E-Ticaret Tıklama Verileri Üzerinde Satın Alma Niyeti Tahmini

Prediction of Purchase Intention on the E-Commerce Clickstream Data

Ahmet Gürbüz, Mehmet S. Aktaş
Bilgisayar Mühendisliği Bölümü, Yıldız Teknik Üniversitesi
ahmet.gurbuz@std.yildiz.edu.tr, aktas@yildiz.edu.tr

***Özetçe* — Her gün dünya çapında milyonlarca müşteri e-ticaret siteleri üzerinden alış-veriş yapmaktadır. Müşterilerin e-ticaret web siteleri kullanımından büyük miktarda tıklama verileri üretilmektedir. Tıklama verilerinden kullanıcıların davranışlarının öğrenilmesi ve satın alma niyetinin tahmin edilmesi, önemli bir ihtiyaç haline gelmiştir. Bu araştırma kapsamında, müşterilerin oturumlarını sonlandırmadan önce, satın alma niyetinin tahmin edilebilmesi üzerinde bir metodoloji önerilmektedir. Önerilen metodoloji, ACM RecSys Sempozyumu tarafından 2015 yılında yayınlanmış e-ticaret veri seti üzerinde denenmiş ve başarılı sonuçlar elde edilebildiği tespit edilmiştir.**

Anahtar Kelimeler—Tıklama akışı verisi; Web kullanım madenciliği; Artımlı kümeleme; Anomali Tespiti

***Abstract* — Every day millions of customers do shopping through e-commerce web sites around the world. Huge amounts of click stream data are generated from customers' use of e-commerce websites. Learning the behavior of users from click data and estimating the intention to buy has become an important need. Within the scope of this research, a methodology is proposed on the estimation of the purchase intention before finalizing the sessions of the customers. The proposed methodology was tested by the ACM RecSys Symposium on e-commerce data set published in 2015 and it was found that successful results could be obtained.**

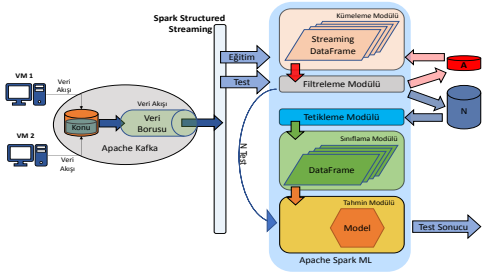
Keywords—Clickstream data; Web usage mining; Incremental clustering; Anomaly Detection

I. GIRIŞ

E-ticaret sitelerinde alışveriş yapan müşterilerin satın alma niyetlerini önceden anlamak ve satın alma olasılığını arttırmak için aksiyonlar alabilmek büyük önem taşımaktadır. Siteler ile etkileşimde bulunan kullanıcıların yaptıkları işlemler, sistem log dosyalarında kayıt altına alınmaktadır. Müşterinin oturum bilgileri, çevrimiçi mağazalar için yararlı olabilecek bilgilerin önemli bir kaynağıdır. Bu veriler tıklama akışı verileri olarak adlandırılmaktadır ve ziyaretçilerin e-ticaret sitesindeki eylemlerinin yapılandırılmış kaydını sağlamaktadır. Tıklama verileri, müşterinin davranışını modellemede kullanılmakta ve e-ticaret sitesinin müşterileri ile ilişkilerini geliştirmek için kullanılabilmektedir.

E-ticaret şirketleri çevrimiçi müşterileri ihtiyaçlarına en uygun yeni stratejiler benimsemelidir. Çevrimiçi müşteriler, geleneksel ticaretteki müşterilerden farklı davranışları gösterebilmektedir. Örneğin;geleneksel ticarette müşteri, mağaza sahibi ile sözlü diyaloğa geçerken e-ticarette müşteriler basit çevrimiçi sorgularla iletişime geçebilmektedir. Müşterilerin ihtiyaçlarına uygun ürünün arayışını e-ticaret sitelerinde yapabilmesi, fiziksel mağazaların aksine, çevrimiçi mağaza müşteriler için, müşteri ve e-ticaret sitesi arasındaki mesafeyi ortadan kaldırmaktadır. Yalnızca birkaç tıklama ile, müşteriler aynı ürünü farklı satıcılardan internet ortamında bulabilmektedir. Bu nedenlerden dolayı, e-ticaret siteleri, müşterileri ile olan bağlarını güçlendirmek için yöntemler geliştirmelidir. E-ticarette müşteri sadakati oluşturmak isteyen şirketler, tüketicinin satın alma olasılığını arttıracak veya gelecekte tekrar çevrimiçi mağazasından satın almaya çekecek çözümler üretmelidir.

