



**SARF MALZEME KULLANIMINDAN VERİ MADENCİLİĞİ
BİRLİKTELİK KURALLARININ ELDE EDİLMESİ, KURALLARIN
ANALİZİ VE SINIFLANDIRILMASI**

Hasan TONTUŞ

**YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

OCAK 2020

Hasan TONTUŞ tarafından hazırlanan “SARF MALZEME KULLANIMINDAN VERİ MADENCİLİĞİ BİRLİKTELİK KURALLARININ ELDE EDİLMESİ, KURALLARIN ANALİZİ VE SINIFLANDIRILMASI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Endüstri Mühendisliği Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Ertan GÜNER

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Başkan: Doç. Dr. Ümit Sami SAKALLI

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Üye: Dr. Öğr. Üyesi Tahsin ÇETİNYOKUŞ

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 30/01/2020

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Hasan TONTUŞ

30/01/2020

SARF MALZEME KULLANIMINDAN VERİ MADENCİLİĞİ BİRLİKTELİK
KURALLARININ ELDE EDİLMESİ, KURALLARIN ANALİZİ VE
SINIFLANDIRILMASI
(Yüksek Lisans Tezi)

Hasan TONTUŞ

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ocak 2020

ÖZET

Her geçen gün kabiliyetleri artan bilgisayar sistemleri sayesinde büyük miktardaki verilerin saklanıp işlenebilmesine rağmen, bunlardan elde edilen bilgi miktarının yeterli düzeyde olmadığı veya amaca yönelik kullanışlı bilgiye odaklanılamadığı bilinmektedir. Kurumlar veri toplama, saklama ve analiz işlemleri için yatırımlarına devam ederlerken, verilerin hacimlerinin çok büyük olması ve yapılarının da etkin bir veri analizi yapılmasına uygun olmaması nedeniyle, uygulamalarda bu verilerin bir kısmı kullanılabilmektedir. Sanayi üretimi sektöründe de üretimin bütün detaylarının saklanabildiği görülmektedir. Bu detaylardan bir tanesi de, iş emirlerinde kullanılan sarf malzemelerdir. Sarf malzeme kullanımı neredeyse tüm üretim çeşitleri için ortak payda olduğundan, bu alana yönelik bir veri analizine ihtiyaç duyulması kaçınılmazdır. İşlenmemiş verilerden işe yarayacak bilgi ya da örüntü elde etme sürecine veri madenciliği denir. Veri madenciliği çeşidi olan ve veriler içinde birlikte hareket eden öğelerin keşfedilmesini sağlayan birliktelik kurallarının tespiti için geliştirilen algoritmalar arasında en sık kullanılan model Apriori algoritmasıdır. Bu çalışmada Apriori algoritması seçilerek, bir kamu kurumunun yürüttüğü bakım-onarım faaliyetlerinde ihtiyaç duyulan sarf malzemelere yönelik bir uygulama gerçekleştirilmiştir. Son beş yıl içinde meydana gelen arızalar için kullanılan malzemelerin verileri incelenerek yapılan çalışma sonucunda elde edilen bilgilerin kullanılmasıyla, etkin ve hızlı bir planlama süreci, uygun depo düzeni, malzeme tedarik maliyetinde ve standart birim adam saatte azalma, iş yükü ve iş gücü konularında iyileştirmeler sağlanacağı değerlendirilmiştir. Ayrıca, veri setinin birbiriyle ilişkili gruplar halinde ele alınmasının önemi anlatılmıştır. Bu bağlamda bütün veri setinden elde edilen sonuçlarla, alt grup veri setlerinden elde edilen sonuçlar birbiriyle kıyaslanmıştır. Çalışmada büyük boyutlu veri seti kullanıldığından, elde edilen çok sayıdaki birliktelik kurallarına yönelik ikinci bir analiz yapılarak, kuralların amaca yönelik sınıflandırılması yapılmıştır. Çalışmada, genellikle analizcinin sezgisel olarak seçtiği minimum destek değerinin tespiti için bir formül geliştirilerek algoritmada bu formül aracılığıyla elde edilen değerler kullanılmıştır. Ayrıca, veri setinin sunduğu verilerle, ortaya çıkan kural sayısı arasındaki ilişki araştırılmıştır.

Bilim Kodu : 90607

Anahtar Kelimeler : Veri Madenciliği, Birliktelik Kuralları, Apriori Algoritması, Malzeme Kullanımı, Birliktelik Kuralları Sınıflandırılması

Sayfa Adedi : 111

Danışman : Prof. Dr. Ertan GÜNER

OBTAINING DATA MINING ASSOCIATION RULES FROM USE OF
CONSUMABLES, ANALYSIS AND CLASSIFICATION OF THESE RULES

(M. Sc. Thesis)

Hasan TONTUŞ

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

January 2020

ABSTRACT

It is known that even though large amounts of data can be stored and processed by computer systems with increasing capabilities every day, the amount of information obtained from them is not sufficient or cannot be focused on useful information. While organizations continue to invest in data collection, storage and analysis operations, only some of these data can be used in applications because the data volumes are very large and their structures are not suitable for efficient data analysis. It is seen that all the details of production can be kept in the industrial production sector as well. One of these details is the consumables used in work orders. Since the use of consumables is the common denominator for almost all types of production, it is inevitable that a data analysis is needed for this area. The process of obtaining information or patterns that will work from raw data is called data mining. Apriori algorithm was the most commonly used model for the identification of association rules, which is a type of data mining and which enables the discovery of elements that move together within the data. In this study, an application was made for the consumables that a public institution would need for maintenance-repair activities. As a result of the study, the data obtained during the last five years were evaluated by using the information obtained, effective and fast planning process, appropriate warehouse layout, material procurement cost and unit man hours reduction, workload and labor force improvements were evaluated. The importance of examining the data set in related groups is explained. Since the results obtained from the analysis without sub-group were assumed to be insufficient at the point of service, the results obtained from the whole data set and the results obtained from the sub-group data sets were compared. Since a large-scale data set was used in this study, a second analysis was made for the large number of association rules and the classification of the rules was made. In this study, a formula has been developed to determine the minimum support value that the analyst has intuitively selected and the values obtained by this formula are used in the algorithm. In addition, the relationship between the data set and the number of rules was investigated.

Science Code : 90607

Key Words : Data Mining, Association Rules, Apriori Algorithm, Material Use, Classification of Association Rules

Page Number : 111

Supervisor : Prof. Dr. Ertan GÜNER

TEŞEKKÜR

Beni kendisine yüksek lisans öğrencisi olarak kabul eden ve tez çalışmam sırasında hiçbir önerimi reddetmeyerek kıymetli bilgi, birikim ve tecrübeleri ile bana yol gösteren, yapmış olduğum hataları anlayışla karşılayarak desteğini sürdüren aynı zamanda manevi desteğini de hiçbir zaman esirgemeyen değerli danışman hocam sayın Prof. Dr. Ertan GÜNER’e sonsuz teşekkür ve saygılarımı sunarım.

Başarılarımı abartan, başarısızlıklarımı başarı gören, beni benden daha iyi savunan, üzüntülerimle üzülen, tüm eğitim hayatım süresince bana hep destekçi olan ve çalışmalarım boyunca manevi desteğiyle beni hiçbir zaman yalnız bırakmayan anneme sonsuz sevgi, saygı ve teşekkürlerimi bildirmek isterim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xii
RESİMLERİN LİSTESİ	xiii
SİMGELER VE KISALTMALAR.....	xiv
1. GİRİŞ.....	1
2. LİTERATÜR ARAŞTIRMASI.....	5
3. VERİ MADENCİLİĞİ.....	11
3.1. Veri, Bilgi ve Metaveri	12
3.2. Veritabanı Sistemleri.....	13
3.3. Veri Ambarları	13
3.3.1. Datamart ve veri küpleri	14
3.3.2. Veri ambarının oluşturulması	14
3.4. Veri Madenciliği Metodolojisi.....	15
3.5. Veritabanında Bilgi Keşfi Süreci	15
3.6. Veri Madenciliği Modelleri ve Kullanılan Algoritmalar	17
3.6.1. Sınıflama ve regresyon	18
3.6.2. Kümeleme.....	23
3.6.3. Birliktelik kuralları ve ardışık zamanlı örüntüler	24

Sayfa

4. BİRLİKTELİK KURALI	25
4.1. Birliktelik Kuralının Matematiksel Gösterimi	26
4.1.1. Destek ve güven kavramları.....	28
4.2. Birliktelik Kurallarında Kullanılan Algoritmalar.....	32
4.2.1. AIS (Agrawal, Imielinski, Swami) algoritması	32
4.2.2. SETM algoritması.....	33
4.2.3. Apriori algoritması	33
4.2.4. AprioriTid algoritması.....	39
4.2.5. FP-Growth algoritması.....	39
4.2.6. Apriori-Hybrid algoritması	40
4.2.7. OCD algoritması (Off-line candidate determination / Off-line ürün kümesi tespiti).....	40
4.2.8. Partition algoritması (Bölümleme algortması).....	40
4.2.9. DIC algoritması (Dynamic itemset counting/Dinamik ürün kümesi sayımı)...	41
4.2.10. CARMA (Continuous association rule mining/sürekli birliktelik kuralı madenciliği algoritması)	41
4.2.11. Diğer algoritmalar ve algoritmaların sınıflandırılması	41
4.3. Birliktelik Kuralları Algoritmalarının Değerlendirilmesi	42
5. BİR FABRİKADA KULLANILAN SARF MALZEMELERE İLİŞKİN BİRLİKTELİK DAVRANIŞLARININ TESPİT EDİLMESİ.....	47
5.1. Genel Bilgiler	47
5.2. Verilerin Tanımı.....	49
5.2.1. Her faaliyet grubu için ayrı ayrı normalite testleri	54
5.3. Veri Setinin Alt Gruplara Ayrılmasındaki Amaç.....	58
5.3.1. Faaliyet grupları ile kullanılan sarf malzeme arasındaki ilişkinin tespiti.	59

	Sayfa
5.4. Veriler İçin Kullanılan Yöntem ve Algoritma	62
5.4.1. Minimum destek ve minimum güven değerinin belirlenmesi	63
5.5. SPSS Modeler Programın Kullanımı ve Veri Seti Üzerinde Çalıştırılması	64
5.6. Bulgular	65
5.6.1. Bütün veri setinden elde edilen bulgular	65
5.6.2. Faaliyet gruplarının veri setlerinden elde edilen bulgular	74
5.6.3. Bütün veri setinden elde edilen bulgular ile alt gruplardan elde edilen bulguların karşılaştırılması	79
6. SONUÇLAR VE ÖNERİLER	87
KAYNAKLAR	91
EKLER	97
EK-1. Uygulamda Kullanılan Veri Seti Şablonu ve Malzeme Listesi	98
EK-2. SPSS Modeler Programın Kullanımı ve Veri Seti Üzerinde Çalıştırılması	101
EK-3. Sarf Malzeme Kullanımına Ait Pareto Analizinin Tüm Listesi	107
ÖZGEÇMİŞ	111

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 4.1. Olası birliktelik kuralları	28
Çizelge 4.2. Ürün satış bilgileri	29
Çizelge 4.3. Müşteriler ve aldıkları ürünler	35
Çizelge 4.4. Algoritmaların kıyaslanması	45
Çizelge 5.1. Veri setinin tanımlayıcı istatistikleri.....	51
Çizelge 5.2. Malzeme çeşidi sıklıkları	52
Çizelge 5.3. Faaliyet grubu bazında malzeme çeşidi sıklıkları.....	53
Çizelge 5.4. Faaliyet gruplarının tanımlayıcı istatistikleri.....	55
Çizelge 5.5. Normalite test sonuçları	56
Çizelge 5.6. Logaritmik dönüşüm sonrası faaliyet gruplarının tanımlayıcı istatistikleri	57
Çizelge 5.7. Logaritmik dönüşüm sonrası normalite test sonuçları.....	58
Çizelge 5.8. Tanımlayıcı bilgiler	60
Çizelge 5.9. Kruskal wallis test sonuçları	60
Çizelge 5.10. Mann-Whitney Test sonuçları	61
Çizelge 5.11. Mann-Whitney test sonuç özetleri	62
Çizelge 5.12. Bütün veri setinden elde edilen kurallar	67
Çizelge 5.13. Öncül, ardıl ilişkisi.....	74
Çizelge 5.14. Yükseltilmiş minimum destek oranları.....	75
Çizelge 5.15. Elektrik grubundan elde edilen güçlü ve kesin kurallar	76
Çizelge 5.16. Elektroniğin elektriği grubundan elde edilen güçlü ve kesin kurallar	78
Çizelge 5.17. İmalat grubundan elde edilen güçlü ve kesin kuralların bir kısmı.....	79
Çizelge 5.18. Veri setlerine ait karşılaştırmalı özet bilgiler.....	80
Çizelge 5.19. A Sınıfı'na giren sarf malzemeler.....	84

Çizelge	Sayfa
Çizelge 5.20. Alternatif yöntemlerden elde edilen kural sayıları	85

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 3.1. Veri madenciliğinin benzer disiplinlerle ilişkisi.....	12
Şekil 3.2. Veri ambarı yapısı.....	13
Şekil 3.3. Veri madenciliği metodolojisi	15
Şekil 3.4. Veritabanında bilgi keşfi süreci	16
Şekil 3.5. Veritabanından bilgi keşfi aşamaları	16
Şekil 3.6. Veri madenciliği modelleri	18
Şekil 3.7. Örnek karar ağacı.....	20
Şekil 4.1. Örnek Apriori algoritmasının aşamaları	37
Şekil 4.2. Apriori algoritmasının matematiksel modeli	38
Şekil 4.3. Birliktelik kuralı algoritmaları	42
Şekil 5.1. Kuralların ardılları grafiği.....	73
Şekil 5.2. Kuralların öncülleri grafiği	73

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 5.1. Veri setinin excel programındaki görüntüsü.....	50
Resim 5.2. Modeller boş ara yüzü	63
Resim 5.3. Destek sıralı birliktelik kuralları	66
Resim 5.4. Güven sıralı birliktelik kuralları	66

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

C_k	Her bir aday nesne kümeleri
d	Nesne sayısı
$G/\mathcal{C}$	Veritabanı taramalarının azami sayısı
I	Nesnelerin kümesi
i	Nesne
L_k	k-nesnekümenin sık geçen kümeleri
m	Tarama sayısı
R	Ortaya çıkması muhtemel birliktelik kuralı sayısı
X_j, Y_j	Yapılan iş veya nesnelerdir

Kısaltmalar

Açıklamalar

Elk.	Elektrik
Elkt.	Elektronik
Malz.	Malzeme
min_des	Minimum destek
min_güv	Minimum güven
Ort.	Ortalama
Parç.	Parçalar
VTBK	Veri tabanında bilgi keşfi
YMDO	Yükseltilmiş minimum destek oranı

1. GİRİŞ

Günümüzde işletmelerin, yoğun teknoloji ve bilgisayar kullanımlarının artmasıyla birlikte veriler, elektronik ortamda tutulmaya ve işlenmeye başlanmış, böylece elektronik veri saklama ve analiz araçları gelişmiştir. Büyük miktarlardaki veriyi işleme yeteneğine sahip teknolojiye duyulan gereksinim, işlemci hızları ve disk kapasiteleri artan bilgisayar sistemleri sayesinde kolaylıkla karşılanabilir hale gelmiştir. Buna rağmen bu devasa veri yığınlarından elde edilen bilgi miktarının yeterli düzeyde anlaşılır olduğu söylenemez. Toplanan verilerin hacimlerinin çok büyük olması ve yapılarının da etkin bir veri analizi yapılmasına uygun olmaması nedeniyle uygulamalarda bu verilerin küçük bir kısmı kullanılabilir.

Günümüz şartları gereği daha dinamik ve esnek yapıda olmaları gereken iş sektörleri, sadece temel bazdaki bilgiye değil, belli bir takım işlemlerden geçerek bir çaba karşılığı elde edilen bilgilere de ihtiyaç duymaktadırlar. Örneğin eskiden bilgisayarlar sadece üretim miktarı veya süresi gibi basit raporlar hazırlamak için kullanılırken günümüzde üretimin bütün detaylarının bilinmesi bir gereksinim haline almıştır. İşletmeler geleceğe yönelik yaptıkları her türlü planın gerçeğe daha çok yaklaşabilmesi için daha önce yapmış oldukları işlere ait bilgileri elektronik ortamlarda kayıt altına almakta ve bu verilerden anlamlı bilgilere ulaşmayı hedeflemektedirler. Saklanan bu binlerce ürün hareket bilgileri sayesinde herhangi bir zamanda üretim verilerine ulaşmak ve analiz etmek mümkün olabilmektedir.

Veri madenciliği, veri setlerinden ilk bakışta anlaşılamayan gizli bilginin ortaya çıkarılmasıdır. Veri madenciliği teknikleri, iş dünyasında ve akademik alanda sıklıkla kullanılmaktadır. Dünyada bu ihtiyaca olan talep hızla artmaktadır. Veri bilimini kullanmadan firmaların başarılı olmaları çok zorlaşmıştır. İstatistik gibi geleneksel yöntemlerin yeterli olmadığı durumlarda, kaynaklarının etkin kullanılabilmesi için veri madenciliğinden faydanılması gerekmektedir. Ticaret, biyoloji, telekomünikasyon, eğitim, üretim, tıp, bilgisayar ve sosyal bilimler gibi birçok alanda veri madenciliği teknikleri etkin çözümler sunmaktadır. Günümüzde artık veri madenciliğinin önemi bilindiğinden bu alana yapılan yatırımlar gün geçtikçe artmaktadır.

Veri madenciliği, bilgisayarlar yardımıyla büyük veri yığınlarından anlamlı bilgiler elde etmek için yapılan bilgi çözümleme işlemidir. Birliktelik kuralları ise veritabanındaki fark edilmeyen bilgileri ortaya çıkarmaya yarayan bir veri madenciliği çeşididir. İşe yarar, tutarlı bilgiler elde etmeyi amaçlayan birliktelik kuralları, veriler içinde birlikte hareket eden öğelerin keşfedilmesini, keşfedilen bu bağıntılar ile geleceğe yönelik tahminlerin yapılmasını sağlar.

Apriori algoritması, verilerde sık geçen öğelerin keşfedilmesinde en fazla kullanılan bir birliktelik kuralı algoritmasıdır. Sık geçen öğeleri bulmak için birçok kez veritabanını taramak gerekir, bu taramalar aşamasında Apriori algoritmasının birleştirme, budama işlemleri ve minimum destek ölçütü yardımı ile birliktelik ilişkisi olan öğeler bulunur.

Bu çalışmada, bir kamu kurumunun yürüttüğü bakım ve onarım faaliyetlerinde ihtiyaç duyulan sarf malzemelerin önceden tahmin edilmesine yönelik bir çalışma gerçekleştirilmiştir. Son beş yıl içinde meydana gelen arızaların ve bu arızalar için kullanılan malzemelerin verileri incelenerek yapılan çalışma sonucunda tedarik sürecinin kısaltılması ve üretim planlarında malzeme kaynaklı gecikmelerin önlenmesi amaçlanmıştır. Bu amaç doğrultusunda yapılan uygulamada, veri madenciliğinin birliktelik kuralları yöntemi seçilmiş ve Apriori algoritması ile işe yarar ve tutarlı kuralların elde edilmesi için gerçek verilerle çalışılmıştır.

Birçok alanda birliktelik kuralları kullanılarak yapılmış çalışmalar literatürde mevcuttur. Bu alanlara örnek olarak; mühendislik, tıp, eğitim, bankacılık, finans, telekomünikasyon, pazarlama, e-ticaret, sigortacılık gösterilebilir. Literatürde sadece fen bilimlerinde değil, sosyal bilimler alanında da birçok örneğine rastladığımız konu üzerinde, daha önce malzeme planlamasına yönelik olarak bir çalışma bildiğimiz kadarıyla gerçekleştirilmemiştir. Bu bakımdan, malzeme planlama ile birliktelik kuralları ilk kez bu çalışmada ele alınacaktır.

Çalışma sonucunda elde edilmesi amaçlanan kurallar doğrultusunda, arıza ve bakım durumlarına yönelik hangi sarf malzemelerin birlikte kullanıldıklarının tespitinin yapılarak bu malzemelerin temininin birlikte yapılması suretiyle temin masraflarında sağlanacak azalmanın yanısıra uygun depolara ya da bir depoda uygun yerlere yerleştirilerek onarım faaliyeti sırasında iş gücü ve zaman tasarrufu sağlanacaktır. Ayrıca kullanılan verilerin

nicelikleri ile elde edilen kural sayıları arasındaki ilişki sorgulanacaktır. Fazla sayıda birliktelik kuralı elde edilmesi durumunda, bu kurallara yönelik ikinci bir analizin yapılması ve elde edilecek bu kuralların amaca yönelik sınıflandırılması da sağlanacaktır.

Çalışmanın ikinci bölümünde literatür taraması yapılacaktır. Bu bölümde daha çok Apriori algoritmasının kullanıldığı günlük yaşama ilişkin çeşitli uygulamalardan bahsedilecektir. Üçüncü bölüm veri madenciliğine genel bir bakışın ve veri madenciliği yöntemlerinin sunulduğu bir bölüm olacaktır. Dördüncü bölümde ise veri madenciliği yöntemlerinden birisi olan birliktelik kuralları ve bu kuralların çıkarılması için geliştirilen algoritmalarından en çok tercih edilenleri, özellikle Apriori algoritmasının çalışma mantığı aktarılacaktır. Çalışmanın beşinci bölümünde bir işletmede kullanılan sarf malzemelere ilişkin birliktelik kuralları tespit edilmeye çalışılacaktır. Ayrıca, bu kuralların çeşitli yöntem ve sınırlandırmalar ile analizi ve tasnifi yapılacaktır. Son bölümde ise elde edilen bulgular değerlendirilerek çalışmanın literatüre kazandıracığı farklı yanları anlatılacak ve bu noktadan sonra konunun nasıl geliştirilebileceğinden bahsedilerek çalışma sonuçlandırılacaktır.

