

# IS HAYATI VERI MADENCILIGI ILE ISTATISTIK UYGULAMALARINI

## YENIDEN KESFEDİYOR.

Doç. Dr. Sule ÖZMEN  
Marmara Üniversitesi İ.İ.B.F.  
İngilizce İletme Bölümü

### GİRİS

Günümüz dünyasında işletmeler başarılı olabilmek için yeni ekonominin ve yeni iş modellerinin kurallarına uygun hareket etmek durumundadırlar. Bilgisayar ve iletişim teknolojilerindeki gelişmeler çok daha fazla verinin hızlı toplanmasına, depolanmasına, işlenmesine ve bilgiye dönüştürülüp yeniden istenilen noktalara iletilmesine olanak sağlamıştır. Yeni teknolojilerden önemli ölçüde etkilenen ve dijital ekonomi olarak da adlandırılan bu yeni ekonominin kuralları gereği “bilgi” ve “zaman” boyutlarının önemi çok daha artmış ve dolayısıyla işletmelerin karar verme süreçlerinde doğru ve anlamlı bilgiye dayalı hızlı karar alma gereği her zamankinden daha fazla ön plana çıkmıştır.

Bunun için de müşteriler basta olmak üzere işletme ile ilgili taraflar ve süreçler hakkında bilgi sahibi olmak, bilgi sahibi olmanın gereği olarak da veri toplamak gerekir. Bu amaçla hazırlanan veri ambarlarında tutulan veriler çeşitli istatistiksel yöntemlerle analiz edilerek anlamlı, kullanılabilir ve karar verme sürecinin etkinliğini ve verimliliğini artırıcı bilgilere dönüştürülür. Müşteriler hakkında tutulan veriler sadece demografik özelliklerle ilgili verilerden ibaret değildir. Müşterinin firmayla yaptığı işlemlerle ilgili kayıtlar tutulmaktadır. Örneğin işlem saati, aldığı ürün veya hizmet, bunların tutarı gibi veriler müşteri bazında veya ürün bazında veriler tutulmakta ve bu verilerden müşterinin satın alma tutum ve davranışları hakkında bilgiler üretilebilmektedir. Firma için karlı bir müşteri olup olmadığı ya da ne derece karlı bir müşteri olduğu, sadık bir müşteri olup olmadığı yönünde bilgiler elde edilebilmekte, bunun sonucunda müşterilerin hangi yaşam evresinde oldukları tesbit edilip ve ona göre pazarlama stratejileri geliştirilebilmektedir.

Tutulması gereken verilerin miktarının ve çeşitliliğinin artması, analizlerin daha hızlı yapılması gereği ve sonuçta anlamlı ve eyleme yönelik bilgiler ortaya çıkarılması yeni ekonominin ve değişen piyasalardaki yoğun rekabet ortamının bir gereğidir. Rekabetin yoğunlaşması; değişim ve uyum sürecinin gerektirdiği hızı yakalayabilmenin, müşteri odaklı olmanın ve verimliliğin önemini her zamankinden daha çok arttırmıştır. Bunları mümkün kılan ise, yine yeni ekonomide sunulan teknik olanaklardır. Milyonlarca bitlik verilerin depolandığı veri ambarlarının tutulduğu donanımlar ve bunlardan çok hızlı bir şekilde anlamlı bilgiler elde edebilmek için kullanılacak yazılımlar yeni teknolojinin sunduğu olanaklardır.

### VERI MADENCILIGI

Büyük miktarlarda ve oldukça hızlı toplanan verilerin çeşitli analizler sonucunda anlamlı bilgilere dönüştürülmesi noktasında “veri madenciliği” süreci devreye girmektedir. Veri madenciliği tanımları incelendiğinde, bu tanımlardan ortak olan

unsurlardan ilki “çok fazla” miktarlarda verinin veri ambarlarında tutulması ikincisi ise bu verilerden “anlamli” bilgiler elde edilmesidir.

*Veri madenciligi; veri ambarlarındaki tutulan çok çeşitli verilere dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarmak, bunları karar vermek ve eylem planını gerçekleştirmek için kullanma sürecidir.* (Swift, 2001) Bu noktada Kendi basına bir çözüm değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli olan bilgileri sağlamaya yarayan bir araçtır.

Veri madenciligiyle ilgili yazılım ürünleri ve uygulamalarında görülmektedir ki veri madenciligi esasen istatistigin kullandığı bir tekniktir. Dolayısıyla işletmelerde önemli bir kullanım alanı bulması çok doğaldır. Özellikle çok müşterisi olan firmalar, bankalar tarafından büyük bir hızla benimsenmektedir. İşte bu nedendir ki işletmeler istatistik uygulamalarını özellikle pazarlama amaçlarına yönelik uygulamalarını adeta yeniden keşfetmektedirler.

