

Birliktelek Kuralı Yöntemleri ile E-Ticaret Satıřlarının Analizi




Yavuz DEMİROK
161402203

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliđi Anabilim Dalı
Bilgisayar Mühendisliđi Tezli Yüksek Lisans
Danışman: Dr. Öğr. Üyesi Volkan TUNALI

İstanbul
T.C. Maltepe Üniversitesi
Fen Bilimleri Enstitüsü
Aralık, 2018

JÜRİ VE ENSTİTÜ ONAYI


Yavuz DEMİROK'un "BirlikteKuralı Yöntemleri ile E-ticaret Satışlarının Analizi" başlıklı tezi 28.12.2018 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Maltepe Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği"nin ilgili maddeleri uyarınca, Bilgisayar Mühendisliği Anabilim Dalında Yüksek Lisans/Doktora tezi **oy birliğiyle / oy çokluğuyla** olarak kabul edilmiştir.

Unvanı, Adı ve soyadı	İmza
Üye (Tez Danışmanı) : Dr. Öğr. Üyesi Volkan TUNALI	
Üye : Dr. Öğr. Üyesi Mehmet Ali Aksoy TÜYSÜZ	
Üye : Doç. Dr. Turgay Tugay BİLGİN	



Prof. Dr. İter BÜYÜKDIĞAN

Enstitü Müdürü

 maltepe üniversitesi	ETİK İLKE VE KURALLARA UYUM BEYANI	Doküman No	FR-178
		İlk Yayın Tarihi	01.03.2018
		Revizyon Tarihi	
		Revizyon No	00
		Sayfa	1/1

Revizyon Takip Tablosu

REVİZYON NO	TARİH	AÇIKLAMA
00	01.03.2018	İlk yayın.

ETİK İLKE VE KURALLARA UYUM BEYANI

28/12/2018

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarından bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilmeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; çalışmamın Maltepe Üniversitesinde kullanılan "bilimsel intihal tespit programı" ile tarandığını ve öngörülen standartları karşıladığını beyan ederim.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.



Yavuz DEMİROK

Hazırlayan İlgili Birim	Kalite Koordinatörü Dr. Öğr. Üyesi Şafak GÜNDÜZ	Kurumsal Yetkili Prof. Dr. Belma AKŞİT
----------------------------	--	---

(Doküman No: FR-178; Yayın Tarihi: 01.03.2018; Revizyon Tarihi: ; Revizyon No:00)

İNTİHAL RAPORU

Birliktelik Kuralı Yöntemleri ile E-Ticaret Satışlarının Analizi

ORJİNALLİK RAPORU

%**5**

BENZERLİK ENDEKSİ

%**3**

İNTERNET
KAYNAKLARI

%**0**

YAYINLAR

%**4**

ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	%1
2	sbe.maltepe.edu.tr İnternet Kaynağı	%1
3	Submitted to Marmara University Öğrenci Ödevi	%1
4	Submitted to Al-Madinah International University (MEDIU) Öğrenci Ödevi	<%1
5	Submitted to Istanbul Aydın University Öğrenci Ödevi	<%1
6	hentbolokulu.com İnternet Kaynağı	<%1
7	Submitted to Anadolu University Öğrenci Ödevi	<%1
8	ERPOLAT, Semra. "Otomobil yetkili servislerinde birliktelik kurallarının	<%1

28.12.2018



Dr. Öğr. Üyesi Volkan TUNALI

TEŐEKKÜR

Bu tez alıőmamda, tez danıőmanım olan ve her fırsatta desteęini esirgemeyen ve doęru ynlendirmeleriyle alıőmaya teővik eden Sayın Dr. ęr. yesi Volkan TUNALI'ya teőekkrlerimi sunuyorum.

Aynı zamanda bu alıőmaya konu olan ve analizini yaptığım veriyi bana temin eden ve kullanmam konusunda izin veren Sayın Glden OKUR'a teőekkr ediyorum.

Yavuz DEMİROK

Aralık 2018



ÖZ

Birliktelik Kuralı Yöntemleri ile E-Ticaret Satışlarının Analizi

Yavuz Demirok

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Tezli Yüksek Lisans

Danışman: Dr. Öğr. Üyesi Volkan TUNALI

Maltepe Üniversitesi Fen Bilimleri Enstitüsü, 2018

Veri madenciliği yöntemleri ile her geçen gün büyüyen veritabanı bilgilerinin analiz edilmesi daha da kolaylaşmıştır. Bu çalışma ile bir e-ticaret firmasındaki müşterilerin aldığı ürünlerin veri madenciliği teknikleri ile en çok tercih edilen ve birlikte alınan ürünler olarak analizi yapılmış ve sonuçları incelenmiştir. Bu analiz veri madenciliğinde kullanılan bir yöntem olan ve birliktelik kuralları algoritmalarından Apriori Algoritması kullanılarak yapılmıştır. Birliktelik kuralı; müşteri tarafından hangi sıklıkla hangi ürünlerin aynı anda bir arada alındığını ve bu ürün setlerini belirlemek için kullanılan bir yöntemdir. Müşterilerin karakteristik özelliklerine ve kullanım alışkanlıklarına yönelik hangi müşteri, hangi ürünlere, ne zaman ihtiyaç duyduğu incelenmiş ve analiz sonuçları değerlendirilmiştir. Burada iki türlü amaç vardır; yaygın olarak tercih edilen ürünler ve müşterilerin en sık tercih ettiği ürün gruplarının tespiti.

Anahtar Kelimeler: Birliktelik Kuralı, Birlikte Alınan Ürünler, Market Sepet Analizi, Apriori, KNIME

ABSTRACT

Analysis of E-Commerce Sales with Association Rule Methods

Yavuz Demirok


Master Thesis

Computer Engineering

Master of Science in Computer Engineering With Thesis

Thesis Advisor: Assist. Prof. Dr. Volkan TUNALI

Maltepe University Graduate School of Science and Engineering, 2018



With data mining methods, it is even easier to analyze database information that grows day by day. In this study, an e-commerce customer's products were analyzed with data mining techniques by analyzing the most preferred and co-purchased products. This analysis is done by using the algorithm of Apriori Algorithm, which is a method used in data mining. The association is constitutional; is a method used by the customer to determine how often products are taken together at the same time and to determine these product sets. The customer's characteristics and usage habits were investigated and the results of analysis were evaluated. There are two purposes here; the most commonly preferred products and the most frequently preferred product groups of customers.

Keywords: Association Rules, Co-Purchased Products, Market Basket Analysis, Apriori, KNIME

İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI	i
ETİK İLKE VE KURALLARA UYUM BEYANI	ii
TEŞEKKÜR.....	iv
ÖZ	v
ABSTRACT.....	vi
İÇİNDEKİLER	vii
TABLolar LİSTESİ.....	ix
ŞEKİLLER LİSTESİ	x
ÖZGEÇMİŞ	xi
BÖLÜM 1. GİRİŞ.....	1
1.1. Problem	1
1.2. Amaç	1
1.3. Önem.....	2
1.4. Sınırlıklar	3
BÖLÜM 2. BİRLİKTELİK KURALI ANALİZİ.....	4
2.1. Araştırma Modeli	4
2.1.1. Veri Nedir, Bilgi Nedir?.....	4
2.1.2. Veri ve Bilgi arasındaki temel farklılıklar:	4
2.1.3. Veritabanı Kavramı	5
2.1.4. Veri Ambarı Nedir?.....	5
2.1.5. Veri Ambarı Bileşenleri	6
2.2. Veri Madenciliği	9
2.2.1. Veri Madenciliği Ne Demektir?	9
2.2.2. Veri Madenciliğinin Uygulama Alanları.....	11
2.2.3. Veri Madenciliği Süreçleri	13
2.2.4. Veri Madenciliği İşlevleri ve Yöntemleri	14
2.2.5. Sınıflama ve Regresyon Modelleri.....	15
2.2.6. Kümeleme Analizi Modeli	16
2.2.7. Birliktelik Analizi Modeli	16
2.3. Birliktelik Kuralları ve İlişki Analizi	17
2.3.1. Market Sepet Analizi.....	17
2.3.2. Birliktelik Kuralları	17
2.3.3. Birliktelik Kuralı Algoritmaları	19

2.3.4. Apriori Algoritması	20
2.3.5. Apriori ve FP-Growth Algoritmaları.....	24
2.3.6. Apriori ve FP-Growth Algoritmaları Karşılaştırma	25
BÖLÜM 3. YÖNTEM ve UYGULAMA.....	27
3.1. Veriler ve Toplanması.....	27
3.1.1. KNIME Uygulaması ile Detayları.....	27
3.1.2. KNIME Analytics Platform	29
3.1.3. KNIME Server	31
3.1.4. Birlikte Alınan Ürünler Analizi ve İnceleme	31
BÖLÜM 4. BULGULAR VE YORUMLAR	52
4.1. Bulgular.....	52
4.2. Yorumlar	53
BÖLÜM 5. SONUÇ	54
5.1. Özet	54
5.2. Yargı.....	54
5.3. Öneriler	55
KAYNAKÇA.....	56

TABLULAR LİSTESİ

Tablo 2.1 - Market Basket Ürün Tablosu	18
Tablo 2.2 - Market Basket Ürün Tablosu	21
Tablo 2.3 - Müşteri Alışveriş Sepeti	23
Tablo 2.4 - Ürün Frekans Değerleri	23
Tablo 2.5 - İkili Ürün Frekans Değerleri	23
Tablo 2.6 - Üçlü Ürün Frekans Değerleri	24
Tablo 2.7 - Apriori ve FP-Growth Karşılaştırma Tablosu [27]	25
Tablo 3.1 - Tez Çalışmasına Konu Olan Veri Örneği	27
Tablo 3.2 - Analizi Yapılan Örnek Veri	32
Tablo 3.3 - GroupBy Node'u İle Tekilleştirilecek Order Number Örneği	37
Tablo 3.4 - 6'lı Birliktelik İlk 10 Listesi	43
Tablo 3.5 - 5'li Birliktelik İlk 10 Listesi	44
Tablo 3.6 - 4'lü Birliktelik İlk 10 Listesi	45
Tablo 3.7 - 3'lü Birliktelik İlk 10 Listesi	46
Tablo 3.8 - 2'li Birliktelik İlk 10 Listesi	47
Tablo 3.9 - En Fazla Satılan Ürüne Göre Destek ve Güven İlk 10 Listesi	49
Tablo 3.10 - En Fazla Güven Değerine Göre İlk 10	50
Tablo 3.11 - En Fazla Satılan İlk 10 Ürün Listesi	51

ŞEKİLLER LİSTESİ

Şekil 2.1 - Veri Ambarı Konsepti	6
Şekil 2.2 - OLAP Küp [8]	8
Şekil 2.3 - Veri Ambarı Bileşenleri	9
Şekil 2.4 - Veri Madenciliği ve Bilgiye Ulaşma [13]	10
Şekil 2.5 - Verilerin Bilgiye Dönüştürülme Süreci.....	10
Şekil 2.6 - Veri Madenciliği Süreci	14
Şekil 3.1 - KNIME Uygulamasının Yapısı [30]	28
Şekil 3.2 - KNIME Ana Ekranı	29
Şekil 3.3 - KNIME Yeni Workflow Açma	32
Şekil 3.4 - KNIME Excel Reader	33
Şekil 3.5 - KNIME Excel Reader Node.....	33
Şekil 3.6 - KNIME Excel Reader Node Configure	34
Şekil 3.7 - KNIME Node Sarı ve Yeşil Statüler	35
Şekil 3.8 - KNIME Metanode.....	35
Şekil 3.9 - KNIME Metanode İçerik Node'ları	35
Şekil 3.10 - KNIME String Replace Node Configure	36
Şekil 3.11 - KNIME GroupBy Node	37
Şekil 3.12 - KNIME GroupBy Node Configure 1	38
Şekil 3.13 - KNIME GroupBy Node Configure 2	39
Şekil 3.14 - KNIME GroupBy output view	39
Şekil 3.15 - KNIME Cell Splitter Node.....	40
Şekil 3.16 - KNIME Cell Splitter Node Configure	40
Şekil 3.17 - KNIME Cell Splitter Output View.....	41
Şekil 3.18 - KNIME Item Set Finder Node ve Association Rule Learner Node	41
Şekil 3.19 - KNIME Item Set Finder Node Configure	42
Şekil 3.20 - KNIME Association Rule Learner Output View	48
Şekil 3.21 - KNIME Association Rule Learner Node Configure	48

ÖZGEÇMİŞ

Yavuz Demirok

Bilgisayar Mühendisliği Anabilim Dalı

Eğitim

Y.Ls.	2018	Maltepe Üniversitesi, Fen Bilimler Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı
Ls.	2008	Anadolu Üniversitesi, İşletme Fakültesi, İşletme Bölümü
Ö.Ls.	2003	Adıyaman Üniversitesi, Bilgisayar Programcılığı Programı
Lise	2000	İbrahim Turhan Lisesi Yabancı Dil Ağırlıklı

İş/İstihdam

<i>Yıl</i>	<i>Görev</i>
2018 -	Turkcell Teknoloji – İş Analisti
2012 - 18	Ericsson Araştırma ve Geliştirme A.Ş. – İş Analisti
2003 - 12	Turkcell Global Bilgi A.Ş. – BT Self Servis Kanalları IVR Yönetimi

Kişisel Bilgiler

Doğum yeri ve yılı	: İstanbul, 1982	Cinsiyet: Erkek
Yabancı diller	: İngilizce	
GSM / e-posta	: 0 533 210 95 23 / demirok19@gmail.com	

BÖLÜM 1. GİRİŞ

1.1. Problem

Gerek kamu sektörleri, gerekse özel sektörler olsun her gün büyük boyutlarda veriler üretip depolamaktadırlar. Depolanan bu verilerin hacmi büyüdükçe, yönetmek ve bunlardan farklı analiz ve sonuçlar üretmek zorlaşmaktadır. Klasik yöntemler bu analizleri yapmaya zaman zaman yetersiz kalmaktadır. Bu durumu yönetmekte, veri madenciliği teknikleri önemli bir araçtır. Bu çalışmada e-ticaret sitesindeki satış verileri kullanılarak ürünler arasındaki birliktelik kuralları incelenmiştir. Mevcut yapıda belki de hiç farkında olunmayan ama en fazla satılan ürün kümeleri belirlenerek müşterilere bu ürün kümeleriyle ilgili yeni kampanyalar ve promosyonlar oluşturulması hedeflenmiştir. Böylelikle bu çalışmaya konu olan e-ticaret firmasının satış planlaması ve stratejileri üzerinde de değişiklikler söz konusu olacaktır.

1.2. Amaç

Günümüzde birçok sektör ve kuruluşlar büyük veriler elde etmekte ve her geçen gün artan veri trafiğini gelişen teknoloji ile büyük boyutlu depolama alanlarında saklayabilir duruma gelmiştir. Dolayısıyla büyük boyutlardaki bu verilerin incelenmesi ve bu verilerin analiz edilerek fayda sağlayacak sonuçlara ulaşılabilmesi için çeşitli bilgisayar programlarına ve aracı uygulamalara ihtiyaç duyulmaktadır. Veri madenciliğinin günümüzdeki önemi ve uygulama alanları her geçen gün artış göstermektedir. Veri madenciliğinin uygulama alanlarından biri de birliktelik kuralı ile birlikte alınan ürünlerin analizi yani market sepet analizidir. Market sepet analizi ile, müşterilerin birlikte aldığı ürünlerin satış bilgileri kullanılarak, aldığı ürünlerin birbirleri arasındaki ilişki ve kural seti bulunur. Bu birlikteliğin bulunması, sektörün pazardaki karını arttırmakta önemli bir araç olarak kullanılabilir. Aynı zamanda birliktelik kuralı, satılan ürünler içinde birlikte satılabilen veya birlikte satılma potansiyeli olan ürünler arasındaki ilişki çıkartılarak ileride olabilecek satışların tahminlemesinde de kullanılabilir.

Birliktelik kurallarının bulunması için birçok algoritma geliştirilmiş ve denenmiştir. Bu algoritmalar kendi içinde çalışma yöntemleri açısından farklılıklar göstermektedir. Ancak bu algoritmaların hepsi de hemen hemen aynı amaca hizmet eder ve bir sepet analizi yapıldığında bir birliktelik sonucu üretir.

Bu algoritmalarından Apriori Algoritması kullanılarak bir birliktelik analizi çıkartıp sonuçları ve detayları incelenmiştir. Buradaki amaç bu birlikteliğin sektörde kullanılması ve çıkartılan analiz sonuçları neticesinde doğru müşteriye doğru ürün setini sunarak doğru hizmeti vermek ve satış trendini arttırmaktır.

1.3. Önem

Satış verilerinin doğru analiz edilerek firma gelirini artırmak her sektör için hayati değer taşıyan bir unsurdur. Sadece karlılık değil aynı zamanda müşteri memnuniyeti de hayati önem taşımaktadır. Çünkü günümüzde müşteri memnuniyeti ve bağlılığı en az satış kadar önem taşıyan bir unsurdur. Bağlı müşteri veya memnun müşteri ileride de satış yapılabilecek potansiyel müşteridir. Satış verilerinin analizi sektöre çok kıymetli bilgiler sunar. Örneğin bu veriler ile bir firma en çok sattığı veya en az sattığı ürünleri mevcut yöntemlerle kolaylıkla bulabilir. Klasik yöntemlerin yetersiz kaldığı veya tam olarak amaca hizmet edemediği durumlar da yaşanmaktadır. Her geçen gün gelişen teknolojiler ile sektöre en basit analizlerin dışında aslında görünmeyen kuralları görme imkanı sağlamaktadır. Ancak bu doğru yöntemleri kullanarak mümkündür.

Bu çalışmada müşterilerin birlikte almayı tercih ettiği ürünlerin analizi Apriori Algoritması kullanılarak çıkartılmıştır. Bu analiz sonucunda çıkartılan birliktelikler ile sektörün satış stratejilerine etki edebilecek sonuçları incelenmiştir. Bu çalışmanın önemi birlikte satılan ürünleri ön plana çıkartarak bu ürün kümeleri ile beraber, en az satılan ürünlerin de satışını artırabilecek aksiyonlara ulaşmaktır. Bu sayede yapılabilecek değerlendirmeler ile müşteri memnuniyeti ve karlılığın artması hedeflenmektedir.

1.4. Sınırlıklar

Bu alıřmada kullanılan veri E-Ticaret sitesinin 2,5 yıllık satıřlarını kapsamaktadır. Veri madenciliğinde veri ne kadar büyürse elde edilen analiz ıktılarının güvenilirlięi ve doęruluęu o derecede artar. Bu alıřmada kullanılan veri boyutu yaklaşık 33.000 satırdan oluřmaktadır. Veri madencilięi olarak düşünöldüğünde 33.000 satırlık veri aslında pek de büyük hacimli bir veri deęildir. Dolayısıyla verilerin az olması bu alıřmanın sonuçlarını etkileyen önemli bir etken olmuřtur.



BÖLÜM 2. BİRLİKTELİK KURALI ANALİZİ

Bu bölümde veri madenciliği metodlarından birliktelik kuralı incelenerek veri madenciliği hakkında detaylı bilgilere yer verilmiştir. Veri madenciliğinin birliktelik kuralı algoritmalarından Apriori Algoritmasının detayları, çalışma mantığı ve örneklemeleri ile benzer çalışmalar ve akademik yayınlar incelenerek bu çalışmada yer verilmiştir.

Çalışmaya başlarken teorik bilgilerin ve kavramların incelenmesi ile matematiksel formüllerinin detaylandırılması oldukça önemlidir. Bu bağlamda Araştırma Modeli içerisinde bu detaylardan bahsedilmiştir.