Bu araştırma kapsamında, yukarıda bahsettiğimiz motive edici nedenlere çözüm sağlayabilecek bir metodoloji önerilmektedir. Önerilen metodoloji; e-ticaret sitesi müşterilerinin oturum verilerini kullanarak, müşteri davranışı öğrenme modelleri oluşturmaktadır. Oluşturulan modeller kullanılarak, müşterilerinin satın alma niyeti, henüz oturumun başlarında tahmin edilmektedir. Müşteri oturum verileri üzerinde öğrenme gerçekleştirilmeden önce, aykırı veri analizi yapılmakta ve aykırı değerler içeren oturumlar öğrenme modeli eğitiminde kullanılmamaktadır. Aykırı veri tespiti için arttırımlı bölme tabanlı kümeleme analizi yöntemi (arttırımlı k-means algoritması) kullanılmıştır. Satın alma yönelimi tahmini için, örnek-tabanlı ikili sınıflandırma yöntemi (Lojistik Regresyon algoritması) kullanılmıştır. Önerilen metodoloji, 2015 yılında ACM RecSys Sempozyumu tarafından yayınlanmış e-ticaret sitesi tıklama veri seti [1] üzerinde uygulanmış ve metodolojinin, satın alma eğilimini tahmindeki başarısı test edilmiştir. Elde edilen sonuçlar, sistemin satın alma niyetini tahmin etmede başarılı tahminler yapabildiğini göstermektedir.



Şekil 1. Sistemin Çalışma Yapısı

Bu bildirinin 2. bölümünde, e-ticaret müşterilerinin tıklama akışı verileri üzerinde yapılan benzer çalışmalar anlatılmaktadır. 3. bölümde önerilen metodoloji aktarılmaktadır. 4. bölüm, önerilen sistemin kullanılabilirliğini göstermek için geliştirdiğimiz prototip uygulamanın detaylarını sunmaktadır. Geliştirilen sistem üzerinde yapılan değerlendirme sonuçları 5. bölümde yer almaktadır. Son bölüm olan 6. bölümde, çalışmanın sonuçları tartışılmakta ve geleceğe dönük çalışmalar özetlenmektedir.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©2019 IEEE

II. LİTERATÜR TARAMASI

Literatür incelendiğinde, e-ticaret tıklama verileri üzerinden bilgi çıkarımı alanında yoğunlaşan farklı araştırmalar görülmektedir. Örneğin; Sheetal, Shailendra ve Ashwin [2] tarafından yapılan çalışmada, web kullanım kayıtları üzerinden, kullanıcıların siteyi ziyaret ederken oluşturdukları seyahat deseni çıkartılmaktadır. Bu çalışmada, web sunucusu kayıtlarında; hem farklı kullanıcıları hem farklı oturumları ayırt edebilmek için bir yaklaşım önerilmiş ve benzer yöntemlerle karşılaştırılmıştır. Diğer bir örnek; Samiah, Asger ve Muheet [3] tarafından yapılan çalışmada; Glassman ve Albarracin [4] tarafından ilk olarak önerilen Tutum-Davranış İlişkisi hesaplaması taban alınarak, Tutum-Davranış Eş-İlişkisi formülü önerilmiştir. Geliştirdikleri formül ile iki farklı e-ticaret tıklama veri seti üzerinde gerçekleşen satın alımlardaki davranış ilişkileri benzerliği ortaya konmaktadır. Bu araştırmada; geçmişteki bu örnek çalışmalardan farklı olarak, e-ticaret sitesi müşterilerinin satın alma niyetlerinin, henüz oturumun başlarında, tahmin edilmesi üzerinde çalışılmaktadır.

