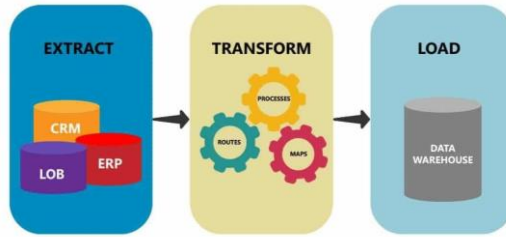


## DSML VE ETL

Dijital çağın hızla ilerlemesiyle birlikte, veri artık yalnızca bir yan ürün olmaktan çıkarak, stratejik kararlar almayı sağlayan temel bir varlık haline gelmiştir. Büyük veri çağında, verilerin anlamlı hale getirilmesi ve iş süreçlerine entegre edilmesi, işletmelerin rekabet avantajını koruyabilmesi için kritik bir gereklilik olmuştur. Bu noktada, ETL (Extract, Transform, Load) ve DSML (Data Science and Machine Learning) platformları, ham verinin değerli bilgilere dönüştürülmesi sürecinde kilit rol oynar. ETL süreçleri, verilerin çeşitli kaynaklardan toplanarak temizlenmesi, dönüştürülmesi ve analiz için uygun hale getirilmesi sürecini yönetirken, DSML platformları ise bu verileri kullanarak derinlemesine analizler yapmayı ve tahmin modelleri geliştirmeyi mümkün kılar. Bu yazıda, ETL ve DSML kavramlarının temellerini, kullanım alanlarını ve geleneksel yöntemlere kıyasla sundukları avantajları keşfedecek, aynı zamanda bu teknolojilerin iş dünyasında nasıl bir dönüşüm yarattığını inceleyeceğiz.

### 1)ETL

#### a) ETL Nedir?



ETL, verilerin kaynaktan hedef depoya aktarılması sürecini ifade eder. **Extract**, **Transform** ve **Load** kelimelerinin kısaltmasıdır. Veriler öncelikle mevcut kaynaklardan çıkarılır ve bu veriler daha sonra istenen formata dönüştürülür ve ardından daha fazla analiz için depoya yüklenir.

**Veri Çekme (Extract):** Farklı kaynaklardan (veritabanları, dosyalar, API'lar) veri toplama.

**Veri Dönüştürme (Transform):** Verilerin temizlenmesi, filtrelenmesi, zenginleştirilmesi ve analiz için uygun formatlara dönüştürülmesi.

**Veri Yükleme (Load):** Dönüştürülmüş verilerin hedef bir sisteme veya veri ambarına aktarılması.

#### b) Piyasadaki Örnekleri Nelerdir?

##### Apache NiFi:

- **Özellikleri:** Verilerin otomatik olarak toplanması, yönlendirilmesi ve dönüştürülmesi için açık kaynaklı bir veri akışı otomasyon platformu.

##### Talend:

- **Özellikleri:** Açık kaynaklı ve ticari ETL çözümleri sunar. Büyük veri entegrasyonu ve bulut tabanlı hizmetler için güçlü yeteneklere sahiptir.

##### Informatica PowerCenter:

- **Özellikleri:** Kurumsal düzeyde veri entegrasyonu ve veri kalitesi yönetimi için gelişmiş araçlar sunar.

##### Microsoft SSIS (SQL Server Integration Services):

- **Özellikleri:** SQL Server için veri entegrasyonu ve iş akışı otomasyonu sağlar. Kullanıcı dostu bir arayüz ile geniş veri kaynakları desteği sunar.

#### **Apache Kafka:**

- **Özellikleri:** Gerçek zamanlı veri akışı işleme ve veri entegrasyonu için ölçeklenebilir ve güvenilir bir platform.



Gartner'ın Temmuz 2020'de yayınladığı araştırma sonucuna göre sektörde en çok Informatica, IBM(InfoSphere DataStage), Talend, SAP(BODS) ,Oracle(ODI) kullanılmaktadır.