2.1. Araştırma Modeli

2.1.1. Veri Nedir, Bilgi Nedir?

Veri, işlenmemiş veya işlenmeye hazır kaynak olarak tanımlanabilir. Bilgisayar ortamında ise nümerik olmayan veya nümerik değerler ile her türlü resim ve değer niteliği taşıyan karakterler veri olarak ifade edilir. Bilgi ise bu verileri kullanarak anlamlı sonuçların çıkarılması ile elde edilir. Örneğin string olarak ifade edilen bir metin kullanılarak bir bilgi elde edilebilir, (string ifade – B bu ifadeyi B Blok haline çevrildiğinde bu Bilgi olarak ifade edilebilir).

Veri ve bilgi kavramlarına şöyle bir örnek de verilebilir; Şişenin ham maddesi camdır. Cam veri, şişe ise bu örnekte bilgidir.

2.1.2. Veri ve Bilgi arasındaki temel farklılıklar:

- Veri tek başına her zaman birşey ifade etmez. Farklı veriler bir araya gelerek bir bilgi oluşturabilir.
- Verinin herhangi bir bağımlılığı yoktur ancak bilgi veriye muhtaçtır, veri olmadan bilgiye ulaşamaz.
- Veriler tek başına iken tam olarak kesin sonuçlar içermez, ancak bilgi çok daha nettir. Çünkü bilgi veriler ışığında bir sonuca varmaktadır [1].

2.1.3. Veritabanı Kavramı

Veritabanı, herhangi bir sektördeki verilerin düzenli olarak tutulabildiği gerektiğinde sorgulanabildiği ve çeşitli çıktıların üretilebildiği bir sistemdir. Sadece sektör olarak düşünmenin yanısıra bireysel olarak da kullanılabilir. Her türlü kişisel bilgi, ürün stok ve adetleri, müşteri özlük bilgileri, ürün fiyatları ve birçok detayın tutulabildiği bir çeşit veri depolama merkezi olarak da tabir edilebilir. İşletmelerin günlük yaşam döngüsünde sadece bu verileri deploması ve saklaması, faaliyetlerini yönetebilmesi için yeterli değildir. Depoladığı verilerin hacmi büyüdükçe yönetmek de zorlaşmaktadır. Her geçen gün büyüyen veritabanını yönetmek veritabanı tasarımının önemini ortaya çıkarmaktadır. Tasarımın en önemli prensiplerinden birisi veritabanında birbiriyle ilişkili verilerin gruplandırılması ve bu şekilde bir tablo yapısı oluşturmaktır. İlişkili verilerin tutulduğu tablolar arasında bir ilişkilendirme yapılması gerekir. Veritabanı tasarımında bu ilişkilendirmeler düzenlenirken, tablolardan veri sorgulama hızının iyi olması, herhangi bir güncelleme veya yeni bir veri ekleme işlemi yapıldığında veritabanının tamamen taranmadan yapılabiliyor olması, tekrarsız ve kirli verilerin barındırılmıyor olması, kısıtlı imkanlara sahip olmaması önemli etkenlerdir, bu tip veritabanları İlişkisel Veritabanı olarak tanımlanmaktadır [2].

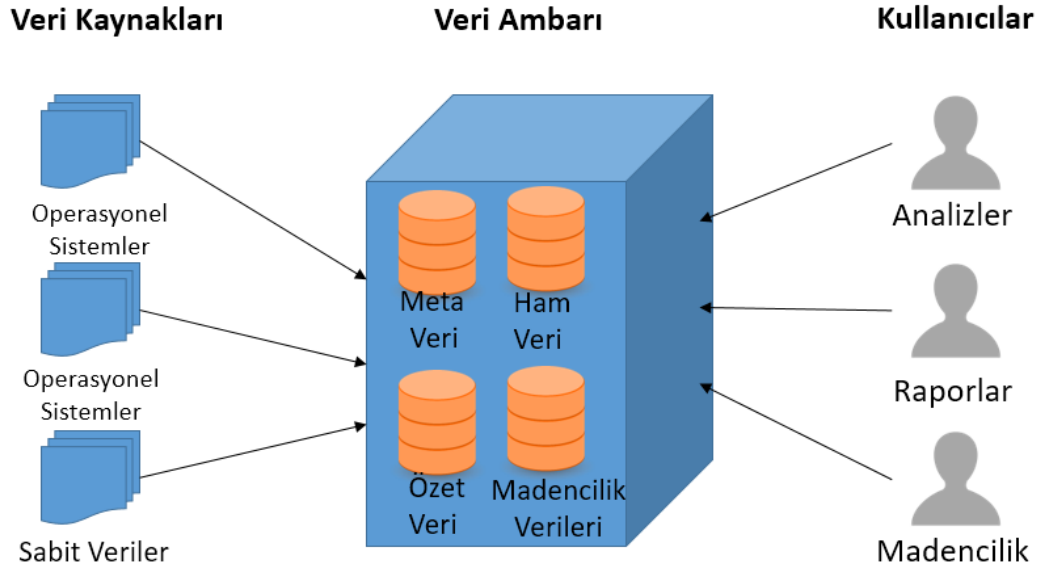
Veritabanındaki tablolarda saklanan veriler, ihtiyaç duyulduğunda birtakım sorgulama yöntemleri ile veritabanından alınabilir. Veri madenciliği yöntemleri ve teknikleri ile veritabanında tutulan verilerin sadece okunması veya sorgulanması değil, bu veriler kullanılarak çeşitli sonuçlar ve modeller de üretilebilmektedir.

2.1.4. Veri Ambarı Nedir?

Veri ambarı kavramı ilk olarak 1990 yılında Bill Inmon tarafından ortaya atılmıştır. Inmon'un tanımına göre veri ambarı konuya odaklı, kalıcı, zamana bağlı ve entegre veriler kümesidir. Bu veriler, organizasyon içerisinde bilinçli kararlar alınmasında önemli bir rol oynar. Veri ambarları çok boyutlu olarak genelleştirilmiş ve birleştirilmiş verileri sağlar. Bu sayede veri ambarları Çevrimiçi Analitik İşleme (OLAP) araçlarını sağlar. Bu araçlar, çok boyutlu alanlarda verilerin daha etkin analiz edilme imkanı sunar [3].

Bilgi her kuruluş için oldukça değerli bir olgudur, dolayısıyla bilgiyi depolamak ve kolay bir şekilde ulaşılabilir olmalıdır. Veri her zaman bilgi demek değildir. İnternette

herhangi bir şey aratıldığında bile binlerce veri gelir ancak bunlardan biri veya birkaç tanesi gerçekten işe yarayan bilgidir [3].



Şekil 2.1 - Veri Ambarı Konsepti

2.1.5. Veri Ambarı Bileşenleri

Veri ambarı kavramında 4 ana bileşen vardır. Bu bileşenler aşağıdaki gibidir.

- Çok Boyutlu Veritabanı
- ETL
- OLAP
- Meta Veri

Çok Boyutlu Veritabanı: Çok boyutlu veritabanı teknolojisi, büyük boyut ve ölçülerdeki verilerin analiz edilmesinde önemli bir rol oynar. Bu teknolojiye veriler çok boyutlu küp diye tabir edilen kalıplar olarak gösterilmektedir. Ayrıca bu modellerde veriler ilgili sayısal ölçüt veya bir işi ayırt edici metinsel boyutlarla kategorize edilir. Bu teknoloji, mevcut sistemler tarafından yeteri seviyede analiz edilemeyen veri türlerine uygulanabilmektedir [4].

ETL (Extract – Transform – Load): ETL (Çıkart – Dönüştür – Yükle) araçları, birçok farklı kaynaktan elde edilen verilerin alınmasından, temizlenmesine,

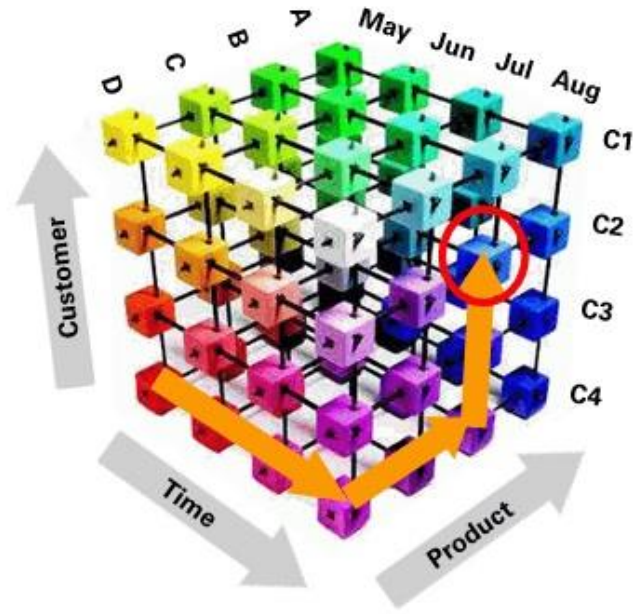
biçimlendirilmesinden, bütünleştirilmesine ve veri ambarına aktarılmasına kadar geçen süreçten sorumlu olan uygulamalar bütünüdür. Genel anlamda ETL süreçleri karmaşık yapılardır, bu süreçler üç başlık altında incelenebilir; Kaynak alan, Hedef Alan ve Eşleştirme alanı.

Kaynak alan; standart olarak varlık ilişki diyagramı modellerine sahipken, Hedef alan; standart olarak yıldız şeması modellerine sahiptir. Fakat Eşleştirme alanı için belirli bir standart model bulunmamaktadır [5].

OLAP (On-Line Analytical Processing): OLAP (Çevrimiçi Analitik İşleme), Esas olarak günlük işlemlerle, iş verilerini birlikte analiz etmek için kullanılır. OLAP, karar vermeyi desteklemek için güçlü bir araçtır. OLAP sistemi, kullanıcının farklı bakış açılarından verileri kolayca çıkarmasına ve görüntülemesine izin verir. Ayrıca, kullanıcıların büyük miktarda veri üzerinde hızlı ve etkili analiz yapabilmelerini sağlar. OLAP sistemleri, verileri çok boyutlu biçimde saklar. OLAP özet verileri verimli bir şekilde sağlayabilir ve kullanıcıların bu özet verilere daha hızlı ve daha kolay erişmesini sağlar.

OLAP küpü, verileri çok boyutlu biçimde depolamak için kullanılan bir yöntemdir. Bu yöntem verilerin daha hızlı analiz edilmesine olanak sağlar. OLAP küpü, verileri farklı veya çoklu açılardan çalıştırma ve analiz etme yeteneğine sahiptir. Küp, boyutlar olarak nitelendirilen ve ölçüt adı verilen sayısal verileri içerir. Çok boyutlu bir küp veya OLAP küpü, verileri çeşitli kaynaklardan birleştirir ve depolar. Küp içerisine veri düzenlemesi gereken durumlarda, ilişkisel veritabanının olası sınırlandırmalarının da üstesinden gelir [6].

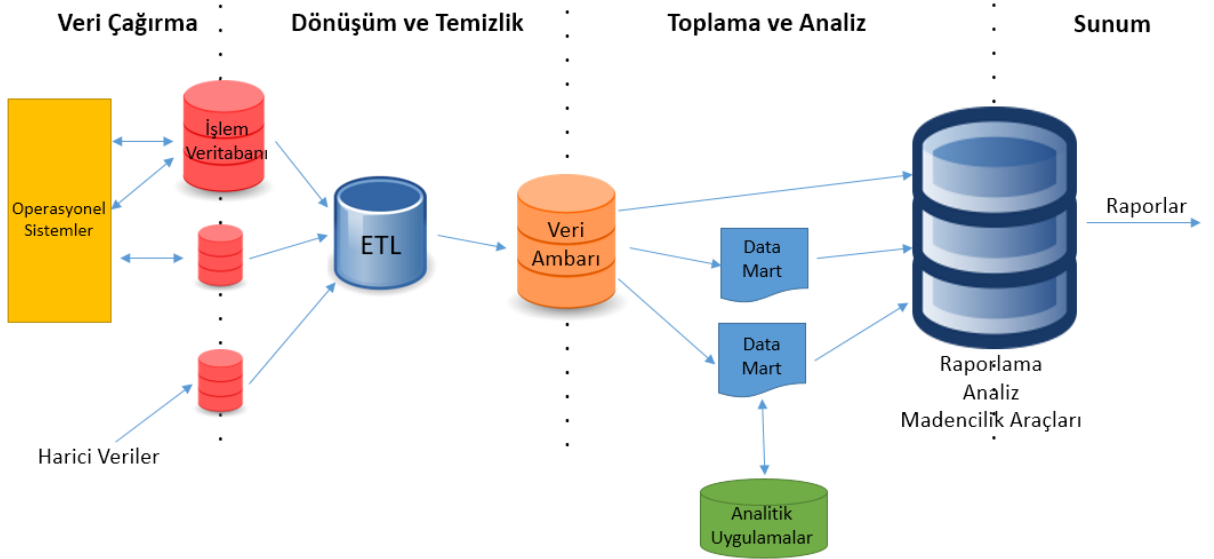
- **Roll-Up:** Bir çeşit veri toplama işlemidir. Verileri özetleyerek daha küçük boyutlara getirir.
- **Drill-Down:** Roll Up ile yapılan işlemin tersini yapar. Toplanan verileri ayrıştırarak detaylandırır.
- **Slice and Dice:** Küpün bir bölümünü alır ve aynı şemaya ait bir alt küp oluşturur. Küp içerisinden bir bölüm seçerek bu bölüm üzerinde işlem yapılmasını sağlar [7].



Şekil 1.2 - OLAP Küp [8]

Drill-down veya roll-up işlemleri her boyutta gerçekleştirilebilir. Örnek; firma adı, ürün adı ve tarihten oluşan bir küp'te, satılan ürünlerin sayısı tarihe göre incelenebilir, firmanın şubeleri bazında ve tarih bazında satılan ürünlerin satış grafiği bu küp aracılığıyla çıkartılabilir [9].

Meta Veri: Meta Veri, Yunanca ve Latince kökenli bir kelime olup bir bilgi kaynağını tanımlar. Bir meta veri kaydı, bir kaynağı belirlemek için gereken bir takım özellik veya öğelerden oluşmaktadır. Örneğin, kitaplıklarda ortak olan bir meta veri sistemi, bir kitabı veya kütüphane öğelerini tanımlayan öğelerle birlikte bir takım meta veri kaydı içerir; Kitabın yazarı, kitabın içeriği ve konusu, kitabın yayınlanma tarihi, önsöz ve açıklamaları, konusu ve kitabın kütüphanenin hangi rafında olduğu bilgisi, vb [10].



Şekil 2.3 - Veri Ambarı Bileşenleri

2.2. Veri Madenciliği

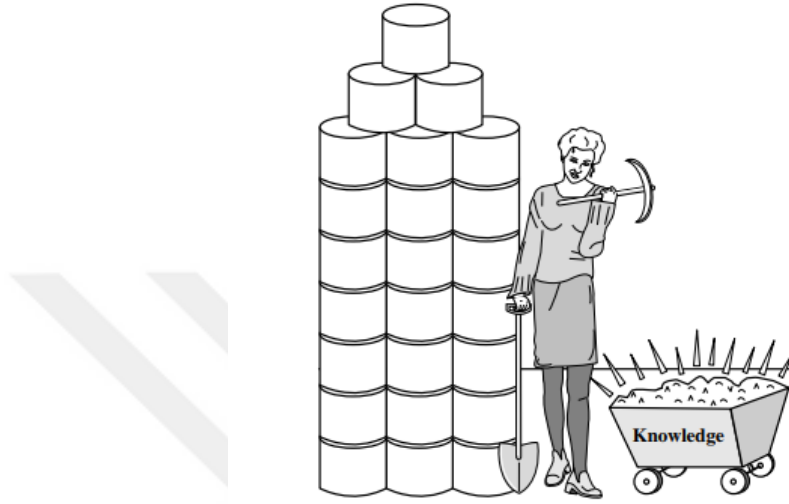
2.2.1. Veri Madenciliği Ne Demektir?

İçinde bulunduğumuz bu dönemde her gün birçok sektörde veriler elde ediliyor ve depolanıyor. Bu verileri analiz etmek ve bu verilerden bir sonuç çıkartmak her sektör için oldukça önemlidir. Tıp, mühendislik alanları, toplum ihtiyaçları, bilim, günlük hayatın birçok alanında küçük veya büyük işletmelerin bilgisayar ortamlarında depoladığı veriler terabayt, petabayt seviyelerinde tabir edilmekte. Bu veri ambarının her geçen gün artmasının bir sebebi de bilgisayarlaşmanın bir sonucu olmasının yanı sıra, veri saklama yöntemlerinin her geçen gün gelişen teknoloji ile daha da kolaylaşmasından kaynaklıdır.

Veri madenciliği, kısaca veriden bilgiye ulaşmaktır denilebilir. Başka bir tanıma göre veri madenciliği, elde edilen verilerle (gürültüsüz ve temiz veriler ile), bir takım yöntem ve algoritmalar kullanılarak fayda sağlayacak sonuçlar üretmek ve bu veriler arasındaki ilişkiyi bulmayı amaçlayan bir analiz yöntemidir [11]. Mevcut verileri kullanarak yeni bilgilere ulaşmak, bilinmeyeni keşfetmek veri madenciliğinin asıl ve önemli amaçlarından birisidir [12].

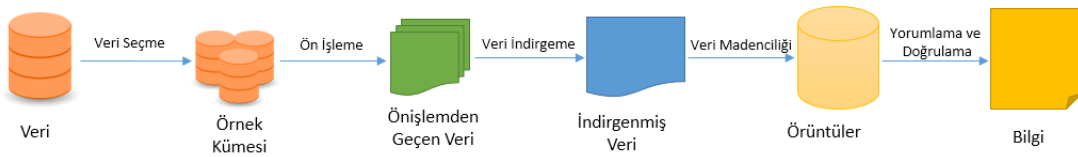
Amaca hizmet etmediği sürece veri kendi başına değersiz bir olgudur. Veri işlenip bilgiye çevrilmeye başladıkça değer kazanır. Örnek; Büyük kaya kütleleri veya toprak yığınları içerisinde altın madeninin ayrıştırılması ve altın madenine ulaşmak veri

madenciliğinde faydalı verilere ulaşmak ile benzerdir. Kaya kütleleri veya toprak yığınlarını veri olarak ifade edilirse, onu ayrıştırmak ve içerisinde altın madenini çıkartmak da veri madenciliği yöntemleri ve sonuçları olarak ifade edilebilir. Burada önemli olan veri dağları içerisinde altın değeri taşıyan veri sonuçlarına ulaşabilmektir.



Şekil 2.4 - Veri Madenciliği ve Bilgiye Ulaşma [13]

Veri madenciliğinde veriden bilgiye nasıl ulaşılır, verinin hangi aşamalardan geçerek bilgiye dönüştürüldüğü Şekil 2.5'te gösterilmiştir.



Şekil 2.5 - Verilerin Bilgiye Dönüştürülme Süreci

1. **Veri Temizleme:** Kirli olarak tabir edilen gürültülü ve tutarsız verileri temizler. Veri temizleme, veri analizi tekniğinin bir uygulamasıdır. Verilerin gerçek içeriğinin ne olduğunu yapısını ve kalitesini belirlemede kullanılır. Genellikle veri temizliği, temizlenmiş verilerle bir kaydı günceller ancak ciddi temizlik, verilerin ayrıştırılmasını ve yeniden toplanmasını içerir.
2. **Veri Entegrasyonu:** Birden fazla veri kaynağının birleştirilebilmesidir. Tanımlama genel olarak, çoklu bilgi sistemlerinin entegrasyonu, seçilmiş

sistemleri, birleştirilmiş yeni bir bütün oluşturacak şekilde birleştirmeyi ve kullanıcıların tek bir bilgi sistemi ile etkileşime girmesini kolaylaştırmayı amaçlamaktadır. Belirli bir bilgi gereksinimi göz önüne alındığında, farklı tamamlayıcı bilgi sistemlerinden gelen verilerin, ihtiyaca cevap vermek için daha kapsamlı bir temel elde etmek üzere birleştirilmesi gerekir.

