

T.C.
BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI
ENDÜSTRİ MÜHENDİSLİĞİ BİLİM DALI

**E-TİCARETTE PAZAR SEPETİ ANALİZİ İLE SATIŞ
YÖNETİMİ**
Yüksek Lisans Tezi

Tezi Hazırlayan
Dilara VARDARLI

İstanbul, 2022

T.C.
BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI
ENDÜSTRİ MÜHENDİSLİĞİ BİLİM DALI

**E-TİCARETTE PAZAR SEPETİ ANALİZİ İLE SATIŞ
YÖNETİMİ**
Yüksek Lisans Tezi

Tezi Hazırlayan
Dilara VARDARLI
1908092010
0000-0001-5193-5816

Dr.Öğr.Üyesi Sabahattin Kerem AYTULUN

İstanbul, 2022

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “E-ticarette Sepet Analizi ile Satış Yönetimi” başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere uygun şekilde tarafımdan yazıldığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, yararlandığım eserlerin tamamının kaynaklarda gösterildiğini ve çalışmanın içinde kullandıkları her yerde bunlara atıf yapıldığını, patent ve telif haklarını ihlal edici bir davranışımın olmadığını belirtir ve bunu onurumla doğrularım. 31.05.2022

Dilara VARDARLI

T.C.
BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜ
TEZLİ YÜKSEK LİSANS SINAV TUTANAĞI

31.05.2022

Enstitümüz *Endüstri Mühendisliği* Anabilim Dalı *Endüstri Mühendisliği* Programı yüksek lisans öğrencilerinden 1908092010 numaralı **Dilara VARDARLI**'nin "*Beykent Üniversitesi Lisansüstü Eğitim – Öğretim Yönetmeliği*"nin ilgili maddesine göre hazırlayarak, Enstitümüze teslim ettiği "*E-Ticarette Pazar Sepeti Analizi İle Satış Yönetimi*" konulu tezini, Yönetim Kurulumuzun 31/05/2022 tarih ve 2022/22 sayılı toplantısında seçilen ve On-Line toplanan biz jüri üyeleri huzurunda, Beykent Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 29. maddesinin 3. fıkrası gereğince 45 dakika süre ile Zoom programı aracılığıyla on-line olarak aday tarafından savunulmuş ve sonuçta adayın tezi hakkında "**OYBİRLİĞİ**" ile "**KABUL**" kararı verilmiştir.

İşbu tutanak, 2 nüsha olarak hazırlanmış ve Enstitü Müdürlüğü'ne sunulmak üzere tarafımızdan düzenlenmiştir.

DANIŞMAN
Dr. Öğr. Üyesi Sa*** Ke*** AY***
(Beykent Üniversitesi)

ÜYE
Dr. Öğr. Üyesi Ke*** Gö*** NA***
(Beykent Üniversitesi)

ÜYE
Doç. Dr. Ar*** ES*** ÖN***
(Marmara Üniversitesi)

Adı ve Soyadı : Dilara VARDARLI
Danışmanı : Dr. Öğr. Üyesi Sabahattin Kerem AYTULUN
Türü ve Tarihi : Yüksek Lisans Tezi, 2022
Alanı : Endüstri Mühendisliği
Anahtar Kelimeler : E-Ticaret, Veri Madenciliği, Pazar Sepet Analizi, Satış Yönetimi

ÖZ

E-TİCARETTE PAZAR SEPETİ ANALİZİ İLE SATIŞ YÖNETİMİ

Çalışmanın temel amacı yenedünya ortamı ile ticaret hayatının internete uyarlanmasında, veri madenciliği yardımı ile satın alınan ürünlerin birbirleriyle olan bağlantılarını araştırmak ve böylece doğru satış stratejileri oluşturmaktadır.

Bu araştırmada, Kaggle sitesinden alınan orta ölçekli bir çevrimiçi mücevher mağazasındaki Aralık 2018 ile Aralık 2021 (3 yıl) arasındaki toplam 95912 veriyi barındıran veri seti kullanılmıştır. Öncelikle veriye bir ön işleme yapılmış, sorunlar bulunarak ortadan kaldırılmış ve veri seti analize hazır hale getirilmiştir. Veri setinden sık ürün setleri ve buna bağlı olarak birliktelik kuralları bulunmuştur. Ayrıca veri setindeki zaman özelliğinden faydalanarak genelleştirilmiş sıra özelliği yardımıyla ürünlerin satın alma sıraları da tespit edilmiştir. Bütün bu çıkarımlar sonucunda pazarlama bölümleri için ürün ve kullanıcı odaklı satış stratejilerinin belirlenmesinde önemli rol oynayan bilgilerin derlenmesi sağlanmıştır.

Name and surname : Dilara VARDARLI
Advisor : Dr. Lecturer Sabahattin Kerem AYTULUN
Type and Date : Master Thesis, 2022
Field : Industrial Engineering
Keywords : E-Commerce, Data Mining, Market Basket Analysis, Sales
Management

ABSTRACT

SALES MANAGEMENT WITH MARKET CART ANALYSIS IN E-COMMERCE

The main purpose of the study is to investigate the connections between the products purchased with the help of data mining and thus to create the right sales strategies in the adaptation of the new world environment and the commercial life to the internet.

In this research, a dataset containing a total of 95912 data from December 2018 to December 2021 (3 years) from a medium-sized online jewelry store from Kaggle was used. First of all, a preprocessing of the data was made, the problems were found and eliminated, and the data set was made ready for analysis. Frequent product sets and association rules were found from the data set. In addition, using the time feature in the data set, the order of purchase of the products was determined with the help of the generalized order feature. As a result of all these inferences, information that plays an important role in determining product and user-oriented sales strategies for marketing departments has been compiled.

İÇİNDEKİLER

Sayfa No:

ÖZ

ABSTRACT

TABLolar LİSTESİ.....	iii
ŞEKİLLER LİSTESİ.....	iv
KISALTMALAR	v
SÖZLÜK	vi
GİRİŞ	1

BİRİNCİ BÖLÜM

1. LİTERATÜR ÇALIŞMASI	2
------------------------------	---

İKİNCİ BÖLÜM

2. VERİ MADENCİLİĞİNE GENEL BAKIŞ	7
2.1. Veri Madenciliği Uygulama Süreçleri	8
2.2. Veri Madenciliği Modelleri.....	9
2.2.1. Tahmin Edici Modeller.....	9
2.2.2. Tanımlayıcı Modeller	9
2.3. Veri Madenciliği Yöntemleri	10
2.3.1. Sınıflandırma Yöntemi	10
2.3.2. Kümeleme Analizi	11
2.3.3. Birliktelik Kuralları	12
2.3.4. FP-Growth Algoritması	13
2.4. Genelleştirilmiş Sıralı Problem Modeli (GSP).....	15
2.5. Sepet Analizi	16
2.6. Veri Madenciliğinin Uygulama Alanları.....	17

ÜÇÜNCÜ BÖLÜM

3. E-TİCARETE GENEL BAKIŞ	19
3.1. E-Ticaretin Türkiye’deki Durumu.....	20

DÖRDÜNCÜ BÖLÜM

4. VERİ MADENCİLİĞİNİN E-TİCARETE UYGULANMASI	24
4.1. Veri Hakkında	24
4.2. Eksik Verilerin Tamamlanması.....	25
4.3. Sık Ürün Setlerinin Bulunması.....	26
4.4. Ürünlerin Satın Alınma Sırasının Belirlenmesi.....	35
SONUÇ.....	40
KAYNAKÇA	41



TABLÖLAR LİSTESİ

Sayfa No.

Tablo 4.1. Veri Setinde Yer Alan Özellikler (Yazar tarafından oluşturulmuştur.)...	24
Tablo 4.2. Farklı Parametreler İçin Sık Ürün Seti Sonuçları (Rapidminer).....	31
Tablo 4.3. Sık Ürün Setleri (RapidMiner)	32
Tablo 4.4. “Create Association Rules” Operatöründe Kullanılan Parametreler (RapidMiner).....	33
Tablo 4.5. “Create Association Rules” Operatörü Sonuçlarının Bir Kısımını	33
Tablo 4.6. “Diffrence” Özelliği Eklenmiş Veri Seti Örneği	35
Tablo 4.7. Veri Setinden Üretilen Pivot Tablo Örneği (RapidMiner)	36
Tablo 4.8. “1” ve “?” Yerine Sırasıyla “True” ve “False” Dönüşümü (RapidMiner)36	
Tablo 4.9. GSP Algoritması Sonuçları(RapidMiner).....	39

ŞEKİLLER LİSTESİ

	Sayfa No.
Şekil 2.1. Veri Uygulama Süreçleri (Http1, 2020)	8
Şekil 2.2. Veri Madenciliği Teknikleri (Http2, 2020)	9
Şekil 2.3. Karar Ağacı (Http3, 2017).....	10
Şekil 2.4. Kümeleme Dağılım Grafiği (Http4, 2018)	11
Şekil 2.5. FP-Growth 1.Adım (Http6, 2020)	14
Şekil 2.6. FP-Growth 2.Adım (Http7, 2020)	14
Şekil 2.7. FP-Growth 3.Adım (Http8,2020)	15
Şekil 2.8. Birlikte Alınma Olasılık Hesaplaması (Http9, 2020)	16
Şekil 3.1. 2017-2018 Yıllık Büyüme Verileri (TÜSİAD,2019:S.Y.).....	20
Şekil 3.2. Perakende Satış Dağılımları (TÜSİAD,2019:S.Y.).....	21
Şekil 3.3. Dünya Geneline E-Ticaret Dağılımı (Http13,2019:S.Y.).....	22
Şekil 4.1. Verideki bilgilerin bir kısmı (RapidMiner)	25
Şekil 4.2. Eksik Veri Tamamlama Akışı (RapidMiner)	26
Şekil 4.3. “Order_Id” Altında Ürünler Birleştirildiğinde Elde Edilen Veri Seti Örneği (İlk 10) (RapidMiner).....	27
Şekil 4.4. FP-Growth Operaötörü Akışı (RapidMiner)	27
Şekil 4.5. Sipariş Başına Satın Alınan Ürün Sayısı (RapidMiner)	28
Şekil 4.6. Filtre Examples Operatörünün Kullanımı Ve Filtre Ekleme (RapidMiner)	29
Şekil 4.7. “FP-Growth Kullanımı (RapideMiner)	29
Şekil 4.8. “FP-Growth” Kullanımı (RapidMiner)	30
Şekil 4.9. “Create Association Rules” Operatörünün Kullanımı.....	34
Şekil 4.10. GSP Operatörü Parametreleri (RapidMiner)	37
Şekil 4.11. Kullanılan Rapidminer İş Akışı.....	39

KISALTMALAR

AIS Algoritması: Agrawal, Imielinski ve Swami Tarafından Geliştirilen Algoritma

GBT : Gradient bosted trees (Karar Ağacı Sınıflandırma Algoritması)

GSP : Genelleştirilmiş Sıralı Problem Modeli

Max Gap : En Fazla Boşluk

Min Gap : En Az Boşluk

TÜBİSAD : Bilişim Sanayiciler Derneği

Usd : Amerikan Doları

SÖZLÜK

E-Ticaret: Elektronik ortamda yapılan ticaret hayatıdır.

FP-Growth Algoritması: Apiori algoritmasının geliştirilmiş halidir. Veri kümelerinde sık görülen kalıpları veya ilişkileri bulan analitik süreçtir.

Genelleştirilmiş Sıralı Problem Modeli (GSP): Paralel doğrular varsayımının reddedildiği durumda kullanılan bağımlı değişkenin sıralı yapısını dikkate alan modeldir.

Grandient Boosted Trees (GBT): Regresyon ve sınıflandırma problemleri için çalışan bir başka topluluk makine öğrenme algoritmasıdır.

Kollektif: Çok fazla kişi ve nesneyi içine alan veya bir araya gelmesi sonucu oluşan durumdur.

Korelasyon: Olasılık ve istatistik kuramının iki ya da daha fazla değişken arasındaki doğrusal ilişkinin yönünü ve gücünün göstermek amaçlı tabir edilebilir.

Literatür: Herhangi bir bilim dalında yazı ve yapıtların tümü anlamına gelir.

Veri Madenciliği: Büyük ölçekli veriler arasından faydalı bilgiye ulaşma, bilgiyi madenleme işidir.

GİRİŞ

Dünya genelinde internet kullanımının yaygınlaşması ve artması, gündelik yaşantıyla birlikte ticaret hayatının da dengelerini değiştirerek, online alışverişin pazar payının günden güne artmasına sebep olmaktadır. Yeme-içme sektöründen, giyim sektörüne, aksesuar sektöründen, mücevherat sektörüne herkes satışlarını e-ticaret sistemine uygun hale getirmektedir. Online ortama Pazar yeri diyecek olursak, bu pazar yerinin çok iyi analiz edilmesi gerekmektedir.

Bu tez çalışmasında, internet üzerinden mücevherat satışı yapan bir işletmenin verilerinden faydalanarak veri madenciliği yöntemleri ile sepet analizi yapılacaktır.

Çalışmanın birinci bölümünde literatür çalışması bulunmaktadır. Veri madenciliği ile detaylı bilgiler ikinci bölümde, e-ticaret ile ilgili bilgilere ise üçüncü bölümde yer verilmektedir.

Dördüncü bölümde ilk olarak tezde kullanılacak veri ile ilgili detaylı bilgi edinilmektedir. Veri setinde eksik olan bilgilere ulaşabilmek için karar ağacı sınıflandırma algoritması (“Gradient Boosted Trees” GBT) yöntemi kullanılmaktadır. Veri setindeki eksikler giderildikten sonra, FP-Growth operatörü yardımıyla sık ürün setleri tespit edilmektedir. Dördüncü bölümün son evresinde ürünlerin hangi sırayla satın alındığı birliktelik kuralları ile bulunmaktadır.

