



T.C.
ALTINBAŞ ÜNİVERSİTESİ
Lisansüstü Eğitim Enstitüsü
Endüstri Mühendisliği

SEZGİSEL YÖNTEMLERLE MARKET SEPET ANALİZİ

Fuat GENÇ

Yüksek Lisans Tezi

Danışman: Prof. Dr. Osman Nuri UÇAN

İSTANBUL, 2020

SEZGİSEL YÖNTEMLERLE MARKET SEPET ANALİZİ

FUAT GENÇ

Endüstri Mühendisliği

Lisansüstü Eğitim Enstitüsü'ne

Yüksek Lisans

Tezi olarak sunulmuştur.

İstanbul, 2020

Fuat GENÇ tarafından hazırlanmış ve 21.08.2020 tarihinde sunulmuş “ SEZGİSEL YÖNTEMLERLE
MARKET SEPET ANALİZİ” başlıklı tez Endüstri Mühendisliği Anabilim Dalında Yüksek Lisans Tezi olarak
oy birliği ile kabul edilmiştir.

Prof. Dr. Osman Nuri UÇAN
Danışman

Tez Savunma Sınavı Jüri Üyeleri:

Prof. Dr. Osman Nuri UÇAN
Mühendislik ve Doğa
Bilimleri,
Altınbaş Üniversitesi

Doç. Dr. Niyazi Onur BAKIR
Mühendislik ve Doğa
Bilimleri,
Altınbaş Üniversitesi

Dr. Öğr. Üyesi Kemal Dinçer
DİNGEÇ
Mühendislik Fakültesi, Gebze
Teknik Üniversitesi

Bu tezin Yüksek Lisans Tezi olarak bütün şartları sağladığını beyan ederim.

Lisansüstü Eğitim Enstitüsü kabul tarihi:
....../....../....

Bu dökümandaki tüm bilgilerin akademik kural ve etiğe bağılı kalınarak yazıldığını ve tez yazım kuralları kapsamında bu çalışmada bulunan ve özgün olmayan bütün bilgi ve materyallerin referanslandırıldığını temin ederim.

Fuat GENÇ



ÖNSÖZ

Yüksek lisans öğrenimim sırasında ve tez çalışmalarım boyunca gösterdiği her türlü destek ve yardımdan dolayı çok değerli hocam ***Prof. Dr. Osman Nuri UÇAN'a*** en içten dileklerle teşekkür ederim.

Tezimin hazırlık ve yapım aşamaları boyunca yardımlarını esirgemeyen sevgili eşim ***Çiğdem GENÇ'e*** teşekkürü borç bilirim.



ÖZET

SEZGİSEL YÖNTEMLERLE MARKET SEPET ANALİZİ

GENÇ, Fuat

Yüksek Lisans, Endüstri Mühendisliği, Altınbaş Üniversitesi,

Danışman: Prof. Dr. Osman Nuri UÇAN

Tarih: Ağustos, 2020

Sayfa Sayısı: 69

Bu çalışmada yapay zekanın temel çalışmalarından ve uygulamalarından olan Birliktelik Kuralları, Market Sepet Analizi, Apriori Algoritması konularının teorik açıklamasının yanı sıra Weka programı özelindeki teori ve uygulamaları gerçekleştirilmiştir. Bununla birlikte; araştırma özeline mevcut şekilde Weka programında uygulaması yapılarak değerlendirilmesi gerçekleştirilmiştir. Büyük boyutlardaki veri, verilerin analiz edilme yöntem ve araçları, zengin veriler olarak belirtilmiştir. Toplanan datalar büyük veritabanlarının "Veri Yığınları" haline dönüşmesi; ender görülen bir olay ve olgu halini almıştır. Veritabanları sadece yalnız olmakla birlikte, ayrıca yönlendirme hissi çalışması uygulamaktadır. Bunun nedeni ise hüküm vericinin data araçlarının dahil edilmediği birçok sayıda büyük verinin içindeki önemli verileri nitelemektedir. Ek olarak, şu anki yetkin sistem teknolojilerini manuel olarak çözümlemek için genellikle alanında yetkin kullanıcılara güvenilmesi gerekmektedir. Oldukça sık olan bu öğeleri bulma, data mining'in en çok üzerinde çalışılan alanlarından biri olma yolunda hızla ilerlemektedir. Apriori algoritması ise (FIM)-sık kullanılan veri madenciliği için en çok kullanılan algoritmadır. Apriori algoritmasının çeşitli uygulamaları rapor edilmiş ve değerlendirilmeye alınmıştır. Data yapısının özellikle Bodon ile bir trie sayesinde maksimum verimli hale getiren çalışmalardan biri olarak önem arz etmektedir. Bodon'un oldukça sık bir şekilde öğe bulma çalışmalarının sonucu olarak, Borgelt ve Goethals'ın çalışma sonuçlarından daha hızlı olduğu saptanmıştır. Özellikle yapılan çalışmada, girdi sağlanan işlemlerin eşit bir PC(personal computer) tarafından incelenen paralel bir modele Bodon çalışmaları düzenlenerek aranje edildi. Revize ile değiştirilmiş uygulamaya eş ve paralel PC(personal computer) etkisi değerlendirilmeye alınmıştır. Özellikle son zamanlarda firmalar çok fazla veriye sahiptir, ancak bu verilerden işlenen ve verim alma noktasında

eksik kalmaktadırlar. Big Data'lar değerli bir veri özü olarak görülmekte ve data mining konusu daha yeni yeni gün yüzüne çıkmasıyla birlikte, farklı üretim alanlarında çalışan firmaların stratejik karar alma noktasında güven duymaktadırlar. Buna zıt bir düşünce olarak; görünmeyen veya gözardı edilen realiteler, saklanan dataları ivedilikle süzen tekniklerle ortaya koyulabilir. Pazar sepeti analizi, consumer alışverişin bütün unsurlarını saptamak adına en uygun tekniklerden birisi olarak görünmektedir. Bunun gibi informasyonlar; stock kontrolü, promosyon desteği ve cross sale fırsatları gibi pazarlama çalışmaları içeren kararlar için odak noktası anlamında ele alınabilir. Bu çalışmadaki temel amaç ve kapsam; bir süpermarketin ürün gamındaki farklı malzemelerin birbiriyle olan ilişkisinin nasıl olduğu ve aynı zamanda satış pazarlama faaliyetleri aracılığıyla bu ilişkileri nasıl uygulayacaklarını görmektir. Uygulama ve yöntem dataları ile ürünlerin eş güdümlü olarak match olması ve aynı zamanlı müşteri ürün satın alımları esasında temel ipuçları ve bilgiler verecektir. Bu ve buna benzer bilgiler, promosyon verisi, stock kontrolü ve cross sale fırsatları gibi satış faaliyetleri ile ilgili kararlarında odak noktası olarak alınabilir.

Anahtar Kelimeler: Sezgisel Yöntemler, Basket sepet analizi, Weka.

ABSTRACT

MARKET BASKET ANALYSIS WITH HEURISTIC METHODS

GENC, Fuat,

M.Sc., Industrial Engineering, Altınbaş University,

Supervisor: Prof. Dr. Osman Nuri UÇAN

Date: August, 2020

Pages: 69

In this study, besides the theoretical explanation of Association Rules, Market Basket Analysis, Apriori Algorithm, which is one of the basic studies and applications of artificial intelligence, a theory study within the Weka program was conducted. In addition; In accordance with the theory, the application was made in the Weka program and evaluated. Large amounts of data, methods and means of analyzing data are specified as rich data. The data collected will turn large databases into "Data Stacks"; it has become a rare event and phenomenon. Databases are working not only alone but also the feeling of a decision maker. The reason for this is that it characterizes valuable data in many big data that fixity center not included data tools. In addition, in order to manually analyze current competent system technologies, it is often necessary to rely on competent users in their field. Finding these items, which are very common, is rapidly progressing to become the best one studied places of data mining. Apriori algorithm(FIM) is wided used algorithm for frequently used data mining. Various applications of the Apriori algorithm have been reported and evaluated. It is important as one of the studies that makes the data structure maximum efficient especially with a triad with Bodon. As a result of Bodon's efforts to find an item in a fairly frequent figure, it was found that Borgelt and Goethals were faster than the study results. In particular, in the study, Bodon studies were arranged and arranged in a parallel model, where the inputs provided were examined by an equal PC (personal computer). The effect of peer and parallel PC has been taken into consideration. Especially recently, companies have a lot of data, but they are lacking at the point where they are processed from this data and yield efficiency. Big Data are considered as a valuable data essence and with the fact that data mining has just emerged, they are confident in the strategic decision-making point of companies working in

different production areas. As a contrary thought; invisible or ignored realities can be brought to light with techniques that filter the stored data immediately. Market basket analysis appears to be one of the most suitable techniques to detect all the elements found in consumer shopping carts. Information such as this; It can be considered as the focal point for decisions that involve marketing studies such as stock control, promotional support and cross sale opportunities. The main purpose and scope in this study; is to see how the different materials in a supermarket's product range relate to each other and also how to apply these relationships through sales and marketing activities. With the application and method data, the products will be coordinated match and simultaneous customer will provide basic tips and information based on product purchases. This and similar information can be taken as a focal point in sales-related decisions such as promotional data, stock control and cross-sale opportunities.

Keywords: Heuristic Methods, Basket - basket analysis, Weka.

İÇİNDEKİLER

	<u>Pages</u>
ÖZET	vi
ABSTRACT... ..	viii
TABLO LİSTESİ.....	xiii
ŞEKİL LİSTESİ	xiii
SEMBOL LİSTESİ	xv
1. GİRİŞ.....	1
2. GENEL KISIMLAR.....	3
2.1. BİRLİKTELİK KURALLARI	3
2.1.1. Araç ve Metotlar	9
2.1.2. Birliktelik Kuralı Genel Değerlendirilmesi	17
2.2. APRIORI ALGORİTMASI.....	17
2.2.1. Apriori Algoritması ve Ayrıntıları.....	18
2.2.2. Apriori Algoritması Uygulamaları	20
2.2.3. Örnek Apriori Yazılım Kodları ve Uygulama Örneği.....	20
2.2.4. Çevik Apriori Uygulaması	25
2.2.5. Eş Uyumlu Uygulamanın Hesaplanması.....	26
2.3. MARKET SEPET ANALİZİ	31
2.3.1. Market Basket Analizinin Güçlü ve Zayıf Yönleri	32
2.3.2. Market Sepet Analizi Data Araştırmaları	34
3. MALZEME VE YÖNTEM.....	41
3.1. WEKA YAZILIMI	41
3.1.1. Data Önışleme	42
3.1.2. Kayıp Datalar.....	42
3.1.2.1. Yanlış veya aşırı datalar	43
3.1.2.2. Gereksiz datalar	44
3.1.2.3. Değişken değerlerinin bağımsız olduğu durumlar	45
3.1.2.4. Kullanılacak Aalgoritmaya bağlı yapılandırmalar	46
3.1.3. Sınıflandırma	47
3.1.4. Kümeleme.....	50
3.1.5. Birliktelik Kuralları	50

3.2. DATA SETİNİN TANITILMASI.....	52
3.3. DATANIN HAZIRLANMASI.....	55
4. BULGULAR	58
4.1. WEKA EXPLORER MODÜLÜ SONUÇLARI	58
5. TARTIŞMA VE SONUÇ	65
REFERANSLAR	66
EK A	68
ÖZGEÇMİŞ	69



TABLO LİSTESİ

Sayfa

Tablo 2.1. İşlem Dataları	11
Tablo 2.2. Karma Ürün Kombinasyonları	12
Tablo 2.3. Ürünler Kombinasyonu(Satın Alınmış)	12
Tablo 2.4. 2’li Dataya Dönüşüm (İşlem Datası).....	13
Tablo 2.5. Destek ve Güven Data Ölçümlemesi	15
Tablo 2.6. Bağıntılama Sonucu Elde Edilen Uyum.....	15
Tablo 2.7. Data Mining İşlem Sıklık Örneği(a).....	18
Tablo 2.8. Data Mining İşlem Sıklık Örneği(b)	19
Tablo 2.9. Apriori Uygulaması-1	21
Tablo 2.10. Apriori Uygulaması-2.....	22
Tablo 2.11. Apriori Uygulaması-3.....	22
Tablo 2.12. Apriori Uygulaması-4	23
Tablo 2.13. Apriori Uygulaması-5.....	23
Tablo 2.14. Apriori Uygulaması-6.....	23
Tablo 2.15. Apriori Uygulaması-7.....	24
Tablo 2.16. Apriori Uygulaması-8.....	24
Tablo 2.17. Apriori Uygulaması-9.....	24
Tablo 2.18. Apriori Örnek Uygulaması-10.....	24
Tablo 2.19. Toplu Datalardan Ürün İlişki Tablosu.....	35
Tablo 2.20. Market Sepet Analizi Satın Alma Tavsiyeleri.....	37
Tablo 2.21. Birliktelik Kuralı ve Ürün Raflandırması.....	37
Tablo 2.22. Market Satınalma Tavsiyeleri	38
Tablo 2.23. Mağaza İçin Alışveriş Sepeti Önerileri	39
Tablo 2.24. Market Sepet Analizi Tavsiyeleri.....	39
Tablo 3.1. Data Setinin Tanıtılması.....	53

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Aşamalı Data Room Mimarisi.....	5
Şekil 2.2. Data Mining Adımlaması	6
Şekil 2.3. Data Set Tree	266
Şekil 2.4. Data Set Uygulama Gravite Oranları	Error! Bookmark not defined. 8
Şekil 2.5. Data Kümesi Ağacı	299
Şekil 2.6. 1 ve 2'li Unsurlu Data Kümesi Tree	30
Şekil 2.7. İlişkili Ürün Sıralaması Örneği	Error! Bookmark not defined. 7
Şekil 2.8. Satışta Etkili Ürün İlişkisi Örneği	387
Şekil 2.9. Birbiriyle İlişkili Ürünler ve Raf Yeri Belirlemesi.....	40
Şekil 3.1. Database Kaybolan Datalar	422
Şekil 3.2. Replace MissingModülünün Kullanımı	433
Şekil 3.3. Aşırı Uç Verilerin Numeric Cleaner Modülü Kullanılmadan Önce ve Sonraki Durumu.....	433
Şekil 3.4. Numeric Cleaner Modülünün Kullanımı	434
Şekil 3.5. Principal Components Uygulama Öncesi Analizi	444
Şekil 3.6. Principal Components Modülü ile Boyut Küçültme	455
Şekil 3.7. Normalize Modülünün Kullanımı	456
Şekil 3.8. Numeric To Binary Modülünün Kullanımı	466
Şekil 3.9. J48 Modülünün Kullanımı ile Elde Edilen Kurallar	477
Şekil 3.10. J48 Modülü ile Elde Edilen Sınıflandırma Doğrulukları.....	478
Şekil 3.11. Cost Sensitive Classifier Modülünün Kullanımı	489
Şekil 3.12. Cost Sensitive Classifier Modülünün Kullanımı ile Elde Edilen Sınıflandırma ..	489
Şekil 3.13. Simple-KMeans Uygulaması Sonrası Kümeleme Bilgileri.....	50
Şekil 3.14. Simple-KMeans Modülünün Uygulaması Sonrası Kümeleme Oranları	51
Şekil 3.15. Apriori Modülü Uygulaması Sonrası Birliktelik Kuralları.....	52
Şekil 3.16. Datanın Wordpad’de düzenlenmesi	535
Şekil 3.17. Datanın .arff formatında kaydedilmesi.....	546
Şekil 3.18. Aşırı değerlerin kontrolü	557
Şekil 3.19. Aşırı değerler	568
Şekil 4.1. Weka’da Apriori Algoritması’nın seçilmesi	589
Şekil 4.2. Apriori Algoritması Uygulaması Sonucu	60

Şekil 4.3. Filtered Associator Algoritması	61
Şekil 4.4. FilteredAssociator Algoritması Uygulaması Sonucu	62
Şekil 4.5. FPGrowth Algoritması Seçimi	63
Şekil 4.6. FPGrowth Algoritması Uygulaması Sonucu	63



SEMBOL LİSTESİ

FIM	:	Yüksek Frekanslı Unsur Madenciliği
WEKA:		Waikato Environment for Knowledge Analyses
OLAP	:	Online Analitik İşleme
FP	:	Yüksek Frekanslı Pattern Genişlemesi
KDD	:	Database Data Keşfi
TDB	:	Database Unsurlar Kümesi
GPL	:	Database Lisansı

Big data kümelemeleri birbiri ile olan ilişkileriyle, unsurlarıyla birlikte realiteye geçme olasılıklarını önceki dataları analiz edip, sunuma çıkarıp ileriye yönelik çalışmaları destekleyen, data mining yön ve yöntemi Birliktelik Kuralları olarak tanımlanmaktadır. Birliktelik Kuralları; e-ticaretten sanayiye, perakende satıştan eğitime birçok alanda aktif olarak rol alan ve alabilecek bir yöntem setidir.

Data mining metotlarında datanın nitelenmesine sağlayan niteleyici (descriptive) ve nihai hali belirlenemeyen dataların tahmini için uygulanan tahminleyici (predictive) yöntemler olarak 2'ye bölünmektedir. Regresyon (regression), bölümlendirme (classification), kümeleme(clustering) ve Birliktelik Kuralları ise niteleyici yöntemler arasında gösterilebilir.

Apriori Algoritması ise Agrawal ve Srikant'ın 1994 yılı itibariyle geliştirdiği bir algoritmadır. Birliktelik kuralı sonuç algoritmaları arasında ünü en çok olan ve kullanılandır. Algoritma adını yaygın unsurlar kümelerinin ilk bilgilendirilmesi olarak kullanılmasından mütevellî priori olarak betimlenen yani önceki anlamını içeren “Apriori” dir.

n ögeli bir öge kümesi için geniş unsur ve geniş unsur adaydır denilebilir. Apriori Algoritması geniş unsurları indirgeme durumunda görev almaktadır.

Müşterinin ürün alım esnasındaki alışveriş gamını gösteren uygulamaya ise; “Pazar Sepeti” denir. Müşteri satın alma davranışlarının irdelenmesi çoğunlukla market basket analysis olarak geçmektedir. Market sepet analysis data mining olarak geniş olarak tercih edilen bir metottur. Kullanıcıların hangi şeyi satın alma davranışında olabileceği veya aldığı bir hizmet veya ürün ile birlikte bunların dışında hangi hizmet/ürünleri satın alabileceğinin analysis edilmesi satış oranı arttıracaktır. Ayrıca; Pazar Sepeti Analizi ek satış ürünü analizi, ürün rafı düzeni, müşteri satın alma davranış analizi, ürün mağazası düzeni, envanter kontrolü, etkin ürün/hizmet satış metotlarının yenilenmesi ve gelişimi gibi durumlarda fayda sağlamaktadır.

WEKA, bir veri madenciliği ve machine öğrenmesi olarak Yeni Zelanda menşeli bir yazılımdır. WEKA programı Java dili ile yazılmış bir programdır. Nesneye yönelik programa

dillerinden olan Java yazılımı farklı öğrenme algoritmaları adına regular bir ortam şartları sağlamaktadır. WEKA'yı öne çıkaran özelliği ise 1'den fazla sınıflandırma metodunu içermesidir. Ek olarak diğer özelliği de; yazılımların komut girilerek realiteye geçmesine olanak tanınmasıdır. WEKA'nın genel user arayüzü WEKA'da; Preprocess(önişleme), Select Attribute(nitelik seçme), Associate (birliktelik kuralları), Classify(sınıflama), Cluster (kümeleme) ve Visualize (görselleştirme) bölümleri yer almaktadır.



2. GENEL KISIMLAR

2.1. BİRLİKTELİK KURALLARI

80'li yılların sonlarına doğru database teknolojisi birbiri ile global eş uyumu ile betimlendirilmiş ve kategorize edilen AR&GE(Araştırma ve Geliştirme) modelleriyle ve temel kök değişimi ile güncel ve sağlam database düzeni üzerine uygulamalarda bulunmaktadır. Bu uygulamalar ileri düzey data modelinde kullanabilir. PC(personal computer) donanımı ve sisteminin üstel olarak artmasıyla birlikte yazılım dünyasının son 30 yılda büyük, sağlam ve değerlendirilir maliyetli bilgisayarlar olmasıyla birlikte, data source toplama araçları ve stok ortamı olarak da öne taşınmıştır. Bu teknoloji sayesinde database ve ana informasyonlara yüksek oranda destek sağlar. Üretim alanında ve birçok database sisteminin yanında data uygulaması informasyon stoklaması, management informasyonu kullanımı ve data incelemesi adına öncü konuma yerleşir.

Veriler belirli bir süredir çok sayıda farklı database üzerinde saklanır pozisyonudur. Özellikle son dönemlerde ön planda olan bir database mimarisi geniş bir alan olarak ticari hale getirilen data room, çok sayıda heterojen veri source gelişimini kolaylaştırmak üzere 1 sitede birleştirilmiş olarak yönetim stratejisi konularında yön vermek üzere kullanılmaktadır. Data room kampüslerde teknolojik data cleaning(veri temizleme), data entegration ve canlı analiz uygulaması(OLAP) şeklinde işlenmektedir. Canlı analiz işlemesine ek bir şekilde; analiz metotları özetleme, farklı set birleşimi ve biraraya getirme(toplama) gibi farklı pozisyonlarda informasyonları görüntüleme yeteneğine sahiptir. Canlı analiz işlemesi data classification ve data zarflaması dahil bir çok konuda deeply analysis olarak fayda sağlamaktadır.

Fazla sayıda data, sağma data işleme programları, geniş bir data şeklinde betimlenmiştir. Genel çalışmada eksponansiyel büyüme data ve stok alanlarının genişletilmesi, sağlam data analiz araçları ile elde edilmektedir. Nihai olarak toplanan datalar big data databaselerinin "Data Yığınları" haline dönüşmesi; ender rastlanan bir olay ve olguya dönüşmüştür. En nihateyinde ise, büyük değerli amaç doğrultusunda saklanan informasyonlar geniş datalara özgü değildir. Databaseler tek başına olmamakla birlikte, yalnızca bir karar mercii hissi yani

sezgisel olarak uygular, karar vericinin sonuçlandırma araçlarına hakim olmadığı çok sayıda datanın içindeki önemli informasyonları göstermektedir. İçerik olarak ise mevcut yetkin sistem teknolojilerini manuel olarak giriş için alanında yetkin kişilere güvenmektedir. İnfomasyon tabanına bilgiyi paylaşımı; maalesef bu kurallar çerçevesinde taraflı eğilimli ve vakit eriten ve masraflı forma girmektedir. Data mining araçları datayı uygular ve analiz eder, değerli data modellerini gün yüzüne çıkarır, iş izlemleri, bilimsel ve tıbbi çalışmalar yapabilir. Datalar ile data arasında büyüyen informasyon eksikliğini, sistematik bir data tekamül çağrısı data gömülerini "value nuggets" şekline dönüştürecek madencilik enformasyon araçları formatında işlev görür.[1]

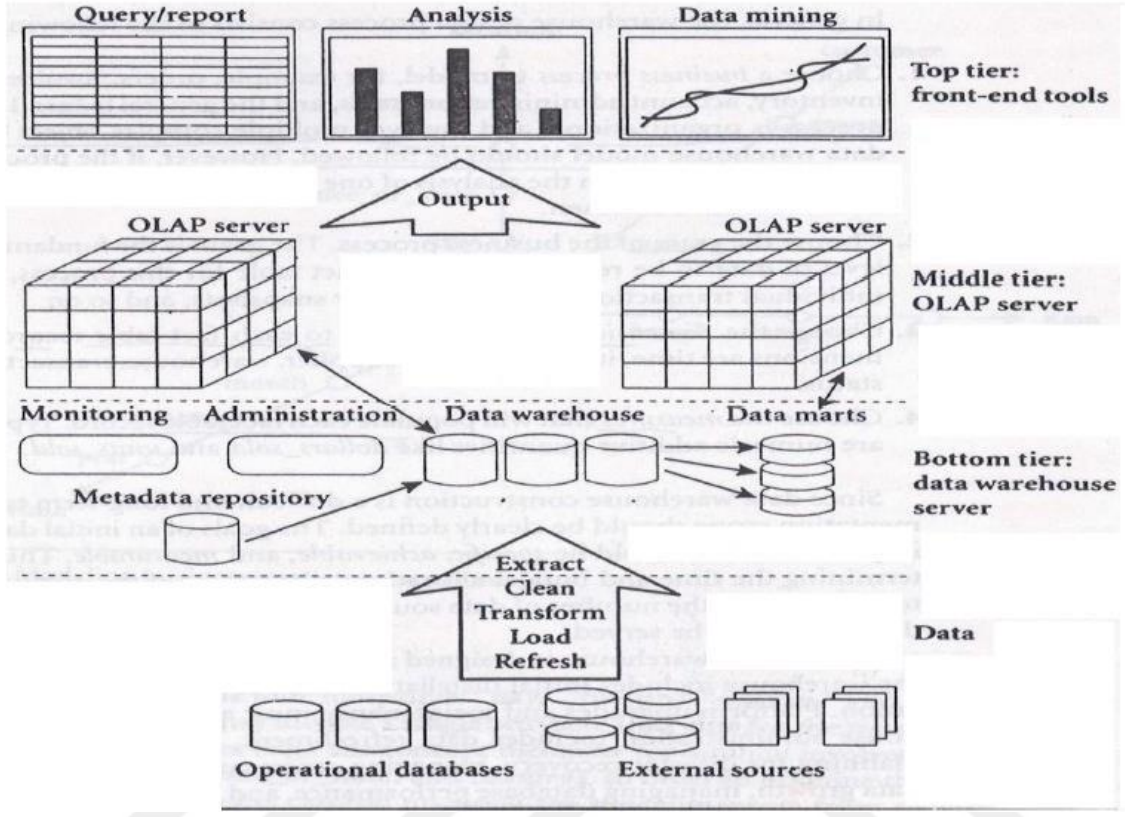
Data mining; çok fazla sayıda datayı bahsedilen altın madenine taş veya kum ikamesi olarak taş/altın madenciliği denir diğer bir deyişle kum madenciliği olarak da bahsedilir. Böylece datamining mutabık bir şekilde davet edilmiş olur ve "Information Mining Datas" olarak maalesef biraz fazla "Information Mining" kavramına nazaran daha “short format mining” fazla sayıda datanın anlamını betimler. Ayrıca, madencilikte fazlaca sensitive birimden içeren altın datalar yer almaktadır. Birbiriyle ilişkili diğer yöntemlerden aktarılan farklı çok sayıda ilkeyi bulunmaktadır. Informasyon benzeri data mining adına çeşitli anlamı olan database data mining, enformasyon eldesi, data ve desen analizi, data arkeolojisi vb. çok sayıda çalışma konusu yer almaktadır.