2. LİTERATÜR ARAŞTIRMASI

Birliktelik kurallarının literatürde çok farklı amaçlar için kullanıldığı belirlenmiştir. Temel veri madenciliği görevlerinden biri olan birliktelik kurallarının elde edilmesi en çok apriori algoritması aracılığıyla gerçekleştirilmiştir. Birliktelik kuralları ve apriori algoritması konularını ele alan çalışmaları yayınlandıkları yılları göz önünde bulundurmak suretiyle şu şekilde özetlemek mümkündür.

Karabatak ve İnce (2004) öğrenci başarılarının analizi için apriori algoritmasından yararlanarak, bazı birliktelik kuralları oluşturmuşlardır. Çalışma sonucunda, öğrencilerin üniversiteye giriş sınavında aldıkları puanlar ile ilk dönem başarı durumları ve sonraki dönem başarı durumları arasında dikkate değer ilişkiler keşfedilmiştir [1]. Flank (2004) birliktelik kurallarını, birbirleriyle ilişkili çoklu tablolar üzerinde çalıştırdığı algoritmayla ortaya çıkarmıştır [2].

Witten ve Frank (2005), birliktelik kuralları modelleriyle sınıflandırma modellerini kıyaslamış, sınıflandırma modelinin sadece bir sınıfı baz alarak kuralları ortaya çıkarırken, birliktelik kurallarının veritabanındaki tüm değişkenler arasındaki ilişkileri incelediğini savunmuştur [3].

Cavique (2007), büyük veri setlerinde birliktelik kuralı algoritmalarından bazılarının çok uzun zamanda sonuçlandığını vurgulamakta olup, çalışmasında büyük veri setlerinde hızlı bir şekilde birliktelik kurallarını bulacak bir algoritma tanıtmaktadır. Önerilen yöntemin başarılı olduğu vurgulanmaktadır [4]. Özçakır ve Çamurcu (2007) yaptıkları çalışmada apriori algoritmasını kullanan bir yazılım tasarımı gerçekleştirmişlerdir. Bir firmanın pastane satış verileri üzerinde geliştirdikleri yazılımı çalıştırmaları neticesinde, aynı nitelikteki ürünlerin genellikle birlikte satıldığı sonucuna varmışlardır [5].

Chen ve diğerleri (2008), birliktelik kuralları uygulamalarıyla trafik kazasına sebep olan faktörler arasındaki ilişkileri incelemişlerdir [6]. Ay ve Çil (2008), bir perakende zincirinde birliktelik kuralları oluşturmaya yönelik bir çalışma yapmışlardır. Elde edilen birliktelik kurallarından yararlanılarak aynı zamanda farklı bir yerleşke düzeni de önerilmiştir [7]. Erdem ve Özdağoğlu (2008) yaptıkları çalışmada acil hasta verilerine apriori algoritmasını uygulamaları neticesinde hastalara ilişkin anlamlı kurallar ortaya koymuşlardır [8].

Oktay (2010), çok ziyaret edilen ve güncel bir portal sitesinden alınan kullanıcı log kayıtları üzerinde web kullanım madenciliği yöntemini uygulayarak, saklı bilgileri ortaya çıkarmak istemiştir. Bunu gerçekleştirmek için, veri madenciliği yöntemlerinden olan birliktelik analizinin en sık kullanılan algoritması Apriori algoritması ile bazı bilim insanları tarafından ortaya atılan TFPR algoritması seçilmiştir. Apriori algoritması ile sonuçlara ve kurallara ulaşılmış, TFPR algoritması ile de en sık ziyaret edilen ve edilme olasılığı en yüksek olan sayfalar bulunmuştur [9].

Wang ve Liu (2011), Apriori algoritmasının zayıf yönlerine dikkatleri çekmiş ve gelişmiş bir birliktelik kuralı algoritması sunmuştur. Önerilen algoritma, aday kümelerini tarama zamanlarını azaltmaya ve aday öğelerini saklamak için hash ağacı kullanmaya dayanıyor. Algoritmanın sonucunda, madenciliğin işlem süresi azalıyor ve algoritmanın etkinliği artıyor [10]. Ekim (2011) tez çalışmasında, Selçuk Üniversitesinde kullanılan öğrenci işleri otomasyonundan sağlanan veriler üzerinden, öğrenciler hakkında geleceğe yönelik tahmin yapılabilmesi için gerekli birliktelik kurallarına ulaşmıştır. Bu tezde apriori algoritması ve karar ağacı algoritması kullanılarak elde edilen kurallar sayesinde, Selçuk Üniversitesine henüz yerleşen bir öğrencinin, üniversite başarısına etki eden faktörler araştırılmıştır. Araştırma sonucunda, öğrencinin ailesinin eğitim seviyesinin ve gelir düzeyinin öğrencinin başarısında en etkili faktörler olduğu belirlenmiştir [11].

Alaeddinoğlu (2012), Van Gölü'ne ilişkin göl seviyesi, göle düşen yağış, göldeki buharlaşma vb. gibi geçmiş yıllarda kayıt altına alınmış verileri Apriori algoritması ile inceleyerek, Mekânsal-Zamansal bir veri madenciliği uygulaması gerçekleştirmiştir. Bu uygulama ile Van Gölü'nün kıyı kesimlerinde meydana gelen taşmalar ve oluşan yeni kıyı arazilerinin tespit edilmesi amaçlanmıştır. Böylece Van Gölü'nün kıyı kesimlerinde alınması gereken tedbirlerin ve yapılması öngörülen yatırımların değerlendirilmesi sağlanmıştır [12]. Erpolat (2012) Apriori ve Fast Grow algoritmalarını kullanarak yaptığı çalışmada, otomobil yetkili servislerinde birlikte satın alınan ürünleri belirlemiştir [13].

Gosain ve Maneela (2013), nicel birliktelik kurallarının çok önemli olduğunu ve gerçek hayatta veri kaynaklarında mutlaka var olacağını savunarak, bu kurallara özgü yeni algoritmalar ortaya koymuşlardır [14]. Shweta ve Garg (2013), Apriori, Tahmin edici Apriori ve Tertius algoritmalarını karşılaştırmıştır. Bu üç algoritmanın karşılaştırılması sonucunda, Apriori algoritmasının diğer iki algoritmadan daha iyi performans gösterdiği

tespit edilmiştir [15]. Aytaç ve Bilge (2013) yaptıkları çalışmada, tele pazarlama verilerini kullanmış ve birliktelik kuralları oluşturmuşlardır. Çalışmalarında sayısal değişkenler, apriori algoritmasının işleyebilmesi için kategorik hale getirilmiş ve pazarlama stratejileri için değerli olan bazı bilgilerin ortaya çıktığı ifade edilmiştir [16].

Doğan, Erol ve Buldu (2014), bir sigorta şirketinin müşterilerine ilişkin toplanan bir veri seti üzerinde birliktelik kurallarını uygulamışlardır. Çalışma sonunda birlikte satın alınan sigorta çeşitleri belirlenmiş ve pazarlama faaliyetlerine yön gösterebilecek bilgilerin ortaya çıktığı raporlanmıştır [17]. Keleş ve Kaya (2014) ise yaptıkları çalışmada duvar işçiliğinde verimliliği etkileyen faktörleri apriori algoritması ile incelemişlerdir [18].

Doğrul, Diyar ve Kurt (2015), trafik kazası verilerini analiz etmek için birliktelik kurallarından faydalanmışlardır. Çalışmada kazaların sıklıkla yaşandığı yerler ve zamanlar belirlenmiş ve bu gibi durumlarda önlemlerin artırılması ile birlikte kazaların önlenebileceği ifade edilmiştir [19]. Kurt, Pehlivanoglu ve Duru (2015) çalışmalarında, bir ortaokulda okuyan öğrencilerin sosyal medya kullanma alışkanlıkları üzerine veri toplamış ve sonuçları birliktelik kuralları yardımıyla analiz etmişlerdir. Çalışmanın sonucunda ortaya çıkan kurallar raporlanmıştır [20].

Chen ve diğerleri (2016) tarafından nicel veri kaynaklarında birliktelik kuralları elde edebilmek için nicel birliktelik kuralları yönteminde “Geçici Bulanıklaştırma” adlı bir yaklaşım ileri sürülmüştür [21]. Kaur ve Kang (2016), apriori algoritmasının zaman faktörünün yok sayıldığı statik veri setleri üzerinde çalıştığına dikkat çekmişler ve bu olumsuzluğu gidermek adına yeni bir algoritma önermişlerdir. Önerdikleri algoritmada sadece statik özellikler değil, zamanla değişen özellikler de dikkate alınmıştır [22].

Yuan (2017), geliştirdiği Apriori algoritmasıyla, tarama sayısını azaltarak, işlem sırasında harcanan zamanı %98'in oranında düşürmüştür. Bu nedenle, yeni geliştirilen Apriori algoritması, orijinal Apriori algoritmasından daha verimli hale getirmiştir [23]. Ateş (2017), minimum destek değerinin uygun seçilememesi, veri setindeki ilginç ve değerli bazı kuralların ya üretilmemesine ya da değerli olmayan çok sayıda gereksiz kural üretilmesine neden olduğundan yola çıkarak, çalışmasında nicel veri setleri üzerinde destek değerinin seçilmesi sorununu giderebilmek için çoklu destek değeri kullanan yeni bir yaklaşım önermiştir [24]. Özçalıcı (2017), Türkiye’deki ikinci el otomobil piyasasında ise

araçların satış fiyatlarını tahmin etmeye yönelik bir çalışmada birliktelik kurallarından faydalanmıştır [25].

Sağın (2018), perakende sektöründe faaliyet gösteren büyük bir hırdavat şirketinin beş buçuk yıllık verileri üzerinde market sepet analizi uygulayarak ilişkili ürün kategorilerini belirlemiştir. Birliktelik kurallarının belirlenmesinde Apriori ve FP-Growth algoritmalarının her ikisi de ayrı ayrı çalıştırılarak bu tip bir veri setindeki kullanışlılıkları kıyaslanmıştır. Değişik boyuttaki veri setleri için yapılan kıyaslamada FP-Growth algoritmasının daha hızlı fakat Apriori algoritmasının daha kapsamlı sonuç verdiği belirtilmiştir [26].

Karamaşa (2018) veri madenciliği yöntemi kullanarak, 2009-2014 yılları aralığında Türkiye’de bayram dönemlerinde trafik polisinin sorumluluğu altındaki bölgelerde meydana gelen trafik kazalarını ve bu kazaların ortak özelliklerini gösteren birliktelik kuralları elde etmiştir. Bulunan kuralların öncül kısmı ile ardıl kısmı arasında pozitif ilişkinin bulunduğu yani kuralın öncül kısmının gerçekleşmesi ile ardıl kısmının olması arasında pozitif yönlü ilişki bulunduğu ifade edilmiştir [27].

Dümrek (2018) çalışmasında, İstanbul Anadolu yakasında bulunan büyük bir perakende zinciri müşterilerinin bir haftalık satış verilerini kullanarak birliktelik kuralları çıkarmış ve Apriori algoritması ile sepet analizi yapmıştır. Satışı arttırmaya yönelik reyonlar da uygulanan çapraz ürün diziliminin tüketici davranışlarıyla uyumu araştırılmıştır. Çalışmanın sonucunda mevcut çapraz ürün listesine öneri bir liste oluşturulup aynı mağazada bir hafta süre ile öneri ürün dizilimi uygulanmıştır. Öneri ürün teşhir listesinin mevcut teşhir listesine %60 oranında satış artışı sağladığı sonucuna varılmıştır [28].

Bu tezin uygulama alanı ise, onarım ve bakım ağırlıklı çalışan fabrikalarda sarf malzeme tedarikinin planlanmasına yardımcı karar mekanizması olarak birliktelik kurallarından faydalanılmasıdır. Üretimin plansız olarak gerçekleştiği bu tip fabrikalarda hangi malzemelerin birlikte temin edilmesi gerektiğinin ve hangi malzemelerin birlikte kullanıldığının tespit edilmesiyle öncelikle temin masrafından ve iş gücünden tasarruf edilecektir. Bununla birlikte ilk defa bu çalışmada, bütün veri seti birbiriyle ilişkili alt gruplara ayrılacak ve bütün veri setinden elde edilen birliktelik kurallarının güvenilirliği, alt veri gruplarından elde edilen kurallarla kıyaslanmak suretiyle test edilecektir.

Böylelikle daha güvenilir olan kurallar elde tutularak, diğerlerinin elenmesi sağlanacaktır. Bununla birlikte, elde edilen nihai kuralların amaca yönelik olarak sınıflandırılması sağlanacaktır. Ayrıca veri setinden mevcut bulunan bazı teknik sayısal değerlerin, elde edilen kural sayılarıyla ilişkili olup olmadığı üzerinde durulacaktır. Bahsi geçen konular literatürde daha önce incelenmemiş olduğundan, bir uygulama eşliğinde yürütülecek olan çalışma, bu konularda özgün bir araştırma örneği sunmaktadır.

3. VERİ MADENCİLİĞİ

Günümüzün yüksek kapasiteli donanımları, büyük boyuttaki verilerin elde edilmesini ve bunların saklanması sağlayabilmektedir. Ama bu kadar büyük hacimli veri içerisinde yer alan önemli bilginin nasıl açığa çıkarılacağı çok önemli bir sorundur. Çünkü bilgi, karar vermenin en önemli girdisidir. Gerçek zamanlı bir bilgi akışının sağlanması ise verilerin çok hızlı toplanmasının yanında, çok hızlı düzenlenmesi ve aynı hızda verilere ulaşılabilmesine bağlıdır. İhtiyaç duyulan anlamlı bilgiye ulaşma işlemi, daha sonra stratejik karar verme sürecinde veya farklı bilimsel bulguların oluşturulmasında kullanılabilir. Binlerce kayıttan oluşan herhangi bir veri seti içinden, anlamlı bilgiye klasik metotlarla ulaşamayacağı ve analizlerin gözlemle yapılamayacağı aşikardır. Bundan dolayı yararlı bilgiye ulaşmak için yapılacak analizlerin, bilgisayar programları aracılığıyla yapılması gerektiği ortaya çıkar. Veri madenciliği bu noktada devreye girer.

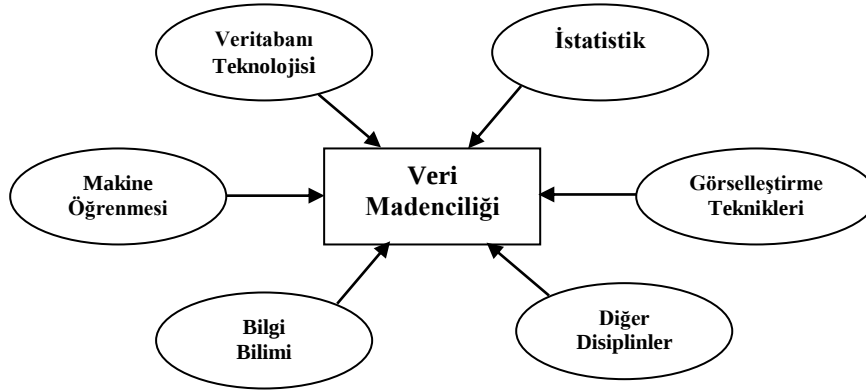
Hand'e göre veri madenciliği; istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ve diğer alanların arayüzünde yatan yeni bir disiplindir. Veri tabanı sahiplerinin ilgisini çekecek ya da değerli olan daha önce beklenmeyen ilişkileri bulmak için büyük veritabanlarının ikincil analizi ile ilgilenmektedir [29]. Jacobs ise veri madenciliğini, işlenmemiş verinin tek başına veremediği bilgiyi sunabilen veri analizi süreci şeklinde açıklamıştır [30]. Başka bir tanımda ise veri madenciliği, kuruluşların karar almaları için ihtiyaç duydukları yardımcı ve yeni bilgileri sunan ya da geleceğe ilişkin tahminler ve planlar yapılabilmelerini sağlayan teknikler bütünü olarak sunulmuştur [31].

Veri madenciliğini Kumar ve Wahidabanu ise, veri setlerinin, daha önce varlığı bilinmeyen, kullanıcının anlayabileceği ve ona fayda sağlayacak ilişkilerin tespit edilmesi ve verinin anlaşılmasını kolaylaştırmak için ve özetlenmesi ve analizinin yapılması olarak tanımlamışlardır. Veri madenciliği sayesinde keşfedilen ilişkiler ve elde edilen metinler, model ya da örüntü olarak adlandırılırlar [32]. Bu tanımlamalar doğrultusunda veri madenciliğinin kullanım amaçları şöyle özetlenebilir:

- Veri setlerinin içerisinde bulunan gizli bilgiyi ortaya çıkartma ve anlamlı bilgileri üretme,
- Verinin özelliklerinden yararlanarak davranış eğilimlerini anlama,
- Tahminlerde bulunarak, bilgiyi gelecekteki ilişkileri yönetmek amacıyla değerlendirme.

Veri madenciliği sayesinde, büyük veri kümelerinden oluşan veritabanı sistemleri içerisindeki anlamlı bilgilere ulaşılabilir. Bu işlem, istatistik, modelleme teknikleri, veritabanı teknolojisi, matematik disiplinleri ve çeşitli bilgisayar programları kullanılarak yapılabilir.

Makine öğrenimi, istatistik ve veri madenciliğinin her biri veri içindeki örüntüleri bulmayı amaçlar. Veri madenciliği algoritmalarında kullanılan yöntemlerin çekirdeğini makine öğrenimi yöntemleri oluşturur. Örneğin, karar ağaçları ve kural çıkartımı hem veri madenciliği algoritmalarında hem de makine öğreniminde kullanılan kullanılmaktadır. Söz konusu ilişki Şekil 3.1’de görülmektedir [34].



Şekil 3.1. Veri madenciliğinin benzer disiplinlerle ilişkisi [34]

3.1. Veri, Bilgi ve Metaveri

Günlük hayatta veri, bilgi ve data gibi kelimeler eş anlamlı olarak kullanılmaktadır. Ancak bilgi, verinin düzenlenmiş ve işlenmiş halidir. Veri, kendi başına hiçbir anlam ifade etmeyen rakam, metin veya semboller kümesidir. Verinin belli bir amaca yönelik olarak işlenmesi sonucunda ise bilgi elde edilir. Dolayısıyla bilgi, bir soruya yanıt vermek için veriden çıkarılan sonuç olarak tanımlanabilir [33]. Başka bir ifadeyle bilgi, potansiyel olarak ilginç ve faydalı olan veriler arasındaki ilişki olarak tanımlanmaktadır. Veriden bilgi elde etme işlemi ise veri çözümlemesi olarak adlandırılmaktadır [34].

$$\text{Veri} + \text{Açıklama} (+ \text{Analiz}) = \text{Bilgi} \quad (3.1)$$

Veriyi bilgiye dönüştürme işlemine veri analizi denir. Veriyi oluşturan sayılara, harflere ve anlamlarına metaveri denir. Metaveri, “veri hakkındaki veri” olarak tanımlanabilir [35].

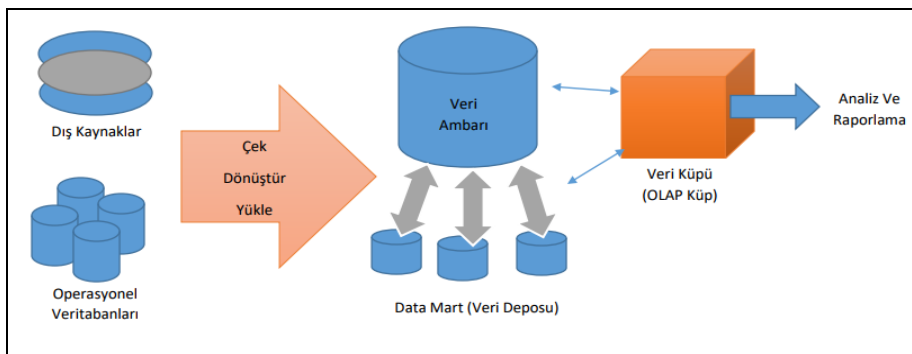
3.2. Veritabanı Sistemleri

Veritabanı düzenli bilgiler topluluğudur. Bilgisayar terminolojisinde veritabanı, sistematik erişim imkânı olan, yönetilebilir, güncellenebilir, taşınabilir, birbirleri arasında tanımlı ilişkiler bulunabilen bilgiler kümesi olarak tanımlanır. Bir başka ifadeyle, bir bilgisayarda sistematik şekilde saklanmış, veri analizi programlarınca işlenebilecek veri yığınıdır [36].

Veritabanı sistemlerinin gelişimi üç ayrı fonksiyon ile değerlendirilmektedir: Veri toplama ve veritabanı oluşturma, Veri yönetimi (veri depolama, bilgi keşfi), Veri çözümlemesi ve veriyi anlama (veri ambarcılığı ve veri madenciliği gerektirir) [34].

3.3. Veri Ambarları

Veri ambarı, canlı bir sistem vasıtasıyla toplanan, sorgulama ve analiz için kullanılmak üzere dizayn edilmiş ilişkisel ve büyük ölçekli verilerin saklandığı depodur [37]. Veri ambarında asıl amaç tarihsel bir yapı oluşturarak verileri depolamak ve bu tarihselliği kullanarak üzerinden daha doğru analizler ve raporlar alabilmektir. Yapılma amacı doğrultusunda kullanılmadığı sürece kendi başına değeri yoktur. Büyük miktarda verilerin saklanmasını sağlayan veri ambarı, ilişkili verilerin sorgulanabildiği ve analizlerinin yapılabildiği geleneksel yöntemlerin yanında veri madenciliği ve çok boyutlu çözümleme gibi güçlü veri çözümleme teknikleri için altyapı sağlar [34]. Genel hatlarıyla bir veri ambarının mimarisi Şekil 3.2’de gösterilmiştir.



