

# Bir Reklam Aracısı Yazılımının Makine Öğrenmesi Yetenekleriyle İyileştirilmesi için Planlama

Görkem Giray<sup>1</sup>, Murat Osman Ünalır<sup>2</sup>, Cem Çatıkkaş<sup>1</sup>

<sup>1</sup> Kokteyl A.Ş., İstanbul, Türkiye  
gorkem.giray@kokteyl.com, cem.catikkas@kokteyl.com  
<sup>2</sup> Ege Üniversitesi, Bilgisayar Mühendisliği Bölümü, İzmir  
murat.osman.unalir@ege.edu.tr

**Özet.** Bir Reklam Aracısı Yazılımı, hangi reklam ağından hangi sırada reklam talep edeceğine dair kararı verirken uygulamanın reklam gelirini azami seviyeye çıkarmayı hedefler. Bu kararı ne kadar iyi verebileceği, reklam gelirini ölçmek için kullanılan eGBM (Etkin Bin Gösterim Başına Maliyet) değerini ne kadar iyi tahmin edebileceğine bağlıdır. Mevcut durumda bu tahmin, önceden kodlanmış bir karar mekanizması ve konu alanı uzmanları tarafından belirlenen parametreler doğrultusunda yapılmaktadır. Bu tahminin daha iyi yapılabilmesi ve zaman içinde oluşan veri doğrultusunda güncellenebilmesi için makine öğrenmesi algoritmalarının kullanılması mümkündür. Bu bildiride, mevcut sisteme makine öğrenmesi yetenekleri kazandırmak için yapılan planlama çalışmaları sunulmaktadır. Bu kapsamda bir makine öğrenmesi süreci, bu süreci gerçekleştirmek için tasarlanan bir mimari ve geliştirme takımının sahip olması gereken yetkinlikler anlatılmıştır.

**Anahtar Kelimeler:** Yazılım Mimarisi, Makine Öğrenmesi Süreci, Mobil Reklamcılık, Reklam Aracısı Yazılımı.

## Planning for the Enhancement of an Ad Mediation Software with Machine Learning Capabilities

**Abstract.** An Ad Mediation Software aims to maximize the ad revenue of an application while deciding from which ad network to request an ad. How well it can make this decision depends on how well it can predict the value of eCPM (Effective Cost Per Mille) that is used to measure advertising revenue. Currently, this prediction is made according to a pre-coded decision mechanism and parameters determined by domain experts. It is possible to use machine learning algorithms so that this estimation can be made better and updated in accordance with the data that has been generated over time. In this context, a machine learning process, an architecture designed to realize this process and the competencies that should be possessed by the development team are described.

**Keywords:** Software Architecture, Machine Learning Process, Mobile Advertising, Ad Mediation Software.

## 1 Giriş

Son yıllarda mobil cihazların kullanımı oldukça yaygınlaşmıştır. Bu cihazların kullanıcılar tarafından benimsenmesinin temel nedenlerinden biri bu cihazlar üzerinde çalışan uygulamaların sayısının artması, çeşitlenmesi ve kullanıcıların birçok gereksinimini karşılamalarına yardımcı olmalarıdır. Mobil uygulamaların önemli bir bölümünün gelir modeli reklam gösterimine dayanmaktadır. Mobil cihaz ve uygulama sayısındaki artış reklam sektörünün de mobil kanallara ağırlık vermesine neden olmuştur. Mobil uygulamalara reklam sağlayan çok sayıda reklam ağı ortaya çıkmıştır. Bu reklam ağlarının temel amacı reklam verenler ile reklam yayımlayanlar (bu durumda mobil uygulamalar) arasında bir köprü kurmaktır. Reklam ağlarının sayısının artmasıyla birlikte de çok sayıda reklam ağı ile reklam yayımlayanlar arasında köprü kurma görevini reklam aracı yazılımları üstlenmeye başlamıştır.

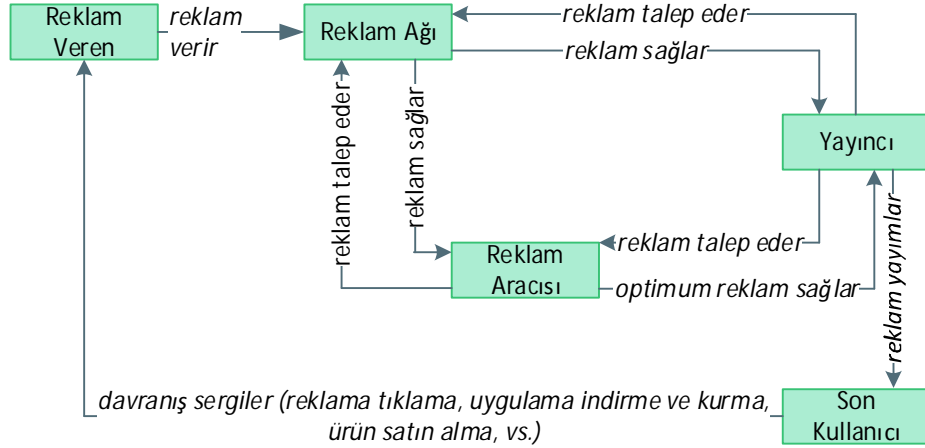
Bir reklam aracı yazılımı, bir uygulamaya reklam sağlarken, hangi reklam ağlarından hangi sırada reklam talep edeceğine dair karar verirken uygulamanın reklam gelirini (aynı zamanda da kendi reklam gelirini) azami seviyeye çıkarmayı amaçlamaktadır. Bunun için de reklam gösteriminden en fazla gelir sağlanabilecek reklam ağından başlayarak reklam talebinde bulunur. Mobil reklam sektöründe reklam gösterimi başına geliri ifade etmek için eBGBM (Etkin Bin Gösterim Başına Maliyet; İngilizcesi: eCPM – Effective Cost per Mille) kavramı kullanılmaktadır. Mevcut durumda Kokteyl şirketinin Reklam Aracı Yazılımında gerçekleşecek eBGBM değerleri alan uzmanları tarafından sezgiye dayalı olarak tahmin edilmektedir ve sistemde güncellenmektedir. eBGBM değerini etkileyebilecek değişkenler tespit edilerek, bu değişkenlerle eBGBM arasındaki örüntülerin makine öğrenmesi algoritmalarıyla tespit edilmesi mümkün görünmektedir. Bu çalışma kapsamında Kokteyl şirketinin mevcut Reklam Aracı Yazılımına makine öğrenmesi yetenekleri kazandırılması için yürütülen projenin planlama aşamasında gerçekleştirilen faaliyetler paylaşılmaktadır. Planlama aşamasında gerek reklam sektöründeki gerekse başka sektörlerdeki benzer çalışmalar [1 - 4] ve makine öğrenmesi yetenekleri bir sisteme kazandırılırken kullanılacak yazılım mühendisliği pratikleri [5, 6] incelenmiştir.