Data mining, prodesteki önemli bir başlangıç adımıdır. Database üzerinden informasyon keşfinin, informasyon bir proses olarak keşfi, bir proses olarak ise aşağıdaki aşamalar uygulamaya alınır:

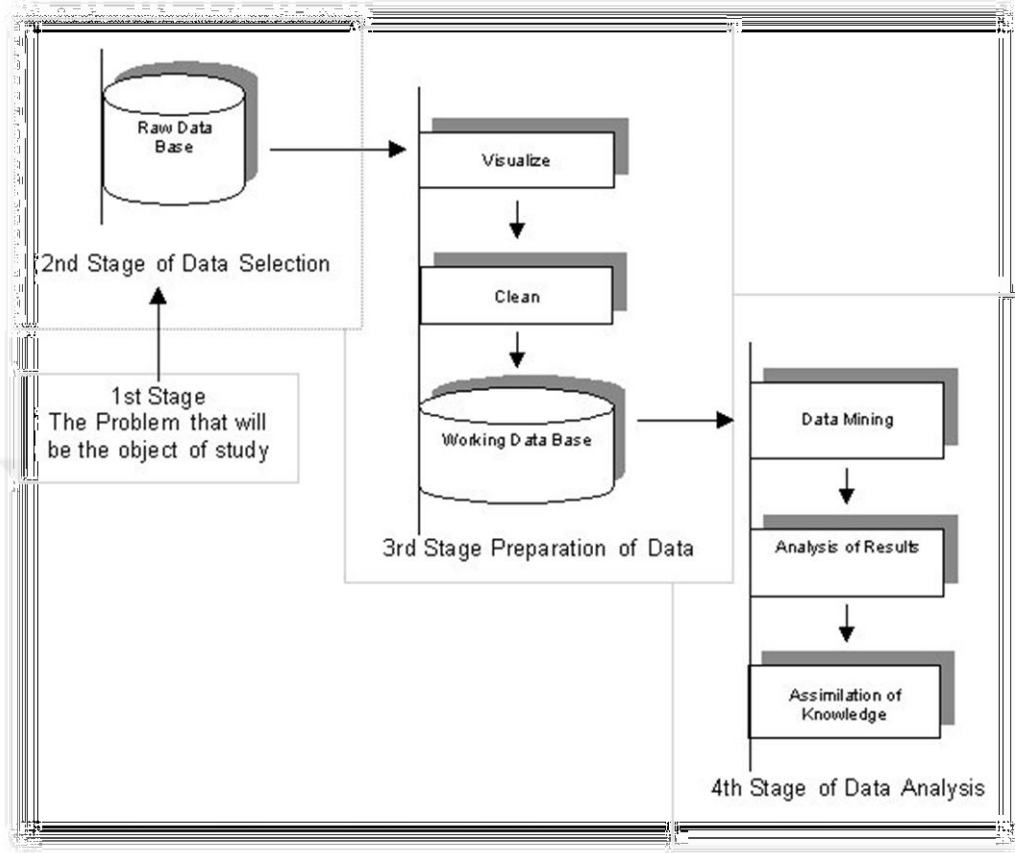
- i) Data Cleaning (Veriyi daha temiz ve kullanılabilir hale getirmek).
- ii) Data Entegrasyonu (Çok sayıda datanın kombin edilmesi)
- iii) Data Seçimi (Database özelinde data alımı ve seçimlendirilmesi).
- iv) Data Transformasyonu (Dataların değere dönüştürüldüğü veya mining uygun formatına evrilmesi)
- v) Data Mining (Üst düzey data modellemeleri için kullanılır).
- vi) Pattern Transformasyonu (Farklı pattern seçimleri ve değerlendirilmesi için)
- vii) Informasyon Sunumu (Görseller ve görsellerin hangi ilgili alanlarda olması)

gerektiğinin seçimi, görselleştirilmesi için) [2]

viii)viii)



Şekil 2.1: Aşamalı Data Room Mimarisi



Şekil 1.2: Data Mining Adımlaması

Kullanabileceğiniz farklı database kuramları vardır. Datalardan gelen informasyonları temsili olarak bulunur. Informasyon tabanı – işlem gören alan informasyonu arama kısmına iletim sağlamak veya nihai bilginin şaşırtıcılığını değerlendirmesini pattern ile sağlar. Bu ve buna benzer türdeki informasyonlar kuram üstünlüğünü nitelik / nitelik değer bazında değişik soyutlama aşamalarında userlara sunmaktadır. User yararı bazında; herhangi bir şekil ilgisini bunun özünde kullanmak için değerlendirilebilir olması ek olarak farklı şekilde dahili konumlandırılması olabilir. Diğer modeller ise alan informasyonu, ek farklılık sınırlandırılmaları veya aşamalar - meta datalardır.

Data mining motoru - bu data mining için ihtiyaçtır – ekolojisi anlamında da uygundur, uygun biçimde şunlar adına bir bakıma uygulanabilir kuramdan oluşur: Öz yapılaşyon, bağlantı, sınıflandırma, sistem incelemesi ve evülasyon - değişkenlik incelemesi. Pattern ilişkilendirme

aracı ise birleşimi farklı ölçümlendirmelerinde işlenir, datayla kuramsallaşır. Araştırma kuramlarını farklı biçimlendirmeye yönelik normları sınırlandırmak farklılık aşamaları şeklinde uygulanabilir, bulunan normlar opsiyonel seçeneleştirilir, pattern inceleme aracı, mining modülüyle eş zamanlı olabilir, data mining metotlarının hayata geçirilmesine bağlı olarak işlenmiş ve müsmir bir data mining için, pattern ilginçliğinin işleme alınması deeply bir şekilde uygulanmalıdır. Search işlemini kısıtlamak için mining uygulamasına olası farklı normlarla çalışılmaya yöneltmektedir.

Graphic user arayüzü – Bahsi geçen modül ile user ve data mining sistemi arasında user izni vermek için data mining sorgusuyla sistemle etkileşim içine girebilir veya duty olarak search odağında olmak için bilgi girdisi ve data mining sonuçlarına itafen buluş amaçlı data mining uygulaması yapılabilir. Ayrıca bu yapının user database ve data room incelemesine olanak sağlar, şekiller ya da data structure, tehlikeli yapılarla birlikte patternleri değişik formlarda sunar.

Data room bakış açısından, data mining OLAP analitiği için ilerlemiş bir kademe olarak saptanmaktadır. Ek olarak, data mining kısıtlı dataların kısaltılmış haliyle analitik incelenmesi içeriğiyle stoklama yapısının çokca ileri data bilgilendirmesine olanak sağlar.[3]

Araştırma özelindeki bir husus: "User Data Mining ulaşım davranışını belli aralıklarla kullanım patternleri" sıralı şekilde izah eder. Uygulanan database T, bağıntı sınırlandırması ve çerçevesinde $X \rightarrow Y$ itimat olarak belirlenir. Uygulama çerçevesinin oranı(yüzde olarak)T olarak; bu oran(yüzde olarak) T dayanağı ise X birleşim Y şeklinde konumlandırılır. Kompozit rule; 2 birime bölünebilir. İlk olarak fazla destek adımına göre pattern minimal dayanaklarıdır. Ayrıca, ilişki olarak rules itimat dönümü itimatın minimal itimatsal şekilde açıklanmaktadır.[4]

Parvinder S. Sandhu, çalışmalarında "Data mining yararına topluluk rules" olarak belirtiyor, kazanç ve bütünsel sayıya özgü kuvvetli şekilde çalışma yaklaşımı olarak da " Birliktelik kurma data mining aktif research paydaşı olmakla birlikte enformasyon bulgusu ve fazla sayıda yapısal düzen alanında belirtilen hedefte çalışmalar oluşturuldu. Özellikle bu çalışmada, önem kriteri faktörü ve faydasına dayalı bilinçli bir kuram olarak ilişkilendirme

kurallarının bir düzende farkedilmesinde büyük rol oynamaktadır. İlk başlarda belirtilen yaklaşım, traditional Apriori'yi tavsiye etmektedir. Algoritmanın ayrıntılı bir database normunda tavsiye edilen yaklaşım konusunda, demonoton olan şekli bir k unsur setinin fazla sık şekilde bu unsur setinin tamamını (k-1) şekilde için de bulunduğu alt kümelerde ifade eden Apriori'nin de asıl yapıtaşlarındandır. Sürekli olmak kaydıyla, uygulamanın sonraki aşamasında, birliktelik kuralı normu kompozit yüklüğü elde eder ve yararlı kısıtlamalar ve yararsız bağ rules için kompozit yararlı sonuç notlandırılır. En nihayetinde, önemli bağın bir alt segmenti konumlandırılır. Sonuçlanan UW'ye göre belirlenen kısıtlar ise deneysel nihayete tavsiye edilen etkinliğini belirtmektir. Yüksek yarar bağıntılama setleri belirlemede iş-ge için pozitif marjlı bir şekilde hayata geçirilmelidir.[5]

"2 Kıyaslama" önerisi olmadan kısıt seti data mining algoritmaları güncel "Birliktelik kuralı mining metotları data mining research alanında önemli fayda sağlayan database'deki unsurlar kümeleri arasında farklı kıyaslamalar içinde bulunmaktadır. Araştırma özelinde Matrix Apriori & FP-Büyüme konuları detaylı olarak incelenmiştir. Algoritmaları kritik eden 2 case araştırmasında 2 yapay data kümesi ele alınarak kademeli olarak hayata geçirilmiştir.

- i. Data kümesiyle fayda marjını tespit etmek için üretilen değişik özellikler,
- ii. Sebeplerinin analizi için farklı kademelerdeki fayda marj sonuçları,
- iii. Algoritmik başarımlar, veriiletilen data kümesinin özellikleri ve başlangıç noktası değeri,
- iv. Başlangıç noktası değerinin yüzden onun seyrederken Matrix Apriori, bütünsel başarımın FP-Artışından çok daha iyi başarımlar gösteriyor
- v. Apartman matrisi data structure fazla maliyet gösterir, daha küçük datalara ulaşmak daha kolaydır. [6]

Genetik algoritmalar kullanarak farklı ortaklık kuramlarının bulunmasında, big datadan farklı ve ansızın ortaya çıkan setler keşfetmek, bağ kaidelerini data mining şekliyle akıllarda kalmıştır. Karakteristik yaklaşım konusunda sağlam ve sadeleştirici hipotezler uygulamaktadır. Bu setlerin yapısal biçimini tespit etmek ve set niteliğinin ölçüsünü engellemek dayanak ya da itimat gibi sade özellikleri içermektedir. Dayanak ve itimat, ortaya çıkarılan alakanın aşamasının kurallarına engel koymaktadır.[7]

"Itemet destekleyici izlencesine ithafen center of user kuramı" Pazar Sepeti Analizinde Data Mining" bir şirketin öncesi hakkındaki informasyonun ileriye dönük zamanda öngörmek için aracı edilen başarı kazanım kıstasıdır. Nihai davranış yönelimlerini ortaya çıkarır ve şirket bundan fayda sağlayabilir. Data mining, user için temiz data girdisi ve içindeki harika sayıda enformasyon data room sayesinde; eleme prosesinden başarıyla çıkabilir. [8]

"Market Sepet Birliktelik Kuralı Data Mining İncelemesi; Data Mining çalışma detaylarının eğilim gösterdiği bir araştırmadır. Bile-isteye verilmiş hüküm bilgi eksikliğinden daha iyi hükümler verir ve takdir sonuçlarının ana olgusudur. Data Mining vesaitleri ve metotları çalışmak için bir proses ve bir yöntemdir. Bu minvalde olan araştırma, bu algoritmaları incelemesinin uygun pozitif marj sağlayacağı yöntemle ilerlemesine olanak tanımaktır. Bunların olumlu anlamda yenilenmesi için en iyi informasyona yönelik hükümlerde bulunmak en iyi yöntem olacaktır. [9]

2.1.1. Araç ve Metotlar

Data Toplama Metotları: Salahammura ismindeki marketten, Pakistan'daki İslamabad ilinde bulunan Petskari Pazarı, bu günlerde işlem dataları eldesi konusunda başarılı. Mamüllerin elden çıkması müşterinin gerçekleştirdiği uygulamalardır ve farklı mamüllerden oluşur. Bu farklı mamüller müşteri için ayrılan işlemler olarak sekonder saklama alanı diğer bir deyimle standart disk ya da DVD'ler vardır. Değişik birimlerde saklanır, birim satışı sayacıyla, ülke birim satışı günlük, haftalık, aylık satış olarak separate edilmiştir. İlgili matematiksel analizlere dayanarak, ihtiyaç halindeki datalar bu databaseler üzerinden incelenir.[12]

Yöntem: Süpermarker özelinde, supervisor satın alma davranışları özelinde enformasyon detaylandırılmasıyla bilgi talebinde bulunabilir. Mağazaya iletilen bir süreçte ürünü satın alma olasılığı nedir? Sorusuna yanıt bulabilmek adına; Birliktelik Kuralı ve Pazar Sepeti, satın alım yapan kişilerin bireysel satış dataları üzerinden çalışma yapılabilir. Ek olarak, satış ve pazarlama özelinde bir organizasyon planı için düşünebilir ya da reklamcılık kararlarıyla fihrist tasarımı değişik plan normlarında gerçekleşir. Örnek olarak; sürekli kullanılan unsurlar kompozit satın alınmış orderlar kendilerine göre daha yakın konumlandırılabilir. Bu sayede ilgili ürünlerin satış ve pazarlama teşviki başarıyla gerçekleşmiş olur. Pazar Sepet

Analizi olarak; indirim yapılabilecek ürünler ve onların stratejileri ile satış oranları artırılabilir.[12]

Küme ve kümenin içinde bulunan ürünler bazında ele alındığında markette; her malzemenin etrafında temsil eden farklı değişken içerir. Bu değişkenlere iletilen yalın normdaki değer vektörü, satın alma patternleri özelinde detaylandırılır. Sürekli birbiriyle eşleşmesi yapılan ya da satın alım işlemi gerçekleşmiş ürünler belirtir. Aynı zamanda bunlar ilişki setlerini içermektedir. Mesela; PC(personal computer) alan kişiler bununla birlikte yazıcı da alma davranışı sergilemektedir. Ek olarak; biraz sonraki satırda belirtilen birliktelik kuramında bahsedilen konu sayısal olarak şöyle şekillenmiştir[11];

PC = Printer ; Destek =% 10, Güven =% 90.

Ana unsurlar: $K = \{i_1, i_2, \dots, i_m\}$ unsurlarını düşünelim. E(Görevler), bir database uygulamaları kümesi olarak nitelendirilsin. Belirtildiği gibi her bir adım $Y, Y \subseteq K$ gibi unsurları içermektedir. Yeni uygulamalar ve işlemler ise; $Y \subseteq E$ adı verilen bir betimleyici ile bağlantı kurulur. Bir malzeme setini inceleyelim; U bir işlem, B 'yi ve ancak $B \subseteq U$ 'yu kapsıyorsa, bir bağıntı kurma setinde B biçiminin ilişkisi C , burada $B \subseteq K, C \subseteq K$ ve $B \subseteq C = .B \Rightarrow C$ kuralında işlem şemasında E kararlılığı sağlanmasıdır. Destek ile hangi uygulama adımı E kümesinin oransal yüzdesi belirtilir. $B \subseteq C$ 'yi (yani B ve C 'yi) kapsıyorsa bu ihtimal dahilinde; $P(B \subseteq C)$ kural $B \Rightarrow C$ 'nin itimatının D İşlem, E 'ye uygun olarak D uygulamalarının oransal yüzdesidir. E, B de içerir ve bu da aynı zamanda C kapsar. Şartlı ihtimal olarak; $P(C|B)$ şöyle ki; destek = $(B \Rightarrow C) = P(B \subseteq C)$ Güven = $(B \Rightarrow C) = P(C \subseteq B)$.

Asgari destek aşaması (asg_sup) ve asgari güven aşamasını (asg_conf) eşiti olarak şartlara güçlü iması yapılmaktadır. Bütünsel olarak, 0 ve 1.0.11 ikamesi olarak %0 ve %100 aralığında oluşacak formatta destek ve güven sayılarını(oranlarını) buluruz.[18]

Bazı unsurlar bir unsur seti olarak belirtilir. L unsurlarını kapsayan bir unsur set düzeni bir L -unsur setidir aynı zamanda. Format (PC, Printer) 2 unsurluk bir düzenseldir. Bir unsur düzenselinin oluşum aralığı ise; unsur setini kapsayan uygulama adımlarıdır, bu durum ek olarak şunu da basitçe şöyle açıklamaktadır; frekans dayanağı sayısı ya da unsur düzenselinin sayımsal durumunu özetlemektedir. Unsur düzenseli; bir unsur kümesinin belirlenme

frekansını ilgili ürüne asg_sup'a eşit ya da ürüne eşleşimi bir diğer değişle total uygulama sayısı E şeklinde de alınabilir. Yani ürün dayanağının minimum dayanağı karşılaması için ihtiyaç olan uygulama adımları, asgari dayanak sayısı şeklinde belirtilir. Bir unsur düzenseli minimum dayanağı sağlıyorsa eğer, frekansı sık bir unsur düzenselidir. Frekans yoğunluğu yüksek unsur seti normalde Lk şeklinde ifade edilir. Data Mining kuramında Birliktelik Kuralı 2 adımlı bir prosestir;

- i) Frekans yoğunluğu yüksek bulunan unsurlarda tanım gereği, her unsur seti minimum önceden belirtilmiş bir asgari destek sayısı kadar frekans aralığında oluşur.
- ii) Frekans yoğunluğu yüksek unsur setinden sağlam birliktelik kuralı şunu zengileştirir ve bu sınırlamalar minimum desteği ve minimum güvenin oluşmasına olanak getirir.[16]

Tablo 2.1: İşlem Dataları

İşlem ID	1'den Fazla Ürün Satın Alan Müşteri Ürünleri
1	Kurabiye, Makarna, Fasülye, Bulgur
2	Kurabiye, Fasülye
3	Makarna, Fasülye
4	Fasülye, Makarna, Bulgur
5	Makarna, Fasülye
6	Kurabiye, Makarna
7	Kurabiye, Bulgur, Fasülye

İşlem datalarına baktığımızda, Pazar Sepeti Analizi aracılığıyla bağıntı setlerinin sonuçlarına ulaşabiliriz;

- i) Pazar Sepeti aşama aşama incelemesi aşağıda belirtildiği şekildedir;
- ii) İhtimal dahilindeki tüm bağıntıları kurun.
- iii) İhtimal dahilindeki tüm bağıntıların güven ve destek hesaplamalarını gerçekleştirin.
- iv) Minimum destek ve minimum güven eşik ölçütlerini bağıntı kuramı eldesi için uygulayın.[16]

Hala satın alım işlemi yapan kişiler tarafından dikkat çeken unsurları X (bağımsız değişken) ve bağıntısı olan, geçerli farklı unsurları Y (bağımlı değişken) ifadeleri ile hatırlayalım. Eğer sadece 2 unsurumuz varsa; (B ve C), sadece 2 ihtimal bağıntı kuralı mevcuttur:

[B] $\rightarrow$ [C] ve [C] $\rightarrow$ [B]

B, C ve D; bu 3 unsur özelinde olası on iki(12) bağıntılandırma kuralı mevcuttur:

Tablo 1.2: Karma Ürün Kombinasyonları

	X	Y
1	[B] $\rightarrow$	[C]
2	[B] $\rightarrow$	[D]
3	[B] $\rightarrow$	[C,D]
4	[C] $\rightarrow$	[B]
5	[C] $\rightarrow$	[D]
6	[C] $\rightarrow$	[BD]
7	[D] $\rightarrow$	[B]
8	[D] $\rightarrow$	[C]
9	[D] $\rightarrow$	[BC]
10	[B,C] $\rightarrow$	[D]
11	[B,D] $\rightarrow$	[C]
12	[C,D] $\rightarrow$	[B]

Bağımsız değişkenin (X), e unsurlarının sayısının içindeki e-1'e kadar olan unsurların kompoziti olduğuna emin olunmalıdır. Bağımlı değişken (Y), bu listedeki tüm unsurlar ve unsurlar arası farkların kompozitlendirilmiş halidir. [10,11]

4 öge için 50 farklı kombinasyon birleşimi aşağıdaki gibi listelenmiştir;

Tablo 2.3: Ürünler Kombinasyonu(Satın Alınmış)

No	X	Y	No	X	Y	No	X	Y	No	X	Y	No	X	Y
1	B	C	11	C	B D	21	D	B C E	31	B C	D E	41	CE	B
2	B	D	12	C	D E	22	E	B	32	B D	C	42	CE	D
3	B	E	13	C	B E	23	E	C	33	B D	E	43	CE	B D
4	B	C D	14	D	B D E	24	E	D	34	B D	C E	44	DE	B
5	B	D E	15	D	B	25	E	B C	35	B E	C	45	DE	C
6	B	C E	16	D	C	26	E	C D	36	B E	D	46	DE	B C
7	B	C D E	17	D	E	27	E	B D	37	B E	C D	47	B C	E
8	B	B	18	D	C C	28	E	B C D	38	C D	B	48	B C	D
9	B	D	19	D	C E	29	B C	D	39	C D	E	49	B D	C
10	B	E	20	D	B E	30	B C	E	40	C D	B E	50	C D	B

Toplam Ürün Sayıları	1	2	3	4	5	10	100	500
Birlikteli Kuralı Dahilindeki Ürün	0	2	24	48*0	360	77004	6.15378E+38	4.636E+176

Genel olarak, ihtimal dahilindeki bağıntı setlerinin totalde= R kadar ve aşağıdaki ifade edilen formülzasyona uygun şekilde sayısı; e üstel: $R = 3e - 3e + 1 + 1$. Aslında, binlerce unsur olduğunda, tüm ihtimal dahilindeki bağıntı setlerinin ve set frekans sıklığının hesaplamak oldukça zordur ve ihtiyaç dahilindeki yeterliliğe de maalesef sahip olamayacaktır. Destek ve güveni bulmak için ilk olarak işlem datalarının 2'li dataya dönüştürmeliyiz.

