or the Future of AI? | by Sciforce | Sciforce | Medium](https://medium.com/sciforce/synthetic-data-a-passing-trend-or-the-future-of-ai-65a64e3234b7#:~:text=What%20is%20synthetic%20data%3F)). Finans ve bankacılıkta metin tabanlı sentetik veri kullanımı birkaç açıdan önem taşır:

* **Müşteri Etkileşim Verileri:** Bankaların çağrı merkezi kayıtları, müşteri şikayet/görüş mailleri veya chatbot konuşmaları gibi metin verileri genellikle kişisel bilgileri içerir ve paylaşımı kısıtlıdır. LLM tabanlı metin üretimi kullanılarak, gerçek etkileşimlerden öğrenen bir model yardımıyla benzer yapıda **sentetik diyaloglar** veya yazışmalar üretmek mümkündür. Örneğin, bir bankanın sık karşılaşılan müşteri talep ve sorularını içeren yapay konuşma kayıtları oluşturulup, yeni chatbot modelleri bu verilerle eğitilebilir. Böylece gerçek müşteri verisi ifşa edilmeden, benzer içerikli eğitim verisi elde edilmiş olur.
* **Finansal Dokümanlar:** Bankacılık sektöründe sözleşmeler, raporlar, piyasa bültenleri, analiz yazıları gibi çok sayıda metin dokümanı bulunur. Büyük dil modelleri, belirli bir finansal üslup ve formata uygun sentetik dokümanlar yaratmak için kullanılabilir. Örneğin, bir kredi değerlendirme raporunu örnek alarak benzer formatta birçok sentetik rapor üretip yeni işe alınan analistlerin eğitiminde kullanmak ya da bir piyasa haber özetini farklı varyasyonlarla sentetik olarak türetip algoritmik işlem modellerini test etmek mümkün hale gelmiştir.
* **Nadir/Olağanüstü Durum Senaryoları:** Doğal dilde nadir görülen veya geçmişte sınırlı örneği bulunan olaylara ilişkin metin verileri, LLM'ler kullanılarak çeşitlendirilebilir. Örneğin, tarihi birkaç finansal kriz dönemi haberi varsa, bunlardan yola çıkarak yeni **sentetik haber makaleleri** üretmek ve bir finansal risk algılama modelini daha zengin senaryolarla eğitmek değerlidir. Bu sayede model, sadece gerçek verideki birkaç örnekle sınırlı kalmayıp, olası farklı kriz anlatılarına da maruz kalabilir.

Metin tabanlı sentetik veri üretiminde LLM’lerin yanı sıra daha basit teknikler de kullanılabilir. **Markov zincirleri**, **n-gram modelleri** veya karakter seviyesinde RNN/LSTM modeller, LLM öncesinde sıkça kullanılan yöntemlerdi. Ancak günümüzde bu teknikler LLM'lerin gerçeğe yakınlık konusundaki başarısının gerisinde kalmıştır. Yine de, eğer hedef sadece anlamlı görünen ancak kurgu metinler oluşturmak ise (örneğin bir test veritabanını doldurmak için rastgele adres, paragraf, email içeriği üretmek gibi), Faker kütüphanesi benzeri araçlar metin üretimi için de hazır bileşenler sunar (rastgele cümleler, Lorem Ipsum paragraf üretimi vb).

Metin üretiminde dikkat edilmesi gereken bir konu, modelin eğitim verisinden ezberleyip hassas bilgileri sızdırmamasıdır. Örneğin büyük bir dil modeli, eğitilirken gördüğü gerçek bir kredi kartı numarasını veya müşteri ismini aynen üretmemelidir. Bu riskleri azaltmak için geliştiriciler, LLM tabanlı sentetik metin üretiminde gerektiğinde **differential privacy** tekniklerini uygulamaya veya model çıktısını filtrelemeye özen göstermektedir. Nitekim bankacılıkta sentetik veriye yönelimin ana motivasyonlarından biri gizlilik olduğu için, üretilen metnin de bu beklentiyi karşılaması gerekir.