Bildirinin geri kalan bölümleri şu şekilde düzenlenmiştir: İkinci bölüm mobil reklam sektörü, şirketin reklam aracı yazılımı ve makine öğrenmesi hakkında genel bir bilgi vermektedir. Üçüncü bölümde makine öğrenmesi yeteneklerine sahip bir mimari tasarım anlatılmaktadır. Dördüncü bölümde ilgili çalışmalar özetlenmiştir. Son bölümde bir özet ve gelecek çalışmalar sunulmuştur.

## 2 Arka Plan

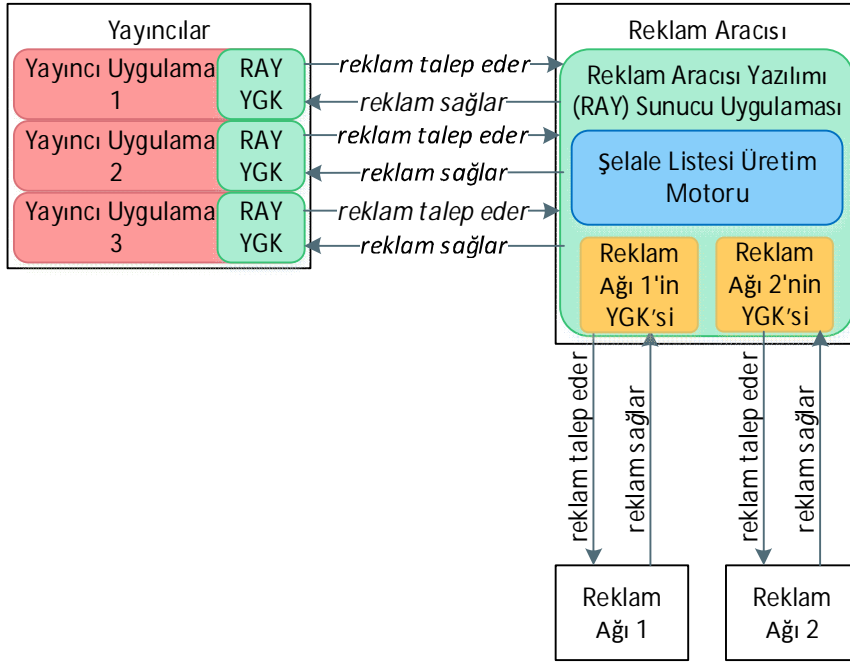
### 2.1 Mobil Reklam Sektörü ve Reklam Aracısı Yazılımı

Mobil reklam sektöründeki ana paydaşlar Şekil 1’de gösterilmektedir. **Reklam verenler**, **reklam ağlarına** ürünleri ve/veya hizmetleri hakkında reklam vermektedir. **Yayıncılar**, **reklam ağlarından** reklam talep etmekte ve reklam ağları yayıncıya o andaki en uygun reklamı sağlamaktadır. Uygun reklamın seçimi konusundaki karar birçok değişkene (yayıncının hangi uygulama olduğu, **son kullanıcı** hakkındaki bilgiler, uygulamanın çalıştığı cihaz gibi) bağlı olarak değişmektedir. **Yayıncıların reklam ağlarından** doğrudan reklam alması durumunda bazı zorluklar ortaya çıkmaktadır. Uygulamanın geliştirilmesi açısından ortaya çıkan bir zorluk, **yayıncı** uygulamanın reklam talep edeceği tüm **reklam ağlarıyla** ayrı ayrı entegre olma gereksinimidir. İş amaçları açısından bakıldığında bir **yayıncının** birçok **reklam ağından** en uygun reklamı (en fazla geliri getirecek reklam) alamaması durumunda gelir kaybı yaşayacaktır. Dolayısıyla **yayıncının** en uygun reklamı seçecek karar algoritmasını uygulamanın bir parçası olarak geliştirmesi gerekmektedir. **Reklam aracıları** birçok **reklam ağı** ile entegre olarak en uygun reklamı sağlama konusunda uzmanlaşmaktadır. Uygulama geliştirme açısından **yayıncı** uygulamanın sadece bir **reklam aracısı** ile entegre olarak birçok **reklam ağından** reklam alabilmektedir ve çok sayıda **reklam ağı** ile entegre olmanın getirdiği karmaşıklık **reklam aracısı** tarafından yönetilmektedir. Bunun yanında en uygun reklamı seçme konusunda uzmanlaşan **reklam aracıları** bu işi de **yayıncı** uygulamaların yerine yapmaktadır. Böylece **yayıncıların** reklam gelirlerini arttırmasına yardımcı olmaktadır. **Son kullanıcılar** ise **yayıncı** uygulamaları kullanarak reklamları görüntülemekte ve bu reklamlara karşı zaman zaman bir davranış (reklama tıklama, uygulama indirme ve kurma, ürün satın alma gibi) sergilemektedir. Bu tür davranışlar **yayıncılara** gelir sağlamaktadır.