Veri madenciliginde vurgulanan unsurlar istatistigin tanımı içinde zaten yer almaktadır. İstatistik verilerin toplanması, sınıflandırılması, özetlenmesi, grafik ve tablolarla sunulması, analiz edilerek ana kütle hakkında anlamli bilgiler elde edilmesi ve yorumlar yapılmasıdır. Veri madenciliginde ulaşmak istenen amaç aslında istatistik biliminin amacı ile aynı doğrultudadır. Verilerden bilgiyi keşfetmek. Zaten veri madenciliginde kullanılan temel aracın istatistiksel yöntemler olduğu birçok tanımda ve uygulamada vurgulanmaktadır. Her ikisinde de temel olan öğeler veri ve bilgidir. Bu nedenle birbiriyle oldukça örtüşen konulardır.

Veri madenciligi 1990’larda ortaya çıkan ve dünya da yeni yaygınlaşan bir kavram. Veri madenciligi uygulamalarında kullanılan ve büyük miktarlarda verilerin tutulduğu veri ambarı projeleri de benzer şekilde son yıllarda önem kazanan projelerdir. Bilisim sektöründe veri ambarı ve veri madenciligiyle ilgili gerek donanım gerek yazılım ürünleri üstünde çalışan firmalar son yıllarda artış göstermektedir. Bunların arasında istatistikçilerin hiç yabancı olmadıkları SPSS firması da bulunmaktadır. SPSS veri madenciligi uygulamaları için ileri düzey istatistik bilgisi gerektirmeyen, kullanıcıya uygun arayüzlerle sunulan Clementine, AnswerTree gibi ürünler çıkartmıştır.

Veri ambarlarında verileri tutmak ve veri madenciligi uygulamaları kullanmak, geleneksel işleri elektronik ortama taşıma ve elektronik işletme olmanın da gereklerinden ve ileri aşamalarından biridir. Bu aşama internet bağlantısına sahip olma, elektronik posta sahibi olma, web sitesine sahip olarak sanal dünyada var olma, hatta İnternet üzerinden alım ve satım işleminin gerçekleştirilmekten daha ileri bir aşamadır. Tüm bunları gerçekleştirmenin yanısıra elektronik işletme olabilmek için satış ve kar artırıcı amaçla yönelik uygulamalara yönelmek gerekir. Bu da, ancak müşteri odaklı çalışmakla, müşterilerle ilgili gerekli verileri tutarak anlamli bilgiler elde etmekle dolayısıyla çok sayıda müşteriyle birebir ilişki kuracak teknolojik ve yönetsel altyapının kurulmasıyla mümkün olabilmektedir.

Veri Madenciligi istatistik biliminin teknolojiyle bütünleşmesi sonucu oluşturulan bir araçtır. Uygulamada istatistigin yeniden keşfedilmesine olanak sağlayan hiç kuskusuz yeni teknolojilerin sunduğu fırsatlardır. Teknik olarak eskisinden çok daha fazla miktarlarda verilerin toplanması, depolanması ve hızlı bir şekilde işlenmesi mümkündür. Ancak veri madenciliginde unutulmaması gereken önemli kavramlardan biri bunun paket programdan

ibaret olmadıdır. Her ne kadar kullanıcı için kolaylık sağlayıcı arayüzler hazırlansa da en azından temel istatistik bilgisine ihtiyaç vardır. Bu ihtiyaç elbette bu programı sağlıklı kullanmak ve isabetli kararlar alabilmek açısından önemlidir. Örneğin iki değişken arasında korelasyon bulmak istediğiniz zaman paket programın buna itiraz etme sansi olmayabilir, üstelik korelasyon derecesinin anlamlı olduğu sonucu elde edilebilir, ancak sonuç bu iki değişken arasında korelasyon katsayısının gerçekten istenen bilgi olduğu anlamına gelmez.

## VERİ MADENCİLİĞİNİN İŞLETMELERDE KULLANIM AMACI

İstatistğin amacı nasıl ana kütle hakkında anlamlı bilgiler elde etmek ve yorum yaparsa veri madenciliğinin amacı da anlamlı bilgiler elde etmek ve bunu eyleme dönüştürecek kararlar için kullanmaktır. İlgilendiği ana kütle örneğin mevcut veya potansiyel müşteriler olabilir. Müşterilerin profillerini, satın alma eğilimlerini, bir ürünü veya hizmeti kabul etme veya etmeme ihtimallerini tahmin etme veri madenciliğinde hedeflenen amaçlar arasındadır. Bu tahminlerin ise ana hedef olan strateji belirlemek için çeşitli kararlar vermek için kullanılır.

