

Müşteri Yolculuğu Analitiğini Makine Öğrenmesi Modeliyle Tahminleme

Yağız YAMAN

^aKırıkkale University, Computer Engineering Department

Özet

Bu çalışmada, müşterilerin davranışlarından ve müşterileri tanımlayan özneliklerden müşteri serüveninin belirlenmesi, belirlenen müşteri tiplerine göre bir sonraki gelecek müşteri veya mevcut müşterilerin gelecekteki davranışlarını tahmin eden makine öğrenmesi modeli geliştirilmiş ve bu modelin çıktıları ve kullanımı hakkında bilgi verilmiştir. Pazarlamada önemli konulardan birisi olan müşteri davranış analizi için, müşterinin yapmış olduğu geçmiş davranışlar üzerinden gelecekte yapabileceği olası davranışların, firmalar tarafından ön görülmesi, hem müşterinin memnuniyeti hem de firmanın finansal kazancı için önemlidir. Bu çalışmada sentetik bir veri kümesi kullanılmıştır. Çalışma, CRISP-DM (Cross-Industry Standard Process for Data Mining) metodolojisi aşamaları ile ilerletilmiştir. Veri, çeşitli ön işleme adımlarından sonra makine öğrenmesi için hazırlanmış, modeller birbiri ile bağlantılı olacak şekilde kurulmuş ve daha sonra tahminler yapılmıştır. Eğitilen model, sisteme yeni katılacak bir müşterinin başlayacağı aktiviteyi ve girilen adım parametresi ile kaç adım sonra hangi aktivitede olacağı hakkında tahminler yapmıştır. Yapılan tahminler sonucu model başarıları değerlendirilmiş ve başarıların, oluşturulan modellere göre değiştiği gözlemlenmiştir, bu gözlemler ve çıktılar, çalışmada sunulmuştur. En kötü tahmin modelinin başarısı %75 olup en iyi tahmin eden model başarısı %100'dür. Çalışma, müşteri yolculuğu analitiği sürecinin düzgün bir şekilde uygulanması durumunda firmalara kaybettikleri müşteriler için aksiyon noktalarını belirleyip üzerinde düzeltme veya müdahale imkanı sağlamaktadır. Şayet kayıp ile sonuçlanmayan bir serüven söz konusu ise, firmalar gelirlerini ve memnuniyetlerini arttırmak amacı ile bir sonraki müşteri davranışı için aksiyon alıp kampanyalar düzenleyebilirler.

Anahtar Kelimeler: Müşteri Yolculuğu Analitiği, Yapay Zekâ, Makine Öğrenmesi

Summary

In this study, a machine learning model was developed that predicts the future behavior of the next future customer or current customers according to the determined customer types, and information about the outputs and usage of this model is given. For customer behavior analysis, which is one of the important issues in marketing, companies need to predict possible future behaviors based on past behaviors of the customer, both for customer satisfaction and financial gain of the company. A synthetic dataset was used in this study. The study was advanced with the steps of the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology. The data was prepared for machine learning after various preprocessing steps, the models were set up to be interconnected, and then predictions were made. The trained model made predictions about the activity that a new customer to join the system will start and how many steps he will take with the entered step parameter. As a result of the predictions made, the model successes were evaluated and it was observed that the successes changed according to the models created. These observations and outputs are presented in the study. The success of the worst predictor model is 75%, and the success of the model that predicts the best is 100%. The study provides companies with the opportunity to determine action points for lost customers and correct or intervene on them if the customer journey analytics process is implemented properly. If there is an adventure that does not result in loss, companies can take action and organize campaigns for the next customer behavior in order to increase their income and satisfaction.

Keywords: Customer Journey Analytics, Artificial Intelligence, Machine Learning

1. Giriş ve Tanım

Müşteri yolculuğu analitiği kavramı, özünde müşteri yolculuğunu içermektedir ve bu kavramın tanımı şu şekilde yapılabilir: Müşteri yolculuğu, müşterinin bir ürün veya bir hizmet ile ilgili olarak satın alma kararı veya satın almama kararı öncesinde geçtiği sıralı adımları ifade eder. Müşteri yolculuğu birden çok temas noktasına yayılabilir. Temas noktaları, ürün veya hizmet deneyimlerini teşvik eden deneyimsel süreçler kümesi olarak adlandırılabilir (Berry ve diğerleri, 2006). Müşteri yolculukları, sektör, pazar ve iş süreçlerine bağlı olarak, günler veya haftalar sürebilir. Öyle ki; gelişen dijital çağ ile beraber müşteriler, firmalar ile birden fazla kanaldan etkileşime girerek daha karmaşık müşteri yolculukları oluşmasına sebep olmuştur (Verhoef, Lemon, 2016). Bu nedenle müşteri yolculuğu analitiğinin popülerliğinin ve ihtiyacının artacağını düşünebiliriz.

Müşteri yolculuğu analitiği, müşteri yolculuklarının nasıl ilerlediklerinin analizini, müşterilerin bulunduğu adımlara göre ihtiyaçlarını değerlendirmeyi ve bunları müşterilere sunmayı amaçlar. Müşteri yolculuğu analitiği, kullanıcıları için müşteri deneyimini anlamayı ve bu sayede belirli saptanan etkileşim noktalarında müşterilerine kampanya sunmayı hedefleyen şirketler için oldukça önemlidir. Müşteri yolculuğu analitiği ile şirketler, her önemli müşteri etkileşiminde görünürlük kazanır olmuştur (Verhoef, Lemon, 2016). Örneğin, bazı şirketler müşteri memnuniyeti anketi yolu ile müşteri geri bildirimini sağlamaya çalışırken, çoğu şirket bunu yapmayı beklediği dönüşleri alamayıp müşterilerinin hangi nedenden vazgeçtiğini öğrenemeyebilir (Verhoef, Lemon, 2016). Bu da şirkete maliyet olarak geri dönmektedir. Çünkü, bir müşteriyi elinde tutmak, yeni bir müşteri kazanmaktan çok daha az maliyetlidir (Dahiya, 2015).