Şekil 1. Mobil reklam sektöründeki ana paydaşlar.

Şekil 2’de reklam aracı yazılımının (yazılım mimarisi için bkz. [7]) yayıncılarla ve reklam ağlarıyla olan ilişkisi gösterilmektedir. Reklam aracı yazılımını oluşturan iki bileşen, şekilde yeşil renk ile gösterilmiştir. Reklam Aracı Yazılımının (RAY) sunucu uygulaması, reklam ağlarının yazılım geliştirme kitlerini (YGK) içermektedir. Bu YGK’ler (Şekil 2’deki Reklam Ağı 1’in ve 2’nin YGK’si) aracılığıyla sunucu uygulaması reklam ağlarından reklam talep etmektedir ve reklam almaktadır. Yayıncı uygulamalar ise RAY’nin YGK’sini uygulamalarının içine yerleştirerek RAY’den reklam talep etmektedir ve almaktadır. Daha önce anlatıldığı gibi, teknik açıdan RAY birçok reklam ağı ile entegre olmanın getirdiği karmaşıklığı yayıncı uygulamalara saydam hale getirmektedir. İş amaçları açısından ise RAY birçok reklam ağından sağladığı reklamlar arasından en uygun reklamı yayıncı uygulamaya göndererek reklam gelirlerinin artırılmasını sağlamaktadır.



Şekil 2. Reklam aracı yazılımının yayıncılar ve reklam ağları ile ilişkisi.

Şekil 2’de RAY’nin içerdiği bir bileşen olarak Şelale Listesi Üretim Motoru gösterilmektedir. Şelale listesi, yayıncının hangi reklam ağlarından hangi sırada reklam talebinde bulunması gerektiğini belirleyen bir öncelik listesidir (Bkz. Tablo 1). Şelale listesi, her bir yayıncı için belirli aralıklarla (örneğin günlük) çeşitli değişkenler (reklam formatı, reklam alanı, geçmiş eBGM değerleri gibi) göz önüne alınarak şelale listesi üretim motoru tarafından üretilir. Her yayıncı uygulamanın RAY YGK’si bu şelale listesini belirli aralıklarla sunucu uygulamasından indirerek reklam taleplerini bu güncellenmiş listeye göre yapar. Şelale listesindeki sıralama gelecekte gerçekleşecek eBGM değerlerini ne kadar iyi tahmin edebilirse reklam gelirleri o

kadar artacaktır. eBGBM reklam gösterimi başına geliri ifade etmektedir ve “(Gelir / Gösterim Sayısı) x 1000” formülü ile hesaplanmaktadır. Mevcut durumda şelale listesi oluşturmak için kullanılan girdiler (eBGBM tahminleri gibi) alan uzmanları tarafından yapılan gözlemler sonucunda sezgisel olarak belirlenmektedir ve güncellenmektedir. Mevcut bu çözüme alternatif olarak eBGBM değerini etkileyen parametrelerin tespit edilmesi ve makine öğrenmesi algoritmaları kullanılarak eBGBM tahmin edilmesi hedeflenmektedir.

**Tablo 1.** Örnek bir şelale listesi.

---

1. sıradaki reklam açısından reklam talep et
Eğer reklam alındıysa
Reklamı göster
Çık
2. sıradaki reklam açısından reklam talep et
Eğer reklam alındıysa
Reklamı göster
Çık
3. sıradaki reklam açısından reklam talep et
Eğer reklam alındıysa
Reklamı göster
Çık
...

---

## 2.2 Makine Öğrenmesi

Makine öğrenmesi, bilgisayar sistemlerini açık bir şekilde programlamadan, bilgisayar sistemlerine “öğrenme” yeteneğini (yani belirli bir görevi yerine getirirken gösterdiği performansı kademeli olarak geliştirme) kazandırmak için istatistiksel teknikleri kullanan bilgisayar biliminin bir alanıdır. Öğrenme süreci gözetimli ya da gözetimsiz olabilmektedir. Gözetimsiz (unsupervised) öğrenmede, açık bir geri bildirim sağlanmasa bile bilgisayar sistemi girdideki örüntüleri öğrenir [8]. En yaygın gözetimsiz öğrenme görevi kümelemedir, yani girdi örneklerindeki alt kümeleri tespit etmektir. Gözetimli (supervised) öğrenmede ise, bilgisayar sistemi bazı örnek girdi-çıkı çiftlerini gözlemler ve girdiden çıktıya eşlenen bir fonksiyonu öğrenir [8].

Veri analizi farklı amaçlar için, farklı sorulara cevap bulmak için yapılabilir. Bu konuda yaygın olarak kullanılan bir sınıflandırma veri analitiği türlerini dört ana başlıkta toplamaktadır [9]: (1) Tanımlayıcı (descriptive), (2) Teşhis koyucu (diagnostic), (3) Tahminleyici (predictive), (4) Öngörücü (prescriptive). Tanımlayıcı analiz, “ne oldu?”; teşhis koyucu analiz, “neden bu oldu?”; tahminleyici analiz, “ne olacak?” ve öngörücü analiz, “ne yapmalıyım?” sorusuna cevap arar.

Bu proje kapsamında tahminleyici analiz ile hangi parametrelerin eBGBM değerinin oluşumuna etkisi olduğu tespit edilmeye çalışılacaktır. Sonrasında mevcut veri kullanılarak gözetimli öğrenme yaklaşımıyla gelecekte oluşabilecek eBGBM

değerleri tahmin edilmeye çalışılacaktır. Bu görevleri yerine getirebilmek için yapılan mimari tasarım ve literatürde dayandığı temeller bir sonraki bölümde anlatılmaktadır.