3. **Veri Seçimi:** Veri Seçimi, analiz görevine ilişkin verilerin veritabanından alındığı süreçtir. Bazen veri dönüştürme işleminden önce veri dönüşümü ve konsolidasyonu gerçekleştirilir.
4. **Veri Dönüşümü Süreci:** Bu aşamada, veriler, özet veya toplama işlemleri gerçekleştirilerek, madencilik için uygun formlara dönüştürülür veya birleştirilir.
5. **Veri Madenciliği:** Veri kalıplarını çıkarmak için akıllı yöntemlerin uygulandığı aşamadır.
6. **Örüntü Değerlendirme:** Veri modelleri değerlendirilir.
7. **Bilgiye Ulaşma:** Bilginin elde edildiği, sonuca ulaşıldığı aşamadır [13].

2.2.2. Veri Madenciliğinin Uygulama Alanları

Günümüzde kamu kuruluşlarından özel sektörlere, küçük işletmelerden büyük işletmelere, kadar birçok sektörde veri madenciliği kullanılmaktadır. Veri madenciliğinin kullanım alanları aşağıdaki şekilde kategorize edilebilir;

Bilim ve Mühendislik Alanları: Bilimsel araştırmaların yapıldığı ortamlar veya bilgisayar ortamlarının benzer koşullarının sağlanması ve analizi sürecinde büyük miktarlarda çıktılar elde edilmektedir. Bu çıktıların değerlendirilip bilgiye dönüştürülmesi ancak veri madenciliği kullanılarak mümkün olabilir [14]. Elde edilen veriler farklı çalışmalara da ışık tutabilir veya yeni çağrışımlar da yapabilir. Tek başına önemsiz gibi görünen veri kümeleri bu doğrultuda bilim ve teknolojik alanda yeniliklere ışık tutabilir.

Bankacılık ve Finans: Bankacılık ve finans sektörlerinde müşteriler üzerinde araştırmalar yapılarak belirli sonuçlara ulaşılabilir. Müşteri bazında risk analizi, müşterinin kronolojisine bakılarak kredibilitesinin belirlenmesi sağlanabilir. Belirli dönemler incelenerek müşteri güvenilirliği belirlenip ödeme veya ödememe durumlarıyla

ilgili bir tahmin analizi çıkartılabilir [15]. Riskli müşteri kitlesi belirlenerek sektörde risk oluşturabilecek durumların önüne geçilebilir.

Müşteri ilişkileri Yönetimi: Müşterilerin davranış ve hareketleri izlenerek buradan elde edilen verilerden müşterinin ileri zamanlarda yapacağı hareketleri veya davranışları tahmin etmek için kullanılabilir [15]. Örneğin belli bir müşteri kitlesinin en çok tercih ettiği ürün yelpazesi incelenerek bu ürünler için bir kampanya veya promosyon yapılabilir. Bu ürünleri hiç tercih etmemiş müşteri kitlesine de bu ürünler sunulabilir.

Müşteri Fraud (Sahtekarlık) Tespiti: Müşteri verilerinin incelenip bu veriler içerisindeki tutarsızlıklar veya sahte bilgilerin kullanılarak bir işlem yapılması durumu tespit edilebilir. Özellikle sigorta ve bankacılık sektörlerinde yapılan fraud (sahtekarlık) analizleri ile kredi kartı sahtekarlıklarının tespit edilmesinde önemli bir rol oynar [15]. Bankacılık veya sigorta sektörleri dışında herhangi bir resmi kurumda yapılan işlemde verilen kimlik bilgilerinin, başka bir işlemde verilen kimlik bilgileri ile tutarlılığı da bu kapsamda değerlendirilebilir. Örneğin; İki sene önce resmi bir kurumdan yapılan bir işlemde müşterinin verdiği kimlikteki bilgileri ile, yeni yaptığı bir işlemde verdiği kimlikteki bilgilerindeki farklılıklar bu inceleme ile kolaylıkla ortaya çıkartılabilir.

Güvenlik ve Emniyet (İstihbarat): Kişilerin veya birtakım grupların suç ve olay potansiyeli incelenerek, terör veya herhangi bir suç faaliyetinde bulunabilecek kişilerin arasındaki bağı ve bağlantıyı analiz ederek olası bir faaliyeti tahminlemek söz konusudur [14, 15]. Belirlenmiş bölgeler veya semtler ile belirli gruplar veya şahıslar bu kapsamda incelenebilir. Herhangi bir suçu tetikleyecek faaliyetler için oluşabilecek koşullarında bu kapsamda incelenmesi önemlidir. Her türlü suç faaliyeti koşullar ve bir takım etkileşimlerin sağlanması ile oluşur. Veri madenciliği kişi ve grupların yanısıra koşulları oluşturacak parametreleri de inceleyerek bir tahminde bulunabilir.

Eğitim Alanı: Öğrenci profilleri analiz edilerek başarı durumlarının yanısıra başarısızlıklar ve nedenleri, başarının daha üst seviyelerde gelmesi için nelere daha çok önem vermeli veya nelerin eğitim sistemine eklenip/çıkartılması gerektiğini, üniversite puanları ile okuldaki başarı arasında bir bağlantı olup olmadığını inceleyerek eğitim kalitesi artırılabilir [14].

Öğrencilerin geçmiş dönemleri incelenerek başarını tahminleme ile öğrencinin başka okula geçme ihtimalinin tahmini, aynı özelliklere sahip öğrencilerin koşullarının iyileştirilmesi veya daha da geliştirilmesi adına yönlendirici bilgi vermesi de veri madenciliği ile mümkün olabilir [16].

Sağlık Sektörü: Sağlık sektörü veri madenciliğine aslında en çok ihtiyacı olan ve de en çok kullanan sektördür denilebilir. Hastaların test ve tahlil veri setleri kullanılarak kanser teşhisinin önceden belirlenmesi veya bir kalp krizi ihtimalinin önceden tahminlenmesi mümkündür [17]. Bu kapsamda bakıldığında veri madenciliği aslında her sektör için önem arz etmektedir ancak Tıp ve sağlık sektöründe birçok hastalığın önceden tespiti ile tedavi sürecinin hastalık başlamadan tahminlenmesi veri madenciliğinin kullanımını oldukça önemli kılmaktadır.

2.2.3. Veri Madenciliği Süreçleri

Veri madenciliği süreçleri Şekil 2.6'daki gibi incelenebilir. Bu şekile göre aslında veri madenciliği bir süreçten ibarettir. Doğru verilerle veya doğru algoritmalar ile çalışmak başlı başına yeterli olmamaktadır. Bu süreçler aşağıdaki gibidir;

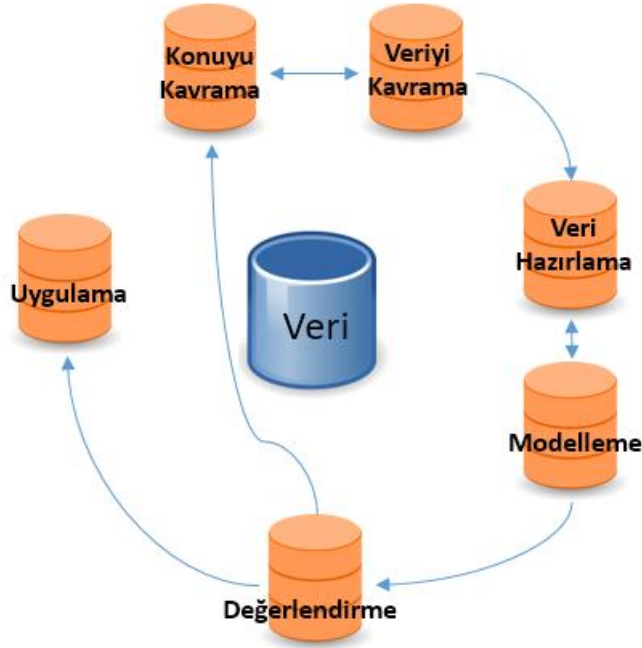
Problem tanımı: Her şeyden önce üzerinde çalışılacak olan veya hangi veriyi kullanarak hangi veriye ulaşılmalı? konusunun belirlenmesi gerekir. Problem diye nitelendirilen durum aslında ne yapılmak istendiğinin bir karşılığıdır.

Verilerin Hazırlanması: Kullanılacak olan verinin seçilmesi, temizlenmesi ve veya başka verilerle birleştirilerek amaca uygun hale getirilmesi, verinin belli formatlara çevrilmesi süreci bu adımda yapılır.

Modelin Kurulması: Bu aşamada kullanılacak olan veri üzerinde hangi veri madenciliği tekniğinin kullanılacağı belirlendiği veya hangi algoritma ile çalışılacağı belirlenir.

Modelin Değerlendirilmesi: Model ile ihtiyaç duyulan bilgiye ulaşılabildi mi? Ya da model ile hangi bulgulara ulaşılmış bu aşamada incelenebilir.

Uygulama (Modelin İzlenmesi): Sürecin modelleme sonrasındaki en son adımdır. Burada sonuçlar değerlendirilir ve istenilen bilgiye ulaşılan adımdır [18].



Şekil 2.6 - Veri Madenciliği Süreci

2.2.4. Veri Madenciliği İşlevleri ve Yöntemleri

Veri madenciliği işlevleri, veri madenciliği görevlerinde bulunacak kalıp türlerini belirtmek için kullanılır. Genel olarak, veri madenciliği görevleri iki kategoriye ayrılabilir: tanımlayıcı madencilik ve tahminleyici madencilik. Tanımlayıcı madenciliğin görevi, veritabanındaki verilerin genel özelliklerini karakterize eder. Tahmine dayalı madenciliğin görevi ise, öngörülerde bulunmak için mevcut veriler üzerinden çıkarım yapmaktır.

Bazı durumlarda, kullanıcıların kendi verilerindeki hangi tür kalıpların ilginç olabileceği konusunda bir fikri olmayabilir bu nedenle, paralel olarak çeşitli farklı desen türlerini aramak isteyebilir. Bu sebeple, farklı kullanıcı beklentilerini veya uygulamalarını karşılamak için çok sayıda modele sahip olabilen bir veri madenciliği sistemine sahip olmak önemlidir. Ayrıca, veri madenciliği sistemleri, çeşitli sayıdaki (yani, farklı soyutlama seviyelerinde) kalıpları bulabilmelidir. Etkileşimli ve keşif amaçlı madenciliği teşvik etmek için, kullanıcıların fare tıklaması gibi çıktı kalıplarıyla kolayca oynayabilmeleri gerekir. Basit fare tıklamaları ile belirtilebilecek işlemler arasında, bir boyut (veya bir özellik) eklenmesi veya bırakılması yer alır. Satırlar ve sütunlar (dönme veya eksen dönüşü), boyut gösterimlerini değiştirme (Örnek; 3-D küpünden 2-D çapraz

sekanslar veya çapraz çizgiler dizisine) veya boyutlar boyunca OLAP roll-up veya drill-down işlemlerini kullanma. Bu tür işlemler, veri desenlerinin farklı görüş açılarından ve çoklu soyutlama seviyelerinde ifade edilmesine izin verir.

Veri madenciliği sistemleri, kullanıcıların aramayı ilginç desenlere yönlendirecek veya ipuçlarını belirtmelerine de izin vermelidir. Veritabanındaki tüm veriler için bazı kalıplar tutmayabileceğinden, genellikle her bir tespit edilen modelle bir kesinlik veya güvenilirlik ölçüsü ilişkilendirilir [13].

2.2.5. Sınıflama ve Regresyon Modelleri

Sınıflandırma, sınıf etiketinin bilinmeyen nesnelere tahmin etmek için modeli kullanabilmek amacıyla veri sınıflarını veya kavramlarını tanımlayan ve ayıran bir dizi modelin (veya fonksiyonun) işlenmesidir. Türetilen model, bir dizi eğitim verisinin analizine dayanmaktadır (yani, sınıf etiketi bilinen veri nesnelere).

Türetilen model, sınıflandırma (IF-THEN) kuralları, karar ağaçları, matematiksel formüller veya sinir ağları gibi çeşitli biçimlerde temsil edilebilir. Bir karar ağacı, her bir düğümün bir öznitelik değeri üzerinde bir testi, her bir dalın testin bir sonucunu temsil ettiği ve ağaç yapraklarının sınıfları veya sınıf dağılımlarını temsil ettiği akış şemasına benzer bir ağaç yapısıdır. Karar ağaçları, sınıflandırma kurallarına kolaylıkla dönüştürülebilir. Bir sinir ağı, farklı sınıfların nesnelere ayırmak için eğitilebilen doğrusal eşik birimleri topluluğudur.

Sınıflandırma, veri nesnelere sınıf etiketini tahmin etmek için kullanılabilir. Bununla birlikte, birçok uygulamada, sınıf etiketlerinden ziyade, bazı eksik veya mevcut olmayan veri değerlerinin de tahmin edilmesi istenilebilir. Bu genellikle tahmin edilen değerler sayısal veriler olduğunda ve genellikle tahmin olarak adlandırıldığı zaman geçerlidir. Her ne kadar tahmin, hem veri değeri tahmini hem de sınıf etiketi öngörüsüne atıfta bulunsun da, genellikle veri değeri tahmininin sonucudur ve bu nedenle sınıflandırmadan farklıdır. Tahmin, mevcut verilere dayanarak dağıtım eğilimlerinin tanımlanmasını da kapsar. Sınıflandırma ve tahmin işlemine, sınıflandırma veya tahmin sürecine katkıda bulunmayan özniteliklerin belirlenmesini amaçlayan uygunluk analizinden önce gelmek gerekebilir. Bu nitelikler daha sonra hariç tutulabilir [13]. Sınıflandırma ve Regresyon modellerinde kullanılan en önemli teknikleri aşağıdaki gibi sıralanabilir;

- Karar Ağaçları
- Yapay Sinir Ağları,
- Genetik Algoritmalar,
- K-En Yakın Komşu,
- Bellek Temelli Nedenleme,
- Naive-Bayes,
- Lojistik Regresyon [18].

2.2.6. Kümeleme Analizi Modeli

Kümeleme, bilinen bir sınıf etiketine başvurmadan veri nesnelarini analiz eder. Genel olarak, sınıf etiketleri, başlangıçta bilinmedikleri için eğitim verilerinde mevcut değildir. Kümeleme, bu tür etiketler üretmek için kullanılabilir. Nesnelar, sınıf içi benzerliği en üst düzeye çıkarmak ve sınıflar arası benzerliği en aza indirgeme ilkesine dayalı olarak kümelanmiş veya gruplandırılmıştır. Yani, nesneların kümeleri, bir kümelanme içindeki nesneların, birbirine göre yüksek benzerliğe sahip olmaları, ancak diğer kümelerdeki nesnelardan çok farklı olmaları için oluşturulmuştur. Oluşturulan her küme, kuralların türetilebileceği bir nesne sınıfı olarak görülebilir [13].

2.2.7. Birliktelik Analizi Modeli

Birliktelik analizi, belirli bir veri kümesinde sık sık birlikte ortaya çıkan özellik-değer koşullarını gösteren ilişkilendirme kurallarının keşfidir. Birliktelik analizi, market sepeti veya işlem verileri analizi için yaygın olarak kullanılmaktadır [13].

Birliktelik kuralları (ilişkilendirme analizi) ilişkilendirme kurallarını bulmak için kullanılır. Bu kurallar, öznitelik veya değerlerin sık sık birlikte görünmesi ve verilerin arasındaki ilişkilendirme kuralları ile öğeler arasındaki ilişkiyi keşfetmede kullanılır.

Birliktelik kuralının bulunması, destek ve güven düzeyinin kural eşığı değerini aştığı ilişki kuralını bulmaktır. İlişkilendirme kurallar süreç olarak 2 aşamaya ayrılır:

- Tüm sık kullanılan takım kümelerini bulmak: tüm desteklerin kullanıcı tarafından en küçük destek eşik değeri tasarlandığından daha düşük olmayan öğe kümelerini bulmak;

- Sık kullanılan öge güçlü bağlantı kuralı oluşturur: Bu kurallar en küçük desteği ve en küçük güveni tatmin etmelidir. Ancak tüm sık kullanılan öge kümelerini bulmak tüm bağlantı kuralı çekirdeğidir [19].

2.3. Birliktelik Kuralları ve İlişki Analizi

2.3.1. Market Sepet Analizi

Market Sepeti Analizi, müşterilerin alışveriş sepetlerinde bulunan ürünler arasındaki bağlantıyı ve markette alışveriş yaparken müşteri davranışlarının tespiti üzerine gerçekleştirilen bir analiz metodudur [20]. Spesifik olarak Market Sepet Analizi, müşteriler tarafından en çok satın alınan öğeleri eşzamanlı olarak tanımlamayı amaçlamaktadır. Burada, ürün süpermarkette birkaç çeşit ürün olarak tasvir edilmektedir. Başka bir deyimle bir arada en sıklıkla tercih edilen ürünlerin tespit edilmesini sağlar.

Market sepet analizini kullanarak, müşterilerin eş zamanlı olarak satın aldığı ve satın alma potansiyeline sahip ürünler hakkında bilgi edinilebilir. Müşterilerin aynı anda hangi ürünleri satın aldıklarına karar vermek için kullanılan market sepet analizinin amacı; alışveriş ürünlerinin, alışveriş sepetlerine veya alışveriş listesine yerleştirilmesinde müşterilerin davranışlarını gözlemlemek ve bu veriler ile sonuçlar üretmektir. Bir müşterinin alışveriş sepeti modelini tanımlamak, bir şirketin iş stratejisi ve gereksinimlerinin belirlenmesinde yardımcı olabilecek bir unsurdur [20].

2.3.2. Birliktelik Kuralları

Birliktelik kuralı “ne” ile “ne” ifadesiyle ilgili bir kavramdır. Bu konu, bir markette müşteriler tarafından gerçekleştirilen işlemler hakkında bir açıklama şeklinde olabilir. Bu açıklamayla birliktelik kuralı, satın alınan bir ürünün, hangi ürünle birlikte alındığı ve müşteri alışkanlıklarının belirlenmesi olarak ifade edilebilir, dolayısıyla birliktelik kuralı çoğunlukla market sepeti analizi olarak adlandırılır [21, 22].

Birliktelik kuralının iki önemli unsuru vardır; bunlar Destek ve Güven’dir. Destek (destekleyici değer), veritabanındaki ürün öğelerinin kombinasyonlarının yüzdesidir. Güven (kesinlik değeri), ilişki kuralları içindeki ürün ilişkilerin gücünü belirleyen bir değerdir.

$$\text{Destek} = \frac{n(A \cup B)}{N} \quad (1)$$

$$\text{Güven} = \frac{n(A \cup B)}{n(A)} \quad (2)$$

1 nolu formüle göre, A ve B ürünlerinin birlikte satıldığı işlem sayısının, toplam satış işlemleri sayısına (N) bölünerek elde edildiğini gösterir.

2 nolu formüle göre, A ve B ürünlerinin birlikte satıldığı işlem sayısının, A ürünün bulunduğu tüm işlem sayılarına bölünerek elde edildiğini gösterir. Güven değerinin 0 çıkması demek, A ürününün bulunduğu işlemlerin hiçbirinde B ürününün bulunmadığı anlamına gelir [23].