Tablo 2.4: 2'li Dataya Dönüşüm (İşlem Datası)

İşlem	Müşterinin Birden Fazla Satın Aldığı Ürünler	İşlem	B	C	D	E
1	Kurabiye , Makarna , Fasülye , Bulgur	1	0	1	1	1
2	Kurabiye , Fasülye	2	1	1	0	1
3	Makarna ,Fasülye	3	0	1	1	1
4	Makarna , Bulgur , Fasülye	4	1	0	1	1
5	Makarna , Fasülye	5	0	1	1	1
6	Kurabiye , Makarna	6	1	0	1	1
7	Kurabiye , Bulgur , Fasülye	7	1	1	1	0
Toplam			4	5	6	6

Yalınlık için ilgili ürünlerde kısaltmalar kullanıyoruz. 2 kutuplu uygulama kayıt tablosuna

1000 Record ismini kullanalım. 100 Record; bütün her bir sütunda B_, C_, D_ ve E_ olarak ifade edilir. Ek olarak, 1000 Record'un işlem kayıtlarına destek olmak ve 1000 olarak isimlendirmek için sol tarafındaki bir(1) dizisini de belirlemiş ve işleme koymuş olduk.[10]

$N(X \cap Y)$ ile ifade edilen dayanak rakamı matematiksel olarak ölçülebilir durumdadır. İşlem sayısı olan N hep 7'dir, bunun nedeni ise sadece 7 data mevcuttur. Dayanak (bağıntılama sınırlandırması için yüzde olarak ve sadece dayanak rakamı olarak inceleme sayısının birbiri ile ilişkili matematiksel oranıdır.

$$\text{Destek } (X - Y) = n(X \cap Y) / N \quad (2.1)$$

Hür değişken olan n tane-x, sadece bağıntılama uygulamasının totalde hür olmadığından kompozitleme adına dayanak olarak basit bir norma aittir. İtimat ise hür olmayan parametrenin kompozitinin dayanak rakamlarının hür olmayan değişken dayanak rakamlarına matematiksel kıyaslarıyla çok basit bir şekilde tahminlenir.

$$\text{İtimat } (X \cap Y) = n(X \cap Y) / n(X) \quad (2.2)$$

Market basket analysis bağıntılama setlerini bulmak için ise 3. Adım; minimum dayanakla birlikte minimum itimatla 2 adım seçeneği hesaplamasıdır. Bu sayede; bu 2 adımı hücrelere ekledikten sonra mesela minimum % 60 dayanak ve % 90 asgari itimat seçiyoruz. [11]

Asgari Dayanak =% 60, Asgari İtimat =% 90

Değişik adım values çok yaygın ya da yüksek frekanslı rules setleri oluşturur. Mesela; asgari dayanağı % 30'e, minimum itimatı % 65'a uyarlırsak sekizli ilişkilendirme ve bağıntı kuram sonuçlarını aşağıdaki gibi elde edebiliriz;

Tablo 2.5: Destek ve Güven Data Ölçümlemesi

No.	X	Y	N(XUY)	N	%Destek	n(X)	Güven	Kural Uyumu
1	B	C	2	7	30%	5	51%	0
2	B	D	2	7	30%	5	51%	0
3	B	E	3	7	44%	5	76%	1
4	B	C , D	1	7	15%	5	26%	0
5	B	D ,E	2	7	30%	5	50%	0
6	B	C , E	1	7	15%	5	28%	0
7	B	C , D , E	1	7	15%	6	27%	0
8	C	B	2	7	30%	6	42%	0
9	C	D	2	7	30%	6	42%	0
10	C	E	4	7	58%	6	82%	1
11	C	B, D	1	7	15%	6	22%	0
12	C	D, E	2	7	30%	6	42%	0
13	C	B, E	1	7	15%	6	22%	0
14	C	B , D , E	1	7	15%	5	22%	0
15	D	B	2	7	30%	5	69%	1
16	D	C	2	7	30%	5	69%	1
17	D	E	3	7	44%	5	99%	1
18	D	B, C	1	7	15%	5	34%	0
19	D	C , E	2	7	30%	5	68%	1
20	D	B, E	2	7	30%	5	68%	1
21	D	B , C , E	1	7	15%	5	34%	0
22	E	C	3	7	44%	8	51%	0
23	E	C	4	7	58%	8	68%	1
24	E	D	3	7	44%	8	51%	0
25	E	B, C	1	7	15%	8	18%	0
26	E	C D	2	7	30%	8	34%	0
27	E	B, D	2	7	30%	8	34%	0
28	E	B , C, D	1	7	15%	8	18%	0
29	B C	D	1	7	15%	11	12%	0
30	B C	E	1	7	15%	10	12%	0
31	B C	D, E	1	7	15%	10	12%	0
32	B D	C	1	7	15%	8	15%	0
33	B D	E	2	7	30%	8	30%	0
34	B D	C , E	1	7	15%	8	15%	0
35	B E	C	1	7	15%	11	11%	0
36	B E	D	2	7	30%	11	21%	0
37	B E	C , D	1	7	15%	11	11%	0
38	C D	B	1	7	15%	9	14%	0
39	C D	E	2	7	30%	9	26%	0
40	C D	B E	1	7	15%	9	14%	0
41	C E	B	1	7	15%	12	10%	0

42	C E	D	2	7	30%	12	19%	0
43	C E	B D	1	7	15%	12	10%	0
44	D E	B	2	7	30%	10	23%	0
45	D E	C	2	7	30%	10	23%	0
46	D E	B C	1	7	15%	10	12%	0
47	B C	D E	1	7	15%	13	9%	0
48	B C	E D	1	7	15%	16	8%	0
49	B D	E C	1	7	15%	14	9%	0
50	C D	E B	1	7	15%	15	8%	0

Tablo 2.6: Bağlantılama Sonucu Elde Edilen Uyum

No	X	->	Y	%Dayana	İtimat	İntibak Onayı
3	B	->	E	44%	76 %	√
10	C	->	E	58%	81 %	√
15	D	->	B	28%	68 %	√
16	D	->	C	29%	67 %	√
17	D	->	E	43%	100 %	√
19	D	->	C E	29%	67 %	√
20	D	->	B E	29%	67 %	√
23	E	->	C	57%	67 %	√

Tablo örneklemeden yola çıkarak çeşitli uygulamalar ve ayrıntılı analizlemesini yapıyoruz; X ve Y malzemelerinin birlikte satıldığı, güven – ürünü olan X’i alan kişinin Y ürününi yani şartlı ihtimal dahilinde satın alma frekans yoğunluğunu belirtmektedir.[19]

Ek olarak; bağlantılama setinden güven ve destek oranlarının aşama adımları verilebilir. Baktığımızda destek için asgari bir değer üzerinde durulduğu yani yüzde yirmi bir(%21), X-Y malzemelerinin beraber işlenme yoğunluğu olan asgari consumer satın alma güven aralık değerlemesinde ise satın alınan malzemeler arasında X de Y de malzemelerinin satın alınma işlem yoğunluk aralığını ifade etmektedir.[19]

Tabloyu genel olarak incelediğimizde yalnızca elli(50) uygulama sırasında 4,11,16,18,20, 21 ve 24. uygulama adımlarının dayanak ve itimat alt sınır aralık değerlerinin fazlalık gösterdiği tespit edilmiştir bu da demek oluyor ki; %20 ve %60 uygulamadaki malzemedir. Kuralda dahil olarak kati suretle müşteri tarafından ürün alınır ve ilgili ürün/ürünleri sepetine

eklemektedir.

2.1.2. Birlikte Kullanılan Kuralları Genel Değerlendirilmesi

Yapılan çalışmalar neticesinde data mining araç ve yöntemlerinin, satın alma işlemi yapılan bazı ürünlerin alım esnasında yapılan işlemlerin hareketli eğilim davranışlarıyla alakalı kalıpları optimum hale getirmek için etkin halde uygulayabileceği ve kullanılabileceği tespit edilmiştir. Data Mining bağıntı kuralı geniş ve daha önemli olarak uygulanan daha çok algoritmik bazda ise Pazar Sepeti uygulama algoritması olarak uygulandı. Bu uygulama sayesinde satın alma işlemi yapan kişiler tarafından oldukça yoğun frekansla yapılan işlemler müşterilerin bağıntılı işlemlerinde satın almalarına olan desteği aynı zamanda da güveni kullanarak irdelemiş ve raporlamıştır. Bu yöntemle birlikte ürünler arasında satın alımı yapan kişiler arasında malzeme satın alınırken belli bağıntılar olduğu açıkça görülmektedir.[14]

Ek olarak; bu analizle birlikte raflarda malzeme düzeninde en yüksek verimle kullanılabileceği görülmektedir. Bu metot ise ürün hizmetini sunan satıcının yüksek marjlı kar elde edebileceğini göstermektedir. Bu sayede Data Mining araçlarıyla ürünün raflara iletilme karar vericiliğini iletirmek ve aynı zamanda geliştirmek adına kullanılabilir.[14, 15]

2.2.APRIORI ALGORİTMASI

Yüksek frekanslı kullanım oranı yüksek unsurları bulma Data Mining'te sıkça araştırılan konulardandır. Apriori; (FIM)-sık unsurlar algoritması- bu konuda en çok kullanılan algoritmadır. Bu algoritma farklı işlemleri raporize etmiş ve incelemiştir. Data structure Bodon ve trie ile max faydalı işlem ve uygulama olarak ön plana çıkmıştır. Bodon'un yüksek frekanslı unsur bulmasının sonuçları olarak; diğer araştırmacı bulgularından daha çevik bir formda olduğu nitelenmektedir. Araştırmada girdi adımlarının eş bir PC tarafından tarandığında eş bir model ile bu sisteme uygun şekilde revize edildi. Değişen işlem ve uygulama adımlarına eş uyumlu PC etkisi sunulmuştur.[12]

İşlem database sık kullanılan elemanların kullanılması, 1'den fazla işlemde faydalı olduğu ispatlanmıştır [12]. Farklı geniş bir databasede sık sık unsur setleri için çok sayıda algoritma teklif edilmiştir. Ayrıca, her destek aşamasıyla her database için her bir işlem adımından daha

iyi performans ile gösterilmiş bir uygulama bulunmamaktadır. [13] Genelde çok sayıda uygulama 2 temel algoritmaya göre uyarlanmıştır ve bu temel üstüne yazılmıştır; Apriori [14] ve sık pattern genişlemesi (FP Genişlemesi). [15] Apriori, itere edilerek çok geniş databasede sık unsurları bulur. Apriori algoritması, aday unsurları bulmak, desteği matematiksel olarak ölçmek ve her iterasyonda yüksek frekanslı elemanlar için aday elemanlar yalın hale getirmek için önemli bir ihtiyaçtır. FP genişleme ve ilerleme algoritması çok vakit almayan ve sık unsurları bulan bir algoritmadır.

Birçok durumda FP genişme algoritması Apriori algoritmasına göre daha iyi olsa da yüksek frekanslı ve sık olan elemanların data mining prosesini çevikleştirmek için Apriori Algoritmasını kullanarak farklı geliştirmeler uygulamıştır ve işleme almıştır.

Bu çalışmada, Bodon'un pratiklerine ve araştırmalarına ilişkin paralel bir Apriori uygulandı ve eş-uyumlu bir PC'de performansını değerlendirmesi yapıldı. Bodon'un eş-uyumlu matematiksel ölçümü için bu işlem adımlarının üstünde durulmasının nedeni ise Bode'nin trie data structure olarak işleme alınan çalışmasının, hash tree kullanan uygulamalardan daha iyi durumda gösterilmesidir.[16] Çalışmanın sonraki bölümleri de şu şekilde detaylandırılmış ve organize edilmiştir; Part 1; sık frekanslı mining çalışma tanıtımı, Part 2; sık kullanılan eleman mining eş-uyumlu matematik ölçüm için uygulama kullanımı. Bu sayede simetrik multi işlemcili bir PC'de çalışmanın bilimsel verilerini sunarak bu çalışmayı özetlemektedir.

2.2.1. Apriori Algoritması ve Ayrıntıları

Databaselerde enformasyon bulgusu(KDD) olarak bilinen data mining, çok geniş veritabanlarından faydalı, saklı informasyonların otomatik şekilde ortaya konması işlemidir. [12]. Data mining temel templatelerinden birisi olan apriori değişken geniş databaselerdeki elemanlar arasındaki mining bağıntı kuralları olarak bilinmektedir. Birlikte kurallarının detaylı irdelenmesi aracılığıyla yüksek frekanslı kullanılan unsur setleri bulmak için de faydalı bir yöntemdir. [4]

TDB = {T1, T2, T3, ..., Tn} uygulamaları database temeli şeklinde düşünelim. Uygulama $T_i = \{i_1, i_2, i_3, \dots, i_m\}$ $I = \{i_1, i_2, i_3, \dots, i_p\}$ olarak belli kısım unsurlar kapsıyor olsun ve r / n ,

Ki / B gibi. L unsurlu Y unsuru bir l-eleman olarak gösterilsin. Bir diğer uygulama adımı ve işlemi olarak T_i Y unsur setini kapsıyorsa sadece X kapsar T_i şeklinde desteği oluşur.

Sup şeklinde ifade edilen T satırı elemanı olarak X, TDB'de X kapsayan adım sayısını ifade eder. Destek, $Sup(X)$ asgari destek olarak ifade edilen bir destek adımından daha büyükse bu eleman setinin sıkı olduğunu gösterir. Mesela, $T1 = \{C, D, E, F\}$ olduğu bir $TDB = \{T1, T2, T3, T4, T5, T6\}$ ve $I = \{B, C, D, E, F\}$

$T2 = \{C, D, E\}$, $T3 = \{B, C, E\}$, $T4 = \{B, C, D, E, F\}$

$T5 = \{B, C, D\}$ ve $T6 = \{C, F\}$. Tablo 2.7'de gösterildiği gibi yüksek frekanslı data setlerinin her bir satırdaki data mining işlem-data seti tablosu;

Tablo 2.7: Data Mining İşlem Sıklık Örneği(a)

İşlem	Data Seti
T1	{C, D, E, F}
T2	{C, D, E}
T3	{B, C, E}
T4	{B, C, D, E, F}
T5	{B, C, D}
T6	{C, F}

Tablo 2.8: Data Mining İşlem Sıklık Örneği(b)

{1-Data Seti}	İşlemler
{B}	T3, T4, T5
{C}	T1, T2, T3, T4, T5, T6
{D}	T1, T2, T4, T5
{E}	T1, T2, T3, t4
{F}	T1, T4, T6

Tablo 2.7(a) içindeki unsurlarıyla işlem listeleri ve Tablo 2.8.(b) 1 eksi 2 mesela

$Sup(\{B, F\}) = 1$, ek olarak $Sup(\{B, C, E, F\}) = 1$, Tablo 2.7. (a)'da ise 1-eleman türetmek adına benzer biçimde matematiksel ölçümü yapılabilir.

2.2.2. Apriori Algoritması Uygulamaları

Apriori öncelikle Srikant ve Agrawal tarafından tasarlanmıştır[12]. User tanımlı asgari desteğe uygun yüksek frekanslı unsur setleri belirir. Algoritma öncelik geçişinde geçiş adayı olan 1-eleman setleri kurulur. Sonraki süreçte ise algoritmanın destek değerleri asgari desteğin değerinin altında seyrediyorsa diğer aday 1-eleman setlerini temizlemek suretiyle yoğun olarak kullanılan 1-elemanlık yeni setler oluşturulur. Algoritma içerisinde yüksek yoğunlukla gerçekleşen 1-elemanlık tüm unsurlar bulunmasının ardından 2-elemanlık setler arasından sıklık göstermeyen unsurların temizlenmesi ardından 1-eleman setleri ile kompozit hale getirilir. Gerçekleştirilen bu işlem aday unsur setlerinin olgun bir şekilde oluşana dek itere bir şekilde devam etmektedir.

2.2.3. Örnek Apriori Yazılım Kodları ve Uygulama Örneği

Apriori Algoritması alanında ilk çalışmaları yapan Agrawal ve Srikat tarafından algoritmanın başlangıç kod çalışmaları ve uygulamaları aşağıdaki gibi verilmiştir;

$L_1 = \{\text{sık tekrarlanan öğeleri içeren küme}\}$

for(k=2; L k-1 != 0 ; k++){

Ck = apriori-üret (L k-1);

forall (işlemler t ID){

Ct $\subset$ (C(k),t);

forall adaylar c $\in$ Ct{

c.sayaç++;

}

}

$L_k = \{ c \in C(k) \mid c.sayaç = \text{enküçük} \}$

Sonuç = $U(k) L(k)$;

Algoritmadaki L , database ya da data setimizde bulunan unsurları ifade eder [22]. Apriori - üret fonksiyonundan- çıkan her bir $C(k)$ öge setinde t bir işlem olma şartıyla birlikte ortaya çıkarılan bu $C(k)$ öge setinin t işlemini kaydeden sayacın her ilişkisinde 1 işlem artışı sağlanır. Akabinde sıradaki L 'yi işleme koyarak algoritma işlemlerini sürdürür. Sonuç itibarıyla de üretilen bu L kompozit kümesini itere ediyoruz.

➤ **Apriori-üret fonksiyonunu ise şu şekilde açıklayabiliriz;**

C_k 'ya ekle

select p.eleman1, p.eleman2, ... p.eleman k-1, q.eleman k-1

for(L_{k-1} p'den L_{k-1} q'ya kadar)

where p.eleman1, p.eleman2, ... p.eleman k-2 = q.eleman k-2 , p.eleman k-1 < q.eleman k-1

forall elemanlar $c \in C_k$ {

forall c 'nin $(k-1)$ -alt kümesi s için{

if($s \notin L_{k-1}$)

{

c 'yi $C(k)$ 'dan sil

}

}

}



Yukarıdaki algoritma yalın bir şekilde aday elemanların üretilip itere bir şekilde aday kümeler ile eşleşip iletilmesini sağlamaktadır. [22]

Tablo 2.9: Apriori Uygulaması-1

İşlemler	Data Seti
T1	{C, D, E, F}
T2	{C, D, E}
T3	{B, C, E}
T4	{B, C, D, E, F}
T5	{B C, D}
T6	{C, F}

Tablo 2.10: Apriori Uygulaması-2

Malzeme	Destek
{B}	3
{C}	6
{D}	4
{E}	4
{F}	3

Tablo 2.11: Apriori Uygulaması-3

1-Malzeme	Dayanak
{B}	3
{C}	6
{D}	4
{E}	4
{F}	3

Tablo 2.12: Apriori Uygulaması-4

2-Malzeme	Dayanak
{B}	3
{C}	6
{D}	4
{E}	4
{F}	3

Tablo 2.13: Apriori Uygulaması-5

2-Malzeme	Dayanak
{B, C}	3
{C, D}	4
{C, E}	4
{C, F}	3
{D, E}	3

Tablo 2.14: Apriori Uygulaması-6

2-Ürün Seti	Destek
{B, C}	3
{C, D}	4
{C, E}	4
{C, F}	3
{D, E}	3

Tablo 2.15: Apriori Uygulaması-7

Aday 3-Ürün Seti	Destek
{B, C, D}	2
{B, C, E}	1
{B, C, F}	1
{C, D, E}	3
{C, D, F}	2
{D, E, F}	2

Tablo 2.16: Apriori Uygulaması-8

3-Ürün Seti	Destek
{C, D, E}	3

Tablo 2.17: Apriori Uygulaması-9

3-Ürün Seti	Destek
{C, D, E}	3

Tablo 2.18: Apriori Örnek Uygulaması-10

Aday 4-Ürün Seti	Destek
{}	

Tablo 2.9 – Uygulama (1) 'de, beş 1-öge grubu, tüm bir maddelik öğelerin asgari destek 3'ten daha fazla desteği olduğu uygulama databasesinden kurgulanmıştır. Tablo 2.12. Uygulama-(4), 2 unsur seti sunar, Tablo 2.15. Uygulama-(7) sık kullanılan 3 unsur seti önermektedir. Ayrıca, geçiş unsuru 4'te oluşturulan 4 unsurluk küme bulunmuyor, sebebi ise herhangi bir önerilen aday 4 unsur kümesi yapamadı. [17]

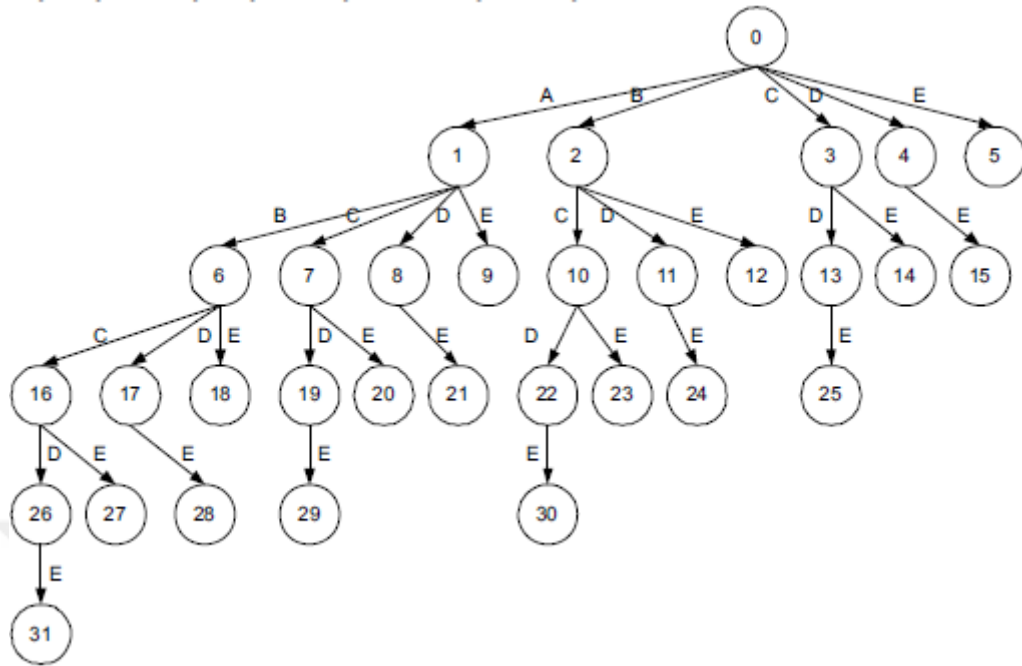
2.2.4. Çevik Apriori Uygulaması

Çevik apriori uygulaması ile ilgili pek çok sayıda araştırması ve çalışmaları bulunmaktadır.

[17] Bodon çalışmalarında genellikle adayların unsurların üretilmesine ve data yapılarının işlenmesine ve uygulamaların detaylarına özgü bir şekilde frekans yoğunluk aralığı özelinde data mining algoritmalarının kazançları üzerine tartışmalar açarak konunun gelişimi içinde katkıda bulunmuştur. Bodon şu anki algoritmalarda gerekli olan çalışma aşamalarının gözardı edildiğini vurgulamıştır. Buna istinaden Bodon tüm adayların sürece dahil olabilecekleri çevik bir apriori uygulaması sunmuştur. [16]

Apriori algoritması normalinde hash ağacı [13] örnek alınarak uygulandı. Bodon trie yapısı özelinde bir Apriori algoritmasını işleme aldı. Bu bir üçgenin köklü bir şekilde eğilimlendirilmiş halidir. Kökün derinliği ise 0'dır. 2 dönüm noktası arasında bir uzunluk vardır ve her bir uzunluk için 1 harf ayrılmıştır. Bu bağlamda; $K = \{k_1, k_2, k_3, \dots, i(m)\}$ kapsamındaki belli unsuru içerir. 'A' un, 'B' meyve suyunu, 'C' yoğurdu, 'D' bademi, 'E' sütü ifade eder ve küme; $K = \{A, B, C, D, E\}$ şeklindedir. Bu sayede, $C(k) = \{i_1 < i_2 < i_3 < \dots < i_k\}$ şeklinde bir namzet unsur seti oluşur, $I = \{i_1, i_2, i_3, \dots, i_m\}$ unsurlarının bulunduğu şekilde öge setleri kurulabilir. Örneğin $C(2)$ 'deki 10 aday unsur 2'li veya 3'lü şekilde kümelemeler ve alt kümelemeler aşağıdaki gibi olabilir.