Bir müşterinin bir şirket ile etkileşime geçmesini görselleştirmenin en iyi yolu, müşteri serüvenini, görsel bir yolculuk haritası şeklinde sunmaktır. Bu görselleştirmeye, müşteri yolculuğu haritalaması adı verilir. Bir müşteri yolculuğu haritası hazırlamak, müşterilerin şu an da hangi adımda olduğunu, hedeflere ulaşmış olup olmadığını görmek için ilk temas noktasından son temas noktasına kadar mevcut süreci haritalandırmaktır. Müşteri yolculukları farklı alternatifler, farklı davranışlar ve farklı kanallar içerdiğinden her zaman doğrusal olmayabilir. Yolculuklar genelde ileri, geri; sürekli bir karar aşaması içeren (satın alma ya da vazgeçme) aşamalardan ve bu kararlar arasında döngüsel bir tekrar içerebilen çok kanallı yolculuklardan oluşur. Müşteri yolculuklarında çeşitlilik, özellikle teknolojik gelişmeler ile birçok farklı ortamlardan gelebilecek müşteri temas noktalarını ile birlikte artmaktadır. Bu nedenle yolculuk haritalarını oluşturmak zor olabilir. Bununla birlikte haritalar oluşturulmadan önce, veri toplanması gerekmektedir.

Müşteri yolculuğu analitiğinde, yapay zekâ önemli bir yere sahiptir. Müşteri yolculuklarını iyileştirmek, karmaşıklığı ve büyük veriyi kullanmak yapay zekâ ile mümkündür. Yapay zekâ devasa veri yığınlarının işlenmesi ve analiz edilmesinde, bir zamanlar işin içinden çıkılması zor gibi gözükken veri tabanı bilimi ve yürütme arasındaki boşluğu doldurabilir. Hem yapılandırılmış veriler hem de yapılandırılmamış veriler, yapay zekâ tarafından çok hızlı bir şekilde yorumlanabilir (Tanveer, Khan, Ahmed, 2021).

Müşteri yolculuğu analizi ile ilgili yapılan çalışmalar, başlık olarak literatürde iş geliştirme, süreç madenciliği, müşteri deneyimi ve yapay zekâ destekli pazarlama başlıkları adı altında görülmektedir (Lemon, Verhoef, 2016; Terragni, Hassani, 2018; Bernard, Andritsos, 2018; Zaki, Neely, 2019; Tanveer, Khan, Ahmed, 2021). Yapay zekâ destekli pazarlama çalışması sonucunda, iş geliştirme için farklı pazarlama hizmetlerinin yapay zekâ üzerindeki etkisi test edilmiş, sonuçların güvenilir ve sonraki analizler için kullanılabilir olduğu öne sürülmüştür (Tanveer, Khan, Ahmed, 2021). Süreç madenciliği çalışması sonucu, müşteri yolculuğu analizinin kişiselleştirilmiş önerileri geliştirmek için; süreç madenciliği algoritmalarına uygun bir formatta ön işleme, filtreleme, model keşfetme ve analiz etme yeteneğine sahip tekrarlanabilir yapı oluşturulduğu öne sürülmüştür (Terragni, Hassani, 2018). Müşteri deneyimi çalışması sonucunda ise müşteri deneyimi verilerinin müşteri deneyimi analitiği aracılığı ile müşteri deneyimi öngörülerinin geliştirildiği kapsamlı ve bütüncü bir çerçeveye sağlandığı öne sürülmüştür (Holmlund, Vaerenbergh, Ciuchita, Raval, Sarantopoulos, Ordenes, Zaki, 2020).

Yapay zekânın pazarlama stratejileri ve özellikle müşteri davranışları üzerindeki etkisini anlaması değerlidir (Agrawal, Gans, Goldfarb, 2019); Yapay zekânın pazarlamadaki yeri, artan bilgi işlem yeteneği, donanım maliyetlerinin düşürülmesi, büyük verilere erişilebilirlik ve makine öğrenimi algoritmaları ve modellerinin artan etkinliği ile şu an da hız kazanmaktadır.

Literatürde müşteri yolculuk haritaları oluşturmak ve müşteri yolculuğu analizini gerçekleştirmek için veriye dayalı bir yöntem halen eksik (Terragni, Hassani, 2018). Bu çalışmada, müşteri yolculukları incelenerek yolculuğu

gerçekleştiren müşterilerin öznelikleri dikkate alınır. Öznelikler kullanılarak makine öğrenmesi yardımı ile sisteme yeni dahil olan bir müşteri için yolculuk rotası oluşturulup, müşterinin birkaç adım sonra gerçekleştirmesi muhtemelen olayların tahmini yapılacaktır. Bu model sadece yeni müşteriler için değil, sistemde zaten var olan müşteriler için de yolculuk öngörüsünde bulunacaktır. Modelin çalışması için önce aktiviteler arası geçiş olasılıkları hesaplanıp bir aktiviteden sonra müşterinin önündeki diğer aktivite seçenekleri belirlenir. Daha sonra her bir aktivite için modeller kurulur ve bu modeller birbirine bağlanarak yolculuk tahminleri yapabilecek hale getirilir. Çalışma, CRISP-DM metodolojisi aşamalarına göre ilerlemiş, veri üzerinde bu metodolojiye göre işlemler yapılmıştır. Model, veri setinden kaynaklanan ve müşteriler arasındaki özelliklerin çok açık farklılıklarının sonucunda yapılan tahminler her bir modele göre değişmektedir.

2. Veri Madenciliği ve Müşteri Yolculuğu Analitiği (Data Mining and Customer Journey Analysis)

Veri madenciliği; önceden bilinmeyen, geçerli ve uygulanabilir veri yığınlarından, bilgi elde edilmesi süreci olarak tanımlanabilir (Baykal, 2006). Bu süreçte veriler üzerinde çözümlenmeler yapılarak bilgiye ulaşılmaya çalışılır. Veri madenciliği milyonlarca veri ve çok sayıda değişken ile ilgilidir (Coşlu, 2013).