Şekil 3.2. Veri ambarı yapısı [37]

3.3.1. Datamart ve veri küpleri

Veri ambarlarının alt bileşenlerinden olan datamart, daha küçük boyutta veri barındıran özelleştirilmiş veri depolarıdır. Konularına göre yapılan analiz sorgulamalarında, alt kümelerine ayrılmış yapılar olması anlaşılabilirlik ve çözümlenebilirlik açılarından kolaylık sağlar. Veriye daha hızlı erişim sağlanır, performans artar. Veri ambarlarından daha hızlı yaratılıp daha hızlı yanıt döndürmesi sorgulama süreç performanslarına etkileri nedeniyle tercih edilmişlerdir [38].

Geleneksel veri modellerinden farklı olarak veri küpleri, veriyi zaman içerisinde sadece bir kez oluşan bir yapı olarak değil, verinin zaman içerisindeki farklılıkları bir seri değişimi olarak modellenir. Veriye bir seri halinde bakılmasını sağlar dolayısıyla detaylı analizin temelini oluşturur [34].

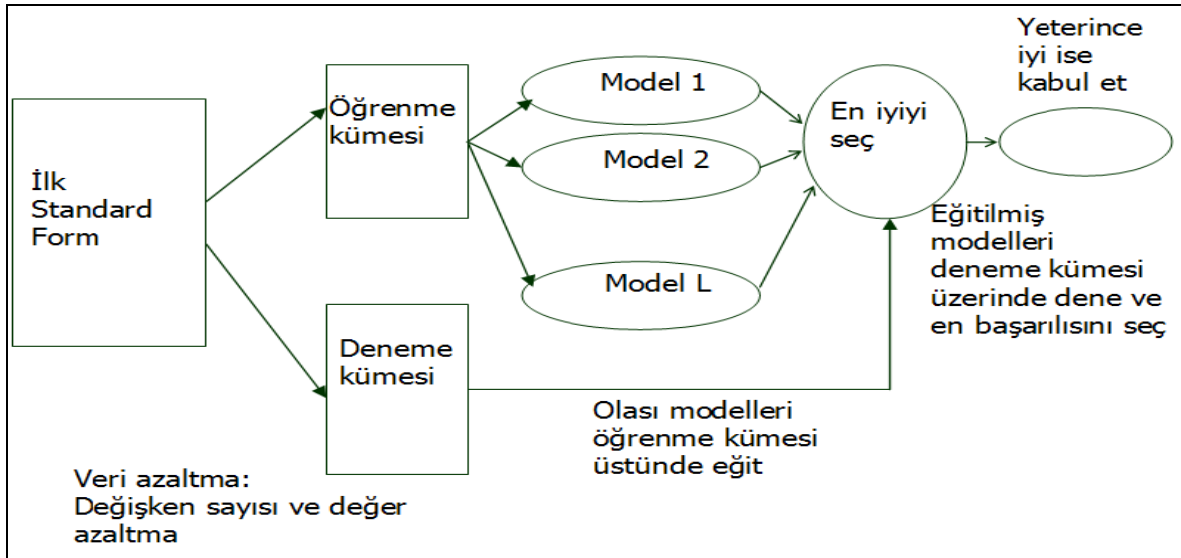
3.3.2. Veri ambarının oluşturulması

Veri ambarının oluşturulması, veri madenciliğinin maliyet ve zamanın önemli bir kısmını oluşturduğundan önemli bir aşamadır. Veri madenciliği uygulanacak verileri öncelikle tek bir çatı altında toplanması gerekmektedir. Sadece verinin tek bir çatı altında toplanması çözüm değildir. Aynı zamanda veriler içindeki hata ve belirsizliklerin temizlenmesi de gerekmektedir. Bu aşamadaki işlemler veri toplama, uyumlandırma, birleştirme ve temizlenme, seçme ve dönüştürmedir [39]. Veri ambarı oluşturma işlemleri aşağıdaki gibi özetlenebilir:

- İlişkili verilerin bir platformda bir araya toplanması,
- Verilerin temizlenmesi, işlevselleştirilmesi, birleştirme ve birbirleriyle ilişkilendirilmesi,
- Uygun formata getirilen verilerin datamart veya veri ambarına yerleştirilmesi,
- Gönderilen tüm verilerin bir veritabanında toplanmasının sağlanması,
- Metaveride bulunan ilişkili bilgilerin saklanması ve kullanıcıların sunulması,
- Veri depolarında bulunan bilgilerin, karar destek amaçlı kullanımının sağlanması [35].

3.4. Veri Madenciliği Metodolojisi

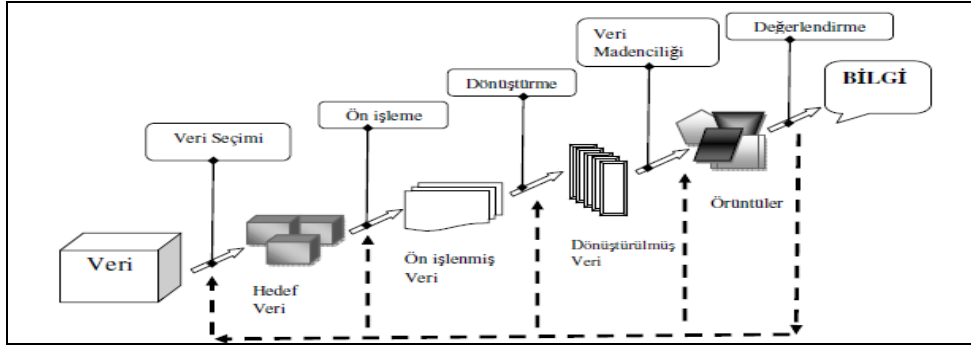
Standart form içinde verilen veri, öğrenme ve deneme olmak üzere ikiye ayrılır. Her uygulamada birden fazla teknik kullanılabilir. Bu tekniklerden hangisinin en başarılı sonucu saylayacağı önceden bilinmemektedir. Bu yüzden öğrenme kümesi üzerinde çok sayıda değişik teknik kullanılarak çok sayıda model oluşturulur. Sonra bu çok sayıdaki model deneme kümesi üzerinde denenerek en başarılı olanı seçilir. Seçilen model yeterince başarılıysa kullanılır, aksi takdirde algoritmanın başına dönülerek çalışma tekrarlanır. Bu sırada başarısız olan örnekler incelenerek bunlar üzerindeki başarının nasıl arttırılabileceği incelenir. Örneğin standart forma yeni alanlar eklenerek programa verilen bilgi arttırılabilir veya var olan bilgi değişik biçimde kodlanabilir ya da amaç farklı şekilde tanımlanabilir [33]. Bir veri madenciliği çalışmasında kullanılan metodoloji Şekil 3.3'te verilmiştir.



Şekil 3.3. Veri madenciliği metodolojisi [33]

3.5. Veritabanında Bilgi Keşfi Süreci

Veri madenciliği, veri yığınları arasında soyut kazılar yaparak veriyi ortaya çıkarmaya yarayan bir süreçtir. Örüntüleri ayrıştırıp süzerek sonraki adıma hazır hale getirmek de bu bilgi keşfi sürecinin bir parçasıdır. Bu süreç Şekil 3.4'te gösterilmiştir. İncelenen veri özelliklerinin bilinmemesi durumunda ise hiç bir veri madenciliği algoritmasının fayda sağlaması beklenemez. Bu sebeple, veri madenciliği sürecine başlamadan önce veri özelliklerinin detaylı analiz edilmesi gerekmektedir [40].

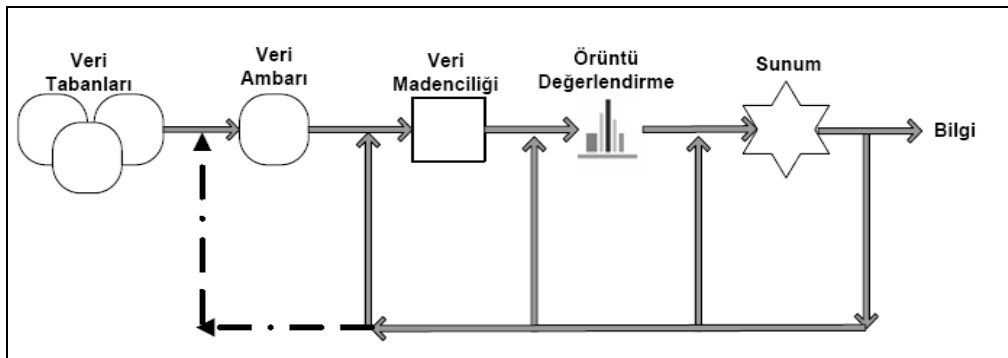


Şekil 3.4. Veritabanında bilgi keşfi süreci [40]

Veri Tabanında Bilgi Keşfi (VTBK) sürecinde izlenmesi gereken temel aşamalar şunlardır;

- Problemin tanımlanması,
- Verilerin hazırlanması,
- Modelin kurulması ve değerlendirilmesi,
- Modelin kullanılması ve
- Modelin izlenmesidir [40].

Veri madenciliği, VTBK işleminin temel bileşenlerinden birisini oluşturmaktadır. VTBK ise veri madenciliği dışında kendi içerisinde farklı süreçleri barındırmaktadır. Genel olarak veri tabanından bilgi keşfi işlemi Şekil 3.5'te görüldüğü gibi 5 aşamadan oluşur [41].



Şekil 3.5. Veri Tabanından bilgi keşfi aşamaları [41]

VTBK sürecini oluşturan adımlar aşağıdaki şekilde sıralanmaktadır [42]:

Veri ön işlemleri: Bu aşamada veri içerisindeki tutarsızlık ve düzensizlikler giderilir. Veri tabanındaki eksik verilere ilişkin düzenlemeler yapılır. Bu işleme veri temizleme denir.

İkinci aşamada veri birleştirme denilen, tutarsızlıklardan arındırılmış verilerin birleştirilmesi işlemi yapılır. Böylece farklı kaynaklardan toplanan verilerin tek bir veri ambarında toplanabilmesi için gerekli düzenleme ve standartlaştırma işlemleri yapılır.

Veri seçme ve dönüştürme: Bu aşamada yapılan işlemler; veri madenciliği konusu ile ilgili bilgi seçimi, verinin türünün belirlenmesi, veriler arasındaki ilişkilerin ve genellemelerin belirlenmesi ve elde edilecek sonuçlar için görselleştirme ve sunum araçlarının belirlenmesi şeklindedir.

Veri madenciliği: Veriler arasındaki saklı ilişkileri ve anlamlı veri örüntülerini ortaya çıkarmak için çeşitli analiz yöntemlerinin ve algoritmaların kullanıldığı aşamadır.

Örüntü Değerlendirme: Bu aşamada elde edilmiş bilgiyi temsil eden ilginç örüntüler tanımlanır ve bu örüntülerin ilginçlik seviyeleri tespit edilir.

Bilgi Sunumu: Elde edilen verilerin, çeşitli sunum ve raporlama araçlarından faydalanılarak kullanıcılara ulaştırıldığı aşamadır.

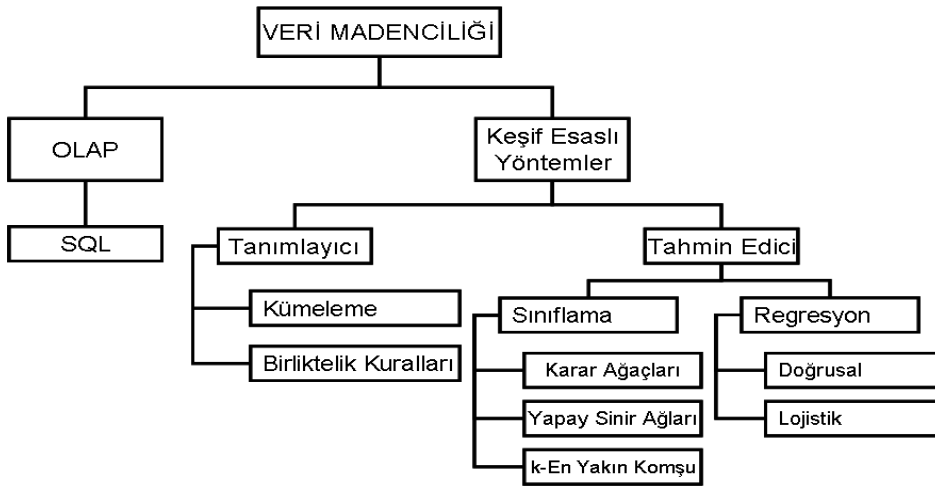
VTBK süreci defalarca tekrar ve aşamalar arası atlamalar ve bu aşamalar arasında ileri geri hareketler içerebilmektedir. VTBK sürecinin tüm aşamaları, işleminin bütünlüğünün sağlanması açısından en az veri madenciliği aşaması kadar önem taşımaktadır [42].

3.6. Veri Madenciliği Modelleri ve Kullanılan Algoritmalar

Veri madenciliğinde kullanılan modeller, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında incelenmektedir. Tahmin edici modeller keşfe dayalı modellerdir. Sonuçları bilinen verilerden faydalanılarak bir model geliştirilir ve geliştirilen bu model kullanılarak sonuçları bilinmeyen veri kümeleri için sonuç tahminleri oluşturulur [40]. Örneğin bir market önceki dönemlerde müşterilere satmış olduğu ürünlere ilişkin gerekli verilere sahiptir. Bu verilerde bağımsız değişkenler ürün alan müşterinin özellikleri, bağımlı değişken ise hangi ürünlerin satıldığıdır. Bu verilere uygun olarak kurulan model, daha sonraki müşteri özelliklerindeki değişime göre ürün tedarik planlamasının yapılması tahmininde kullanılmaktadır.

Tanımlayıcı modellerde ise karar vermeye yardım edebilecek, mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır [40]. Anadolu'daki orta gelirli bir aile ile İstanbul'daki yüksek gelirli bir ailenin yaptıkları eğitim harcamalarının birbirine yakın olmasının belirlenmesi, tanımlayıcı modele bir örnektir. Şekil 3.6'da sunulan veri madenciliği modelleri işlevlerine göre 3 temel grupta toplanır:

- Sınıflama ve regresyon,
- Kümeleme,
- Birliktelik kuralları ve sıralı örüntüler.



Şekil 3.6. Veri madenciliği modelleri [43]

Çeşitli kaynaklarda bu modellerden başka modellere de rastlanmaktadır. Karar verme, ayırma, istisnalar (fark saptaması), metin madenciliği bunlardan bazılarıdır.

3.6.1. Sınıflama ve regresyon

Aynı amaç doğrultusunda kullanılan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Ancak çok değişkenli regresyon gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan tekniklerden de yararlanılması mümkün olmaktadır [44].

Sınıflandırma, yeni bir nesnenin özelliklerinin incelenmesi ve bu nesnenin daha önceden özellikleri belirlenmiş bir gruba atanmasıdır. Veritabanındaki kayıtlar, sınıflandırma işi sonunda bir sınıf kodu içerecek biçimde yeni bir sütun olarak tabloya eklenir.

Sınıflandırmada, sınıflara örnek olarak verilen bir eğitim veri setinin olması ve sınıf tanımlarının iyi yapılması önemlidir. Buradaki asıl amaç, sınıflandırılmamış bir veri girildiğinde bu veriyi sınıflandırabilecek bir model oluşturmaktır. Bu modelde bir tane bağımlı değişken, bir veya birden çok bağımsız (tahmin edici) değişkenler bulunmaktadır. Modeldeki bağımsız değişkenler girdileri meydana getirerek, bağımlı değişkenin yani çıktı değerinin elde edilmesinde kullanılırlar. Bağımlı değişkenin kategorik olduğu durumlarda problem sınıflama probleminden, sayısal olması halinde ise regresyon probleminden söz edilebilir [44].

Buradaki amaç bir malın özellikleri ile müşteri özelliklerini eşleştirmektir. Böylece bir müşteri ile bir ürün arasındaki ilişki anlaşılabilir. “A ülkesinin vatandaşları çoğunlukla üst segment cep telefonu kullanma eğiliminde iken, B ülkesinin vatandaşları orta ve düşük segment cep telefonu kullanma eğilimindedirler.” Örneğin bir cep telefonu firması geçmiş müşteri hareketlerinin analizi ile yukarıdaki gibi kurallar keşfederse A ülkesinde üst segment model telefon reklamına ağırlık verirken, B ülkesinde ise daha düşük segment telefonlarının reklamına ağırlık verecektir. Samsung firmasının, reklamlarını, Türkiye’deki ekonomik kriz sonrasında farklılaştırması bu duruma örnek olarak verilebilir. Aşağıda birkaç sınıflandırma örneği verilmiştir [36]:

- Kredi için başvuran kişilerin düşük, orta, yüksek riskli olarak sınıflandırılması,
- Bir web sayfasında yer alacak içeriğin seçilmesi,
- Hileli sigorta başvurularının belirlenmesi.

Tüm bu örneklerde, herhangi bir kaydın (girdinin) daha önce belirlenmiş sınırlı sayıdaki sınıflardan birisine girebileceği beklenmektedir [45].

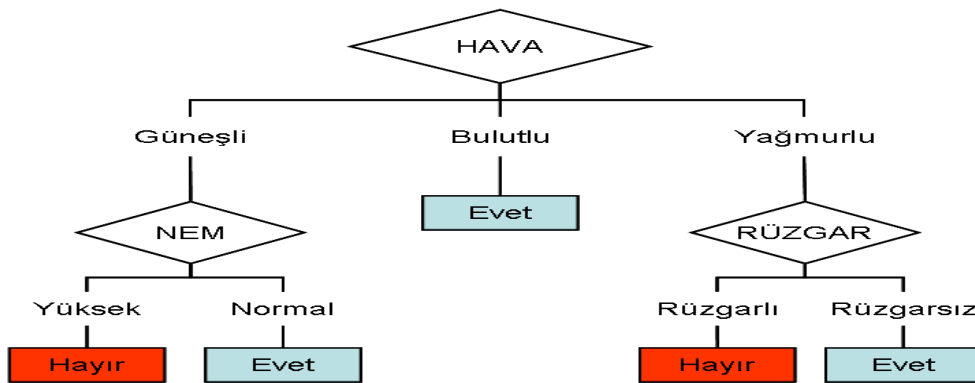
Regresyon modelleri ise, bir değişkenin diğer değişkenler cinsinden tahmin edilmesine yarayan ilişkilerin bulunması amacıyla geliştirilen modellerdir [33].

“Evi olan, aynı iş yerinde beş yıldan uzun zamandır çalışmakta olan, geçmiş kredi borçlarında geç ödeme süresi bir ayı geçmemiş evli bir erkeğin kredi skoru 800’dür.” Kredi başvurusunun değerlendirilmesinde, başvuran kişinin finansal güvenilirliğini tespiti yarayan, örneğin 0 ile 1000 arasında bir skor hesaplanır. Bu skor kişinin özellikleri ve geçmiş kredi hareketlerinden faydalanılarak hesaplanır. Sınıflandırma ve regresyon

modellerinde en çok kullanılan teknikler , karar ağaçları, yapay sinir ağları, genetik algoritmalar, k-en yakın komşu, bellek tabanlı yöntemler, naive-bayes ve lojistik regresyondur [40].

Karar ağaçları: Veri madenciliğinin tahmin edici modeller sınıfında bulunan ve bir ağaç görünümündeki teknik, ağaç yapısı ile kolay anlaşılır kurallar oluşturabilen ve bilgi teknolojileri ile kolay entegre olabilen en sık kullanılan sınıflama tekniğidir. Karar ağaçlarında ana fikir, veri kümesinin elemanlarının gruplara bölünmesine dayanır. Amaç grubun bütün elemanları aynı sınıf etiketine sahip olana kadar işlemi devam ettirmektir. Veri kümesinden karar ağacı oluşturup belirli durumlarda bunun ne yapacağını belirlemek için kullanılır. Ağaç kökten yaprağa doğru inilerek kurallar yazılır. Bu kurallar, sonucun anlamlı olup olmadığı noktasında uzman kişi tarafından denetlenebilir. İstatistiksel yöntemlerde veya yapay sinir ağlarında ise, veriden elde edilen fonksiyonun, anlaşılabilen bir kural olarak yorumlanması zordur [33].

Karar ağaçlarında, her düğüm bir nitelik üzerinde gerçekleştirilen testi, dallar bu testin çıktısını, yaprak düğümler ise sınıfları belirtir. En üstteki düğüm kök düğümdür. Karar ağaçları, kök düğümden yaprak düğüme doğru çalışır [43]. Şekil 3.7’de hava durumuna göre piknik yapıp yapmamaya karar veren karar ağacı gösterilmiştir. Şekil 3.7’de sunulan karar ağacından aşağıdaki kurallar türetilir:



Şekil 3.7. Örnek karar ağacı [43]

- Eğer hava güneşli ve nem de yüksek ise piknik yapma,
- Eğer hava güneşli ve nem de normal ise piknik yap,
- Eğer hava bulutlu ise piknik yap,

- Eğer hava yağmurlu ve rüzgârlı ise piknik yapma,
- Eğer hava yağmurlu ama rüzgârsız ise piknik yap.

Yapay sinir ağları: Yapay sinir ağları, yapay zekâ kavramının alt konularından biridir. Biyolojik sinir ağlarından esinlenilerek geliştirilen bir sınıflandırma işlemidir. Yapay sinir ağları örneklerle ilgili topladığı bilgilerden yararlanarak genellemeler yapmakta ve daha sonra hiç görmediği örneklerle karşılaştığı zaman, öğrendiği bilgileri kullanarak o örnekler hakkında karar verebilen bir modeldir. Yapay sinir hücreleri birbirleriyle çeşitli şekillerde bağlantılar içerirler. Bu bağlantılar gerçekleşme noktası olan yapay nöron bağlantıları arasındaki katsayılar hesaplanarak gerçekleştirilir. Oluşan nöronların bir de iç katsayıları mevcuttur. İç katsayıya aktivasyon seviyesi denir. Aktivasyon seviyeleri ve bağlantı katsayıları ile birlikte oluşturulan yapay sinir ağları sayesinde de sınıflandırma işlemi yapılabilir. İleri yayımlı ve geri yayımlı olmak üzere iki yöntemle oluşturulabilirler [44].