### 3 Makine Öğrenmesi Yeteneklerine Sahip Bir Mimari Tasarımı

[10], bir bilgi teknolojileri organizasyonunun sistemlerine makine öğrenmesi yetenekleri kazandırabilmesi için üç ana alanda planlama yapmasını önermektedir: (1) süreç, (2) mimari ve (3) yetkinlik. Buna göre Kokteyl şirketi çatısı altında yapılan çalışmalar üç başlık altında incelenecektir:

1. Bir makine öğrenmesi sürecinin tasarımı
2. Bir makine öğrenmesi mimarisinin tasarımı
3. Yukarıdaki iki maddedeki etkinliklerin gerçekleştirilebilmesi ve sürdürülebilir kılınması için gerekli yetkinliklerin kazanılması

#### 3.1 Makine Öğrenmesi Süreci

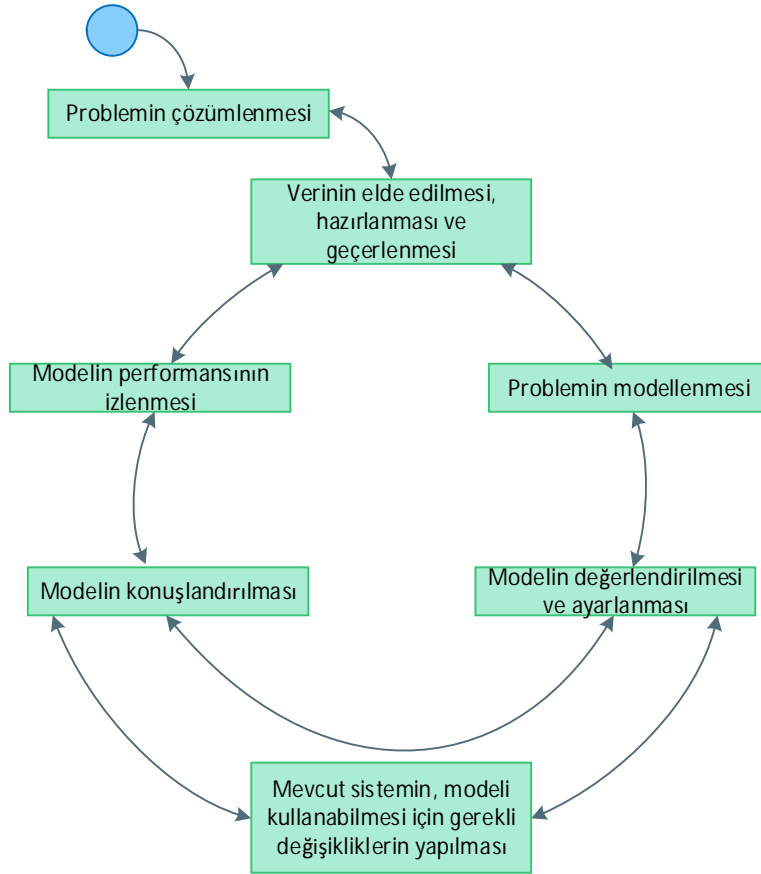
Bir sistemin makine öğrenmesi yeteneklerini kazanabilmesi için veri ve makine öğrenmesi (yaklaşımlar, teknikler, algoritmalar gibi) ile bağlantılı bazı işlemlerin yapılması gerekmektedir. Bu işlemler için literatürde [10 – 13] önerilen adımlar yedi başlık altında toplanmıştır ve aşağıdaki gibi özetlenmiştir (5 numaralı adım bu proje kapsamında elde edilen deneyim sonucunda eklenmiştir):

1. **Problemin çözümlenmesi:** Çözülmesi gereken problem kavranarak makine öğrenmesi açısından değerlendirilir. Makine öğrenmesi açısından yapılacak değerlendirme sonucunda eldeki problemin nasıl bir çözümlenme modeli gerektirdiğine (tanımlayıcı, teşhis koyucu, tahminleyici, öngörücü gibi), gözetimli ya da gözetimsiz öğrenme algoritma gruplarından hangisinin uygun olduğuna karar verilir.
2. **Verinin elde edilmesi, hazırlanması ve geçerlenmesi:** Problemin çözümü için kullanılacak verinin kaynakları belirlenir, bu verinin neleri temsil ettiği anlaşılır (verinin anlamı), veri işlenebilecek şekilde temizlenir ve düzenlenir, elde edilen nihai verinin problemin çözümünde kullanılabileceği geçerlenir.
3. **Problemin modellenmesi:** Makine öğrenmesi için kullanılacak algoritma(lar) belirlenir ve bu modeller gerçekleştirilir. Bu gerçekleştirim için hazır kod kütüphaneleri aynen ya da değiştirilerek kullanılabilir.
4. **Modelin değerlendirilmesi ve ayarlanması:** Modelin sonuçları uygun ölçülere göre değerlendirilir; bazı durumlarda birden fazla algoritma kullanılarak deneyler yapılır ve bu algoritmalar arasından en uygunu ya da bir kombinasyonu seçilir. Algoritmaların çeşitli parametrelerini (örneğin karar ağaçlarında ağaç derinliğinin ne kadar olacağı gibi) ayarlamak için eldeki veri kullanılarak çeşitli deneyler yapılacaktır.
5. **Mevcut sistemin modeli kullanabilmesi için gerekli değişikliklerin yapılması:** Makine öğrenmesi yeteneklerinin mevcut bir sisteme kazandırılması planlanıyorsa

bu durumda mevcut sistemde gerekli deęişiklerin yapılması ve test edilmesi gerekmektedir.

6. **Modelin konuşlandırılması:** Makine öğrenmesi modelinin gerçekleşmiş hali kullanılmak üzere bir altyapıya (bir sunucu, birçok sunucudan oluşan veri merkezi, bulut gibi) konuşlandırılır. Bu model başka uygulamalarla iletişim kurarak görevini yerine getirebilir.
7. **Modelin performansının izlenmesi:** Makine öğrenmesi modelinin görevini ne kadar iyi yerine getirdiđi oluşan yeni veriler izlenerek gözlemlenir. Düzenli ya da düzensiz aralıklarla bu model güncellenerek modelin başarımının artırılması sağlanabilir.