Müşteri yolculuğu analitiğinde daha verimli ve doğru sonuçlara ulaşabilmek adına veri madenciliğinden faydalanılmış, müşteri yolculukları belirlenmiş ve yolculukları oluşturan müşteri özelliklerine göre hem sisteme yeni dahil olabilecek müşteri için yolculuklar tahmin edilmiş, hem de zaten sistemde olan müşteriler için de sonraki adımlar makine öğrenmesi modeli sayesinde tahmin edilmiştir.

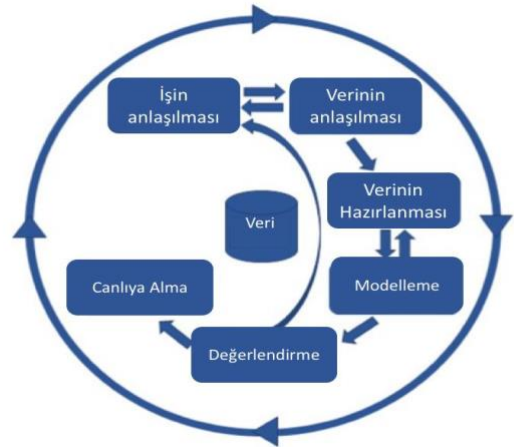
2.1. Metodoloji

Araştırma metodolojisi, çalışma ile ilgili araştırmalardan belirli bilgileri seçmek, ilişkilendirmek, gözlemek ve analiz etmek için kullanılan strateji olarak belirlenir. Bu çalışmada araştırma ve uygulama sürecinde CRISP-DM metodolojisi kullanılmış, veri üzerindeki işlemler bu metodolojiye göre yapılmıştır.

CRISP-DM metodolojisi veri madenciliği projelerini planlama ve geliştirmede kullanılan bir süreç metodolojisidir.

Veri madenciliği sürecinde ve çalışmanın tamamında CRISP-DM metodolojisine takip edilmiş, aşağıdaki metodoloji adımları sıra ile uygulanmıştır.

1. İşin Anlaşılması (Business Understanding)
2. Verinin Anlaşılması (Data Understanding)
3. Verinin Hazırlanması (Data Preparation)
4. Modelleme (Modeling)
5. Değerlendirme (Evaluation)
6. Canlıya Alma (Deployment)



Şekil 1: CRISP-DM Diyarımı

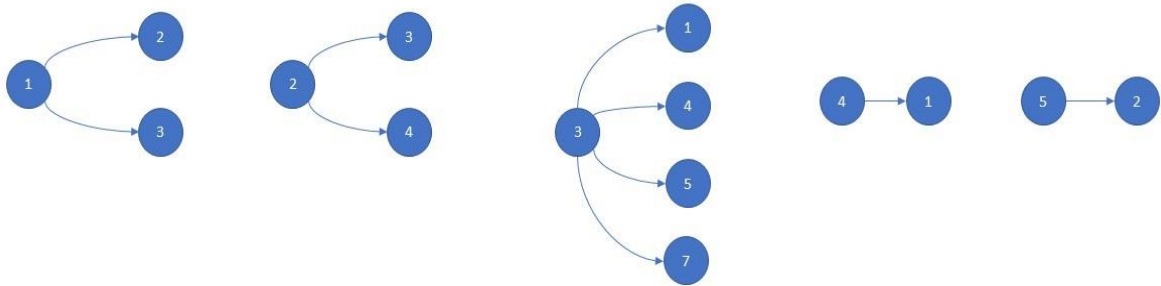
3. CRISP-DM (Cross-Industry Standard Process for Data Mining) Metodolojisinin Müşteri Yolculukları Verisine Uygulanması

Veri madenciliğinin birçok farklı disiplini barındıran yapısı ve uygulama alanlarındaki çeşitliliği, veri büyüklüğü sebebi ile farklı ve kirli veri kaynakları ile çalışmadaki zorluklardan dolayı standart bir metodolojiye ihtiyaç

duyulmaktadır. CRISP-DM metodolojisi, Daimler Chrysler AG, SPSS, NCR ve OHRA gibi lider veri madenciliği kullanıcıları ve onların tedarikçilerinden oluşan bir topluluk tarafından geliştirilmiş ve dünyanın en büyük veri madenciliği çözümlerinin kullanıldığı adımlardan oluşan bir süreci ifade eder (Dolgun, Ersel, 2014). Çalışmada müşteri yolculukları analitiği için bütün adımlar sırası ile takip edilip, gereken veri madenciliği çalışmaları Python'da Spyder ortamında yapılmıştır.

3.1. İşin Anlaşılması (Buisness Understanding)

Müşteri yolculukları analizi yapılarak, müşterilerin gelecekte yapabileceği eylemleri önceden sezerek buna göre bir kampanya ya da süreç planlaması belirlenmesi amacı ile hem müşteri memnuniyetini hem de firmaların finansal memnuniyetini arttırmak hedeflenmiştir. Bu analiz ayrıca, eğer bir müşteri kaybı yaşanacaksa bunun hangi adımda ve ne kadar sürede gerçekleşeceği gibi geri dönüşleri de keşfetme ve bunlara göre yeni planlamalar yapma imkânı sunmaktadır. Çalışmada kullanılan veri setine göre şekil 2'de müşterilerin davranışlarına göre geçiş yapabileceği durumlar gözükmemektedir.