Genetik algoritma: Evrim teorisine dayanan genetik algoritmalar ve evrimsel programlama, doğal evrimdeki gözlemlenen prensiplerden esinlenerek oluşturulmuş en iyileme yöntem bazlı stratejilerdir. Bu yöntemde problem çözümüne, rastgele bir başlangıç çözümü olarak başlanır. Devamında alternatif çözümler üretilir. Daha sonra elde edilen bu çözümler birbirleri ile eşleştirilerek performansı daha yüksek çözümler oluşturulmaktadır. Aynı biçimde elde edilen sürekli çözümlerin birbirleriyle birleştirilmesi tüm çözüm kümeleri için tekrarlanarak yeni çözümler aranır. Böylelikle, tıpkı evrim teorisi mantığında olduğu gibi, iyi çözümlerin bulunduğu çözüm kümesinin daha iyi hale getirilmesi amaçlanmaktadır. Özetle, tüm çözüm uzayından rastgele seçilen bir çözüm üzerinden operatörler aracılığıyla daha iyi çözümlere ulaşılması amaçlanır. Bu arama daha iyi çözüm elde edilemeyece kadar sürer. Çözüm uzayı büyük olan problemlerde, hızlı bir şekilde çözüme ulaşmak için, arama sırasında iyi sonuç vermeyecek alanların dikkate alınmaması gerekmektedir [46].

K-en yakın komşu: Veri uzayında birbirine yakın benzer tip kayıtların, birbirlerinin komşusu olacağından hareket ederek, kolay ve güçlü bir algoritma olan k-en yakın komşu algoritması geliştirilmiştir. K-en yakın komşu algoritmasının işleyişi, komşu davranışlarının girdi olarak kullanılması ve sonucunda hedeflenen kaydın davranışının tahmin edilmesine dayanır. K-en yakın komşu algoritmasındaki k harfi ilgilenilen komşu sayısını tanımlar. Özetle, bir bireyin (kayıtın) davranışı tahmin etmek istenirse, veri

uzayında o bireye en yakın k bireyin davranışı incelenmelidir. Bu k komşunun davranış ortalaması hedef kaydın davranış tahmini olacaktır [46].

Bellek tabanlı yöntemler: Öğrenen organizma mantığına dayanmaktadır. Örneğin bir araba tamircisi elde ettiği bulguları ve daha önce tamir ettiği benzer arıza deneyimlerinden faydalanarak çözüme ulaşır. Benzer şekilde bellek tabanlı yöntemler de önceki deneyimlerini kullanırlar. Bu yöntemde, önceki kayıtların özelliklerinden faydalanarak bir veritabanı oluşturulur. Yeni gelen bir kayda benzer olan kayıtları sistem tarafından belirlenir ve bu benzer kayıtlar kullanılarak ilgili kayda yönelik tahminde bulunur veya bir sınıflama işlemi yapılır. Bellek tabanlı yöntemlerin en önemli özelliği, veriyi olduğu haliyle kullanabilmesidir. Bellek tabanlı yöntemler diğer veri madenciliği yöntemlerinden farklı olarak, kayıtların şekliyle değil, iki kayıt arasındaki uzaklığı belirleyen bir uzaklık fonksiyonuyla ve komşu kayıtların işlenmesi neticesinde bir sonuç üreten bir kombinasyon fonksiyonuyla ilgilenir [37]. Bellek tabanlı yöntemler sahtekârlık tespiti ve klinik işlemler gibi alanlarda kullanılmaktadır.

Naive bayes: Bu yöntemde, her olasılığın öğrenme kümesinde kaç kere meydana geldiği hesaplanır. Öncelikli olasılık olarak adlandırılan bu değer, sınıflandırma işlemine esas olur. Örneğin; bir banka kredi başvurularını “olumlu” ve “olumsuz” olarak sınıflandırmayı amaçlamaktadır. Olumlu risk çıktısı 10 olay içinde 2 defa oluşmuşsa, olumlu risk için öncelikli olasılık 0,2’dir. Böylelikle “Kredi başvurusunda bulunan birine ilişkin hiçbir şey bilinmiyorsa, bu kişi 0,2 olasılıkla olumlu risk sınıfındadır” şeklinde yorumlanır. Naive Bayes aynı zamanda öncelikli olasılıklarla birleştirilmek suretiyle tahminde kullanılan, bağımsız/bağımlı değişken kombinasyonunun meydana gelme sıklığını da bulur [46].

Lojistik regresyon: Regresyon analizi, hedef değişken ile ilgili bağımsız değişken arasındaki ilişkiyi sembolize eden matematiksel modeli elde etmeye yarayan bir yöntemdir. Veri madenciliğinde sıkça kullanılan doğrusal regresyondan farkı; doğrusal regresyonda tahmin edilecek hedef değişken sürekli değer alırken, lojistik regresyonda hedef değişkenin kesikli bir değer almasıdır. Diğer bir ifadeyle, doğrusal regresyonda hedef değişkenin değeri tahmin edilmektedir. Buna karşılık lojistik regresyonda ise hedef değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilmektedir [47].

3.6.2. Kümeleme

Dağınık yapıda bulunan bir kitleyi, nispeten daha homojen yapıdaki alt gruplara ayırma işlemine kümeleme denir. Kümeleme modellerinde amaç, aynı küme içerisindeki üyelerin birbirlerine benzediği, ancak farklı özellikte birden çok kümenin belirlenmesi ve veri tabanındaki kayıtların bu farklı kümelere atanmasını sağlamaktır. Kümeleme analizi, nesnelerin alt dizinlere gruplanmasını yapan bir işlemdir. Kümeleme, bir denetimsiz öğrenme yöntemidir [43].

Sınıflama ile kümelemeyi birbirinden ayıran en önemli fark, sınıflama da yeni sınıflar tanımlanırken, kümelemede yeniden tanımlanmış sınıflar yoktur. Sınıflamada veriler, bir eğitim neticesinde belirlenen bir modele göre önceden tanımlanmış bir sınıfa atanmaktadır. Kümeleme işlemi, sınıflama işlemindeki gibi önceden belirlenmiş sınıflara göre bölme yapmaz. Verilerin kümelenmesi işlemi, verilerin birbirlerine olan benzerliklerine göre yapılmaktadır. Kümeleme neticesinde meydana gelen sınıfların hangi özellikte olduğunun belirlenmesi, analizcinin çıkan sonuçları yorumlamasıyla ortaya çıkar [43].

Başka bir ifadeyle, kümeleme modelinde, sınıflama modelinde olan veri sınıfları yoktur [48]. Sınıflama modelinde, verilerin önceden belirlenmiş sınıfları vardır ve yeni gelen bir verinin hangi sınıfa dahil olabileceği tahmin edilmektedir. Kümeleme modelinde ise, önceden tanımlı bir sınıfın olmadığı durumda, verilerin bir model vasıtasıyla belirlenen benzerlik-yakınlık kriterlerine göre gruplar halinde kümelere ayrılması işlemi yapılır. Aynı küme içindeki elemanların birbirlerine benzerliği fazla, kümelerin ise bir bütün olarak birbirlerine benzerlikleri az olmalıdır [43].

Büyük veritabanlarının kümelenmesinde oldukça etkin bir algoritma olan k ortalamalar algoritması en sık kullanılan kümeleme algoritmasıdır. Bu algoritmada ortaya çıkan yeni bir vakada, algoritma tüm veriyi inceler ve bu vakaya en çok benzeyen vakaların bir alt kümesini oluşturur. Oluşturulan alt küme çıktıyı tahmin etmek için kullanır [49].

K ortalamalar yönteminde öncelikle tüm nesnelerin içerisinde rastgele olarak k adet nesne seçilir. Seçilen k adet nesnenin her biri, yeni oluşan kümelerin merkezi kabul edilir. Geriye kalan nesneler oluşturulan bu kümelerin hangisinin merkezine daha yakınsa o kümenin elemanı olur. Eleman sayısı değişen her küme için ortalama değer hesaplanır ve

hesaplanan bu değer o kümenin yeni merkezi olur. Bu işlem açıkta nesne kalmayıncaya kadar devam eder.

3.6.3. Birliktelik kuralları ve ardışık zamanlı örüntüler

Birliktelik kuralları, büyük veri kümelerinden ilk bakışta çıkarılamayacak, işe yarar, tutarlı ve ilginç ilişkilerin tespiti için kullanılır. Birliktelik kuralları diğer bir ifadeyle, veri kümesi içindeki sıkça görülen durumların tespiti işine yarar. Birliktelik kurallarından en sık faydalanan, tipik ve genel kullanım alanı market sepet analizidir [34].

Birliktelik kuralı algoritmalarının temel amacı, iki ya da daha çok olayın bir arada meydana gelme olasılıklarının belirlenmesidir. Bu yöntemin en sık kullanıldığı market sepet analizinde satın alma eğilimlerinin tanımlanması amaçlanır. Bununla birlikte tıp, finans, pazarlama, sigorta, bankacılık, eğitim ve daha birçok alanda bu tekniklerden yararlanılmaktadır. Veritabanındaki verilerin sürekli veya kategorik olmalarına bakılmaksızın birliktelik ilişkilerinin ortaya çıkarılabilmesi, bu tekniğin fen bilimleri, sosyal bilimler, sağlık bilimleri vb. gibi alanlarda kullanımını kolaylaştırmış ve artırmıştır. Birliktelik kuralları, eş ya da ardışık zamanlı gerçekleşen ilişkileri tanımlar. Aşağıda bu teknikte çıkarılan sonuçlara ilişkin birkaç örnek sunulmuştur.

- Müşteriler gazlı içecek satın alırlarsa, % 85 olasılıkla cips de satın alırlar.
- Yarım yağlı süt alan müşteriler, % 60 olasılıkla kepek ekmek satın alırlar [40].

Ardışık zamanlı örüntüler ise birbirini izleyen dönemlerde gerçekleşen ilişkilerdir. Aşağıda birkaç ardışık zamanlı örüntü için birliktelik kuralları örnek olarak sunulmuştur.

- Erkek hastalara X ameliyatı yapıldıktan bir ay sonra % 75 ihtimalle Y enfeksiyonu oluşur.
- İMKB endeksi yükselirken X hisse senedi % 20'den daha değer kazanırsa, bir hafta içerisinde Y hisse senedinin değeri de % 75 ihtimalle artacaktır.
- Çekiç satın alan bir müşteri o anda, % 90 ihtimalle, ilk altı ay içerisinde % 10 ihtimalle, bir yıl içerisinde % 45 ihtimalle çivi satın alacaktır [40].

4. BİRLİKTELİK KURALI

Birliktelik kuralları, bir veritabanındaki bazı olayların birlikte görülmesinin olasılık yapılarıyla gösterilmesidir. Birliktelik kuralları verinin potansiyel ilişkilerini tanımlar. Amaç birlikte gerçekleşmesi muhtemel olayların kurallarının ortaya konmasıdır [32].

Birliktelik kurallarının anlaşılabilirliği her zaman bu yöntemin en güçlü yönünü oluşturmuştur. Bununla birlikte kullanıcı, meydana gelecek birliktelik kurallarının tipini, sayısını, olasılık oranını, öncülünü ve ardılını kontrol edebilmektedir. Buna örnek olarak kuralların destek ve güvenilirlik seviyelerine göre filtrelenmesi ya da girdi ve çıktı sayılarının istenen düzeyde tutulabilmesi gösterilebilir. Klasik anlayışta güçlü bir kural büyük destek ve yüksek güven oranına sahiptir. Birliktelik kurallarının iki aşamalı bir eleme sistemi vardır. Bu yöntemde kullanıcı tarafından belirlenen destek ve güven eşik değer oranlarını yakalayan veya geçebilen kurallar bulunur. Başka bir eleme yöntemi de kullanıcı tarafından sınırlanmış nesne tanımlanmasıdır. Sınırlanmış nesne, girdide veya meydana gelecek kurallarda bulunması istenilen kayıtları ifade eder. Örneğin 100 çeşit malzemenin bulunduğu bir veri seti sadece beş çeşit malzeme ile sınırlanırsa bu model sadece ilgili beş çeşit malzemeye ilişkin kurallar çıkarır [50]. Agrawal ve Srikant bu yöntemle, sınırlanmış nesne ile kurallar için çeşitli etkin çözümleme teknikleri geliştirmişlerdir [51].

Birliktelik kuralı yönteminde geçmiş veriler analiz edilir. Bu aşamada sağlanan veriler sayesinde birliktelik davranışlarının tespiti yapılarak geleceğe yönelik karar almayı kolaylaştıran bir disiplindir. Günümüzde veri toplama işlemi artık çok kolaylaşmıştır. Bilgi sistemlerindeki gelişmeler sayesinde sadece veri toplama değil, veriye ilişkin nicelik ve nitelik yönünden tüm ayrıntılar da elde edilebilmektedir. Böylelikle tek bir cins veriye ait onlarca alt veri de elde edilebilmektedir. Bununla birlikte firmaların satış noktalarında, yeni teknoloji otomatik ürün veya müşteri tanıma sistemleri yaygınlaşmıştır. Böylece, herhangi bir olaya ait verilerin olay esnasında toplanması ve elektronik ortamlara aktarılması otomatik olarak sağlanmaktadır. Bahsedilen bu olaylardan (örneğin satış) sağlanan verilerde, birçok bilgi yer alır. Ancak bu bilgilerin birçoğu ilk bakışta anlaşılabilir. Bu kapsamda veri madenciliğine, bilgi keşfi sürecine ihtiyaç vardır. Günümüzde birçok kuruluş bu tip bilgileri içeren veritabanlarını, kurumun en önemli parçalarından biri olarak görmektedir. Çünkü tüm gelecek planlarını bu verilerden

sağladıkları bilgi ve tahminlere göre yapmaktadırlar. Dolayısıyla bu verileri düzgün ve doğru bir şekilde kullanabilmek için ciddi yatırımlar yapmaktadırlar.

Birliktelik kuralında, örneğin market sepet analizindeki amaç, müşterilerin alışveriş esnasında satın aldıkları ürünlere ilişkin birliktelik ilişkilerini keşfederek, müşterilerin alışveriş eğilimlerinin tespit edilmesidir. Kuruluşlar, keşfedilen bu eğilimler sayesinde çeşitli yönlerden kazançlarını artırabilmektedirler. Örneğin bir markette eğer patates cipsi ile kola satışı arasında güçlü ve anlamlı bir bağlantı ortaya çıkarılmış ise, bu iki ürünün satışının artırılması için bu ürünlere yönelik mekansal düzenlemeler veya kampanyalar yapılabilir. Market sepet analizinin gerçekleştirilmesiyle, birliktelik ilişkilerinin keşfedilmesi, market yöneticilerinin daha etkili satış stratejileri geliştirmelerine olanak sağlar. Market yöneticileri kola ve patates cipsi raflarını yan yana koyarak ya da kola alanlara patates cipsi fiyatında indirim yaparak kampanya düzenleyip bu ürünlerin satışlarını arttırabilir. Nitekim birliktelik kurallarının uygulandığı en tipik örnek market sepeti analizidir.

4.1. Birliktelik Kuralının Matematiksel Gösterimi

Birliktelik kurallarında sıklıkla kullanılan önemli terimler;

Öncül; kuralın sol tarafını ifade eder. Koşul durumunu belirtir. Öncül meydana gelmişse sonuç olasılıkları incelenmeye başlanır.

Ardıl; kuralın sağ tarafını ifade eder izleyen de denir. Öncülün meydana gelmesini takip eden durumlardır.

Minimum destek değeri; belli bir kaydın tüm veri setinde görülme sıklığının eşik değerini ifade eder, “min_des” olarak gösterilir.

Minimum güven değeri; belli bir olayın oluşması durumunda, oluşabilecek diğer olayın minimum eşik değerini ifade eder, “min_güv” olarak gösterilir.

Birliktelik kuralının matematiksel modeli 1993 yılında Agrawal, Imielinski ve Swami tarafından ifade edilmiştir. Çalışmada birliktelik kuralları, X ve Y'nin nesne kümeler

olduğu $X \rightarrow Y$ (X birliktelik Y) şeklinde ifade edilmiş ve birliktelik kurallarının matematiksel şekli belirlenmiştir. Birliktelik kurallarının bir anlam ifade etmesini, kullanışlı ve güçlü olmalarını sağlamak amacıyla minimum destek ve minimum güven değerlerini geliştirmişlerdir. Böylece kullanıcı tarafından keyfi belirlenen minimum destek ve devamında minimum güven değerlerini sağlayan bütün birliktelik kurallarının keşfedilmesini sağlamışlardır [52].

Bu modelde; $I = \{i_1, i_2, i_3, \dots, i_m\}$ nesnelerin kümesini, D ise işlemler kümesini ifade eder. Her i , bir nesne (kayıt) olarak adlandırılır. D veritabanında her işlem T , $T \subseteq I$ olacak şekilde tanımlanan nesnelerin kümesi (nesne küme) olsun. Her işlem bir tanımlayıcı alan olan TID ile temsil edilir. A ve B nesnelerin kümeleri olsun. Bir T işlemler kümesi ancak ve ancak $A \subseteq T$ ise yani A , T 'nin alt kümesi ise A 'yı kapsıyor denir. Bir birliktelik kuralı $A \Rightarrow B$ formunda ifade edilir. A öncül ve B ardıl (izleyen) olarak adlandırılır.

$$\text{Burada; } A \subset I, B \subset I \text{ ve } A \cap B = \emptyset \text{ dır.} \quad (4.1)$$

Elde edilen kuralın, eldeki hareketler kümesi tarafından belirli bir oranda desteklenmesi istenir. Bu nedenle, birliktelik ilişkisini oluşturan eşleştirme kuralı, kullanıcı tarafından belirlenmiş destek ve güvenilirlik eşik değerlerini sağlayacak ya da aşacak biçimde üretilir. İki kayıt (ürün) X ve Y için destek ve güven değerlerinin hesaplanması aşağıda sunulan formüller aracılığıyla yapılır [40].

$$\text{Destek: } P(X \text{ ve } Y) = X \text{ ve } Y \text{'nin kullanıldığı kayıt sayısı} / \text{Toplam kayıt sayısı} \quad (4.2)$$

$$\text{Güven: } P(Y|X) = X \text{ ve } Y \text{'nin kullanıldığı kayıt sayısı} / X \text{'in toplam kayıt sayısı} \quad (4.3)$$

Destek oranı, bu bağıntının veritabanında hangi sıklıkta olduğunu, güven oranı ise X ürününü kullanmış bir kişinin hangi olasılıkla Y ürününü kullanacağını ifade eder. Bu iki değerinde yüksek olması söz konusu bağıntının ciddiyetini artırır [40].

Birliktelik kuralı formüsel olarak şu şekilde tanımlanabilir;

$$X_1, X_2, \dots, X_m \Rightarrow Y_1, Y_2, \dots, Y_n \quad (4.4)$$

Buradaki, X_j ve Y_j yapılan iş veya nesnelerdir. Bu kural genellikle “ $X_1, X_2, \dots, X_m$ gibi herhangi bir X olayı meydana geldiğinde, buna bağlı olarak $Y_1, Y_2, \dots, Y_n$ gibi bir olayın da aynı olay veya hareket içinde yer aldığını söyler [40]. Örneğin aşağıdaki kural bir mağazanın satış hareketlerinden elde edilmiştir.

Ürün(X , cep telefonu) $\Rightarrow$ Ürün(Y , koruyucu kılıf)

Burada X bir hareketteki değişkeni simgelemektedir. Bu kural, cep telefonu alan müşterinin aynı zamanda koruyucu kılıf da almaya yöneldiği anlamına gelir. Çizelge 4.1’de 3 farklı ürün (A, B, C) için farklı olasılıklar tablosu sunulmuştur. Aynı şekilde farklı sayılardaki kombinasyonlar için de birliktelik kuralları çıkarılabilmektedir.

Çizelge 4.1. Olası birliktelik kuralları

Nu.	X	$\rightarrow$	Y
1	A	$\rightarrow$	B
2	A	$\rightarrow$	C
3	A	$\rightarrow$	B, C
4	B	$\rightarrow$	A
5	B	$\rightarrow$	C
6	B	$\rightarrow$	A, C
7	C	$\rightarrow$	A
8	C	$\rightarrow$	B
9	C	$\rightarrow$	A, B
10	A, B	$\rightarrow$	C
11	A, C	$\rightarrow$	B
12	B, C	$\rightarrow$	A

Nesne sayısı arttıkça elde edilmesi muhtemel birliktelik kuralı sayısı da artar. Örneğin, nesne sayısı 4 olan bir kümeden, minimum destek ve minimum güven kriterleri sağlandığı takdirde 50 adet kural çıkabilir. Nesne sayısına (d) göre ortaya çıkması muhtemel birliktelik kuralı sayısı (R) aşağıdaki formüle göre hesaplanabilir [53].

$$R = 3^d - 2^{d+1} + 1 \quad (4.5)$$

4.1.1. Destek ve güven kavramları

Kullanıcı tarafından belirlenen destek ve güven değerleri, kuralın meydana gelebilmesi için ulaşılması veya aşılması gereken iki engeldir. İlgili bir kayıt hakkında birliktelik kuralı elde edilmek isteniyorsa öncelikle destek eşik değeri, bu aşılabılırsa güven eşik değeri

aşılmalıdır. Destek değeri, keşfedilen kuralların yararlılığını, kullanmaya değer olup olmadığını ortaya koyarken, güven değeri de keşfedilen kuralların kesinliğini (yüzdesel olasılığını) ifade eder. Destek ve güven değerlerinin örnek gösterimi şu şekildedir:

$$A \Rightarrow B \text{ [destek} = \% 5, \text{güven} = \% 80] \quad (4.6)$$

Destek değerinin hesaplanması için geliştirilen formül şu şekildedir:

$$\text{Destek } (A \Rightarrow B) = (A \text{ ve } B \text{'nin bulunduğu satır sayısı}) / (\text{toplam satır sayısı}) \quad (4.7)$$

Destek değerinin %5 olması; satılan tüm satışların (veritabanı kayıtlarının) en az %5’inde A ve B birlikte bulunmaktadır.