## Finans ve Bankacılık Sektöründe Sentetik Veri Kullanım Örnekleri

Finans sektöründe sentetik veri kullanımının temelinde, müşteri mahremiyetini koruyarak veri odaklı inovasyonu sürdürme ihtiyacı yatmaktadır. Aşağıda bankacılık ve finans dünyasında öne çıkan bazı sentetik veri kullanım alanları ve örnekleri listelenmiştir:

* **Veri Paylaşımı ve İşbirliği:** Bankalar, düzenleyici kısıtlar nedeniyle müşteri verilerini kurum içindeki farklı departmanlar veya dış iş ortaklarıyla (fintech’ler, tedarikçiler) kolayca paylaşamamaktadır ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=Under%20GDPR%2C%20the%20possibility%20of,potential%20of%20data%20decreases%20significantly)). Sentetik veri, GDPR gibi yasalara tam uyumlu olması sayesinde bu engeli aşmaya yardımcı olur ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=Synthetic%20data%20is%20compliant%20with,or%20even%20with%20the%20government)). Örneğin bir banka, iş birliği yapmak istediği bir fintech şirketine gerçek müşteri verilerini veremese de, onlara istatistiksel olarak benzer sentetik müşteri verisi sağlayarak yeni ürün geliştirme sürecini hızlandırabilir. Birleşik Krallık’ta finans otoritesi FCA, sentetik veriyi hassas finansal verilerin paylaşımında kullanılabilecek önemli bir araç olarak tanımlamış ve 2023’te kurduğu *Sentetik Veri Uzman Grubu* ile finansal hizmetlerde veri paylaşım zorluklarını aşmak üzere sentetik veri kullanımını araştırmaya başlamıştır ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=1,the%20Synthetic%20Data%20Expert%20Group)) ([FCA issues report by Synthetic Data Expert Group on using synthetic data in financial services | Global Regulation Tomorrow](https://www.regulationtomorrow.com/eu/fca-issues-report-by-synthetic-data-expert-group-on-using-synthetic-data-in-financial-services/#:~:text=,sharing%20for%20fraud%20controls)). Özellikle **açık bankacılık** kapsamında, gerçek müşteri hesap verilerini ifşa etmeden fintech uygulamalarını test etme amacıyla sentetik veri kullanımının yaygınlaşması beklenmektedir.
* **Dolandırıcılık Tespiti (Fraud Detection):** Finansal dolandırıcılık olayları (kredi kartı sahtekarlığı, kimlik hırsızlığı, sahte hesap açma vb.) nispeten seyrek gerçekleşen ancak tespit edilmesi kritik öneme sahip durumlardır. Makine öğrenimi tabanlı dolandırıcılık tespit modelleri genellikle *dengesiz veri* problemine sahiptir: çok az sayıda sahte işlem, çok sayıda normal işlem arasında kaybolur. Sentetik veri bu sorunu hafifletmek için kullanılmaktadır. Örneğin, az sayıdaki gerçek dolandırıcılık vakasından yola çıkarak benzer özelliklere sahip ek sentetik vakalar üretilip eğitim setine katılarak modelin bu tür vakaları tanıma becerisi artırılır ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=For%20example%2C%20to%20detect%20certain,help%20to%20achieve%20optimal%20results)). American Express şirketi, kredi kartı işlemlerinde sahtekarlığı saptamak için tabular sentetik veri üzerinde Ar-Ge çalışmaları yapmış; gerçek işlemlerden üretilen istatistiksel olarak tutarlı sentetik verilerle sahtekarlık tespit modellerini eğitip başarılı sonuçlar elde edilebileceğini göstermiştir ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=The%20financial%20company%20American%20Express,help%20train%20the%20detection%20models)) ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=generate%20statistically%20accurate%20synthetic%20data,help%20train%20the%20detection%20models)). Hatta yapılan bir araştırmada, üretilen sentetik finansal verilerin dağılımının orijinaliyle aynı olduğu ve bu verilerle eğitilen modellerin performansının gerçek veriyle eğitilenlerle eş düzeyde olduğu ortaya konulmuştur ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=,to%20Synthesize%20Artificial%20Financial%20Datasets)). Bu sonuçlar, sentetik verinin dolandırıcılık tespitinde gerçek veri yerine geçebilecek kadar kaliteli olabileceğini göstermektedir.