Makine öğrenmesi sürecinde aşamalar arasında genellikle sık geçişler olabilmektedir ve süreç sıralı bir düzende ilerlememektedir. Bundan dolayı Şekil 3'te gösterildiđi gibi yinelemeli ve artırımlı bir süreç takip edilmektedir (benzer bir süreç önerisi için [14] çalışmasındaki Şekil 6'ya bakınız).



Şekil 3. Yinelemeli ve artırımlı makine öğrenmesi süreci.

Yukarıda belirtilen süreç kullanılarak bu proje kapsamında her adımda yapılması planlananlar aşağıdaki gibidir:

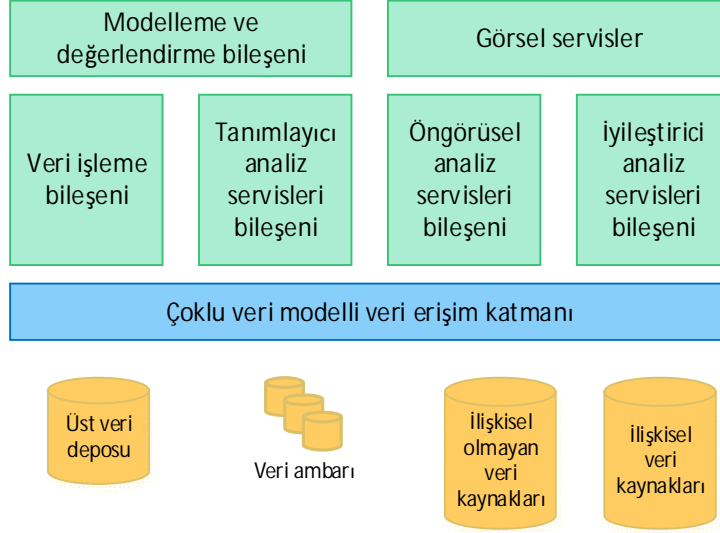
1. Bu çalışma kapsamında çözülmesi hedeflenen problem eBGBM değerinin tahmin edilmesidir. Bu tahmin etme işi için geçmişte oluşan eBGBM değerleri, bu değerlerin oluşmasını etkilediği düşünülen özelliklerin (feature) değerleri kullanılarak gözetimli öğrenme yaklaşımı kullanılacaktır.
2. Bunun için çeşitli kaynaklarda saklanan veriler birleştirilecek, temizlenecek ve makine öğrenmesi algoritmalarının kullanımı için hazır hale getirilecektir.
3. Sonrasında bu problemin çözümü için uygun olabilecek algoritmalar gerçekleştirilecektir. Literatürdeki bulgulara göre bu problem için çoklu bağlanım (multiple regression) modelinin uygun olabileceği değerlendirilmektedir [15].
4. Gerçekleştirilen modeller geçmiş veri kullanılarak eğitilecek ve test edilecektir.
5. Mevcut sistemde bulunan Şelale Listesi Üretim Motoru bileşeni makine öğrenmesi algoritmasıyla oluşturulmuş tahminleri servis edecektir. Bunun için mevcut sistemde gerekli değişiklikler yapılacaktır.
6. Seçilen model canlı ortama konuşlandırılarak sınırlı sayıda kullanıcı için kullanıma açılacaktır.
7. Sonrasında modelin başarımı izlenerek modelde iyileştirmeler yapılarak aşamalı olarak tüm kullanıcılar için kullanıma açılacaktır. Sonrasında modelin başarımı sürekli olarak izlenerek belirli aralıklarla modelde iyileştirmeler yapılmaya devam edilecektir.

Proje kapsamında bu süreci desteklemek için tasarlanan mimari bir sonraki bölümde anlatılmaktadır.

### 3.2 Makine Öğrenmesi Yeteneklerine Sahip Bir Mimari

Makine öğrenmesi yeteneklerinin geliştirildiği sistemlerde karmaşıklığın büyük bir kısmı verinin elde edilmesi, hazırlanması ve geçerlenmesi, sonrasında ise makine öğrenmesi modellerinin oluşturularak ayarlanması aşamalarında ortaya çıkmaktadır [13, 16]. Sculley ve arkadaşları makine öğrenmesi sistemlerinin çok küçük bir bölümünün makine öğrenmesi kodu olduğunu; bu kodu çevreleyen sistemin diğer bileşenlerinin çok daha geniş ve karmaşık olduğunu belirtmiştir [16]. Bu bağlamda böyle bir sistemin mimarisini oluşturması beklenen bileşenler iki ana grupta sınıflandırılabilir: (1) Veriyi makine öğrenmesi algoritmaları için hazırlayan bileşenler (veri toplama, veri geçirme, özellik belirleme, konfigürasyon, analiz araçları) ve (2) bu büyük ölçekli verinin işlenmesi için gerekli altyapının sağlanması (hizmet altyapısı, makine kaynak yönetimi, görev yönetim araçları, izleme). Literatürdeki bulgular ve bu projedeki gereksinimler doğrultusunda makine öğrenmesi algoritmalarının çalıştırılması için geliştirilecek mimari Şekil 4'teki gibi tasarlanmıştır.