Literatürde, akan veri üzerinde aykırı durum tespiti konusunda çalışmalar gözlenmektedir. Mohiuddin ve diğerleri [5] tarafından yapılan çalışmada VMware-tabanlı performans verileri üzerinde gerçek zamanlı dağıtık anomali tespiti yapılmıştır. Artımlı K-Ortalama algoritması ile anomali tespitinde %98' e varan doğruluk oranı ile aykırı veri tahmini yapabilmişlerdir. Bu çalışma kapsamında, e-ticaret müşterisi oturum verileri içinde (gerek satın alma gerçekleşmesi gerekse satın almanın gerçekleşmemesi durumları için) aykırı verilerin gerçek zamanlı olarak tespiti üzerinde çalışılmaktadır. Dağıtık ortamda yerleşik ve büyük ölçekli veri yönetimi, ya da web servis yönetimi üzerinde, farklı dağıtık sistem çözümleri literatürde önerilmiştir [9, 10, 11, 14, 15]. Bu çalışmada dağıtık veri işleme ve analizi için Spark platformu kullanılmaktadır. Verinin hayat döngüsünü takip etmek için, provenans araştırma alanında, farklı çözümler sunulmaktadır [12, 13, 16, 17]. Bu çalışmada ise, kullanıcıların, bir web sitesi üzerindeki hareketleri tıklama verileri kullanılarak takip edilmektedir.

III. METODOLOJİ

Bu araştırma kapsamında önerdiğimiz metodoloji, e-ticaret sitesi müşterilerinin, ürün satın alma niyetlerini, web oturum süresi devam ederken, gerçek zamanlı olarak, tahmin etmektedir. E-ticaret sistesi müşterilerinin, oturum verileri üzerinde, makine öğrenme algoritmaları çalıştırılarak, müşterinin satın alma davranışı modellenmektedir. Oluşturulan öğrenme modeli kullanılarak, yeni müşteri web oturumlarında, müşterilerin satın alma eğilimleri tahmin edilmektedir. Önerdiğimiz metodoloji Şekil-1' de resmedilmektedir. Önerilen metodoloji, müşteri oturum verileri üzerinde aykırı veri analizi de yapmaktadır. Şekil-1' de resmedildiği üzere, sadece normal oturum verileri, satın alma tahmininde kullanılan öğrenme modelllerinin eğitiminde kullanılmaktadır. Akan oturum verisi üzerinde uygulanan aykırı veri tespiti işlevini, oturum verisi üzerinde satın alma davranışı tahmini işlevi takip etmektedir. Bu araştırma kapsamında, tıklama verileri üzerinde, veri dönüşümü işlemi uygulanarak, oturum verileri oluşturulmuştur. Oturum verileri üzerinde, aykırı veri tespiti ve satın alma davranışı tahmini için de özellik çıkarımı işlemi gerçekleştirilmiştir. Bu işlemleri sırasıyla aşağıda aktarıyoruz.

Veri Ön İşlemesi: E-ticaret tıklama kayıtı içinde; tıklamanın gerçekleştiği zaman damgası, tıklanan ürün kodu, tıklanan ürünün kategorisi, satın alınan ürünün miktarı ve satın alınan ürünün fiyatı, bulunmaktadır. Tıklama kayıtlarının bazıları, ürün satın alımlarını içerirken; diğerleri ürün satın alımı söz konusu olmadığından, ürünlerle ilgili verileri içermemektedir. Veri Ön İşleme aşamasında; oturum belirleme çalışması yapılmıştır. Oturumlar, zaman tabanlı sezgizel yaklaşımla belirlenmiştir. Buna yaklaşımla, tıklama kayıtları, oturum kayıtlarına dönüştürülmüş ve tüm oturum kayıtlarında ürün verilerinin de olması sağlanmıştır. Satın alma gerçekleşmemiş oturum kayıtlarında ürünle ilgili veriler sıfır olarak atanırken, satın alma gerçekleşmiş kayıtlarda ürün adeti ve fiyatı bilgileri olduğu gibi kalmıştır. Oturum kayıtlarına, oturum tanımlayıcısı alanı da eklenmiştir. Oturum süresi, konfigüre edilebilir olmakla birlikte, ilk başlangıçta 30dk olarak belirlenmiştir. Oturum kayıtlarındaki, oturum tanımlayıcısı alanına, sistemin t_0 anından başlayarak; her 30dk' da bir ardışık olarak artacak şekilde yeni bir oturum tanımlayıcısı atanmıştır. Eş oturum süresinde yer alan tüm kayıtlarda, oturum tanımlayıcısı alanı aynı değeri almıştır.