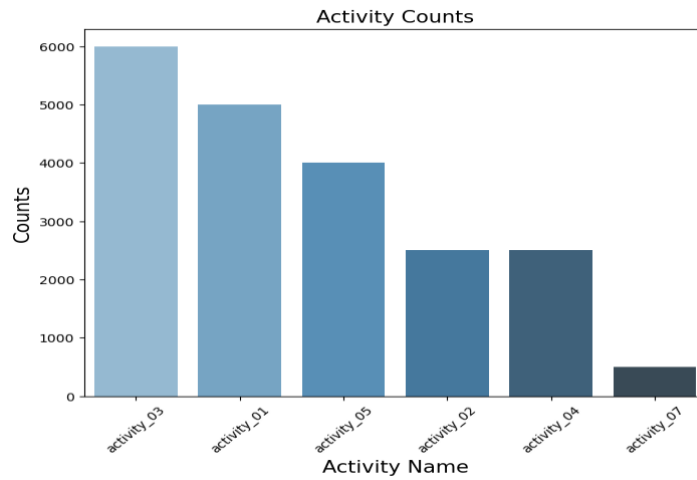


Şekil 2: Aktivitelerin birbirlerine geçiş durumları

Veri setinde müşterileri tanımlayan özelliklere göre tespit edilen serüvenler bulunmaktadır. Bu serüvenleri, şekil 2'de gözüktüğü gibi aktiviteler arasında gerçekleşen geçişler ve müşterileri tanımlayan özellikler oluşturur. Sayı ile ifade edilen her bir durum müşterinin yaptığı bir aktivitedir (Activity_01, Activity_02, ...). Aktivitelerin birbirleri arasındaki geçişlerine, müşterinin özelliklerine göre karar verilmektedir. Bu karar, makine öğrenmesi kullanılarak; müşterilerin özellikleri ve aktivitelerin geçiş olasılıkları göz önünde bulundurularak verilir. Her bir geçişte yine müşteri özelliklerine göre yeni bir karar verilecek şekilde müşteri serüvenleri oluşmaya başlar. Oluşan bu serüvenler sayesinde artık bir müşteri yolculuğundan bahsedilebilir. Müşteri yolculukları oluşturulduktan sonra farklı amaçlara göre; kayıpların olduğu durumları öğrenme ve o duruma gelindiğinde farklı bir kampanya uygulama veya 3 numaralı durumdan sonra en fazla seçilen durumu keşfetme ve yine buna göre karar alma süreçleri gibi durumlar için kullanılabilir.

3.2. Verinin Anlaşılması (Data Understanding)

Öncelikle müşteri yolculuğu veri setine customer-journey.me sitesinden ulaşılmıştır. Veri seti; trace_id, activities, employed, age ve income gibi bilgileri içeren 5 farklı sütundan ve 2.500 satırdan oluşmaktadır. Şekil 2'de veride bulunan aktivitelerin gerçekleşme sayısı gösterilmiştir.

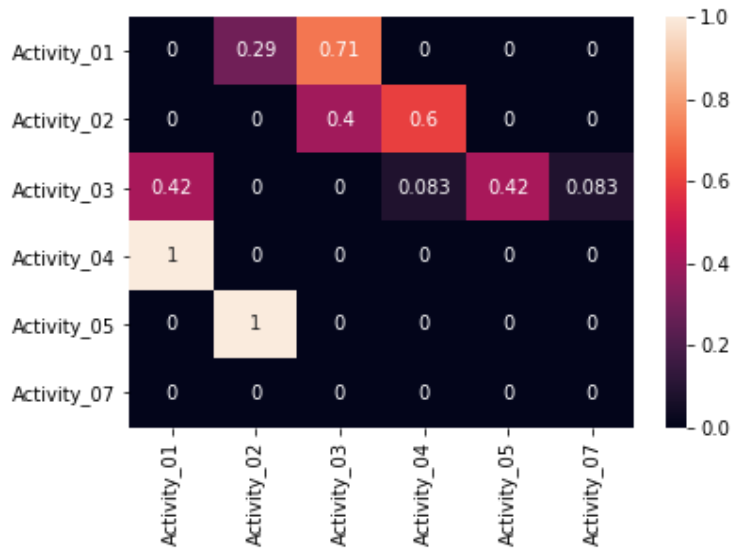


Şekil 3: Verinin aktivite dağılımı

Veri setinin içerdiği veri tipleri kategorik verilerden oluşmaktadır. Veride bulunan aktivitelerin birbiri ile arasında bulunan olasılık değerleri, hangi aktiviteden sonra hangisinin geleceği *Bayes teoremi* ile denklem 1'deki formüle göre hesaplanmış, olasılık değerleri bu koşullu olasılık ilkesine göre bulunmuştur. Şekil 4'de aktivitelerin olasılık matrisi gösterilmiştir.

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Denklem 1: Bayes teoremi formülü



Şekil 4: Aktivitelerin birbiri ardına gerçekleşme olasılığı

Olasılık dağılımları ve aktiviteler arası geçişleri belirledikten sonra modelleme aşamasında her bir aktivite için, bir diğerine geçmesinin tahmini yapılacak ve bu model tahminleri, birbirine bağlanarak müşteri yolculukları belirlenecektir.

3.3. Verinin Hazırlanması (Data Preparation)

Veri hazırlama, veri bilimi çalışmalarının en dikkat edilmesi gereken, en fazla zaman harcanan ve en önemli aşamalarından birisidir (Kılınç ve Başeğmez, 2018:3). Bu süreçten sonra, veriler analiz edilebilir ve sonuçlara ulaşılabilir hale getirilir (Dasu ve Johnson, 2003:1). Veri kalitesi veri madenciliğinde değerli bir konudur. Hazırladığımız veri, eğitilecek olan algoritmanın temel yapı taşı olacaktır (Kılınç ve Başeğmez, 2018:3).

Veriyi hazırlama aşamaları bu çalışmada verilerin birleştirilmesi, verilerin temizlenmesi ve verilerin dönüştürülmesi olmak üzere üç adettir.

3.3.1. Verilerin Birleştirilmesi ve Temizlenmesi

Bu iki aşamanın tek bir başlıkta verilmesinin nedeni farklı farklı dosyalarda bulunan verilerin bir araya getirilmesi, her verinin kendi içerisinde eksik, gürültülü verilerin bulunmasından kaynaklanmaktadır. Bu nedenle Python'da Spyder ortamı kullanılarak önce veriler tek bir veri seti olacak şekilde birleştirilmiş daha sonra ise veri temizleme aşamasına geçilmiştir.