$\{A, B\}, \{A, C\}, \{A, D\}, \{A, E\}, \{B, C\}, \{B, D\}, \{B, E\}, \{C, D\}, \{C, E\}$ ve $\{D, E\}$.



Namzet elemanları saklamak adına oluşturulan 3'lü ağaç aracılığı ile namzet eleman oluşturulması ile etkili bir birliktelik kuralı aynı zamanda data yapılarının basit işlemesi sağlanabilir. Bununla birlikte data yapılarının yalın uygulaması ve canlı bir şekilde bağıntılanması için olumsuz etki oluşturabilecek örneklem tabanlı algoritmaları da kendi içine dahil edebilir. [16]

2.2.5. Eş Uyumlu Uygulamanın Hesaplanması

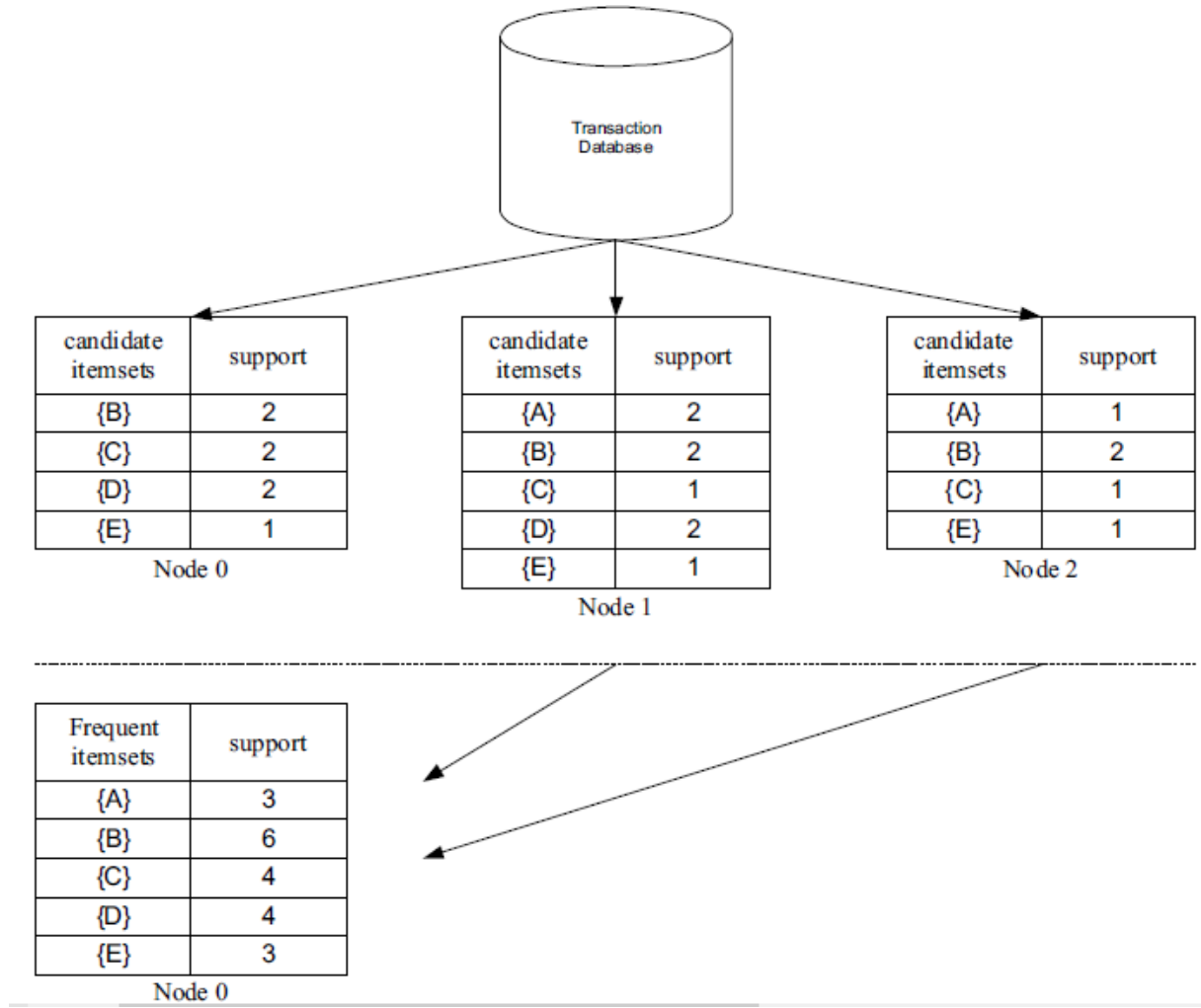
Apriori algoritması bu zamana kadar farklı şekillerde düzenlenmiştir. Bu revizasyon bir uygulama database'ini bağımsız bir şekilde TDB1, TDB2, TDB3, ..., TDBn şeklinde ayırmaktadır. Bu database ayırımı ile dataya hızlı erişim ve yüksek frekanslı unsur setlerine artarak büyümesi sağlanır. Ek olarak da bölümlenme sayesinde ana belleğe uyum ve erişimiyle ilgili birimlerin çalışması ve analizi kolaylaşabilir.

Database'i N birimlere ayırmanın akabinde J birimini her bir dönüm noktasında lokal k-
elemanlar tarafından tahminlenen ve sonuçlanan şekilde N dönüm noktalarına yayılan birim

“Apriori” uygulamasıdır. Dönüm noktalarının lokal namzet k-elemanları tamamlandığında namzet lokal k-elemanları dönüm noktasını sıfıra indirger. Dönüm noktası sıfır sonrasında da tüm aday k-eleman kümelerinin totalini matematiksel olarak ölçer ve aday k-eleman kümelerini sık sık k-eleman kümelerini temizlemektedir. Şekil 5, eş uyumlu hesaplama için gerekli çalışma özetini göstermektedir. Eş uyumlu bir bilgisayarda N dönüm noktasının varolduğunu düşünelim.[19]

- Eş Uyumlu İşlem Database (TDB) tarayarak frekansı yüksek 1 unsurluk kümelerin eldesi için L1;

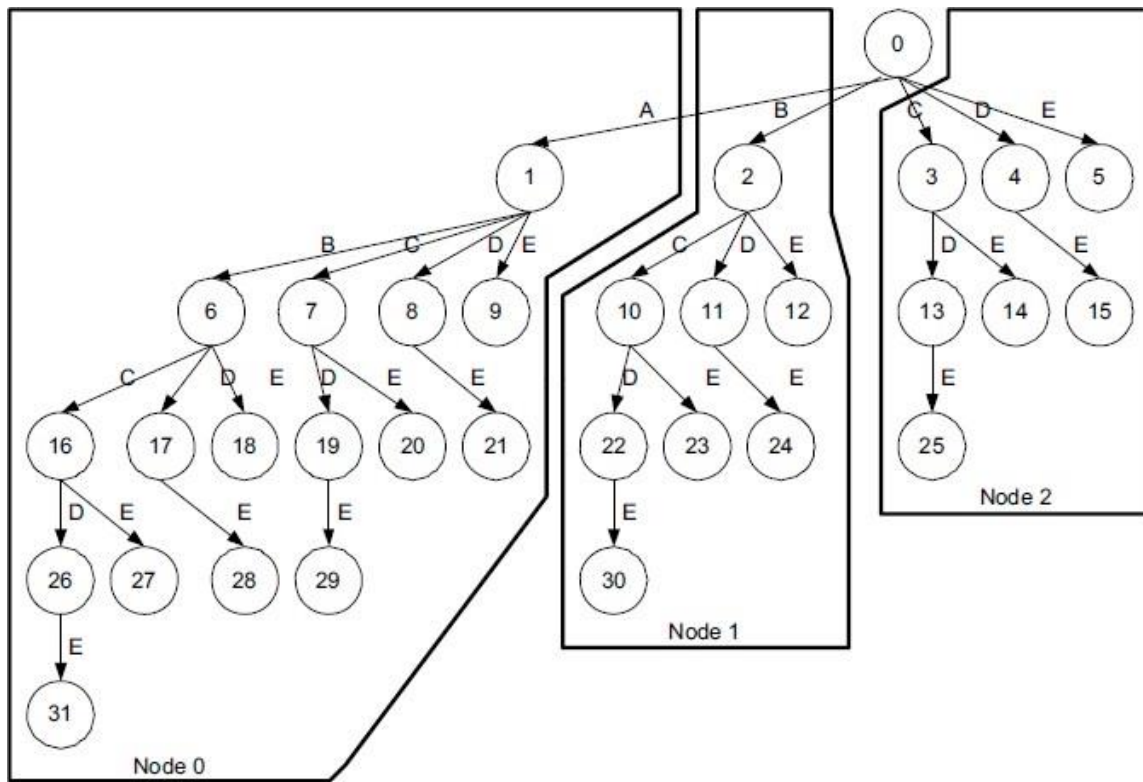
- TDB; N birimime bölünür,
- İşlemci ise P_i biriminden aday 1-eleman oluşturmak TDB [i] okur,
- P_i , lokal unsur adayını 1 unsurluk kümelerini P_0 'a iletir,
- P_0 , bütün lokal aday 1-elemanları matematiksel ölçümü ve yüksek frekanslı 1 elemanlık seti L1'i oluşturmak için temizlenmesi,
- Yüksek frekanslı 3 unsurlu kümeleri kapsamak adına eş uyumlu analiz J2,
- P_i lokal adayı 3 unsurlu kümeleri oluşturmak için analiz eder,
- P_i , lokal 2 unsurlu kümelerini ana unsura iletir,
- Ana unsur, tamamının lokal namzet 3-elemanlarını tahminler ve J3 için temizleme(budama) işlemini uygular. [18]



Şekil 2.4: Data Set Uygulama Gravite Oranları

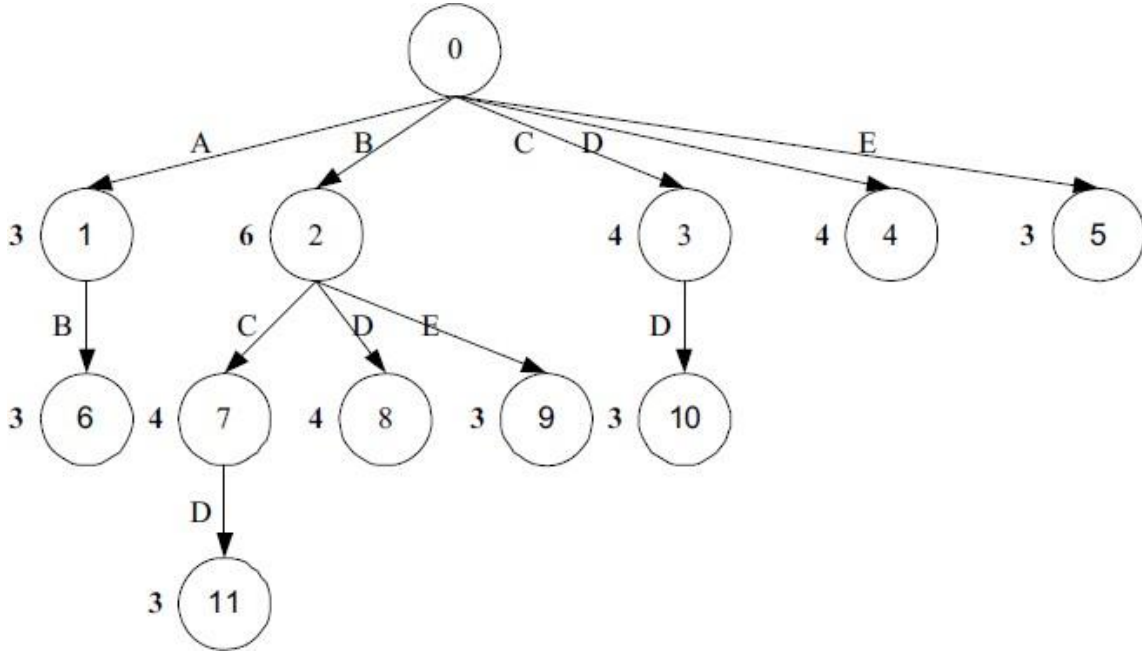
Lokal namzet unsur-2 kümelerini seçmek adına da aynı proses işler. Bir dönüm noktası adına verilen data birimine ilişkin şekilde lokal namzet 2-madde kümelerini hesaplamaktadır. Sonuçlar yüksek frekanslı 2 unsurluk kümelerin oluşturulması için dönüm noktaları sıfıra geri iletilir. Bu çalışma özelinde Bodon'un çalışmasıyla birlikte benzer biçimde trie data modelini bir veya iki unsurluk kümeleri saklamak için kullanılmadı. [17] Bu uygulamanın ikamesi olarak namzet unsur-bir kümelerini saklamak adına yalın bir namzet iki maddelerini depolamak adına iki aşamalı işleme koyulur. Yüksek frekanslı k-unsurlu kümeleri oluşturmak adına da k üçlü ağaç veri modeli k-elemandan oluşur. Dönüm noktası $L(k)-1$ 'e bağlı olarak Her bir dönüm noktası $L(k-1)$ özelinde lokal k-eleman kümeleri için lokal trie'sini kademeli bir şekilde oluşturur. Her bir unsur kümesinin sayısı en nihayetinde özet olarak raporize edilir, sonrasında da her dönüm noktası yüksek frekanslı k-elemanlarını oluşturmak için kendi lokal

aday k-elemanlarını temizler. Global olarak yüksek frekanslık- elemanları her dönüm noktasından sonra tüm lokal yüksek frekanslı k elemanlarını toplamak için kullanılır. Proses daha fazla aday unsur kümesi oluşana dek itere olur. Şekil 7'ye bakıldığında 3 dönüm noktası içinde aday k-elemanları saklanamak üzere bir trie tasviri içinde oluşur. Database üzerinde dönüm noktalarını 3 ayrı bölüme ayırarak her bir bölüm kendi içerisinde bağımsız olarak çalışmasına izin verilir. Bu sayede A unsurunu içeren işlemler dönüm noktası sıfıra paylaştırılır ve bu unsurun içinde barındıran uygulamalar dönüm noktası bire iletilir. Uygulamalardan artan bölümleri ise dönüm noktası 2'ye iletilir. [17, 18]



Temelde Appriori Algoritması olan uygulamada aday elemanları saklamak üzere kullanılan data structure ve data sayıları bir trie aracı ile sağlanır. 3'lü data structure unsur kümelerini saklamak, ulaşmak ve sayım işlemi için etkin bir yöntem şeklini sağlar. Aday elemanlara bağlı bir şekilde sıralı bir trie oluşturulur. Aday unsur seti eğer belirli aralıklarla trie üzerine

yerleşmişse trie daha fazla genişletilmez. Bu sayede 3'lü içindeki bir dönüm noktası sadece ilgili olan yüksek frekanslı unsur kümelerini içermektedir. Şekil 2.6'da ifade edilen örnek baz alındığında asgari destek üç(3) yüksek frekanslı unsur kümesini saklamak için oluşturulmuş bir trieyi gösterir. [18]



Şekil 2.6: 1 ve 2'li Unsurlu Data Kümesi Tree

3'lü ağaç yapısıyla asgari destekli üç işlemle aşağıdaki unsurlar ulaşılır;

{A}, {B}, {C}, {D}, {E}, {A, B}, {B, C}, {B, D}, {B},

Sayısal olarak; 4, 7, 5, 5, 4, 4, 5, 5, 4, 4 ve 4 ve ardından dayanağıyla {E}, {C, D} ve {B, C, D} Şekil 2.6'da ifade edildiği gibi database üzerinde işlem gören yapıyı incelediğimizde uygulama sayıları ile yüksek frekanslı unsur kümelerini saklamak kademeli şekilde bir ürün ağacı oluşturmuştur.

Apriori temelli algoritmaların aday unsurlarının oluşturulduğu süreçte unsur setlerini saklamak için data yapılarının optimal hale getirilmesi ve ilgili ayrıntıların üstüne yoğunlaşmaktır. [17, 18]

2.3. MARKET SEPET ANALİZİ

Günümüzde firmalar data sahipliği konusunda oldukça şanslı ve zengin ancak sahip oldukları bu dataları değere dönüştürme konusunda bir o kadar zayıflık göstermekte. Gelişmekte olan dünyayla birlikte büyük boyutlu dataların önemli bir kaynak olduğunun farkedilmesi büyük bir şans ve farklı endüstri alanlarında şirketlerin alacakları stratejik karar noktalarında bu datalara olan güven oranı her geçen gün biraz daha artmakta. Hem de bu sayede geçmişten bugüne gözden kaçan fırsatlar ve imkanlar data tarama yöntemleriyle açığa çıkarılabilir.[16] Pazar sepeti analizi son kullanıcı ve alıcının satın alım işlemlerinde sepetlerinde bulunan ürünleri kolay bir şekilde bulmak ve erişmek için iyi bir metodolojidir. Buradan elde edilen enformasyon dataları değere dönüştürülerek; çapraz satış kampanyaları, stok kontrolü ve promosyon planlamalarına fayda sağlamaktadır. İşlem verileri ile data mining kompozit sınırlandırmala ürünlerin paralel olarak eşleşmeleri ve buna uygun şekilde paralel satın alımlar için değerli bilgi girdisi sağlar. Bu tür girdiler ve uygulamalar neticesinde çapraz satış kampanyaları, stok kontrolü ve promosyon planlamalarına yönelik proje çalışmaları yapılır.[21]

Pazar Sepet Analizi çalışması başlangıç noktası olarak bu tür faaliyetler için eşsiz denebilecek bir uygulamadır. Yön verilmiş data mining araştırmacısı analiz öncesi bazı informasyonlardan habersiz olması çalışmaya katkı sağlamaktadır diğer bir deyimle projenin yaratıcılığını artırır. Ek olarak data kümesi içinde data yönetimi ve işlemesi adına ihtiyaç olan her şey mevcut. Apriori Algoritması ile farklı ürün kategorilerinde bulunan ürünlerin birbirleri ile bağıntılanması da datalar üzerinde somut sonuçlar elde etmek için Pazar Sepeti Analizi için iyi bir temel sağlamaktadır.[20, 21]

Pazar Sepeti Analizi sayesinde üretilen kural setleri çoğunlukla 3 şekilde bölünür; faydalı, faydasız ve açıklanamayan. Bir işlem adına teklif edilebilecek türde enformasyon faydalı bir kuraldan çıkarılabilir. Bu şekilde yapılmış bir çalışma herhangi bir günde birbiriyle alakasız 2 ürünün birlikte satılması sonucunu verebilir. Bu data ve kural setinin sentez bir şekilde analizi, değerlendirilmesi ve yorumlanması sonucu müşteri davranış eğilimlerinin tespiti de yapılabilmektedir. Birbiriyle ilişkisi hiçbir şekilde yok diye düşünülen 2 ürünün yapılan çalışmalar sonucunda yüksek bir korelasyona sahip olduğu tespit edildiğinde ilgili 2 ürünün

market içi yerleşimi ile muhteşem bir satış fırsatı yakalanmış olur.[20] Fayda içermeyen kural setleri yalın bir şekilde ortak sentez ürünlerini ortak informasyonlardan üretebilir. Örneğin, bahçe ürünleri satın alan müşterinin toprak ürünleriyle de ilgilenmesi normaldir. Bu sebeple fayda içermeyen ve önem arzetmeyen kurallar ihtimal dahilinde bir hareketi konusunda faydalı enformasyon oluşturmaz. Belirli kısıtlar sonucu üretilecek diğer bir farklı problem ise ürün paketinden veya ilginç denilebilecek bir özel pazarlama kampanyası sonucu çıkmasıdır. Analizi uygulamadan önceki pazarlama fırsatları hakkında detaylı informasyon sahibi olmak, belirli fırsatlar veya ek ürünlerin hangi kısıtlar altında alındığını aynı zamanda hangi çıktıların tüketici seçimlerinden dolayı yararlı olacağını göstergesidir. Bundan dolayıdır ki Pazar sepeti analizi daha önce yapılmış pazarlama kampanyaları sayesinde yakalanmış satış rakamlarını hem başarı yönünden hem de getiri yönünden ölçme konusunda çok başarılıdır.[20]

Açıklanamayan kurallar veri madenciliğinde en kötü sayılabilecek durumdur ve yalnızca tavsiye edilen bir hareket planı sağlamazlar, bunu anlaması ve anlamlandırması da bir hayli zordur. Bilgi sağlayan bu kısıtlama altındaki kurallar yararsız ve uygulanmaya ters niteliktedir. Bir örnek verecek olursak şayet; Amerika Birleşik Devletleri'nde gerçekleştirilen bir araştırmada donanım ürünleri ile ilgili açılan bir mağazada en çok WC tuvalet kağıdı kabı satışları olmuştur. Bunun anlamı şöyledir ki; mağaza içi deneyimlerde yapılan satış stratejileri, indirimler ve kampanyalar müşteri satın alma davranışlarının öncesinde iyi bir şekilde analiz edilmediğinde farklı sonuçlar doğuracağını ve bu ön analizin çok önemli bir ihtiyaç olduğunu bize göstermektedir.[21]

2.3.1. Market Basket Analizinin Güçlü ve Zayıf Yönleri

Yönsüz data minin kusursuz derecede olması market sepet analizinin en güçlü yönlerindendir. Bu metot ile büyük ve geniş tabanlı bir data kümesinin hangi yol ve aşamadan başlanacağı konusunda fikir sahibi olmuş oluruz. Data Mining metotlarının büyük bir çoğunluğu herhangi bir yön verilmemiş özelinde kullanılmazken, büyük ve geniş tabanlı bir data kümesini analiz edip user uygunluğu güçlü bir analiz için Pazar Sepeti Analizi hızlı bir şekilde uygulanabilir.[20]

İlk olarak dataların biçimleri ve data içeriklerini kaybetme tehdidi ile başlamalıyız. Her bir parametrenin kendi özelinde bir data türü ve her bir nesnenin saklı kalabileceği bir data parametre türü vardır. Parametreler özelinde esnek yapıya sahip olan parametreler kazanç sağlar çünkü alandan kazanma fırsatı sağlamaktadır. Standart bir norm ile parametre esnekliğinde olan datalar arasında fark mevcutta sayılabilecek karakter sayısı kadardır. Standart uzunluktaki norm için her bir birim özelindeki en geniş alan o geniş alanda yer tutacak kadar geniş olması adına çok daha önceden nitelendirilmiş olması gerekmektedir. Bazı durumlarda daha az yer kaplayan data kayıtları boşuna yer kaplıyor şeklinde görüntü sergileyebilir. Değişken yer kaplayan dataların akabinde her bir datanın kapladığı alan ilgili alanın genişliği ve büyüklüğü kadar da olabilir. İşlem datası konu olduğunda unsurları temsil etmenin en natürel şekli ilgili bu dataları değişken alana sahip olan data türleri ile eşleştirmektir. Çok sayıda izlenen çalışma ve çalışma kayıtları ele alındığında Pazar Analiz Sepeti uygulaması birçok detaylı ve önemli informasyonu hem saklayıp hem de işlemeye devam ettiği gözlemlenmiştir.[19, 20]

Analiz temel olarak ele alındığında ise avantaj sağladığı yön olarak işlerliğinin ve işlem ilerleme çevikliğinin yalın olması söylenebilir. Yapay Sinir Ağları ile kıyaslandığında hesaplamalarda oldukça yalındır ve karşılaştığı küçük sorunlar karşısında oldukça çevik bir çözüm refleksi gösterir.[20]

Ek olarak unsurlar, gerçekleştirilen uygulama adımları bağıntı kısıtları üretmek adına ihtiyaç olan hesaplamalara paralellik gösterir ve eksponansiyel şekilde artar. Problemin sonuca kavuşması adına unsurlarda budama işlemi yapılabilir. Bu unsurları yüksek bir sınıflandırma seviyesinde tutarak kolayca toplanabilir. Genelleştirilmiş unsurlar her daim bir çok farklı hareket refleksi gösterebilir seviyede olmasa da kısıt üretme prosesin kontrolü adına metotlar mevcuttur. Asgari dayanak temizleme buna uygundur denilebilir. İlerleyen kısımlarda bu konuya ilişkin formül ve çeşitli örneklerle detaylı bilgilere yer verilecektir.[21]