Özellik Çıkarımı: Bu araştırma kapsamında, bir e-ticaret sitesindeki, oturum verileri üzerinde satın alma tahmini yapılabilmesi için, satın alma davranışını en iyi şekilde modelleyebilecek, tahmin edici, özellikler tespit edilmiştir. Her bir oturum verisi için, bir özellik vektörü oluşturulmuştur. Satın alma niyetini tahmin etmek amaçlı kullandığımız özellikleri şu şekilde özetliyoruz: Oturum Süresi: Ziyaretçinin oturum süresi fazla ise, ürün alma olasılığının yükseldiği; oturum süresinin az olduğu durumda ise, ürün alma olasılığının düştüğü söylenebilir. Oturumda Ziyaret Edilen Sayfa Sayısı: Bir oturum süresinde ziyaret edilen sayfa sayısı çok fazla ise, kullanıcının bir ürünü almaktan ziyade, site üzerinde ürün keşfi yaptığı söylenebilir. Oturumun Kısa Süreli Olup Olmaması: Bir oturumun süresi çok kısa ise (örneğin: sitede sadece bir sayfaya bakılıp, site terkediliyorsa), kullanıcının ürün satın almaya meyilli olmadığı söylenebilir. Oturum süresi belli bir eşik değerinin altında ise (örneğin 5sn), bu özelliğin değerine "1", diğer durumda ise "0" atanmıştır. Oturumda Farklı Kategoride Sayfalar Ziyaret Edilip Edilmemesi: Bir oturum süresinde çok sayıda farklı kategorilerde sayfa incelenmişse, kullanıcının ürün satın almaya meyilli olmadığı söylenebilir. Oturum Sürecinde Tıkalan Üründe Promosyon Olması: Bir oturum sürecinde, tıklanan bir ürünle ilgili promosyon söz konusu ise, bu durumun ürünün satın alma olasılığını arttıracağı söylenebilir. Promosyon var olduğu durumda, bu özelliğin değeri "1", diğer durumda "0" olarak atanmıştır. Özellik Vektörü Sınıf Etiketi: Oturum sürecinde, ürün satın alımı söz konusu olmuşsa (oturum verisinde ürün adet ve fiyat bilgileri varsa), özellik vektörünün sınıf etiketine "1" atanmıştır. Satın alım olmaması durum için sınıf etiketi "0" değerini almıştır.

Aykırı Veri Analizi: Aykırı veri analizi, oturum verileri üzerinde arttırımlı parça tabanlı kümeleme analizi kullanılarak gerçekleştirilmiştir. Bu amaçla, e-ticaret sitesi log dosyasından veri dönüşümü yoluyla oluşturulan web oturum verileri üzerinde; rassal olarak yapılan örnekleme ile tüm kayıtları temsil edebilecek nitelikteki, web oturum kayıtları seçilmiştir. Seçilen kayıtlar manuel olarak incelenmiş ve anomali içermeyen oturum verileri tespit edilmiştir. Bu verilerin üzerinde arttırımlı k-ortalama (k-means) algoritması kullanılarak segmentasyon yapılmıştır.