Ürün ve hizmet sektöründe müşterilerle ilgili veri madenciliği uygulama amaçlarına ilişkin çok çeşitli örnekler vermek mümkün. En karlı pazar bölümlerini, en karlı müşterileri, yeni bir promosyon kampanyasında müşteriye sunulan ürün veya hizmetin kabul edilme oranlarını saptamak pazarlamada veri madenciliği uygulamasında önemli amaçlardır. Veri madenciliği uygulamalarından elde edilecek faydalara ilişkin bazı örnekler aşağıda sıralanmıştır.

- Bir işletme kendi müşterisiyken rakibine giden müşterilerle ilgili analizler yaparak rakiplerini tercih eden müşterilerinin özelliklerini elde edebilir ve bundan yola çıkarak gelecek dönemlerde kaybetme olasılığı olan müşterilerin kimler olabileceği yolunda tahminlerde bulunarak onları kaybetmemek, kaybettiklerini geri kazanmak için strateji geliştirebilir.
- Ürün veya hizmette hangi özelliklerin ne derecede müşteri memnuniyetini etkilediği, hangi özelliklerinden dolayı müşterileri bunları tercih ettiği ortaya çıkarılabilir.
- Müşterilerin kredi riskleri hesaplanarak hangi müşterilerin kredi riskinin yüksek olduğu, hangi müşterilerin geri ödemesini zamanında yapamayabileceği kestirilebilir. Kredi kartı ödemelerini aksatan, gecikmeli olarak yapan veya hiç yapmayanların özelliklerinden yola çıkılarak bundan sonra aynı duruma düşebilecek muhtemel kişiler saptanabilir.
- En karlı mevcut müşteriler saptanarak, potansiyel müşteriler arasından en karlı olabilecekler belirlenebilir. Karlı müşteriler tesbit edilerek onlara özel kampanyalar uygulanabilir. En masraflı müşteriler daha masrafsız müşteri haline dönüştürülebilir. Örneğin en çok bankacılık işlemi yapanlar ortaya çıkarılıp bunlar sube bankacılığı yerine daha masrafsız İnternet bankacılığına yönlendirilebilir.
- Bir ürün veya hizmetle ilgili bir kampanya programı oluşturmak için hedef kitlenin seçiminden başlayarak bunun hedef kitleye hangi kanallardan sunulacağı kararına kadar olan süreçte veri madenciliği kullanılabilir. Bu kararlar için yapılan gerçek

bir uygulama örneği kullanılan istatistiksel yöntemlerle birlikte aşağıdaki bölümde anlatılmıştır.

## VERİ MADENCİLİĞİ UYGULAMASI: KAMPANYA YÖNETİMİ

Bu bölümde bir işletmede veri madenciliği uygulamalarından faydalanılarak gerçekleştirilen bir pazarlama kampanyası örneği sunulacaktır. Gerçek bir vaka olan bu uygulama sonucunda kampanyadan başarılı bir sonuç elde edilmiştir. Burada özellikle üzerinde durulan nokta, veri madenciliği araçlarında kullanılan analizlerin önemli bir bölümünün istatistiksel analizler olduğu ve bu analizlerin hangi amaçlar için kullanıldığıdır.

Kredi kartı pazarlama amaçlı bu kampanya bir çapraz satış kampanyası olup amaç ödeme davranışı “iyi” olan bireysel kredi kullanıcılarına “kredi kartı” sunmaktır.

- 1- İstatistik işinin içine hemen ilk amaç aşamasında girmiştir. Kampanya için öncelikle mevcut müşteriler arasından ödeme davranışı “iyi” olanların belirlenmesi gerekmektedir. “iyi” ödeme davranışının ölçülebilir (operasyonel) tanımının yapılabilmesi için hangi özelliklerdeki müşterilerin iyi ödeme davranışlı müşteriler olduğunun belirlenmesi gerekir. Bu noktada ödeme davranışına etki eden özelliklerin yani değişkenlerin ve bu değişkenlerin ödeme davranışının ne derece etkilediğinin bulunması istatistiksel analizlerle mümkündür. Örneğin discriminant analiziyle veya lojistik regresyon kullanarak ödeme davranışı iyi ve kötü olan müşteriler sınıflandırılabilir. K-kare analizleriyle ödeme davranışı iyi ve kötü olarak kategorize edilen müşterilerin diğer kategorik özellikleriyle ilişkili olup olmadığına bakılabilir. Söz konusu kampanyada ise “iyi” ödeme davranışlı müşterinin tanımı ilgili departman yönetimlerinin tecrübelerine dayanarak karşılıklı görüş alışverişleri sonunda saptanmıştır.