* **Kredi Riski ve Skorlama:** Kredi başvurularını değerlendiren veya müşteri temerrüt riskini hesaplayan modeller de sentetik veriden fayda görebilir. Bankalar genellikle kredi skorlama modellerini geliştirirken yıllara yayılan müşteri finansal verilerine ihtiyaç duyarlar. Bu veriler çok hassastır ve bir ülkeden diğerine aktarımı bile yasal engellere takılabilir. Sentetik kredi verisi kullanarak, farklı coğrafyalardaki ekiplerin iş birliği yapması veya yeni bir modelin güvenle test edilmesi mümkün olmaktadır. Örneğin, bir bankanın geçmiş verilerine dayanarak oluşturulmuş sentetik müşteri finansal durum ve ödeme geçmişi verileri, yeni kredi skorlama algoritmalarını eğitmek için kullanılabilir. FCA tarafından 2022 yılında yapılan bir çalışma, finansal hizmetlerde sentetik verinin kredi skorlama dahil çeşitli risk modelleme alanlarında faydalı olabileceğini vurgulamıştır ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=insights%20on%20each,A%20full%20overview)). Sentetik veriler, modeli aşırı örnekleme (oversampling) yoluyla veri dengesizliklerini giderme veya nadir görülen temerrüt olaylarını çoğaltma gibi amaçlarla da destekler.
* **Nadir ve Aykırı Durum Analizi:** Finansal piyasalarda veya bankacılık operasyonlarında ender rastlanan ancak gerçekleştiğinde büyük etkisi olan olaylar (örneğin kara para aklama girişimleri, ani piyasa çöküşleri, vb.), geçmiş verilerde sınırlı sayıda örneğe sahip olabilir. Sentetik veri, bu tür olayların farklı varyasyonlarını simüle ederek hazırlık seviyesini artırmada kullanılır. Örneğin, bankalar **Anti-Money Laundering (AML)** sistemlerini eğitirken geçmişte yakalanmış birkaç kara para aklama vakasını alıp, para transfer tutarları, gönderici/alıcı profilleri, transfer süreleri vb. açıdan çeşitlendirerek yüzlerce sentetik vaka oluşturabilirler. Bu sentetik vakalar ile eğitilen bir AML modeli, olası farklı senaryolara karşı daha duyarlı hale gelecektir. Benzer şekilde, **yetkisiz ödeme (Authorized Push Payment) dolandırıcılığı** gibi yeni türeyen tehditlerde de, kısıtlı gerçek vaka bilgisinden sentetik türetmeler yapmak faydalıdır ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=insights%20on%20each,A%20full%20overview)). Piyasa riski tarafında ise, geçmişte birkaç kez yaşanmış aşırı oynaklık dönemlerini (ör. 2008 krizi, 2020 pandemi) sentetik verilerle çoğaltarak stres testleri yapmak, finansal kurumların direnç analizinde kullanılan bir yöntemdir.
* **Model Testi ve Validasyonu:** Bankalar yeni geliştirdikleri yapay zeka modellerini veya analitik sistemleri, üretime almadan önce kapsamlı testlere tabi tutarlar. Özellikle karar destek sistemlerinin beklenmedik durumlarda nasıl davranacağını görmek için geniş senaryo setlerine ihtiyaç duyulur. Sentetik veriler, bu test ortamlarını zenginleştirmek için idealdir. Örneğin, bir bankanın sahtekarlık tespit modeli üretime alınmadan önce, içeriği bilinen çeşitli sentetik müşteri ve işlem verileriyle beslenip modelin doğru alarm verip vermediği kontrol edilebilir. **Simülasyon** ortamlarında sentetik veri kullanımı, finansal kurumların "ne olursa" tarzı analizleri için de önemlidir ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=The%20applications%20of%20this%20technology,and%20increased%20collaboration%20across%20teams)). Piyasa simülasyonları, müşteri davranış simülasyonları (ör. mobil bankacılık uygulamasında belirli akışları test etmek için sentetik kullanıcı işlemleri) gibi birçok alanda yapay olarak üretilen veriler, sistemlerin performans ve güvenlik değerlendirmesine katkı sunar.
* **Veri Artırma ve Önyargı Giderme:** Sentetik veri, mevcut gerçek veri setini artırmak (data augmentation) ve veri içindeki önyargıları azaltmak için de kullanılır. Örneğin, eldeki kredi başvuru verisinde belirli bir demografik grup az temsil ediliyorsa, o gruba ait sentetik başvuru verileri üreterek modelin bu gruba karşı olası önyargısını azaltmak mümkündür ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=The%20applications%20of%20this%20technology,and%20increased%20collaboration%20across%20teams)) ([Top 5 Synthetic Data Finance Applications in 2025](https://research.aimultiple.com/synthetic-data-finance/#:~:text=Regulations%20such%20as%20GDPR%20and,between%20teams%20within%20an%20institution)). Yine, çok sınıflı bir finansal tahmin probleminde (örneğin farklı ürünlerin talep tahmini) düşük frekanslı sınıfları sentetik örneklerle artırarak model doğruluğunu iyileştirmek sentetik verinin bir diğer kullanım şeklidir. 2023 yılı itibarıyla yapılan bazı çalışmalar, gerçek veriyle desteklenmiş sentetik verilerle eğitilen derin öğrenme modellerinin sadece gerçek veriye kıyasla daha yüksek doğruluk ve adillik (fairness) değerleri yakalayabildiğini göstermektedir ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=How%20Accurate%20Is%20Synthetic%20Data,Real%20Data)) ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=often%20outperformed%20real,the%20dataset%E2%80%94the%20mean%20%2053)).