Market sepet analizinin temeldeki asıl problemi, doğru ve net bir şekilde toplam düzen seviyesinin belirlenememesidir. İşlemler esnasında bazı veri kayıpları yaşanabilir ve unsurların yüksek frekanslı olağan seviyelerinden farklı bir seyir gösterebilir. İhtimal dahilinde olan çözüm genelleştirilmiş unsurlardan elden giden informasyonları tutabilecek

sanal unsurları eklemektir, ender görülen ihtiyaç da aslında tam olarak budur. [21]

2.3.2. Market Sepet Analizi Data Araştırmaları

Market sepet analizi alışveriş işlemlerini detaylı şekilde irdelemek ve satış okazyonlarını belirlemede faydalı olan önemli bir araçtır. Kural ve kısıtları meydana getirme yöntemleri Apriori ile ana destek-güven normları kullanıldı ve bu destek-güven ölçütleri çalışma yürütücüsü tarafından daha önceden tayin edilmiştir. İlk etapta tayin edilen destek ve güven değerleri olarak; minimum dayanak 0,2 (% 20) ve minimum itimat 0,5 (% 50). Marketlerden(Pazar yeri) dataları kompozit hale getirildi ve her bir markette(Pazar yeri) bu analiz bağımsız olarak işletildi. Devreye alınan bu model hangi unsurların yüksek frekanslı işlemler içine dahil olduğunu tespit etmek amacıyla kullanılır. Ürün adı ve işlem adına ait olan bu yöntem, kısıtlı sayıda parametreye ihtiyaç duymaktadır. Bu sebeple Pazar sepet analizi dataları için bahsi geçen model daha çok seçilir. [17]

Birliktelik kuralı özelinde her bir destek-güven hesaplaması yapılarak belirli bir norma çevrilir. Ayrıca SQL Server’da data mining aracı için Excel’de her bir kural için ihtimal ve değer hesaplaması mevcuttur. Bağıntı kuralları tablosu unsurları arasındaki bağıntılaşma oranlarını(yüzdelere) buradan bulabilir ve erişebiliriz. Bu tablo özelinde 2 kural öne çıkarak önem arz etmektedir; ihtimal ve değer.[17]

İhtimal dahilinde herhangi birbirine benzer aynı ürün grubundan 2 ürün alan müşteri 3. Olarak alacağı ürün yine aynı ürün grubundan olma olasılığını göstermektedir. Diğer söylemle; ihtimal sadece bir kısıtın itimatıdır.[18]

Kısıtın değeri ölçüsü, kısıt başlığının bir kısıt maddesi ile beraber oluşma olasılığının ölçüsüdür. Değer derecesi pozitif ise, kısıt maddesi ve kısıt başlığı birbirinden bağımsız olarak oluşma ihtimali daha yüksektir. Pozitif bir değer puanı kısıt gövdesi kısıt maddesi doğru olmasından mütevellî kısıt başlığının da ihtimalinin arttığını gösterir.[18]

Bağımlılık ağı unsurların birbirleri ile olan bağlantılarının derecesini grafik formunda görülmesini sağlar. Satın alınan malzemelerin yüksek frekans yoğunluğu ile birbirleri ile

ilişkili olduğu ve hangi malzemenin diğer hangi malzeme ile hangi derecede ilişkisi olduğunu basitçe gösterir. [18]

Tablo 2.19: Toplu Datalardan Ürün İlişki Tablosu

İhtimal	Değer	Kısıt
90 %	1.40	Yüz Maskesi, Yüz Kremi -> Güneş Kremi
78 %	0.75	Sabun, Saç Bakım Jeli -> Duş Jeli
72 %	0.70	Sabun, Saç Bakım Yağı -> Duş Jeli
67 %	0.67	Saç Bakım Jeli, Saç Bakım Yağı -> Duş Jeli
66 %	0.66	Yüz Maskesi, Saç Bakım Yağı -> Duş Jeli
61 %	0.64	Saç Bakım Yağı, Ağız Bakım Seti -> Duş Jeli
56 %	0.68	Ağız yıkama, Ağız Bakım Seti -> Ağız Bakım Seti
57 %	0.61	Saç Bakım Yağı, Vücut Kremi -> Duş Jeli
53 %	1.22	Diş Fırçası -> Ağız Bakım Seti
47 %	0.60	Sabun, Ağız Bakım Seti -> Ağız Bakım Seti
43 %	0.54	Saç Bakım Yağı -> Duş Jeli
44 %	0.50	Saç Bakım Yağı, Saç Boyası -> Duş Jeli
41 %	0.54	Sıvı Sabun, Sabun -> Ağız Bakım Seti
39 %	0.53	El kremi, Ağız Bakım Seti-> Ağız Bakım Seti
39 %	0.52	Kağıt Havlu, Diş fırçaları -> Ağız Bakım Seti

Satış oranı ve satın alma davranışına eğilimli olan ürünlerin daha fazla sattığı tespit edilmiştir.

- Bir müşterinin [Yüz Maskesi] ve [Yüz Kremi] satın alması durumunda bunlarla birlikte [Yüz kremi günü] satın alma olasılığının % 90 olduğu tespit edilmiştir..
- Bir müşterinin [Sabun] ve [Saç Bakım Jeli] satın aldığı anda % 78 ihtimalle [Duş Jeli] de alacağı tespit edilmiştir.
- Bir müşterinin [Sabun] ve [Saç Bakım Yağı] aldığı anda [Duş Jeli] de satın alacağı % 72 olasılıkla tespit edilmiştir.
- Bir müşterinin [Saç Bakım Jeli] ve [Saç Bakım Yağı] satın aldığı anda [Duş Jeli] de alacağının % 67 ihtimali olduğu tespit edilmiştir.
- Bir müşterinin [Diş Fırçası] aldığı anda yarısından fazla ihtimalle (% 53) [Ağız Bakım Seti] de alacağı tespit edilmiştir.

Prekande olarak satış yapan satıcılarla birlikte, e-ticaret sitesi üzerinden satış yapan firmalar bu analizlerden faydalanabilir ve yüksek oranda satış fırsatları yakalayabilir.

[Yüz Kremi] ve [Yüz Maskesi] gibi birbirine benzer ve aynı kategoriden sayabileceğimiz ürünleri alan müşteriler aynı benzer kategoriden diğer bir ürün olan [Diş Fırçası] alabilir ve beraberinde de [Ağız Bakım Seti] satın alma davranışı gösterebilir.

Birbirleri içinde ayrılan ve diğer durumlarda benzerlik gösteren ürünlerin market içi yerleşimleri de bu analizler göz önünde bulundurularak yapıldığında hem ikili veya üçlü ürünlerin yerleşimini belirlerken diğer ürünlerin de satın alınmasına teşvik edici bir durumda olabileceğini ispat etmiştir. Örneğin; [Ağız Bakım Seti] satın alan müşterilerin aynı zamanda [Diş Fırçası] gördüğünde onu da satın almak isteyebileceği gibi. [21]

En çok alınan malzemelere bakıldığında;

[Saç Bakım Jeli] - [Sabun] - 252 adet

[Diş fırçası] ve [Ağız Bakım Seti] - 245 adet

[Diş macunu] ve [Sabun] - 204 adet

[Sabun] ve [Saç Spreyi] - 165 adet

[Yüz Maskesi] ve [Yüz Kremi] - 159 adet

[Yüz Kremi] ve [Sabun] - 133 adet

[Yüz Kremi] ve [Ağız Bakım Seti] - 106 adet

[Diş Fırçası] ve [Sabun] - 107 adet

[Bebek Pudrası] ve [Yüz Kremi] - 102 adet

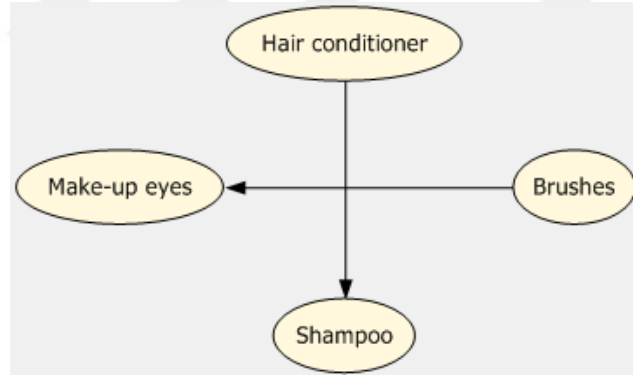
[Ağız Bakım Seti] ve [Diş Fırçası] - 100 adet

Reyon ürün diziliminin başta nasıl olduğuna dair bir veri olmamakla birlikte, birliktelik kurallarına göre elle ettiğimiz verilere göre bu sayede market sahibine raf dizilimi konusunda tavsiyede bulunulabilir. Verilerimiz alışverişlerden elde edilen sonuçlara göre uygulamaya geçirilmiştir.

Tablo 2.20: Market Sepet Analizi Satın Alma Tavsiyeleri

Alınan Ürün	Öneri	Alınan Ürünün Satışı	Bağlı Satışlar	% Bağlı Satış Yüzdesi	Önerilen Ürünlerin Ortalaması	Bağlı Satışların Toplamı
Saç Bakım Jeli	Sabun	611	252	41.41 %	1	863
Diş Fırçası	Ağız Bakım Seti	200	100	50 %	2	300

Market Sepet Analizi tavsiye raporunda ürünler arası bağlantıların nasıl olduğu ve birim satış yapan firmalar için faydalı olabilecek tavsiyelerde bulunur. Her bağlantı setinde ürünler özelinde birbirini destekleyecek şekilde sayısal değerlerin olduğu istatistiki veriler olur ve setler arasında ihtimal dahilindeki sınır aşımında dikkate alınır. Mevcudu değerlendirdiğimizde tavsiye raporunda {[Saç Bakım Jeli] ve [Sabun]} ve [Diş Fırçası] ve [Ağız Bakım Seti] gruplarındaki malzemelerin satışı söz konusu olduğunda satışı artışına yönelik bir eğilim göstereceğini tavsiye etmektedir.[19]

**Şekil 2.7:** İlişkili Ürün Sıralaması Örneği**Tablo 2.21:** Birliktelik Kuralı ve Ürün Raflandırması

İhtimal	Değer	Beraberlik Kısıtı
44 %	0.53	Saç Bakım Jeli -> Sabun



Şekil 2.8: Satışta Etkili Ürün İlişkisi Örneği

• Data Mining algoritması dahilinde sadece tek bir kısıtlama üretildi. Bu kısıta göre müşteri [Saç Bakım Jeli] alırsa şayet [Sabun] satın alma olasılığı %44 olmaktadır. Bu veri sadece kendi başına anlaşılır ve işlenir bir enformasyon girdisi sağlamayan değersiz bir kısıttır. Satıcının tek bir kural üzerinden böyle bir strateji vermesi anlamsız kalmaktadır. [20]

• Satıcının bu sebepten dolayı satışlarının az olması düşük kar marjı nedeniye dükkanını kapatması olağandır bir durumdur. Minimum aşama değerine bakıldığında tek bir bağlantı kurulmasının ve tek bir kurulmasının nedeninin mantığını anlayabiliriz. Tüm yapılan çalışmalarda minimum destek %40 olarak belirlendi ve bu değer altındaki belirlenen kurallar ortak değerlendirilmemektedir. Satış oranını ortalama olarak etkileyen birbiriyle ilişkili ürünlerin yüksek güven ve destek kaideleri ile ilgilenilir. Tablo 2.22’de bakıldığında seçili ürün olan saç bakım jelinin satışı ile birlikte satışı artırması ve bağlı satışı artırması ön görülen sabunun detaylı hesaplaması yer almaktadır.[20]

Tablo 2.22: Market Satın Alma Tavsiyeleri

Seçilen Ürün	Tavsiye	Seçili Ürünün Satış Adedi	Bağlı Satış Adedi	% Bağlı Satışları n Yüzdesi	Referans Değeri	Beraberlik Değeri
Saç Bakım Jeli	Sabun	35	21	60 %	1.64	56

Bu market özelinde saç bakım jelinin sabununla satmak arasında bağlı satışların yüzdesi %60,

bağlı satışların genel değeri ise 65'tir.

Kural-Kısıta Rastlanmadı -> Yüksek Bağlantı Değeri Yoktur

• Çalışmanın temel amaçlarından bir tanesi de araştırmanın temelinde kuralları bulmak ve uygunluğunu aramaktır. Asgari destek aşaması %10 ve asgari güven aşamasının %40 olarak belirlendiği durumla birlikte herhangi bir 3. Ürün ile tarayıcı datası ile yapılan işlemde bir ilişki bulunamamıştır. Yeterli sayıda uygulama yapılmış olmasına karşın halihazırda bulunan herhangi bir kural setine ise rastlanmamıştır. Bunun ana sebebi ise daha önceden tanımlanmış olan güven ve destek aşamalarından dolayı bu kısıta takılmaktır. Market Sepet Analizi özelinde mevcudun talebine karşı uygun miktarda işlem görülmesine karşın müşteriler belirli ürünleri beraber satın alırken bazı ürünleri de birlikte almayı tercih etmemekte. Bununla birlikte tahmin edildiği kadar yüksek frekanslı olarak satın almıyorlar. [20]

• Diğer bir örnekle birlikte alınabilecek ürünler;

- [Saksı] ve [Gübre]
- [Fındık] ve [Ceviz]
- [Çöp Torbası] ve [Tuvalet Kağıdı]
- [Puding] ve [Süt]
- [Meyve suyu] ve [Büskivi]
- [Bulgur] ve [Makarna]
- [Şeker] ve [Çay]

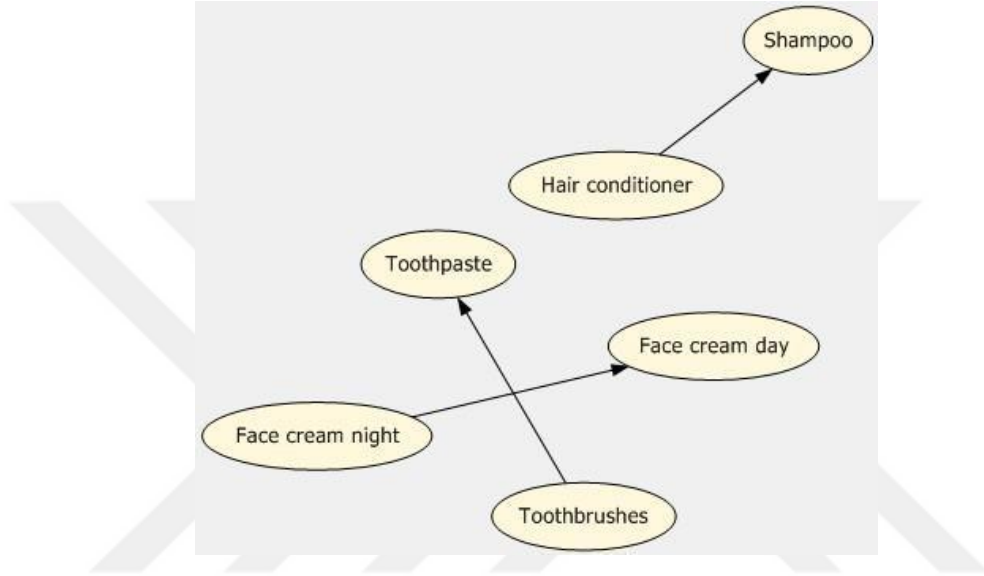
Tablo 2.23: Mağaza İçin Alışveriş Sepeti Önerileri

Alınan Ürün	Referans	Alınan Ürün Satışları	Birlik Satışlar	%	Tahmini Tavsiye Değeri	Birlik Ürün Satışları
Diş Fırçası	Ağız Bakım Seti	28	12	42.00 %	1.5	40

• Bu araştırma tavsiyesi olarak [Diş Fırçası] ve [Ağız Bakım Seti]'ni %42 düzeyinde beraber satış yapılması tavsiye edilmektedir.[20]

Tablo 2.24: Market Sepet Analizi Tavsiyeleri

İhtimal	Değer Derecesi	Birliktelik Kuralı
65 %	0.92	Saç Bakım Jeli -> Sabun
55 %	0.93	Yüz Maskesi -> Yüz Kremi
44 %	0.80	Diş Fırçası -> Diş Macunu



Şekil 2.9: Birbiriyle İlişkili Ürünler ve Raf Yeri Belirlemesi

- Diğer bir market reyonu ise kozmetik ürün grubunun olduğu raflar ve orada da birbirinin satışlarını doğrusal şekilde artış etkisi bulunan ürün olduğu tespit edilmiştir. Bunun yanı sıra marketin bulunmuş olduğu konum ve satın alma davranışları da bunda bir o kadar etkilidir. Mevcut kurallar özelinde değerlendirildiğinde önceki örneklerle benzerlikler göstermiştir;
- Müşteri [Saç Bakım Jeli] alırsa,% 65 ihtimalle [Sabun] da alacaktır.
- Müşteri [Yüz Kremi] alırsa [Yüz Maskesi] 'yi alma ihtimali ise %55'tir.[20]

3. MALZEME VE YÖNTEM

3.1. WEKA YAZILIMI

1996 yılında ilk kez resmi şekilde Waikato Üniversitesi merkezli piyasaya sürülen WEKA programı hem “machine learning” hem de bir “data mining” yazılımıdır. Bilimsel çalışmalardan, akademik eğitime, pek çok endüstri de dahil olmak üzere pek çok farklı alanda kullanımı olan WEKA yazılımı temelinde data incelemesi(analizi) ve yazılımsal boyutta pek çok tahminleyici modelin dahilinde algoritmaların bulunduğu bir platform olarak düşünülebilir esasında. Mobil bir sisteme sahip olan WEKA, platformlar arası taşınabilirliği sayesinde geniş bantlı data izleme ve modellemesiyle kullanıcı dostu bir arayüze sahiptir. Bunlara ek olarak JAVA ile de beraber çalışabilir durumdadır bu da WEKA’nın mobiliteliğini desteklemektedir. Tez çalışmasının teori bölümünde de konusu geçen data işleme ve data mining algoritmaları da WEKA’nın yapısal sıralamasında da etkili olmuştur.

“Waikato Environment for Knowledge Analysis” baş harflerini temsilen oluşturulan kısaltmadan adını alan WEKA, makine dili kapsamında işlenmiş yazılım araçlarından biridir. JAVA yazılım dili özelinde açık kaynak kodlu olarak yazılan WEKA ek olarak lisanslı olarak da paylaşımı yapılmaktadır.

Yalın bir dosyadan dataları işleyen ve buradaki parametreleri nümerik ve non-numeric olarak kabul edip işlemektedir. Bununla birlikte database’den de data alabilir ama bunun mümkün olabilmesi için data dosyasının bir doküman datası formatında olması gerekmektedir.

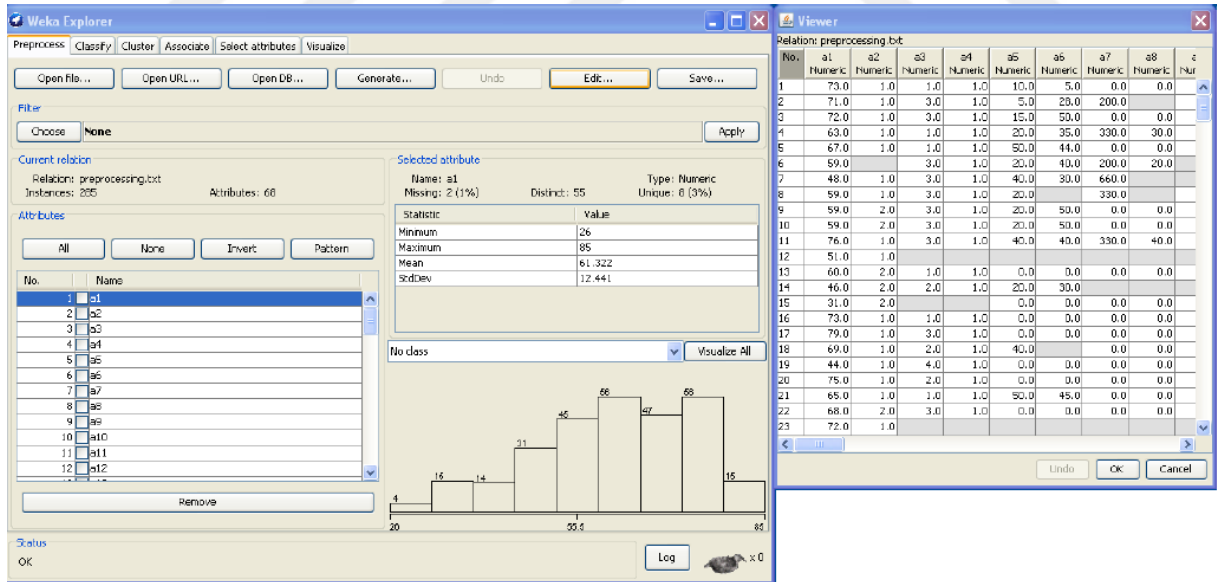
Data ön işleme, sınıflandırma, regresyon ve özellik kriterleme gibi birbirinden farklı matematiksel ve istatistiksel kütüphane WEKA’da makine öğrenmesi tabanlı olarak hazır bulunmaktadır. Tüm bunların nihayetinde ise yapılan proje ve çalışmaları görsel olarak yansıtan bir özelliği de vardır. Bu sayede işlem, uygulama ve sonuçlandırma sürecini bütünsel olarak başarı ile sağlamış olur.

3.1.1. Data Önışleme

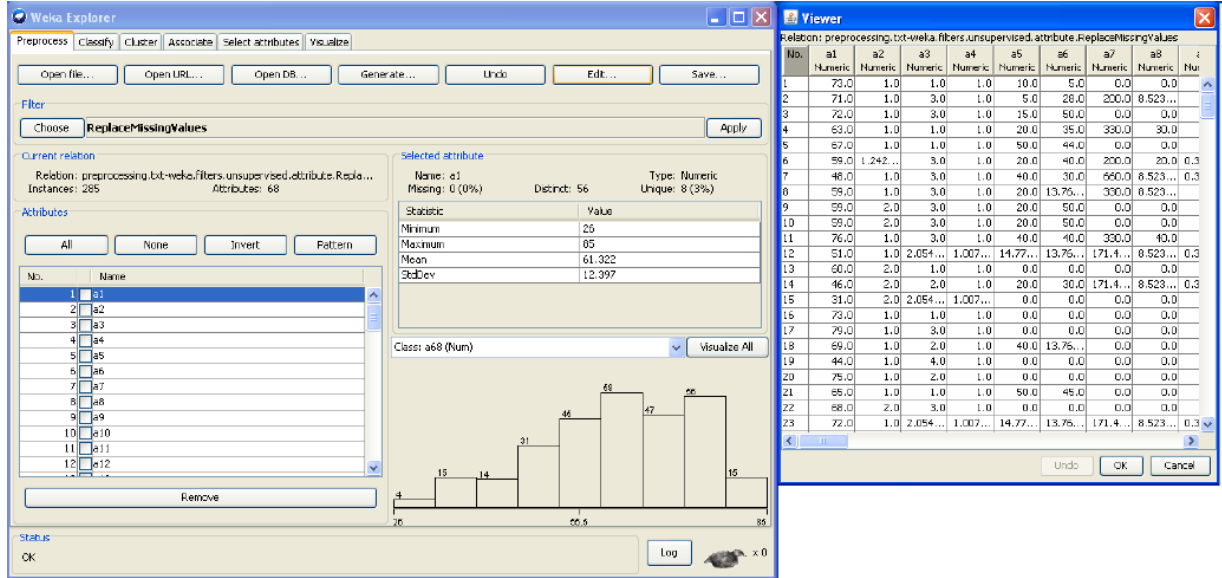
Data önışleme geniş birtabana yayılmış datanın üzerinde bulunduđu database ya da data romlardaki dataları işlem ve uygulama öncesi istatistiki olarak yalınlaştırmayı hedeflemektedir. Data üzerinde farkında olmadan bilgi keşfi yapılan bu uygulamada; data seti üzerinde bir takım düzeltmeler, eksik daha var ise ekleme, mükerrer dataları temizleme, transforme etme, cleaning, normalizasyon, boyut revizesi gibi işlemler yapılır.