Güven değeri aşağıdaki formüle göre hesaplanır:

$$\text{Güven } (A \Rightarrow B) = (A \text{ ve } B \text{'nin bulunduğu satır sayısı}) / (A \text{'nın bulunduğu satır sayısı}) \quad (4.8)$$

Güven değerinin %80 olması; “A ürünü kullananların (satın alanların) %80’i, B ürününü de kullanmışlardır (satın almışlardır).” anlamını taşır. Güven değerinin %100 olması demek A ürünün alan her kişinin mutlak suretle B ürününü de aldığı anlamına gelir. Böyle kurallara kesin kural adı verilir.

Örneğin, bir markette alışveriş yapan 5 müşterinin aldığı ürünler Çizelge 4.2’deki gibi olduğu varsayılırsa;

Çizelge 4.2. Ürün satış bilgileri

Müşteri	Alınan ürünler
1	Pirinç, yağ, meşrubat, zeytin
2	Pirinç, şehriye
3	Şehriye, yağ, meşrubat, süt
4	Pirinç, şehriye, yağ, meşrubat
5	Pirinç, şehriye, yağ, süt

Bu veriler kapsamında gerçekleştirilen birliktelik analizi neticesinde aşağıda sunulan tarzda kurallar elde edilir:

- {şehriye, meşrubat} $\rightarrow$ {yağ} (Destek=0,4 Güven=1,0),
- {şehriye, yağ} $\rightarrow$ {meşrubat} (Destek=0,4 Güven=0,67),
- {yağ, meşrubat} $\rightarrow$ {şehriye} (Destek=0,4 Güven=0,67),
- {meşrubat} $\rightarrow$ {şehriye, yağ} (Destek=0,4 Güven=0,67),
- {yağ} $\rightarrow$ {şehriye, meşrubat} (Destek=0,4 Güven=0,5),
- {şehriye} $\rightarrow$ {yağ, meşrubat} (Destek=0,4 Güven=0,5). Bu sonuçlar da aşağıdaki şekilde yorumlanabilir:

- Şehriye ve meşrubatı birlikte alanlar mutlaka yağ da almaktadır.
- Şehriye ve yağ satın alan müşteriler %67 olasılıkla meşrubat da almaktadır.
- Yağ ve meşrubatı birlikte alan müşteriler %67 olasılıkla şehriye da almaktadır.
- Meşrubat alanlar %67 olasılıkla şehriye ve yağ da satın almaktadır.
- Yağ alanlar %50 olasılıkla şehriye ve meşrubat da almaktadır.
- Şehriye alanlar %50 olasılıkla yağ ve meşrubat da almaktadır.

Destek ve güven değerlerinin matematiksel gösterimi aşağıdaki gibidir.

$$\text{Destek } (A \Rightarrow B) = P(A \cup B) \quad (4.9)$$

$$\text{Güven } (A \Rightarrow B) = P(B \setminus A) \text{ veya} \quad (4.10)$$

$$\text{Güven } (A \Rightarrow B) = \text{Destek } (A \Rightarrow B) / \text{Destek } (A) \quad (4.11)$$

$$\text{Burada, Destek } (A) = \text{Destek } (A \Rightarrow A)' \text{ dir.} \quad (4.12)$$

Başka bir ifade ile destek ve güven değerleri aşağıdaki şekilde tanımlanır:

$$\text{Destek } (A) = |A| / |D| \quad (4.13)$$

$$\text{Destek } (A \Rightarrow B) = |A.B| / |D| \quad (4.14)$$

$$\text{Güven } (A \Rightarrow B) = \text{Destek } (A \Rightarrow B) / \text{Destek } (A) \quad (4.15)$$

Burada; $|A|$, incelenen kayıtlardaki A ürününü içeren işlemlerin sayısını, $|A.B|$, incelenen kayıtlardaki A ve B ürünlerini birlikte içeren işlemlerin sayısını, $|D|$, veritabanındaki bütün işlemlerin sayısını ifade etmektedir.

Daha önce de ifade edildiği üzere, destek değeri, A ve B nesnelerinin birlikte bulunma olasılıklarını ifade ederken, güven değeri ise, A'yı içeren kayıtların B'yi de içermesi olasılığını ifade eder. Yüksek destek ve güven değerine sahip kurallara güçlü kurallar adı verilir [52].

Kullanıcı tarafından belirlenen minimum destek ve minimum güven eşik değerlerini aşmak suretiyle oluşan birliktelik kuralları ilginç olarak ifade edilir. Ortaya çıkarılan bu ilginç örüntü, bilgi olarak kullanıcıya sunulur. Bu değerler genel olarak %0 - %100 aralığında ifade edilirler [34].

Eğer bir küme k tane nesne içeriyorsa bu kümeye “k-nesneküme” denir. Minimum destek eşik değerini sağlayan nesneküme de “sık geçen nesneküme” adı verilir. K-nesnekümenin sık geçen kümeleri L_k şeklinde ifade edilmektedir [34].

Minimum destek ve güven eşik değerlerinin aşılması iki farklı aşamada sağlanır. Bu iki eleme kriterini geçen nesnekümelerinden birliktelik kuralları çıkarımı yapılabilir. Söz konusu iki aşama aşağıda özetlenmiştir.

- Sık geçen nesnekümeler bulunur: Kullanıcı tarafından belirlenmiş olan minimum destek eşik değerini sağlayan nesnekümelere sık geçen nesneküme adı verilmektedir [52]. Buradaki sıklık ifadesi kullanıcı tarafından belirlenen destek eşik değerine göre şekillenen subjektif bir terimdir.

- Sık geçen nesnekümelerin içerisinden güçlü birliktelik kuralları oluşturulur: Sık geçen nesnekümelerini kullanarak minimum güven eşik değerini sağlayan birliktelik kurallarının bulunmasıdır. Kullanıcının ilgilendiği ve potansiyel olarak önemli bilgiye, sık geçen nesnekümelerinin içinden, minimum güven eşik değeri kriterine uygun olan birliktelik kurallarının çıkarılmasıyla ulaşılır [52].

Veri kümelerinin ortaya çıkarılmasını sağlayan algoritmalarda, bir veri birçok defa sayılır. İlk sayımda her bir verinin destek oranları teker teker belirlenir. Destek oranı en az olan veri kümeleri belirlenir. Sonraki sayımlarda bulunmuş veri kümelerinden başlanır. Her sayımda bulunmuş kümeler, yeni aday veri kümesi diye adlandırılan daha da büyük kümeleri yaratmakta kullanılır. Bulunan her yeni aday kümenin destek oranları belirlenir. Sayım sonunda en büyük aday veri küme ortaya çıkarılır. Bu aday küme sonraki sayımında kullanılmak üzere hazır tutulur. Bu işleme, yeni ve daha büyük veri kümesi elde edilemeyene kadar devam edilir [51].

4.2. Birliktelik Kurallarında Kullanılan Algoritmalar

Büyük veri yığınlarından birliktelik kurallarının ortaya çıkarılabilmesi için günümüze kadar çok çeşitli algoritmalar geliştirilmiştir. Bunlar içerisinde, konuya ilişkin ilk algoritma olan AIS algoritması 1993'te [52], bundan daha sonra SETM algoritması [54] ve 1994'de ortaya atılan ve günümüze kadar en sık kullanılan algoritmalar olan Apriori ve AprioriTid [51] algoritmaları geliştirilmiştir.

4.2.1. AIS (Agrawal, Imielinski, Swami) algoritması

1993 yılında geliştirilen AIS algoritması, geniş nesne kümeleri üretme mantığına dayanan bir algoritmadır. AIS algoritması veritabanını birçok defa tarar. Bu taramalar sırasında tüm işlemleri okur. İlk taramada veritabanındaki nesneleri, diğer nesnelerden bağımsız olarak her birini tek başına sayar. Böylece hangi nesnelerin, geniş nesneler (çok rastlananlar) olduğu belirlenerek, bunlardan geniş olanlar aday nesne kümeleri olarak işaretlenir. Bir tarama işlemi gerçekleştirildikten sonra elde edilen nesne kümeleri ile bir önceki taramada geniş oldukları belirlenen nesne kümelerinin arasındaki ortak nesne kümeleri belirlenir. Belirlenen bu ortak nesne kümeleri, işlemdeki diğer nesnelerle eşleştirilerek yeni aday kümeleri oluşturulur. Bu eşleştirmenin yapılabilmesi için iki koşul vardır. Birinci koşul, birleştirilecek nesnenin geniş olmasıdır. İkinci koşul ise alfabetik kısıt koşuludur. Birleşecek nesne harf sırası açısından nesne kümesi içindeki tüm nesnelerden sonra geliyor olmalıdır. AIS algoritması budama tekniğini kullanarak birleştirme adımını gerçekleştirir. Budama tekniği, aday kümeleri içindeki gereksiz kümelerin silinerek elenmesi işlemidir. Bu adımdan sonra, her aday kümesinin destek oranı hesaplanır ve destek değerleri, eşik destek değerine ulaşan veya bu değeri aşan kümeleri,

geniş nesne kümesi olarak belirlenir. Bu belirlenen nesne kümeler bir sonraki taramada birleşme işlemine tabi tutularak bir sonraki aday kümelerin belirlenmesi için kullanılırlar. Bu işlem eşik destek değeri sağlanamayana kadar sürdürülür [52]. Her durum için veri setinin üzerinden tekrar tekrar geçmesi, AIS algoritmasının zayıf yönüdür.

4.2.2. SETM algoritması

SETM algoritması, 1995 yılında Houtsma tarafından geliştirilmiştir. Bu algortmada, L_k geniş nesne kümelerinin her biri iki parametreyle tanımlanır. Bunlar, nesnenin ismi ve bu nesnenin ayırt edilmesini sağlayan özellik numarasıdır. Bu özellik numarası algortmada TID (Transaction Identification) şeklinde gösterilir. Algoritma içerisinde her bir aday nesne kümesi (C_k), {TID, Nesne Kümesi İsmi} şeklinde, ikili bir kodlamayla tutulur [54].

SETM algoritması da, tıpkı AIS algoritması gibi veritabanını birçok kez tarar. Birinci taramada veritabanındaki nesneleri teker teker sayar ve hangilerinin geniş nesneler olduğunu belirler. Sonraki taramalarda, bir önceki taramada belirlenen geniş nesne kümeleri kullanılarak yeni aday kümeleri belirlenir. AIS algoritmasından farklı olarak SETM algoritması SETM aday nesne kümelerinin TID'lerini de tutar. Daha sonra, aday nesne kümeleri nesne ismine göre sıralanarak küçük nesne kümeleri silinir. Eğer veritabanı TID numarasına göre sıralanmışsa, bir sonraki taramada herhangi bir işlemdeki geniş nesne kümeleri L_k 'nin TID numarasına göre sıralanmasıyla elde edilir. Bu şekilde veritabanı bir kaç kez taranır ve başka bir geniş nesne kümesi bulunamadığında algoritma sonlandırılır [54].

SETM algoritmasında TID bilgisinin de tutulması ve aday nesne kümesinin desteği hesaplanırken C_k sıralanmamış olmasından dolayı nesne kümelerinin bir kez daha sıraya dizilmesi, algortmada yer ve zaman karmaşıklığına neden olmaktadır.

4.2.3. Apriori algoritması

1994 yılında Agrawal ve arkadaşları tarafından geliştirilen Apriori algoritması, bu alandaki en başarılı, en çok bilinen ve en çok kullanılan algoritma olmuştur. Apriori algoritması ismini çalışma mantığından almıştır. Bu mantık, yaygın nesnelerin bilgilerini bir önceki (prior) adımdan almasıdır [51].

Apriori algoritması, daha önce geliştirilmiş AIS ve SETM algoritmalarından her bir geçişte aday nesne kümelerinin üretilmesi ve sayılması bakımından farklılık gösterir. Bu algoritmalarada, tarama sırasında veriler okunurken aday nesne kümeleri üretilir ve bir işlem (T, transaction) okunduktan sonra, geniş nesne kümelerinin bu işlemlerde olup olmadığına bakılır. Aday nesne kümeleri oluşturulduktan sonra yeni aday nesne kümelerinin üretilmesi, aday nesne kümelerinin, işlemlerdeki diğer nesnelerle elde edilen geniş nesne kümeleriyle birleştirilmesi sonucu elde edilir. Fakat bu işlemler, küçük nesne kümesi olan birçok aday nesne kümesinin üretilmesi ve sayılması sonucunu doğurduğundan, bu algoritmalarda zaman karmaşıklığının artmasına sebep olur [52].

Apriori algoritması ise aday nesnelerin üretimi için sadece bir önceki taramada geniş olarak işaretlenen nesne kümelerini kullanır. Bu işlemin mantığı, geniş bir nesne kümesinden elde edilen bir nesne kümenin de geniş olacağı kabulüne dayanır. Böylece k elemanlı bir kümeden geniş nesne kümelerinin elde edilmesi, k-1 adet nesneye kadar geniş nesne kümelerinin birleştirilmesi ve geniş olmayan alt kümelerin silinmesiyle mümkün olmaktadır. Bu birleşme ve silme (budama) işlemi sonunda daha az sayıda aday nesne kümesi oluşacağından, daha az performans yeterli olacaktır [52].

Apriori algoritmasında, seviye mantığı arama olarak bilinen tekrarlama işlemlerinden oluşan bir yaklaşım kullanır. Bu yaklaşımda, (k+1)-nesne kümelerin tespit edilmesi için k-nesnekümeler, girdi olarak kullanılır. İlk olarak, sık geçen bir elemanlı nesnekümeler tespit edilir. Bu nesne kümeler L_1 olarak sembolize edilir. Yukarıda değinildiği üzere, L_2 'nin (2 elemanlı sık geçen nesnekümelerin kümesi) bulunması için L_1 kullanılır. L_2 de L_3 'ün bulunmasında kullanılır. Daha fazla elemanlı sık geçen k-nesnekümeler bulamayınca kadar bu işleme devam edilir [34].

Seviye mantığına göre sık geçen nesne kümelerin tespitini yapabilmek amacıyla apriori (önceki) özelliği kullanılır. Bu sayede arama alanı küçültülerek işleminin etkinliği artırılmış olur. Apriori algoritmasında birleştirme ve budama işlemleri birbirini izleyen aşamalar şeklinde gerçekleşir. Bu aşamalar aşağıdaki gibi ifade edilebilir.

Birleştirme Adımı: L_{k-1} 'in kendi arasında birleştirilmesi ile aday k-nesne kümelerinin kümesi (C_k) oluşturulur, buradan da L_k bulunur. I_1 ve I_2 , L_{k-1} 'de bulunan nesnekümeler olsun. $I_1[j]$ gösterimi, I_1 'deki j'inci nesneyi temsil eder (örneğin; $I_1[k-2]$, I_1 deki sondan 2.

nesneyi ifade eder). Apriori algoritması, bir işlem veya nesnekümedeki nesnelerin sözcük sırasına göre sıralandığını varsayar. L_{k-1} 'in ilk $(k-2)$ nesneleri ortak ise L_{k-1} 'in elemanları birleştirilebilir. Ancak bu durum altında $L_{k-1} \times L_{k-1}$ birleştirme işlemi gerçekleştirilebilir.

Bu durum; L_{k-1} 'in I_1 ve I_2 nesnekümeleleri eğer;

$$(I_1[1] = I_2[1]) \cap (I_1[2] = I_2[2]) \cap \dots \cap (I_1[k-2] = I_2[k-2]) \cap (I_1[k-1] = I_2[k-1]) \quad (3.16)$$

ise birleştirme işlemi gerçekleştirilebilir demektir. $I_1[k-1] < I_2[k-1]$ koşulu aynı adayların oluşmamasını sağlar. I_1 ve I_2 'nin birleştirilmesiyle oluşan nesnekümesinin sonuçları $I_1[1]I_1[2] \dots I_1[k-1]I_2[k-1]$ dir [34].

Budama Adımı: C_k , aday kümesi L_k 'nın bir üst kümesidir. Aday kümenin (C_k) elemanlarının tamamı için sık geçen ifadesi kullanılamazken, bunun tam tersi; sık geçen nesne kümeler (C_k)'nın elemanıdır denebilir. Yani, sık geçen k -nesnekümelerinin hepsi C_k da yer almak zorundadır. Bir tarama neticesinde L_k 'nın belirlenmesi, C_k 'daki her bir adayın sayısının belirlenmesiyle mümkün olmaktadır. Örneğin, minimum destek sayısına ulaşan veya bu sayıyı aşan bütün adaylar sık geçen olarak ifade edilirler ve böylece L_k 'da yer alırlar. C_k 'nın büyük boyutta olması, taramaların ağır biçimde gerçekleştirilmesine neden olur. Bu sebepten dolayı apriori özelliği kullanılarak C_k 'nın boyutunun azaltılması hedeflemiştir. Apriori özelliği aşağıdaki anlatılan mantık çerçevesinde kullanılır. Sık geçen olmayan $k-1$ elemanlı bir kümeden sık geçen k elemanlı bir alt küme elde edilemez. Bu sebepten, aday k -nesnekümenin $(k-1)$ ögeli altkümeleri, L_{k-1} kümesinin ögesi değilse, aday nesneküme sık geçen olarak tanımlanmaz. Bu aday nesneküme C_k aday kümesinden çıkarılır. Bu çıkarma işlemleri neticesinde azaltılan C_k aday kümeleri için yapılacak tarama işlemlerinin hızında atış sağlanır [34]. Çizelge 4.3'te alışveriş fişleri sunulan ve yukarıda anlatılan algoritmaya ilişkin bir örnek aşağıda açıklanmaktadır [46].

Çizelge 4.3. Müşteriler ve aldıkları ürünler

Müşteriler	Alınan Ürünler
1	I_1, I_2, I_5
2	I_2, I_4
3	I_2, I_3
4	I_1, I_2, I_4
5	I_1, I_3
6	I_2, I_3
7	I_1, I_3
8	I_1, I_2, I_3, I_5
9	I_1, I_2, I_3

Çizelge 4.3'te bir marketten yapılan dokuz alışverişin bilgileri sunulmuştur. Bu tabloda yapılan alışverişlerin numaraları ve her alışverişte satılan ürünler görülmektedir. Bu bilgiler doğrultusunda Apriori algoritmasının aşamaları aşağıda anlatılmıştır.

Algoritmada öncelikle, her nesne C_1 aday kümesinin bir elemanıdır. C_1 aday kümesi, 1-nesnekümeler, yani bir elemanlı kümeler olarak tanımlanır. Algoritma, her ürünün sayısını bulmak için bütün alışveriş fişlerini (D, veritabanını) tarar. Her ürün için tek tek bulunan bu sayılar, destek sayıları olarak belirlenir. Örneğin, çizelge 4.3'ten de görüldüğü gibi tüm veritabanında (D) I_1 ve I_3 'ten 6 adet, I_2 'den 7 adet, I_4 ve I_5 'ten de 2 adet satılmıştır.

Bu işlemlerin devamında, kullanıcı tarafından belirlenen minimum destek sayısı dikkate alınarak işlemlere devam edilir. Örneğin “2” olarak belirlenen destek sayısı için işlemlere şu şekilde devam edilir. L_1 kümesinde bir elemanlı sık geçen nesneler (1-nesnekümeler) gösterilir. Daha önce sayılan bir elemanlı kümelerdeki (C_1) tüm ürünlerin destek sayısı, destek eşik değeri olarak belirlenen 2 sayısına ulaştığı için C_1 'deki bütün nesneler sık geçen nesnekümeler olarak değerlendirilir. Böylece C_1 'deki bütün nesneler L_1 kümesinin de elemanı olduğundan bu kümeye aktarılır.