Tablo 2.1 - Market Basket Ürün Tablosu

Ürün ID	Ürünler Listesi
1	Domates,Patates,Süt
2	Elma,Süt,Cikolata
3	Domates,Patates,Süt,Su
4	Patates,Zeytin Yağı,Cikolata
5	Domates,Patates,Yogurt

Tablo 2.1'e göre 5 müşterinin sepet analizi yapıldığında Domates ve Patates ürünlerinin Süt ile arasındaki ilişki, Güven Destek ilişkisi olarak aşağıdaki gibi gösterilebilir.

$$\text{Güven} = (A \rightarrow B) = \frac{A \text{ ve } B \text{ toplam işlemini içeren işlem sayısı}}{A \text{ İçeren Toplam İşlem Sayısı}} \quad (3)$$

$$\text{Güven} = (A \rightarrow B) = \frac{\text{Domates,Patates,Süt}}{\text{Domates,Patates}}$$

$$\text{Güven} = (A \rightarrow B) = \frac{2}{3} = \mathbf{0,66}$$

$$\text{Destek} (A) = \frac{A \text{ içeren Toplam İşlem Sayısı}}{\text{Toplam İşlem Sayısı}} \quad (4)$$

$$\text{Destek} (A) = \frac{2}{5} = \mathbf{0,4}$$

2.3.3. Birliktelik Kuralı Algoritmaları

Büyük verileri tanımlamak için kullanılan algoritmalar, sıralı veya paralel olarak 2 sınıfta incelenebilir. Çoğu durumda, verilerin sözlüksel sırada (öge adına göre) tanımlandığı ve depolandığı varsayılır. Bu sıralama, verilerin üretilip sayılabildiği mantıklı bir şekilde sağlanır. Bu sıralı algoritmalarda normal yaklaşımdır. Öte yandan, paralel ve dağıtılmış algoritmalar büyük eşya kümeleri bulma görevini nasıl yapabileceğine odaklanır. Sıralı ve Paralel Algoritma Kümeleri aşağıdaki gibidir [24].

a) Sıralı Algoritmalar;

- AIS Algoritması
- SETM Algoritması
- Apriori Algoritması
- Apriori-TID
- Apriori-Hybrid
- Offline Candidate Determination (OCD) (Çevrimdışı Aday Belirleme)
- Partitioning (Bölümleme)
- Sampling (Örnekleme)
- Dynamic Itemset Counting (DIC) (Dinamik Kümelenme Sayımı)
- CARMA (Continuous Association Rule Mining Algorithm) (Sürekli Birliktelik Kuralı Madenciliği Algoritması)
- FP-Growth (Sık Büyüme Modeli)

b) Paralel ve Dağıtılmış Algoritmalar;

- Count Distribution (CD) (Sayım Dağılımı)
- Parallel Data Mining (PDA) (Paralel Veri Madenciliği)
- Distributed Mining Algorithm (DMA) (Dağıtılmış Madencilik Algoritması)
- Common Candidate Partitioned Database (CCPD) (Ortak Aday Bölümlü Veritabanı)
- Data Distribution (DD) (Veri Dağıtımı)
- Intelligent Data Distribution (IDD) (Akıllı Veri Dağıtımı)

- Hash-Based Parallel Mining of Association Rules (HPA) (Paralel Madencilik Tabanlı Birliktelik Kuralları)
- Parallel Association Rules (PAR) (Paralel Birliktelik Kuralları)
- Candidate Distributed Algorithm (CDA) (Aday Dağıtılmış Algoritma)
- Skew Handling (SH) (Çarpık Dağıtım)
- Hybrid Distribution (HD) (Hibrit Dağıtım)

Yukarıda da belirtildiği çeşitli şekiller ve büyüklükte çok sayıda birliktelik kuralı algoritmaları mevcuttur. Bu çalışmada günümüzde birçok alanda daha çok tercih edilen birliktelik kuralı algoritmalarından olan Apriori Algoritması kullanılmıştır.

2.3.4. Apriori Algoritması

Günlük hayatta iş, bankacılık, sosyal ve sağlık hizmetleri, güvenlik ve siyaset alanlarında günlük olarak büyük miktarlarda veriler toplanmaktadır. Bu veri kümeleri yönetilen değerlerin ve ilişkilerin özelliklerini yansıtır. Bu tür büyük veri kümelerinden bilgiye ulaşmak için sağlam, basit ve hesaplama olarak verimli araçlar gerektirir. Bu tür araçların geliştirilmesi ve anlaşılması, veri madenciliğinin en önemli işlevidir. Bu araçlar bilgisayar bilimi, matematik ve istatistikten alınan bilgilere dayanmaktadır.

Apriori Algoritması, 1994 yılında Agrawal ve Srikant tarafından geliştirilmiştir. Veri Madenciliğine istatistik ve makine öğrenim topluluğunun dahil edilmesi ile bilgisayar bilimleri, matematik, fen bilimleri, işletme ve yönetim gibi birçok alandaki araştırmacılara da yön vermiştir. Veri madenciliğinin veri analizi yönü, istatistiklerden daha açıklayıcıdır ve sonuç olarak, veri madenciliğinde matematiksel kökenler istatistiklere göre biraz daha az belirgindir. Bununla birlikte, veri madenciliğindeki büyük ve karmaşık optimizasyon problemlerinin çözümü için matematiksel yöntemlerden faydalanılmaktadır. Bu özellikle karmaşık kurallar için geniş veritabanlarındaki araştırmayı gerektiren birliktelik kural madenciliğinde geçerlidir. Piyasa sepeti analizinde kullanılan orijinal tekniklerin çok çeşitli uygulamalarda yaygınlaştırılması için matematiksel modelleme gereklidir. Matematiksel analiz, algoritmaların performans durumuyla ilgili bilgileri verir.

Birliktelik kuralları, bir destek ya da veriler tarafından desteklenen kurallardır. Birliktelik kurallarında en çok kullanılan analiz, büyük perakendecilerin noktadan satış işlemlerinin içeriğini ele alan market sepeti analizidir. Buna örnek; “Ekmek ve Süt satın

alan tüm müşterilerin yüzde 90'ı da Gazete satın alabilir” olabilir. Müşteri davranışları hakkındaki öngörüler, müşteri anketleri yoluyla da elde edilebilir, ancak işlem verilerinin analizi, hem daha az maliyetlidir hem de tüm mevcut müşterileri kapsama avantajına sahiptir. Müşteri anketlerine nazaran, işlem verilerinin analiz edilmesi bazı önemli engellere sahiptir. Örneğin, satış noktası verileri tipik olarak kişisel ilgi alanları, yaş ve müşteriler hakkında bilgi içermez. Bununla birlikte, market sepeti analizi, müşteri davranışları hakkında yeni bilgiler sağlamanın yanında daha iyi müşteri ilişkileri, müşteri kazanma, daha iyi ürün yerleşimleri ve reklam, ürün geliştirme ve sahtekârlık tespiti yoluyla daha yüksek kar kazanımını sağlar. Bu bağlamda doğru müşteriye doğru ürünü veya ürünlerin nokta atışı pazarlanması maliyeti düşüren önemli bir etkidir.

Market sepeti analizi, perakende alışverişle sınırlı değil, aynı zamanda diğer iş alanlarında da uygulanabilir. Örnek;

- Kredi Kartı İşlemleri,
- Telekomünikasyon Hizmetleri,
- Bankacılık Hizmetleri,
- Sigortacılık,
- Müşteri Hizmetleri ve Pazarlama Hizmetleri,
- Tıp ve Sağlık

Birliktelik kural madenciliği piyasa sepeti analizi, metin veri analizi ve İnternet saldırı tespiti dahil olmak üzere birçok alanda kullanılmaktadır. Bir müşteri bir satış noktasından geçtiğinde, piyasa sepetinin içeriği kaydedilir. Bu, hangi ürünlerin satıldığı ve özellikle de hangi eşyanın satıldığına dair bilgi veren market sepeti verilerinin büyük birikimleriyle sonuçlanır.

Tablo 2.2 - Market Basket Ürün Tablosu

Market Basket ID	Ürün
1	Ekmek, Su
2	Süt, Ekmek, Gazete
3	Ekmek, Zeytin
4	Ekmek, Gazete, Su
5	Gazete

Her satır, bir popüler perakende ürün içeren bir market sepetine veya işlemine karşılık gelir. Tablo 2.2'deki tablo incelendiğinde aşağıdaki verileri ortaya koyar;

- Beş sepetin dördü Ekmek içerir
- İki sepet su içerir
- Ekmek içeren sepetlerin yarısı da su içerir
- Su içeren tüm sepetler de ekmek içerir

Büyük bir perakende mağazasında binlerce ürün satılmaktadır ve mağaza her gün binlerce müşteriye hizmet vermektedir. Böylece toplanan verilerin büyüklüğü önemlidir ve Tablo 2.2'nin sonucundaki gibi basit kuralların tespiti bile karmaşık algoritmalar gerektirebilir. Algoritmaların verimi, veri kümelerinin özelliklerine bağlı olarak artabilir veya azalabilir.

Müşteriler için en basit model, müşterilerin mağazadaki reyonlardan ürünleri rastgele seçtiğini varsayan modeldir. Bu durumda, her ürünün seçimi başka herhangi bir üründen bağımsız olacaktır. Sonuç olarak, ilişki kuralı keşfi, herhangi bir öğenin seçilmesi olasılığını belirler. Diğer modellerin performansını bu “sıfır hipotezi” ile karşılaştırmak önemli olsa da, alışveriş yapanların yani sepeti dolduranların daha karmaşık bir yaklaşıma sahip oldukları fark edilmiştir. Alışveriş yapan müşterinin kahvaltı malzemeleri, öğlen yemeği, akşam yemeği ve atıştırmalıklar, içecekler ve pazar akşam yemekleri vb ürünler, yüksek kaliteli, taze, az yağlı, özel diyet ve çevre açısından güvenli ürünler, marka ürünler, ucuz ürün tercihleri olacaktır. Alışveriş yapan müşterinin bu gibi hedefleri ve tercihleri seçimleri etkileyecek, ancak doğrudan gözlenemeyecektir. Bir anlamda market sepeti analizi, alışveriş yapanların aldıkları ürünleri nasıl seçtikleri hakkında bilgi sağlamaktadır [25].

Bir başka örnek ile Apriori Algoritmasının çalışma mantığı şu şekilde de ifade edilebilir. Bir marketin 10 müşterisinin sepetindeki ürünlerin listesi Tablo 2.3'deki gibi olsun;

Tablo 2.3 - Müşteri Alışveriş Sepeti

Müşteri	Ürünler
1	Sut,Su,Ekmek,Gazete
2	Sut,Su,Ekmek,Gazete,Yumurta
3	Su,Ekmek,Gazete
4	Su,Ekmek,Yumurta
5	Sut,Su,Gazete
6	Sut,Ekmek,Gazete
7	Su,Ekmek,Gazete,Yumurta
8	Sut,Ekmek,Gazete,Yumurta
9	Ekmek,Gazete,Yumurta
10	Sut,Su,Ekmek,Yumurta

Bu alışveriş listesine göre tekli ürünlü frekans tablosu Tablo 2.4'deki gibidir; bu tabloya göre her ürünün tek olarak kaç adet satıldığı görünmektedir.

Tablo 2.4 - Ürün Frekans Değerleri

Ürün	Tekli Frekans
Sut	6
Su	7
Ekmek	9
Gazete	8
Yumurta	6

Tablo 2.5'te İki frekans tablosunda da 2 ürünün bir arada satılma adetleri gösterilmektedir

Tablo 2.5 - İkili Ürün Frekans Değerleri

Ürün	İkili Frekans
Sut, Su	4
Sut, Ekmek	5
Sut, Gazete	5
Sut, Yumurta	3
Su, Ekmek	6
Su, Gazete	5

Tablo 2.6’da ise 3’lü olarak ürünlerin bir arada satılma adetleri görülmektedir.

Tablo 2.6 - Üçlü Ürün Frekans Değerleri

Ürün	Üçlü Frekans
Sut, Ekmek, Gazete	4
Su, Ekmek, Gazete	4
Su, Ekmek, Yumurta	4
Ekmek, Gazete, Yumurta	4

2.3.5. Apriori ve FP-Growth Algoritmaları

Apriori Algoritması ile beraber en fazla kullanılan birliktelik kuralı algoritması FP-Growth Algoritmasıdır. Bazı durumlar da performans olarak Apriori Algoritmasından daha iyi sonuçlar verse de bu iki algoritma da çalışma mantıkları farklı ancak verdikleri sonuçlar hemen hemen aynıdır.

Apriori Algoritması: Apriori algoritması aşağıdaki modüller ile çalışır.

- Ön işlem modülü
- Apriori veya FP Growth Algoritması Modülü
- Birliktelik Kuralı Üretimi
- Sonuçlar

Önişleme modülü, normalde ASCII biçiminde olan günlük dosyasını, Apriori algoritması tarafından işlenebilen formatta bir veritabanına dönüştürür. İkinci modül iki adımda gerçekleştirilir.

- Sık Kullanılan Öğe kümesi oluşturma
- Kuralların türetilmesi

FP-Growth Algoritması: FP-Growth algoritması aşağıdaki 4 modülde çalışır.

- Ön İşleme Modülü
- FP Tree bir FP Growth Modülü
- Birliktelik Kuralı Üretimi
- Sonuçlar

Ön işleme modülleri, normal olarak ASCII biçimindeki günlük dosyasını, FP Growth algoritması tarafından işlenebilen formatta bir veritabanına dönüştürür. 2. modül iki adımda gerçekleştirilir.

- FP Tree üretimi
- Birliktelik kuralları oluşturmak için FP Growth uygulama

FP-Tree: “Boş” olarak etiketlenmiş bir kökten oluşan sık model ağacıdır. Kökün çocukları olarak bir öge önek alt ağaçları kümesi ve sık öge başlık tablosu vardır. Öge öneki alt ağaçlarındaki her düğümün üç alanı vardır:

- *Öge Adı:* Düğümün hangi ögeyi temsil ettiğini gösterir,
- *İşlem Sayısı:* Bu düğüme ulaşan işlem sayısını gösterir,
- *Düğüm Bağlantısı:* Aynı öge adını taşıyan FP-Tree’deki bir sonraki düğüm bağlantısını gösterir. Düğüm bağlantısı yok ise “Boş” değeri verir [26].

2.3.6. Apriori ve FP-Growth Algoritmaları Karşılaştırma

Tablo 2.7 - Apriori ve FP-Growth Karşılaştırma Tablosu [27]

Algoritma	Teknik	Çalışma Zamanı	Bellek Kullanımı	Paralleleleştirilebilirlik
Apriori	Tekli, çift, üçlü vb. Üretir	Aday nesil aşırı derecede yavaştır. Çalışma zamanı, farklı öğelerin sayısına bağlı olarak katlanarak artar.	teklileri, çiftleri, üçlülere vb. Kaydeder.	Aday nesil çok paraleleleştirilebilir
FP-Growth	Sıralı öğeleri frekansa göre model ağacına ekler	Süreç, işlemlerin sayısına ve öğelerin sayısına bağlı olarak doğrusal olarak artar	Veritabanının sıkıştırılmış bir sürümünü saklar.	Veriler birbirine çok bağlıdır, her düğümün köke ihtiyacı vardır.

Hem Apriori hem de FP Growth algoritması, veritabanından sık kullanılan modelleri araştırmak için kullanılır. Her iki algoritma da sık görülen kalıpları keşfetmek için bazı teknikleri kullanır. Apriori algoritması büyük veritabanları ile iyi çalışır ancak FP Growth algoritması büyük veritabanıyla Apriori algoritmasına göre daha düşük

performans ile çalışır [28]. Bellek kullanımı, tarama sayısı ve tüketilen zaman arasındaki karşılaştırmalar aşağıdaki gibidir;

Tekniği: Apriori algoritması, Apriori özelliğini kullanır ve madenciliği sık kullanılan modellerin katkı yapılmamış ham özelliklerine katılır. FP Growth algoritması, minimum desteği karşılayan veritabanından serbest koşullu örnek ve model tabanı oluşturur.

Arama Türü: Apriori, ilk arama yöntemini kullanırken, FP Growth, bölme ve fethetme yöntemini kullanır.

Bellek Kullanımı: Apriori algoritması, çok sayıda aday veri seti nesli ile uğraşırken büyük bellek alanı gerektirir. FP Growth algoritması, aday veri seti üretimi olmadan sıkça rastlanan öğeleri keşfettiği sıkı yapısı nedeniyle daha az bellek gerektirir.

Tarama Sayısı: Apriori algoritması, aday kümesi oluşturmak için birden fazla tarama gerçekleştirir. FP Growth algoritması veritabanını yalnızca iki kez tarar.

Zaman: Apriori algoritmasında yürütme süresi, her seferinde adayların üretilmesinde daha fazla zaman harcar. FP Growth'un yürütme süresi, Apriori'ye göre daha azdır [29].

BÖLÜM 3. YÖNTEM VE UYGULAMA

3.1. Veriler ve Toplanması

Bu çalışmada www.partistore.com sitesinin Ocak 2016 – Nisan 2018 tarihleri arasındaki satış verileri incelenmiştir. Bu E-Ticaret sitesinden yapılan satış verileri üzerinde birliktelik kuralları Apriori Algoritması kullanılarak bir değerlendirme yapılmıştır. Bu veri tablosunda toplamda 33.061 satış kaydı olup veriler site sahibinin izniyle Excel dosyası olarak temin edilmiştir. Kullanılan veri içerisinde, 894 farklı ürün yer almaktadır, aynı zamanda 7.114 farklı müşteriye yapılan satış kayıtları yer almaktadır.

Tablo 3.1 - Tez Çalışmasına Konu Olan Veri Örneği

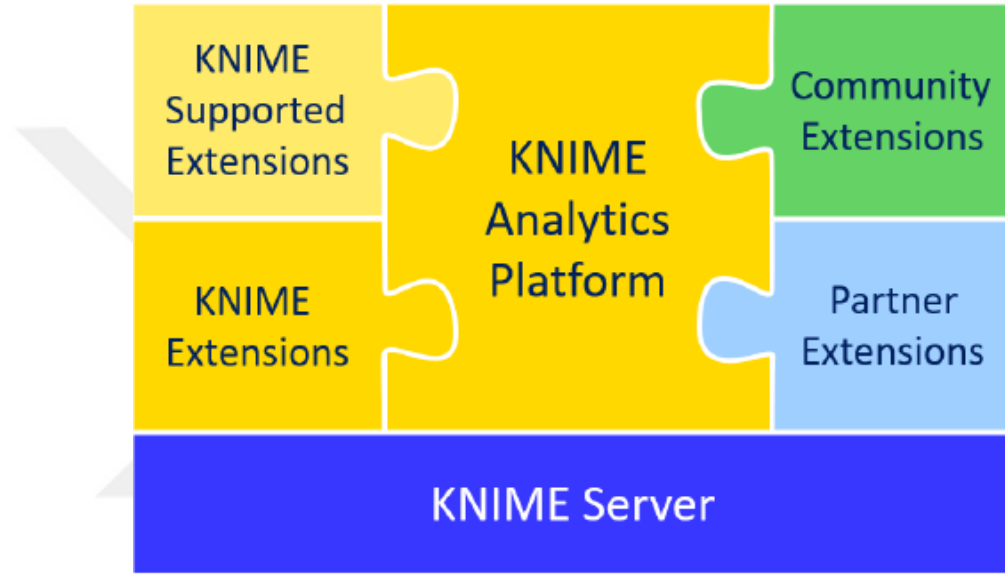
Tarih	Order Number	SKU	Name	Quantity
02.01.16	5634	5202-altin-rengi	2 Şekli Rakam Folyo Balon Renk 10/Renk 11 Renkleri	1
02.01.16	5636	4095	Tema 24 Tabak	3
02.01.16	5636	4103	Tema 24 Peçete	2
03.01.16	5636	4099	Tema 24 Temalı Karton Bardak	3
03.01.16	5911	4134	Tema 31 Tabak	6
03.01.16	5911	4136	Tema 31 Temalı Karton Bardak	6
03.01.16	5911	4150	Tema 31 Cupcake Standı	1
03.01.16	5911	4140	Tema 31 Masa Örtüsü	1
03.01.16	5911	4233	Dratif Kürdan Kız Doğum Günü Parti Malzemesi	1
03.01.16	5941	4751	Tema3 Masa Örtüsü	1
03.01.16	5941	4220	Tema3 Tabak	4
03.01.16	5941	4750	Tema3 Temalı Karton Bardak	3
03.01.16	5941	4749	Tema3 Peçete	2
03.01.16	5941	5670	Tema 34 Baskılı Parti Balonu	2

3.1.1. KNIME Uygulaması ile Detayları

KNIME (Konstanz Information Miner), Java tabanlı açık kaynak platform bir uygulamadır. Verilerde saklı olan potansiyelin keşfi, ve yeni öngörüler elde etmek veya yeni oluşumları tahmin etmek için tasarlanmış bir uygulamadır. Birçok sektörde, yıllardan beri 60'tan fazla ülkede farklı veriler ve platformlar üzerinde çalışmalar

yapılmaktadır. Örneğin; rakamlardan görüntülere, moleküllerden insanlara, sinyallerden karmaşık ağlara ve büyük veri analizine kadar birçok alanda KNIME uygulaması kullanılmaktadır [30].