Yukarıdaki maddeler, finans alanında sentetik verinin çok yönlü kullanımını özetlemektedir. Özetle, **sentetik veri bankacılık sektöründe veri paylaşımını kolaylaştırmak, model geliştirmeyi hızlandırmak, nadir risklere hazırlanmak ve müşteri mahremiyetini korumak** gibi ihtiyaçlara yönelik önemli bir araç haline gelmiştir. Bu nedenle, global çapta büyük finans kuruluşları ve fintech girişimleri sentetik veri teknolojilerine yatırım yapmaya başlamıştır. Örneğin, İsviçre merkezli bir sigorta şirketi olan La Mobilière, müşteri kaybı (churn) tahmin modellerini geliştirirken müşterilerinin orijinal verisinden ürettiği sentetik verilerle model performansından ödün vermeden gizliliği sağladığını bildirmiştir ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=In%20the%20field%20of%20insurance%2C,or%20on%20their%20customer%20privacy)). Benzer şekilde, dünyanın en büyük finansal kurumlarından biri, geleneksel veri anonimleştirme yöntemlerinin yerine sentetik veri ile departmanlar arası veri paylaşım sürecini başarılı bir şekilde uygulamaya koymuştur ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=models%2C%20check%20out%20Gretel%E2%80%99s%20case,identifying%20and%20redacting%20datasets)). Bu örnekler, sentetik verinin laboratuvar çalışmalarından çıkıp fiilen finans dünyasında katma değer yaratmaya başladığını göstermektedir.

## Son Gelişmeler ve Trendler (2022-2025)

Son 3 yılda (2022-2025) sentetik veri alanında önemli gelişmeler yaşanmış ve özellikle finans sektörü için belirgin trendler ortaya çıkmıştır:

* **Regülatör Desteği ve Ekosistem Oluşumu:** 2022 yılından itibaren finansal regülatörler, sentetik veriye yönelik ilgilerini açıkça göstermeye başlamışlardır. Özellikle Birleşik Krallık FCA (Financial Conduct Authority), 2023’te bir *Sentetik Veri Uzman Grubu* kurarak finansal hizmetlerde sentetik veri kullanımını teşvik eden kapsamlı bir rapor yayınlamıştır ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=1,the%20Synthetic%20Data%20Expert%20Group)) ([FCA issues report by Synthetic Data Expert Group on using synthetic data in financial services | Global Regulation Tomorrow](https://www.regulationtomorrow.com/eu/fca-issues-report-by-synthetic-data-expert-group-on-using-synthetic-data-in-financial-services/#:~:text=,sharing%20for%20fraud%20controls)). Bu raporda dolandırıcılık tespiti, kredi skorlama, açık bankacılık, APP dolandırıcılığı ve AML gibi alanlar başta olmak üzere sentetik verinin sağladığı fırsatlar ve zorluklar ele alınmıştır ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=insights%20on%20each,A%20full%20overview)). Regülatörlerin bu konudaki proaktif tutumu, sektörde sentetik veri kullanımının “meşruiyet kazanması” açısından kritik bir gelişmedir. Ayrıca Avrupa Birliği’nin dijital finans stratejileri kapsamında sentetik verinin, güvenli veri paylaşımı ve AI geliştirme amaçlarıyla vurgulandığı görülmektedir. Bu dönemde kurulan endüstri çalışma grupları, yayınlanan kılavuzlar ve gerçekleştirilen konferanslar sentetik verinin finans ekosisteminde giderek daha fazla kabul gördüğüne işaret etmektedir.
* **Teknolojik İlerlemeler ve Yeni Yöntemler:** 2022-2025 arası, generative AI alanındaki hızlı ilerlemeler sentetik veri üretimini de doğrudan etkilemiştir. Görüntü tarafında **diffusion** tabanlı modellerin (DALL-E 2, Stable Diffusion vb.) başarısı, benzer fikirlerin tabular veri sentezine uyarlanabileceği tartışmalarını başlatmıştır. Metin tarafında ise 2023 yılında ChatGPT’nin popülarite kazanması, LLM’lerin sentetik veri üretimindeki rolünü gündeme getirmiştir. Artık geleneksel yöntemlerin yanında, *"Veri için Prompt Mühendisliği"* diyebileceğimiz bir yaklaşım ortaya çıktı: Örneğin bir finans uzmanı, ChatGPT benzeri bir modele uygun talimatlar vererek belirli senaryolara ilişkin sentetik veri ürettirebiliyor (kod veya formül yazmadan). 2022’de yayınlanan bir çalışma, "Reinforcement Prompting" adı verilen teknikle finansal duygu analizi için yüksek kaliteli metin veriler sentezlemeyi başarmıştır ([Reinforcement prompting for financial synthetic data generation](https://www.sciencedirect.com/science/article/pii/S2405918824000229#:~:text=Reinforcement%20prompting%20for%20financial%20synthetic,data%20for%20sentiment%20analysis%20tasks)). Yine 2023’de yayınlanan kapsamlı bir literatür taraması, tabular sentetik veri üretimi konusunda altı farklı yöntem kategorisine ayrılan 70 algoritmayı inceleyerek bu alandaki hızlı büyümeyi belgelemektedir ([Tabular and latent space synthetic data generation: a literature review | Journal of Big Data | Full Text](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00792-7#:~:text=synthetic%20data%20generation%20algorithms,informed%20practices%20with%20synthetic%20data)). Bu akademik ilgi patlaması, her geçen gün yeni ve daha iyi sentetik veri algoritmalarının ortaya çıktığını göstermektedir.