#### **c) Kullanım Alanları Nereledir?**

Aslında kaynaktan veri çıkartıp işimize yaramayacak veri parçalarını siliyor ve yorumlayabileceğimiz formata çeviriyoruz. Sonrasında da bunu kaydediyoruz. Bu süreci (ETL) günlük hayatımızda neredeyse hepimiz yapıyoruz. Elde ettiğimiz verileri bir yere kaydedip daha sonra yorumlamamız da ELT oluyor.

Örneğin ingilizce bir makaleden bilmediğiniz kelimeleri ayıklıyorsunuz. Bu işlem bir extract aslında.

Ayıkladığınız bu verileri türkçeye çeviriyorsunuz, bu da transform oluyor.

Sonunda ise bu kelimeleri bilmediğiniz kelimeler defterinize yazıyorsunuz bu kısım da Load kısmı oluyor.

Yani ETL veya ELT 'yi neredeyse her yerde kullanıyoruz. Kullanıldığı başlıca alanlar:

**Veri Ambarı Besleme:** Kurumların büyük veri setlerini toplaması ve merkezi bir veri ambarında depolaması.

**İş Zekası (BI):** Karar verme süreçlerini desteklemek için analitik ve raporlama sistemleri ile entegrasyon.

**Veri Kalitesi Yönetimi:** Verilerin tutarlılığını, doğruluğunu ve bütünlüğünü sağlamak.

**Müşteri İlişkileri Yönetimi (CRM):** Farklı müşteri temas noktalarından elde edilen verilerin entegre edilmesi.

**E-ticaret Veri Entegrasyonu:** Çeşitli satış kanallarından gelen verilerin analiz edilmesi ve raporlanması.

#### d) **Seri Üretimle Olan Benzerlikleri Nelerdir?**

ETL platformları, tıpkı seri üretim hatları gibi verilerin sistematik ve tekrarlanabilir bir şekilde işlenmesini sağlar. Otomasyon ve süreç optimizasyonu ile veri işleme adımları daha verimli hale getirilir, bu da veri entegrasyonu ve analiz süreçlerinin hızlanmasına olanak tanır.

#### e) **Component ve Mikroişlemcilerle Olan Benzerlikleri Nedir?**

ETL platformları, komponent tabanlı bir yaklaşıma sahiptir. Her bir ETL süreci, farklı bileşenler (komponentler) kullanılarak oluşturulur. Mikroişlemcilerdeki modüler tasarım felsefesine benzer şekilde, ETL platformlarında da farklı işlemler (veri çekme, dönüştürme, yükleme vb.) modüler bir şekilde yapılandırılır.

#### f) **Sağladığı Avantajlar Geleneksel Programlamaya Göre Nelerdir?**

- **Otomasyon:** Tekrarlayan veri işleme görevlerini otomatikleştirir, insan hatalarını minimize eder ve iş gücü tasarrufu sağlar.
- **Hız ve Verimlilik:** Verilerin hızlı bir şekilde işlenmesi ve analiz edilmesi için optimize edilmiş süreçler sunar.
- **Kullanıcı Dostu Araçlar:** Kodlama bilgisi az olan kullanıcıların bile karmaşık veri entegrasyon görevlerini kolayca gerçekleştirmesine olanak tanır.
- **Esneklik:** Farklı veri kaynakları ile kolayca entegre olabilir ve değişen iş ihtiyaçlarına hızla uyum sağlayabilir.

#### g) **Üretim Aracı ve Üretim Yöntemi Olarak Kazandırdığı Yenilikler Nelerdir?**

ETL platformları, verilerin daha etkin bir şekilde yönetilmesi ve işlenmesi için yenilikçi yöntemler sunar:

- **Gerçek Zamanlı İşleme:** Anlık veri işlemleri ile daha hızlı ve etkili karar alma süreçleri sağlar.
- **Veri Kalitesi:** Veri doğrulama ve temizlik süreçleri ile daha kaliteli veri elde edilir.
- **Hızlı Prototipleme:** Yeni veri işleme çözümleri hızlı bir şekilde geliştirilebilir ve test edilebilir.