Sonrasında, $L_1 \times L_1$ kartezyen çarpımı kullanılarak yeni bir aday küme oluşturur. C_2 olarak adlandırılan bu aday küme iki elemanlı nesnekümelerden oluşur. C_2 'nin belirlenmesinin devamında sık geçen 2-nesnekümelerin kümesi olan L_2 'yi belirlenir. L_1 kümesindeki ürünlerin ikili kombinasyonlardan eşik destek değerine ulaşan veya bu değeri aşan kümeler L_2 olarak belirlenir. Yani, veritabanının (D) taranmasıyla C_2 kümesindeki ürünlerin destek sayıları bulunur ve bu değerler destek sayısı sütununda gösterilir. Daha sonra, uygun ürünler (destek sayıları 2 veya daha fazla olanlar) L_2 kümesine aktarılır.

$$C_2 = L_1 \times L_1 = \{I_1, I_2\}, \{I_1, I_3\}, \{I_1, I_4\}, \{I_1, I_5\}, \{I_2, I_3\}, \{I_2, I_4\}, \{I_2, I_5\}, \{I_3, I_4\}, \{I_3, I_5\}, \{I_4, I_5\}$$

$$L_2 = \{I_1, I_2\}, \{I_1, I_3\}, \{I_1, I_5\}, \{I_2, I_3\}, \{I_2, I_4\}, \{I_2, I_5\} \text{ olarak belirlenir.}$$

Sonraki adımda, L_3 (sık geçen 3-nesnekümelerin kümesi) belirlenir. Bunun için algoritma, L_2 kümesinin kendisiyle kartezyen çarpımını ($L_2 \times L_2$) gerçekleştirir. Kartezyen çarpım sonucu oluşan bu küme C_3 (yeni aday küme) olarak adlandırılır. Ancak C_3 aday kümesi oluşturulurken, bir önceki adımda elde edilen L_2 dikkate alınır. L_2 'de mevcut olmayan bir

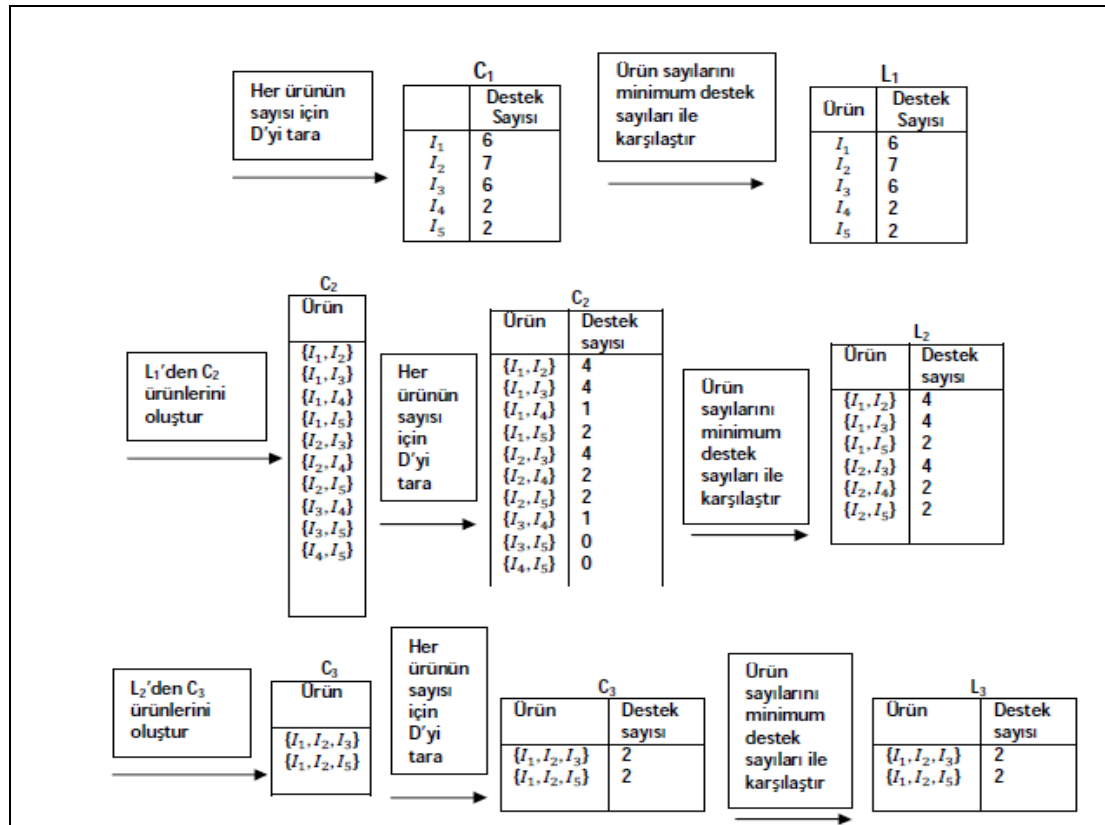
ikili küme, örneğin $\{I_1, I_4\}$ elemanlarının her ikisini içinde barındıran bir küme, C_3 'ün elemanı olamaz. C_3 ortaya çıkarıldıktan sonra, minimum destek sayısı kriteri dikkate alınarak, L_3 tespit edilir. Yukarıda değinildiği üzere, C_3 oluşturulurken L_2 'nin dikkate alınarak bazı aday kümelerin elenmesi ve taramanın sadece kalan kümeler için yapılması, algoritmayı gereksiz işlem yoğunluğundan kurtarır. Böylelikle;

$$C_3 = L_2 \circ L_2 = \{I_1, I_2, I_3\}, \{I_1, I_2, I_5\}, \{I_1, I_3, I_5\}, \{I_2, I_3, I_4\}, \{I_2, I_3, I_5\}, \{I_2, I_4, I_5\}$$

$L_3 = \{I_1, I_2, I_3\}, \{I_1, I_2, I_5\}$, olarak belirlenir.

Yukarıdaki paragrafta anlatılan aynı işlemler bu kez C_4 aday kümesinin belirlenmesi için tekrarlanır. Yani, algoritma tarafından $L_3 \circ L_3$ kartezyen çarpımı yapılır ve $C_4 = \{I_1, I_2, I_3, I_5\}$ olarak belirlenir. Ancak bir önceki adımda L_3 içerisinde $\{I_2, I_3, I_5\}$ elemanlarını içeren bir küme bulunmadığından, bu üç ürünü birlikte içeren bir küme C_4 'ün elemanı olamaz. Bu nedenle bu nesneküme budanır ve C_4 kümesi boş küme olur. Tüm sık geçen nesneler bulunduğu ve artık üretilebilecek yeni aday küme kalmadığından, algoritma sonlanır.

Bahsi geçen bu örneğe ilişkin anlatılan tüm işlemler ve apriori algoritmasının adımları Şekil 4.1'de gösterilmiştir.



Şekil 4.1. Örnek Apriori algoritmasının aşamaları

Son adımda; ikinci, üçüncü ve dördüncü adımlarda bulunan sık geçen nesnekümelere ilişkin birliktelik kuralları oluşturulur. Örneğin dördüncü adımda elde edilmiş, sık geçen bir nesne olan $\{I_1, I_2, I_5\}$ için aşağıdaki birliktelik kuralları çıkarılabilir.

- $I_1 \& I_2 \Rightarrow I_5$, güven = $2/4 = \%50$
(1. ve 2. ürünü kullananlar, %50 olasılıkla 5. ürünü de kullanırlar.)
- $I_1 \& I_5 \Rightarrow I_2$, güven = $2/2 = \%100$
(1. ve 5. ürünü kullananlar, kesinlikle 2. ürünü de kullanırlar.)
- $I_2 \& I_5 \Rightarrow I_1$, güven = $2/2 = \%100$
(2. ve 5. ürünü kullananlar, kesinlikle 1. ürünü de kullanırlar.)
- $I_1 \Rightarrow I_2 \& I_5$, güven = $2/6 = \%33$
(1. ürünü kullananlar, %33 olasılıkla 2. ve 5. ürünü de kullanırlar.)
- $I_2 \Rightarrow I_1 \& I_5$, güven = $2/7 = \%29$
(2. ürünü kullananlar, %29 olasılıkla 1. ve 5. ürünü de kullanırlar.)
- $I_5 \Rightarrow I_1 \& I_2$, güven = $2/2 = \%100$
(5. ürünü kullananlar, kesinlikle 1. ve 2. ürünü de kullanırlar.)

Bu kurallar oluşturulduktan sonra son bir kriterin daha sağlanması gerekmektedir. Bu kriter, olasılık ihtimalini dikkate alan minimum güven değeri kriteridir. Minimum güven değeri de minimum destek değeri gibi kullanıcı tarafından belirlenerek, algoritmaya tanıtılır. Örneğin, üzerinde durulan yukarıdaki durum için kullanıcı tarafından belirlenen minimum güven değerinin %70 olması durumunda, algoritma kullanıcıya birliktelik kuralı olarak sadece ikinci, üçüncü ve altıncı kuralı sunacaktır. Çünkü diğer kurallar eşik güven değerine ulaşamamıştır [34]. Apriori algoritmasının matematiksel modeli Şekil 4.2’de verilmiştir.

```

1)  $L_1 = \{\text{sık geçen 1-öge kümesi}\};$ 
2) for ( $k=2; L_{k-1} \neq \emptyset; k++$ ) do begin
3)    $C_k = \text{apriori-gen}(L_{k-1});$  // Yeni adaylar
4)   forall transactions-hareketler  $t \in D$  do begin
5)      $C_t = \text{subset}(C_k, t);$  // Adaylar t içindedir
6)     forall candidates – adaylar  $c \in C_t$  do
7)        $c.\text{count}++;$ 
8)   end
9)    $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
10) end
11)  $\text{Answer} = \cup_k L_k;$ 

```

Şekil 4.2. Apriori algoritmasını matematiksel modeli [34]

4.2.4. AprioriTid algoritması

Birliktelik kurallarında kullanılan Aprioritid algoritması Agrawal ve Srikant tarafından önerilmiştir [40]. AprioriTid algoritması aday nesne kümelerinin sayısını azaltmak ve dolayısıyla tarama işlemini hafifletmek için apriori-gen ismi verilen bir fonksiyondan yararlanır. Apriori algoritmasından farklı olarak ilk geçişte veritabanının destek seviyesini bulmaya çalışmaz. Bu iş için SETM algoritmasındaki gibi, elemanı $\langle TID, \{X_k\} \rangle$ formunda olan bir C_k kullanılır. Burada, TID numaralı işlemdeki aday geniş k -nesne kümesi X_k 'dir. Böylece $k=1$ durumunda C_1 veritabanına ilişkin işlemler yapılır. Yine, $k>1$ olduğu durumlarda C_k , apriori algoritmasının ileri adımlarında üretildiği gibi üretilir. Ancak farklı olarak t işlemindeki C_k elemanı $\langle TID, c \rangle$ formatında bulunur. Buradaki c , t işleminde C_k 'ya ait bir aday elemandır. Eğer bir işlemin, herhangi bir k nesnekümesi adayı yoksa bu durumda C_k 'nın bu işlem için herhangi bir elemanı olmayacağından bu işlemin TID numarasını taşımayacaktır [51].

4.2.5. FP-Growth algoritması

FP-Growth algoritması, yaygın nesne kümelerini aday nesneler üretmeden bulmak için geliştirilmiş bir yöntem olup, diğer algoritmalara oranla daha iyi performans gösterir. En önemli özelliği olarak, tüm veri tabanını FP-Ağaç denilen sıkıştırılmış bir ağaç veri yapısı şeklinde tutması ve veri tabanını sadece iki kez taraması gösterilebilir. FP-Growth algoritması üç adımdan oluşur. İlk taramayla destek sayılarının hesaplanması, ikinci taramayla Fp-Ağaç yapısının kurulması ve bu ağaç yapısından birliktelik kurallarının çıkarılması. İlk tarama sonunda, destek sayıları minimum destek değerinden küçük olanlar hesaplamaya katılmaz. Öğelerin ağaçtaki başlangıç noktaları "Başlık Tablosu"nda tutulur. Bu tablonun amacı, ağaçta dolaşmayı sağlamak ve ağaçtaki işaretçiler ile birbirine bağlanan düğümlerin takibini kolaylaştırmaktır. Minimum destek değerlerine göre ürünler en büyük destek değerinden en küçük destek değerine sahip olan ürünler şeklinde sıralanır. Böylelikle yaygın olmayan nesnelerin ağaca eklenmesi önlenir. Bunun yanında her bir girdinin öğeleri de sıklıklarına göre sıralanarak ağaca sıkıştırılmış biçimde eklenir. Sıralama işlemi, destek değeri büyük olan nesnelerin köke daha yakın olmasını sağlar. Minimum destek kriterini geçemeyen değerler sıralamaya eklenmez. Oluşturulan ağaç üzerinde Growth algoritması çalıştırılarak ele alınan nesnenin geçtiği dallar belirlenir. Tek bir dalın varlığı halinde, dalı oluşturan nesnelerin kombinasyonu, yaygın nesneler

kümesini verir. Ancak dal sayısı birden fazlaysa, bulunan yollar o öge için koşullu örüntü temelini oluşturur. Her bir koşullu örüntü temelinden koşullu örüntü ağacı elde edilir ve örüntü ağacı üzerinden algoritma yeniden çalıştırılır [34].

4.2.6. Apriori-Hybrid algoritması

Apriori ve Apriori-TID algoritmalarının karışımı olan Apriori-Hybrid algoritması, her döngüde aynı algoritma kullanılmasının gerekmediği mantığına göre çalışır. Daha önce geliştirilen bu iki algoritmanın gözlemlenmesi sonucu ilk döngülerde Apriori algoritmasının kullanılması, daha sonra hafifleyen işlemler için Apriori-TID algoritmasına geçiş yapması tarzının benimsendiği bir yöntemdir. Ancak her döngünün sonunda k 'nın tahmin edilmesi işlemi yapılır. En son döngüdeki geçiş aşaması hariç Apriori algoritmasından performansının yüksek olduğu gözlemlenmiştir [55].

4.2.7. OCD algoritması (Off-line candidate determination / Off-line ürün kümesi tespiti)

Off-Line ürün kümesi tespiti anlamındaki OCD, geniş kümelerin belirlenmesi için, büyük veri tabanından alınan örneklerin iyi sonuçlar verebileceğini savunan bir algoritmadır. Daha önceki döngülerden elde edilen bilgilerin birleşik analiz sonuçlarını kullanarak gereksiz aday kümelerini yok etmeye çalışır. Böylece, veri üzerinden daha az döngü yapılması sağlanacaktır. OCD algoritmasında, aday kümelerin belirlenmesi için önceki döngülerin sonuçlarından elde edilen tüm bilgilerden faydalanılır. Burada $k+1$ büyüklüğünün tüm alt kümeleri incelendiğinden gereksiz aday küme sayısı çok artar [55].

4.2.8. Partition algoritması (Bölümleme algortması)

Bölümleme anlamına gelen ve 1995 yılında Savasere tarafından önerilen algoritma, veri tabanını sadece 2 defa tarar. Veri tabanını küçük bölmelere ayıran Partition algoritmasının işleyebilmesi için veri dağılımının homojen olması gerekmektedir. Çünkü her bir bölmede bir nesne kümesi eşit sayıdaysa, ikinci taramada çoğu nesne kümesinin sayısı büyük olur. Dağınık veri dağılımlarında ise, birçok nesne kümesi, ikinci taramada küçük olarak belirlenir. Böylece, yanlış nesne kümeleri de sayılır, bu da işlem süresini artırır. [35].

4.2.9. DIC algoritması (Dynamic itemset counting/Dinamik ürün kümesi sayımı)

Dinamik ürün kümesi sayımı anlamına gelen DIC, tarama sayısını azaltabilmek için nesne kümelerini önceden oluşturup ve bunları saymaya yönelik bir yöntem izler. Veri tabanı, işlemlerin aralıkları gibi görünür. Aralıklar ardışık olarak taranıp, veri tabanının sonuna ulaşıldığında, veri tabanının en başına dönlür. Daha sonra son bir tarama yapılır ve tam olarak sayılmayan ürün kümeleri sayılır. Tarama sayısı doğal olarak aralığın sayısı ve büyüklüğüne bağlıdır. Küçük bir aralık varsa, tek bir tarama yeterli olabilecek ve ürün kümeleri bu ilk taramada meydana gelecektir. Yapılacak olan ikinci taramada sadece tam olarak sayım işlemi gerçekleştirilecektir [55].

4.2.10. CARMA (Continuous association rule mining/sürekli birliktelik kuralı madenciliği algoritması)

Sürekli birliktelik kuralı madenciliği algoritması CARMA, büyük ürün kümelerinin online olarak hesaplanması için kullanılır. Bu yöntemde mevcut birliktelik kurallarını gösterilmesi online olmanın bir getirisi. CARMA en fazla iki kez tarama yapar. Ürün kümelerinin oluşturulması ilk taramada, bunların sayılması ise ikinci taramada yapılır. Eğer oluşan ürün kümelerinin alt kümeleri de kriterlere uymak kaydıyla potansiyel büyük ürün kümesiyseler, bunlardan da yeni ürün kümeleri oluşturur. Bir ürün kümesinin potansiyel büyük bir ürün kümesi olup olmadığı, belirli bazı hesaplamalarla yapılır [55].

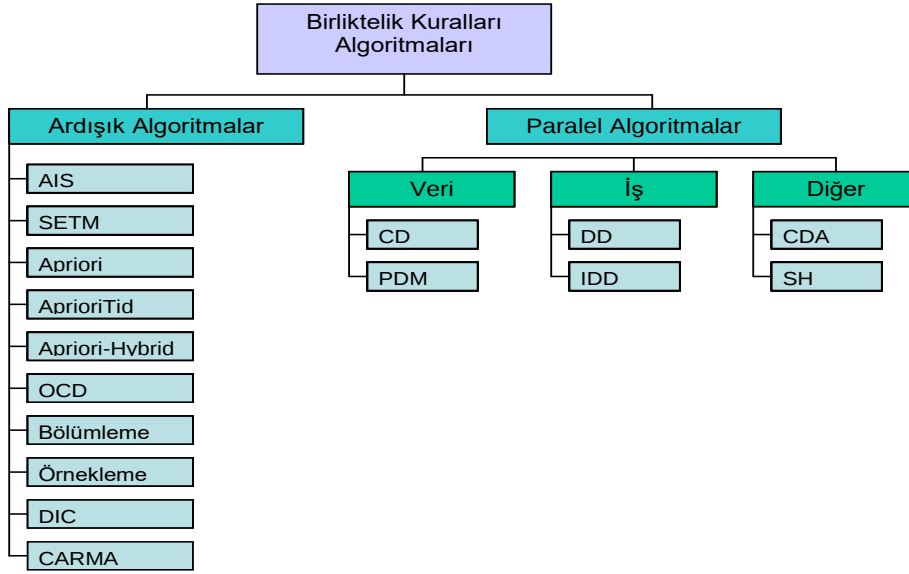
4.2.11. Diğer algoritmalar ve algoritmaların sınıflandırılması

Birliktelik kurallarının analizinde kullanılan diğer algoritmalar da yukarıda anlatılan algoritmalarla benzer mantıkta çalışırlar. Bu algoritmalarından bazıları şunlardır [56]:

- Paralel veri mantığı CD (Count Distribution–Sayım Dağılımı) (Agrawal, Shafer, 1996),
- PDM (Parallel Data Mining - Paralel Veri Madenciliği) (Park, 1995), DMA (Distributed Mining Algorithm - Dağıtılmış Madencilik Algoritması) (Cheung, 1996),
- CCPD (Common Candidate Partitioned Database-Ortak Aday Bölünmüş Veritabanı) (Zaki, 1996),
- Paralel görev mantığı, DD (Data Distribution - Veri Dağılımı) (Agrawal, Shafer, 1995),
- IDD (Intelligent Data Distribution) (Han, 1997),

- HPA (Hash-based Parallel Mining of Association Rules – Bağlantı Kurallarının Çırpı Temelli Paralel Madenciliği) (Shintani ve Kitsuregawa, 1996),
- PAR (Parallel Association Rules - Paralel Bağlantı Kuralları) (Zaki, 1997).

Birliktelik kurallarını tespitinde kullanılan algoritmaların ayırt edici yönlerine göre yapılan sınıflandırma Şekil 4.3'te gösterilmektedir.



Şekil 4.3. Birliktelik kuralı algoritmaları [55]

4.3. Birliktelik Kuralları Algoritmalarının Değerlendirilmesi

Veritabanının taranması esnasında adayların sayısına göre ne kadar büyüklükte bir alana ihtiyaç duyulacağı anlaşılabilir. Veritabanı taramalarının azami sayısı ($G/\mathcal{C}$) ve azami karşılaştırma işlem sayılarına (işlemci tahmini) bakılarak da zaman ihtiyacını belirlenebilir. Artık veritabanlarının ikincil disklerde saklanabildiğinden, performans ölçümü için veritabanının bütününde yapılan tarama sayıları incelenmektedir. Her bir hareketin nesne sayısı m ve k -nesnelerinin yaygın olan nesne kümeleri L_k olarak ifade edilirse, yaygın nesne kümelerinin sayısı 2^m kadar olur. Seviye-mantığını kullanan algoritmalarda, ilk taramayla birlikte L_1 yaygın nesne kümeleri elde edilir. İkinci tarama sonunda L_2 yaygın nesne kümeleri, üçüncü tarama sonunda da L_3 yaygın nesne kümeleri elde edilir ve bu işleme aday nesne küme kalmayana kadar devam edilir. Yaygın nesne kümeleri elde edilemediğinde bir taramaya daha gerek kalmaz ve algoritmalar sonlanır. Yaygın kümelerin bulunduğu son seviye m ise veritabanı en çok $(m+1)$ kez taranır. Apriori-TID

algoritmasında tüm veritabanını sadece birinci geçişte taranır. Daha sonraki herhangi bir taramada tüm veritabanı değil, bir önceki taramada elde edilen C_k taranır. Ancak, işlemler sırasında C_k 'nın tüm hareketlerinin nesneleriyle birlikte ele alınması bu yöntemin zayıf yanıdır. Bir diğer birliktelik kuralı algoritması olan OCD, tek elemanlı yaygın nesne kümelerini (L_1) ortaya çıkarmak için bir defalığına veritabanını tarar. Daha sonra elde edilen bu bilgiler Sampling tekniğinde de olduğu gibi veritabanının bir bölümünü ve L_k 'nin aday nesne kümelerini bulmak için kullanılır. İkinci taramada ise aday nesneküme oluşturulmaya çalışılmadan, sadece ilk taramada elde edilen aday nesne kümelerinin destek değeri hesaplanır. Partition ve CARMA teknikleri de, veritabanının en fazla iki defa taranmasına ihtiyaç duyarlar. Böylece, G/Ç yükünü azaltmış olurlar.