* **Kalite ve Gizlilikte İyileşmeler:** Yeni geliştirilen yöntemler, sentetik verinin kalitesini artırırken gizlilik koruma mekanizmalarını da entegre etmeye odaklanmıştır. Örneğin 2022-2023 yıllarında bazı sentetik veri platformları, **differential privacy** garantilerini modellerine entegre ettiklerini duyurdular. Bu sayede üretilen verinin her bir çıktısı matematiksel olarak tanımlı bir gizlilik seviyesi sağlıyor. Ayrıca 2024 yılında piyasaya sürülen sentetik veri araçlarının birçoğu, üretilen verinin orijinale kıyasla ne kadar iyi olduğunu tek bir skorla özetleyen **bileşik kalite metrikleri** sunmaya başladı. Gretel.ai firmasının tanıttığı *Synthetic Quality Score (SQS)* buna bir örnektir; 0-100 arası bir skorla verinin genel kalite ve gizlilik durumunu raporlamaktadır ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=part%20of%20Gretel%E2%80%99s%20Synthetic%20Quality,from%20a%20typical%20SQS%20report)) ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=Gretel%E2%80%99s%20Synthetic%20Data%20Quality%20Report,synthetic%20data%20generation%20process%20went)). Bunun gibi metrikler, kullanıcıların sentetik veriye olan güvenini artırmakta ve farklı jeneratif modelleri karşılaştırmayı kolaylaştırmaktadır. İlginç bir diğer gelişme, bazı durumlarda sentetik verinin modele katkısının gerçek veriyi aşabildiğinin gösterilmesidir. 2023 yılında Gretel.ai tarafından 8 farklı makine öğrenimi veri kümesi üzerinde yapılan bir denemede, sentetik verilerle eğitilen modellerin doğruluğunun gerçek verilerle eğitilenlere çok yakın olduğu (hatta kimi durumda daha yüksek olduğu) raporlanmıştır ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=How%20Accurate%20Is%20Synthetic%20Data,Real%20Data)) ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=often%20outperformed%20real,the%20dataset%E2%80%94the%20mean%20%2053)). Bu tür bulgular, sentetik veriye yönelik “kalite” endişelerini azaltarak, onu gerçek veriye güçlü bir alternatif olarak konumlandırmaktadır.
* **Sektörde Benimseme ve Yatırımlar:** Son yıllarda finans alanında faaliyet gösteren birçok kurum ve girişim, sentetik veri çözümlerine yatırım yapmaya başlamıştır. 2022 ve 2023 yıllarında çeşitli bankalar, sigorta şirketleri ve kredi büroları; Mostly AI, Synthesized, Statice, Hazy, Tonic AI, Syntho gibi sentetik veri sağlayıcılarıyla ortak projeler duyurdu. Örneğin, Wells Fargo bankasının veri inovasyonu birimi, sentetik veriyi kullanarak hem veri gizliliğini koruyup hem de yenilikçiliği hızlandırdıklarını bir röportajda aktarmıştır ([3 Questions: Innovation in financial services through synthetic data](https://mitibmwatsonailab.mit.edu/research/blog/3-questions-innovation-in-financial-services-through-synthetic-data/#:~:text=3%20Questions%3A%20Innovation%20in%20financial,innovation%20and%20expedite%20technology)). Yine bir başka büyük banka, müşteri verilerini anonimleştirmek yerine sentetik olarak yeniden üretmenin, model geliştirme süreçlerinde onlara aylar kazandırdığını belirtmiştir ([Synthetic Data Will Transform Fintech AI As We Know It](https://blog.mondato.com/synthetic-data-will-transform-fintech-ai-as-we-know-it/#:~:text=powered%20by%20synthetic%20data,over%20the%20next%20few%20years)). Pazar araştırmaları da bu benimseme trendini rakamlarla desteklemektedir: 2023 yılında küresel sentetik veri pazarı büyüklüğü yaklaşık $320 milyon olarak tahmin edilirken, 2025’e gelindiğinde katlanarak büyüme beklenmektedir ([Synthetic Data Generation Research Report, 2023 & - GlobeNewswire](https://www.globenewswire.com/news-release/2025/01/13/3008253/28124/en/Synthetic-Data-Generation-Research-Report-2023-2024-2030-Growing-Development-Platforms-and-Cloud-Based-Solutions-Expanding-Applications-in-Healthcare-Finance-and-Automotive-Sectors.html#:~:text=Synthetic%20Data%20Generation%20Research%20Report%2C,growing%20at%20a%20CAGR)). Özellikle finans sektörü, sağlık sektörü ile birlikte bu pazardaki en büyük talep alanlarından biri olarak öne çıkmaktadır. Sentetik veriye artan talebin arkasındaki itici güçler; mahremiyete yönelik toplumsal ve yasal hassasiyet, makine öğrenimine dayalı ürün sayısının artması ve kaliteli veriye duyulan ihtiyaç olarak sıralanabilir.