Beşinci sonuç bölümü olarak adlandırmaktadır. Tezin başından sonuna kadar yapılan bütün çalışmaların nereye vardığını anlatmaktadır.

BİRİNCİ BÖLÜM

1. LİTERATÜR ÇALIŞMASI

Bu tez çalışması için yapılan literatür araştırması sırasında e-ticaret sektöründe müşteri sadakatini etkileyen faktörler incelenmiştir.

Korelasyon ve k-means yöntemlerini kullanarak internet alışverişinde müşteri sadakatini etkileyen faktörler tespit edilmek istenmiştir. Bunun sonucunda, katılımcıların en çok tercih ettikleri alışveriş siteleri ile ilgili olarak; site kullanımının basit olduğu, kolay ve hızlı şekilde işlem yapabildikleri, satın alma işlemi sırasında ve sonrasında ürün ile ilgili sorun yaşamadıkları görülmektedir. Kişisel bilgilerin ve kredi kartı bilgilerinin korunması konusunda müşterilere güven verdikleri, satın alma sonrasında hizmet ve yardım, iade, tazminat ve garanti gibi konularda her türlü desteği verildiği, sipariş yönetimi sistemlerinin iyi olduğu ifade edilmektedir. (Yıldız vd., 2021:SY)

Çalışmaya katılan tüketicilerin araştırma kapsamındaki çevrimiçi mağazaların site içinde sundukları teklifleri, performansını, sunulan hizmetleri, site özelliklerini beğendiklerini ifade ettikleri görülmüştür. Dolayısıyla çevrimiçi perakende sektöründeki markalar pazarlama faaliyetlerini yürütürken, tüketicilerin diğer firmalar yerine bu mağazaları tercih edeceklerini söylemelerinin ve mağazaya karşı olumlu tutuma sahip olmalarının arkasındaki sebepleri göz önünde bulundurabilirler. Ayrıca bu çalışmada e-hizmet kalitesi aracılığıyla alışveriş deneyiminden memnun kalan ve tatmin olan tüketicilerin söz konusu çevrimiçi perakende sektöründe faaliyet gösteren markalar için yol gösterici nitelikte bir sonuç olduğu düşünülmektedir.(Yaşın vd., 2019:SY)

Değer, müşteri hizmetleri, fayda ve güvenlik algıları ile online müşteri sadakati arasında anlamlı ilişkiler saptanmıştır. Online alışveriş sitesine sahip işletmeler online müşterilerinin tekrar online olarak işletmelerinden alışveriş yapmaları için bu faktörlere önem vermelidir. Elde edilen bulgulardan online müşterilerin sadakatini birinci sırada değer, ikinci sırada müşteri hizmetleri, üçüncü sırada fayda ve son sırada güvenlik algılarının etkilediği görülmüştür.

İlk olarak online alışveriş sitelerinden alışveriş yapan müşterilerin değer algısının müşteri sadakatini etkileyen en önemli değişken olduğu görülmüş olup, bu bulgu çalışma sonuçları ile örtüşmektedir. Parasuraman ve Grewal tarafından yapılan çalışma sonucunda değer algısının tekrar satın alma davranışını tetikleyici rapor edilmiştir.

Aynı doğrultuda Srinivasan vd. algılanan değer yüksek olursa online müşteri sadakatinin de yüksek olacağı bu durumun satıcı ve müşteri olmak üzere her iki tarafı da olumlu yönde etkileyeceği sonucuna ulaşmışlardır. Buna ilaveten Chen yaptığı çalışmanın sonucunda, yapılan alışveriş sonucu müşteride oluşacak değer algısı düşük olduğunda online alışveriş siteleri diğer alışveriş siteleri ile rekabet edebilmek için daha fazla çaba sarf etmeleri gerektiğini belirtmiş ve bu durumun da müşteri sadakatinin oluşmasında itici güçlerden birisi olduğu açıktır.

Analizler sonucu online alışveriş sitelerinden alışveriş yapan müşterilerin müşteri hizmetleri algısının müşteri sadakatini etkileyen ikinci değişken olduğu görülmüştür. Müşteri hizmetleri değişkenine ilişkin elde edilen bulgular araştırma bulgularıyla da tutarlı görülmektedir. Eid tarafından yapılan çalışma sonucunda müşteri sadakatini oluşturma online alışveriş siteleri için en büyük zorluklardan biri olduğu, bunun için müşteri memnuniyet ve güveninin sağlanması uzun vadeli müşteri ilişkilerinin geliştirilmesinin ön şartlar olduğu ifade edilmiştir. Sadık müşteriye ulaşmanın en etkili yöntemlerinden biri kaliteli bir müşteri hizmeti sunmaktır. Müşteri hizmetlerinde yaşanan aksaklıkların en düşük seviyeye indirilmesi gelecekte yüksek müşteri hizmetleri algısına yol açacaktır. Bu doğrultuda online alışveriş siteleri de müşteri gereksinimlerini anlamak ve hizmet geliştirmeye dayalı ve duyarlı bir geribildirim mekanizması geliştirerek bu algıyı artıracığı söylenebilir.

Üçüncü olarak online müşteri sadakatini etkileyen fayda algısı olup gelişecek olumlu fayda algısı müşteri sadakatini belirlemede önemli bir rol oynayacaktır. Online alışveriş sitesi müşteriye alışveriş için bilgi sağladığında ve ürün çeşitliliği sunduğunda müşteriler online alışveriş hizmetleri hakkında olumlu duygular geliştirmeleri daha olasıdır. Gelişecek olumlu fayda algısı kullanıcı memnuniyet ve

güvenini belirlemede önemli bir rol oynayacak ve sadık müşteriler haline getirecektir.

Güvenlik algısı da online müşteri sadakatini anlamlı olarak etkilemektedir. Buna karşın güvenlik algısının değer, müşteri hizmetleri ve fayda algılarından sonra gelmesi ve önem sırasında en sonda yer alması çalışmanın dikkat çeken bulgularından birisidir. Günümüz online alışveriş sitelerinde güvenlik stratejik bir özellik olmaktan çıkmış ve müşteriler için de temel bir gereklilik halini alması bu bulguyu destekler niteliktedir.

Yapılan çalışmalarda da güvenlik algısının müşteri sadakatinde belirleyici rol oynadığı vurgulanmaktadır yaptıkları çalışmada diğer faktörlerin yanında olumlu gizlilik ve güvenlik algılarının müşteri memnuniyeti ve sadakatini etkilediğine dikkat çekmişlerdir. Chang ve Chen bir çok potansiyel müşterinin online alışveriş sitelerinden alışveriş yapmamasının başlıca nedeninin güvenlik algısı eksikliği olduğunu belirtmişlerdir. Güvenlik, online alışveriş siteleri müşterileri için dikkate alınması ve özen gösterilmesi gereken önemli bir konu olup gerekli güvenlik gereksinimleri (kimlik doğrulama, bütünlük, şifreleme ve inkar etmeme) online alışveriş siteleri tarafından sağlanmalıdır. Müşteriler online alışveriş sitelerinden alışveriş öncesinde, sırasında ve sonrasında güvenlik özellikleri (güvenlik politikası, güvenlik sözleşmesi, güvenli alışveriş garantisi gibi) ve koruma mekanizmaları (şifreleme, koruma, kimlik doğrulama, güvenli soket katmanı teknoloji gibi) bulunduğunu gördüklerinde satıcıların güvenlik ihtiyaçlarını karşılamayı amaçladıklarını anlayıp olumlu bir güvenlik algısı geliştirebilirler.

Online müşteri sadakatini etkilediği düşünülen ve bu çalışma kapsamında olan dört müşteri algısı bu çalışmanın kısıtlarından birisidir. Gelecek çalışmalarda araştırmacılar müşteri sadakatini etkileyen diğer müşteri algılarını değerlendirme kapsamına alabilirler. Diğer bir kısıt ölçeğin devlet üniversitesi öğrencilerine uygulanmasıdır. Gelecek çalışmalarda ölçek farklı ya gruplarına ya da özel üniversitelerde uygulanabilir. Birçok farklı mecra ve yerde yapılan müşteri sadakati araştırmaları, benzer sonuçları ortaya sermiştir.(Ateş, 2018:SY)

Tez için yapılan literatür araştırması sırasında veri madenciliğinin e-ticarete uygulanması ile ilgili bir araştırmada, 2014-2019 yılları arasında Türkiye’de yer alan

üçü büyük online kitap satış sitesi üzerinden elde edilen verilerle yapılan bu çalışmada, kümeleme yöntemiyle hangi yılda, hangi yazarın ve o yazara ait hangi kitap kategorisinin çok veya az sattığı bilgilerine ulaşılmıştır. Ayrıca en çok satan yayınevleri de analiz kapsamında değerlendirilmiştir. Çalışmanın bulgularına göre, yerli ve yabancı romanlar en çok satılan kategori olmuştur. Bu çalışmanın en önemli özelliği, dijital çağda zorlu bir rekabeti içinde barındıran online kitap satışlarını veri madenciliği tekniği ile incelenmektedir.(Polat vd.,2021:SY)

Birliktelik kuralları ile ilgili yapılan literatür çalışmasında, hazır giyim sektöründe faaliyet gösteren firmanın alışveriş kayıtları bazı sonuçlara varmıştır. Yapılan analizler sonucunda, gömlek alan kişilerin %34.094'ünün kravat ya da papyon da satın aldığı ve bu kişilerin, toplam müşterilerin %17.618'ini oluşturduğu görülmektedir. Veri setini oluşturan müşterilerin yaş ortalamasının 43, cinsiyete göre dağılıma bakıldığında müşterilerin çoğunluğunun erkek olduğu görülmektedir. Buna dayanarak ticari rekabette üstünlük sağlamak için, verilerin ayrıntılı olarak kayıt altında alınması gereklidir.(Timor vd., 2011:SY)

Anket çalışması sonucunda yapılan analizlere göre internet üzerinden alışveriş yapanların çoğunlukla kadınlardan oluştuğu gözlemlenmiştir. Katılımcıların diğer profil özellikleri ise; 25-34 yaş aralığına sahip, gelir düzeyi yüksek, özel sektör çalışanlarından oluşmalarıdır.

İnternet üzerinden alışveriş yapmak için bilgisayar ya da akıllı telefon sahibi olmak ve internet teknolojilerini bilmek gerektiği için internet üzerinden alışveriş yapanların profili bu özelliklere uygun çıkmıştır. İnternet üzerinden satış yapan e-ticaret firmalarının bu profili öz önüne alarak pazarlama stratejilerini geliştirmeleri gerekmektedir. Araştırmaya katılan kullanıcıların çoğunluğu geç teslimattan şikâyetçi olmuşlardır. İnternet üzerinden satışı bulunan şirketlerin lojistik ağı nedeniyle bu tip şikâyetlerin olması olağan karşılanmıştır. İşletmeler bu sorunu minimize ederek cirolarını artırma yoluna gidebilirler. Katılımcıların yüksek bir oranı ödeme şeklini tek çekim olarak belirtmişlerdir.

Bunun nedeni bankaların taksitlerde vade farkı koyması olmakla birlikte, işletmelerin bankalarla anlaşmasının bulunduğu banka pos cihazlarında kısa vadeli (1-6 ay) taksit seçeneklerini vade farksız ödeme imkânı ya da kısa vadeli taksit seçeneklerinde işletme karını artıracak kampanyalarla destekleyerek tüketici sayısını arttırması hedeflenebilir. (Gürsoy,2019:SY)



İKİNCİ BÖLÜM

2. VERİ MADENCİLİĞİNE GENEL BAKIŞ

Teknolojinin ve bilgisayarın yaşantımıza girmesiyle, yaptığımız her işlem sayısal ortamda kayıt altına alınmaya başladı. Sanal ortamlarda veya mağazalarda yaptığımız alışverişlerdeki her türlü işlem bilgisayarda, veri tabanlarında tutulmaktadır. Tutulan verilerin hepsi veritabanlarından çıkarılmayı bekleyen değerli bir madendir. Bu bilgilerden yola çıkarak veri madenciliğiyle ilgili şunlar söylenebilir;

Veri madenciliği 1990'lerden beri veri depolama araçları, barkot ve RFID teknolojilerine paralel olarak geliştirmekte ve kullanım alanı yayılmakta olan bir konudur. Bu nedenle, kullanılan yer ve zamana göre çeşitli tanımları yapılmıştır. Çünkü her geçen gün daha da geliştiği için bugün yapılan bir tanım yarın yetersiz kalabilmektedir. Veri madenciliği daha önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veritabanlarından elde edilmesi ve bu bilgilerin işletme kararları verirken kullanılmasıdır. “Daha önce bilinmeyen” ya da tahmin edilemeyen örnekleri veri madenciliği çarpıcı bir şekilde önümüze sermeyi hedefler.

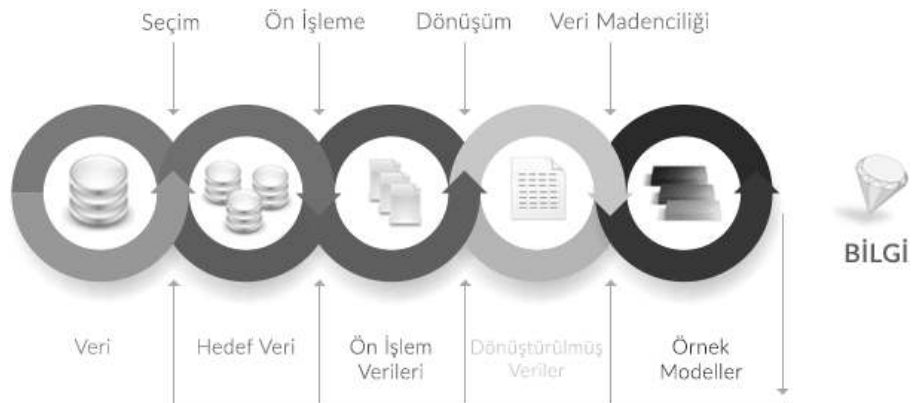
Bir diğer tanımda ise veri madenciliği, istatistik ve matematiksel tekniklerle birlikte örüntü tanıma teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi sürecidir.