#### h) **ETL Platformlarının dünyadaki yaygınlıkları ve Zamana Göre Değişimi**

ETL platformları, veri miktarının hızla artması ve iş ihtiyaçlarının çeşitlenmesi ile giderek daha yaygın hale gelmiştir. Başlangıçta büyük ölçekli işletmeler tarafından kullanılan ETL çözümleri, günümüzde orta ve küçük ölçekli işletmelerde de yaygınlaşmıştır. Bulut tabanlı çözümler ve açık kaynak araçların yaygınlaşması, ETL platformlarının erişilebilirliğini artırmıştır. Zamana göre ETL platformlarının evrimi, teknolojik gelişmeler ve iş dünyasının değişen ihtiyaçları doğrultusunda sürekli bir değişim göstermektedir.

## 2)DSML

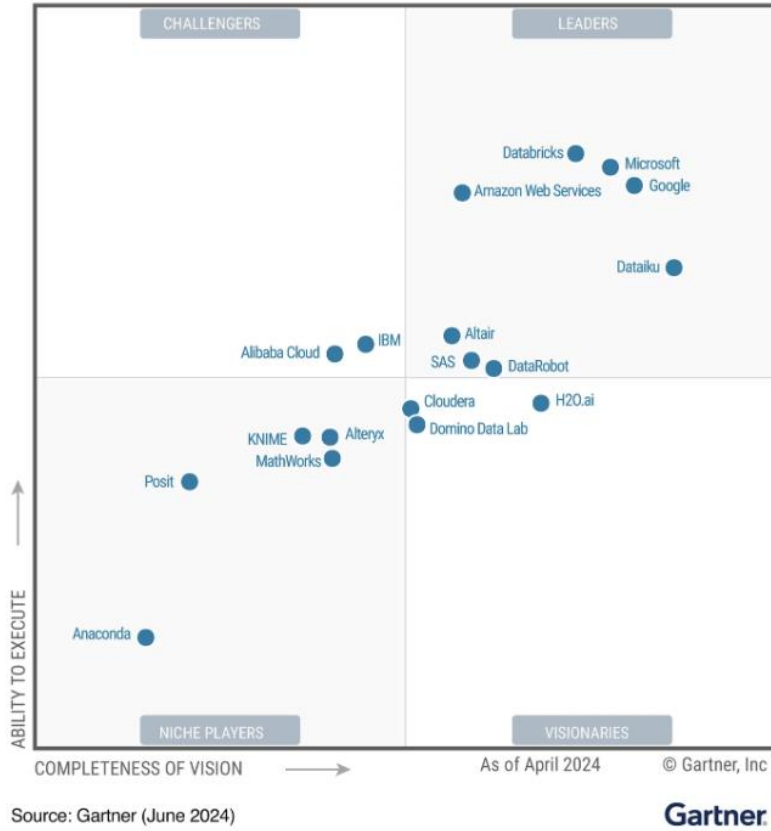
### a) DSML Nedir?

Veri bilimi ve makine öğrenimi (DSML) platformu, veri bilimcileri ve makine öğrenimi mühendislerinin modelleri hızla geliştirmesini ve veri analitiğini kolaylaştırmasını sağlayan bir dizi araç ve hizmettir.

### b) Piyasadaki Örnekleri Nelerdir?

- **KNIME:** Veri analizi, veri madenciliği ve makine öğrenimi süreçlerini entegre eden, açık kaynaklı bir platformdur.
- **RapidMiner:** Veri hazırlığı, modelleme ve değerlendirme süreçlerini bir arada sunan, kullanıcı dostu bir platformdur.
- **Alteryx:** Veri entegrasyonu, veri analizi ve makine öğrenimi için kapsamlı bir platformdur. Veri işleme ve analiz sürecinde çeşitli araçlar sunar.
- **DataRobot:** Otomatik makine öğrenimi (AutoML) platformu, veri bilimi sürecini kolaylaştırarak hızlı model oluşturma ve değerlendirme imkanı sağlar.
- **Orange:** Veri analizi ve görselleştirme için açık kaynaklı bir platformdur. Sürükle-bırak arayüzü ile çeşitli veri bilimsel işlemler yapılabilir.
- **TIBCO Spotfire:** Veri görselleştirme ve analitik araçlar sunan, aynı zamanda makine öğrenimi işlevlerine sahip bir platformdur

Figure 1: Magic Quadrant for Data Science and Machine Learning Platforms



### c) Kullanım Alanları Nereledir?

**İş Analitiği:** Şirketler, iş kararlarını desteklemek için veri analizi ve raporlama araçları kullanır. Satış, pazarlama ve finans gibi alanlarda verileri analiz ederek stratejik kararlar alırlar.