Veri temizleme aşaması veri madenciliği sürecinin en zaman alıcı ve önemli aşamasıdır (Aydoğan, Gencer, Akbulut, 2008). Çalışmada eksik ve gürültülü veriler temizlenmiş ve bir sonraki adım olan veri dönüşümü için hazır hale getirilmiştir.

3.3.2. Verilerin Dönüştürülmesi

Verilerin dönüştürülmesi adımı özellikle makine öğrenmesinin veriyi anlaması için çok önemli bir adımdır. Çalışmada kullanılan veri setinin tüm alanları kategorik değişkenlerden oluştuğu için bunları makinenin anlayacağı dile; yani 1'ler ve 0'lar ile ifade edilecek hale getirilmiştir. Bu yapılan işleme encoding işlemi denilir. Bu aşamada en önemli kısımlardan bir diğeri ise işe yarayan verilen dönüştürülmesidir. Veri setinde kişiye özel, tekil değere sahip olan trace_id alanı bu aşamada silinmiş, herhangi bir encoding işlemine sokulmamıştır. Bu aşamada, modüler bir yapı izlenmiş olup, her aktivite için ayrı ayrı veri önileme ve veri dönüşümü uygulanmamış, bunun yerine girilen aktivite parametresine göre dinamik bir yapı ve modülerlik kullanılmıştır. Bu sayede girilen aktivite parametresine göre; aktiviteden sonra gelebilmesi ihtimali olan tüm aktiviteler "target" adında yeni bir alana yazılmıştır. Yazılacak olan aktiviteler sayılarına göre farklılık göstermektedir. Yani eğer bağımlı değişkene yazılacak değer sayısı 2'den büyükse çoklu sınıflandırma, değilse ikili sınıflandırmaya göre encoding işlemlerinden geçeceklerdir. Daha sonra alandaki aktivite numaralarına göre Numpy kütüphanesi yardımı ile encoding işlemi yapılmıştır. Burada SKLearn kütüphanesi altında LabelEncoder kullanılarak da bu işlem yapılabilirdi, fakat bu işlem her aktivite için ayrı modeller kurularak ve daha sonra o modellerin birbirine bağlanarak yolculuk tahmini yapılacağı için aktivite numaralarına göre encode edilmiş ve bu sayede daha okunabilir ve anlaşılır hale gelmiştir.

Hedef alanına girilen aktiviteye göre değerler yazıldıktan sonra, bağımsız değişkenler için de encoding işlemleri yapılmıştır. Bu işlemler SKLearn kütüphanesi altında bulunan LabelEncoder ve OneHotEncoder sınıfları yardımı ile yapılmıştır. Kategorik verileri model için anlaşılır hale getirecek şekilde LabelEncoder ile nümerik değişkenlere dönüştüreceğiz. OneHotEncoder ise kategorik değişkenlerin bulunduğu sütunu alır ve birden çok sütuna böler. İçerisinde iki farklı parametre bulunduran alan olan employed LabelEncoder yardımı ile içerisinde ikiden fazla parametre bulunduran alanlar; age ve income OneHotEncoding yardımı ile encode edilmiştir.

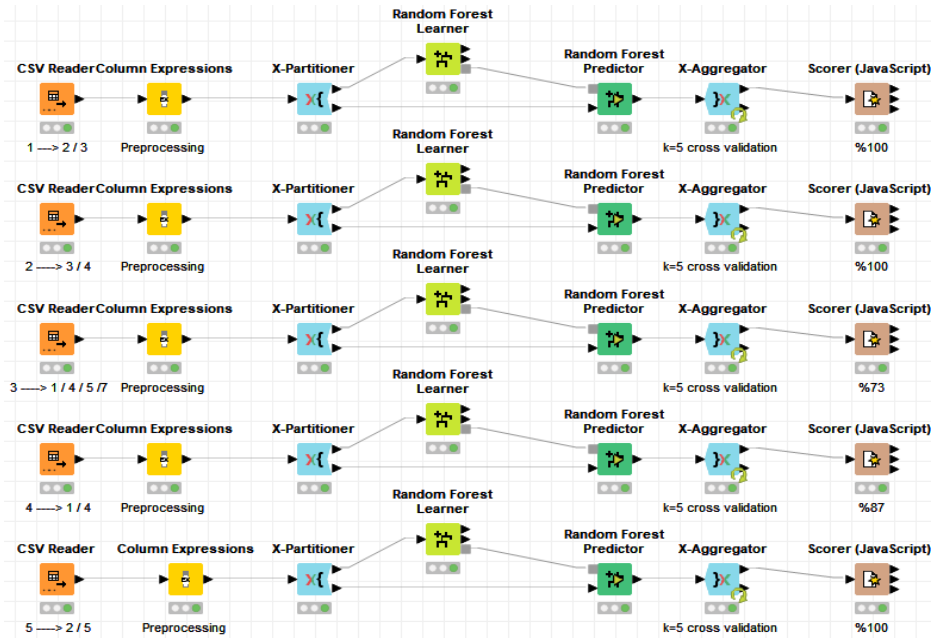
Verilerin dönüştürülmesi işlemi başarı ile tamamlandığında artık veri makine öğrenmesi algoritmaları için hazır durumdadır. Veri ön işleme ve veri dönüştürme aşamasından sonra veri şekil 5'deki hale getirilmiştir.