Örneğin;

- Kredi ödemesini ödeme süresi içinde yapanlar ve
- Kredi ödemesini ödeme süresi içinde yapmayanlar filtreleme yöntemiyle ayrıştırılmıştır.

Ödemesini zamanında yapmayanlar ise gecikme süreleri ve kaç kez geciktirdiklerine göre sınıflandırılmış ve bu sınıflandırma sonunda belirlenen bir gecikme zaman ve sayısı baz alınarak tekrar sınıflandırılmıştır..

- Kredi geri ödemesini belli bir zamandan fazla geciktirenler ve
- Kredi geri ödemesini belli bir zamandan fazla geciktirmeyenler

Ödemesini belli bir zamandan fazla geciktirmeyenler ise bu kez;

- Bir kez geciktirenler
- Birden fazla geciktirenler olarak sınıflandırılmışlardır.

Anlaşılabileceği üzere burada kullanılan yöntemler tanımlayıcı istatistiksel analiz yöntemleridir. Veri ambarlarında depolanan veriler kullanılarak bu sınıflandırmalar binlerce, onbinlerce mevcut müşteri için, teknolojik destekle, dakikalarla ifade edilecek

çok kısa bir zaman içinde yapılmıştır. Bu asama sonunda ödeme davranışı “iyi” olan müşteriler ortaya çıkmış ve kredi kartı hedef pazarı olarak tesbit edilmiştir.

- 2- Hedef kitlenin belirlenmesinin ardından yine istatistiksel analizler kullanarak karar verilmesi gereken iki önemli husus kredi kartı limiti kararı ve kartın cinsidir. Limitin müşterinin hangi özelliklerine göre tayin edilmesi gerekir. Bu bilgi kredi kartının klasik mi yoksa gold kart mı olacağı kararını etkilediğinden dolayı önemlidir. Söz konusu özellikler gelir seviyesi, mesleği, araba, ev sahibi olma gibi özellikler olabilir. Bunların yanı sıra geri ödemesini düzgün yapan müşterilerin aldığı kredi tutarına veya aylık taksit ödeme tutarına ilişkin veriler de kredi limiti kararını vermek için kullanılacak uygun değişkenler olabilir. Bunun cevabını yine istatistiksel analizler sonucu elde etmek mümkündür.
- 3- Birinci ve ikinci aşamalardan sonra saptanan hedef kitleye, belirlenen kanallarla erişip ürünü sunmanın ardından kampanyanın etkinliğini ölçme aşamasında yine istatistik devreye girmektedir. Kampanyanın sonuçlarını değerlendirme aşamasında öncelikle kabul oranları incelenmiş ve bununla ilgili öngörülerin ne denli başarılı olduğu kontrol edilmiştir. Bu aşamada hedef kitle saptarken kullanılan yöntemlerin, çeşitli karar aşamalarında kriter olarak kullanılan değişkenlerin ne denli uygun olup olmadığı bu şekilde test edilebilmiştir. Bu uygulamada başarı oranı oldukça yüksek %75 ler civarında olmuştur. Ürün, sunulan hedef kitlenin % 75’i tarafından kabul etmiştir. Bu kabul oranı geleneksel yöntemlerle yönetilen bir kampanya yönetimine göre oldukça yüksek bir orandır. Bu oran veri madenciliği uygulamasını sunan firma tarafından dahi çok yüksek bulunmuştur. Başarı oranının yüksek olmasında etkin ve verimli biçimde kullanılabilen istatistiksel yöntemlerin payı oldukça fazladır. Kriter değişkenlerin doğru yöntemlerle tesbiti kabul oranını artırmıştır.

Kampanya sonunda hedeflenerek erişilen müşteriler için yepyeni bir değişken daha ortaya çıkmıştır. Kredi kartı kampanyasında ürünü kabul etme ve etmeme kararları yeni bir kategorik değişken olarak veri ambarında yerini alır. Bu veri, yeni bir kampanya için oldukça değerli bir veridir. Çünkü bu değişkeni kriter değişkeni olarak kullanarak ürünü kabul eden müşteri ile etmeyen müşterileri özelliklerine göre tekrar sınıflandırmak ve hangi özellikteki müşterilerin kabul etme olasılığının daha çok olduğunu belirlemek mümkündür. Bu şekilde bir sonraki kampanyanın etkinliği artırılabilir. Hangi müşterilerin kabul etme ihtimalleri yüksekse onlara erişilerek zaman ve maliyet tasarrufu yapılabilir.