Bir sonraki aşamada, log doyasından elde edilen diğer web oturum kayıtları, arttırımlı k-means algortiması ile mevcut kümelere dağıtılmıştır. Hiçbir kümeye yerleştirilemeyen web oturum kayıtları, aykırı değer olarak nitelendirilmiştir. Öğrenme modellerinin eğitiminde sadece normal kullanıcı davranışı sergileyen (iyi huylu davranış kümelerinden birinde segmentasyonu yapılabilen) oturum verileri kullanılmıştır. Burada, her yeni gelen oturum verisi, sistemin kümeleme modülü üzerinden bir defa geçmektedir. Kümeleme esnasında veri noktaları herhangi bir kümeye uymuyorsa noktanın en yakın olduğu kümeye bakılmaktadır. Ve aşağıdaki denklem ile benzerlik oranları ile hesaplanmaktadır.

Benzerlik = Kümenin Yarıçapı / Kümeye olan Uzaklık

Benzerlik eşik değerleri: Benzerlik>=0.9, kümeye dahil edilir. 0.66<Benzerlik<0.9 ise yeni küme yaratılır ve kümenin merkezine noktanın kendisi atanır. Kümenin ilk yarıçapı olarak bu algoritmanın kullanıldığı modül çalıştırılırken kullanıcıdan alınan yarıçap değeri atanmaktadır. Benzerlik<=0.66 ise veri noktası "anomali" olarak etiketlenmektedir. Aksi durumlarda veri "iyi huylu" olarak etiketlenir. Bir veri noktası bir kümeye dahil edilirken benzerlik oranının Benzerlik<1 olarak seçilmesi kümelere yarıçaplarını genişletebilme yeteneği kazandırmaktadır. Benzerlik>=1 olarak verilmiş olsaydı yeni gelen ve kümelere girecek olan noktalar en fazla kümelerin yarıçapları kadar uzakta olabilmekte ve kümelerin yarıçapları genişleyememektedir. Artımlı K-Ortalama algoritmasının sözde kodu aşağıdaki gibidir.

Algoritma: Artımlı k-ortalama

```
% Gösterim: a ← b'nin anlamı a b'ye atandı.
% a ← [] a'nın dizi olduğu gösterir.
% ilkYarıçap: kümeler yeni oluşturulurken verilen yarıçap uzunluğu
% veriNoktaları: noktaların ait olduğu kümelerin dizi elemanları
% sonuc == -1: aykırı veridir, benzerlikOranı < 0.66
benzerlikOranı ← kümelere girecek verinin benzerlik oranı
ilkYarıçap ← kümelerin başlangıç yarıçap uzunluğu
maxKüme ← maks küme sayısı
toplamKüme ← []
veriNoktaları ← []
kümeMerkezleri ← []
kümeYarıçapları ← []
for veri noktası n do
    sonuc = herhangiKümeyeUydur(n, kümeMerkezleri,
kümeYarıçapları)
    if sonuç != -1 then
        if sonuc.benzerlik > benzerlikOranı then
            kümeGüncelle (kümeYarıçapları, veriNoktaları)
        else
            kümeOlustur(kümeMerkezleri)
            toplamKüme ← toplamKüme +1
            if toplamKüme > maxK then
                ikiEnYakınKümeyiBirleştir()
        veriNoktaları.label = "N"
    else
        veriNoktaları.label = "A"
```

Bir veri seti üzerinden Artımlı K-Ortalama algoritması sonucu oluşacak küme sayısı: (1) yeni oluşacak kümelere verilecek ilk yarıçap, (2) veri noktalarının kümelere benzerlik oranı, (3) maksimum oluşacak küme sayısıdır.