Büyük ölçekli veriler içinden faydalı bilgilere ulaşabilme, gelecekle ilgili tahminde bulunabilen ve daha önceden tespit edilememiş veri setlerini oluşturmamızı sağlayan istatistiksel bir yöntemdir. Genel olarak veri madenciliği, veri setleri arasındaki düzenin, veri analizinde ve yazılım tekniklerinde kullanılmasıdır. Aynı zamanda veri madenciliği, kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi teknik yaklaşımları içerir. Asıl amaç, daha önceden fark edilmemiş veri desenlerini gün yüzüne çıkarmaktır.(Silahtaroglu, 2020:11)

2.1. Veri Madenciliği Uygulama Süreçleri

Veri madenciliğinin ortaya çıkması ve günümüzde bu denli kullanılmasının en büyük nedenlerinden bir tanesi, günümüzde büyük veritabanlarının erişebilir olmasıdır. Fakat bu bilgilerin tamamının gerçek ve doğru bilgiler olduğunun garantisi yoktur. Bu sebeple bilgiler belirli işlemlerden geçmelidir. Elde edilen bilgilerin uygulanma süreci, ulaşılmak istenen veri büyüklüğü ve gerçekleşecek işlemin uzunluğuna bağlı olarak değişiklik göstermektedir. Fakat her teknikte olduğu gibi veri madenciliğinde de belli başlı sıralamalar bulunmaktadır. İşleyiş sistemi şu şekilde ifade edilebilir;

- **Veri Filtreleme:** Kullanılacak verilerin belirlenme aşaması
- **Veri Temizliği:** Toplanan veriler arasından gereksiz ve gürültülü olanların kullanım dışı bırakılması.
- **Veri Birleştirilmesi:** Farklı kaynaklardan, benzer içerikli edinilen bilgilerin birleştirilmesi
- **Veri Seçme:** Temizlenen ve birleştirilen bilgilerden, uygun olanların seçilmesi.
- **Veri Dönüştürme:** Bu adıma kadar elde edilen verilerin, veri madenciliğine uygun hale dönüştürme.
- **Madencilik Çalışması:** Hazır hale getirilen veriler üzerinde, amaca uygun algoritmaların uygulanması.
- **Yorumlama ve Doğrulama:** Veri madenciliği uygulamasını gerçekleştirdikten sonra, elde kalan sonuçların doğruluklarının araştırılması. Sonuçların doğruluğu ise farklı uygulamalardan alınan sonuçlar ile mümkün kılınır. (Gtech, 2021)



Şekil 2.1. Veri Uygulama Süreçleri (Http1, 2020)

2.2. Veri Madenciliği Modelleri

Veri madenciliği modellerini tahmin edici ve tanımlayıcı modeller olarak ikiye ayrılır. Tahmin edici modeller özellikle eğitilmiş veri setlerinde verinin sınıfı, gelecek değeri vb. tahmin değerlerini belirlemek için kullanılırken tanımlayıcı modeller eğitilmiş veya eğitimsiz veri setlerinde verinin gruplandırılması kullanılır.



Şekil 2.2 Veri Madenciliği Teknikleri (Http2, 2020)

2.2.1. Tahmin Edici Modeller

“Bu müşteriden ne kadar gelir elde edilebilir?” veya “ Bu işlemde dolandırıcılık var mıdır?” gibi sorulara yanıt arar, Aşağıdaki seçeneklerde bu sorulara yanıt bulunacak seçeneklerdir.

- Sınıflandırma
- Regresyon
- Zaman Serisi Analizi

2.2.2. Tanımlayıcı Modeller

Veriler arasındaki gizli bağı ortaya çıkarmak ve “Bebek bezi alan bir müşterinin, bebek maması alma olasılığı % kaçtır?” sorusuna yanıt bulmayı amaçlar.

Aşağıdaki seçenekler amaca ve soruların cevaplarına ulaşmayı sağlayan seçeneklerdir.

- Kümeleme Yöntemi
- Birliktelik Kuralı

2.3. Veri Madenciliği Yöntemleri

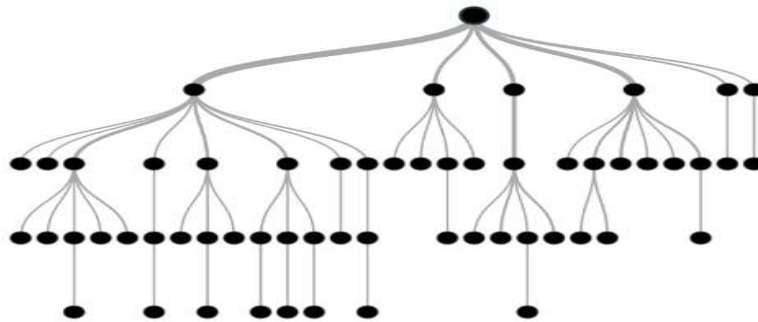
Veri madenciliğinde asıl dikkat edilmesi gereken husus, verileri sisteme en uygun yöntem ile dönüştürmektir. Arzulanan sonuca başarılı bir şekilde ulaşmak için farklı amaçlara sahip birçok yöntem aynı anda kullanılabilir. Kullanılan yöntemlerin detaylı açıklaması şöyledir;

2.3.1. Sınıflandırma Yöntemi

Örüntü tanıma, kalite kontrol denetimleri, hastalık tanıları ve benzeri alanlar, sınıflandırma yönteminin çokça kullanıldığı alanlardır. Aynı zamanda kişilerin tercih ettiği araba modellerini, havanın sıcaklığını, kutudan çekilen topun rengini ve bunlara benzer tahmin esaslı konular bu yöntem için verilebilecek en iyi örneklerdir. Sınıflandırma yöntemi kendi içerisinde algoritmalara ayrılır.

Bunlar;

- Karar Ağacı
- İstatistiğe Dayalı Algoritmalar
- Mesafeye Dayalı Sınıflandırma Algoritması
- Kollektif Öğrenme Kavramı
- Yapay Sinir Ağları

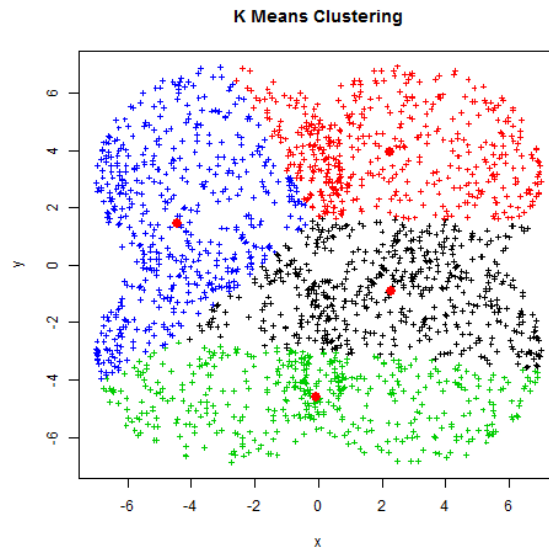


Şekil 2.3. Karar Ağacı

(Http3, 2017)

2.3.2. Kümeleme Analizi

Kümeleme analizi (clustering) veri madenciliğinin en önemli alanlarından birisidir; amacı, nesneleri birbirlerine olan benzerliklerine göre gruplara ayırmaktır. Elde bulunan veriler incelenerek birbirlerine benzeyenler bir kümeye, benzemeyenler ise bir başka kümeye toplanırlar. Verilerin kümeleme analizine göre modellenmesinde matematik, istatistik, makine öğrenimi ve yapay zeka gibi birçok alandan yararlanır. Makine öğrenimi açısından, her bir küme gizli bir örüntüyü temsil eder ve uygulanan öğrenme ise bir denetimsiz öğrenmedir. Bu açıdan bakıldığında ise kümelemeyi, “gizli örüntülerin ortaya çıkartılması için uygulanan bir denetimsiz öğrenme yaklaşımı” olarak tanımlanabilir. Sınıflandırmada olduğu gibi verileri gruplara ayırma işlemidir. Kümeleme analizinde, sınıflar önceden belirli değildir. Verilerin birbirlerine olan benzerlikleri baz alınıp, hangi gruba veya kaç değişik gruba ayrılacağı belirlenir. İki verinin benzerliğinden kasıt ise aralarındaki mesafenin ölçülmesi ve değerlendirilmesidir. Bu değerlendirme, veritabanındaki diğer verilere kıyasla iki verinin ne kadar yakın ya da benzer oldukları açısından yapılacağı gibi önceden belirlenmiş kısıtlar eşik değerleri çerçevesinde de yapılabilir. Fakat çalışılan veri boyutu büyüdükçe bu hesapları yapmak hem güçleşir; hem de varılan sonuç gerektiği kadar hassas olamayabilir. (Silahtaroglu, 2020:72)



Şekil 2.4 Kümeleme Dağılım Grafiği

(Http4, 2018)

Örneğin, müşteriler çeşitli özellikleriyle değerlendirilerek, kimlik, ad, posta kodu veya tanımlanan müşteri numarası baz alınarak gruplanırlar. Tüm müşteriler, özellik açısından kendisine benzeyen kişi ve niteliklere sahip olan müşterilerle aynı gruba atanır. Tıp, pazarlama, ekonomi gibi birbirinden apayrı sektörler bu yöntemi kullanmaktadır.

Kümeleme analizi, kullanılan veri türü ve yapılan işin amacına göre farklılık gösterebilir. Genele olarak hiyerarşik ve bölümlenmeli olarak ikiye ayrılırlar. Fakat yapılan taramalar daha alt bölümlere de ayrılabilceğini göstermektedir.

- Hiyerarşik Yöntemler
- Toplaşım Kümeleme
- Bölümlenmeli Yöntemler
- Yer Değiştiren Algoritmalar
- Olasılıksak Algoritmalar
- K-Medoid Yöntemler
- K-Means Yöntemler
- Yoğunluk Algoritmaları

2.3.3. Birliktelik Kuralları

Veritabanlarındaki bilgi miktarı arttıkça birçok kurum ve kuruluş sahip oldukları bilgiler arasındaki ilişkileri ortaya çıkarma abası içerisine girmiştir. Böylesi yığınlar halindeki bilgiler arasındaki ilişkiler kurumlar için altın değerinde sonuçlar doğurabilecek kararların alınmasına vesile olur. İlişki analizi veritabanındaki bir dizi bilgi ya da kaydın diğer kayıtlarla olan bağlantısını açıklayan işlemler dizisidir. Veritabanlarındaki bilgilerde artış meydana geldiğinde, birbirleriyle ilişkili değişkenler ortaya çıkarılır ve aralarındaki bağlantının derecesi tespit edilir. Belirli tür verilerin arasındaki ilişkileri tanımlayan yöntemdir. Örneğin cinsiyet ile eğitim durumu arasında veya yaş ile gelir seviyesi arasında bağıntı kurularak, müşteriye sunulacak ürünler tespit edilebilir. Veya herhangi bir ürün satın alırken, bu ürünün yanında başka bir ürün ya da ürünlerin satın alınması, bu ürünler arasındaki bağlantıyı ortaya koyar.

Literatürde bu tür çalışmalara “pazar sepeti analizi” denilmektedir. Birlikte kuralları zamanla geliştirilerek alt algoritmalar oluşturulmuştur. Bu algoritmalar aşağıdaki gibidir;

- **AIS Algoritması:** Geniş nesne kümeleri üretmek için geliştirilmiş bir algoritmadır. Veritabanındaki isimlerin, yan, nesne isimlerinin A’den Z’ye sıralanması kısıtını taşır.
- **SETM Algoritması:** Geniş nesne kümesinin her bir elemanı iki parametreden oluşur; bunların bir tanesi nesnenin ismiyken diğeri bu nesneyi ayırt etmeye yarayan bir özellik numarasıdır.
- **Apriori Algoritması (FP Growth):** Bağlantı analizlerinin yapılp bağlantı kurallarının ortaya çıkartılması konusunda en çok bilinen ve kullanılan algoritmasıdır. Geniş nesne kümelerinin ortaya çıkartılması işlemleri için kullanılır.
- **AprioriTid Algoritması:** Bu algoritmada taramadan önce aday nesne kümelerini belirlemek için kullanılır. (Silahtaroglu, 2020:165)

2.3.4. FP-Growth Algoritması

FP Growth, apriori algoritmasının geliştirilmiş versiyonudur. Veri kümelerinde sık görülen kalıpları veya ilişkiler bulan analitik süreç olarak kullanılır. Apriori algoritması, öge setleri oluşturarak “minimum destek sayısı” üzerinden en sık kullanılan seti keşfederek kalıplar üretir. Algoritmanın işleyiş adımları aşağıdaki gibidir:

Adım-1: Veritabanındaki öge setlerinin oluşum aşamalarını bulmak için verileri taramaktadır. Bu adım, Apriori’nin ilk adımı ile aynıdır. Öge setlerinin sayısına destek sayısı veya öge seti sıklığı denir.

Adım-2: Ağaç kökünü oluşturup FP ağacını oluşturmaktadır. Kök, null ile temsil edilir.

Adım-3: Veritabanını tekrar tarayıp işlemleri incelemektedir. Maksimum sayıya sahip öge kümesi en üstte alınır, sonraki öge kümesi daha düşük sayı ile ağaç dallarında azalarak devam eder.

Adım-4: Veritabanındaki bir sonraki işlem incelenir. Öge setleri, azalan sayı sırasına göre işlenir. Bu işlemin herhangi bir öge kümesi başka bir şubede zaten mevcutsa (örneğin 1.işlemde), bu işlem dalı kökle ortak bir dal paylaşacaktır. Bu ortak öge kümesinin bu işlemdeki başka bir öge kümesinin yeni düğümüne bağlı olduğu anlamına gelir.