Bir algoritmanın asıl performansı, önerdiği doğru adayların sayısıyla ölçülmelidir. Bilindiği üzere, algoritmaların tamamı, aday kümelerini oluşturabilmek için daha önceki taramalardan elde edilen yaygın nesne kümelerini kullanırlar. Bu kapsamda, destek değerlerinin sayılabilmesi için bir önceki nesne kümelerinden elde edilen yaygın nesne kümeleri ana belleğe alınırlar. Ana belleğin yetersiz olması ihtimaline karşı, depolama mimarilerine olanak sağlayan algoritmalar ve farklı seçeneklerdeki tampon (buffer) yönetimi de kullanılmaktadır. AIS, sadece ihtiyaç halinde L_{k-1} 'in diskte kalmasını önermişken, SETM, C_k 'nin ana belleğe alınamayacak kadar büyük olması durumunda, FIFO yöntemine göre belleğe yazılmasını önermiştir. Apriori ve bu tarzdaki algoritmalar ise, L_{k-1} 'in diskte saklanmasını önerirler. Ancak devamında C_k 'yı bulmak için bir bloğun ana belleğe alınması gereğini vurgu yaparlar. Bu genellemeyle birlikte, destek değerlerinin hesaplanabilmesi için Apriori-TID ve Apriori-Hybrid algoritmaları C_k 'yı ana bellekte tutarlar. Bunlar dışındaki diğer algoritmalar bellek yetersizliği sorunuyla ilgilenmezler. Yani bellek kapasitesi sorunu olmayan kullanıcılar tarafından tercih edilmelidirler. Bunlar dışındaki diğer tüm sıralı teknikler, ana belleğe sığacak kadar veritabanının elverişli bir kısmının keşfini amaçlarlar.

AIS ve SETM algoritmaları, yaygın nesne kümeleri için herhangi bir depolama mimarisi önermezken, Apriori ve bu tarzdaki algoritmalar, hash ağacı veya dizisi gibi veri yapıları önermişlerdir.

Birliktelik kurallarının keşfiyle ilgilenen ticari araçların birçoğu, ortalama genel durumlar için iyi sonuçlar veren Apriori algoritmasını kullanmaktadır. Buna rağmen bazı özel

durumlar için Apriori algoritmasından başka bir algoritmanın kullanılmasıyla daha yüksek performans sağlanabilmektedir. Örneğin AIS algoritması, çok nesneli veritabanları için uygun olmadığından, yoğun işlem gerektirmeyen veritabanları için kullanılır. Diğer taraftan, ilk geçişte Apriori yöntemi zaman bakımından daha avantajlıyken, sonraki geçişlerde Apriori-TID yönteminin Apriori'ye nazaran, taramaları daha kısa sürede gerçekleştirdiği belirlenmiştir. Dolayısıyla, zaman bakımından bir kısıt bulunması halinde, bu iki tekniğin avantajlı yönleri alarak oluşturulan Apriori-Hybrid algoritması en iyi performansı vermektedir. Ancak, Apriori'den Apriori-TID'e geçiş kapsamlı bir iş olduğundan ek yük getirir. Destek değeri düşük belirlenen durumlar için, tahmin bazlı bir yöntem olmasına rağmen, OCD algoritmasının kullanılması, yaygın nesne kümelerinin bulunmasında etkin sonuçlar verecektir. CARMA, online, kullanıcı etkileşimli ve geri dönüşüm tabanlı bir yöntemdir. Tercih edilme, durumu hareket dizilerinin bir ağdan okunması halidir. Bu halde etkin sonuçlar verir.

Çizelge 4.4'te, tarama sayısı, veri yapıları ve bazı özel durumlar bakımından birliktelik kuralları algoritmalarının kıyaslamaları yapılarak sunulmuştur.

Çizelge 4.4. Algoritmaların kıyaslanması [57]

Algoritma	Tarama Sayısı	Veri Yapısı	Açıklamalar
AIS	m+1	Belirtilmemiş	Belli başlı, az veri içeren seyrek hareketset veritabanları için uygundur.
SETM	m+1	Belirtilmemiş	SQL uyumlu.
Apriori	m+1	L_{k-1} : Hash tablosu C_k : Hash ağacı	Belli başlı, orta düzey veri içeren hareketset veritabanları için uygundur. AIS ve SETM'e oranla daha iyidir. Paralel algoritmaların baz aldığı algoritmadır.
Apriori-TID	m+1	L_{k-1} : Hash tablosu C_k : TID ile dizinlenmiş dizi $\overline{C_k}$: Sıralı mimari ID: bitmap	$\overline{C_k}$ 'nın büyük olması durumunda çok yavaştır. Küçük boyutlu $\overline{C_k}$ ile Apriori'den daha iyidir.
Apriori-Hybrid	m+1	L_{k-1} : Hash tablosu <u>İlk safhada:</u> C_k : Hash ağacı <u>İkinci safhada:</u> C_k : TID ile dizinlenmiş dizi $\overline{C_k}$: Sıralı mimari ID: bitmap	Apriori'den daha iyidir. Buna rağmen, Apriori'den Apriori-TID'e geçiş ek yük gerektirir. Geçiş noktasını saptamak çok önemlidir.
OCD	2	Belirtilmemiş	Düşük destek eşik değerleri ve çok büyük veritabanları için uygundur.
PARTITION	2	Hash tablosu	Büyük veritabanları için uygundur. Homojen veri dağılımını destekler.
Sampling	2	Belirtilmemiş	Düşük destek eşik değerleri ve çok büyük veritabanları için uygundur.
DIC	Aralık boyutuna bağlıdır	Ağaç	Veritabanı, hareketset aralıklar şeklindedir. Yüksek boyutlu adaylar, bir aralığın sonunda oluşturulur.
CARMA	2	Hash tablosu	Hareket dizilerinin bir ağdan okunmasına uygundur. Online. Kullanıcılardan sürekli geri-besleme gelir ve destek ve/veya güven değerlerini işlem sırasında herhangi bir anda değiştirebilirler.
FP-Growth	2	FP-Tree	Aday üretimsiz bir tekniktir. Apriori ve türevlerinden daha iyi performansa sahiptir.
CD	m+1	Hash tablosu ve ağacı	Veri paralelleştirme.
PDM	m+1	Hash tablosu ve ağacı	Veri paralelleştirme, erken aday budaması.
DMA	m+1	Hash tablosu ve ağacı	Veri paralelleştirme, aday budaması.
CCPD	m+1	Hash tablosu ve ağacı	Veri paralelleştirme, paylaşımlı-bellek içeren makinelerde kullanılır.
DD	m+1	Hash tablosu ve ağacı	Görev paralelleştirme, round-robin bölümlendirmesi.
IDD	m+1	Hash tablosu ve ağacı	Görev paralelleştirme, ilk nesnelerle bölümlendirme.
HPA	m+1	Hash tablosu ve ağacı	Görev paralelleştirme, hash fonksiyonu ile bölümlendirme.
SH	m+1	Hash tablosu ve ağacı	Veri paralelleştirme, adaylar her bir işlemcide bağımsız olarak oluşturulur.
HD	m+1	Hash tablosu ve ağacı	Melez veri ve görev paralelleştirme, ızgara (grid) paralel mimarisi

5. BİR FABRİKADA KULLANILAN SARF MALZEMELERE İLİŞKİN BİRLİKTELİK DAVRANIŞLARININ TESPİT EDİLMESİ

5.1. Genel Bilgiler

Birliktelik davranışlarının tespit edilmesine ilişkin uygulamanın, kamu faaliyeti icra eden bir fabrikada gerçekleştirilmesi planlanmıştır. Çeşitli faaliyetlerin yürütüldüğü fabrikada ağırlıklı olarak bakım ve onarım faaliyetleri yapılmaktadır. Kuruluş bu faaliyetleri beş farklı gruba ayırmış olup, kendi içerisinde yaptığı tüm analizleri bu gruplandırmaya göre yapmaktadır. Bu gruplar;

- Elektrikli cihazların arıza onarımı,
- Elektronik cihazların arıza onarımı,
- Elektronik cihazların elektrik arızası onarımı,
- Kalibrasyon,
- İmalat faaliyetlerinden oluşmaktadır.

Bakım ve onarım gibi plansız işlere maruz kalan bir fabrikada tahmin edilmesi en zor fakat gerekli işlemlerden biri, hangi cihazların arıza yapacağı ve bu arızanın onarımı için hangi malzemelere ihtiyaç duyulacağıdır. Özellikle gecikmeye tahammülü olmayan işler için fabrikada onarım için gereken malzemelerin sürekli mevcut bulundurulması gerekmektedir. Kamuya ait işletmelerde malzeme tedariki vb. harcamalar için ilgili yılın başında bütçe tahsis edilir ve acil bir durum olmadığı müddetçe ek ödenek verilmez. Ayrıca talep edilecek bütçe için önceden bir harcama planı gönderilir. Bu sebeplerden dolayı malzeme tedarikinin, söz konusu arıza gerçekleşmeden yapılması zorunda kalmaktadır. Dolayısıyla arıza onarımları için hangi malzemelerin tedarik edilmesi gerektiğinin analizi büyük önem taşımaktadır.

Arıza onarımı esnasında iki tip malzemeye ihtiyaç duyulmaktadır. Bunlardan birincisi, nadiren meydana gelse de, arıza yapan cihazın ana parçalarından birinin tamamen bozularak onarılamayacak hale gelmesi sonucu, o parçanın yenisinin tedarik edilmesi gerekmektedir. İkincisi ise, arızanın onarımı için gerek duyulan sarf malzeme ihtiyacıdır. Uygulama çalışmasında, çok sıklıkla ihtiyaç duyulan ikinci tip malzeme yani sarf

malzemeler incelenmiştir. Ayrıca bu tip malzemelerin büyük çoğunluğu yukarıda bahsedilen 5 ana grup faaliyet için de kullanılan ortak malzemelerdir.

Çalışmanın sonucunda elde edilmesi planlanan birliktelik davranışlarının tespit edilmesiyle farklı alanlarda iyileştirme yapılması amaçlanmıştır. Birinci amaç olarak, iş gücünden tasarruf edilebileceği düşünülmektedir. Arıza onarımı sırasında ihtiyaç duyulan bir sarf malzemenin tespit edilmesiyle, diğer ihtiyaç duyulacak malzemelerin de ihtiyaç oluşmadan önce depolardan iş istasyonuna getirilmesi sağlanacaktır. Böylece depolar ve iş istasyonu arasında geçen toplam süre azaltılarak hem iş gücünden tasarruf hem de arızanın onarımı için planlanan standart birim zamanda azalış sağlanacaktır. Bu da tam kapasite çalışan işletmeler için ek kapasite anlamına gelmektedir.

İkinci amaç olarak, malzeme tedarik maliyetinden tasarruf edilebileceği düşünülmektedir. Uygulama çalışmasının yapıldığı kamu kurumunda malzeme tedariki üç farklı yöntemle yapılabilmektedir. Bunlar; doğrudan temin, açık ve belirli istekliler arasında yapılan ihale yöntemleridir. Geçmiş veriler incelendiğinde her üç yöntem için de, talep edilen malzemenin toplam değeri arttıkça, birim tedarik maliyetinin azaldığı tespit edilmiştir. Dolayısıyla birlikte kullanılan malzemelerin tespit edilmesi, hem alım sürecini hem de tedarikçi firmanın toplam fiyatta kırıma gitmesine neden olacaktır. Böylece malzeme maliyetinde azalma sağlanacaktır.

Üçüncü amaç olarak, iş istasyonlarının bulunduğu atölyelerden ortalama 1 km. uzaklıkta bulunan 12 farklı deponun malzeme içeriğinin, elde edilecek kurallara göre yeniden düzenlenmesiyle depo ve taşıma personelinin iş yükü hafifleyecektir.

Hedeflenen son amaç söz konusu işletmeye özeldir. İşletmede arıza onarım faaliyetlerini yürüten dört farklı müdürlük vardır. Bu müdürlüklerin her biri malzeme ihtiyaç planlarını kendi bünyelerinde sezgisel olarak yapmakta ve tedarik birimine iletmektedirler. Dolayısıyla dönem içerisinde malzeme kıtlığı veya yokluğuyla sıkça karşılaşmaktadır. Birliktelik davranışlarının oluşturulması ve malzeme planlarının tek elden yapılmasıyla birlikte tedarik sürecinde harcanan iş gücünden tasarruf sağlanacağı ve malzeme yokluğundan kaynaklanan bekleme sorunlarında azalma görüleceği açıktır. Sonuç olarak, birliktelik kurallarının tespit edilmesiyle;

- Etkin ve hızlı bir planlama süreci,
- Standart birim zamanda azalma,
- Uygun depo düzeni,
- Malzeme tedarik maliyetinde azalma ve
- İş yükü ve işgücü konularında iyileştirmeler olacağı değerlendirilmektedir.

Veri madenciliği alanında birliktelik kuralları ile ilgili çeşitli uygulamalar bulunmaktadır. Bu uygulamada, fabrikada yapılan bir onarım sırasında hangi sarf malzemelerin birlikte kullanıldıkları tespit edilmeye çalışılmıştır.

Öncelikle SPSS programı yardımıyla veri setine ilişkin anlamlı istatistiki bilgiler incelenecektir. Daha sonra ise için SPSS Modeler programındaki Apriori algoritması kullanılarak verilerin analizi gerçekleştirilecektir. Birliktelik davranışlarının analizi yapılırken öncelikle fabrikada açılan tüm iş emirleri için ilgili algoritma çalıştırılıp sonuçlar incelenecek, sonrasında ise beş faaliyet grubu için de ayrı ayrı bu işlem tekrarlanacaktır. Son olarak, tüm veri setinden elde edilen kurallar ile faaliyet gruplarının veri setinden elde edilen kuralların karşılaştırılması planlanmıştır. Aradaki bağlantının güçlü veya zayıf olma durumuna göre birliktelik ilişkilerinin gücü tartışılacaktır.

5.2. Verilerin Tanımı

Söz konusu uygulama için kullanılan veriler bir kamu işletmesinden elde edildiğinden veri setinde gizlilik ilkelerine bağlı kalmıştır. Bu kapsamda işletmenin adı, veri setinde bulunan iş emri no sütunu ve üretim no sütunu değiştirilmiştir. Ancak birliktelik ilişkilerinin herhangi bir sapmaya maruz kalmaması için malzeme bilgilerinde oynama yapılmamıştır. Malzeme kullanım verisi, fabrikada 2013 ile 2017 yılları arasında onarım faaliyetlerinde kullanılan tüm sarf malzemelerin listesidir. Bu listede 16 036 iş emrinde 180 çeşit malzeme kullanılmıştır. Bu veri setinin bir kısmının Excel paket programındaki görüntüsü Resim 5.1’de gösterilmektedir. Veri seti şablonunun açıklaması ile kullanılan sarf malzeme listesi Ek-1’de verilmiştir.

Microsoft Excel - MALZEME-HasanRED.xls

Dosya

Görünüm

Ekle

Biçim

Araçlar

Yeni

Pencere

Yardım

AY14

0

Anlat Tur

12

K

T

A

Resim 5.1. Veri setinin bir kısmının excel programındaki görüntüsü

Veri setinde 180 çeşit sarf malzeme bulunmaktadır. Birlikte analiz için verinin kategori bazında kodlanmış olması gerekmektedir. Yani tüm veri seti 0-1 sayıları ile kodlanmalıdır. Bu veri setinde bir iş emrinde kullanılan sarf malzeme “1” ile, kullanılmayan sarf malzeme “0” ile kodlanmıştır. Bu ikili kodlama, bir arıza onarımında ilgili malzemenin kullanılıp kullanılmadığını göstermektedir. Ayrıca Resim 5.1’de görülen “İş emri no” sütunu faaliyet grubu ve işin geliş sırasını ifade etmektedir. “İş emri no” sütunu, market sepet analizindeki müşterilere benzetilebilir.

Veri kümesinin büyüklüğüne göre değişen ama sıklıkla en fazla zaman isteyen aşama, verilerin ön işlemden geçirilmesi aşamasıdır. Bu veri seti, analize uygun olacak şekilde ön işlemden geçirilmiş, yanlış girişler temizlenerek toplam 16036 giriş (iş emri satırı) elde edilmiş ve kategori bazında kodlanmıştır. Fabrikada yürütülen faaliyetler daha önce de değinildiği üzere 5 ana grup altında sınıflandırılmaktadır. Bunlar; elektrikli cihazın arıza onarımı, elektronik cihazın arıza onarımı, elektronik cihazın elektrik arızası onarımı, kalibrasyon ve imalat faaliyetlerinden oluşmaktadır.

Söz konusu arızalar için SPSS programı kullanılarak hazırlanan tanımlayıcı istatistik bilgileri Çizelge 5.1’de sunulmuştur.

Çizelge 5.1. Veri setinin tanımlayıcı istatistikleri

İş emri sayısı	16 036
Ortalama mlz.çeşidi	2,23
Medyan	1,00
Tepe değer	1
Std. Sapma	2,265
Varyans	5,131
En az hareket sayısı	1
En fazla hareket sayısı	23
Toplam hareket sayısı	35 694

Bu çalışma kapsamında 16 036 adet arıza onarımı bilgileri dikkate alınmıştır. Analizde kullanılacak bu veri seti ışığında, bir arıza için kullanılan ortalama malzeme çeşidi 2,23, medyan yani orta değeri 1, mod (tepe) yani en çok kullanılan değer 1, standart sapması 2,265, varyansı 5,131 olarak hesaplanmıştır. Veri setinde bulunan tüm arızalar için toplam 35 694 adet malzeme girişi yapılmıştır. Bu da Excel veri setinde 35 694 adet “1” sayısı kodlandığı anlamına gelmektedir. Bir arıza onarımında kullanılan en az malzeme çeşidi 1, en çok 23 olarak hesaplanmıştır. Yani, bir arıza için kimi zaman tek bir sarf malzeme kullanılırken kimi zaman da 23 çeşit malzeme kullanılmıştır. Arızalar için kullanılan sarf malzeme çeşitlerinin sıklıkları Çizelge 5.2’de sunulmuştur.

Çizelge 5.2. Malzeme çeşidi sıklıkları

Malzeme Çeşidi	Sıklık	Yüzde	Kümülatif Yüzde
1	8 495	53,0	53,0
2	3 474	21,7	74,6
3	1 639	10,2	84,9
4	1 004	6,3	91,1
5	463	2,9	94,0
6	257	1,6	95,6
7	163	1,0	96,6
8	97	0,6	97,2
9	84	0,5	97,8
10	65	0,4	98,2
11	42	0,3	98,4
12	71	0,4	98,9
13	44	0,3	99,1
14	39	0,2	99,4
15	26	0,2	99,5
16	18	0,1	99,7
17	10	0,1	99,7
18	21	0,1	99,9
19	14	0,1	99,9
20	4	0,0	100,0
21	1	0,0	100,0
22	2	0,0	100,0
23	3	0,0	100,0
Toplam	16 036	100,0	

Çizelge 5.2’de görüldüğü üzere tek malzeme çeşidi kullanılan iş emri sayısı 8 495 olup, bu değer tüm veri setinin %53’ünü oluşturmaktadır. İki çeşit malzeme kullanılan iş emri sayısı ise 3 474 olup, bu değer tüm veri setinin %21,7’sini oluşturmaktadır. Bununla birlikte 21 çeşit malzeme kullanılan sadece tek bir iş emri mevcuttur.

Arızalarda kullanılan malzeme çeşidi sayılarının faaliyet grubuna göre nasıl değiştiğini gösteren çapraz tablo Çizelge 5.3’te sunulmuştur. Örneğin, 4 çeşit malzeme kullanılan 1004 adet iş emrinin 398 adedi Elk. arızası grubuna, 194 adedi Eln. arızası grubuna, 331

adedi Eln. ünitelerin elk. arızası grubuna, 10 adedi Kalibre grubuna ve 71 adedi de imalat grubuna dahildir.

Çizelge 5.3. Faaliyet grupları bazında malzeme çeşidi sıklıkları

Kaç çeşit malzeme	Arıza Grubu					Toplam
	Elektrik Arızası	Elektronik Arızası	Elektronik ünitenin elektrik arızası	Kalibre	İmalat	
1	1 905	1 873	2 924	1 288	505	8 495
2	1 014	705	1 268	181	306	3 474
3	552	360	550	43	134	1 639
4	398	194	331	10	71	1 004
5	221	40	131	11	60	463
6	89	18	55	2	93	257
7	54	8	71	0	30	163
8	41	10	34	0	12	97
9	33	0	44	0	7	84
10	39	0	21	0	5	65
11	30	0	10	0	2	42
12	50	0	19	0	2	71
13	33	0	5	0	6	44
14	17	0	10	0	12	39
15	15	0	0	0	11	26
16	7	0	1	0	10	18
17	4	0	5	0	1	10
18	1	0	14	0	6	21
19	1	0	1	0	12	14
20	0	0	1	0	3	4
21	0	0	0	0	1	1
22	0	0	0	0	2	2
23	0	0	0	0	3	3
Toplam	4 504	3 208	5 495	1 535	1 294	16 036

5.2.1. Her faaliyet grubu için ayrı ayrı normalite testleri

Her faaliyet grubunda kullanılan malzeme miktarlarının, birbirinden anlamlı derecede farklı olup olmadığının tespit edilmesi, bir takım testlerin yapılmasıyla mümkün olmaktadır. Ancak sözü edilen testlerden hangisinin seçilmesi gerektiği önemli bir sorundur. Çünkü veri yapısına uygun olmayan bir testin seçimi, ciddi hatalara sebep olmakta, hatta elde edilmesi gereken sonucun tam tersi bir sonucun elde edilmesiyle karşı karşıya kalınabilmektedir.

Bu noktada çalışmada kullanılan veri kümesinde bir değişikliğe gidilmiştir. Birliktelik kuralları çıkarımında, sadece tek bir malzeme kullanılan kayıtlar fayda sağlamayacağından, bu aşamadan sonra yapılacak tüm test ve analizlerde indirgenmiş veri seti (7541 iş emri) kullanılacaktır. Dolayısıyla gruplar arası ilişki durumu ile kural çıkarımı ve bu ikisi arasındaki karşılaştırmaların tümü indirgenmiş veri seti üzerinden yapılacaktır.

Öncelikle veri setimizde yer alan grupların normal dağılıma sahip olup olmadıkları araştırılacaktır. Nitekim normal dağılıma sahip bir veri seti için yapılacak testler daha güçlü sonuçlar vermektedir. Bu bağlamda SPSS aracılığıyla sağlanan tanımlayıcı bilgiler Çizelge 5.4'te, yapılan Kolmogorov-Smirnov normalite testinin sonuçları ise Çizelge 5.5'te sunulmuştur.