KNIME uygulaması <https://www.knime.com> adresi üzerinden ücretsiz olarak indirebilir. Sitede “Learning” bağlantısı içerisinde KNIME kullanımı hakkında genel bilgiler ve eğitim dökümanları yer almaktadır. “Software” bağlantısı altında yazılım ile ilgili bir takım genel bilgiler yer almaktadır.



Şekil 3.1 - KNIME Uygulamasının Yapısı [30]

KNIME uygulamasında KNIME in kendine ait Server'ı ve KNIME Analytics Platform bulunmaktadır. Bu çalışma KNIME Analytic Platform kullanılarak yapılmıştır.

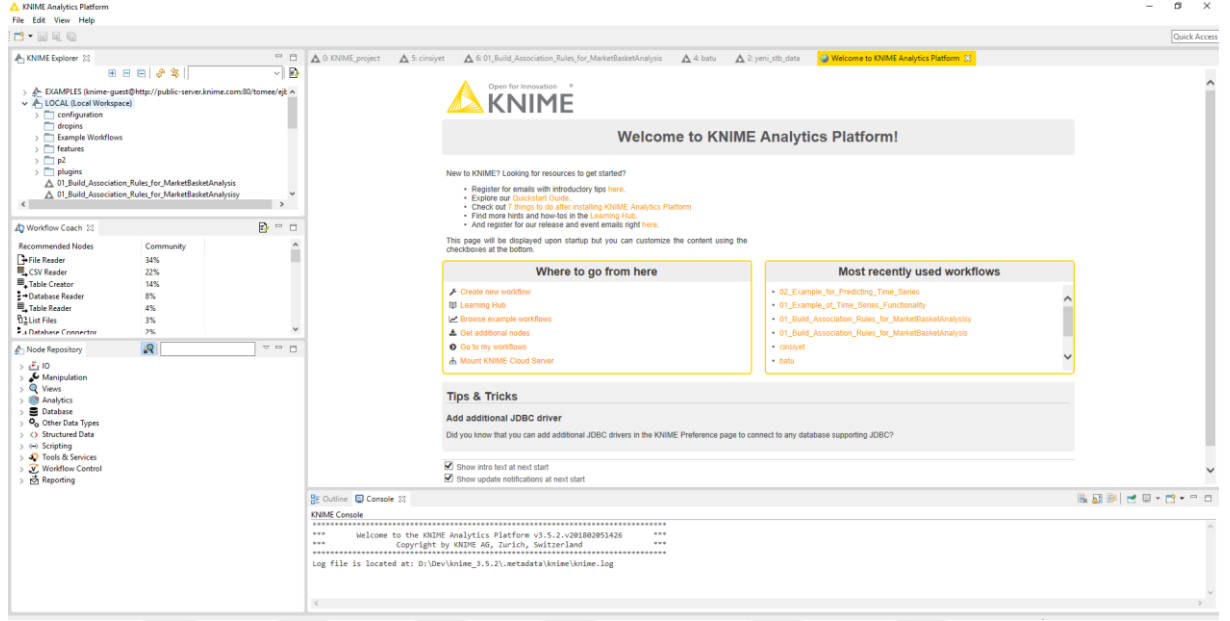
KNIME uygulaması, 2000'den fazla modül, yüzlerce kullanıma hazır örnek, kapsamlı bir entegre araç yelpazesi ve mevcut en geniş kapsamlı algoritma seçeneği sunar.

KNIME uygulaması işletim sistemine göre Windows, Linux, Mac, SDK (Software Development Kit) seçenekleri mevcuttur. Download butonuna tıklanıldığında getting started altında da eğitim bilgileri de yer almaktadır

KNIME uygulaması indirildikten sonra herhangi bir yükleme işlemine gerek duyulmadan çalışabilmektedir. Uygulama açıldıktan sonra workspace alanı belirlenmesi gerekiyor. Ekran açıldığında birkaç alt ekrandan oluşuyor. Sol üst köşede mevcut projeler ve knime içindeki örnekler yer almaktadır. Ana ekranın sol tarafındaki node repository

kısımında uygulama içerisinde kullanılabilir hazır node'lar bulunmaktadır. Bu alandan sürükleyip bırak mantığı ile istenilen node'lar ile çalıştırılmaktadır.

KNIME uygulamasında File → Install KNIME Extensions alanından yeni eklentiler eklenebilir.



Şekil 3.2 - KNIME Ana Ekranı

KNIME Uygulaması KNIME Analytics Platform ve KNIME server olarak 2 başlık altında incelenebilir.

3.1.2. KNIME Analytics Platform

Powerful Analytics (Güçlü Analiz):

- **Güçlü:** Yerel düğümlerin ve araç entegrasyonları için oluşturulmuş geniş bir uygulama, KNIME Analytics Platformunu herhangi bir veri bilimcisi için mükemmel bir araçtır.
- **Güvenilir ve Emniyetli:** 2008 yılından bu yana yılda iki defa oluşturulan yeni yazılım sürümleri ve binlerce özel kullanıcının tercih ettiği bir uygulamadır.
- **Ölçeklenebilir:** Tek bir bilgisayar, akış ve büyük veriler arasında kolaylıkla geçiş yapma imkanı. Yeni yetenekleri mevcut altyapı altında entegre edilebilir olması.

- **Kapasiteyi Geniřletmek:** Daha iyi iřbirlięi, otomasyon ve daęıtım iřlevleri iin mevcut kapasitelerin KNIME Sunucusu ile geniřletilmesi.

Data and Tool Blending (Veri ve Araların Uyum):

- **Büyük Veri Uzantıları:** KNIME Büyük Veri Uzantıları olan, Apache Hadoop ve Apache Spark'ın gücünü KNIME Analytics Platform ve KNIME Server'in kullanım kolaylıęı ile birleřtiriyor.
- **Veri Harmanlama:** Basit metin dosyaları, veritabanları, belgeler, görüntüler, aęlar ve hatta Hadoop tabanlı verileri birleřtirerek alıřma imkanı sunar.
- **Aralar Arası Uyum:** Eski bir kodlama / kod (R & Python) dahil olmak üzere birok fazla aracın entegrasyonu, grafiksel olarak belgelenmesi ve veri bilimcileri arasında paylařılmasını saęlar.
- **Görsel:** Öğrenmesi kolay grafik arayüz, kodlamanın isteęe baęlı olduęu ve görsel olarak belgelenebilen uygulamadır.

Productivity Features (Verimlilik Özellikleri):

- **Metanode Baęlantısı:** Metanodeları kontrol etmek iin pratik bir özellięe sahiptir. İř akıřı kontrol edildikten ve iř akıřına yerleřtirildikten sonra, KNIME bu baęlantıyı kaydeder ve bir metanode güncellendiyse veya yeni sürüm kontrol edildiyse bunu her durumda otomatik olarak günceller.
- **Yerel Otomasyon:** İř Akıřı düęümlerinin herhangi bir iř akıřına dahil edilmesini ve parametre vermesini saęlayarak yeniden kullanılabilir iř akıřlarının oluřturulmasına olanak tanır.
- **İř Akıřı Farkı:** Bir iř akıřında yapılan deęiřiklikleri karřılařtırıp bunların farklı sürümler olup olmadıęını, bir iř arkadařının yaptıęı deęiřiklikler veya kazayla meydana gelen deęiřikliklere karřı güvence verir ve iř Akıřı Farkı karřılařtırmalarını kolaylařtırır.

Open Platform (Açık Platform):

- **Sınırsız Açık Kaynak:** En yeni, eksiksiz kod tabanı GPLv3 lisansı altında, büyük işletim sistemleri desteklenmektedir.
- **Açık Platform:** Dünya genelinde önemli veri bilimciler topluluğunu en zor analitik sorunlarına çözüm getirerek inovasyon çalışmalarını güçlendirir.
- **Taşınabilir ve Dayanıklı:** Geriye dönük uyumluluk, mevcut iş akışlarının yeni sürümlerle çalışmaya devam etmesini sağlar, çalışmaları geleceğe dönük olarak korur. Sektör lideri PMML desteği, sorunsuz model taşınabilirliği ve dağıtımını sağlar [30].

3.1.3. KNIME Server

KNIME Server, KNIME Analytics Platformunun gücünü ve esnekliğini, işbirliği, dağıtım ve yönetim işlevlerini sağlar. KNIME Server, küçük gruplardan büyük şirketlere kadar her boyutta takımın birlikte daha iyi çalışmasını sağlar.

3.1.4. Birlikte Alınan Ürünler Analizi ve İnceleme

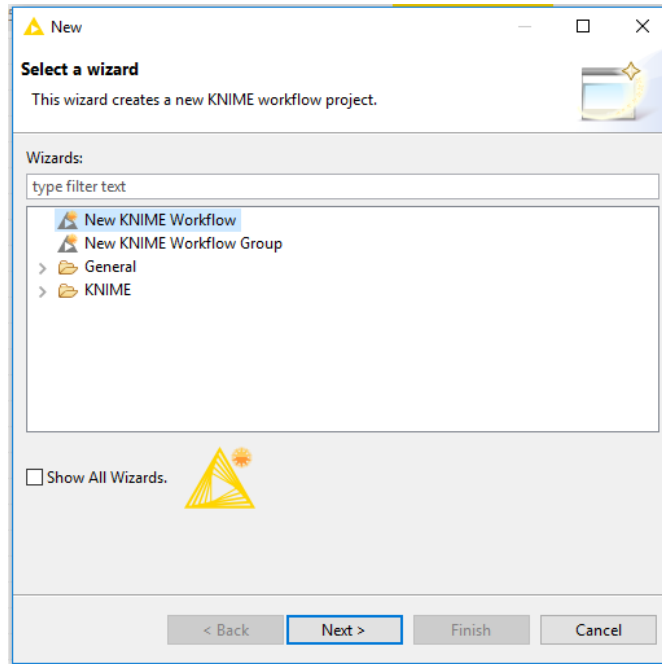
KNIME uygulaması ile bir e-ticaret sitesinin satış verileri üzerinden birliktelik kurallarından Apriori algoritması kullanarak birlikte satılan ürünlerin analizi yapılmıştır. Verinin içeriği incelendiğinde analizi yapılacak olan verideki başlıklar Tablo 3.2'de gösterilmiştir. Order number olarak belirtilen sütunda alışveriş yapan müşteriye atanan bir müşteri id bilgisidir. SKU satılan ürünlerin kodudur. Name satılan ürünlerin adıdır. Quantity sütununda ise o üründen o satış işleminde kaç adet satıldığı bilgisini göstermektedir.

Tablo 3.2 - Analizi Yapılan Örnek Veri

Tarih	Order Number	SKU	Name	Quantity
02.01.16	5634	5202-altin-rengi	2 Şekli Rakam Folyo Balon Renk 10/Renk 11 Renkleri	1
02.01.16	5636	4095	Tema 24 Tabak	3
02.01.16	5636	4103	Tema 24 Peçete	2
03.01.16	5636	4099	Tema 24 Temalı Karton Bardak	3
03.01.16	5911	4134	Tema 31 Tabak	6
03.01.16	5911	4136	Tema 31 Temalı Karton Bardak	6
03.01.16	5911	4150	Tema 31 Cupcake Standı	1
03.01.16	5911	4140	Tema 31 Masa Örtüsü	1
03.01.16	5911	4233	Dratif Kürdan Kız Doğum Günü Parti Malzemesi	1
03.01.16	5941	4751	Tema3 Masa Örtüsü	1
03.01.16	5941	4220	Tema3 Tabak	4
03.01.16	5941	4750	Tema3 Temalı Karton Bardak	3
03.01.16	5941	4749	Tema3 Peçete	2
03.01.16	5941	5670	Tema 34 Baskılı Parti Balonu	2

Bu çalışmaya konu olan veri Excel formatında temin edilmiştir, KNIME uygulaması açıldığında

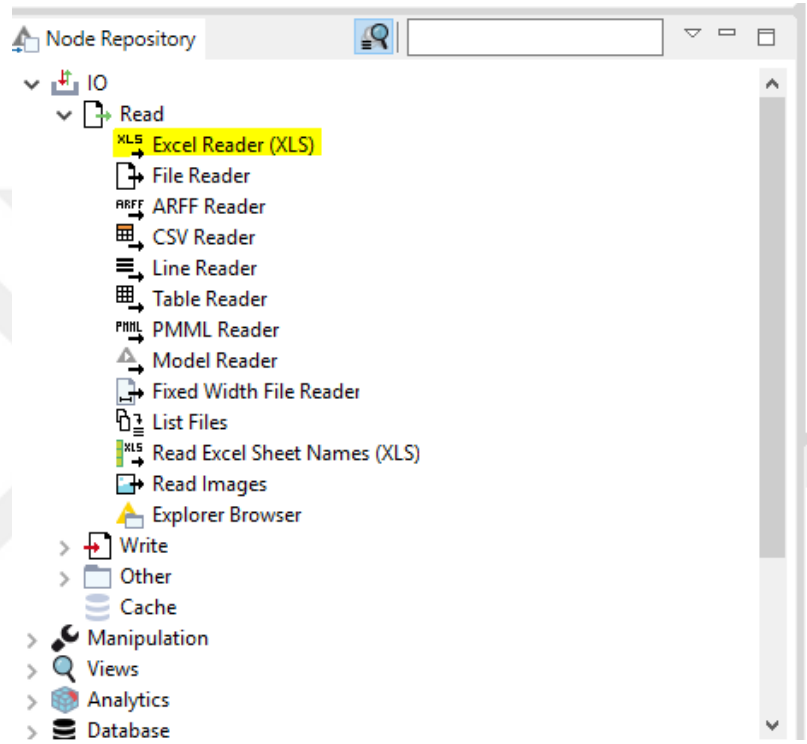
File – new seçeneklerinden Şekil 3.3’deki ekranda yeni bir workflow açılıp KNIME uygulamasında bir çalışma alanı oluşturulabilir.



Şekil 3.3 - KNIME Yeni Workflow Açma

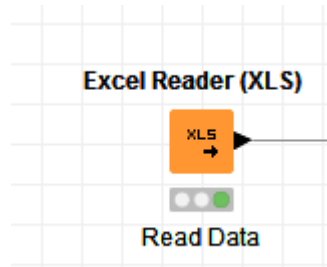
New KNIME Workflow seçeneği seçilip sonra workflow'a bir isim vererek KNIME da yeni bir çalışma alanı oluşturulur.

Kullanılacak olan veriyi Excel formatında olduğu için uygulama içerisinde sol taraftaki “Node Repository” alanından IO → Read → Excel Reader (XLS) node’u sürükleyip bırakılarak workflow içerisine eklenmelidir.



Şekil 3.4 - KNIME Excel Reader

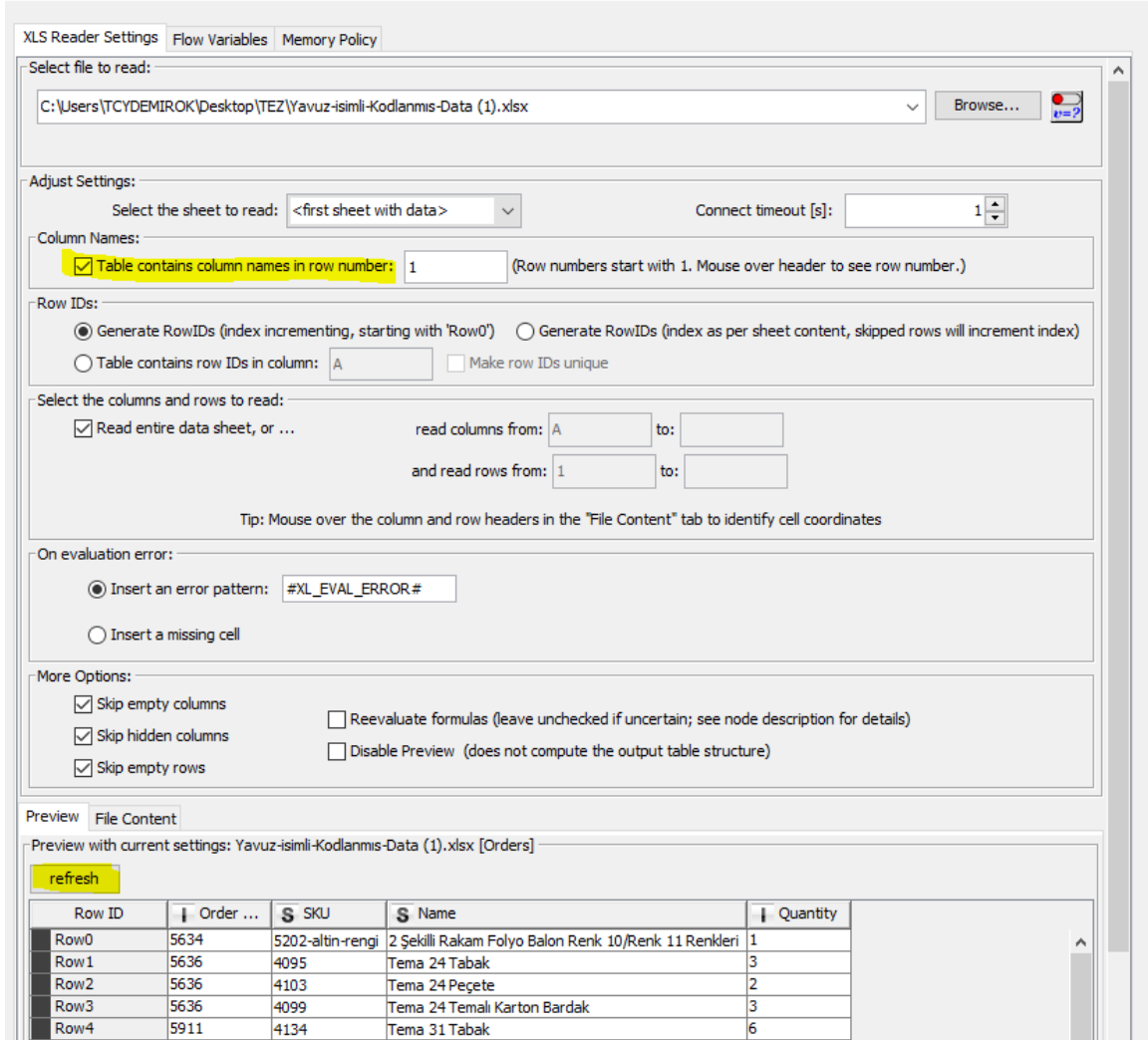
Node’u çalışma alanına eklendikten sonra node üzerinde fare ile sağ tıklanarak – configure seçeneği seçildiğinde üzerinde çalışılacak veri seçilebilir.



Şekil 3.5 - KNIME Excel Reader Node

Şekil 3.6’da seçilen Excel verisinden hangi sayfa kullanılmak isteniyorsa select the sheet to read sekmesinde seçilebilir. Excel dosyasındaki ilk satırdaki verileri başlık olarak

kullanabilmek için “table contains column names in row number” seçeneği işaretlenmelidir. Configure alanında herhangi bir değişiklik yapıldığında değişiklikleri görebilmek için refresh tuşuna basılmalıdır.