Adım-5: Ayrıca, işlemlerde ortaya çıktıkça kalem setinin sayısı artar. Hem ortak düğüm hem de yeni düğüm sayısı, işlemlere göre oluşturuldukça ve bağlandıkça 1 artar.

Adım-6: Bir sonraki adım, oluşturulan FP ağacını çıkarmaktır. Bunun için ilk olarak en alttaki düğüm, en düşük düğümlerin bağlantıları ile birlikte incelenir. En düşük düğüm, frekans örüntü uzunluğunu temsil eder. Koşullu kalıp tabanı, en düşük düğüm (sonek) ile oluşan FP ağacındaki örnek yollarından oluşan bir alt veritabanıdır.

Adım-7: Yoldaki bir dizi öge kümesinden oluşan bir koşullu FP ağacı oluşturun. Eşik desteğini karşılayan öge grupları koşullu FP ağacında dikkate alınır.

Adım-8: Koşullu FP ağacından sık modeller oluşturulur. (Http5, 2020)

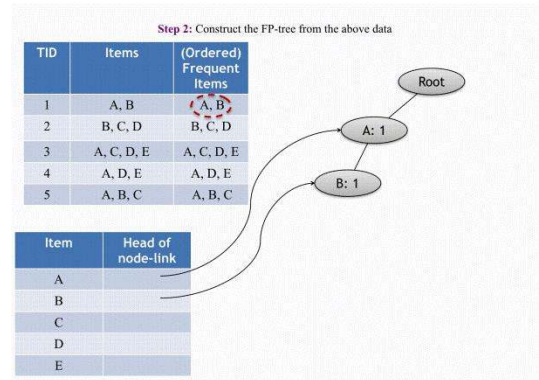
A Simple Example: Step 1

TID	Items	Item	Frequency
1	A, B	A	4
2	B, C, D	B	3
3	A, C, D, E	C	3
4	A, D, E	D	3
5	A, B, C	E	2

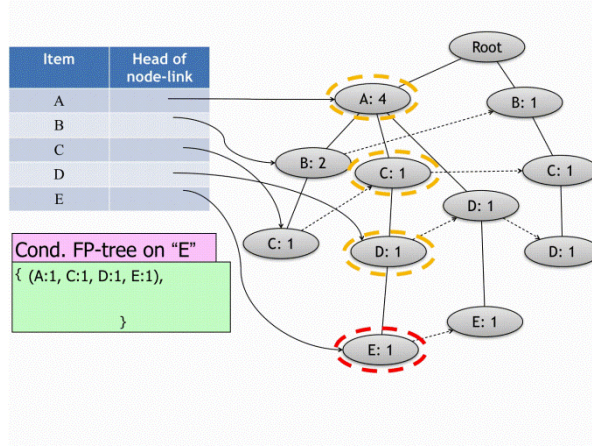
Threshold = 2

Item	Ordered Frequency
A	4
B	3
C	3
D	3
E	2

Şekil 2.5. FP-Growth 1.Adım
(Http6, 2020)



Şekil 2.6. FP-Growth 2.Adım
(Http7, 2020)



Şekil 2.7. FP-Growth 3.Adım
(Http8,2020)

2.4. Genelleştirilmiş Sıralı Problem Modeli (GSP)

Paralel doğrular varsayımının reddedildiği durumda kullanılan bağımlı değişkenin sıralı yapısını dikkate alan modeldir. Genelleştirilmiş sıralı lojistik modelde bağımlı değişkenin bir düzeyinden diğer düzeyine geçiş olasılıkları farklı olacağından bağımlı değişkenin her bir düzeyi için model tahmini yapılmaktadır. Bağımlı değişkeni $k=1,2,\dots, K-1$ olmak üzere $K-1$ kategorisi olan genelleştirilmiş sıralı problem model Eşitlik 2' de gösterilmektedir. (Boes vd., 2006:171)

$$P\left(Y = \frac{K}{X}\right) = F(\mu_1 - \beta_1 X), K = 1 \text{ ise}$$

$$P\left(Y = \frac{k}{X}\right) = F(\mu_k - \beta_k X) - F(\mu_{k-1} - \beta_{k-1} X) \quad 1 < k < K-1 \text{ ise (2)}$$

$$P\left(Y = \frac{k}{X}\right) = 1 - F(\mu_K - 1 - \beta_K - 1X) \quad , k=K \text{ ise}$$

Eşitlik 2'de F, birikimli lojistik dağılımı göstermektedir. B, lojistik katsayı vektörünü, μ kesim noktasını ifade etmektedir.

2.5. Sepet Analizi

Yapılan alışverişlerdeki ürünlerin arasındaki korelasyona bakarak müşterilerin satın alma alışkanlıklarını belirlemeye çalışan yöntemdir. Yapılan bu çalışma sayesinde mağaza yönetimi, birlikte satın alınan ürünleri yakın raflara yerleştirerek satışlarda artış olmasını amaçlar. Sepet analizinde, satışı yapılan ürünlerin ilişkilerini bulabilmek için destek (support) ve güven (confidence) ölçütlerinden faydalanılır. Bu ölçütlerin büyük olması birliktelik kurallarının güçlü olduğuna karar verilir.

“Destek sayısı” bu ölçütlerin hesaplanmasında kullanılan değerdir. **Sayı(X,Y)** A ve B ürününün birlikte alındığını gösterir. **X → Y** gösterimi, A ürün grubunu alanların B ürün grubunu da aldığını ifade eder.

Bir X faktörünün gerçekleşme olasılığı,

$$P(\{X\}) = 6.000 / 10.000 = \%60$$

Bir Y faktörünün gerçekleşme olasılığı,

$$P(\{Y\}) = 7.500 / 10.000 = \%75$$

X ve Y faktörlerinin beraber gerçekleşme olasılığı,

$$P(\{X, Y\}) = 4.000 / 10.000 = \%40$$

Kaldıraç değeri,

$$\text{Kaldıraç } (X \Rightarrow Y) = \frac{0.40}{0.60 * 0.75} = \%89$$

Şekil 2.8. Birlikte Alınma Olasılık Hesaplaması
(Http9, 2020)

Bu algoritmanın uygulanma adımları aşağıdaki gibidir.

- Eşik değerleri belirlenmelidir. (min.destek ve min.güven)
- Küme içerisindeki her ögenin destek değeri bulunmalıdır.
- Minimum destek değerinden küçük olan öge kümeleri devre dışı bırakılmalıdır.
- Elde kalan tekli birliktelikler göz önüne alınarak ikili birliktelikler oluşturulmalıdır.
- Minimum destek değerinden küçük olan öge kümeleri devre dışı bırakılmalıdır.
- Elde kalan ikili birliktelikler göz önüne alınarak üçlü birliktelikler oluşturulmalıdır.
- Ve son olarak elde edilen üçlü birlikteliklerden birliktelik kuralları çıkarılır. (Htt10, 2016)

2.6. Veri Madenciliğinin Uygulama Alanları

Veri madenciliği iş dünyasında sektör fark etmeden kullanılan bir sistemdir. Geniş veritabanlarına sahip, perakende satış, sağlık veya sigortacılık gibi alanların dışında, pazarlama, bankacılık ve benzeri birçok alanda uygulanmaktadır. Hangi alanda hangi şekilde kullanıldığını daha iyi ifade edebilmek için uygulama alanlarını ana başlıkları aşağıdaki gibidir.

Bankacılık:

- Kredi kartı kullanım alanlarına göre müşteri belirlenmesi
- Kredi taleplerine göre kişi değerlendirilmesi

Pazarlama:

- Müşterilerin satın alma örüntülerinin belirlenmesini sağlar.
- Müşterilerin demografik özellikleri arasındaki bağlantıları tespit eder.
- Posta kampanyalarında cevap verme oranını artırır.
- Mevcut müşterilerin elde tutulup, yeni müşteri kazanımını sağlar.
- Risk yönetimi ve dolandırıcılık saptamasını yapar.

CRM:

- Müşteri memnuniyetini artırmayı.
- Kampanyalarda en iyi seviyede yarar sağlayabilmeyi.

E-Ticaret:

- Web sayfasında gezinen kullanıcıların davranışlarının belirlenmesi.
- Müşteri bilgilerine göre kampanya oluşturmak. (Silahtaroglu, 2020:15)



ÜÇÜNCÜ BÖLÜM

3. E-TİCARETE GENEL BAKIŞ

Elektronik ticaret veya e-ticaret, internetin kullanımının arttığı tarihlerden itibaren ortaya çıkmış, ticaretin elektronik bir ortamda yapılması durumudur. Her türlü tanıtım, satış, ödeme ve dağıtım gibi işlemlerin sadece bilgisayar sistemleri üzerinden gerçekleşmesidir.

E-Ticaret sistemi şirketlerin sadece belirli bir kitleye değil, küresel bir alana ulaşmasına yardımcı olur. Aynı zamanda üreticinin herhangi bir aracıya ihtiyaç duymadan, direkt tüketiciye ulaşmasını sağlayan düzendir. Bu düzen, özellikle KOBİ'ler için makul bir ticaret yöntemidir. KOBİ'lerin pazar marjlarının artmasına yardımcı olur. Ayrıca küçük çaplı üreticilerin dünya pazarına girebilmelerini sağlar. E-Ticaret, ürün seçeneklerinin artmasını, kalitenin yükselmesini ve çok hızlı ödemesi alınıp teslim edilmesini sağlamaktadır. Bu durumda rekabetin artıp, ticari işlem maliyetlerinin düşmesine sebep olur. En büyük avantajı ise, normal şartlarda ortalama 10 yıl süren markalaşma süreci, elektronik ticaret sayesinde 2 yıla kadar inmektedir. (Http11, 2022:S.Y.)

E-Ticaret yapabilmek için her sistemde olduğu gibi detaylı bir organizasyon planı yapmak gerekmektedir. İlk adım iyi bir site kurulumu ile başlamaktadır. Pazar yeri platformlarından dahi satış yapacak olsanız, adınıza ait web sitenin olması gerekir. Güçlü bir altyapı ile kurulan web sitesi sonrası satış süreci başlamaktadır. Bu sürecin başarılı olması için her şeyden önce iyi bir fikre sahip olmak gerekir. Sektör ve ürün seçimi başarıyı etkileyen önemli faktörlerden olacaktır. Doğru yapılan seçimler sonrası hitap edilecek kitleye göre bir iş modeli oluşturulmalıdır. İster küçük bir kitleye hitap eden bir platform, isterseniz geniş yelpazeli bir hizmet ağına sahip olabilirsiniz.

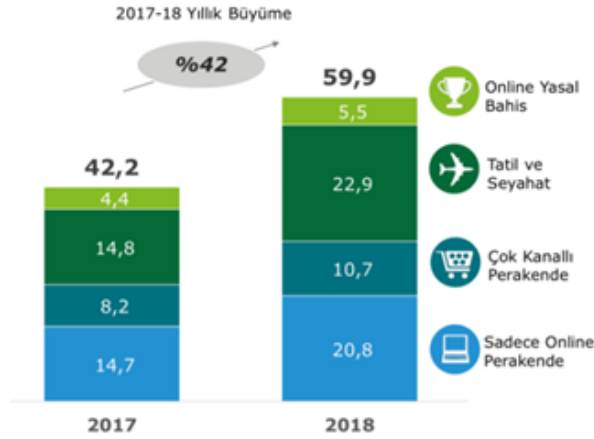
Online satış, müşterilerin ürünleri sipariş etmesiyle başlayan bir süreçtir. Sipariş edilen ürünün sorunsuz bir şekilde müşteriye teslim edilmesi ile süreç sonlanmış olur. Yapılan her satış için hizmet sonrası destek ve teknik servis aşamaları başlar. (Http12,2022:S.Y.)

3.1. E-Ticaretin Türkiye’deki Durumu

1997 yılında “Bilim ve Teknoloji Yüksek Kurulu” toplantısı ile Türkiye e-ticaret sektörüne giriş yapmıştır. 2013 yılı itibari ile belirgin raporlamalar oluşturulmaya başlanmıştır. Bu raporlamalar sayesinde büyümenin ne derece ciddi olduğunu gerçek anlamda gözlemlenmiştir. 2013-2015 yılları arasında yaklaşık olarak %34 oranında büyüme olmuştur. TÜBİSAD verilerine göre 2016 yılı sonrasında 17,5 milyar TL’lik bir hacme ulaşılmıştır.

E-Ticaretin Türkiye’deki hacmi her geçen gün durdurulamaz bir şekilde büyümeye devam etmektedir. Dijital dönüşümün yaşandığı şu dönemde, kamu, büyük işletmeler, KOBİ ve bireylerde bu değişime dâhil olmaktadır. Tüketiciler bu değişim ve dönüşüme kolaylıkla ayak uydurup, dijital ortama yönelmektedir.

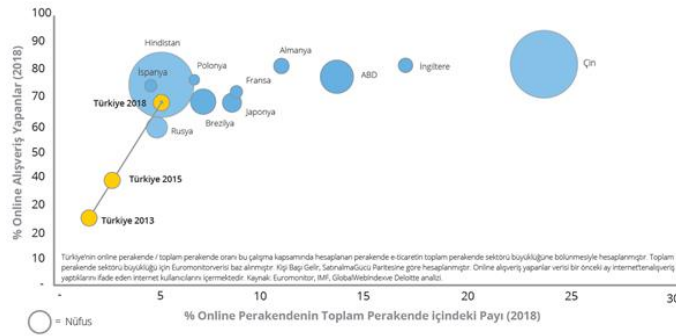
Bu sebeple işletmeler yeni iş modelleri kurarak satışlarında artış elde etmenin yolunu aramaktadır. Dijital hayatın önemli noktalarını tespit ederek, e-ticaretin doğru büyümesine önem verilmelidir. Türkiye’de e-ticaret en çok perakende, turizm ve teknoloji sektöründe büyüme göstermektedir. Ayrıca Türkiye’de akıllı telefon kullanımı artışı da bu alanın büyümesine etki etmektedir.