Index	employed	0-19yo	20-39yo	40-59yo	60-79yo	80yo+	high	low	middle	target
0	0	0	0	0	0	1	0	0	1	3
1	0	0	0	0	0	1	0	0	1	3
2	0	0	0	0	0	1	0	0	1	3
3	0	0	0	0	1	0	1	0	0	2
4	0	0	0	0	0	1	0	0	1	3
5	0	0	0	0	0	1	0	0	1	3
6	0	0	0	0	0	1	0	0	1	3
7	0	0	0	0	1	0	1	0	0	2
8	0	0	0	0	0	1	0	0	1	3
9	0	0	0	0	1	0	1	0	0	2
10	1	0	0	1	0	0	0	0	1	2
11	0	0	0	0	1	0	1	0	0	2
12	0	0	0	0	0	1	0	0	1	3
13	1	1	0	0	0	0	0	1	0	2
14	0	0	0	0	1	0	1	0	0	2

Şekil 5: Veri ön işleme sonrası veri seti

3.4. Modelleme (Modeling)

Modelleme aşamasında öncelikle hesaplanmış olduğumuz olasılık değerleri ve hangi aktiviteden sonra hangilerinin gelebileceği durumlarına göre verinin hazırlanması aşamasında elde ettiğimiz işlenmiş veriyi Knime Analytics Platform kullanarak her bir aktivite için modeller kurulmuş ve modellerin başarıları gözlemlenmiştir. Burada amaç öncelikle her bir aktivite için kurulan ayrı ayrı modellerin başarılarının tatmin edici olup olmadığına bakılmasıdır. Bu kontrolden sonra modeller müşteri yolculuğu oluşturması için birbirine bağlanarak yolculuk oluşturmuştur. Şekil 6'de Knime Analytics Platform kullanılarak oluşturulan modeller ve model başarıları gözükmektedir.



Şekil 6: Knime Analytics Platform kullanılarak oluşturulan modeller

Modeller kurularken sınıflandırma algoritmalarından her biri denenmiştir. Veri setinde bazı parametreler ve aktiviteler arasında açık ve net bir ayırım olduğundan dolayı hemen hemen bütün algoritmalarda aynı sonuçlara rastlanmış ve pek bir değişiklik gözlemlenmemiştir.

Modelleme Aşamasında Knime ile Kullanılan Sınıflandırma Algoritmaları:

1. Karar Ağaçları (Decision Trees): Karar ağaçları, sınıflandırma ve tahmin için sıkça kullanılan bir makine öğrenmesi algoritmasıdır. Kolay yorumlanması ve anlaşılabilirliği açısından avantaj sağlamaktadır (Çalış, Kayapınar, Çetinyokuş, 2014).
2. K En Yakın Komşu (K Nearest Neighbor): İsminde de anlaşılacağı üzere veriyi uzaklık olarak en yakınında bulunan diğer komşularına ve bu komşuların sayısına (k değerine) göre sınıflandırır. KNN algoritması, yeni verilerin hızla geldiği ve eğitim kümesinin hızla değiştiği durumlarda öteki algoritmalara göre daha iyi bir sınıflandırma işlemi yapabilir (Dilki, Başar, 2020).
3. Destek Vektör Sınıflandırıcı (Support Vector Classifier (SVC)): SVC, verilen değerlerin çıktısını sınıflandırmak için SVC Learner düğümü tarafından oluşturulan bir SVC modeli ve algoritmasını kullanır.
4. Lojistik Regresyon (Logistic Regression): Çok değişkenli bir lojistik regresyon modeli uygular.
5. Rassal Orman (Random Forest): İçerisinde belirlenen sayıya ve derinliğe göre karar ağaçları bulunduran ve oy birliğine göre sınıflandırma yaparak tahminler üretir.

Uygulanan sınıflandırma algoritmaları arasındaki başarı sonuçları birbirlerine yakın olmakla beraber en yüksek başarıya Rassal Orman algoritması ulaşmıştır. Çalışmanın Python kısmında Rassal Orman algoritması tercih edilerek modeller Rassal Orman algoritması ile birbirlerine bağlanmıştır. Tablo 1’de kullanılan algoritmaların başarılarının karşılaştırılması gözükmektedir.

Tablo 1: Knime Analytics Platform Algoritma Başarıları

Algoritma	Doğruluk % (Accuracy)	Hata Oranı % (Error Rate)
Decision Tree	86.76	13.24
KNN	87.34	12.66
SVC	86.44	13.56
Logistic Regression	86.12	13.88
Random Forest	87.36	12.64

Görüldüğü üzere en başarılı algoritma Random Forest algoritmasıdır. Bu nedenle model Python’da SKLearn kütüphanesinin ensemble modülünde bulunan RandomForestClassifier sınıfı yardımı ile Random Forest ile göre kurulup, kriter (criterion) ‘gini’ ve ‘entropy’ olarak 2 farklı şekilde denenmiş, ‘entropy’ kriteri daha başarılı tahminler yaptığı için tercih edilip, çalıştırılmıştır. Ayrıca yine SKLearn kütüphanesinin multioutput modülünde bulunan MultiOutputClassifier sınıfı yardımı ile birbirinden ayrı ayrı model kurulan her bir aktivite modelleri için bağımlı değişkenlerinin sayısına göre çoklu sınıflandırma veya ikili sınıflandırma yapılmıştır.

Değerlendirme bölümünde daha sık işleneceği ön bilgisi ile modelin başarı çıktıları için accuracy score, confusion matrix ve cross-validation yöntemleri kullanılmıştır.

Oluşturulan model ile artık sisteme yeni dahil olacak bir müşterinin yolculuğu ön görülebilir olmuş, bulundurduğu parametrelere göre birkaç adım sonrasında nerede olacağını, nerelere gidebileceğinin tahmini yapılmıştır.