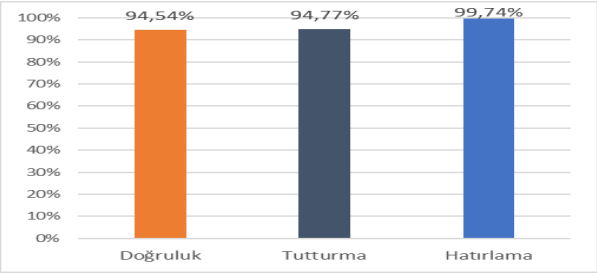
İkili Sınıflandırıcı Öğrenme Modeli Oluşturulması: Normal satın alma davranışı gösteren oturum verileri, eğitim verileri olarak kullanılmıştır. Eğitim oturum verileri içinde, sınıf etiketleri de (satın alma durumu için "1", satın alma gerçekleşmemiş ise "0") yer almaktadır. Etiketli web oturum verileri kullanılarak, kullanıcıların satın alma davranışını tahmin etmede kullanılacak öğrenme modelleri oluşturulmaktadır. Bu araştırma kapsamında, ikili sınıflandırıcı algoritmaları temsilen Lojistik Regresyon yöntemi Batch Gradient Descent yaklaşımı ile gerçeklenerek kullanılmıştır.

İkili Sınıflandırıcı Satın Alma Niyeti Tahmini: İkili sınıflandırma modelinin oluşturulmasıyla birlikte, sisteme her yeni gelen ve etiketi belli olmayan, web oturum verilerinin hangi sınıfa ait olduğunun tahmini yapılabilmektedir. Bir e-ticaret sitesi kullanıcısının, henüz oturumun başlarındayken, oluşturmuş olduğu davranış şekli (özellik vektörü), önerdiğimiz sisteme gönderilmekte ve ikili sınıflandırıcı modeli kullanılarak, satın alma niyeti tahmin edilmektedir. Satın alma niyeti olumsuz yönde olan kullanıcıların belirlenmesi, bu kullanıcılara yönelik pazarlama ve kampanya faaliyetlerinin tetiklenmesini sağlamaktadır.

IV. PROTOTIP UYGULAMA

Çalışma kapsamında, e-ticaret alanındaki müşterilerin satın alma davranışlarını müşteriler henüz oturumlarını bitirmeden önce tahmin edebilen bir prototip uygulama geliştirilmiştir. Geliştirilen proje açık kaynaklı olarak GitHub kod paylaşım sitesinde paylaşılmıştır [6].

E-ticaret sitesi müşterilerinin tıklama verilerini içeren bir veri seti üzerinde, geliştirilen prototip uygulaması denenmiştir. Uygulamada, Apache Kafka ve Apache Spark Structured Streaming platformları kullanılmıştır. Prototip uygulamanın çalışma yapısı Şekil-1' de resmedilmektedir. Uygulama ilk olarak, tıklama verileri üzerinde veri dönüşümü uygulayarak, web oturum verileri oluşturmaktadır. Web oturum verileri Apache Kafka üzerinden yayın yapılarak veri kaybı olmadan Apache Spark Structured Streaming modülüne gönderilmektedir. Bir sonraki aşamada Apache Spark üzerinde sırası ile aykırı veri analizi ve öğrenme modeli oluşturulması işlemleri yapılmaktadır. Son aşamada, sınıflandırıcı tarafından oluşturulan model, sistemin tahmin modülü tarafından kulllanılarak, yeni oturum verilerinin sınıf etiketlerini gerçek zamanlı olarak tahmin etmektedir. Anomali verilerin tespiti için artımlı k-ortalama algoritması kodlanmıştır. Sınıflandırıcı olarak, Spark MLlib Streaming kütüphanesi ile birlikte gelen Lojistik Regresyon sınıfı kullanılmıştır. Sınıflandırıcının yaptığı tahminler kullanılarak, e-ticaret sitesi müşterilerinin oturum içerisindeki satın alma niyetleri tahmin edilebilmiştir. Çalışma kapsamında, geliştirilen sistem için Apache Kafka'nın kafka_2.11-0.10.2.2 versiyonu, Apache Spark'ın 2.4.0 versiyonu kullanılmıştır. Apache Spark'a yazılan uygulamalar Spark'ın Python [7] uygulama ara yüzü kullanılarak geliştirilmiştir. Python programlama dilinin 3.5.2 versiyonu kullanılmıştır. Geliştirilen prototip uygulama, Apache Kafka'dan veri akışını alan modül, ardışık olarak artımlı kümeleme ve sınıflama yapan modül ve tahmin modülü olmak üzere toplam 3 adet spark-submit yazılımı olarak kodlanmıştır.