Sonuçta bir kampanya aslında bir sonraki için çok değerli verilerin ortaya çıkmasını sağlamaktadır ve süreç bu şekilde sürekli kendi kendine geri besleme yaparak devam edebilir. Hem kampanya öncesinde mevcut müşteri bilgilerini hem de kampanya sonunda elde edilen sonuçları istatistiksel analizlerle değerlendirmek gerekir. Verilerin miktarı ve değişkenler arttıkça yeni analizler yapmak kaçınılmaz olacaktır. Veri madenciliği ve veri ambarı projelerinde en büyük sıkıntılardan biri henüz bu kavramlar ve uygulamalar yeni olduğu için geriye dönük verilerin simdilik yeterli olmadığı durumlar söz konusu olmaktadır. Zaman içinde uygulamalar geliştikçe ve piyasada istatistiği daha etkin kullanabilenler arttıkça ve başarılı sonuçlar elde edilmey başladıkça iş dünyası istatistik uygulamalarının önemini daha da iyi anlayacaktır.

## SONUÇ

Yogun rekabet ortamında çalışan işletmeler kendilerini farklılaştırabilmek için bilgiyi etkin ve verimli bir biçimde kullanma yoluna gitmektedirler. Bilgiyi elde etmek için eskiden beri kullanılmakta olan yöntemlerin en önemlilerinden biri olan istatistiksel yöntemler, hakettikleri öneme kavusmaktadırlar. Sağladığı faydayı gören iş dünyası istatistik uygulamalarını adeta yeniden keşfetmeye başlamışlardır.

Veri madenciliği uygulamalarından fayda sağlanmasına neden olan en önemli faktörlerden birisi çok miktarda veriyi istediğiniz şekilde işleyebilme olanıdır. Altını çizilmesi gereken husus ise bu verileri işleme yöntemlerinin aslında senelerdir en temel işletme istatistiği derslerinde de okutulan yöntemler olmasıdır. Yeni olan uygulama, teknolojik olanaklar sayesinde yapılması gereken analizlerin çok daha düşük maliyet ve sürelerde yapılabilmesidir. Dolayısıyla zaten çok önemli olduğu kimse tarafından inkar edilemeyen istatistiksel analizlerin bu denli kolay uygulanabilir olmasından dolayıdır ki istatistik uygulamalarını iş dünyası yeniden keşfetmiştir ve bunun yararlarını her dönemden daha çok ve daha somut bir biçimde şahit olmaktadır. Yukarıda uygulama örneği olarak anlatılan kampanyanın yöneticisi bu konuda şöyle söylemektedir.

*“.....Veri ambarları ve veri madenciliği uygulamaları sayesinde, pazarlamacılar ile bilgi işlemciler ayrılmaz kardaşlar olmaya başlamışlardır. Yepyeni bir “pazarlamacı” kavramı ortaya çıkacaktır. İstatistik ve yöneylem araştırması bilmeyen pazarlamacıların devri kapanacaktır...”*

Bunun istatistik bilimine ve ilgili taraflara çok büyük yararlar sağlayacağı yadsınamaz. İstatistikler ve istatistiksel analizler her alanda her zaman çok önem taşımıştır. Günümüzde farklı olan bilgisayar, internet teknolojilerinin sağladığı olanaklardır. İstatistiğin öneminin daha fazla kişi tarafından anlaşılmasıdır.

## REFERANSLAR

Berry M.J., Linoff G.S.; (2000); *Mastering Data Mining*; John Wiley & Sons, Inc. New York.

Ronald Swift; (2001); *Accelerating Customer Relationship*; Prentice Hall PTR

Glymour C., Madigan D., Pregibon D., Smyth P.; (1997); Statistical Themes and Lessons for Data Mining; *Data Mining and Knowledge Discovery 1*;11-28

Kolsky J.; (1997); Statistics and Data Mining in the Analysis of Massive Data Sets; <http://www.infosense.com/publicatons>

Maindonald J.H.; (2000); New Approaches to Using Scientific –Data Statistics, Data Mining & Related Technologies In Research and Research Training; <http://www.anu.edu.au> ; The Australian National University.

Rocke David M.; (2001); A Perspective on Statistical Tools for Data Mining Applications; <http://lac.uic.edu>