- **Müşteri Analizi:** Müşteri davranışlarını ve tercihlerini anlamak için veri biliminden faydalanılır. Bu, hedefleme stratejilerini geliştirmek ve müşteri deneyimini iyileştirmek için kullanılır.

- **Sağlık Sektörü:** Hasta verilerini analiz ederek teşhis, tedavi ve sağlık hizmetleri yönetimi konusunda iyileştirmeler yapılır. Ayrıca, sağlık araştırmaları ve genetik analizlerde de kullanılır.

- **Finans:** Risk analizi, dolandırıcılık tespiti, piyasa tahminleri ve yatırım stratejileri geliştirme gibi finansal analizler için kullanılır.

- **Perakende:** Stok yönetimi, talep tahmini, fiyat optimizasyonu ve müşteri kişiselleştirmesi gibi alanlarda veri bilimi kullanılır.

- **Üretim ve Operasyonlar:** Üretim süreçlerini optimize etmek, bakım ihtiyaçlarını tahmin etmek ve operasyonel verimliliği artırmak için analizler yapılır.

- **Enerji ve Çevre:** Enerji tüketimini optimize etmek, çevresel etkileri izlemek ve yenilenebilir enerji kaynaklarını değerlendirmek için veri analitiği kullanılır.
- **Eğitim:** Öğrenci performansını değerlendirmek, eğitim materyallerini kişiselleştirmek ve eğitim süreçlerini iyileştirmek için veri bilimi araçları kullanılır.
- **Ulaşım ve Lojistik:** Trafik analizi, rota optimizasyonu ve tedarik zinciri yönetimi gibi alanlarda veri biliminden yararlanılır.
- **Eğlence ve Medya:** İçerik önerileri, kullanıcı davranışları analizi ve medya tüketim trendlerini anlamak için kullanılır.

#### d) Seri Üretimle Olan Benzerlikleri Nelerdir?

DSML (Data Science and Machine Learning) ve seri üretim arasındaki benzerlikler, süreçlerin yapılandırılması, standartlaştırılması ve verimlilik hedeflerinin paylaşılmasıyla ilgilidir. İşte bu iki alan arasındaki bazı benzerlikler:

1. **Standartlaştırma:** Seri üretimde olduğu gibi, DSML projelerinde de belirli standartlar ve protokoller kullanılır. Bu, veri hazırlığından model değerlendirmeye kadar olan süreçlerin tutarlı ve tekrar edilebilir olmasını sağlar.
2. **Otomasyon:** Seri üretimde otomasyon, üretim süreçlerini hızlandırır ve hataları azaltır. Benzer şekilde, DSML'de otomatik veri işleme, model seçimi ve hiperparametre ayarı gibi otomasyon teknikleri kullanılır.
3. **Verimlilik:** Her iki alanda da verimlilik önemli bir hedeftir. Seri üretimde üretim sürecinin verimli olması beklenirken, DSML projelerinde de veri analizi ve model eğitimi süreçlerinin verimli olması istenir.
4. **Tekrar Edilebilirlik:** Seri üretim, aynı ürünün yüksek miktarda ve tutarlı kalitede üretilmesini sağlar. DSML projelerinde de modelin aynı veri kümesi üzerinde aynı sonuçları vermesi, sonuçların güvenilirliğini ve tekrar edilebilirliğini sağlar.
5. **Kalite Kontrol:** Seri üretimde kalite kontrol adımları, ürünlerin belirli standartlara uygun olmasını sağlar. DSML'de de model performansı ve veri kalitesi sıkı bir şekilde kontrol edilir, bu da sonuçların doğruluğunu ve güvenilirliğini sağlar.
6. **Modülerlik:** Seri üretim süreçlerinde çeşitli modüller ve bileşenler bir araya getirilir. Benzer şekilde, DSML projelerinde de veri işleme, modelleme ve değerlendirme gibi süreçler modüler olarak yapılandırılır.
7. **Süreç İyileştirme:** Her iki alanda da süreçlerin sürekli olarak iyileştirilmesi hedeflenir. Seri üretimde bu, üretim yöntemlerinin geliştirilmesi anlamına gelirken; DSML'de bu, model performansını artırmak için yeni yöntemlerin ve tekniklerin uygulanmasını içerir.