## Sonuç

Sentetik veri, finans ve bankacılık sektöründe veriyle ilgili sorunlara yenilikçi bir çözüm olarak ortaya çıkmıştır. Gerçek müşteri veya işlem verilerini kullanmanın riskli ya da imkânsız olduğu durumlarda, istatistiksel açıdan gerçeğe çok benzeyen yapay veriler üreterek hem **gizliliği korumak** hem de **veri odaklı değer yaratmaya** devam etmek mümkündür. Bu raporda tartışıldığı üzere, sentetik verinin başarısı üç temel boyutta değerlendirilir: orijinale benzerlik, kullanım amaçlarına uygunluk ve gizlilik güvence seviyesi. Modern sentetik veri üretim teknikleri (GAN, VAE, LLM vb.) sayesinde, bankalar işlem kayıtlarından müşteri yazışmalarına kadar pek çok alanda gerçeğe oldukça yakın sentetik veri setleri elde edebilmektedir. Üstelik bunu yaparken orijinal verideki önyargıları törpülemek, nadir olayları çoğaltmak gibi avantajlar da kazanılabilir.

Finans sektöründeki örnekler, sentetik verinin **dolandırıcılık tespitinden kredi risk analizine, veri paylaşımından model doğrulamasına** geniş bir yelpazede fayda sağladığını göstermektedir. Örneğin, American Express’in çalışması sentetik işlemlerle sahtekarlık tespit modellerini eğitmenin mümkün olduğunu ortaya koyarken ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=The%20financial%20company%20American%20Express,help%20train%20the%20detection%20models)), bir sigorta şirketinin churn analizinde sentetik veriden yararlanması model performansını düşürmeden gizlilik engelini aştığını kanıtlamıştır ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=In%20the%20field%20of%20insurance%2C,or%20on%20their%20customer%20privacy)). Bu gibi vakalar, sentetik verinin teorik bir kavram olmaktan çıkıp pratik bir araç haline geldiğinin altını çizmektedir.

Önümüzdeki yıllarda, sentetik verinin finansal kuruluşların **veri stratejilerinin ayrılmaz bir parçası** haline gelmesi beklenmektedir. Yapay zeka ve veri analitiği projelerinde sentetik veri kullanımı, kurumlara daha hızlı inovasyon, daha esnek test imkânları ve daha güçlü gizlilik kalkanı sağlayacaktır. Elbette, sentetik veri kullanırken dikkat edilmesi gereken noktalar da vardır: Üretilen verinin temsil gücünün yeterli olması, aşırı uyum (overfitting) barındırmaması, ve kullanım sınırlarının (ör. modelleme vs. raporlama) iyi anlaşılması önemlidir. Ayrıca, sentetik verinin de tamamen hatasız olmayacağı, gerçek veriyle doğrulamanın tamamen yerini almaktansa onu destekleyici bir rol oynayacağı bilinmelidir.

Sonuç itibariyle, finans ve bankacılıkta sentetik veri kullanımı yükselen bir trend olup, veri gizliliği ile yapay zeka geliştirmeyi bir arada mümkün kılması sayesinde sektörün karşılaştığı pek çok zorluğa çözümler sunmaktadır. Gerek regülasyon tarafında gerekse teknoloji cephesinde yaşanan son gelişmeler ışığında, önümüzdeki dönemde daha standartlaşmış metodolojiler, kalite kriterleri ve en iyi uygulama rehberleriyle birlikte sentetik verinin kurumsal bağlamda rutin bir araç haline geleceğini öngörebiliriz. **Veri bilimi dünyasında “yeni petrol” olarak anılan verinin, sentetik versiyonları sayesinde daha erişilebilir, paylaşılabilir ve güvenli hale gelmesi finans sektörünün dijital dönüşümünde kilit rol oynayacaktır.**