Şekil 3.1. 2017-2018 Yıllık Büyüme Verileri
(TÜSİAD,2019:S.Y.)

TÜSİAD raporuna göre son yıllarda Türkiye’ye ait e-ticaret verileri şöyledir;

- 2018 yılında e-ticaret hacmi 59,9 milyar TL’ye ulaşmıştır. 2017 yılına göre %42 oranında büyüme sağlanmıştır.
- Son 5 yılda önde gelen alışveriş sitelerinin pazar payı %40 oranlarından %53 seviyelerine gelmiştir.
- 2014-2018 yılları arasında yıllık büyüme oranı %33 olmuştur.
- 2017 yılında toplam perakende pazarı içinde e-ticaret oranı %4,1 olarak görüldü. 2018 yılında ise bu oran %5,3 olarak değişti.
- Gelişmiş ülkelerde perakende satışında e-ticaret oranı %11,1 iken gelişmekte olan ülkelerde %5,9 olarak gözlenmiştir.
- 2018 yılında turizm sektöründeki satışlar bir önceki yıla göre %54 büyüyerek, 22,9 milyar TL olarak güncellenmiştir.
- 2019 yılında e-ticaret hacmi 136 TL olmuştur
- 2020 yılı bir önceki yıla göre %66 büyüyerek 226,2 milyar TL olmuştur
- 2021 yılı ilk 6 ayında, 161 milyar TL olmuştur. (Tüsiad, 2021:S.Y.)



**Şekil 3.2. Perakende Satış Dağılımları
(TÜSİAD,2019:S.Y.)**

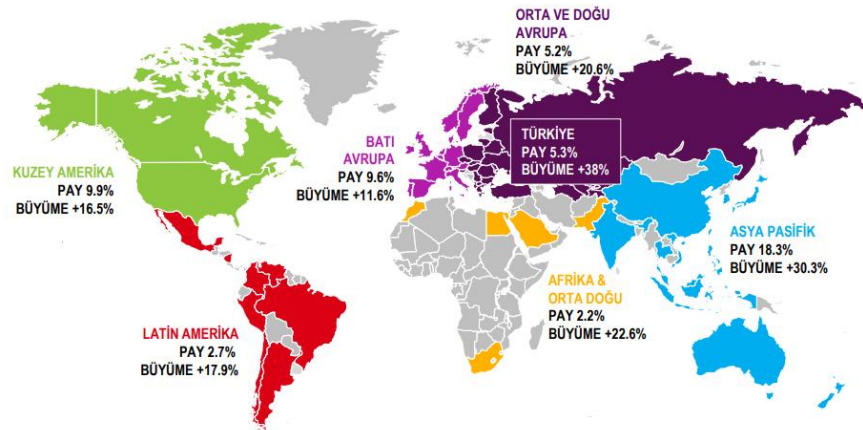
E-Ticaretin Dünyadaki Durumu

E-Ticaret, küresel alanda artış gösterirken, etki ettiği alanlarını her geçen gün büyütmektedir. E-Ticaret kanallarının kullanımı bölgelere göre değişiklik göstermektedir. Dünya genelinde online alışveriş için en çok masaüstü bilgisayarlar kullanılmaktadır. Akıllı telefonlarda bu sektörün tanıtırlığını arttırmada büyük rol oynamaktadır. Sektörün bu denli büyümesindeki en büyük etken, genç nüfus ile artan internet, akıllı cihaz ve kredi kartı kullanımınıdır. Dünya çapında perakende satışlar 2019 yılında 3,53 trilyon dolar olarak gözlemlenmiştir. 2022 yılı içinde 6,54 trilyon olması beklenmektedir.

B2C e-ticaret pazarında Çin ve ABD lider ülke olmaya son gaz devam etmektedir. Çin 2018 yılında 634 milyar dolarlık bir alanla dünyada ilk sırada yeri almıştır. 504 milyar dolar satış ile ABD Çini takip etmiştir. 123 milyar dolar ile Japonya, 86 milyar dolar ile Birleşik Krallık ve 70 milyar dolar ile Almanya bu alanda gelir etmiştir. B2B ticaret hacmi ise, takribi 10 trilyon dolara ulaşmıştır. Rakamlar B2C e-ticaretin 5 katına denk geldiğini göstermektedir.

Bazı ülkelerin sınır ötesindeki e-ticaret davranışları şöyledir;

- Lider Çin, ithalat ürünlerinin %42'sini Japonya'dan, %14'ünü Amerika'dan, %12'sini Hong Kong'dan temin etmektedir.
- ABD'deki tüketiciler Çin'den %57, İngiltere'den %10 ve Kanada'dan %6 ürün satın almaktadır.
- İngiltere'deki tüketiciler Çin'den %50, ABD'den ve Almanya'dan %7 ürün satın almaktadır.(Tüsiad, 2021:S.Y)



Şekil 3.3. Dünya Genelinde E-Ticaret Dağılımı

(Http13,2019:S.Y.)

Dünya genelinde çoğunlukla tercih edilen ürün kategorileri ve oranları aşağıdaki gibidir;

- Ayakkabı ve giyim %31
- Bakım ve güzellik %17
- Saat ve aksesuar %11
- Elektronik %21
- Müzik, medya ve kitap %10
- Son olarak ev aletleri %8

Dünyada e-ticaret satışları:

- 2014 yılı 1,336 trilyon \$
- 2015 yılı 1.548 trilyon \$
- 2016 yılı 1.845 trilyon \$
- 2017 yılı 2.382 trilyon \$
- 2018 yılı 2,9282 trilyon \$
- 2019 yılı 3.535 trilyon \$
- 2020 yılı 4.300 trilyon \$

DÖRDÜNCÜ BÖLÜM

4. VERİ MADENCİLİĞİNİN E-TİCARETE UYGULANMASI

4.1. Veri Hakkında

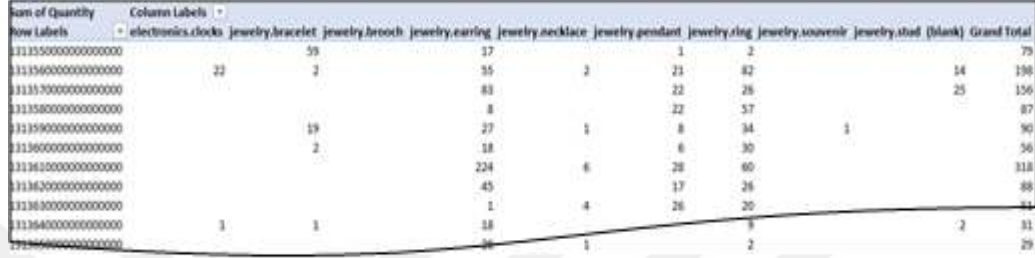
Çalışmaya esas veri Kaggle adresinden elde edilmiş olup, orta ölçekli bir çevrimiçi mücevher mağazasından Aralık 2018 ile Aralık 2021 (3 yıl) arasındaki satın alma işlemlerini içermektedir. (Kaggle,T.y.) Dosyadaki her satır, satın alınan bir ürünü temsil eder. Aynı sipariş/satın alma işleminden birkaç ürün ayrı satırlarda listelenir ve “order_id” alanıyla birleştirilir. Veri kullanıma ve analize açıktır. Veri setinde toplam 95912 alışveriş kaydı bulunmaktadır. Her bir kayıt satın alınan bir adet ürüne ait bilgileri içermektedir. Veri setinde bir adet tarih saat, bir adet reel sayı, bir adet tamsayı ve dokuz adet nominal tipinde özellik bulunmakta olup söz konusu özellik listesi Tablo 4.1’de sunulmuştur.

Tablo 4.1. Veri Setinde Yer Alan Özellikler (Yazar tarafından oluşturulmuştur.)

Özellik Adı	Tipi	Anlamı	Eksik Değer Sayısı
Date/time	Tarih-saat	Alışveriş tarih saati	-
Order_id	Nominal	İşlem kodu	-
Product_id	Nominal	Satın alınan ürün kodu	-
Quantity	Tamsayı	Satın alınan miktar	-
Category_id	Nominal	Ürün kategori kodu	5352
Category_alias	Nominal	Ürün kategori adı	15285
Brand_id	Tamsayı	Marka kodu	10137
Price_USD	Reel sayı	Ürün fiyatı (Dolar)	-
User_id	Nominal	Müşteri kodu	-
Product_gender	Nominal	Ürün kullanıcı cinsiyeti	48169
Color	Nominal	Ürün rengi	7660
Metal	Nominal	Üründe kullanılan metal	5462
Gem	Nominal	Üründe kullanılan taş	34059

4.2. Eksik Verilerin Tamamlanması

Tablo 4.1’de listelenen özelliklerden gri ile renklendirilen özelliklerde çok miktarda eksik bilgi görülmektedir. Örneğin “Gem” özellik adına ait 34059 adet eksik bilgi olduğu görülmektedir. Ayrıca belirli bir ürüne ait “product_id” incelendiğinde Şekil 4.1’ de görüldüğü gibi söz konusu ürünün birden fazla kategori adı ile temsil ettiği görülmektedir.



Sum of Quantity	Column Labels	electronics.clock	jewelry.bracelet	jewelry.brooch	jewelry.earring	jewelry.necklace	jewelry.pendant	jewelry.ring	jewelry.souvenir	jewelry.stud (blank)	Grand Total
11135500000000000000			59		17		1	2			79
11135600000000000000		22	2		55	2	21	82		14	198
11135700000000000000					83		22	26		25	156
11135800000000000000					8		22	57			87
11135900000000000000			19		27	1	8	34	1		90
11136000000000000000			2		18		6	30			56
11136100000000000000					224	6	28	60			318
11136200000000000000					45		17	26			88
11136300000000000000					1	4	26	20			51
11136400000000000000		3	1		18					2	24
11136500000000000000					26	1		2			29

Şekil 4.1. Verideki bilgilerin bir kısmı (RapidMiner)

Söz konusu ürünün birden fazla kategori adı ile temsil edilmesi bu verideki tutarsızlığın bir işareti olarak algılanmış olup, kategori adı ve “id” si veri setinden silinmiştir. Ancak ürün kullanıcı cinsiyeti özelliğinde boş bırakılan yerlerin her iki cinsiyet içinde kullanılan ürünler olduğu kabul edilmiş ve bu yerlere “unisex” değeri atanmıştır. Benzer şekilde üründe kullanılan taş özelliğinde boş geçilen yerlerde de ürünün taşsız olduğu kabul edilmiş aynı durum metal kısmında da metal kullanılmayan ürün olarak belirlenmiştir.

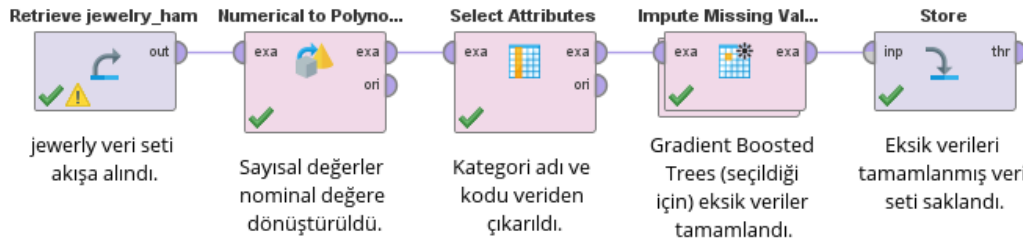
Eksik verinin tamamlanması için karar ağacı sınıflandırma algoritması (“Gradient Boosted Trees” GBT) operatörü kullanılmıştır. Bunun için RapidMiner Studio Ver:9.10(Http14,T.Y.) veri madenciliği yazılımı kullanılmıştır. Yazılımda bulunan “Impute Missing Values” operatörü yardımıyla aşağıda verilen akış tasarımı ile eksik veriler kendilerine en çok benzeyen nesnelerle karşılaştırılarak tamamlanmıştır.

GBT operatörü sınıflandırma ağacı modellerini kullanır. Kademeli olarak iyileştirilmiş tahminlere dayalı sonuçlar elde eden bir öğrenme yöntemidir.

Güçlendirme, ağaçların doğruluğunu artırmaya yardımcı olan esnek, doğrusal olmayan bir prosedürdür. Dereceli artımlı olarak değiştirilen verilere sıralı olarak hızlı ama zayıf sınıflandırma algoritmaları uygulayarak zayıf tahmin modelleri üreten

bir dizi karar ağacı oluşturulur. Ağaçları güçlendirmek, doğruluklarını artırırken, hızı ve insan yorumlanabilirliğini de azaltır.

“Impute Missing Values” operatörü, GBT operatörünü (veya K-NN, karar ağacı, Bayes vb. daha farklı bir sınıflandırma algoritması) ile eksik verileri tamamlamakta kullanılır. Bunun için kullanılan akış Şekil 4.2’de verilmiştir.



Şekil 4.2. Eksik Veri Tamamlama Akışı (RapidMiner)

Verilerin RapidMiner programına uygulanması sonrası oluşan yukarıdaki iş akışını yorumlayacak olursak, veriler sisteme alınıp işlendi. Değerler tek tip değer haline getirildi. Tutarsızlığa sebep olan değerler devre dışı bırakıldı. GBT operatörü yardımıyla eksik olan veriler tamamlanmış oldu. Eksikleri tamamlanan veri ileride kullanılmak üzere depo edildi.