#### e) Component ve Mikroişlemcilerle Olan Benzerlikleri Nedir?

DSML (Data Science and Machine Learning) ile bileşenler (components) ve mikroişlemciler arasındaki benzerlikler, genellikle yapı taşları ve entegrasyon süreçleriyle ilgilidir. İşte bazı benzerlikler:

1. **Modülerlik:**
  - **Mikroişlemciler:** Mikroişlemciler, belirli işlevleri yerine getiren modüller veya bileşenler içerir (örneğin, ALU, registerlar, kontrol birimi). Her bir modül belirli bir işlevi yerine getirir ve birlikte çalışarak genel işlemci işlevini sağlar.

- **DSML:** DSML projelerinde de veri işleme, modelleme, ve değerlendirme gibi modüller bulunur. Her bir modül belirli bir görevi yerine getirir ve birlikte çalışarak veri bilimi sürecinin tamamlanmasını sağlar.
2. **Entegrasyon:**
- **Mikroişlemciler:** Mikroişlemciler, farklı bileşenlerin bir araya gelmesiyle oluşturulur. Bu bileşenler entegre bir şekilde çalışarak karmaşık işlemleri gerçekleştirir.
  - **DSML:** DSML projelerinde veri kaynakları, veri işleme araçları, modelleme algoritmaları ve değerlendirme metrikleri gibi farklı bileşenler entegre edilir. Bu bileşenler bir araya gelerek veri bilimsel analizler ve makine öğrenimi modelleri oluşturur.
3. **Verimlilik ve Performans:**
- **Mikroişlemciler:** Mikroişlemciler, yüksek performanslı işlem yapabilmek için optimize edilmiştir. Performans, işlem hızına ve verimliliğine dayanır.
  - **DSML:** DSML projelerinde de verimlilik ve performans kritik öneme sahiptir. Verimli veri işleme ve hızlı model eğitimi, sonuçların kalitesini etkiler.
4. **Fonksiyonellik:**
- **Mikroişlemciler:** Mikroişlemciler belirli bir işlem fonksiyonunu gerçekleştirir. Farklı işlemciler farklı işlevlere odaklanabilir (örneğin, işlem gücü, enerji verimliliği).
  - **DSML:** DSML araçları ve algoritmaları belirli fonksiyonları yerine getirir (örneğin, veri temizleme, model tahmini). Farklı algoritmalar ve araçlar farklı işlevlere odaklanır.
5. **Tasarım ve Optimizasyon:**
- **Mikroişlemciler:** Mikroişlemciler tasarlandıkları işlevleri en iyi şekilde yerine getirecek şekilde optimize edilir. Tasarım, işlemciye özgü ihtiyaçlara göre yapılır.
  - **DSML:** DSML projeleri de belirli veri setleri ve iş problemleri için optimize edilir. Model seçimi ve hiperparametre ayarları, problem türüne göre yapılır.

**f) Sağladığı Avantajlar Geleneksel Programlamaya Göre Nelerdir?**

- **Hız ve Verimlilik:** Veri hazırlama, modelleme, ve dağıtım süreçlerinin otomasyonu sayesinde zaman tasarrufu sağlar.
- **Kolay Entegrasyon:** Farklı veri kaynakları ve diğer yazılımlarla kolayca entegre olabilir.
- **Kullanıcı Dostu Arayüzler:** Az kod bilgisi olan kullanıcıların bile karmaşık analizler yapabilmesine olanak tanır.
- **Ölçeklenebilirlik:** Büyük veri kümeleri üzerinde çalışabilme yeteneği ile geleneksel yaklaşımlara kıyasla daha hızlı ve esnek çözümler sunar.