Gelen her bir yeni müşteri yazılan forecast fonksiyonu içerisine; müşteri özelliklerini ifade edecek parametreyi, gelen müşterinin kaç adım sonrasında nerede olacağını görmek istediğimiz adım sayısını ve makine öğrenmesi modellerine göre sürekli güncellenerek yeni yolculuk hedefini belirleyen bir değişkeni alır. Yolculuk

tahminleri bir listede tutulur ve sisteme yeni dahil olan müşterinin yolculuk listesi dolu olmayacağı için öncelikle gelen müşteri özelliklerine göre bir tahmin yapılır. Yapılan tahmin sonucunda bulunan ve müşterinin yönelmesi beklenen yeni aktivite (activity_01, activity_02, activity_03, activity_04, activity_05, activity_07) ve bu aktivite ile bağlantısı olan (hesaplanan olasılık tablosunda her aktivite için birbirleri arasındaki bağlantı ve birbirlerine geçiş olasılıkları mevcuttur.) diğer aktivitelere bakılması amacıyla tahmini yapılan aktivitenin veri çerçevesine gidilir. Burada müşterinin özellikleri aranır ve benzer özelliklerde öteki müşterilerin yaptığı ve yeni gelen müşterinin de yapması muhtemel olan yeni hedef aktivite seçilir ve listeye eklenir. Bu süreç girdiğimiz adım sayısı sona erene kadar devam eder. Eğer girilen müşteri parametreleri veya girilen adım sayısı hatalı, eksik veya yanlış ise (Örneğin bir kişi hem 30 yaşında hem de 40 yaşında olamaz veya -1 adım sayısı diye bir öngörü olamaz) yolculuk sistemi çalışmaz.

Modelleme aşamasında modeller kurulup, yapı oluşturulmuştur. Şimdi ise modelin değerlendirilmesi ile modelin başarıları hakkında konuşulması gereklidir.

3.5. Değerlendirme (Evaluation)

Modelin test veri setini ne ölçüde sınıflandırıldığını gösteren karmaşıklık matrisi sonuçları algoritmanın değerlendirilmesi açısından önemlidir. Bunun için modelimizden örnek olarak activity_04 veri çerçevesi ve ondan sonra gelebilecek aktiviteler için modelin karmaşıklık matrisi başarısına bakabiliriz.

Tablo 2: Karmaşıklık matrisi (Confusion Matrix)

	Öngörülen Negatif Sınıf	Öngörülen Pozitif Sınıf
Gerçek Negatif Sınıf	TN: 1202	FP: 81
Gerçek Pozitif Sınıf	FN: 0	TP: 70

True Negative (TN): Gerçek Negatifler; modelin 0 diye tahmin ettiği ve gerçekte de 0 olan tahminler.

False Negative (FN): Yanlış Negatifler; modelin 0 diye tahmin ettiği ve gerçekte 1 olan tahminler.

False Positive (FP): Yanlış Pozitifler; modelin 1 diye tahmin ettiği ve gerçekte 0 olan tahminler.

True Positive (TP): Gerçek Pozitifler; modelin 1 diye tahmin ettiği ve gerçekte de 1 olan tahminler.

Model Duyarlılığı (Sensitivity/Precision): $TP/(TP + FN) = 1$

Model Belirleyiciliği (Specificity): $TN/(TN + FP) = 0.93$

Model Kesinliği (Recall): $TP/(TP + FP) = 0.46$

Modelin Yanlış Sınıflandırma Oranı (Misclassification Rate): $((FP + FN))/TOPLAM = 0.07$ (Error Rate)

F-skoru: $(2 \times Model\ Duyarlılığı \times Model\ Kesinliği)/(Model\ Duyarlılığı + Model\ Kesinliği) = 0.63$

Modelin Doğruluğu (Accuracy): $(TP + TN)/TOPLAM = 0.94$

Karmaşıklık matrisinde negatif 0, pozitif 1 olarak temsil edilmektedir. Fakat biz activity_04 veri çerçevesini ve oradan gidebilmesi mümkün olan aktiviteleri örnek olması için inceledik. Yani activity_04' ten activity_01'e

ve yine kendi üzerinde durması ihtimali anlamında olan activity_04'e olmak üzere iki farklı yol bulunmaktadır. Burada bahsi geçen negatif kısım activity_01'i, pozitif kısım ise activity_04'ü temsil etmektedir.

Yukarıda bulduğumuz değerleri önümüze tek seferde sunan ve her seferinde bunları tek tek hesaplama zahmetinden kurtulmamızı sağlayan ve yine SKLearn kütüphanesinin metrics modülünde bulunan classification_report sınıfını Python kodumuza dahil ettikten sonra modelin sınıflandırma raporunu hesaplayarak bu değerleri önümüze sunmaktadır.

Tablo 3: Sınıflandırma Raporu (Classification Report)

	precision	recall	f1-score	support
1	1.00	0.94	0.97	1283
4	0.46	1.00	0.63	70
accuracy			0.94	1353
macro avg	0.73	0.97	0.80	1353
weighted avg	0.97	0.94	0.95	1353

3.6. Canlıya Alma (Deployment)

CRISP-DM metodolojisinin son aşaması olan canlıya alma yani uygulama aşamasında; müşterilerinin yolculuklarının belirlenmesini isteyip buna göre aksiyon almak isteyen bütün firmalar için değerlendirilebilir ve önem teşkil eden bir süreçtir. Bu çalışmada uygulanan yöntemler ve izlenen yol yardımı ile firmaların hem eski müşterilerinin oluşturduğu hem de gelecekte yeni katılan müşterilerin yolculuklarının belirlenmesi firmanın aksiyon alacağı noktalarını belirleyebilir ve bu durum iki taraf içinde zamansal, firmalar için ise finansal açıdan yarar sağlayabilir.