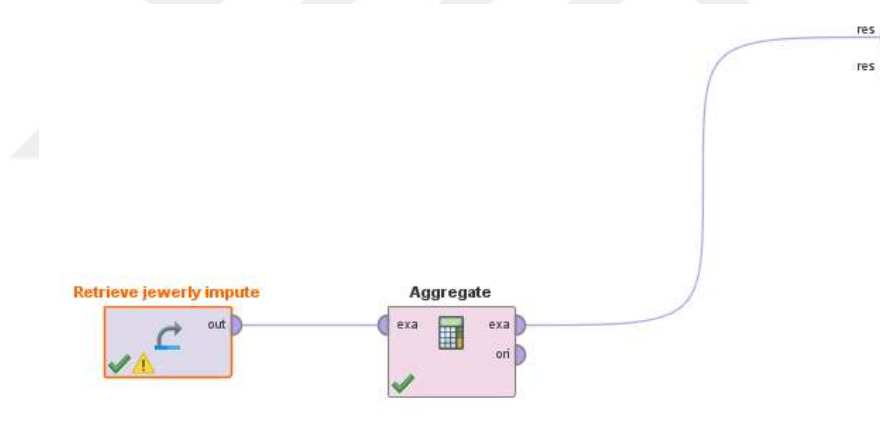
4.3. Sık Ürün Setlerinin Bulunması

Ön işleme yapılan veri seti ile öncelikle sık veri setleri incelenecektir. Veride aynı sipariş numarası (Order_id) ile satın alınan ürünler için ayrı ayrı satırlar oluşturulmuştur. Öncelikle bu satırlar Orde_id altında toplanarak, programın sık ürün setleri için kullandığı “FP-Growth” operatörü yardımıyla söz konusu sık ürün setleri bulunacaktır. Veri setinde “Order_id” altında ürünler birleştirildiğinde elde edilen veri seti örneği (ilk 10 örnek) aşağıda sunulmuştur. Burada “Order_id” altında ürünler, “Product_id” ile temsil edilmektedir. Ürünler “I” yani bar karakteri ile ayrılmıştır.

Row No.	Order_id	concat(Product_id)
1	1924720000...	18422000000000000000
2	1924900000...	18068300000000000000
3	1925510000...	18422100000000000000
4	1925630000...	18355700000000000000
5	1925740000...	18739400000000000000
6	1925760000...	16000200000000000000 18355700000000000000 18355700000000000000
7	1926030000...	18118000000000000000
8	1926110000...	18362800000000000000 18776400000000000000 18783400000000000000
9	1926160000...	18395500000000000000
10	1926380000...	18074600000000000000

**ekil 4.3. “Order_Id” Altında Ürünler Birleştirildiğinde Elde Edilen Veri Seti Örneği
(İlk 10) (RapidMiner)**

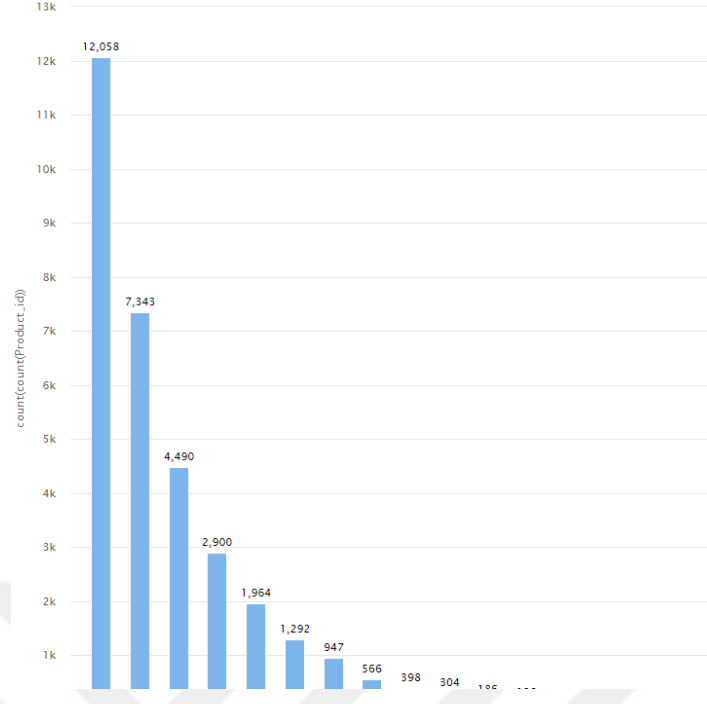
Bu sonuçları veren iş akışı ise Şekil 4.4 ‘de verilmiştir.



Şekil 4.4. FP-Growth Operaötürü Akışı (RapidMiner)

Şekilde 4.4’de verilen akışa ilave olarak “FP-Growth” operatörü yardımıyla sık ürün setleri tespit edilecektir. “FP-Growth” operatörü, Apriori algoritmasının gelişmiş halidir. Çevrimiçi alışverişte “X ürününü alan müşteriler Y ögesini de aldı.” Gibi birliktelik kuralı elde edilmek istenebilir.

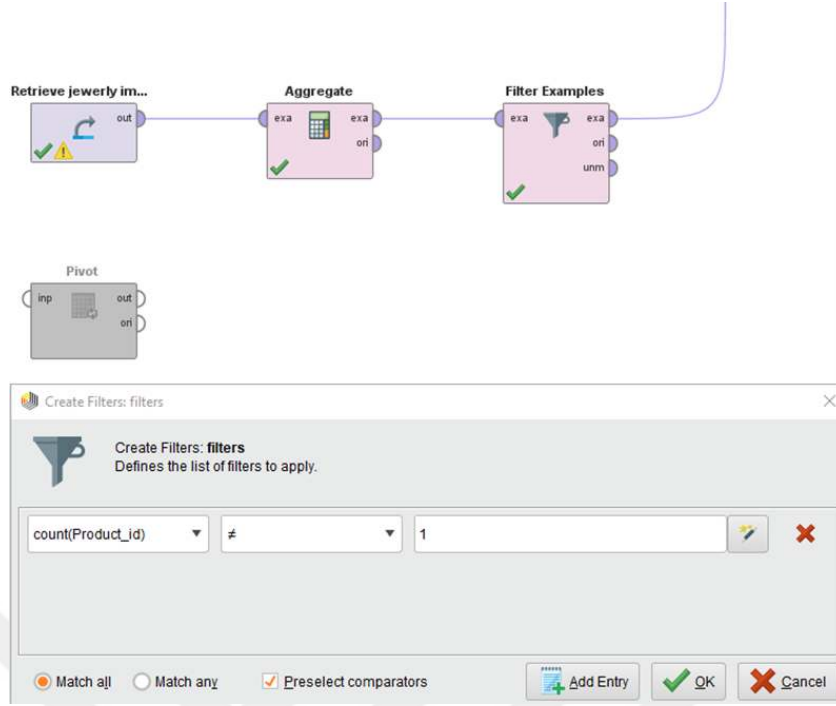
Bunun için, öncelikle müşterilerin sepetlerinde pazardaki hangi ürünlerin sık birlikte bulunduğunu bilmek gerekir ve bunun için de “FP-Growth” algoritması kullanılabilir. FP-Growth kullanılmadan önce “Order_id” bazında alınan ürün sayılarına bakıldığında dağılımın Şekil 4.5’deki gibi olduğu görülmektedir.



Şekil 4.5. Sipariş Başına Satın Alınan Ürün Sayısı (RapidMiner)

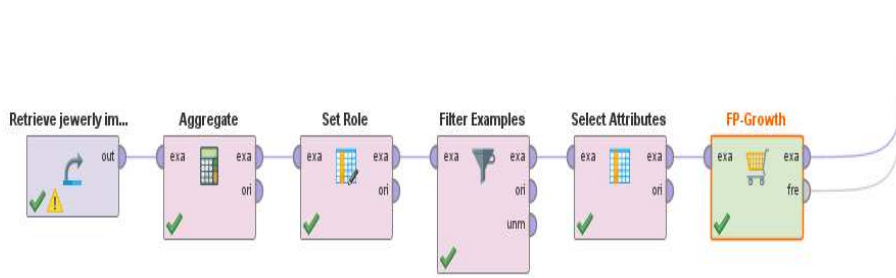
Bu bilgi “Aggregate” operatörüne “Pivot” operatörü eklenmek suretiyle elde edilmiştir. 32910 sipariş içerisinde 12058 siparişte tek bir ürün alınmıştır. Yani bütün siparişlerin %37’sinde tek bir ürün alındığı görülmektedir. Bu sayı iki ürün alınan siparişlerde 7343 olup verinin yaklaşık %60’ını oluşturmaktadır. Üç ürün alınan siparişlerde 4490, dört ürünlerde 2900 ve beşli alınan ürünlerde 1964 olarak devam etmektedir. Bu aşamada amaç sık ürün setlerini bulmak olduğundan sadece bir ürün alınmış siparişlerin veriden çıkarılması gerekmektedir. Bu durum gerçekte birlikte hareket edebilecek ürünleri maskeleymektedir. Aksi takdirde bu veriden bir sık ürün seti bulunması mümkün olmayacaktır.

Bu filtrelemeyi yapmak için “Filter Examples” operatörü kullanılmış, “Aggregate” operatörü tarafından oluşturulan “count(Product_id)” özelliği “1” değerinden farklı olacak şekilde bir ölçüt eklenmiştir. Kullanılan akış ve “Filter Examples” Şekil 4.6’da verilmiştir.



Şekil 4.6. Filtre Examples Operatörünün Kullanımı Ve Filtre Ekleme (RapidMiner)

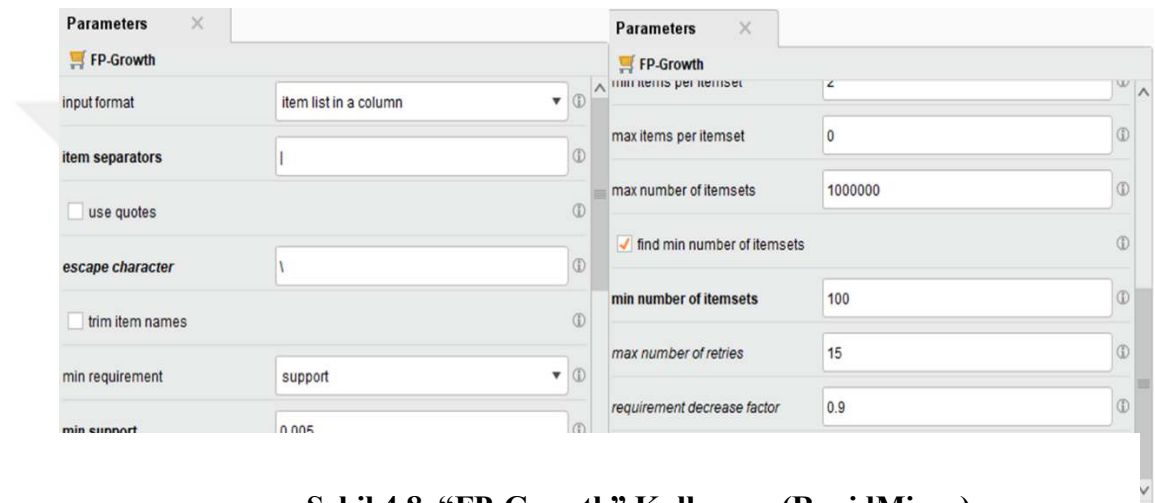
Elde edilen veri setine “FP-Growth” operatörü eklenerek sık ürün setleri elde edilecektir. Bu kapsamda kullanılan akış Şekil 4.7’de verilmiştir.



Şekil 4.7. “FP-Growth Kullanımı (RapidMiner)

Bu akışta “Set Role” operatörü “Order_id” özelliğini “id” yani kimlik özelliğine dönüştürür. Bu “FP-Growth” operatörü için önemlidir. “Select Attributes” operatörü ise “FP-Growth” operatöründe kullanıma gerek kalmayan “count(Product_id)” özelliğini dışarıda bırakmak için kullanılmıştır. “FP-Growth” operatörü için kullanılan parametreler ise Şekil 4.8’deki gibi belirlenmiştir.

Bu parametrelerde “min number of itemsets” seçildiğinde operatör “min support” seçeneğinde seçilen destek eşiğini baz alarak uygun sayıda sık ürün seti hesaplar. Genel olarak, minimum destek eşiği kavramı, bir öge kümesinin sık veya çok sık olmayan oluşumlarının ne anlama geldiğini tanımlayan bir sınır oluşturur. Bir öge veya bir öge seti yalnızca birkaç siparişte görünüyorsa, minimum destek veya minimum frekans parametreleri aracılığıyla hariç tutulur. Nadiren ortaya çıkan öğelerin ve öge kümelerinin hariç tutulması, verilerin sıkıştırılmasına yardımcı olur ve sonuçların istatistiksel önemi artırır.



Şekil 4.8. “FP-Growth” Kullanımı (RapidMiner)

Öte yandan, minimum destek veya minimum frekans değeri çok yüksek ayarlanırsa, algoritma sıfır öge kümesi bulabilir. Bu nedenle, bu operatör, “min number of itemsets” onay kutusu aracılığıyla iki ana mod sağlar. İşaretlenmişse, sabit bir minimum destek değeriyle çalışır, işaretlenmiş ise, sonucun minimum sayıda öge kümesi içerdiğinden emin olmak için dinamik bir minimum destek değeriyle çalışılmış olur. Yine de farklı eşik değerleri için bulunan sık ürün seti sayısı bu setlerdeki ürün sayıları

**Tablo 4.2. Farklı Parametreler İçin Sık Ürün Seti Sonuçları
(Rapidminer)**

Parametre (min suppoort)	Min Parametre Değeri	Maks Parametre Değeri	Toplam Sık Ürün Seti Sayısı	İkili Set Sayısı	3'lü Set Sayısı
0,100	0,022	0,424	6	5	1
0,050	0,011	0,424	10	8	2
0,030	0,007	0,424	16	12	4
0,020	0,004	0,424	28	22	6
0,010	0,00206	0,424	96	77	19
0,005	0,00197	0,424	104	85	19

Bu tablodan da anlaşılmaktadır ki “min support” parametresinin 0,02 değerinden sonra kullanılan parametre değerlerinde sonuçlar artmaktadır. Burada bir kırılma noktasının olduğu görülmektedir. Önemsiz olan ilişkilerin de burada görünmesini engellemek açısından parametre değeri 0,02 olarak seçilmiş ve sık ürün setleri Tablo 4.3’ de sunulmuştur.