Şekil 2. Sistemde Yapılan Testler ve Sonuçları

Python programlama dilinin Numpy [8] 1.11.0 sürümlü kütüphanesi kullanılmıştır. Sistem Ubuntu Linux işletim sisteminin 16.04.3 versiyonu üzerinde çalıştırılmıştır.

V. Değerlendirme

Önerilen metodolojinin başarısı, geliştirilen prototip uygulama üzerinde gerçekleştirilen testlerle ölçülmüştür. Bu amaçla ACM RecSys Sempozyumu tarafından 2015 yılında yayınlanmış e-ticaret veri seti testlerde kullanılmıştır. Veri seti iki ana dosyadan oluşmaktadır. İçerisinde sadece tıklamaları bulunduran dosya içeresinde 9,2 milyon satır tıklama kaydı, satın alımların bulunduğu dosyada ise 2,4 milyon kayıt bulunmaktadır. Dosyaların içinde kayıt yapıları şu şekildedir:1. Veri Dosyası Kayıt Yapısı: Oturum No, Zaman Damgası, Ürün No, Kategori. 2. Veri Dosyası Kayıt Yapısı:Oturum No, Zaman Damgası, Ürün No, Fiyat, Miktar. Veri ön işleme aşamasında, bu iki dosya içerisindeki aynı oturum numarası sahip kullanıcı verileri tek bir dosya altında birleştirilerek, veri dönüşümü gerçekleştirilmiştir. Özellik Çıkarımı aşamasında, tek dosyaya dönüştürülmüş olan, oturum verileri üzerinden, özellik değerlerinin çıkarımı yapılmıştır. Özellik Vektörü Kayıt Yapısı: Oturum No, Oturum Süresi, Kısa Süreli Ziyaret, Sayfa Sayısı, Farklı Sayfa Sayısı, Farklı Kategori, Promosyon ve Özel Tekli, Sınıf Etiketi. Oluşan özellik vektöründeki veri setinde 2.sütundan 7.sütuna kadar olan nitelikler kullanıcıların satın alma davranışlarını betimlerken, son sütundaki nitelik ise kullanıcının ürünü satın alıp almadığını belirtmektedir.

Testler şu özelliklerdeki donaım üzerinde çalıştırılmıştır: LENOVO Z50-70, Intel i5-4210U (1.70-2.40GHz) işlemci, 8GB DRR3 1600Mhz RAM, Ubuntu Linux 16.04.3 LTS işletim sistemi. Testlerde kullanılan veri seti içerisindeki tıklama kayıtlarını içeren veri seti 1.5GB, satın alımları içeren veri seti 55MB büyüklüğündedir. Ön işleme adımı sonrasında oluşturulan veri seti 250MB büyüklüğünde ve bir tanesi sınıf etiketi olmak üzere 7 adet nümerik nitelikten oluşmaktadır. Şekil 2'de görüleceği üzere sistemin performansı, tıklama veri seti üzerinde çalışması ile ürettiği doğruluk (accuracy), tutturma (precision) ve bulma (recall) yüzdelerine göre değerlendirilmiştir. Sistemin başarısı ölçülürken, k-katlamalı çapraz doğrulama yöntemi kullanılmıştır. Elde edilen sonuçlar satın alma niyeti tahmininde başarılı sonuçlar elde edilebileceğini işaret etmektedir.

VI. Sonuçlar Ve Gelecekteki Çalışmalar

Bu araştırma kapsamında, müşterilerin oturumlarını sonlandırmadan önce, satın alma niyetinin tahmin edilebilmesi üzerinde bir metodoloji önerilmektedir. Önerilen metodoloji, ACM RecSys Sempozyumu tarafından 2015 yılında yayınlanmış e-ticaret veri seti üzerinde denenmiş ve başarılı sonuçlar elde edilebildiği tespit edilmiştir. Önerilen metodolojinin çıktıları, e-ticaret sitelerinin satışlarını arttıracak pazarlama faaliyetlerini tetikleyici nitelikte olacaktır.