4. Sonuç

Yapılan çalışmada veri madenciliği teknikleri ve CRISP-DM metodolojisi uygulanarak, firmaların bünyesinde bulundurduğu müşterilere ve daha sonra kendisine katılabilecek olan müşterilerin yolculuklarının belirlenmesi amacı ile makine öğrenmesi modelleri geliştirilmiş, modellerin sonuçlarına ve başarısına bakılmış daha sonra yolculuklar oluşturulması amacı ile modeller birbiri ile ilişkilendirilmiştir. Knime Analytics Platform kullanılarak veri setleri üzerine farklı sınıflandırma algoritmaları (Decision Tree,SVC, Logistic Regression, Random Forest, KNN) uygulanmıştır. Uygulanan algoritmalar arasından başarısı en yüksek olan algoritma Random Forest algoritması olarak saptanmıştır. Daha sonra Python'da SKLearn çerçeve yazılımında, aktiviteler arası ilişkilerden yola çıkarak aktivitelerin birbiri arkasına gelmesi olasılığı hesaplanmış, bu hesaplamalar ile hangi aktiviteden sonra hangisinin gelebileceği hakkında aktiviteler ve olasılık tablosu oluşturulmuştur. Elde edilen bu bilgilerden her bir aktivite için Random Forest algoritması kullanılarak her bir aktivite için ayrı ayrı modeller oluşturulup birbiri ile ilişkilendirilmiştir. Oluşturulan modeller hem veri setinin barındırdığı keskin ayırt edici parametreler hem de yapay bir veri seti olması nedeni ile tahminlerin başarıları %100'e yakın, sadece çok sınıflı sınıflandırma modelinde başarı %75 bulunmuştur Modellerin ilişkilendirilmesi sonucu sisteme yeni dahil olan bir müşteri için müşteri yolculuğu rotası model tarafından, istenilen adım sayısına göre yapabileceği müşteri yolculuğunun tahmini yapılmıştır.

Birçok veri seti dahil olmak üzere müşteri kaybı analizi, müşteri skorlaması, risk skorlaması, market skorlaması, satış ön görü tahminleri ve fiyat optimizasyonu belirleme gibi önemli konularda kullanılan, SKLearn kütüphanesinde kullanılan modellerden bağımsız, çalışmanın başarısını arttırarak daha doğru sonuçlar almak için OptiScorer motoru

kullanılmasının diğer algoritmalara göre daha başarılı olduğu gözlemlenmiştir (Şeker, 2020). Motorun kullanıldığı çalışmalara bakılarak başarı farkları gözlemlenebilir.

Elde edilen sonuçlar dahilinde müşterilerinin hareket edebileceği aksiyon noktalarını değerlendirmek isteyen firmalar için müşteri yolculuklarının analizi (Customer Journey Analytics) zahmetli, değerli ve dikkate alınması tavsiye edilen bir süreçtir. Bu sürecin düzgün bir şekilde uygulanması durumunda firmalar kaybettikleri müşteriler için kayıp olan aksiyon noktalarını belirleyip üzerinde çalışabilir, şayet ortada bir kayıp yoksa daha fazla kazancı sağlamak amacı ile bir sonraki müşteri aktivitesi için aksiyon alıp kampanyalar düzenleyebilirler. Makale; sadece belirli bir finans kuruluşu ya da belirli firmalar için değil, hizmeti insan olan ve ellerinde karmaşık sayılabilecek bir müşteri trafiği barındıran çoğu kurum ve kuruluşlara da uygulanabilir ve tahminlerin daha başarılı, düzgün olması yolunda gelecekteki çalışmalara katkı sağlayabilir.

Kaynakça

- Berry, L. L., Wall, E. A., & Carbone, L. P. (2006). Service clues and customer assessment of the service experience: Lessons from marketing. *Academy of management perspectives*, 20(2), 43-57.
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of marketing*, 80(6), 69-96.
- Dahiya, K., & Bhatia, S. (2015, September). Customer churn analysis in telecom industry. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-6). IEEE.
- Tanveer, M., Khan, N., & Ahmad, A. R. (2021, April). AI Support Marketing: Understanding the Customer Journey towards the Business Development. In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) (pp. 144-150). IEEE.
- Terragni, A., & Hassani, M. (2018, August). Analyzing customer journey with process mining: From discovery to recommendations. In 2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud) (pp. 224-229). IEEE.
- Bernard, G., & Andritsos, P. (2018, June). Cjm-ab: Abstracting customer journey maps using process mining. In International Conference on Advanced Information Systems Engineering (pp. 49-56). Springer, Cham.
- Zaki, M., & Neely, A. (2019). Customer experience analytics: dynamic customer-centric model. In Handbook of Service Science, Volume II (pp. 207-233). Springer, Cham.
- Holmlund, M., Van Vaerenbergh, Y., Ciuchita, R., Raval, A., Sarantopoulos, P., Ordenes, F. V., & Zaki, M. (2020). Customer experience management in the age of big data analytics: A strategic framework. *Journal of Business Research*, 116, 356-365.
- Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: the ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2), 31-50.
- Baykal, A. (2006). Veri madenciliği uygulama alanları. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, (7), 95-107.
- Coşlu, E. (2013). Veri madenciliği. *Akademik bilişim*, 23-25.
- Dolgun, M. Ö., & Ersel, D. (2014). Doğrudan pazarlama stratejilerinin belirlenmesinde veri madenciliği yöntemlerinin kullanımı. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 7(1), 1-13.
- Kılınç, D., & Başeğmez, N. (2018). *Uygulamalarla Veri Bilimi*. İstanbul: Abaküs.
- Dasu, T. ve Johnson, T. (2003) *Exploratory Data Mining and Data Cleaning*, John Wiley & Sons Publication, New Jersey, USA.
- Aydoğan, E. K., Gencer, C., & Akbulut, S. (2008). Churn analysis and customer segmentation of a cosmetics brand using data mining techniques. *Journal of Engineering and Natural Sciences*, 26(1).
- Çalış, A., Kayapınar, S., & Çetinyokuş, T. (2014). Veri madenciliğinde karar ağacı algoritmaları ile bilgisayar ve internet güvenliği üzerine bir uygulama. *Endüstri Mühendisliği*, 25(3), 2-19.

- Dilki, G., & Bařar, Ö. D. (2020). İřletmelerin iflas tahmininde k-en yakın komřu algoritması üzerinden uzaklık ölçütlerinin karşılařtırılması. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 19(38), 224-233.
- Seker, S. E. (2016). Müřteri Kayıp Analizi (Customer Churn Analysis). YBS Ansiklopedi, 3(1), 26-29.
- řeker, ř. E. (2020). OptiScorer: Otomatik Makine Öğrenmesi ile Skorlama.
- AKAR, B. Müřteriye Özel Fiyat Tahmin Çalışması.