Tablo 4.3. Sık Ürün Setleri (RapidMiner)

S/N	Boyut	Destek	Ürün1	Ürün2	Ürün3
	2	0,4243	195666	151597	-
	2	0,0507	195666	189722	-
	2	0,0335	195666	180683	-
	2	0,0160	195666	131368	-
	2	0,0114	195666	131361	-
	2	0,0101	195666	135291	-
	2	0,0074	195666	192458	-
	2	0,0078	195666	194495	-
	2	0,0061	195666	131356	-
	2	0,0053	195666	194442	-
	2	0,0048	195666	131357	-
	2	0,0052	195666	183557	-
	2	0,0044	195666	145509	-
	2	0,0045	195666	193790	-
	2	0,0041	195666	189655	-
	2	0,0316	151597	189722	-
	2	0,0221	151597	180683	-
	2	0,0116	151597	131368	-
	2	0,0078	151597	131361	-
	2	0,0061	151597	135291	-
	2	0,0058	151597	194495	-
	2	0,0042	151597	131356	-
	3	0,0261	195666	151597	189722
	3	0,0191	195666	151597	180683
	3	0,0093	195666	151597	131368
	3	0,0069	195666	151597	131361

Bu tabloda örneğin birinci sıradaki sık ürün setinde 0,4243 destek değeri ile yani işlemlerin %42,43’ünde ürün 195666 ile ürün 151597 birlikte görülmektedir. Aynı tabloda 23’üncü sıradaki sık ürün setinde ise ürün 195666, ürün 151597 ve ürün 189722 işlemlerin %2,6’sında birlikte görülmektedir. Bu sonuçlar firmanın pazarlama stratejilerinde kullanılmak üzere ilgili bölümlerin kullanımına sunulabilmektedir.

Bu aşamada belirlenen sık ürün setlerinden birliktelik kurallarının çıkarılmasına çalışacaktır. Bu amaçla kullanılacak operatör ise “Cretate Association Rules” dur. Operatörün çalışması için Tablo 4.4’de verilen parametrelerin seçilmesi gerekir.

Tablo 4.4. “Create Association Rules” Operatöründe Kullanılan Parametreler (RapidMiner)

Parametre	Anlamı	Değerler	Seçilen Değer
criterion	Kuralın hangi kritere göre bulunacağını gösterir	Confidence, lift, conviction, ps, gain, laplace	Confidence (default)
min confidence	Güven için eşik değeri	0-1 arasında	0,80 (default)
gain theta	Kazanım hesabında kullanılan katsayıdır. θ ile $1 - \theta$ arasında değişen değerler alır. Ne kadar yüksek ise o kadar iyidir.	Serbest	2 (default)
Laplace k	$\frac{1}{1+k} < laplace < \frac{2}{1+k}$ eşitsizliği için seçilen bir $k > 1$ ile ifade edilir. $laplace(X \rightarrow Y) = (supp(X \rightarrow Y) + 1) / (supp(X) + k)$ ile hesaplanır.	Serbest	1 (default)

Bu parametreler ile çalışıldığında elde edilen ilişki kurallarının bir kısmı Tablo 4.5’de verilmiştir.

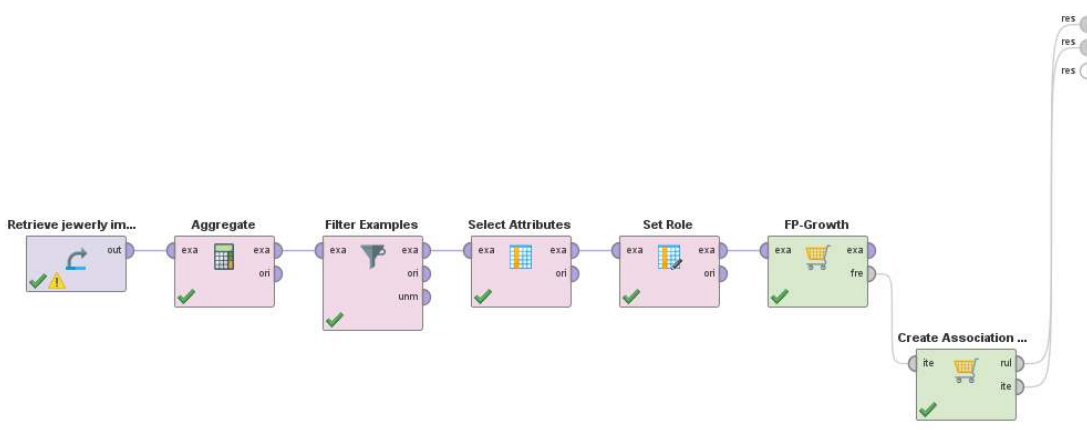
Tablo 4.5. “Create Association Rules” Operatörü Sonuçlarının Bir Kısım

000000000, 13135600000000000000	19566600000000000000	0.00345	0.82759	0.99928	-0.00489	0.00345	∞	5.80000
000000000, 18068300000000000000	19566600000000000000	0.01909	0.86522	0.99709	-0.02503	0.01909	∞	7.41935
000000000, 13529100000000000000	19566600000000000000	0.00542	0.88281	0.99929	-0.00686	0.00542	∞	8.53333
000000000, 13136100000000000000	19566600000000000000	0.00691	0.88344	0.99910	-0.00873	0.00691	∞	8.57895
000000000, 19449500000000000000	19566600000000000000	0.00513	0.89167	0.99938	-0.00638	0.00513	∞	9.23077
000000000	15159700000000000000	0.42432	∞	1.42432	0.42432	0.42432	∞	-0
000000000	19566600000000000000	0.42432	∞	1.42432	0.42432	0.42432	∞	-0

Bu tabloda oluşan kurallar güçlü ilişki kuralları olarak adlandırılır. Ancak burada sorgulanması gereken “lift” değerine bakılır ancak bu değer program tarafından sonsuz olarak hesaplanmaktadır. Bunun sebebi kuralın solunda ve/veya sağında kalan ürün ve/veya ürünlerin destek değerinin sıfır olmasıdır. Yani ürün sadece ürün ikilisi ile görülmektedir.

Tablo 4.5’de “support” ve “confidence” yanında diğer ölçüler de görülmektedir. Bu ölçülerden “LaPlace” $(supp(X \rightarrow Y) + 1)/(supp(x) + k)$ hesaplanır ve “confidence” ile benzerdir. Parametre ekranında değeri verilmektedir. “gain” ölçüsü de “confidence” ile benzerdir ve $gain(X \rightarrow Y) = supp(X \rightarrow Y) - \theta * supp(X)$ i hesaplanır. “p-s” (Piatetski-Shapiro) yapıldığında elde edilen “gain” değeridir. “lift” ölçümü ise $lift(X \rightarrow Y) = supp(X \cup Y)/(supp(Y).supp(X))$ ile hesaplanır. Son olarak “conviction” değeri $conv(X \rightarrow Y) = (1 - supp(Y))/(1 - conf(X \rightarrow Y))$ olarak hesaplanmaktadır.

Esasen “lift” değeri “1” değerini alırsa kuralın solunda ve sağında bulunan ürün setlerinin bağımsız olduğu sonucuna varılmaktadır. Bu durumda “confidence” ile hesaplanan kuralların hepsi ilginç kural da olmaktadır. Kullanılan akış Şekil 4.9’da verilmiştir.



Şekil 4.9. “Create Association Rules” Operatörünün Kullanımı

4.4. Ürünlerin Satın Alınma Sırasının Belirlenmesi

Pazar sepeti araştırmalarında yapılan çalışmalardan biri de satın alma işlemlerinin incelenerek ürünlerin hangi sırayla satın alındığının belirlenmesidir. Bu sonuca ulaşmak için satın alma işlem listesinin yani bu konuda kullanılan veri setinin aşağıdaki işlemlerden geçirilmesi gerekir.

- a) İşlem tarih/saat özelliğinin unix tarihlendirme sistemine göre düzenlenmesi: Bu düzenlemede veri setine, 01 Ocak 1970 tarihi ile işlem tarihi arasındaki fark eklenmektedir.(Http15,T.Y.) Kullanılan iş akışında söz konusu tarih yerine en erken yapılan işlem tarihi ile diğer alışveriş tarihleri arasındaki fark alınmıştır. Bu işlemin amacı yapılan işlemin sıralanmasında kullanılacak olan sabit bir noktadan uzaklık olduğundan, en erken tarihin sabit tarih olarak belirlenmesi bir soruna sebep olmamıştır. Bu özellik veri setinde “Diffrence” olarak isimlendirilmiştir. “Diffrence” özelliğinin eklendiği veri seti Tablo 4.6’da gösterilmiştir.

Tablo 4.6. “Diffrence” Özelliği Eklenmiş Veri Seti Örneği

Row No.	Date/time	Earliest	Difference	Order_id	Product_id	User_id
1	Jan 21, 2019 ...	Dec 1, 2018 ...	51.35034	1961580000...	1313550000...	1515920000...
2	Mar 13, 2019 ...	Dec 1, 2018 ...	102.35417	1998550000...	1313550000...	1515920000...
3	May 9, 2019 5...	Dec 1, 2018 ...	159.61713	2040050000...	1313550000...	1515920000...
4	May 20, 2019 ...	Dec 1, 2018 ...	170.14442	2047680000...	1313550000...	1515920000...
5	Jun 7, 2019 7...	Dec 1, 2018 ...	188.67228	2061110000...	1313550000...	1481120000...
6	Jun 19, 2019 ...	Dec 1, 2018 ...	200.18458	2069460000...	1313550000...	1515920000...

- b) “Order_id” , “Product_id” ve “User_id” birer “id” veri tipi olarak işaretlenmelidir.
- c) Yukarıdaki adımlardan sonra sütunların “User_id” , “Difference” ve “Product_id” olarak belirlendiği ve “User_id” lerin ise sayıldığı bir pivot tabloya dönüştürme işlemi yapılmıştır. Oluşan pivot tablo Tablo 4.7’de gösterilmiştir. Veri çok büyük olduğundan sadece örnek bir kısım gösterilmiştir.

Tablo 4.7. Veri Setinden Üretilen Pivot Tablo Örneği (RapidMiner)

Row No.	User_id	Difference	count(User_id)_1313550000000000000	count(User_id)_1313560000000000000	c
1	1515920000...	51.35034	1	?	?
2	1515920000...	102.35417	1	?	?
3	1515920000...	159.61713	1	?	?
4	1515920000...	170.14442	1	?	?
5	1481120000...	188.67228	1	?	?
6	1515920000...	200.18458	1	?	?
7	1541240000...	206.61307	1	?	?
8	1515920000...	259.53718	1	?	?

Tablo 4.7’den de anlaşılacağı gibi, pivot tablo kullanıcının belirli bir işlemde belirli bir ürünü alması durumunda bu ürüne ait özelliği “1” yapmakta, eğer ürün alınmadı ise “?” ile göstermektedir. Ürünlerin satın alınma sırasını bulabilmek için bu pivot tablonun ürünleri gösteren özellikleri altındaki bilginin sırasıyla “True” ve “False” şeklinde işaretlenmesi gerekmektedir. Bu durum, sonraki operatörlerin veriyi işlemesi için gereklidir.

Verideki “?” ile işaretlenmiş yerlerin sonraki adımlarda işlenebilmesi için “0” değeri ile değiştirilmesi: Bu amaçla “Replace Missing Values” operatörü kullanılmıştır. Bu aşamada verinin son hali Tablo 4.8’ de sunulmuştur.

Tablo 4.8. “1” ve “?” Yerine Sırasıyla “True” ve “False” Dönüşümü (RapidMiner)

Row No.	User_id	count(User_...	count(User_...	count(User_...	count(User_...
1	1515920000...	true	false	false	false
2	1515920000...	true	false	false	false
3	1515920000...	true	false	false	false
4	1515920000...	true	false	false	false
5	1481120000...	true	false	false	false
6	1515920000...	true	false	false	false
7	1541240000...	true	false	false	false

- d) Tablo 4.7'deki "1" ve "?" yerine sırasıyla "True" ve "False" kullanılması: Bunun için veri setine "Numerical to Binomal" operatörü uygulanarak istenen değişiklik yapılmıştır. Verinin değişiklik yapılmış hali Tablo 4.8'de verilmiştir.
- e) Son olarak sütun isimlerindeki fazlalık silinmiş ve sadece "Product_id" lerin görünmesi sağlanmıştır. Bu işlemin sıra bulma süreci üzerindeki etkisi sadece sadeleştirmeden ibarettir.

Yukarıda tanımlanan işlemler aslında bir ön işleme çalışması olarak tanımlanabilir. Bu aşama veri madenciliği çalışmalarında çalışmanın performansını belirlemede ilk sırada yer alır. Bu aşamanın kalitesi çalışmanın da kalitesini belirlemesi açısından son derece önemlidir.

Sonraki aşamada "Generalized Sequential Patterns" kısaca "GSP" olarak bilinen operatör kullanılacaktır. Bu operatör, GSP (Genelleştirilmiş Sıralı Model) algoritmasını kullanarak bir dizi işlemde sıralı kalıpları arar.

GSP, dizi madenciliği için kullanılan popüler bir algoritmadır. Bu operatör, bir dizi işlemde sıralı kalıpları arar. Bu operatörün parametreleri Şekil 4.10'da gösterilmiş ve aşağıda açıklanmıştır.