3.1.2. Kayıp Datalar

Kayıp datalar üzerinde yapılacak işlemler, üzerinde çalışılan projenin sağlıklı sonuçlara ulaşabilmesi adına büyük bir handikap yaratmaktadır. Bu sebeple kayıp dataların belirli işlemlerden geçerek düzeltilmesi gerekmektedir. Bu vesileyle WEKA’da bulunan Replace Missing Values modülü sayesinde kayıp datalar düzenlenebilmektedir. Bu metot sayesinde databasedeki kayıp datalar içindeki değerler sahip olması gereken niteliğin diğer datalar içindeki değerler ile değişmektedir ve böylece full data set düzenlenmiş olur.



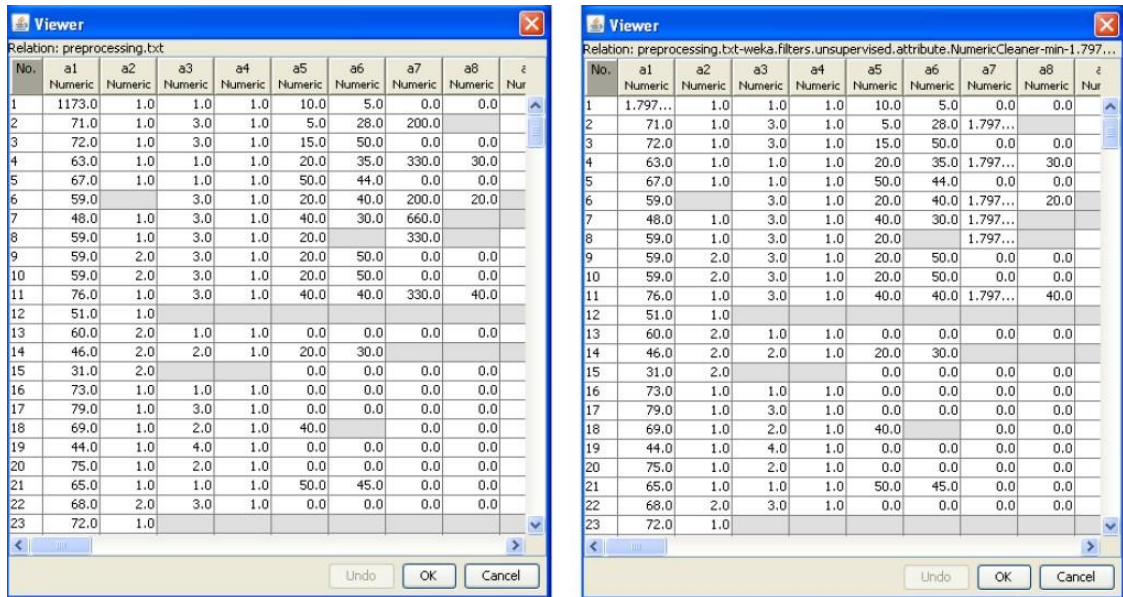
Şekil 3.1: Database Kaybolan Datalar



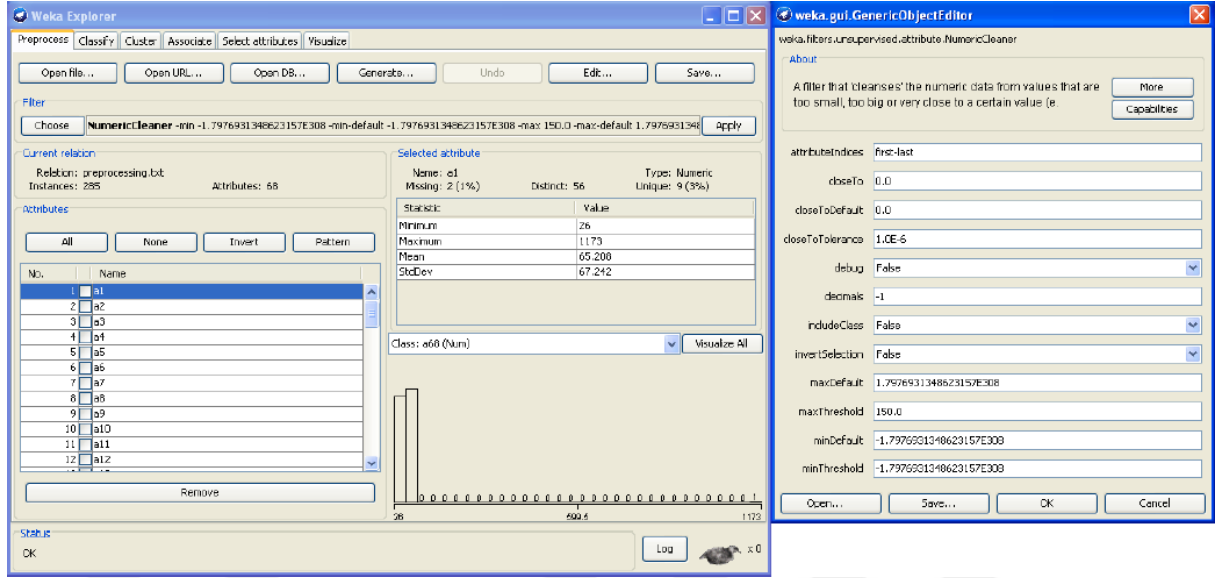
Şekil 3.2: Replace Missing Values Modülünün Kullanımı

3.1.2.1. Yanlış veya aşırı datalar

Data seti içindeki yanlış veya aşırı olarak nitelendirilen bu dataların düzenlenmesi ve temizlenmesi için Numeric Cleaner modülü kullanılmaktadır. Bu metot boyut olarak belli bir form değerine sahip olan datanın databaseden temizlenmesi ile ilgili dataya daha önce kararlaştırılmış olan diğer bir değerin yerleştirilmesini sağlar.



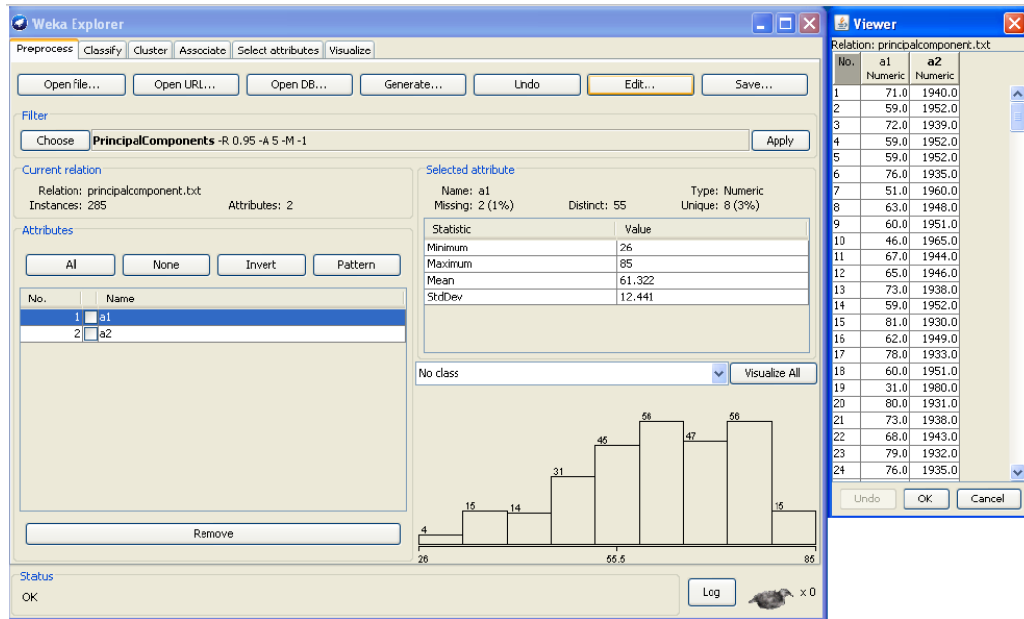
Şekil 3.3: Aşırı uç verilerin NumericCleaner Modülü Kullanılmadan Önce ve Sonraki Durumu



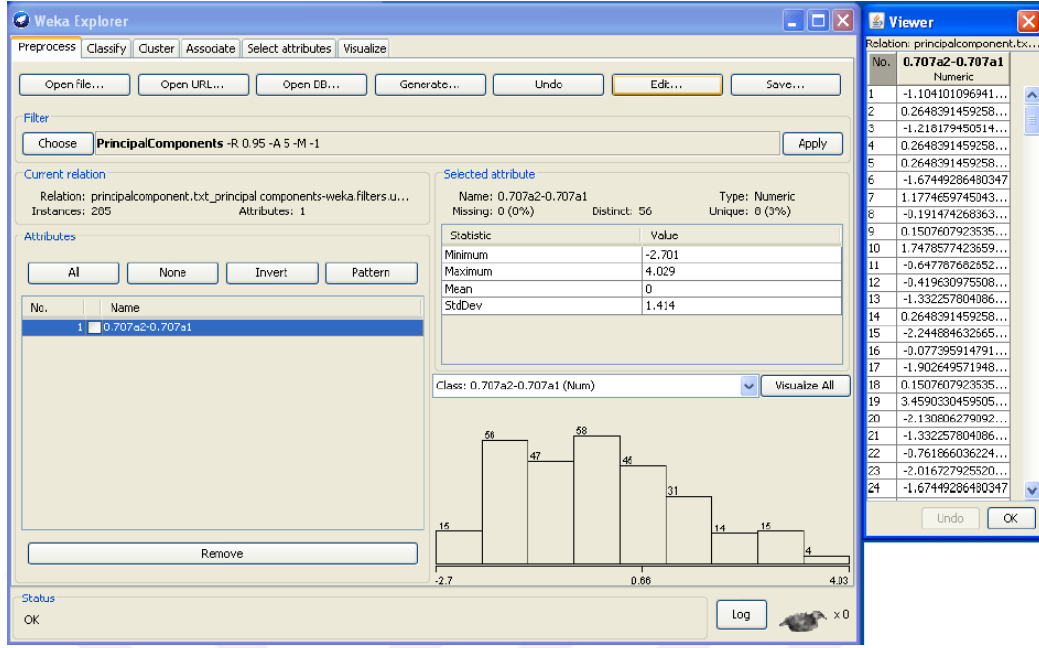
Şekil 3.4: Numeric Cleaner Modülünün Kullanımı

3.1.2.2. Gereksiz datalar

Aynı database içinde birbirini destekler nitelikteki 2 datanın var olması demek gereksiz dataların yer işgal etmesi demektir, bu durum da data işleme süresini uzatıp uygulamayı geciktirmektedir. (Örneğin; hem yaş hem doğum tarihi datası). Bu şekilde bir duruma mahal vermemek adına Principal Components modülü devreye girmektedir ve datanın temizliği ile kaplanan hacim indirgenir.



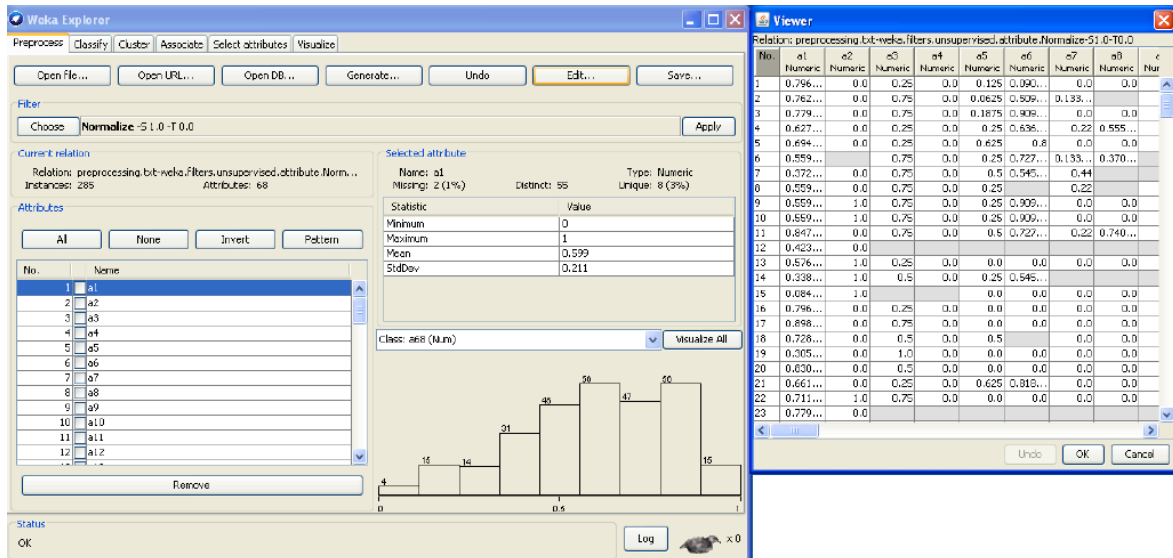
Şekil 3.5: Principal Components Uygulama Öncesi Analizi



Şekil 3.6: Principal Components ile Boyut Küçültme

3.1.2.3. Değişken değerlerinin bağımsız olduğu durumlar

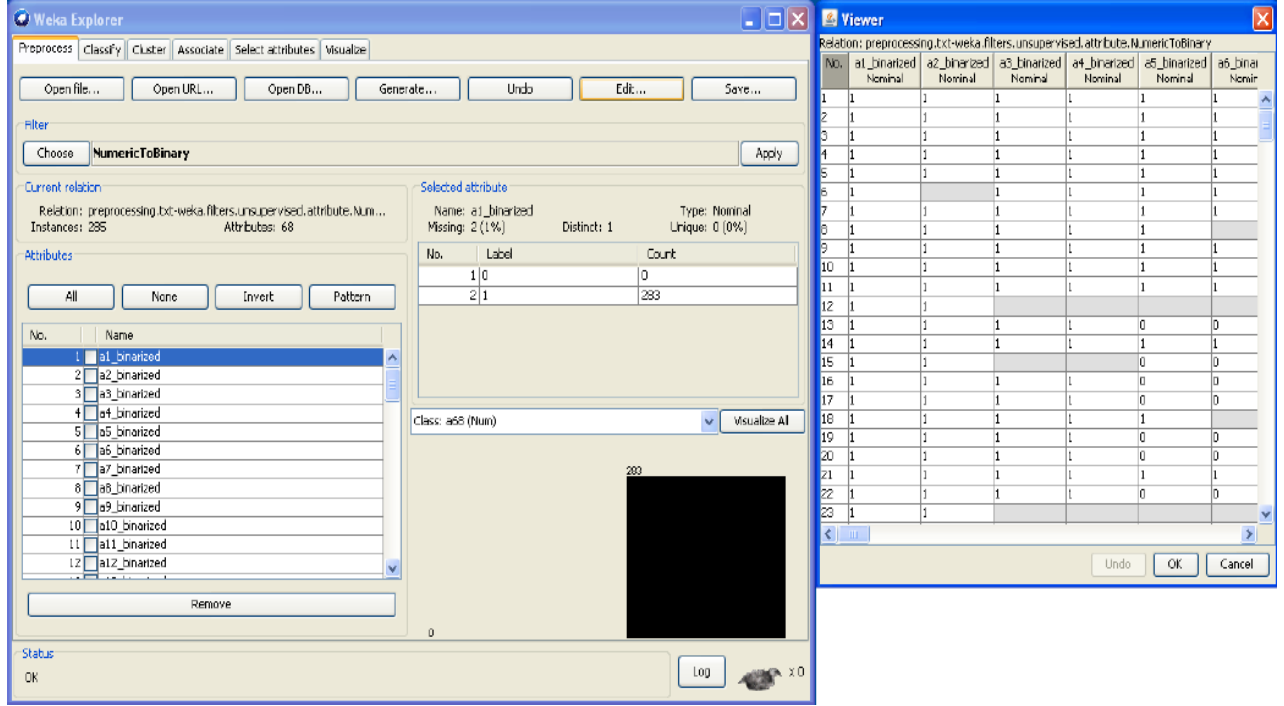
WEKA’da bulunan bir diğer modül ise Normalize modülüdür. Parametre değerlerinin bağımsızlık pozisyon gösterdiği yani ayrıklaştığında bu modül devreye alınarak “nümerik” olan değerler arası mesafe azaltılarak normalizasyon işlemi uygulanmaktadır.



Şekil 3.7: Normalize Modülünün Kullanımı

3.1.2.4. Kullanılacak algoritmaya bağlı yapılandırmalar

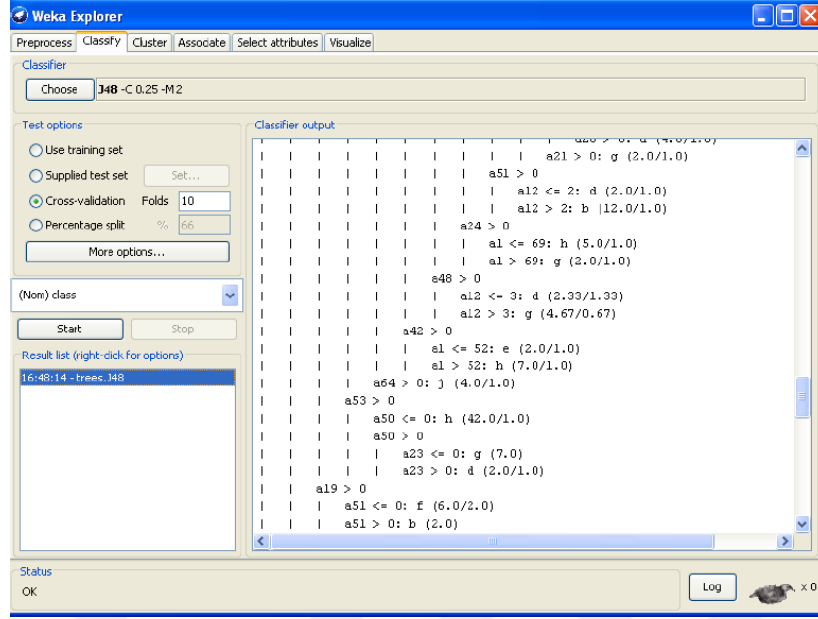
Datanın sağlıklı işlenmesi önündeki en büyük engellerden birisi olan algoritmaya dayalı üretilen kuralların data işlemlerini zorlaştırılması. Bu ve bunun gibi durumlarda Numeric to Binary modülü devreye alınarak databasedeki değerler 0-1 yani binary formatına evrilir.



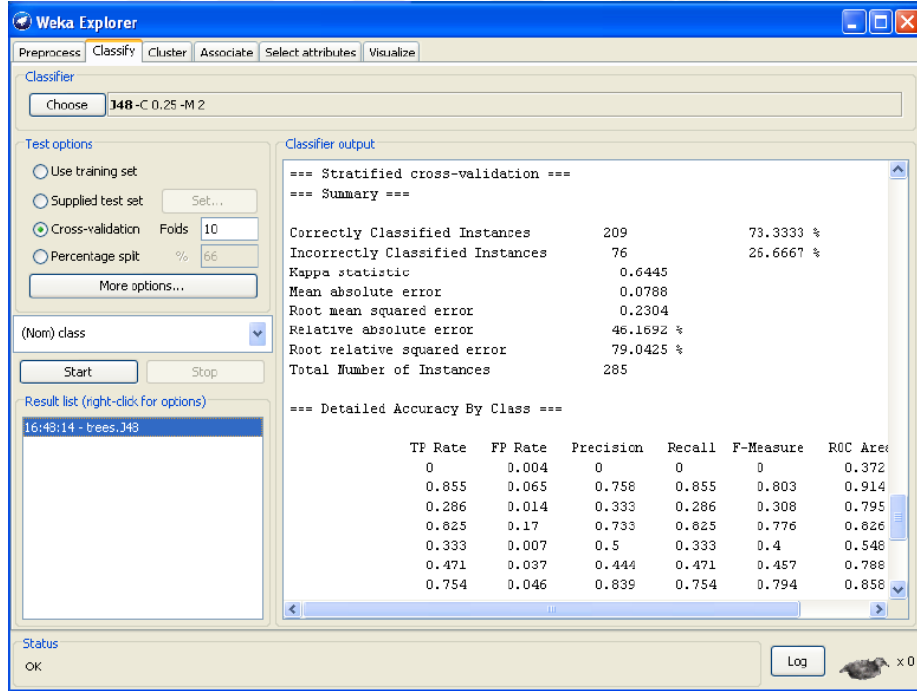
Şekil 3.8: Numeric To Binary Modülünün Kullanımı

3.1.3. Sınıflandırma

Veri ön işleme sonrası sınıflandırma işlemi başlayabilir. WEKA'da bunun için pek çok araç vardır; karar ağaçları, SVM, MLP, J48 vb.. Data seti yüklendikten sonra “Classify” tıklayarak bu araçlara erişilebilir. J48 modülü kullanılarak örnek bir gösterim Şekil 18’de yer almaktadır.



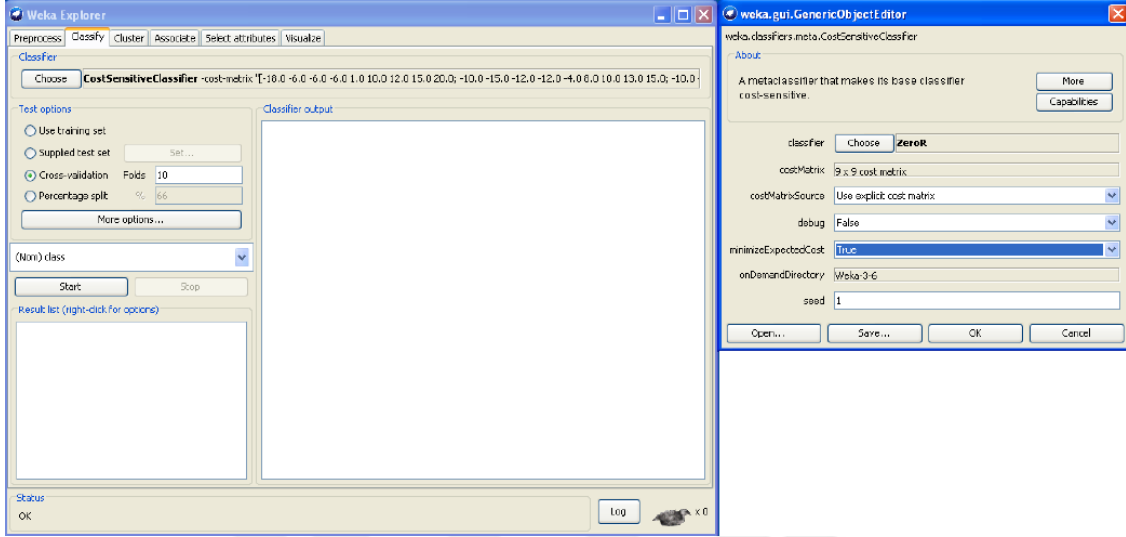
Şekil 3.9: J48 Modülünün Kullanımı ile Elde Edilen Kurallar



Şekil 3.10: J48 Modülü ile Elde Edilen Sınıflandırma Doğrulukları

Data ön işleme sonrası uygulanan sınıflandırma işlemi boyunca ve sonrasında en dikkat edilmesi gereken noktalardan birisi yanlış uygulama sonrası oluşabilecek ağır maliyetler ve maliyetlendirmelerdir. Hatalı sınıflandırma sonrası oluşan maliyetin göreceli bir şekilde hatalı

sınıflandırmada daha yüksek tehlike arz ettiğini bilmek gerekir. Bu denli zarar marjı yüksek olan bir hatanın önüne geçmek için Cost Sensitive Classifier modülü devreye alınır ve bu sayede oluşacak sınıflandırma maliyeti en küçük değere indirgenir ve en doğru tahminleme yapılmış olur.