XLS Reader Settings | Flow Variables | Memory Policy

Select file to read:
C:\Users\TCYDEMİROK\Desktop\TEZ\Yavuz-ismli-Kodlanmis-Data (1).xlsx [Browse...]

Adjust Settings:
Select the sheet to read: <first sheet with data> | Connect timeout [s]: 1

Column Names:
 Table contains column names in row number: 1 (Row numbers start with 1. Mouse over header to see row number.)

Row IDs:
 Generate RowIDs (index incrementing, starting with 'Row0') | Generate RowIDs (index as per sheet content, skipped rows will increment index)
 Table contains row IDs in column: A | Make row IDs unique

Select the columns and rows to read:
 Read entire data sheet, or ... | read columns from: A to: |
and read rows from: 1 to: |

Tip: Mouse over the column and row headers in the "File Content" tab to identify cell coordinates

On evaluation error:
 Insert an error pattern: #XL_EVAL_ERROR# | Insert a missing cell

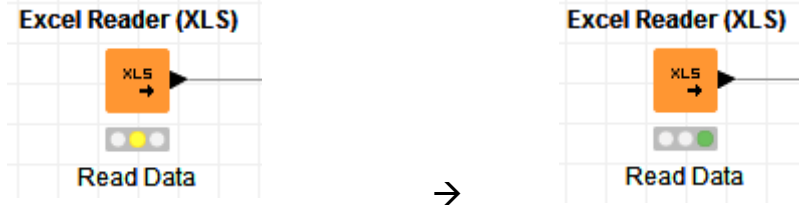
More Options:
 Skip empty columns | Reevaluate formulas (leave unchecked if uncertain; see node description for details)
 Skip hidden columns | Disable Preview (does not compute the output table structure)
 Skip empty rows

Preview | File Content
Preview with current settings: Yavuz-ismli-Kodlanmis-Data (1).xlsx [Orders] [refresh]

Row ID	Order ...	S SKU	S Name	Quantity
Row0	5634	5202-altin-rengi	2 Şekli Rakam Folyo Balon Renk 10/Renk 11 Renkleri	1
Row1	5636	4095	Tema 24 Tabak	3
Row2	5636	4103	Tema 24 Peçete	2
Row3	5636	4099	Tema 24 Temalı Karton Bardak	3
Row4	5911	4134	Tema 31 Tabak	6

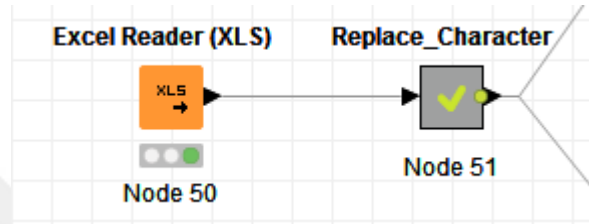
Şekil 3.6 - KNIME Excel Reader Node Configure

Configure ayarları tamamlandıktan sonra Excel reader node'unu sağ tıklayıp Execute yapıldığında Şekil 3.7'deki gibi “Sarı” durumdan “Yeşil” duruma geçerek veri bir sonraki node'a taşınabilir. Bu alan sarı veya kırmızı olması durumunda diğer node'lara veri akışı olmamaktadır. Kırmızı olması durumu yapılan konfigurasyon ayarlarında bir hata olduğu anlamına gelmektedir.



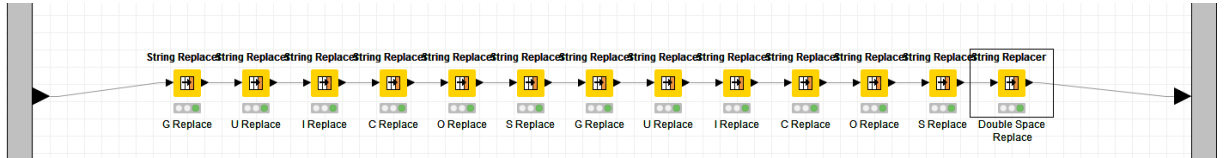
Şekil 3.7 - KNIME Node Sarı ve Yeşil Statüler

Excel data reader node'undan sonra Metanode seti ile kullanılan veri üzerinde temizleme işlemleri yapılabilir.



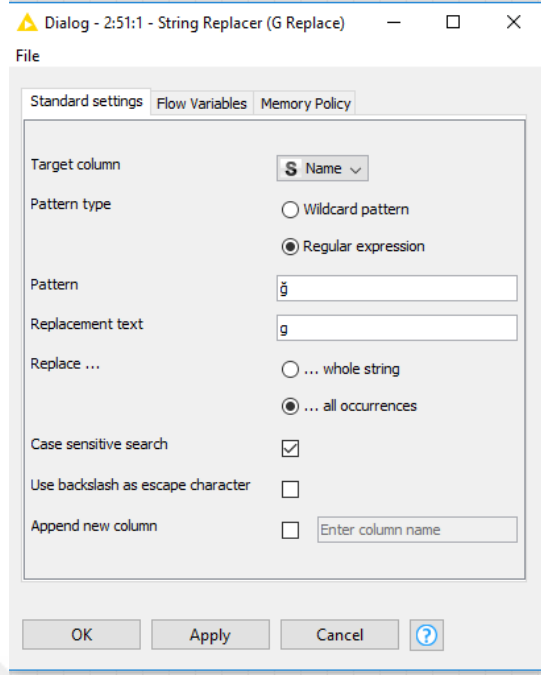
Şekil 3.8 - KNIME Metanode

Replace_Character olarak eklenen node'un içerisinde Şekil 3.9'daki gibi node seti vardır. Bu node seti aracılığıyla veriler hem Türkçe karakterlerden temizlenip hem de çift boşlukların tek boşluk olarak düzenlenmesi sağlanabilir.



Şekil 3.9 - KNIME Metanode İçerik Node'ları

Bu node'ların içeriğini node üzerinde sağ tıklanıp konfigurasyon ayarları Şekil 3.10'daki gibi set edilebilir. Target column olarak string bir alan seçilmesi gerekiyor ve bu alan içerisinde replace işlemi yapılabilmektedir.

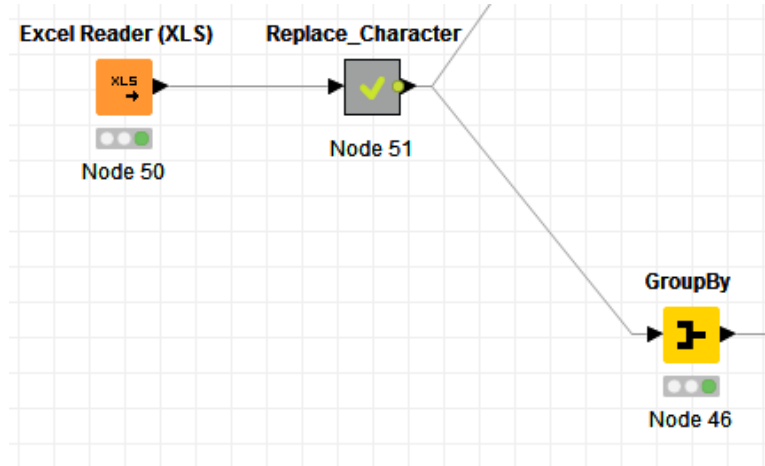


Şekil 3.10 - KNIME String Replace Node Configure

Çalışma alanına eklenen bir sonraki node GroupBy node’u aslında bu modeldeki veriyi birliktelik kuralı uygulanabilecek ilk formata getirebilmek için önemli bir node’dur. Bu node’un konfigurasyonu içerisinde 2 önemli unsura dikkat edilmesi gerekir. Çünkü kullanılan verilerde eğer aynı müşteri birden fazla ürün almış ise aldığı her ürün ayrı ayrı satırlarda gösterilmelidir. Bu verileri her müşteriyi tek satır, ve aldığı ürünleri de aralarında virgül olacak şekilde yanyana olacak şekilde düzenlenmesi gerekecektir.

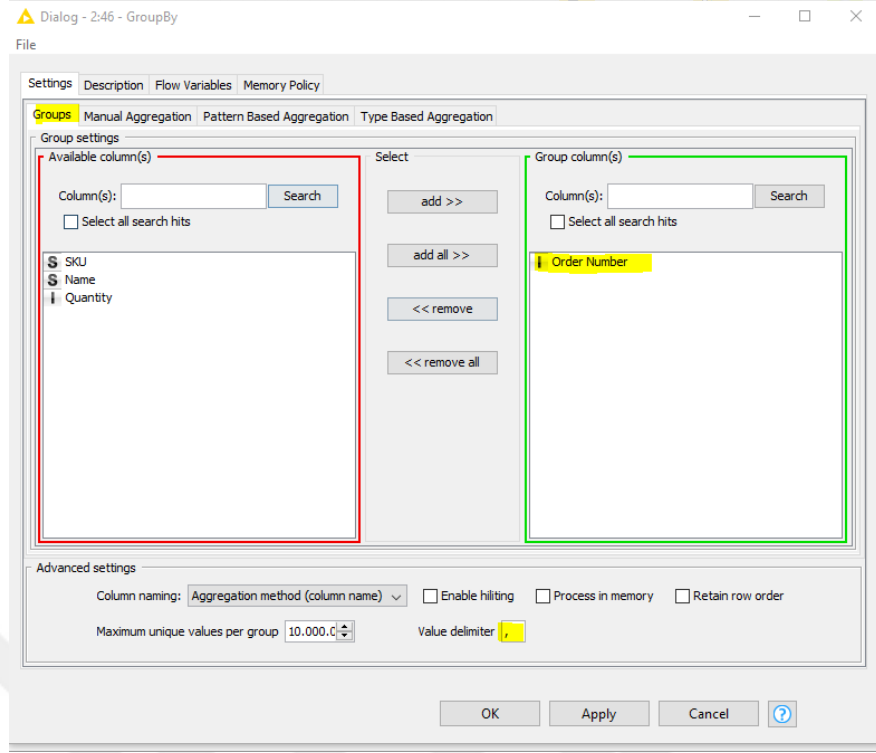
Tablo 3.3 - GroupBy Node'u İle Tekilleştirilecek Order Number Örneği

Order Number	SKU	Name	Quantity
5634	5202-altin-rengi	2 Şekli Rakam Folyo Balon Renk 10/Renk 11 Renkleri	1
5636	4095	Tema 24 Tabak	3
	4103	Tema 24 Peçete	2
	4099	Tema 24 Temalı Karton Bardak	3
5911	4134	Tema 31 Tabak	6
	4136	Tema 31 Temalı Karton Bardak	6
	4150	Tema 31 Cupcake Standı	1
	4140	Tema 31 Masa Örtüsü	1
	4233	Dratif Kürdan Kız Doğum Günü Parti Malzemesi	1
5941	4751	Tema3 Masa Örtüsü	1
	4220	Tema3 Tabak	4
	4750	Tema3 Temalı Karton Bardak	3
	4749	Tema3 Peçete	2
	5670	Tema 34 Baskılı Parti Balonu	2



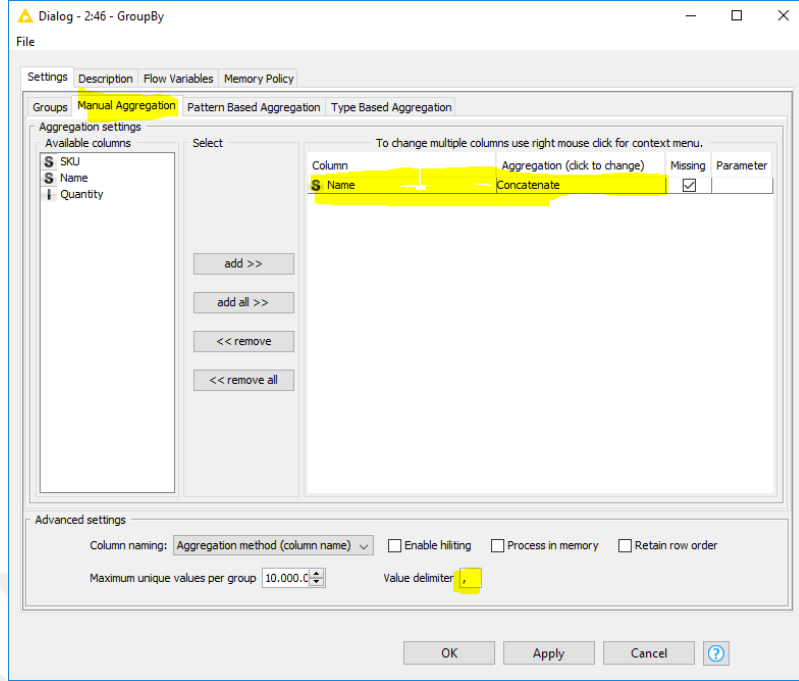
Şekil 3.11 - KNIME GroupBy Node

GroupBy node'u konfigurasyonu içerisinde Groups alanında order number seçilmesi demek, müşteri bilgilerini tek satır halinde gösterilmek istendiğini ifade eder.



Şekil 3.12 - KNIME GroupBy Node Configure 1

Groups kırılımında işlem yaptıktan sonra Manuel Aggregation kırılımında Name kolonunu seçerek aggregation parametresi olarak Concatenate yi seçmek gerekiyor. Value delimiter kısmında verileri hangi simge ile ayrılması gerektiği belirtilir, bu çalışmada veriler, arasında virgül ile ayrılmıştır. Böylelikle müşteri bazında tek tek olacak şekilde ve ürün ismi bazında ürünleri yanyana, aralarında virgül olacak şekilde birleştirilmektedir.

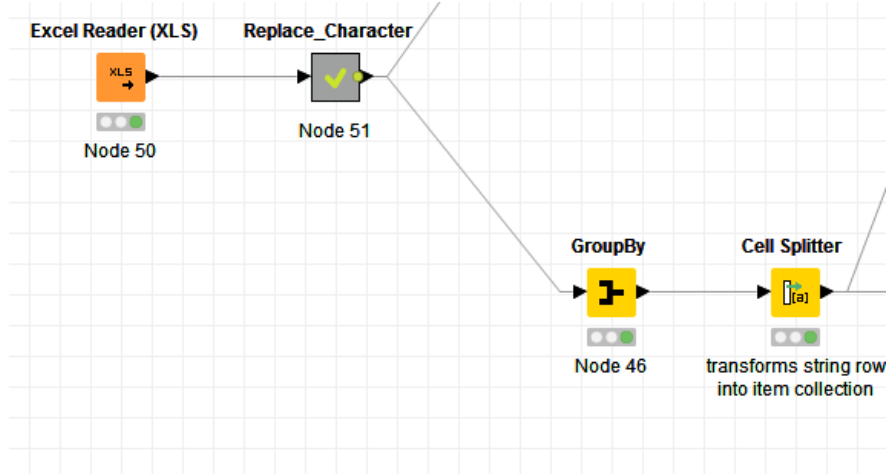


Şekil 3.13 - KNIME GroupBy Node Configure 2

Bu işlemler tamamlandıktan sonra node üzerinde fare ile sağ tıklayıp execute and open views seçildiğinde node çalıştırılacak ve sonrasında verinin oluşan son hali açılacaktır. Şekil 3.14'te olduğu gibi veriler gruplanmıştır.

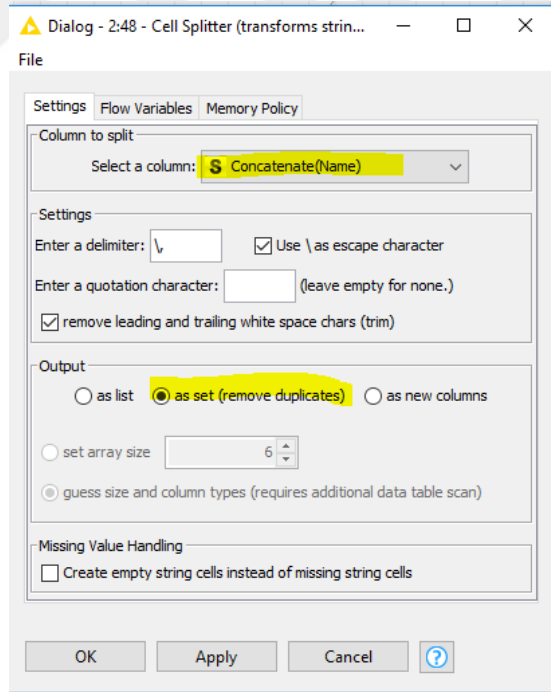
Row ID	Order	Concatenate(Name)
Row0	5634	2 Sekilli Rakam Folyo Balon Renk 10/Renk 11 Renkleri
Row1	5636	Tema 24 Tabak,Tema 24 Pecete,Tema 24 Temali Karton Bardak
Row2	5911	Tema 31 Tabak,Tema 31 Temali Karton Bardak,Tema 31 Cupcake Standi,Tema 31 Masa Ortusu,Dratif Kurdan Kiz Dogum Gunu Parti Malzemesi
Row3	5941	Tema3 Masa Ortusu,Tema3 Tabak,Tema3 Temali Karton Bardak,Tema3 Pecete,Tema 34 Baskili Parti Balonu
Row4	5957	O Sekilli Harf Folyo Balon Renk 10/Renk 11 Renkleri,2 Sekilli Rakam Folyo Balon Renk 10/Renk 11 Renkleri,Karisik Renkli Puanli Parti Balonu
Row5	5973	Plastik Masa Etegi Renk 8 Renkli
Row6	5998	Parti Seti Tema 24 8 Kislilik,Tema 24 Temali Karton Bardak,Tema 24 Tabak,1 Sekilli Rakam Folyo Balon Renk 10/Renk 11 Renkleri,Tema 24 Mum
Row7	6224	Dogum Gunu Renkli Dratif Kurdan,Renk5 Puanli Karton Tabak,Kullan at Karton Bardak Renk5 Puanli,Renk5 Kagit Pecete
Row8	6254	Tema3 Dogum gunu Afisi,Tema3 Masa Ortusu,Tema3 Pecete,Tema3 Harf Afisi,Tema3 Kaynana Dili,Tema3 Parti Taci,3 Yas Dogum Gunu Renkli Parti Balonlari,Tema3 Tabak,Seker RenkSsi Plastik Catal
Row9	6324	Renk 10 Plastik Catal
Row10	6480	Plastik Masa Etegi Renk4 Renkli
Row11	6481	Tema 19 Temali Kagit Pecete,Tema 19 Temali Bardak,Tema 19 Temali Latex Balon,Tema 19 Temali Tabak
Row12	6483	Renk 9 Puanliyele Karton Tabak,Renk 9 Zigzag Karton Tabak,Kullan at Karton Bardak Renk 9 Zigzagli,Kullan at Karton Bardak Renk 9 Puanliyele,Renk 8 Puanli Kagit Pecete
Row13	6565	1 Yas Kiz Dogum Gunu Renk5 Balon,Tema 39 Temali Latex Balon,Iyi Ki Dogdun Baskili Balon
Row14	6568	Tema3 Masa Ortusu,Tema3 Kalp Folyo Balon
Row15	6569	Tema5 Masa Ortusu,Tema5 Temali Karton Tabak,Tema5 Temali Bardak,Tema5 Pecete,Tema5 Kulah Sapka

Şekil 2.14 - KNIME GroupBy output view



Şekil 3.15 - KNIME Cell Splitter Node

Veri Şekil 3.14'deki hale geldikten sonra kümeleme işlemi için Cell Splitter node'u eklenmelidir ve konfigurasyon ayarları Şekil 3.16'daki gibi yapılmalıdır. Kümelemek istenilen kolon seçilip tekrarlanan veriyi temizlemek için as set (remove duplicates) seçeneği seçilmelidir. Ardından bu node yeniden execute edilerek çalıştırılır.



Şekil 3.16 - KNIME Cell Splitter Node Configure

Bu node çalıştırdıktan sonra Şekil 3.17'deki gibi bir veri elde edilmektedir. Bu kolon bir sonraki node'a eklenecek olan giriş değeridir. Birliktelik kuralı analizini yapacak olan bir sonraki node'a veri Şekil 3.17'deki

(Concatenate(Name)_SplitResultSet) formatta verilmesi gerekiyor. Bu işlemleri yapının bu şekile getirilmesindeki temel amaçta aslında budur.