Gelecekteki çalışmalar arasında, farklı örnek-tabanlı öğrenme modelleri oluşturan farklı sınıflandırıcı algoritmaları kullanılarak, satın alma niyetinin tahminlerindeki başarı oranlarının arttırılması üzerinde çalışılacaktır.

Kaynakça

[1] RecSys 2015 Challenge Clickstream Data Web Sitesi: https://2015.recsyschallenge.com/challenge.html, Son Ulaşım Tarihi: 31.1.2019.

[2] Raiyani, S.A., Jain, S., Raiyani, A.G., "Advanced Preprocessing using Distinct User Identification in Web Log Usage Data.", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol.1, Issue.6, pp.418-422, 2012.

[3] Nasti, S.J., Asger, M., Butt, M.A., *"Prognostication of Customer Behaviour using Web Data", International Journal of Advance Research in Computer Science and Management Studies, Vol.4, Issue.5, pp.41-45*, 2016.

[4] Glasman L.R. et al., "Forming Attitudes That Predict Future Behavior: A Mete-Analysis of the Attitude-Behavior Relation", *Psychological Bulletin, Vol 132, Number 5, pp. 778-822*, 2006.

[5] Solaimani, M. et al., "Spark-based anomaly detection over multi-source VMware performance data in real-time", *CICYBS-2014*, 2014.

[6] Prediction of Purchase Intention on the E-Commerce Clickstream Data Web Sitesi: https://github.com/purchase-intention-prediction/-Prediction-of-Purchase-Intention-on-the-E-Commerce-Clickstream-Data, Son Ulaşım Tarihi: 31.1.2019.

[7] Apache Spark Python API Web Sitesi: https://spark.apache.org/-docs/latest/api/python/index.html, Son Ulaşım Tarihi: 31.1.2019

[8] Python Numpy Web Sitesi: http://www.numpy.org, Son Ulaşım Tarihi: 31.1.2019.

[9] Yildiz, B., Wu, K., Byna, S., Shoshani, A., "Parallel membership queries on very large scientific data sets using bitmap indexes," *Journal of Concurrency and Computation: Practice & Experience, 2019. DOI: 10.1002/cpe.5157*, 2019.

[10] Yildiz, B., Fox, G., "Toward a modular and efficient distribution for web service handlers," Journal of Concurrency and Computation: Practice & Experience, vol. 25 iss. 3, pp. 410-426, DOI: 10.1002/cpe.2854, 2013.

[11] Aktas, MS., Fox, GC., Pierce, M., "Information services for dynamically assembled semantic grids", *First International Conference on Semantics, Knowledge and Grid*, 2015.

[12] Jensen, S., Plale, B., Aktas, MS, Luo, Y, Chen, P, Conover, H., "Provenance capture and use in a satellite data processing pipeline", *IEEE Transactions on Geoscience and Remote Sensing 51 (11), 5090-5097*, 2013.

[13] Aktas, MS., Plale, B., Leake, D., Mukhi, NK., "Unmanaged workflows: Their provenance and use", *Data Provenance and Data Management in eScience, 59-81*, 2013.

[14] Aktas, MS, Pierce, M., "High-performance hybrid information service architecture", *Concurrency and Computation: Practice and Experience 22 (15), 2095-2123*, 2010.

[15] Aktas, M.S., Pierce, M., Fox, G.F., "Designing ontologies and distributed resource discovery services for an earthquake simulation grid", *Proc. of the GGF11 Semantic Grid Applications Workshop, Honolulu, USA*, 2004.

[16] Y Tas, MJ Baeth, MS Aktas, "An approach to standalone provenance systems for big social provenance data", *12th International Conference on Semantics, Knowledge and Grids*, 2016.

[17] Aktas, M.S., Astekin, M., "Provenance aware run‐time verification of things for self‐healing Internet of Things applications", Concurrency and Computation: Practice and Experience 31 (3), 2019.