Şekil 3.11: Cost Sensitive Classifier Modülünün Kullanımı

16:11:01 - meta.CostSensitiveClassifier

Correctly Classified Instances	103	36.1404 %
Incorrectly Classified Instances	182	63.8596 %
Kappa statistic	0	
Mean absolute error	0.1410	
Root mean squared error	0.3767	
Relative absolute error	63.1086 %	
Root relative squared error	129.2496 %	
Total Number of Instances	285	

--- Detailed Accuracy By Class ---

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	0.5	a
	0	0	0	0	0	0.5	b
	0	0	0	0	0	0.5	c
	1	1	0.361	1	0.531	0.5	d
	0	0	0	0	0	0.5	e
	0	0	0	0	0	0.5	g
	0	0	0	0	0	0.5	h
	0	0	0	0	0	0.5	j
	0	0	0	0	0	0.5	f
Weighted Avg.	0.361	0.361	0.131	0.361	0.192	0.5	

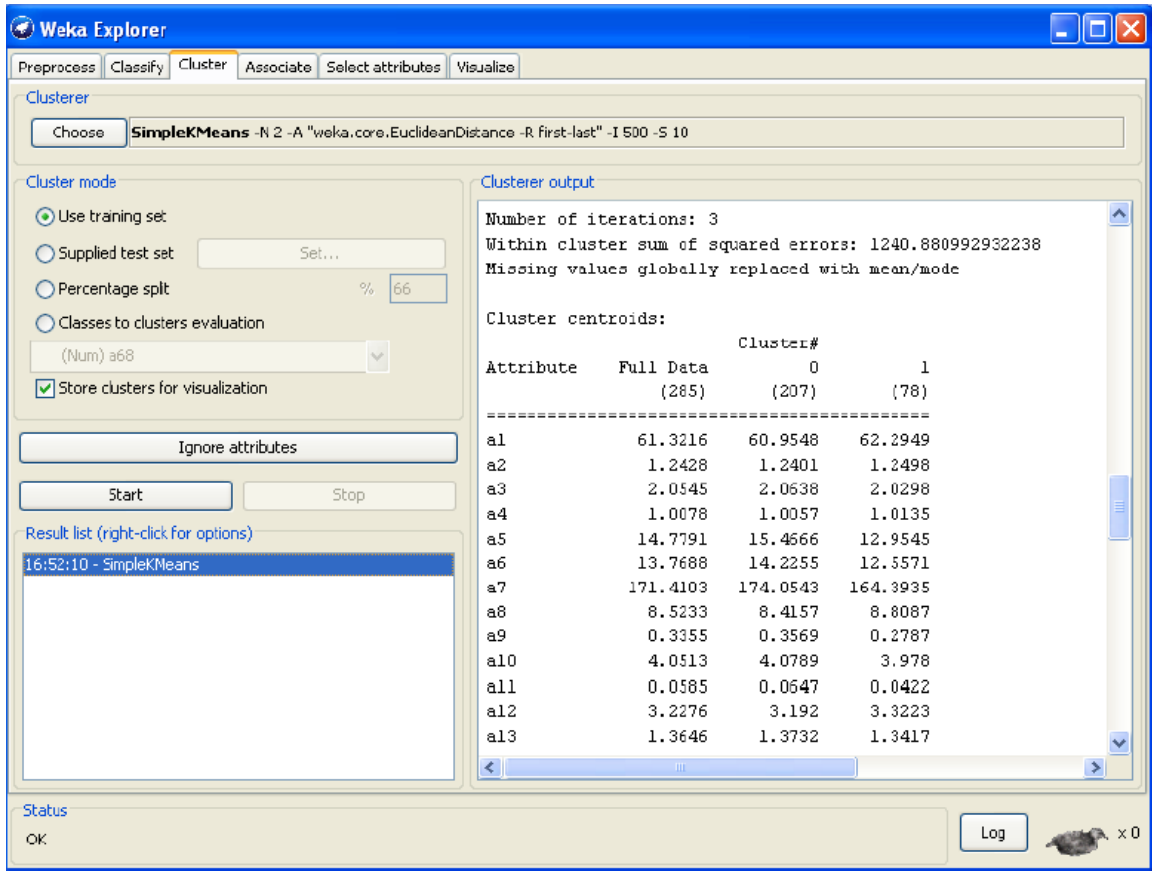
=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	←-- classified as
a	0	0	0	3	0	0	0	0	0	a = a
b	0	0	0	55	0	0	0	0	0	b = b
c	0	0	0	7	0	0	0	0	0	c = c
d	0	0	0	103	0	0	0	0	0	d = d
e	0	0	0	6	0	0	0	0	0	e = e
f	0	0	0	17	0	0	0	0	0	f = g
g	0	0	0	59	0	0	0	0	0	g = h
h	0	0	0	19	0	0	0	0	0	h = j
i	0	0	0	6	0	0	0	0	0	i = f

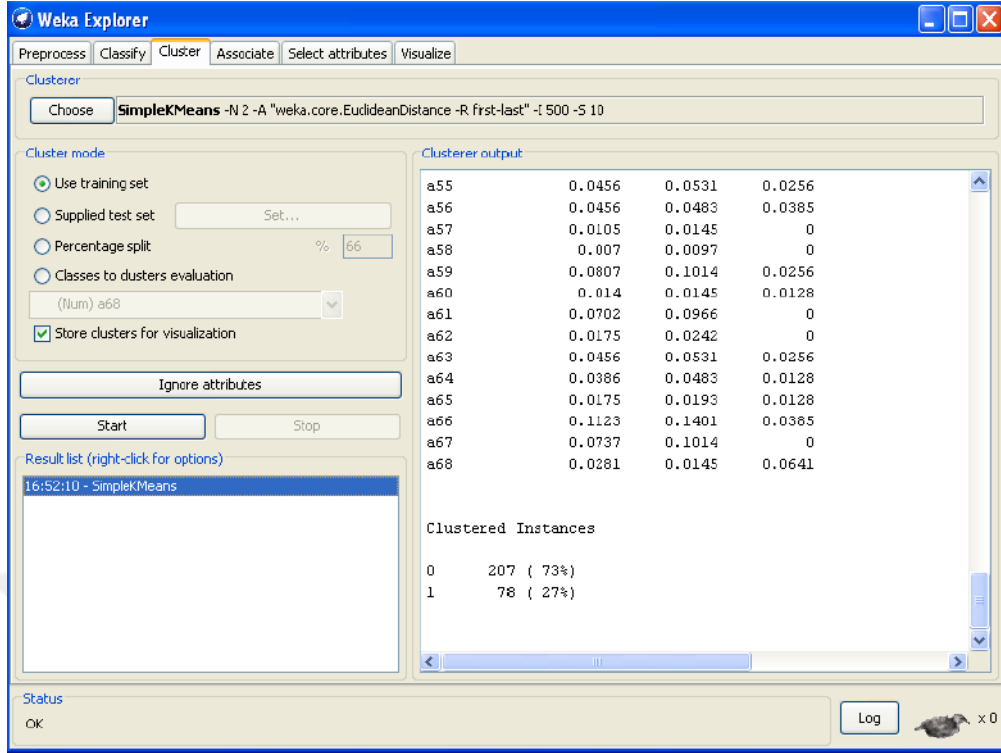
Şekil 3.12: Cost Sensitive Classifier Modülünün Kullanımı ile Elde Edilen Sınıflandırma

3.1.4. Kümeleme

Bir data kümesindeki dataların benzer niteliklerine göre gruplandırılmasına kümeleme denir. K-Means metodu ile n adet datalı kümesini, önceden kararlaştırılan k adet kümeye böler. Bu K-Means algoritmasının işlemesi ve uygulanması diğer metotlara göre daha çeviktir ve kesin bir kümeleme yapar.



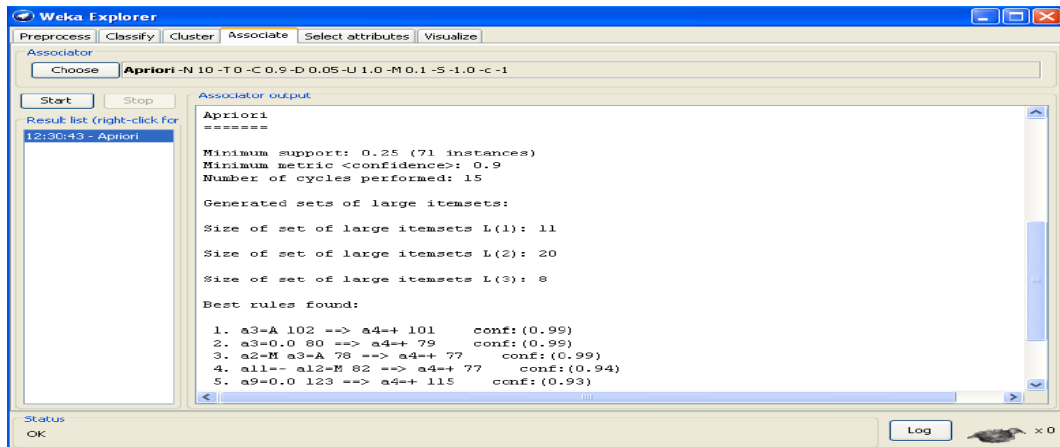
Şekil 3.13: Simple-KMeans Uygulaması Sonrası Kümeleme Bilgileri



Şekil 3.14: Simple-Kmeans Modülünün Uygulaması Sonrası Kümeleme Oranları

3.1.5. Birliktelik Kuralları

Belirli bir data setindeki dataların birliktelik olması ve/veya birlikte olmama ihtimallerinin seçili metotlar ve algoritmalar ile bulunması ve tespit edilmesidir. Belirli güven aralığı çerçevesinde talep edilen sayıda kuralı yakalayana kadar minimum desteği kademeli olarak indirgeyen Apriori algoritması uygulaması uygulaması Şekil 3.15'te gösterilmiştir.



Şekil 3.15: Apriori Modülü Uygulaması Sonrası Birliktelik Kuralları

3.2. DATA SETİNİN TANITILMASI

İstanbul'un Beyoğlu ilçesinde hizmet veren Eren Market'in satışlarından oluşan market verisi uygulama için kullanılmıştır.

Bu veriler 2 aylık süreç sonunda Microsoft Exel yardımıyla kaydedilmiştir. Veriler kaydedilirken müşterilerin sepetlerinde ki doluluğa göre ürün kaydı gerçekleştirilmiştir. Yani en az 4 çeşit ürün alan müşterilerin aldıkları, bir sepet olarak değerlendirilip kayıt tutulmuştur. Toplamda 309 müşterinin 13 farklı ürün grubundan hangilerini aldıklarının kaydı gerçekleştirilmiştir.

Bu bilgilerin alınmasında market sahibi olan Mehmet Ali ÇATAL'dan yazılı onam alınmıştır.

Tablo 3.1: Data Setinin Tanıtılması

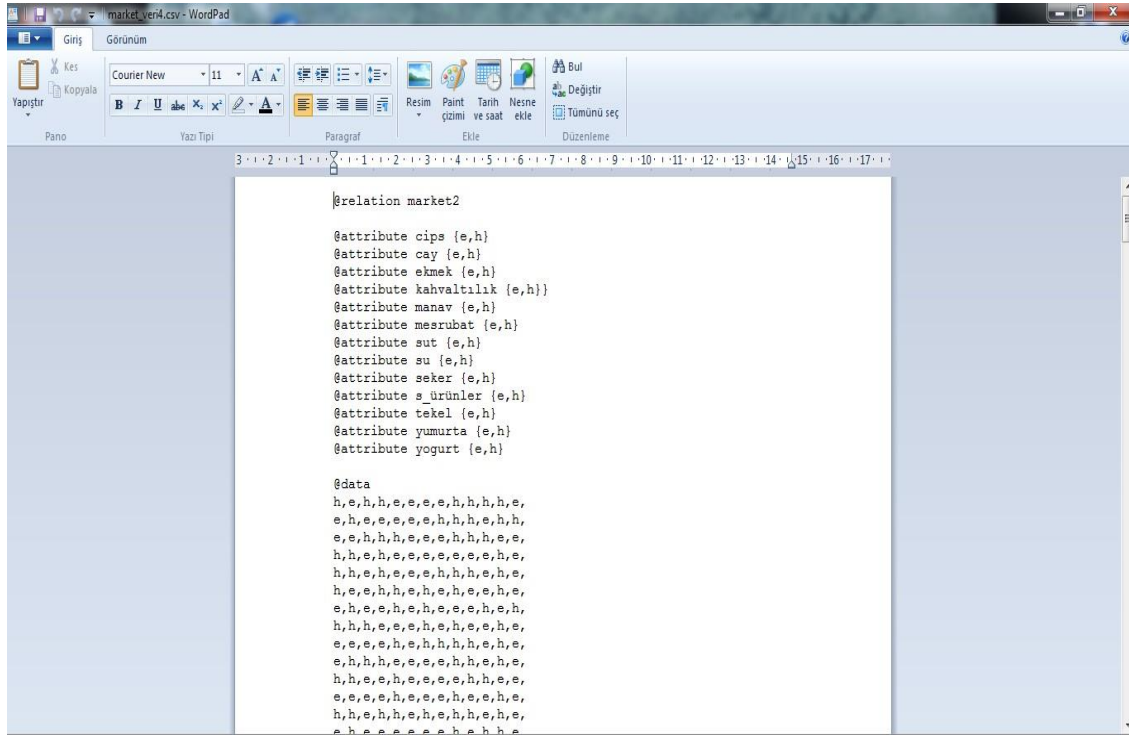
Öznitelik adı	Veri tipi
Cips	Kategorik: evet hayır
Çay	Kategorik: evet hayır
Ekmek	Kategorik: evet hayır
Kahvaltılık	Kategorik:

	evet hayır
Manav	Kategorik: evet hayır
Meşrubat	Kategorik: evet hayır
Süt	Kategorik: evet hayır
Su	Kategorik: evet hayır
Şeker	Kategorik: evet hayır
Ş_ürünler	Kategorik: evet hayır
Tekel	Kategorik:

	evet hayır
Yumurta	Kategorik: evet hayır
Yoğurt	Kategorik: evet hayır

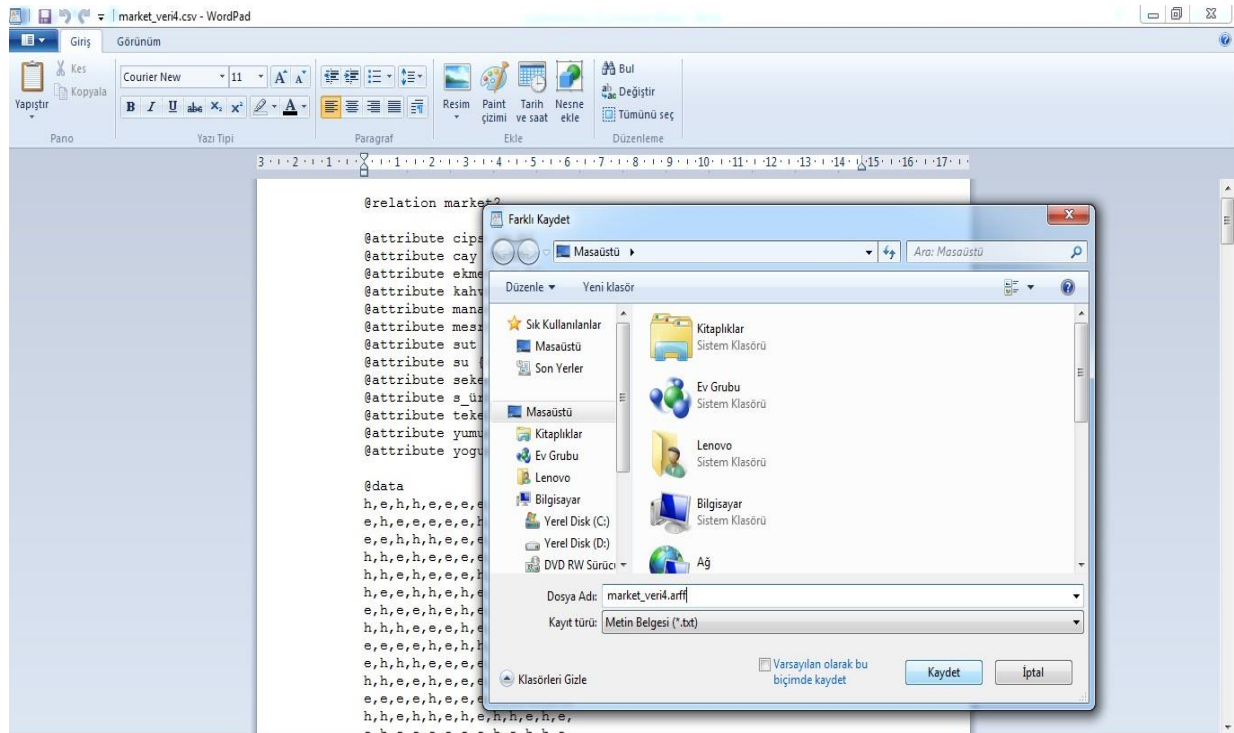
3.3. DATANIN HAZIRLANMASI

Veri madenciliği uygulamalarına geçmeden öncelikle veriyi weka programına uygun hale getirmemiz gerekli. İlk olarak .csv formatındaki veri dosyasını Wordpad’de açarak tüm ‘;’ noktalamalarını ‘,’ yaptık. Sonra veri, Şekil 3.16’daki gibi düzenlenir.

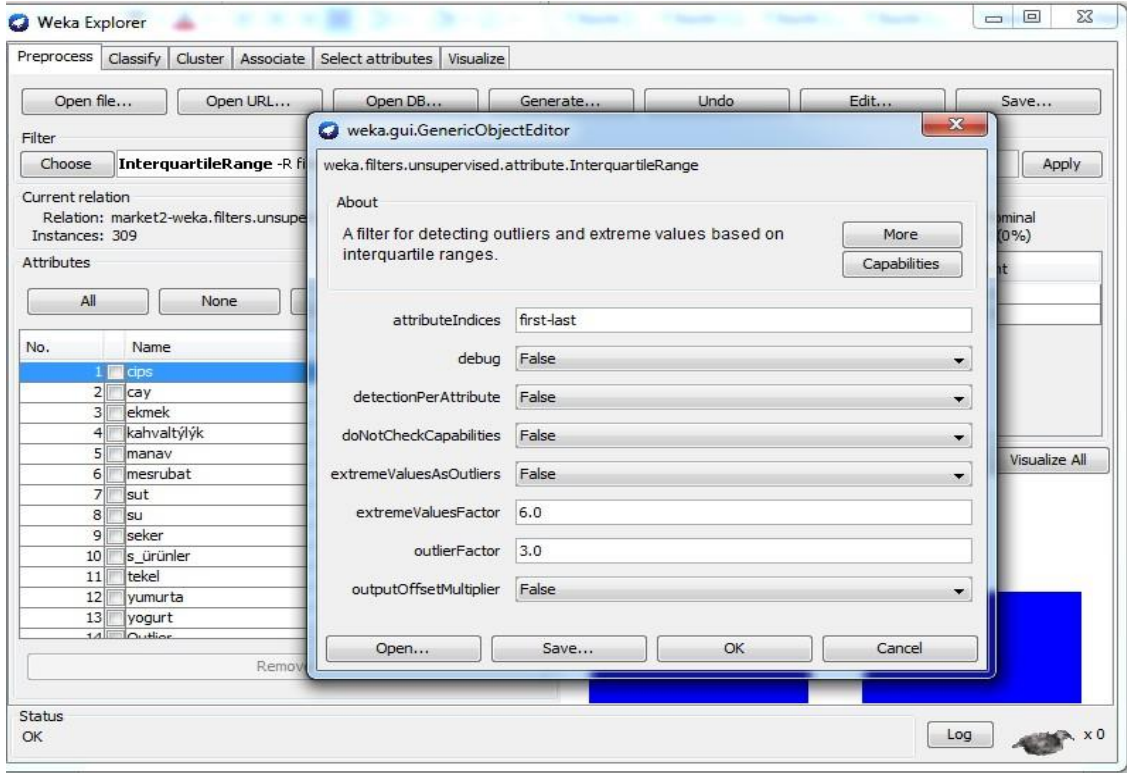


Şekil 3.16: Datanın Wordpad’de düzenlenmesi

Daha sonra dosyayı farklı kaydet diyerek WEKA programının uzantısı olan .arff formatında kaydederiz (Şekil 3.17).

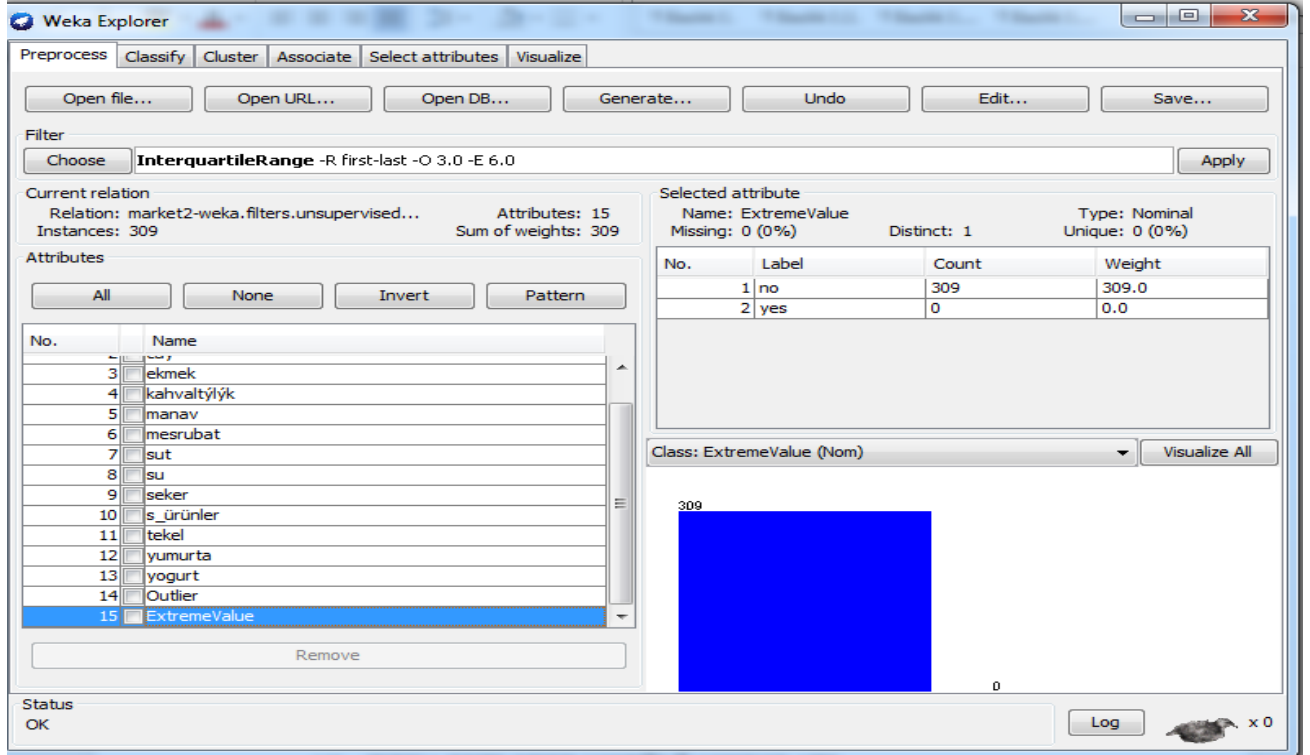


Sonra InterQuarterRange fonksiyonu ile veride aşırı değerler olup olmadığının kontrolü yapılmıştır(Şekil 3.18).



Şekil 3.18: Aşırı Değerlerin Kontrolü

Burada ExtremeValuesFactor için genellikle kullanılan 6, OutlierFactor için ise 3 değerini alınmıştır. AttributeIndices'i 'first-last' olarak almamız tüm verileri kontrol ettiğimizi gösteriyor.



Şekil 3.19: Aşırı Değerler

Şekil 3.19’da görüldüğü üzere aşırı değer saptanmamıştır.

Görüldüğü üzere veri ilk önce Weka yazılımına uygun hale getirildi ve daha sonra da aşırı değerler içerip içermediği kontrol edilerek son halini almış oldu.