Output Table - 2:48 - Cell Splitter (transforms string row)

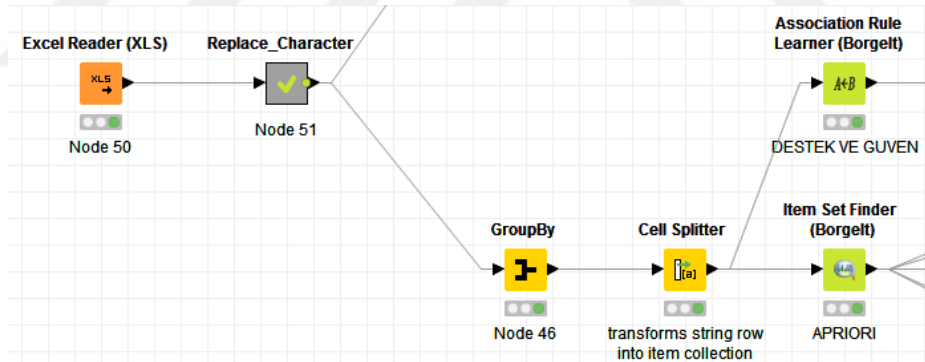
File Hilite Navigation View

Table "default" - Rows: 7114 | Spec - Columns: 3 | Properties | Flow Variables

Row ID	Order ...	Concatenate(Name)	Concatenate(Name)_SplitResultSet
Row0	5634	2 Sekilli Rakam Folyo Balon Renk 10/Renk 11 Renkleri	[2 Sekilli Rakam Folyo Balon Renk 10/Renk 11 Renk...
Row1	5636	Tema 24 Tabak,Tema 24 Pecete,Tema 24 Temali K...	[Tema 24 Tabak,Tema 24 Pecete,Tema 24 Temali K...
Row2	5911	Tema 31 Tabak,Tema 31 Temali Karton Bardak,Te...	[Tema 31 Tabak,Tema 31 Temali Karton Bardak,Te...
Row3	5941	Tema3 Masa Ortusu,Tema3 Tabak,Tema3 Temali K...	[Tema3 Masa Ortusu,Tema3 Tabak,Tema3 Temali K...
Row4	5957	0 Sekilli Harf Folyo Balon Renk 10/Renk 11 Renkleri...	[0 Sekilli Harf Folyo Balon Renk 10/Renk 11 Renkleri...
Row5	5973	Plastik Masa Etegi Renk 8 Renkli	[Plastik Masa Etegi Renk 8 Renkli]
Row6	5998	Parti Seti Tema 24 8 Kisiik,Tema 24 Temali Karton B...	[Parti Seti Tema 24 8 Kisiik,Tema 24 Temali Karton ...]
Row7	6224	Dogum Gunu Renkli Dratif Kurdan,Renk5 Puanli Kar...	[Dogum Gunu Renkli Dratif Kurdan,Renk5 Puanli Kar...
Row8	6254	Tema3 Dogum gunu Afisi,Tema3 Masa Ortusu,Tem...	[Tema3 Dogum gunu Afisi,Tema3 Masa Ortusu,Tem...
Row9	6324	Renk 10 Plastik Catal	[Renk 10 Plastik Catal]
Row10	6480	Plastik Masa Etegi Renk4 Renkli	[Plastik Masa Etegi Renk4 Renkli]
Row11	6481	Tema 19 Temali Kagit Pecete,Tema 19 Temali Bard...	[Tema 19 Temali Kagit Pecete,Tema 19 Temali Bard...
Row12	6483	Renk 9 Puantiyeli Karton Tabak,Renk 9 Zigzag Kart...	[Renk 9 Puantiyeli Karton Tabak,Renk 9 Zigzag Kart...

Şekil 3.17 - KNIME Cell Splitter Output View

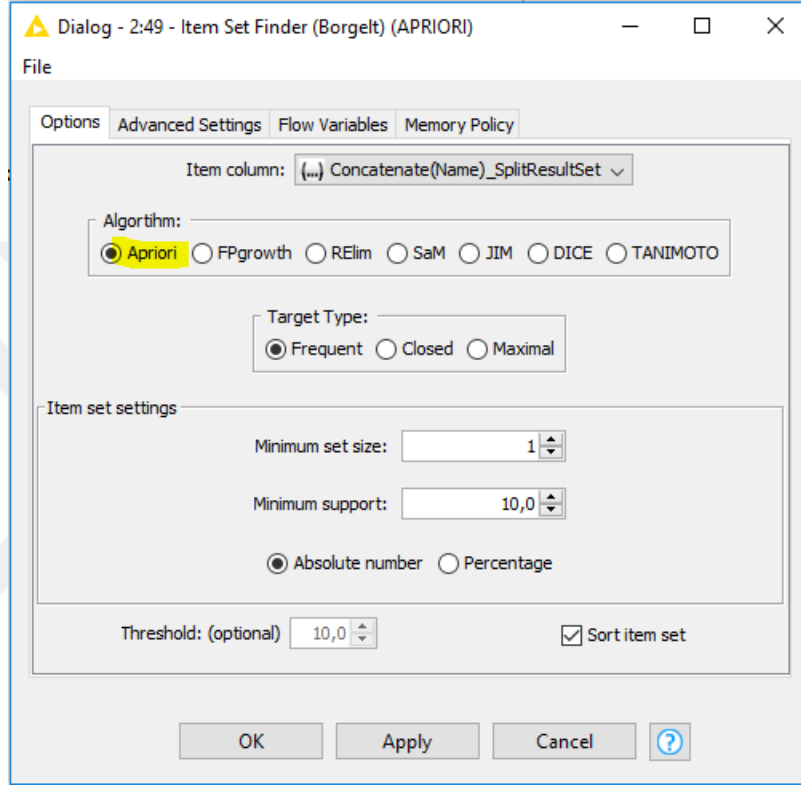
Birliktelik kuralı için 2 ayrı node kullanılmıştır, bunlardan bir tanesiyle birlikte alınan ürünlerin seti bulunmaktadır. Diğeriyle ise ürünler arasında destek ve güven değerleri bulunmaktadır.



Şekil 3.18 - KNIME Item Set Finder Node ve Association Rule Learner Node

Bu 2 node'dan birisi Item Set Finder diğeri ise Association Rule Learner node'larıdır. Item Set Finder node'u ile Apriori ve içeriğindeki farklı algoritmaları kullanarak bir birliktelik seti elde edilebiliyor. Configurasyon ayarlarında gelen seçeneklerden Apriori seçilerek bu çalışmada Apriori Algoritması kullanılacağı belirtilmiştir. Target type olarak frequent yani ürün seti bazında bir birliktelik çıkartılabilir. Minimum set size alanında tekil olarak satılan ürünlerinde görünmesi isteniyor ise 1 seçilebilir. Eğer ki 2 veya daha yüksek bir değer seçilirse 2 veya daha yüksek sayıdaki ürünün bir arada satılma verileri gelecektir. Minimum support alanında

ise en az kaç satış ve yukarısı isteniyor ise o rakam girilerek bir sonuç elde edilebilir. Örneğin 10 değeri girildiğinde A ürünü 10 adetten az satıldı ise gelen sonuç listesinde A ürünü görünmeyecektir. Absolute number seçeneği ise verilerin yüzdesel olarak değil sayısal olarak listelenmesi istendiği için seçilmiştir. Belli yüzdelikler olarak görmek için percentage seçeneği seçilebilir.



Şekil 3.19 - KNIME Item Set Finder Node Configure

Şekil 3.19'daki konfigurasyon ayarlarına göre execute edildiğinde sonuçlarla ilgili şöyle bir sonuca varılmıştır. Birlikte satılan ürün seti olarak en fazla 6'lı ürün setlerine ulaşılmıştır. 6'lı, 5'li, 4'lü, 3'lü ve 2'li ürün setleri için ilk 10 satış verileri aşağıdaki tablolarda gösterilmiştir.

Tablo 3.4 - 6'lı Birliktelik İlk 10 Listesi

ItemSet	ItemSetSize	ItemSetSupport
[Tema13 Bayrak Afis, Tema13 Dogum Gunu Tabagi, Tema13 Karton Bardak, Tema13 Masa Ortusu, Tema13 Temali Latex Balon, Tema13 Temali Pecete]	6	14
[Renk1 Renkli Latex Parti Balonlari 10 Adet, Renk2 Renkli Latex Parti Balonlari 10 Adet, Tema8 Kurdan 10 Adet, Tema8 Temali Bardak, Tema8 Temali Kagit Pecete, Tema8 Temali Karton Tabak]	6	12
[Renk1 Renkli Latex Parti Balonlari 10 Adet, Renk2 Renkli Latex Parti Balonlari 10 Adet, Tema8 Temali Bardak, Tema8 Temali Kagit Pecete, Tema8 Temali Karton Tabak, Tema8 Temali Masa Ortusu]	6	12
[Plastik Masa Etegi Renk1 Renkli, Renk1 Petek Sus Seti, Renk1 Renkli Latex Parti Balonlari 10 Adet, Renk2 Asma Ponpon Sus Seti, Renk2 Plastik Masa Ortusu, Renk2 Renkli Latex Parti Balonlari 10 Adet]	6	12
[Renk1 Renkli Latex Parti Balonlari 10 Adet, Renk2 Renkli Latex Parti Balonlari 10 Adet, Tema8 Kopuk Baloncuk, Tema8 Temali Bardak, Tema8 Temali Kagit Pecete, Tema8 Temali Karton Tabak]	6	11
[Renk4 Renkli Latex Parti Balonlari 10 Adet, Tema5 Happy Birthday Harf Afis, Tema5 Pecete, Tema5 Temali Bardak, Tema5 Temali Karton Tabak, Tema5 Temali Kaynana Dili]	6	10
[Renk 11 Renkli Latex Parti Balonlari 10 Adet, Renk4 Renkli Latex Parti Balonlari 10 Adet, Tema5 Happy Birthday Harf Afis, Tema5 Pecete, Tema5 Temali Bardak, Tema5 Temali Karton Tabak]	6	10
[Renk 11 Renkli Latex Parti Balonlari 10 Adet, Renk4 Renkli Latex Parti Balonlari 10 Adet, Tema5 Pecete, Tema5 Temali Bardak, Tema5 Temali Karton Tabak, Tema5 Temali Kaynana Dili]	6	10
[Tema3 Kurdelali Harf Afis, Tema3 Temali Dogum Gunu Tabagi, Tema3 Temali Dratif Kurdan Sus, Tema3 Temali Kulah Parti Sapkasi, Tema3 Temali Pecete, Tema3 Temali Plastik Bardak]	6	10
[Renk2 Renkli Latex Parti Balonlari 10 Adet, Tema8 Kurdan 10 Adet, Tema8 Temali Bardak, Tema8 Temali Bayrak Afis, Tema8 Temali Kagit Pecete, Tema8 Temali Karton Tabak]	6	10

Tablo 3.5 - 5'li Birliktelik İlk 10 Listesi

ItemSet	ItemSetSize	ItemSetSupport
[Tema4 Bayrak Afis, Tema4 Dogum Gunu Masa Ortusu, Tema4 Dogum Gunu Tabagi, Tema4 Pecete, Tema4 Plastik Bardak]	5	13
[Tema 31 Masa Ortusu, Tema 31 Pecete, Tema 31 Tabak, Tema 31 Temali Dratif Mum, Tema 31 Temali Karton Bardak]	5	12
[Renk 11 Renkli Latex Parti Balonlari 10 Adet, Renk4 Renkli Latex Parti Balonlari 10 Adet, Tema5 Happy Birthday Harf Afis, Tema5 Pecete, Tema5 Temali Bardak]	5	12
[Renk 7 Renkli Latex Parti Balonlari 10 Adet, Tema 31 Pecete, Tema 31 Resimli Dratif Kurdan Sus, Tema 31 Tabak, Tema 31 Temali Karton Bardak]	5	11
[Renk 7 Renkli Latex Parti Balonlari 10 Adet, Renk5 Plastik Catal, Tema 31 Pecete, Tema 31 Resimli Dratif Kurdan Sus, Tema 31 Temali Karton Bardak]	5	11
[Tema5 Happy Birthday Harf Afis, Tema5 Pecete, Tema5 Temali Bardak, Tema5 Temali Karton Tabak, Tema5 Temali Kaynana Dili]	5	10
[Renk4 Renkli Latex Parti Balonlari 10 Adet, Tema5 Happy Birthday Harf Afis, Tema5 Pecete, Tema5 Temali Bardak, Tema5 Temali Kaynana Dili]	5	10
[Renk4 Renkli Latex Parti Balonlari 10 Adet, Tema5 Happy Birthday Harf Afis, Tema5 Pecete, Tema5 Temali Karton Tabak, Tema5 Temali Kaynana Dili]	5	10
[Renk4 Renkli Latex Parti Balonlari 10 Adet, Tema5 Happy Birthday Harf Afis, Tema5 Temali Bardak, Tema5 Temali Karton Tabak, Tema5 Temali Kaynana Dili]	5	10
[Renk 11 Renkli Latex Parti Balonlari 10 Adet, Tema5 Happy Birthday Harf Afis, Tema5 Pecete, Tema5 Temali Bardak, Tema5 Temali Karton Tabak]	5	10

Tablo 3.6 - 4'lü Birliktelik İlk 10 Listesi

ItemSet	ItemSetSize	ItemSetSupport
[Tema 31 Cupcake Standi, Tema 31 Masa Ortusu, Tema 31 Tabak, Tema 31 Temali Karton Bardak]	4	15
[Tema7 Tema14 Temali Bardak, Tema7 Tema14 Temali Dogum Gunu Masa Ortusu, Tema7 Tema14 Temali Dogum Gunu Tabagi, Tema7 Tema14 Temali Pecete]	4	13
[Tema3 Happy Birthday Bardak, Tema3 Happy Birthday Dogum Gunu Tabagi, Tema3 Temali Happy Birthday Harf Afis, Tema3 Temali Parti Taci]	4	12
[Tema 31 Masa Ortusu, Tema 31 Tabak, Tema 31 Temali Dratif Mum, Tema 31 Temali Karton Bardak]	4	12
[Tema14 Dogum Gunu Iyi Ki Dogdun Bayrak Afis, Tema7 Tema14 Temali Dogum Gunu Tabagi, Tema7 Tema14 Temali Pecete, Tema7 Tema14 Temali Plastik Bardak]	4	11
[Tema 19 Temali Bardak, Tema 19 Temali Kagit Pecete, Tema 19 Temali Latex Balon, Tema 19 Temali Tabak]	4	11
[Tema 24 Masa Ortusu, Tema 24 Pecete, Tema 24 Tabak, Tema 24 Temali Karton Bardak]	4	11
[Asa, Eldiven Seti, Tema9 Hediye1, Tema9 Tac]	4	11
[Tema3 Happy Birthday Bardak, Tema3 Happy Birthday Dogum Gunu Tabagi, Tema3 Temali Dratif Kurdan Sus, Tema3 Temali Parti Taci]	4	10
[Tema2 Dogum Gunu Tabagi, Tema2 Karton Bardak, Tema2 Pecete, Tema2 Temali Dogum Gunu Masa Ortusu]	4	10

Tablo 3.7 - 3'lü Birliktelik İlk 10 Listesi

ItemSet	ItemSetSize	ItemSetSupport
[Tema 19 Temali Bardak, Tema 19 Temali Kagit Pecete, Tema 19 Temali Tabak]	3	14
[Tema 19 Temali Bardak, Tema 19 Temali Latex Balon, Tema 19 Temali Tabak]	3	11
[Tema 19 Temali Kagit Pecete, Tema 19 Temali Latex Balon, Tema 19 Temali Tabak]	3	11
[Tema 32 Temali Bardak, Tema 32 Temali Kagit Pecete, Tema 32 Temali Karton Tabak]	3	10
[Tema 35 Dogum Gunu Tabagi, Tema 35 Karton Bardak, Tema 35 Parti Cantasi]	3	10
[Tema 28 Masa Ortusu, Tema 28 Temali Karton Bardak, Tema 28 Temali Karton Tabak]	3	10
[Iyi Dinazor Temali Bardak, Iyi Dinazor Temali Tabak, Tema 30 Temali Kagit Pecete]	3	10
[Transformers Dogum Gunu Karton Bardak, Transformers Dogum Gunu Tabagi, Transformers Temali Pecete]	3	10
[Dratif Tema 24 Mandal Parti Malzemesi, Tema 24 Pecete, Tema 24 Tabak]	3	10
[Iyi Dinazor Temali Bardak, Tema 30 Temali Kagit Pecete, Tema 30 Temali Parti Cantasi]	3	10

Tablo 3.8 - 2'li Birliktelik İlk 10 Listesi

ItemSet	ItemSetSize	ItemSetSupport
[Tema18 Kiz Dratif Kurdan Sus, Tema18 Kiz Harf Afis - 4 Harf]	2	11
[Tema3 Dogum Gunu Harf Afis - 13 Harf, Tema3 Kopuk Baloncuk]	2	11
[6 Sekilli Rakam Folyo Balon Renk 10/Renk 11 Renkleri - Renk 10 Rengi, Tema 28 Temali Harf Afis - 5 Harf]	2	11
[Renk 10 Renkli Metalik Parti Balonu 10 Adet, Tema 23 Yapiskanli Afis]	2	11
[Transformers Dogum Gunu Tabagi, Transformers Temali Pecete]	2	10
[Tema 32 Temali Bardak, Tema 32 Temali Karton Tabak]	2	10
[Tema 32 Temali Kagit Pecete, Tema 32 Temali Karton Tabak]	2	10
[Transformers Dogum Gunu Karton Bardak, Transformers Temali Pecete]	2	10
[Tema 35 Dogum Gunu Tabagi, Tema 35 Folyo Balon]	2	10
[3 Yas Dogum Gunu Renkli Parti Balonlari, Tema12 Balon]	2	10

Bir diğ er node olan Association Rule Learner Apriori algoritmasına göre ç alıř an bir node'dur. Itemcolumn alanında birliktelik kurallarının bulunabileceđ i sütün seç ilir. Minimum set size ve minimum support parametreleri eklenip execute edilebilir. Execute iř lemi verinin b y u k l u ğ u ne g o re biraz uzun s u re b i l i y o r. Execute t a m a m l a n d ı ğ ı n d a node'un u z e r i n d e f a r e i l e sađ t ı k l a y ı p a s s o c i a t i o n r u l e s s e ç e n e ğ i s e ç i l e r e k o l u ř a n v e r i g o r u n t u l e n e b i l i r. B u n o d e ' u n ç ı k t ı s ı g o r u n t u l e n d i ğ i n d e b u s e f e r d a h a f a z l a s u t u n u o l a n b i r t a b l o o l u ř m a k t a d ı r. B u t a b l o d a k i v e r i l e r i n a n l a m l a r ı ř u Ő e k i l d e d i r; c o n s e q u e n t a l a n ı n d a k i u r u n l e r i n, a n t e c e d e n t a l a n ı n d a k i u r u n l e r i l e b i r l i k t e l i k k u r a l l a r ı k a r ř ı l a ř t ı r ı l m a k t a d ı r. İ t e m s e t s u p p o r t, o u r u n l e r i n k a ç d e f a b i r l i k t e s a t ı l d ı ğ ı y a n ı c o n s e q u e n t a l a n ı n d a k i t e k i l u r u n u n A n t e c e d e n t a l a n ı n d a k i u r u n s e t i y l e b i r l i k t e s a t ı l m a a d e d i d i r. B u a l a n l a r i ç e r i s i n d e c o n f i d e n c e v e s u p p o r t d e ğ e r l e r i n i n y a n ı s ı r a l i f t d e ğ e r l e r i

de önemli bir değerdir. Rule confidence alanında ürünlerin kendi aralarındaki güven yüzdesini verir.