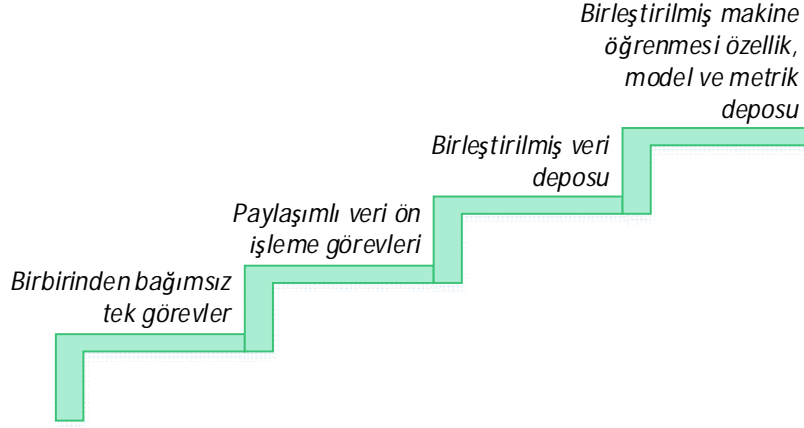




**Şekil 4.** Makine öğrenmesi yeteneklerini mevcut sisteme kazandırmak için gerekli bileşenler.

Makine öğrenmesi algoritmalarına gözetimli öğrenme için girdi olarak verilecek veri *ilişkisel ve ilişkisel olmayan kaynaklarda* saklanmaktadır. Bu kaynaklardaki veriye birleştirilmiş bir şekilde ulaşabilmek için bir *çoklu veri modelli veri erişim katmanı* [17] oluşturulacaktır. Bu katman aynı zamanda birleştirilmiş verinin saklandığı *veri ambarına* erişim için de kullanılacaktır. Bu aktarım işlemleri *veri işleme bileşeninin* sunduğu hizmetler aracılığıyla yapılacaktır. Bu bileşenin görevi veri ambarı geliştirme projelerinde görülen verinin çıkarılması, dönüştürülmesi ve yüklenmesi (ETL: Extract Transform Load) görevlerini yerine getirmektir. Bu bileşen geliştirilirken tüm veri kaynaklarına bütüncül bir bakış açısıyla yaklaşmak ve verinin aktarımı için geliştirilen prosedürlerin iyi tasarlanması gerekmektedir. Aksi takdirde oluşabilecek karmaşık aktarım prosedürlerinin (“pipeline jungles” [16]) test edilmesi, hataların ayıklanması oldukça zor ve maliyetli olacaktır [18].

Şekil 5’te veri hazırlama görevlerinin ve makine öğrenmesi modellerinin tasarım açısından olası gelişim aşamaları gösterilmektedir [6]. Buna göre birbirinden bağımsız veri hazırlama görevlerinden, verinin birleştirilmiş şekilde bir veri ambarında saklandığı ve makine öğrenmesi ile ilgili özelliklerin, modellerin, metriklerin yine birleştirilmiş bir şekilde depolandığı bir yapıya doğru bir gelişmenin olması, bu yapıların bakımının daha kolay yapılması, hataların daha kolay yakalanması ve düzeltilmesi için gereklidir. Bu proje kapsamında da bütüncül bir bakışla bir veri ambarı oluşturulması ve makine öğrenmesi ile ilgili iş ürünlerinin de birleştirilmiş bir şekilde saklanması hedeflenmektedir. Bu hedef için önemli bileşenlerden biri de *üst veri deposu*. Üst veri deposu, veri kaynaklarındaki veriler hakkında bilgiler (verinin ne anlama geldiği, veri tipi, gibi) sunarak gerek geliştiricilerin gerek bu verileri makine öğrenmesi görevleri için kullanacak kişilerin veriyi anlamasına yardımcı olacaktır.



**Şekil 5.** Veri hazırlama görevlerinin ve makine öğrenmesi modellerinin evrimi.

**Tanımlayıcı analiz servisleri** bileşeni, makine öğrenmesi modeli oluşturulmak için gerekli olan mevcut veriyi anlamaya yönelik çözümlenmeleri sunan hizmetleri sağlayacaktır. Bu çözümlenmeler sonucunda eBGBM değerini etkileyen özellikler belirlenecektir. **Tahminleyici analiz servisleri** bileşeni eBGBM değerinin tahmin edilmesine yönelik hizmetleri sağlayacaktır. Bunun için gözetimli öğrenme yaklaşımı ve çoklu doğrusal bağlanım modeli kullanılacaktır. **İyileştirici analiz servisleri** bileşeni, konuşlandırılan modelin performansını izlemek için gerekli hizmetleri sunacaktır. Zamana bağlı olarak değişen eBGBM değerinin zaman serisi analizi teknikleri kullanılarak daha iyi tahmin edilmesine yönelik iyileştirme çalışmaları yapılacaktır. Farklı ortam verilerine sahip makine öğrenmesi modelleri arasında kıyaslama yapılabilecek ve ortam verisinin değişimine göre makine öğrenmesi modelinin otomatik olarak evrilmesi sağlanacaktır.

**Görsel servisler**, kullanıcıların amaçlarına uygun biçimde veriyi görselleştirerek sunacaktır. **Modelleme ve değerlendirme bileşeni** ise tanımlayıcı, tahminleyici ve iyileştirici analiz servislerini kullanarak çeşitli modeller oluşturulması ve bunların değerlendirilmesi için veri analiz uzmanlarına bir arayüz sunacaktır. Bunun için Jupyter Notebook (<http://jupyter.org/>) gibi araçlar kullanılabilir.

Tasarlanan mimarideki bileşenleri geliştirmek ve sonrasında kullanmak için çeşitli yetkinlikler gerekmektedir. Bu yetkinlikler çeşitli roller altında sınıflandırılarak bir sonraki bölümde anlatılmaktadır.