**g) Üretim Aracı ve Üretim Yöntemi Olarak Kazandırdığı Yenilikler Nelerdir?**

**Otomasyon:** DSML veri süreçlerini otomatikleştirir; üretimde otomasyon verimliliği artırır.

**Tahmin ve Planlama:** DSML, talep ve bakım tahminleri sağlar; üretim planlamasını optimize eder.

**Kalite Kontrol:** DSML kalite kontrolü ve anomali tespiti yapar; üretimde kalite kontrol hataları azaltır.

**Veri Analizi:** DSML büyük veri analizi sunar; üretim verileri süreç iyileştirmelerine yardımcı olur.

**Enerji ve Kaynak Yönetimi:** DSML enerji ve kaynak kullanımını optimize eder; üretim enerji verimliliğini artırır.

**Kişiselleştirme:** DSML müşteri verileriyle kişiselleştirilmiş öneriler sunar; üretim kişiselleştirilmiş ürünler üretir.

**Simülasyon ve Modelleme:** DSML süreçleri simüle eder; üretim süreçlerini test ve optimize eder.

### **DSML Platformlarının dünyadaki yaygınlıkları ve Zamana Göre Değişimi**

#### **Başlangıç Dönemi (2000'lerin başı):**

- **Yaygınlık:** DSML araçları sınırlıydı ve çoğunlukla akademik ve araştırma ortamlarında kullanılıyordu.
- **Değişim:** Veri analitiği ve makine öğrenimi yöntemleri üzerine çalışmalar başladı; bu dönemde ilk makine öğrenimi algoritmaları ve veri analizi araçları geliştirildi.

#### **Yükseliş Dönemi (2010'lar):**

- **Yaygınlık:** DSML araçları ve yöntemleri hızla yaygınlaştı. Büyük veri ve bulut bilişim kavramları popüler hale geldi.
- **Değişim:** Python ve R gibi programlama dilleri veri bilimi için yaygın olarak kullanılmaya başlandı. Apache Hadoop ve Spark gibi büyük veri araçları gelişti.

#### **Olgunlaşma Dönemi (2020'ler):**

- **Yaygınlık:** DSML araçları geniş bir endüstri yelpazesinde kullanılmaya başlandı; finans, sağlık, perakende, üretim gibi birçok sektörde entegre edildi.
- **Değişim:** Otomatik makine öğrenimi (AutoML), derin öğrenme ve yapay zeka uygulamaları popüler hale geldi. Google Colab, AWS SageMaker ve Azure Machine Learning gibi bulut tabanlı platformlar yaygınlaştı.

#### **Gelecek Trendleri:**

- **Yaygınlık:** DSML araçlarının kullanımı daha da yaygınlaşacak; daha fazla sektörde ve küçük ölçekli işletmelerde kullanılacak.
- **Değişim:** Kişiselleştirilmiş AI çözümleri, daha gelişmiş otomasyon ve etik veri kullanımı gibi trendler öne çıkacak. Ayrıca, kuantum hesaplama ve yapay zeka entegrasyonları gibi yenilikler bekleniyor.

### **Kaynakça**

- <https://uzmanposta.com/blog/etl/>
- <https://tr.linkedin.com/pulse/etl-ile-elt-aras%C4%B1ndaki-fark-umut-can-kocaman>
- <https://medium.com/@ashnumpy/etl-elt-ve-veri-ambar%C4%B1-nedir-nerelerde-kullan%C4%B1l%C4%B1r-ve-nas%C4%B1l-kullan%C4%B1l%C4%B1r-ce6e5ac765e1>
- <https://www.informatica.com/in/ipaas-magic-quadrant.html>
- <https://www.equalexerts.com/blog/tech-focus/lessons-data-science-machine-learning/>



- <https://lotusaibilisim.com/dsml-araclari-knime-sas-via-alteryx-dataiku/>
- <https://www.datanami.com/2024/06/21/cloud-giants-gain-in-dsml-platform-while-genai-fans-the-flame-gartner/>
- <https://chatgpt.com/>