Association Rules - 2:117 - Association Rule Learner (Borgelt) (DESTEK VE GUVEN)

File Hilite Navigation View

Table "default" - Rows: 367179 Spec - Columns: 11 Properties Flow Variables

Row ID	Consequent	Antecedent	ItemSe...	D Relativ...	D RuleCo...	D Absolut...	D Relativ...	D RuleLift	D RuleLift%	D Absolut...	D Relativ...
Row0	Renk 9 Plastik Bical	Tema 10 Baskili Lavanta Kes...	3	0.042	13	23	0.323	309.3	30,930	3	0.042
Row1	Renk 9 Plastik Bical	[Renk 9 Plastik Catal,Tema11...	1	0.014	12.5	8	0.112	296.42	29,642	3	0.042
Row2	Tema15 Formula ...	[Tema14 Happy Birthday Har...	1	0.014	12.5	8	0.112	296.42	29,642	3	0.042
Row3	Tema15 Formula ...	[Tema14 Happy Birthday Har...	1	0.014	16.7	6	0.084	395.22	39,522	3	0.042
Row4	Tema15 Formula ...	[Tema14 Happy Birthday Har...	1	0.014	14.3	7	0.098	338.76	33,876	3	0.042
Row5	Tema15 Formula ...	[Tema14 Happy Birthday Har...	1	0.014	11.1	9	0.127	263.48	26,348	3	0.042
Row6	Tema15 Formula ...	[Tema14 Happy Birthday Har...	1	0.014	14.3	7	0.098	338.76	33,876	3	0.042
Row7	Tema15 Formula ...	[Tema14 Happy Birthday Har...	1	0.014	12.5	8	0.112	296.42	29,642	3	0.042
Row8	Tema15 Formula ...	[Tema14 Koyuk Baloncuk,Te...	1	0.014	20	5	0.07	474.27	47,427	3	0.042
Row9	Tema15 Formula ...	[Tema14 Koyuk Baloncuk,Te...	1	0.014	14.3	7	0.098	338.76	33,876	3	0.042

Şekil 3.20 - KNIME Association Rule Learner Output View

Rule Lift (Kaldıraç): Birliktelik kuralları algoritmaları bu birliktelikleri belirlerken genellikle destek ve güven değerlerini kullanarak bir sonuç üretir. Ancak güven ve destek değerleri yüksek olduğu durumlarda kaldıraç değeri ön plana çıkmaktadır. Kaldıraç değeri güven değerinin destek değerine bölünmesiyle oluşan bir katsayı değeridir. Aşağıdaki formüller ifade edilebilir [31].

$$\text{Kaldıraç } (A \rightarrow B) = \frac{\text{Güven}(A \rightarrow B)}{\text{Destek } (B)} \quad (5)$$

Dialog - 2:117 - Association Rule Learner (Borgelt) (DESTEK VE...

File

Options Advanced Settings Flow Variables Memory Policy

Item column: (...) Concatenate(Name)_SplitResultSet

Item set settings

Minimum set size: 1

Minimum support: 5,0

Absolute number Percentage

Minimum rule confidence 10,0 Sort antecedent list

OK Apply Cancel ?

Şekil 3.21 - KNIME Association Rule Learner Node Configure

En fazla satılan ürünlere göre birliktelikler incelendiğinde ilk 10 ürün listesi Tablo 3.9'daki gibidir. Örneğin bu tablodaki ilk satırdaki veriler incelendiğinde şu şekilde yorumlanabilir; “Renk1 Renkli Latex Parti Balonlari 10 Adet” ürününü alanların sepetinde “Renk2 Renkli Latex Parti Balonlari 10 Adet” ürünün olma olasılığı %63,5 tir. Ayrıca rulelift katsayısı da 97,643’tür.

Tablo 3.9 - En Fazla Satılan Ürüne Göre Destek ve Güven İlk 10 Listesi

Consequent	Antecedent	ItemSetSupport	RuleConfidence%	RuleLift
Renk1 Renkli Latex Parti Balonlari 10 Adet	[Renk2 Renkli Latex Parti Balonlari 10 Adet]	265	63,5	97,64
Renk2 Renkli Latex Parti Balonlari 10 Adet	[Renk1 Renkli Latex Parti Balonlari 10 Adet]	265	57,2	97,64
Tema3 Temali Pecete	[Tema3 Temali Dogum Gunu Tabagi]	208	59,8	92,03
Tema3 Temali Dogum Gunu Tabagi	[Tema3 Temali Pecete]	208	45,0	92,03
Tema1 Konsepti Dogum Gunu Tabagi	[Tema1 Konsepti Karton Bardak]	199	71,1	17,55
Tema1 Konsepti Karton Bardak	[Tema1 Konsepti Dogum Gunu Tabagi]	199	69,1	17,55
Tema1 Temali Pecete	[Tema1 Konsepti Dogum Gunu Tabagi]	181	62,8	12,59
Tema1 Konsepti Dogum Gunu Tabagi	[Tema1 Temali Pecete]	181	51,0	12,59
Tema1 Temali Pecete	[Tema1 Konsepti Karton Bardak]	169	60,4	12,09
Tema1 Konsepti Karton Bardak	[Tema1 Temali Pecete]	169	47,6	12,09

Güven katsayısına göre ilk 10 ürün listesi çıkartıldığında güven değeri 100 olan birliktelikler bulunduğu görülmektedir. Güven değeri 100 olan ilk satır incelendiğinde; “Renk2 Plastik Masa Ortusu, Plastik Masa Etegi Renk1 Renkli, Renk2 Renkli Latex Parti

Balonlari 10 Adet” ürünlerini alanların “Renk1 Renkli Latex Parti Balonlari 10 Adet” ürünü alma olasılığı %100 dür.

Tablo 3.10 - En Fazla Güven Değerine Göre İlk 10

Consequent	Antecedent	ItemSetSupport	RuleConfidence%	RuleLift
Renk1 Renkli Latex Parti Balonlari 10 Adet	[Renk2 Plastik Masa Ortusu, Plastik Masa Etegi Renk1 Renkli, Renk2 Renkli Latex Parti Balonlari 10 Adet]	34	100,0	15,36
Tema3 Temali Pecete	[Tema3 Temali Karton Bardak, Tema3 Temali Parti Taci, Tema3 Temali Dogum Gunu Tabagi]	29	100,0	15,39
Asa	[Tema9 Tac, Eldiven Seti]	28	100,0	254,07
Eldiven Seti	[Tema9 Tac, Asa]	28	100,0	254,07
Tema9 Tac	[Eldiven Seti, Asa]	28	100,0	254,07
Eldiven Seti	[Tema9 Tac]	28	100,0	254,07
Tema9 Tac	[Eldiven Seti]	28	100,0	254,07
Asa	[Tema9 Tac]	28	100,0	254,07
Tema9 Tac	[Asa]	28	100,0	254,07
Asa	[Eldiven Seti]	28	100,0	254,07

Tüm veriyi herhangi bir birliktelik kuralı gözetmeksizin tekil olarak satış rakamlarına bakıldığında en fazla satılan ürünler Tablo 3.11’deki gibi olmuştur. Bu ürünleri birliktelik kurallarında çıkan sonuçlara göre değerlendirildiğinde birçok ürünle birlikte satıldığı görülmektedir. Tek başına olan satış değerleri tek yönlü bir bilgi verirken birliktelik kuralları uygulanarak çıkan birliktelikler çok yönlü bilgi vermektedir.

Tablo 3.11 - En Fazla Satılan İlk 10 Ürün Listesi

İlk 10 Ürün Satış Listesi		
Name	ItemSetSize	ItemSetSupport
Renk1 Renkli Latex Parti Balonlari 10 Adet	1	463
Tema3 Temali Pecete	1	462
Renk2 Renkli Latex Parti Balonlari 10 Adet	1	417
Renk 7 Renkli Latex Parti Balonlari 10 Adet	1	369
Tema1 Temali Pecete	1	355
Tema3 Temali Dogum Gunu Tabagi	1	348
Tema3 Temali Dogum Gunu Masa Ortusu	1	313
Plastik Masa Etegi Renk1 Renkli	1	310
Tema1 Konsepti Dogum Gunu Tabagi	1	288
Renk2 Plastik Masa Ortusu	1	282

BÖLÜM 4. BULGULAR VE YORUMLAR

4.1. Bulgular

Bir E-Ticaret sitesinin satış verileri üzerinden yapılan birliktelik kuralları analizinde öncelikle birliktelik kümeleri araştırılmış, sonrasında ise ürünlerin kendi aralarındaki destek ve güven birliktelikleri incelenmiştir. Yapılan inceleme neticesinde müşterilerin birlikte almayı tercih ettiği ürün kümeleri en fazla 6 üründen oluştuğu tespit edilmiştir. Bu ürünlerin bir kısmı en fazla satılan tekil ürünler arasında olduğu görülmektedir. Bu analiz neticesinde sektöre aşağıdaki önerilerde bulunulabilir;

- Birliktelik kurallarına göre oluşan 6'lı, 5'li, 4'lü, 3'lü veya 2'li ürün setleri için müşterilere bir kampanya yapıp siteye reklam olarak eklenebilir.
- Birlikte alınan ürün setleri arasında en ucuz olan ürünü hediye olarak promosyon veya kampanya yapılabilir.
- En az satılan ürünler ile en fazla birlikte satılan ürün setleri arasında bir kampanya sepeti oluşturularak en az satılan ürünlerin satışı artırılabilir.
- Siteye en fazla birlikte tercih edilen ürünlerin kataloglarını daha yukarıya ekleyerek müşterinin bu ürünlere ulaşması kolaylaştırılabilir.
- Birlikte alınan ürün set grubundan 2 veya daha fazla ürün alan müşteriye site üzerinde diğer tercih edilen ürünler sunulabilir.
- Tekil satış olarak ilk 10 listesinde olup da, birlikte satılanlar arasında olmayan ürünler için farklı birliktelikler oluşturulup bir kampanya yapılabilir.

Destek ve güven birlikteliği yapılan ürünlerde hangi ürünler için promosyon yapabileceği konusunda fikir verebilir sonuçlara ulaşılmıştır. Örneğin güven değeri %100 olan ürün ve ürün setleri olduğu görülmüştür.

4.2. Yorumlar

Birliktelik kuralları neticesinde ıkartılan analiz doęrultusunda ilgili sektr aısından faydalı olabilecek birtakım bulgu ve sonulara ulaşılmıřtır. Bu sonular doęrultusunda e-ticaret sitesi zerinde birtakım aksiyonlar nerilerek, sektre karlılık ve mřteri memnuniyeti aısından faydalı sonular getireceęi sonucuna ulaşılmıřtır. Aynı zamanda yapılan analiz sonularına gre hem karlılık artacaktır hem de site zerinde aranan rnlerin daha kolay bulunabileceęi bir yapıya geilerek mřteri memnuniyeti artırılmıř olacaktır.



BÖLÜM 5. SONUÇ

5.1. Özet

Bu çalışma E-ticaret alanında hizmet veren bir sektörün satış verileri üzerinden gerçekleştirilmiştir. Satış verileri üzerinden birliktelik kuralları incelenmiştir. Birliktelik kuralları ve analizi çıkartılan sektör, bu analiz sonuçları doğrultusunda satış stratejilerini farklılaştırabilirler. Aynı zamanda bu çalışma ile müşterilerinin alışkanlık ve davranışları da incelenerek hangi ürünlere daha çok odaklanıldığı konusunda fikir veren sonuçlar doğmuştur. Aynı zamanda matematiksel ve standart yöntemlerle belirlenemeyen sonuçlar Apriori Algoritması kullanılarak bir takım tespitler elde edilmiştir.

Bu çalışmada yapılan birliktelik analizi, kullanılan verilerin temizlenip, düzenlendikten sonra Apriori Algoritması kullanılarak gerçekleştirilmiştir. Analiz için KNIME uygulaması kullanılmış olup kullanılan verinin gelişimi KNIME uygulaması içerisinde node'lar eklenerek adım adım izlenmiştir.

Birliktelik kuralları neticesinde çıkan analiz doğrultusunda ilgili sektör açısından faydalı olabilecek birtakım bulgu ve sonuçlara ulaşılmıştır. Bu sonuçlar doğrultusunda yapılan çalışmanın sektöre karlılık ve müşteri memnuniyeti açısından faydalı sonuçlar getireceği ön görülmüştür.

5.2. Yargı

İncelenen verilerin analiziyle bir takım önemli sonuçlara ulaşılmıştır. Bu analiz sonuçlarına göre kullanılan veride birlikte satılan en fazla 6'lı ürün kümesi olduğu gözlenmiştir. Bunların yanı sıra destek ve güven değerlerine göre de analizler yapılmış olup güven değeri %100 e kadar ulaşan ürün birliktelikleri olduğu yargısına ulaşılmıştır. Yani %100 güven katsayısına sahip ürün setleri şimdiye kadar hep bir arada alınmış olduğu anlaşılmaktadır. Bu ürünlerin yanında az satılan ürünleri de ekleyerek kampanya veya promosyon yapılabilir. Böylelikle en az satılan ürünleri en çok satılan ürün setlerine yerleştirerek satışlarının artırılacağı öngörülmektedir. Birlikte tercih edilen ürünler için belli kampanyalar yapılabilir. Örneğin 5'li seti alana 6. 7. Ürün indirimli gibi promosyonlar da önerilebilir. Birlikte tercih edilen ürün setleri arasında en ucuz olan ürün hediye olarak da önerilebilir. Site üzerindeki ürün kategorilerinin birlikte tercih edilen ürünlere daha kolay erişilebilir olacak şekilde düzenleme yapılması bu ürünlerin satışını artırabilir bir etken olabilir. Birlikte tercih edilen ürün setlerinden bazı ları sepete

eklendiğinde birliktelik kuralına uyan dięer ürünler müşteriye önerilerek satışlar artırılabilir yargısına varılmıştır.

5.3. Öneriler

Bu çalışma kapsamında yapılan analiz ürün bazında yapıldı ancak ileride bu analizi farklı bir çalışmada bu ürünleri kategorize ederek kategori bazlı bir çalışma yapılabilir. Bu şekilde kategori bazlı birliktelikler ve müşteri alışkanlıkları tespit edilerek kategori bazında analizler yapılabilir.

Aynı zamanda bu analizlere deęişik parametreler eklenerek de yapılabilir. Örneğin özel gün, milli bayram, dini bayramlar, yılbaşı vb. günlerde yapılan satışlar ve birliktelikler için ayrı analizler yapılarak özel günlere göre farklı stratejiler de belirlenebilir.

KAYNAKÇA

- [1] M. Yılmaz, "Enformasyon ve bilgi kavramları bağlamında enformasyon yönetimi ve bilgi yönetimi," *DTCF Dergisi*, vol. 49, 2017.
- [2] K. Köseoğlu, *Veritabanı mantığı*: Pusula Yayıncılık, 2007.
- [3] M. Golfarelli and S. Rizzi, *Data warehouse design: Modern principles and methodologies* vol. 5: McGraw-Hill New York, 2009.
- [4] T. B. Pedersen and C. S. Jensen, "Multidimensional database technology," *Computer*, vol. 34, pp. 40-46, 2001.
- [5] S. H. A. El-Sappagh, A. M. A. Hendawi, and A. H. El Bastawissy, "A proposed model for data warehouse ETL processes," *Journal of King Saud University-Computer and Information Sciences*, vol. 23, pp. 91-104, 2011.
- [6] K. G. Neha, "Efficient Information Retrieval Using Multidimensional Olap Cube," 2017.
- [7] A. Matei, K.-M. Chao, and N. Godwin, "OLAP for multidimensional semantic web databases," in *Enabling Real-Time Business Intelligence*, ed: Springer, 2015, pp. 81-96.
- [8] Url 1: <https://intellipaat.com/tutorial/msbi-tutorial/ssas/> , Erişim Tarihi : 2018 Mayıs
- [9] J. Han, "OLAP mining: An integration of OLAP with data mining," in *Proceedings of the 7th IFIP*, 1997, pp. 1-9.
- [10] E. Quam. (2000) Dublin Core Metadata. *Minnesota Metadata Guidelines For Dublin Core Metadata*. 2.
- [11] D. T. Larose, *Data mining methods & models*: John Wiley & Sons, 2006.
- [12] R. Lior, *Data mining with decision trees: theory and applications* vol. 81: World scientific, 2014.
- [13] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.
- [14] Ö. Akgöbek and F. Çakır, "Veri Madenciliğinde Bir Uzman Sistem Tasarımı," *Akademik Bilişim*, vol. 9, pp. 801-806, 2009.
- [15] C. Özcan, "Veri madenciliğinin güvenlik uygulama alanları ve veri madenciliği ile sahtekarlık analizi," Yüksek Lisans Tezi, İstanbul Bilgi Üniversitesi, 2014.
- [16] S. Aydın and A. E. Özkul, "Veri madenciliği ve anadolu üniversitesi açıköğretim sisteminde bir uygulama," *Journal of Research in Education and Teaching*, vol. 4, pp. 36-44, 2015.
- [17] G. Dalkılıç and F. Türkmen, "Karınca Kolonisi Optimizasyonu," *Yüksek performanslı bilişim sempozyumu, Gebze*, 2002.
- [18] H. Akpınar, "Veri tabanlarında bilgi keşfi ve veri madenciliği," *İÜ İşletme Fakültesi Dergisi*, vol. 29, pp. 1-22, 2000.
- [19] T.-r. Li and W.-m. Pan, "Intrusion detection system based on new association rule mining model," in *Granular Computing, 2005 IEEE International Conference on*, 2005, pp. 512-515.
- [20] F. Kurniawan, B. Umayah, J. Hammad, S. M. S. Nugroho, and M. Hariadi, "Market Basket Analysis to Identify Customer Behaviours by Way of Transaction Data," *Knowledge Engineering and Data Science*, vol. 1, pp. 20-25, 2017.

- [21] M. Kaur and S. Kang, "Market Basket Analysis: Identify the changing trends of market data using association rule mining," *Procedia computer science*, vol. 85, pp. 78-85, 2016.
- [22] A. Mansur and T. Kuncoro, "Product inventory predictions at small medium enterprise using market basket analysis approach-neural networks," *Procedia Economics and Finance*, vol. 4, pp. 312-320, 2012.
- [23] M. Özçalıcı, "Veri Madenciliğinde Birlikte Kuralları ve İkinci El Otomobil Piyasası Üzerine Bir Uygulama," *Sosyal Bilimler Arastirmalari Dergisi*, vol. 7, 2017.
- [24] M. H. Dunham, Y. Xiao, L. Gruenwald, and Z. Hossain, "A survey of association rules," *Retrieved January*, vol. 5, p. 2008, 2001.
- [25] M. Hegland, "The apriori algorithm—a tutorial," in *Mathematics and computation in imaging science and information processing*, ed: World Scientific, 2007, pp. 209-262.
- [26] B. S. Kumar and K. Rukmani, "Implementation of web usage mining using APRIORI and FP growth algorithms," *Int. J. of Advanced networking and Applications*, vol. 1, pp. 400-404, 2010.
- [27] M. Shah, N. Shah, A. Shetty, D. Shah, and P. Gotmare, "A Comparative Study of Pattern Recognition Algorithms on Sales Data," *International Journal of Computer Applications*, vol. 141, 2016.
- [28] S. Kotsiantis and D. Kanellopoulos, "Association rules mining: A recent overview," *GESTS International Transactions on Computer Science and Engineering*, vol. 32, pp. 71-82, 2006.
- [29] M. Kavitha and S. T. Selvi, "Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons," *International Journal of Computer Science Trends and Technology (IJCS T)–Volume*, vol. 4.
- [30] Url 2: <https://www.knime.com/knime-analytics-platform>, Erişim Tarihi : 2018 Mayıs
- [31] A. Bekki, *Veri Madenciliği: Anadolu Üniversitesi Basımevi*, 2006.