### 3.3 Makine Öğrenmesi Yeteneklerinin Geliştirilmesi İçin Gerekli Yetkinlikler

Makine öğrenmesi yeteneklerinin geliştirilmesi için gerekli yetkinlikler genellikle veri analitiği uzmanı, makine öğrenmesi uzmanı, veri bilimci gibi unvanlar altında toplanmaktadır [13, 19, 20, 21]. Bu roller altında toplanan sorumlulukların ve dolayısıyla sahip olunması gereken yetkinliklerin farklılaştığı görülmektedir [19, 20]. Bu proje kapsamında, makine öğrenmesi sürecindeki 3., 4. ve 7. adımlardaki görevleri yerine getirebilecek bir **veri analitiği uzmanı** görev alacaktır. Mevcut veriyi anlamak

için üst veri deposunu, makine öğrenmesi modelleri için girdileri elde etmek için veri ambarını kullanacaktır. Tanımlayıcı, tahminleyici ve iyileştirici analiz servislerinin makine öğrenmesi ile ilgili bölümlerini geliştirecektir. Modelleri çalıştırmak ve sonuçları görmek için modelleme ve değerlendirme bileşenini kullanacaktır. Bu görevleri yerine getirirken görsel servisleri kullanarak elde edilen analiz sonuçlarını ***konu alanı uzmanları*** ile değerlendirecektir.

Çoklu veri erişim katmanı ve veri işleme bileşenleri ***yazılım mühendisleri*** tarafından geliştirilecektir. Bu bileşenler geliştirilirken konu alanı uzmanları gereksinimleri belirleyecek ve geliştirilen bileşenleri test edecektir.

Tüm bu bileşenlerin barındırılması için gerekli altyapıyı ise ***altyapı mühendisi*** oluşturacak ve işletilmesini sağlayacaktır.

Veri analitiği uzmanı, temel istatistik ve matematik bilgisinin [22] yanında, makine öğrenmesi kütüphaneleri ve model geliştirmek için kullanılacak programlama dilleri (Python, R gibi) hakkında bilgi sahibi olmalıdır. Konu alanı uzmanları, mobil reklamcılık sektörü ve bu sektörde oluşan verinin ve bilginin yorumlanması konusunda deneyimli olmalıdır. Yazılım mühendisleri, sunucu tarafındaki ve mobil cihazlar üzerinde çalışacak bileşenleri geliştirecek yetkinliklere sahip olmalıdır. Bunun yanında gerek yazılım mühendislerinin gerekse altyapı mühendisinin büyük hacimli verinin işlenebileceği altyapıyı tasarlayıp hayata geçirebilecek bilgi birikimine ve deneyimine sahip olmaları gerekmektedir.

Özet olarak, bu proje kapsamında farklı yetkinliklere sahip olmayı gerektiren dört ana rol bulunacaktır. Veri analitiği uzmanı, yazılım mühendisleri, konu alanı uzmanları ve altyapı mühendisi birlikte çalışarak projenin başarılı bir şekilde tamamlanmasını sağlayacaktır.

## 4 İlgili Çalışmalar

NIST'in yayımladığı "büyük veri birlikte çalışabilirlik çerçevesi"nde [23] önerilen referans mimaride verinin kaynağından alınıp kullanılacağı noktaya ulaştırılıncaya kadar gerekebilecek işlevler beş başlık altında toplanmıştır. Toplama (collection), verinin elde edilmesine; hazırlama (preparation), verinin uygun formata dönüştürülmesine; analitik (analytics), veriden bilgi ve içgörü elde edilmesine; görselleştirme (visualization), bilginin uygun şekillerde sunulmasına; erişim (access), veriyi kullanacak paydaşlarla iletişimin sağlanmasına dair görevleri kapsamaktadır. NIST'in referans mimarisindeki bu başlıklarda ele alınması önerilen görevler Şekil 4'teki mimaride kapsamaktadır.

Elde edilen makine öğrenmesi modellerinin konuşlandırılması ve başka paydaşlar tarafından kullanılabilir duruma getirilmesi çeşitli çözümler sunulmaktadır. Örneğin, Algorithmia gibi platformlar makine öğrenmesi modellerinin konuşlandırılması için gerekli altyapıyı sağlamaktadır [24]. Bu platform, modellerin kolayca konuşlandırılması, API ile başka uygulamaların kullanımına sunulması gibi hizmetleri sağlamaktadır [24]. Bu çalışma kapsamında makine öğrenmesi modellerini servis olarak sunan bir platformun da geliştirilmesi hedeflenmektedir.

[5] çalışmasında makine öğrenmesi modellerinin Uber şirketi bünyesinde servis olarak nasıl sunulduğu anlatılmaktadır. Verinin, modellerin girdisi olacak şekilde hazırlanma aşaması Şekil 3'teki süreçle örtüşmektedir. Aynı zamanda makine öğrenmesi modellerinin servis olarak sunulması da bu çalışma kapsamında verilen tasarım kararlarından biridir.

## 5 Sonuçlar ve Gelecek Çalışmalar

Mobil cihazların ve uygulamaların yaygınlaşması ve bu uygulamaların önemli bir bölümünün gelirlerini reklam gösterimiyle elde etmeye başlamasıyla birlikte mobil reklam sektörünün hacmi büyümüş ve sektördeki paydaşların sayısı artmıştır. Bir uygulamada, bir kullanıcı için hangi reklamların gösterilmesinin nasıl bir gelir getireceğini tahmin etmek, mobil reklam sektörü için önemli bir problemdir. Bu proje kapsamında, geçmişte oluşan veriden yola çıkarak gözetimli öğrenme yaklaşımıyla, elde edilebilecek gelirin tahmin edilmesi ve hangi reklamın gösterileceği kararının bu tahmine göre verilmesi amaçlanmaktadır. Mevcut durumda konu alanı uzmanları tarafından sezgisel olarak yerine getirilen bu tahmin görevinin, makine öğrenmesi modelleriyle daha iyi tahmin edilmesi ve bu modellerin sürekli iyileştirilerek tahmin başarısının zaman içinde daha da artırılması beklenmektedir. Bu görevleri yerine getirebilmek için mevcut yazılım sisteminde makine öğrenmesi ile ilgili bileşenlerin geliştirilmesi gerekmektedir. Bu çalışmada bu amaca hizmet edecek bir mimari tasarım sunulmuştur.