Parameter	Value
customer id	User_id
time attribute	Difference
min support	0.05
window size	1.0
max gap	30.0
min gap	1.0
positive value	true

**Şekil 4.10. GSP Operatörü Parametreleri
(RapidMiner)**

Veri seti, zaman ve müşteri için bir özellik içermelidir. Ayrıca her işlem tek bir nesne olarak kodlanmalıdır. Zaman ve müşteri özellikleri, her bir işlemin zamanına göre sipariş edilen her işlemi içeren müşteri başına bir sıra oluşturmak için kullanılır. “customer_id” müşteri kimliğini belirtmek için kullanılır. Veri setimizde bu bilgi “User_id” ile verilmektedir. “time attribute” parametresi ise unix formatında zaman farkını gösteren özelliğin girilmesi için kullanılır.

Veri setimizde bu özellik “Difference” ile tanımlanmıştır. “min support” parametresi sıranın, verilen en küçük destek değerini aşması durumunda operatör tarafından desen olarak algılanmaktadır. Algoritma sıralı kalıpları şu şekilde arar: “Bir müşteri bir işlemde ‘a’ ve ‘c’ öğelerini satın aldıysa, bir sonraki işlemde ‘b’ öğesini satın aldı.” Bu kalıp şu biçimde temsil edilir: <a, c> sonra . En küçük destek, sık görülmesi için böyle bir modeli kaç müşterinin desteklenmesi gerektiğini açıklar. Seyrek desenler atılır. Bir müşteri, dizisinin bu kalıbı içeren bazı bölümler varsa, böyle bir kalıbı destekler. Yukarıdaki model bir müşteri tarafından örneğin şu işlemlerle desteklenir : <s, g> sonra <a, s, c> sonra sonra <f, h>.

Minimum boşluk (min gap), maksimum boşluk (max gap) ve pencere boyutu (window size) parametreleri, satın alma işlemlerin nasıl ele alınacağını belirler. Örneğin, yukarıdaki müşteri ‘c’ öğesini almayı unutursa ve satın almak için 5 dakika sonra geri dönmek zorunda kalırsa, işlemleri şöyle görünür: <s, g> sonra <a, s> sonra <c> sonra sonra <f, h>. Bu, <a, c> ve ardından modelini desteklemez. Bu sorunu önlemek için, pencere boyutu, sonraki bir işlemin ne kadar süreyle aynı işlem olarak ele alınacağını belirler. Pencere boyutu 5 dakikadan büyükse, <c> ikinci işlemin parçası olarak kabul edilir ve bu nedenle bu müşteri yukarıdaki modeli destekler. Maksimum boşluk parametresi, bu kalıbı içeren işlemler zaman içinde çok geniş bir şekilde ayrılmışsa, bir müşteri dizisinin bir kalıbı desteklememesine neden olur. Minimum boşluk parametresi, çok yakınlarsa aynı şeyi yapar. Bu üç değer sırasıyla “1” , “30” ve “1” olarak seçilmiştir.

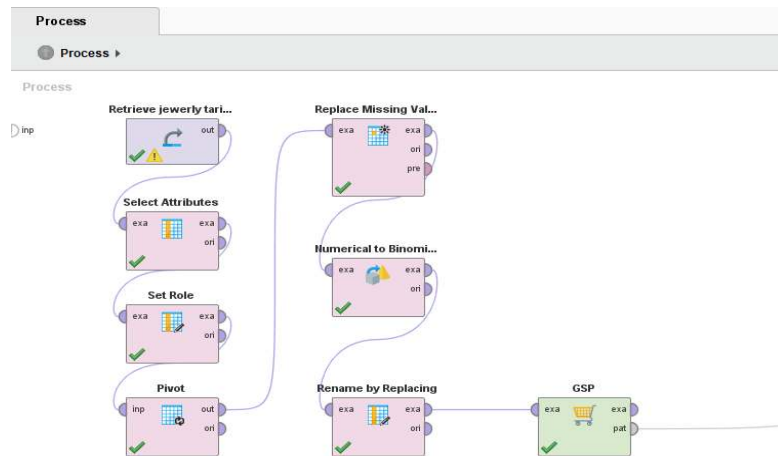
“Difference” özelliği gün cinsinden verildiğinden zaman penceresinin “1” seçilmesi aynı gün geri dönüp satın alma işlemini aynı satın alma işlemi içerisinde değerlendirmesini sağlar. “max gap” 30 seçildiğinde herhangi bir kalıp 30 gün içerisinde ise desteklenir. “min gap” değerinin 1 seçilmesi bu amaçla ise alt sınır

olarak belirlenir. Bu şekilde belirlenen parametreler ile elde edilen sonuç Tablo 4.9’ da verilmiştir.

Tablo 4.9. GSP Algoritması Sonuçları(RapidMiner)

Support ↓	Transa...	Items	Transaction 0	Transaction 1	Transaction 2	Transaction 3	Transaction 4	Transaction 5
0.17424	2	2	19566600000000000000	19566600000000000000				
0.14899	1	2	15159700000000000000, 19566600...					
0.13384	2	2	15159700000000000000	19566600000000000000				
0.12626	2	2	19566600000000000000	15159700000000000000				
0.11364	2	2	15159700000000000000	15159700000000000000				
0.08586	3	3	19566600000000000000	19566600000000000000	19566600000000000000			
0.07071	2	3	19566600000000000000	15159700000000000000, 19566600...				
0.06566	3	3	15159700000000000000	19566600000000000000	19566600000000000000			
0.05808	3	3	19566600000000000000	15159700000000000000	19566600000000000000			
0.05808	4	4	19566600000000000000	19566600000000000000	19566600000000000000	19566600000000000000		
0.05556	2	3	15159700000000000000, 19566600...	19566600000000000000				
0.05556	5	5	19566600000000000000	19566600000000000000	19566600000000000000	19566600000000000000	19566600000000000000	
0.05303	6	6	19566600000000000000	19566600000000000000	19566600000000000000	19566600000000000000	19566600000000000000	19566600000000000000
0.05051	2	3	15159700000000000000, 19566600...	15159700000000000000				

Yukarıdaki süreci ve GSP operatörü kullanılan süreçte aşağıdaki şekilde verilmiştir. Bu sonuçlara göre sadece “15166” ve “151597” arasında bir sıra ilişkisi görülmektedir. Örneğin en son satırda verilen ve destek değeri 0,05051 olan sıralamada 2 işlemde 3 ürün önce “151597” ve “195666” dan sonra “151597” gelmektedir. Bu sonuçların girilen parametrelerle elde edilen sonuçlar olduğunu, farklı parametreler ile daha farklı ancak uzun işlem süreleri gerektiğini bir kez daha hatırlamak gerekir. Verilerin ilk adımından son adımına kadar geçen süreç Şekil 4.11’de gösterilmiştir.



Şekil 4.11. Kullanılan Rapidminer İş Akışı

SONUÇ

Orta ölçekli bir çevrimiçi mücevherat mağazasından alınan veriler incelenerek RapidMiner programı ile düzenlendi. Bu düzenleme sonucunda ürün özelliklerinde eksik veriler tespit edildi. Eksik görülen verileri tamamlamak için GBT(“Gradient Boosted Trees”) operatörü kullanıldı. Eksik verilerin tamamlanma aşaması şöyledir: Sayısal değerler nominal değere dönüştürüldü. Kategorin adı ve id’si verilerden çıkarıldı. GBT operatörü seçilerek eksikler giderildi. Tamamlanan set kullanıma hazır hale geldi. Verileri “Order_id” adı altında birleştirip sık ürün setleri oluşturuldu. Oluşturulan sık ürünlerin %37,9’si tekli alınan ürünlere ait veriler idi. Bunlar “Filter Examples” operatörü ile devre dışı bırakılarak iki ve katları ile alınan ürün alışverişlerini tespit etmek kolaylaştırıldı. Bir ürün veya ürün seti yalnızca birkaç siparişte görünüyorsa, minimum destek parametreleri tespit edildi. Ve bu tespit sonucunda 0,02 değeri sonrasında kırılmaların oluştuğu görüldü. Min. Support değeri 0,02 olarak belirlenmiş oldu. Bu değer sonucunda 195666 ile 151597 isimli ürünler %43 birlikte alındığı bilgisi edinilmiştir. Son aşamada ürünlerin satın alınma sıraları ele alınmıştır. Bu aşamada GSP operatörü kullanılarak, bir dizi sıralı işlemler arandı. Araştırmalar sonucunda 151666 ve 151597 isimli ürünlerin arasında ilişki olduğu görüldü. İşlemler devam ettikçe muhakkak daha farklı ikili setlerde meydana gelecektir. Fakat bizim için bu kadarı yeterli oldu. Genel anlamda yapmış olduğumuz çalışma için seçtiğimiz yöntemlerin ve programların doğru olduğunu söyleyebiliriz. Doğru teknik ve operatörler, çevrimiçi satış yapan bu firmanın hangi ürünlerin birlikte gösterilmesi gerektiği veya pazarlanması gerektiğini bulmuş oldu.

KAYNAKÇA

Kitap yayınları:

SİLAHTAROĞLU Gökhan, 2020, Bilgisayar Bilimleri Veri Madenciliği Kavram ve Algoritmaları (4.Basım), Papatya Bilim Akademik Yayınevi, Esenyurt/İstanbul

İnternet Yayınları:

ATEŞ Vildan, 2018, Online Müşteri Sadakatini Etkileyen Müşteri Algılarının İncelenmesi: Gazi Üniversitesi Örneği, Dergipark, Ankara

GÜRSOY Umman, AKÇAY Özlem, ATALAY Kutluk, , 2019, R Programlama ile Birliktelik Kuralları Analizi: Tüketicilerin İnternet Üzerinden Yaptıkları Alışveriş Verisinin Apriori ve Eclat Algoritmalarıyla İncelenmesi, Dergipark

TİMOR Mehpere, EZERÇE Ayşegül, GÜRSOY U. Tuğba, 2011, Müşteri Profili Ve Alışveriş Davranışlarını Belirlemede Kümeleme Ve Birliktelik Kuralları Analizi: Perakende Sektöründe Bir Uygulama, Dergipark

ULUSAN POLAT Meltem, BABAOĞLU Mehmetcan, 2021, Türkiye’de Online Kitap Satışlarının Veri Madenciliği Yöntemiyle Analizi, Dergipark

YAŞIN Bahar, GÜLEÇ Melike, ZAİD Esra, 2019, İnternet Perakendeciliğinde Elektronik Hizmet Kalitesinin Müşteri Tatmini ve Müşteri Sadakati Üzerindeki Rolü, Dergipark

YILDIZ Arif, KUYUCU Şeyda, 2021, İnternet Alışverişlerinde Algılanan E-Hizmet Kalitesi, Müşteri Tatmini ve Müşteri Sadakati İlişkisinin İncelenmesi: Adıyaman Üniversitesi Öğrencileri Üzerine Bir Uygulama, Dergipark, Adıyaman

İnternet/Kurum:

ETİCARET, 2022, 2021 Yılı E-Ticaret Verileri Açıklandı, <https://www.eticaret.gov.tr/haberler/10040/detay#:~:text=2021%20y%C4%B1l%20B11%20C3%BCIkemizde%20e%2Dticaret,%17%2C7%20olarak%20ger%C3%A7ekle%C5%9Fti.> , 8/04/2022

GTECH,2021,Veri Madenciliği Nedir ve Nasıl Yapılır?, <https://www.gtech.com.tr/veri-madenciligi-nedir-ve-nasil-yapilir/>, 30/12/2021

Http1, 2020, Veri Uygulama Süreçleri, <https://www.medium.com>

Http2,2020,Veri Madenciliği Teknikleri, <https://medium.com/@eminurnacar/veri-madencili%C4%9Finde-pazar-sepeti-analizi-de44962f2086>

Http3,2017,Karar Ağacı, <https://www.medium.com>

Http4, 2018, Kümeleme Dağılım Grafiği, <https://www.medium.com>

Http5, 2020, FP-Growth Açıklaması, <https://tr.myservername.com/frequent-pattern-growth-algorithm-data-mining>

Http6, 2020, FP-Growth 1.Adım, <https://towardsdatascience.com/understand-and-build-fp-growth-algorithm-in-python-d8b989bab342>

Http7, 2020, FP-Growth 2.Adım, <https://towardsdatascience.com/understand-and-build-fp-growth-algorithm-in-python-d8b989bab342>

Http8, 2020, FP-Growth 3.Adım, <https://towardsdatascience.com/understand-and-build-fp-growth-algorithm-in-python-d8b989bab342>

Http9, 2020, Birlikte alınma olasılık hesaplaması,

<https://www.datascienceearth.com/birliktelik-kurallari-yonteminde-matematiksel-surec/>

Http10, 2016, Pazar Sepet Analizi, <https://www.slideshare.net/uslumetin/birliktelik-kurallar-kullanılarak-pazar-sepeti-analizi-market-basket-analysis-using-association-rules>

Http12, 2022, E-Ticaret Nasıl Yapılır?, <https://www.ideasoft.com.tr/e-ticaret-nasil-yapilir-asamalari-nelerdir/>

Http13, 2019,Dünya Geneline E-Ticaret Dağılımı, <https://www.cankiripostasi.com/bilimteknoloji/nielsen-raporu-2019-h8883.html>

Http14, 2022, Yardım Dosyası, <https://rapidminer.com/help/>, 30/12/2021

Http15, 2019, Unix Tarihlendirme Sistemi, T.Y., <https://www.unixtimestamp.com/>

KAGGLE, T.Y, Kullanılan Esas Veri,<https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-jewelry-store?select=jewelry.csv>, 25/12/2021

Vizyoner GENÇ, 2019, 5 Temel Soruda Veri Madenciliği, <https://vizyonergenc.com/icerik/5-temel-soruda-veri-madenciligi-data-mining-nedir>

Wikipedia:

Http11, E-Ticaret Nedir? 2022, https://tr.wikipedia.org/wiki/Elektronik_ticaret, 20/04/2022