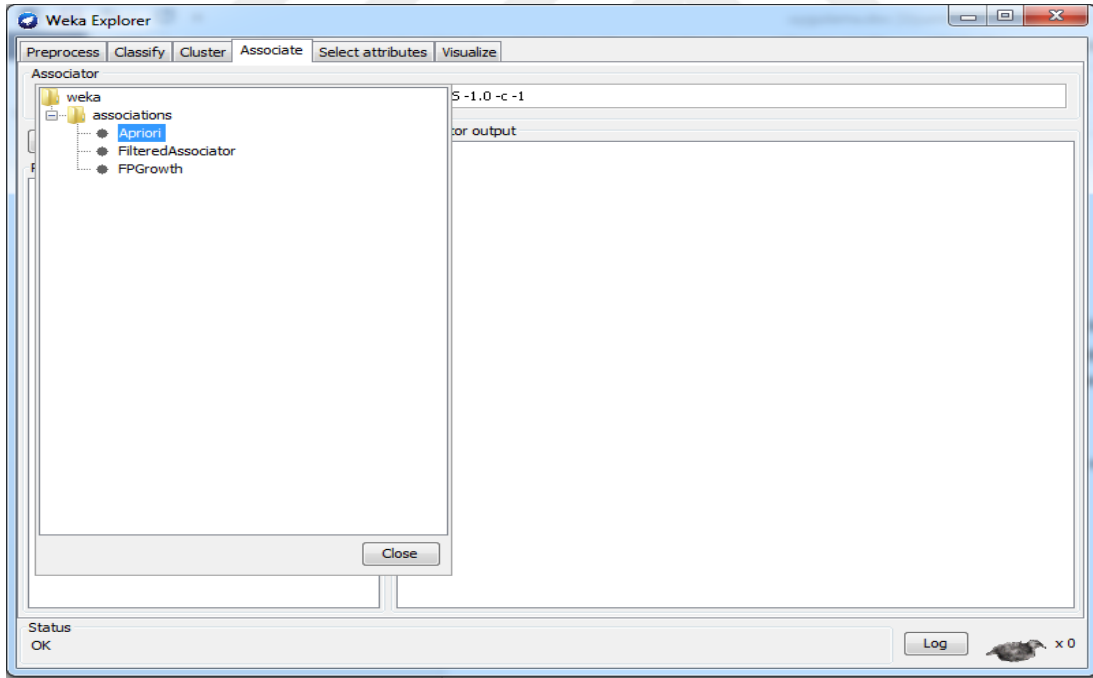
4. BULGULAR

4.1. WEKA EXPLORER MODÜLÜ SONUÇLARI

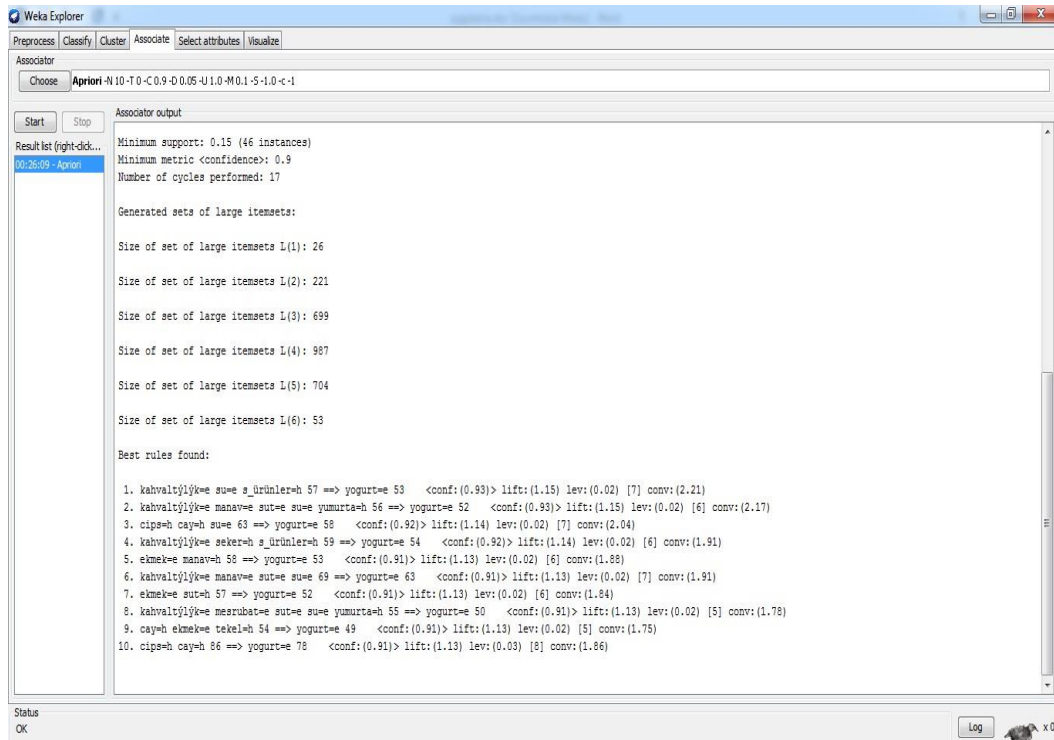
Dataların analizi ve projenin uygulanabilir bir şekilde hayata geçirilmesi için açık kaynak kodlu Waikato Üniversitesi menşeli WEKA yapay zeka yazılım uygulaması kullanılmıştır. Bu projede on basamaklı çapraz tasdik yöntemine göre, eğitim ve deneme kümelerine bölünmüş datalar üzerinde çok bilinen algoritmalara göre 309 adet veri analiz edilmiştir.

Wekada birliktelik analizi yapmak için Associate menüsündeki algoritmalarından yararlanacağız. Bu algoritmalar; Apriori Algoritması, FilteredAssociator Algoritması ve FPGrowth Algoritması'dır.

Apriori algoritmasının kullanılmasıyla (Şekil 4.1) elde edilen sonuçlar Şekil 4.2'de gösterilmiştir.



Şekil 4.1: Weka'da Apriori Algoritması'nın seçilmesi



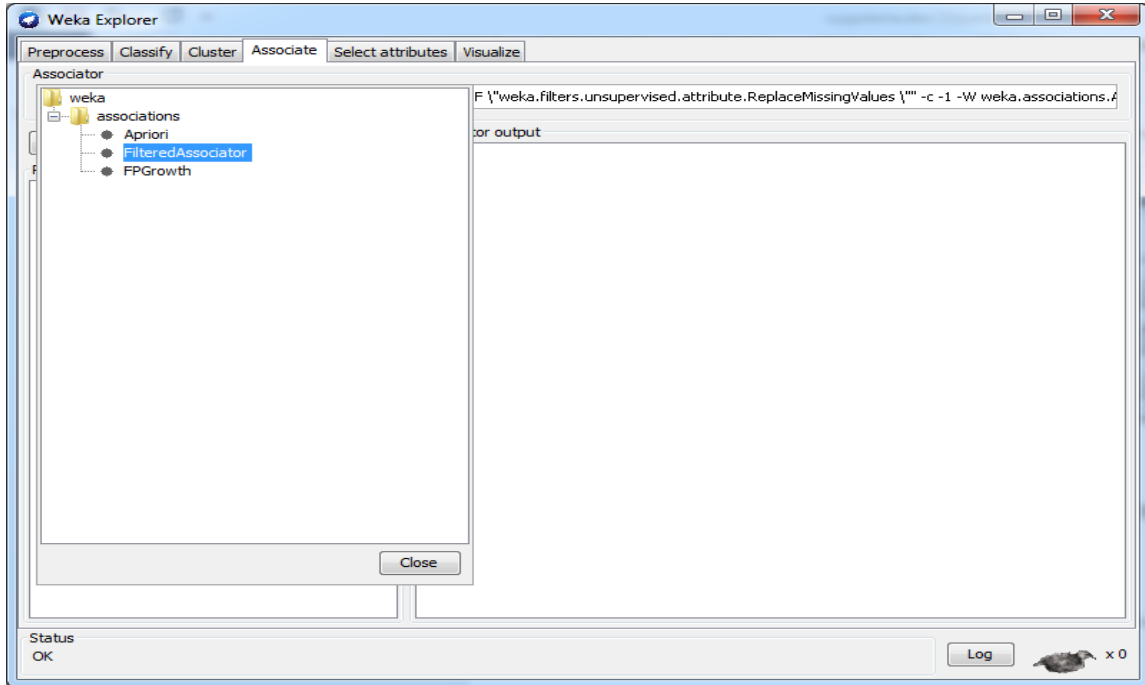
Şekil 4.2’de görüldüğü üzere Apriori Algoritması kullanılarak birliktelik analizi gerçekleştirilmiştir. Sonuçları yorumlayabilmemiz için çıktıda karşımıza gelen conf, lift, lev, conv değerlerinin ne ifade ettiğini bilmemiz gerekir. Conf.(confidence) değeri bulmuş olduğumuz kuralın olasılığının sistemdeki en büyük olasılığa bölünmesiyle elde edilir. Con.(conviction) değeri hesaplanırken x ürünlerinin, y ürünü olmaksızın görülme olasılıkları bulunur. Lift.(interest) matematiksel hesaplama bazında x ve y’nin birbirinden ayrı olması durumunda hangi süreyle birarada işlem göreceğini saptamaktadır buna istinaden bir sonuca varmaktadır. Lev.(leverage) ise x ve y’nin birarada olmaları durumunda oluşan farklılığı saptamaktadır. Ayrıca istatistiki olarak x ve y değerlerinin birbirlerine ilintili olma durumundaki nihai sonucunu bulmaktadır.

Yukarıda belirtilen pay ölçümleri ne kadar büyük olursa elde edilen kurallar arasındaki ilişkide o kadar yüksek olacaktır. Weka en güçlü 10 tane ilişkiyi çıktı olarak vermiştir. Burada önceliği belirlerken ilk olarak confident değerinin büyüklüğü baz alınmaktadır.

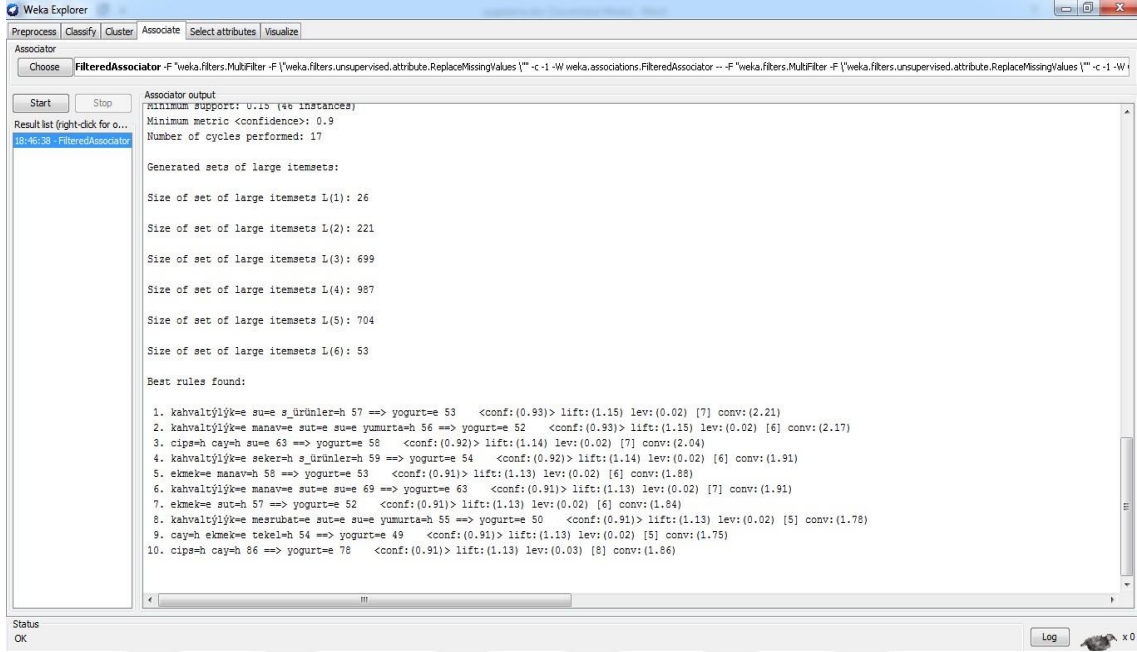
Sonuçlardan bir kaçına baktığımızda kahvaltılık malzeme ve su alan 57 tane müşteriden 53 tanesi aynı zamanda yoğurttan almıştır. 2. en güçlü ilişki ise kahvaltılık malzeme, manav alışverişi, süt ve su alan 56 müşteriden 52 tanesi aynı zamanda yoğurttan almıştır. Sonuçlara bakarak benzer çıkarımlarda bulunabiliriz.

Yani sonuç olarak ele alınan market bir kampanya düzenlemek isterse ilk olarak kahvaltılık malzemeler, su ve yoğurttan oluşan üçlüyü seçmelidir. Bu şekilde satışlarını arttırmayı başarabilir.

FiteredAssociator algoritmasının kullanılmasıyla (Şekil 4.3) elde edilen sonuçlar Şekil 4.4'te gösterilmiştir.



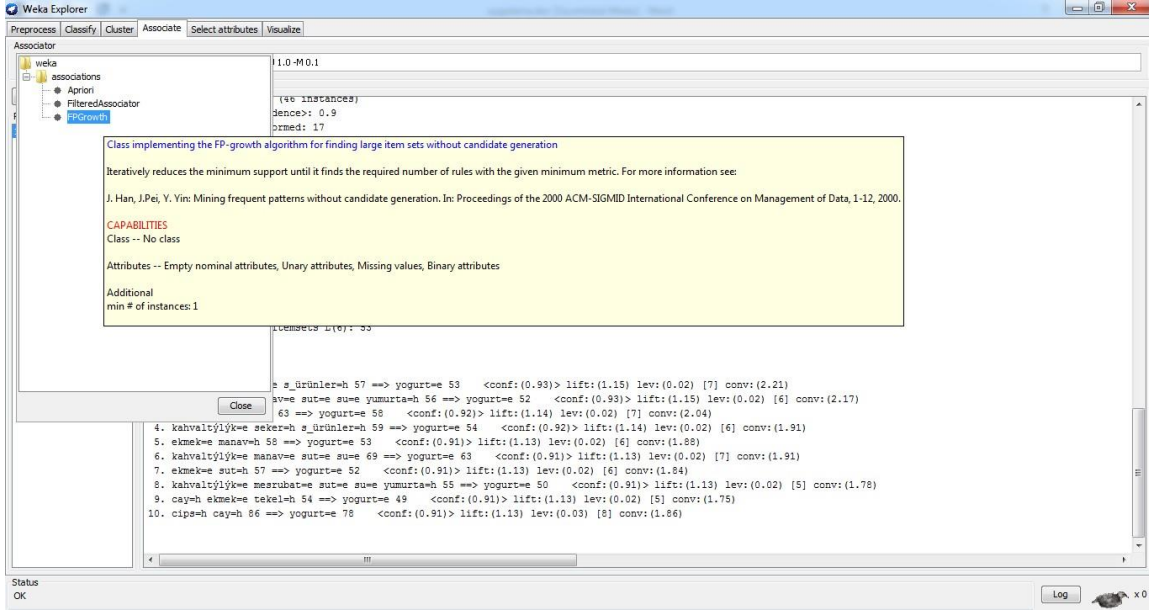
Şekil 4.3: Filtered Associator Algoritması



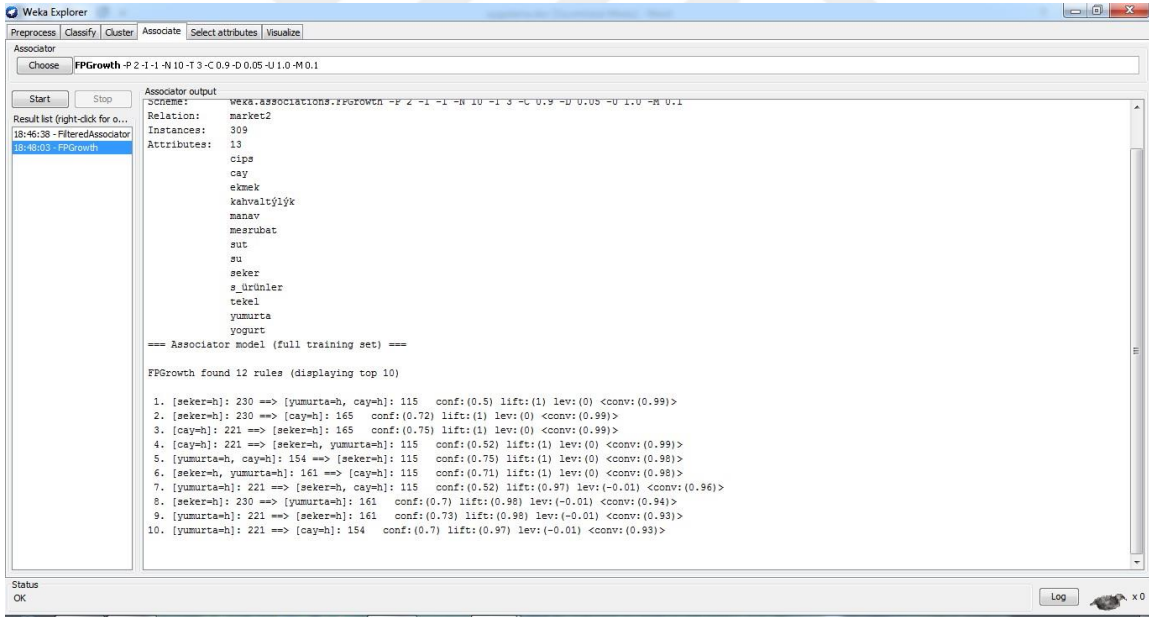
Şekil 4.4: Filtered Associator Algoritması Uygulaması Sonucu

FilteredAssociator Algoritmasında Apriori ile aynı sonuçları vermiştir. Sonuçlardan bir kaçına baktığımızda ekmek alan 58 tane müşteriden 53 tanesi de aynı zamanda yoğurttan almıştır. Burda ise şöyle bir yorum yapabiliriz; ekmek ve yoğurt market içerisinde birbirlerine yakın konumlandırılmalıdır. Çünkü müşteriler bu 2 ürünü genellikle birlikte aldıklarından yakın konumlandırılmaları market satışlarını arttıracaktır. Yine kurallara bakarak benzer çıkarımlarda bulunabiliriz.

FPGrowth algoritmasının kullanılmasıyla (Şekil 4.5) elde edilen sonuçlar Şekil 4.6’da gösterilmiştir.



Şekil 4.5: FPGrowth Algoritması Seçimi



Şekil 4.6: FPGrowth Algoritması Uygulaması Sonucu

Bu algoritmanın çıktılarında ise birlikte alınmama ilişkisinin yüksek olduğu kurallar karşımıza çıkmıştır. Örneğin; şeker almayan 230 tane müşteriden 115 tanesi aynı zamanda yumurta ve çay da almamıştır. Diğer bir kural ise şu şekildedir; çay almayan 221 müşteriden

aynı zamanda 165 tanesi de şeker almamıştır. Yani şeker ve çaydan herhangi birini almayan müşteriler diğer üründe genellikle almamıştır. Bu yüzden bu iki ürün arasında bir kampanya düzenlenmesi satışlarımızı arttıracaktır.



5. TARTIřMA VE SONUÇ

Yapılan alıřmalar ve uygulamalar sonucunda analiz edilen dataların nihai sonucuna gre veri madencilięi aralarının, satın alım iřlemi yapan kiřiler tarafından, rnlerin alınma ařamasında satın alma eęilimleriyle iliřkili yapıları idealize etmek iin gl bir řekilde kullanılabileceęi tespit edilmiřtir.

Veri madencilięindeki baęıntı kuramı sayesinde daha ok kullanılan ve tercih edilen apriori ve dięer birliktelik kuralları uygulamasını hayata geiren algoritmalar kullanılmıřtır. Bu algoritmalar sayesinde satın alımı yapan kiřilerin yksek frekans yoęunluęu ile yapılmıř olan ve onlarla ilintili satın alma davranıřları zelinde gven-destek analizi uygulanarak bazı tespitlerde bulunulmuřtur, bulunan bu tespitlere alıřma iinde yer verilmiřtir. Ayrıca mřterilerin satın alma davranıřları zelindeki ayak izleri takip edildięinde almıř oldukları malzeme veya rnler arasında gerek pozitif gerekse negatif ynde bir baęıntı saptanmıřtır.

Nihayetinde analiz ve alıřmalar sonucunda elde edilen verilere gre rn-raf entegrasyonu esnasında bu anlamlı iliřkiler gz nnde bulundurulularak optimum yer seimi yapılabilir. Bu metot satıř yapan kiřinin standart satıř iřlemlerine nazaran bilimsel verilere dayanarak iřlem grmř ve hayata geirilmė metodoloji ile daha yksek kar marjına ulařabileceęine iřaret etmiřtir.

REFERANSLAR

- [1] Giudici Paulo. Applied Data mining :Statistical Methods for business and industry, -ISBN 9812-53- 178-5 (2003)
- [2] Peter P. Wakabi-Waiswa Venansius Baryamureeba, Extraction Of Interesting Association Rules Using Genetic Algorithms *International Journal of Computing and ICT Research*, 2(1), (2008)
- [3] Nan-chan Hsich, Kuo-Chang cha Enhancing consumer behavior analysis by data mining techniques (2009)
- [4] Mike Chapple, Data mining an introduction Classification (2001)
- [5] Junzo Watada and Kozo Yamashiro, A Data Mining Approach to consumer behavior- *Procedings of the first International Conference on Innovative computing Information* (2006)
- [6] Jiauei Han and Michele Kamber, Data mining Concepts and Techniques, Simon Fraser University, ISBN 1-55860-489-8-(2001)
- [7] Yıldız B. and Ergenç B., (Turkey) in Comparison of Two Association Rule Mining Algorithms without Candidate Generation, *International Journal of Computing and ICT Research*, 674(131), 450-457 (2010)
- [8] Pillai Jyothi, User centric approach to itemset utility mining in Market Basket Analysis, *International Journal on Computer Science and Engineering (IJCSE)*, 3, (2011)
- [9] Parvinder S. Sandhu Dalvinder, Dhaliwal S. and Panda S.N., Mining utility-oriented association rules: An efficient approach based on profit and quantity, *International Journal of the Physical Sciences*, 6(2), 301-307 (2011)
- [10] Agrawal R. and Srikat, Fast Algorithms for Mining Association Rules- 1. Sept (1994)
- [11] Agrawal R., Imilienski T. and Swami A., Mining Associations Rules between Sets of Items in large databases. Proc. of the ALM SIGNOD. *Int'l conf. on management of Data*, 207-216 (1993)
- [12] M. S. Chen, J. Han, and P. S. Yu, "Data Mining: An Overview from a Database Perspective," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6,

- 1996, pp. 866-883.
- [13] B. Goethals and M. J. Zaki, "Advances in Frequent Itemset Mining Implementations: Introduction to FIMI03," In Goethals, B. and Zaki, M. J., editors, *Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, Vol. 90 of CEUR Workshop Proceedings, 2003.
 - [14] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th International Conference on Very Large Data Bases*, 1994, pp. 487-499.
 - [15] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, 2000, pp. 1-12.
 - [16] F. Bodon, "A Fast Apriori Implementation," In B. Goethals and M. J. Zaki, editors, *Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, Vol. 90 of CEUR Workshop Proceedings, 2003.
 - [17] F. Bodon, *A Survey on Frequent Itemset Mining*, Technical Report, Budapest University of Technology and Economic, 2006.
 - [18] S. Thomas, S. Bodagala, K. Alsabti, and S. Ranka, "An efficient Algorithm for the Incremental Updation of Association Rules in Large Databases," *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, 1997, pp. 263-266.
 - [19] GUHA, S., RASTOGI, R., SHIM, K., "CURE: An Efficient Clustering Algorithm for Large Databases", in: *Proceedings of the 1998 ACM SIGMOD International Conference on Management Data*, vol. 27(2), 73-84, 1998.
 - [20] QUINLAN, J.R., "Induction of Decision Trees ", *Journal of Machine Learning*, vol. 1, 81-106, 1986.
 - [21] ZHANG, T., RAMAKRISHNAN, R., LIVNY, M., "BIRCH: An Efficient Data Clustering Method for Very Large Databases", in: *Proceedings of the 1996 ACM SIGMOD International Conference on Management Data*, vol. 25(2), 1996.
 - [22] <http://bilgisayarkavramlari.sadievrenseker.com/2011/09/07/apriori-algoritmasi/>