Gelecek çalışmalar kapsamında bu mimari tasarım doğrultusunda geliştirmeler yapılacaktır. Bu geliştirmeler yapılırken Kokteyl firmasının yazılım geliştirme süreci (mevcut süreç için bkz. [25]) makine öğrenmesi yetenekleri geliştirilebilecek şekilde düzenlenecektir.

## Teşekkür

Bu çalışma TÜBİTAK'ın desteğiyle 3171053 numaralı proje kapsamında yapılmıştır.

## Kaynakça

1. Llach, E.: System for automatically selling and purchasing highly targeted and dynamic advertising impressions using a mixture of price metrics. U.S. Patent Application No. 10/767,050 (2004).
2. O'Kelley, B.: Method and system for pricing electronic advertisements. U.S. Patent Application No. 11/006,121 (2006).
3. Paunekar, A., Hochberg, M.: Dynamic pricing for content presentations. U.S. Patent No. 8,706,547. 22 (2014).
4. Chapelle, O., Manavoglu, E., Rosales, R.: Simple and scalable response prediction for display advertising. ACM Transactions on Intelligent Systems and Technology (TIST) 5(4), Article 61 (2015).

5. Li, L. E., Chen, E., Hermann, J., Zhang, P., Wang, L.: Scaling machine learning as a service. In: International Conference on Predictive Applications and APIs, pp. 14-29. (2017).
6. Taifi, M.: Lessons Learned from Building Scalable Machine Learning Pipelines, In: PAPIs Europe 2018. <https://techblog.appnexus.com/lessons-learned-from-building-scalable-machine-learning-pipelines-822acb3412ad>, son erişim tarihi: 30 Eylül 2018.
7. Kırkağaçlıoğlu, Ö., Giray, G., Ersin, A., Çatıkkaş, C., Koçer, S., Şeremet, T., Ünalır, M. O.: Bir Reklam Aracısı Yazılımının Mimari Evrimi. In: 12. Ulusal Yazılım Mühendisliği Sempozyumu. İstanbul (2018).
8. Russell, S. J., Norvig, P.: Artificial intelligence: a modern approach. Pearson Education Limited, Malaysia (2016).
9. Gartner Report: Forecast: Enterprise software markets. Worldwide, 2011-2016, 4Q12 Update. (2012). <http://www.gartner.com/resId=2054422>, son erişim tarihi 30 Eylül 2018
10. Sapp, C. E.: Preparing and Architecting for Machine Learning. Gartner. ID: G00317328. (2017).
11. Azevedo, A., Filipe Santos, M.: KDD, semma and CRISP-DM: A parallel overview. In: IADIS European Conference on Data Mining 2008, pp. 182-185. Amsterdam, The Netherlands (2008).
12. Labrinidis, A., Jagadish, H. V.: Challenges and opportunities with big data. In: Proc. VLDB Endow, pp. 2032-2033. (2012). DOI=<http://dx.doi.org/10.14778/2367502.2367572>.
13. Pruss, L.: Infrastructure 3.0: Building blocks for the AI revolution. (2017). <https://venturebeat.com/2017/11/28/infrastructure-3-0-building-blocks-for-the-ai-revolution/>, son erişim tarihi: 24 Eylül 2018.
14. Schleier-Smith, J.: An architecture for Agile machine learning in real-time applications. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2059-2068. ACM (2015).
15. Provost, F., Fawcett, T.: Data Science for Business. O'Reilly Media, Inc. (2013). ISBN: 9781449374273
16. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo J., Dennison, D.: Hidden technical debt in machine learning systems. In: Advances in neural information processing systems, pp. 2503-2511. (2015).
17. Diker, N., Giray, G., Ünalır, M. O.: Servis Tabanlı Bir Melez Veri Erişim Mimarisi Önerisi. In: 10. Ulusal Yazılım Mühendisliği Sempozyumu. Çanakkale, (2016).
18. Ananthanarayanan, R., Basker, V., Das, S., Gupta, A., Jiang, H., Qiu, T., Reznichenko, A., Ryabkov, D., Singh, M., Venkataraman, S.: Photon: Fault-tolerant and scalable joining of continuous data streams. In: SIGMOD'13: Proceedings of the 2013 international conference on Management of data, pp. 577-588. New York, NY, USA (2013).
19. Kim, M., Zimmermann, T., DeLine, R., Begel, A.: The emerging role of data scientists on software development teams. In: Proceedings of the 38th International Conference on Software Engineering, pp. 96-107. ACM (2016).
20. Kim, M., Zimmermann, T., DeLine, R., Begel, A.: Data Scientists in Software Teams: State of the Art and Challenges. IEEE Transactions on Software Engineering. (2017). doi: 10.1109/TSE.2017.2754374
21. Simard, P. Y., Amershi, S., Chickering, D. M., Pelton, A. E., Ghorashi, S., Meek, C., Ramos, G., Suh, J., Verwey, J., Wang, M., Wernsing, J.: Machine teaching: A new paradigm for building machine learning systems. (2017). arXiv preprint arXiv:1707.06742.
22. Waller, M. A., Fawcett, S. E.: Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics 34(2), pp. 77-84. (2013).

23. NIST Big Data Public Working Group, Reference Architecture Subgroup: NIST Big Data Interoperability Framework: Volume 6, Reference Architecture. NIST Special Publication 1500-6r1, Version 2, (2018). <https://doi.org/10.6028/NIST.SP.1500-6r1>
24. Algorithmia: Deploying Machine Learning at Scale with Serverless Microservices. Whitepaper, (2018).
25. Giray, G., Yilmaz, M., O'Connor, R. V., Clarke, P. M.: The Impact of Situational Context on Software Process: A Case Study of a Very Small-Sized Company in the Online Advertising Domain. In: European Conference on Software Process Improvement, pp. 28-39. Springer, Cham, (2018).