## Kaynakça

1. Hashemi-Pour, C., Yasar, K., & Laskowski, N. (2023). *What is synthetic data? Examples, use cases and benefits*. TechTarget. ([What is Synthetic Data? Examples, Use Cases and Benefits | TechTarget](https://www.techtarget.com/searchcio/definition/synthetic-data#:~:text=Synthetic%20data%20is%20information%20that%27s,and%20%2034%20models)) ([What is Synthetic Data? Examples, Use Cases and Benefits | TechTarget](https://www.techtarget.com/searchcio/definition/synthetic-data#:~:text=Why%20is%20synthetic%20data%20important%3F))
2. YData (2023). *YData-Synthetic GitHub Repository README*. ([GitHub - ydataai/ydata-synthetic: Synthetic data generators for tabular and time-series data](https://github.com/ydataai/ydata-synthetic#:~:text=What%20is%20synthetic%20data%3F)) ([GitHub - ydataai/ydata-synthetic: Synthetic data generators for tabular and time-series data](https://github.com/ydataai/ydata-synthetic#:~:text=Synthetic%20data%20is%20artificially%20generated,identifiable%20information%2C%20ensuring%20individuals%27%20privacy))
3. Devaux, E. (2021). *Types of synthetic data and real-life examples*. Statice Blog (Medium). ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=This%20post%20presents%20the%20different,and%20tabular%20synthetic%20data)) ([Types of synthetic data and real-life examples | by Elise Devaux | Statice | Medium](https://medium.com/statice/types-of-synthetic-data-and-real-life-examples-1ff0a89a9f5b#:~:text=The%20financial%20company%20American%20Express,help%20train%20the%20detection%20models))
4. AWS Machine Learning Blog (2022). *How to evaluate the quality of synthetic data – fidelity, utility, privacy*. ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=The%20three%20dimensions%20of%20synthetic,data%20quality%20evaluation)) ([How to evaluate the quality of the synthetic data – measuring from the perspective of fidelity, utility, and privacy | AWS Machine Learning Blog](https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/#:~:text=Exploratory%20statistical%20comparisons))
5. Fonseca, J., & Bacao, F. (2023). *Tabular and latent space synthetic data generation: a literature review*. *Journal of Big Data, 10*(115). ([Tabular and latent space synthetic data generation: a literature review | Journal of Big Data | Full Text](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00792-7#:~:text=synthetic%20data%20generation%20algorithms,informed%20practices%20with%20synthetic%20data))
6. Financial Conduct Authority – FCA (2023). *Using Synthetic Data in Financial Services (Report)*. ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=1,used%20to%20create%20advanced%20modelling)) ([Report: Using Synthetic Data in Financial Services](https://www.fca.org.uk/publication/corporate/report-using-synthetic-data-in-financial-services.pdf#:~:text=insights%20on%20each,A%20full%20overview))
7. Mondato (2023). *Synthetic Data Will Transform Fintech AI As We Know It*. ([Synthetic Data Will Transform Fintech AI As We Know It](https://blog.mondato.com/synthetic-data-will-transform-fintech-ai-as-we-know-it/#:~:text=,Chief%20Trust%20Officer%2C%20Mostly%20AI)) ([Synthetic Data Will Transform Fintech AI As We Know It](https://blog.mondato.com/synthetic-data-will-transform-fintech-ai-as-we-know-it/#:~:text=powered%20by%20synthetic%20data,over%20the%20next%20few%20years))
8. DEDOMENA (2022). *Synthetic Data use cases in Financial Industry*. ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=The%20applications%20of%20this%20technology,and%20increased%20collaboration%20across%20teams)) ([DEDOMENA • Synthetic Data use cases in Financial Industry](https://dedomena.ai/blog/synthetic_data_use_cases_financial_industry#:~:text=For%20example%2C%20to%20detect%20certain,help%20to%20achieve%20optimal%20results))
9. Sciforce (2024). *Synthetic Data: A Passing Trend or the Future of AI?* (Medium). ([Synthetic Data: A Passing Trend or the Future of AI? | by Sciforce | Sciforce | Medium](https://medium.com/sciforce/synthetic-data-a-passing-trend-or-the-future-of-ai-65a64e3234b7#:~:text=With%20Gartner%20predicting%20it%20will,concept%20%E2%80%94%20it%E2%80%99s%20the%20future)) ([Synthetic Data: A Passing Trend or the Future of AI? | by Sciforce | Sciforce | Medium](https://medium.com/sciforce/synthetic-data-a-passing-trend-or-the-future-of-ai-65a64e3234b7#:~:text=%2A%20Healthcare%3A%20Generates%20realistic%2C%20privacy,demand%2C%20and%20test%20new%20products))
10. Gretel.ai (2023). *What is Synthetic Data? Definition and FAQs*. ([What is Synthetic Data? Definition and FAQs | Gretel.ai](https://gretel.ai/technical-glossary/what-is-synthetic-data#:~:text=How%20Accurate%20Is%20Synthetic%20Data,Real%20Data))