Çizelge 5.4. Faaliyet gruplarının tanımlayıcı istatistikleri

	Arıza Grubu		Değerler	Std. Hata
MLZ_ÇŞT	Elektrik Arızası	Ortalama	3,98	0,056
		Medyan	3,00	
		Minimum	2	
		Maksimum	19	
		Çarpıklık	2,166	0,048
		Basıklık	4,566	0,096
	Elektronik Arızası	Ortalama	2,78	0,029
		Medyan	2,00	
		Minimum	2	
		Maksimum	8	
		Çarpıklık	1,881	0,067
		Basıklık	4,724	0,134
	Elektronik ünitenin elektrik arızası	Ortalama	3,42	0,048
		Medyan	3,00	
		Minimum	2	
		Maksimum	20	
		Çarpıklık	3,070	0,048
		Basıklık	11,825	0,097
	Kalibre	Ortalama	2,42	0,053
		Medyan	2,00	
		Minimum	2	
		Maksimum	6	
		Çarpıklık	2,240	0,155
		Basıklık	4,684	0,309
	İmalat	Ortalama	4,74	0,148
		Medyan	3,00	
		Minimum	2	
		Maksimum	23	
		Çarpıklık	2,278	0,087
		Basıklık	4,963	0,174

Çizelge 5.4’te her faaliyet grubu için ayrı ayrı tanımlayıcı bilgilere yer verilmiştir. Esasen “Mean” (ortalama) değerine bakıldığında, her bir faaliyet grubunun kullandığı malzeme

çeşidi miktarlarının farklı olduğu gözlemlenebilmektedir. Ancak bu farklılığın anlamlı olup olmadığının, bilimsel olarak gösterilebilmesi gerekmektedir. Daha önce de değinildiği üzere uygun testin seçilmesi için veri setinin normal dağılıma sahip olup olmadığının tespit edilmesi gerekmektedir. Çizelge 5.4’te sunulan “Çarpıklık” değerinin (-1,1) aralığında olması normal dağılıma bir işaret olmasına rağmen, bilimsel yeterlilik için mutlak olarak normal dağılım testlerinin yapılması gerekmektedir. Bu kapsamda uygulanan Kolmogorov-Smirnov normalite testinin sonuçları ise Çizelge 5.5’te sunulmuştur.

Çizelge 5.5. Normalite test sonuçları

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Değerler	df	Sig.	Değerler	Df	Sig.
MLZ_ÇŞT	Elektrik Arızası	0,253	2 599	0,000	0,698	2 599	0,000
	Elektronik Arızası	0,294	1 335	0,000	0,723	1 335	0,000
	Elektronik ünitenin elektrik arızası	0,281	2 571	0,000	0,614	2 571	0,000
	Kalibre	0,426	247	0,000	0,567	247	0,000
	İmalat	0,256	789	0,000	0,673	789	0,000

Çizelge 5.5’teki veriler için iki adet hipotez oluşturulur.

H₀: Verilerde normal dağılımdan sapma yoktur,

H₁: Verilerde normal dağılımdan sapma vardır.

Eğer “Sig.” değeri seçilen güvenilirlik değerine göre, 0,05’ten büyük çıkarsa H₀ yokluk hipotezi, diğer durumda H₁ hipotezi seçilir. Çizelge 5.5’te görüldüğü üzere “Sig.” değerleri sıfıra çok yakın değerler olduğu için, hiçbir faaliyet grubunun verilerinin normal dağılıma uymadığı anlaşılmaktadır.

Normal dağılıma sahip veriler için kullanılabilen parametrik testler daha güçlü sonuçlar sağladığından, veri seti normal dağılıma uygun hale getirilmeye çalışılmalıdır. Bu aşamada Çizelge 5.4’teki “Çarpıklık” değerlerine bakılarak, veri setine uygulanması gereken işleme karar verilmelidir. Çizelge 5.4’te görüldüğü üzere “Çarpıklık” değerleri 1’den büyük pozitif değerlere sahip olduğundan, veriler için logaritmik dönüşüm işlemi seçilip, normalite testinin yeniden yapılması gerekmektedir. Bu kapsamda yapılan ikinci normalite testinin tanımlayıcı istatistikleri Çizelge 5.6’da, sonuçları ise Çizelge 5.7’de sunulmuştur.

Çizelge 5.6. Logaritmik dönüşümden sonra faaliyet gruplarının tanımlayıcı istatistikleri

	Arıza Grubu		İstatistikler	Std. Hata
Log_veri	Elektrik Arızası	Ortalama	0,5233	,00467
		Medyan	0,4771	
		Varyans	0,057	
		Std. Sapma	0,23802	
		Çarpıklık	0,996	0,048
		Basıklık	0,243	0,096
	Elektronik Arızası	Ortalama	0,4184	0,00388
		Medyan	0,3010	
		Varyans	0,020	
		Std. Sapma	0,14169	
		Çarpıklık	0,955	0,067
		Basıklık	0,217	0,134
	Eln.ünitenin elk.arızası	Ortalama	0,4694	0,00421
		Medyan	0,4771	
		Varyans	0,046	
		Std. Sapma	0,21341	
		Çarpıklık	1,351	0,048
		Basıklık	1,428	0,097
	Kalibre	Ortalama	0,3655	0,00749
		Medyan	0,3010	
		Varyans	0,014	
		Std. Sapma	0,11776	
		Çarpıklık	1,738	0,155
		Basıklık	2,058	0,309
	İmalat	Ortalama	0,5662	0,01009
		Medyan	0,4771	
		Varyans	0,080	
		Std. Sapma	0,28331	
		Çarpıklık	0,936	0,087
		Basıklık	0,062	0,174

Çizelge 5.6’da her ne kadar “çarpıklık” değerinde bir normalleşme görülse de Çizelge 5.7’de sunulan ikinci normalite testlerinin sonucuna göre, verilerin normal dağılıma sahip olmadığı anlaşılmaktadır.

Çizelge 5.7. Logaritmik dönüşümden sonra normalite test sonuçları

	Arıza Grubu	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Değerler	Df	Sig.	Değerler	Df	Sig.
Log_veri	Elektrik Arızası	0,215	2 599	0,000	0,843	2 599	0,000
	Elektronik Arızası	0,324	1 335	0,000	0,776	1 335	0,000
	Elektronik ünitenin elektrik arızası	0,278	2 571	0,000	0,782	2 571	0,000
	Kalibre	0,441	247	0,000	0,598	247	0,000
	İmalat	0,213	789	0,000	0,847	789	0,000

Bu aşamadan sonra, faaliyet gruplarının arasında anlamlı bir ilişkinin varlığını parametrik olmayan testlerle sınanması gerekecektir. Veriler 5 grupta incelendiğinden dolayı, söz konusu test gruplarından “K independent samples” testlerinden “Kruskal Wallis” testi seçilecektir.

5.3. Veri Setinin Alt Gruplara Ayrılmasındaki Amaç

Market sepet analizi için birliktelik kuralları çıkarılmaya çalışıldığında, tüm müşterilerin homojen bir yapıda olduğu varsayılır ve bu doğrultuda sonuçlar elde edilir. Belli bir marketten alışveriş yapan müşterilerin markete yakın yerlerde ikamet ettikleri ve bu sebepten dolayı benzer eğilimlere sahip olması, bu varsayımın kabul edilebilir tarafıdır. Ancak tüm müşterilerin homojen olmadığı gerçeği ile birlikte, yüksek miktarda veri girişi sağlanamamasıyla, yüksek güven düzeyinde birliktelik kuralı elde etmek zorlaşacaktır. Örnek olarak Almanya’da elde edilen, bira ile çocuk bezinin bir arada satılıyor olduğu kuralı, Türkiye’den giden bir müşteri için büyük ihtimalle bir anlam ifade etmeyecektir. Eğer yüksek güven düzeyinde bir birliktelik kuralı elde edilebiliyorsa müşterilerin homojen olmaması kuralın etkisini zayıflatmaz, sadece kural elde edebilmeyi zorlaştırır. Ancak, birliktelik davranışlarının tespiti yapılırken, veri setinin daha özel gruplardan oluşturulması, analizin sonuçlarını gerçek eğilimlere daha çok yaklaştıracaktır. Örneğin tüm müşterilere ait bir veri seti kullanmaktansa, bu veri setini örneğin cinsiyete göre

ayırdıktan sonra analizin yapılmasının akabinde erkek ve kadın müşterilerin alışveriş eğilimlerinin tespit edilmesiyle daha gerçekçi birliktelik kurallarına ulaşılabilir. Hatta analizi daha da özele indirgeyerek her müşteri için ayrı sonuçlar elde edilebilir ve her müşteriye ayrı tekliflerde bulunulabilir.

Bu kapsamda düşünüldüğünde, yapılacak malzeme analizi öncelikle tüm veri seti üzerinden sonrasında da faaliyet gruplarının özel veri setleri üzerinden gerçekleştirilecektir. Sonuç olarak ise her iki durumda elde edilen birliktelik kuralları karşılaştırılarak, genel ve özel veri seti sonuçları arasındaki farklar yorumlanacaktır. Faaliyet gruplarına ilişkin analiz yapılmadan önce, malzeme kullanma çeşitleri açısından bu grupların arasında anlamlı bir farkın olup olmadığı araştırılacaktır. Daha önce değinildiği üzere normal dağılıma uygun olmayan ve normal dağılıma dönüştürülemeyen veri setine sahip olduğumuzdan, bu çalışmada parametrik olmayan Kruskal Wallis H ve devamında Mann Whitney U testi kullanılacaktır. Böylelikle elde edilmesi düşünülen gruplar arası anlamlı bir farkın varlığı ile, tespit edilen birliktelik kuralı sayısı arasında bir ilişkinin var olup olmadığı test edilecektir. Yani kullandığı malzeme çeşidi açısından diğer gruplardan anlamlı bir şekilde farklı olan gruptan elde edilen birliktelik kuralı sayısının diğer gruplardan elde edilen kural sayısı ile arasındaki ilişki sorgulanacaktır.

5.3.1. Faaliyet grupları ile kullanılan sarf malzeme arasındaki ilişkinin tespiti

Bu kısımda faaliyet gruplarının kullandıkları sarf malzeme çeşit sayılarının birbirlerinden anlamlı derecede farklı olup olmadıkları sorgulanacaktır. Her bir faaliyet grubunun kullandığı ortalama sarf malzeme sayılarının birbirinden anlamlı derecede farklı olup olmadığının tespitini yapmak için SPSS yardımıyla parametrik olmayan Kruskal Wallis H ve devamında Mann Whitney U testi yapılacaktır. Kruskal Wallis H testi çoklu grupların arasındaki anlamlı farklılığı ortaya koyabilmekte ancak bu anlamlı farklılığın hangi iki grup arasında olduğunu göstermemektedir. Bundan dolayı bu testin devamında ikili gruplar arasındaki anlamlı farkı ortaya koyabilen Mann Whitney U testiyle gruplar arası ilişkiler tek tek tespit edilecektir. Öncelikle Çizelge 5.8’de faaliyet gruplarına ilişkin tanımlayıcı bilgiler sunulmuştur.

Çizelge 5.8. Tanımlayıcı bilgiler

Gruplar	İş Emri Sayısı	Ortalama Kullanılan Malzeme Çeşidi	Std. Sapma	Std. Hata	95% Güvenlik Aralığında Ort.		Minimum	Maksimum
					Alt Sınır	Üst Sınır		
Elektrik Arızası	2 599	3,98	2,878	0,056	3,87	4,09	2	19
Elektronik Arızası	1 335	2,78	1,072	0,029	2,72	2,84	2	8
Elektronik ünitenin elektrik arızası	2 571	3,42	2,455	0,048	3,33	3,52	2	20
Kalibre	247	2,42	0,832	0,053	2,32	2,53	2	6
İmalat	789	4,74	4,168	0,148	4,45	5,03	2	23
Toplam	7 541	3,61	2,704	0,031	3,55	3,67	2	23

Çizelge 5.8’te faaliyet gruplarının örneklem sayıları, ortalama malzeme kullanım değerleri, standart sapma ve hata değerleri, % 95 güven aralığında alt ve üst sınır değerleri, minimum ve maksimum kullanılan malzeme sayıları gösterilmiştir. Buradaki amaç ortalama değerlerinin birbirinden farklı olup olmadığının tespit edilmesidir. Bu değerlerin Çizelge 5.8’de farklı oldukları görülmesine rağmen, bu farklılığın anlamlı olup olmadığını anlamak için Kruskal Wallis H testi uygulanmalıdır. Çizelge 5.9’da bu testin sonuçları sunulmuştur.

Çizelge 5.9. Kruskal wallis test sonuçları

	Arıza Grubu	İş Emri Sayısı	Ortalama Puan
MLZ_ÇŞT	Elektrik Arızası	2 599	4 120,84
	Elektronik Arızası	1 335	3 266,48
	Elektronik ünitenin elektrik arızası	2 571	3 620,73
	Kalibre	247	2 585,19
	İmalat	789	4 333,14
	Toplam	7 541	
		MLZ_ÇŞT	
Chi-Square		311,418	
Df		4	
Asymp. Sig.		0,000	

Çizelge 5.9’da bulunan “Ortalama Puan” sütununa bakarak hangi gruplar arasında anlamlı farkın olduğu sezgisel olarak yorumlanabilse de, bilimsel sonuç için yine iki adet hipotez oluşturulup, bu hipotezlerden bir tanesi “Sig.” değerine bakılarak seçilir.

H_0 : 5 grup da aynı medyanlı sürekli dağılıma sahip toplumlardan alınmış örneklerdir,

H_1 : En az bir grubun medyan değeri diğer gruplardan farklıdır.

Çizelge 5.9'daki "Sig." değeri 0,05'ten küçük olduğu (sıfıra çok yakın olduğundan 0,000 görünmesine rağmen gerçek değeri 3.729141903920777E-66 dır) için H_1 hipotezi kabul edilir. Bu da söz konusu beş faaliyet grubunun en az ikisi arasında sarf malzeme kullanma açısından anlamlı bir farkın olduğunu gösterir. Bu aşamadan sonra hangi gruplar arasında anlamlı farkın olduğunu tespiti için "2 independent samples" testlerinden biri olan Mann Whitney U testi kullanılacaktır. Mann Whitney U testi ikili gruplar arasında sonuç verdiğinden, bu beş grubun birbirleri ile olan ilişkisi tek tek incelenecektir. Böylelikle C(5,2) toplam 10 adet Mann Whitney U testi uygulanacaktır. Bu testlerin sonuçlarının karşılaştırılması için bahsi geçen test örnek iki grup arasında yapılmış ve elde edilen sonuç Çizelge 5.10'da sunulmuştur.

Çizelge 5.10. Mann-Whitney test sonuçları

Tanımlayıcı İstatistikler						
	İş Emri	Ort.Mlz.	Std. Sapma	Minimum	Maksimum	
MLZ_ÇŞT	7 541	3,61	2,704	2	23	
Arıza Grubu	7 541	2,38	1,272	1	5	
Puanlar				Test İstatistikleri		
	Arıza Grubu	İş Emri	Ort.Puan	Toplam Puan	MLZ_ÇŞT	
MLZ_ÇŞT	Elektrik Arızası	2 599	2 123,55	5 519 116,00	Mann-Whitney U	1,329E6
	Elektronik Arızası	1 335	1 663,69	2 221 029,00	Wilcoxon W	2,221E6
					Z	-12,672
	Toplam	3 934			Asymp. Sig. (2-tailed)	0,000

Birinci grup (elektrik arızası) ile ikinci grup (elektronik arızası) arasında anlamlı farklılığın olup olmadığına ilişkin Çizelge 5.10'da gösterilen Mann-Whitney Testi sonuçlarına göre, %95 güven aralığında bu iki grubun arasındaki anlamlı farkı ortaya koyan Sig.değeri 0,05'ten küçük olduğundan, H_1 hipotezi kabul edilir. Yani elektrik arıza grubu ile elektronik arıza grubu anlamlı derecede birbirlerinden farklıdırlar. Böylelikle grupların ortalama malzeme çeşidi sayıları ile gruplardan elde edilecek birliktelik kural sayısı karşılaştırılabilecektir. Daha önce de bahsedildiği üzere tabloda "0,000" olarak görünen değerler aslında sıfır olmayıp, sıfıra çok yakın değerleri temsil etmektedirler.

Yukarıda iki grup için yapılan Mann-Whitney testi, diğer gruplar için de yapılarak, testlerin sonuçları Çizelge 5.11’de özet şekilde sunulmuştur.

Çizelge 5.11. Mann-Whitney test sonuç özetleri

Grup Kodları	Grup İsimleri	Ortalama Puan		Asymp. Sig. (2-tailed)	Anlamlı Fark
1-2	Elektrik-Elektronik	2123,55	1663,69	0,000	Var
1-3	Elektrik-Elektronik’in Elektrikçi	2756,18	2412,97	0,000	Var
1-4	Elektrik-Kalibre	1473,06	902,00	0,000	Var
1-5	Elektrik-İmalat	1668,05	1781,61	0,003	Var
2-3	Elektronik- Elektronik’in Elektrikçi	1836,96	2014,02	0,000	Var
2-4	Elektronik-Kalibre	817,50	650,97	0,000	Var
2-5	Elektronik-İmalat	952,33	1248,91	0,000	Var
3-4	Elektronik’in Elektrikçi- Kalibre	1443,57	1054,82	0,000	Var
3-5	Elektronik’in Elektrikçi- İmalat	1608,18	1916,17	0,000	Var
4-5	Kalibre-İmalat	349,40	571,44	0,000	Var

Çizelge 5.11’deki özet tablo sonuçlarına göre tüm %95 güven aralığında grupların birbirleri arasındaki anlamlı farkı ortaya koyan Sig.değeri 0,05’ten küçük olduğundan, H_1 hipotezi kabul edilir. Yani her grup, diğer gruplardan anlamlı derecede farklıdır.

5.4. Veriler İçin Kullanılan Yöntem ve Algoritma

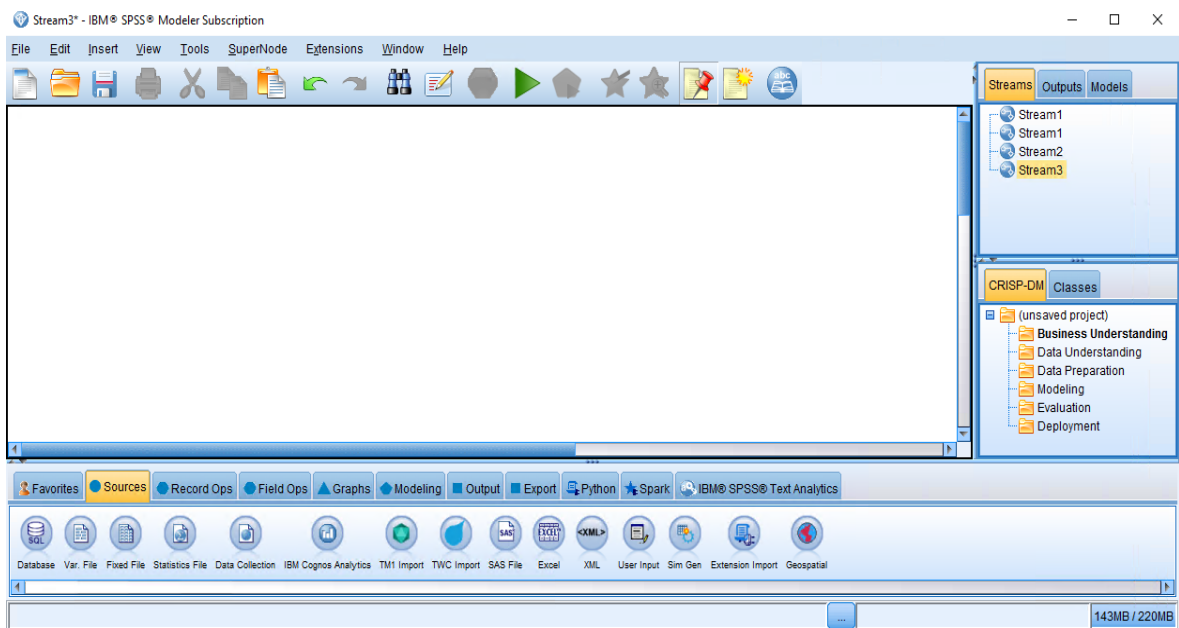
Analizde SPSS Modeler programı kullanılacaktır. SPSS Modeler programındaki “Modelling” özelliğinde bulunan “Apriori” algoritması modeliyle birliktelik analizi gerçekleştirilecektir. SPSS Modeler programının kolay kullanabilecek bir arayüzü ve kısa sürede sonuca ulaşabilecek görsel bir çalışma ortamı mevcuttur. Araştırmacının kullanmak isteyeceği tüm fonksiyonlar ara yüzün alt tarafında bulunan bir araç çubuğunda ikonlar halinde yer almaktadır. Birliktelik kurallarının elde edilmesi için yapılması gereken, gerekli ikonların analiz sayfasına taşınarak birbirlerine bağlantılarının yapıldıktan sonra çalıştırılmasıdır.

Yaygın kullanılması ve tercih edilen bir program olması, kullanım kolaylığı, farklı kaynaklardan elde edilen veri setlerinin kolay entegre edilebilmesi ve orta ölçek

büyüklüğündeki veri setine uygunluğundan dolayı bu tezin uygulama bölümündeki analizlerin SPSS Modeler programı ile gerçekleştirilmesine karar verilmiştir.

Uygulama çalışmasında, öncelikle tüm veri seti için SPSS Modeler programı çalıştırılıp genel birliktelik kurallarının elde edilmesi, ikinci aşamada faaliyet gruplarının her biri için program çalıştırılarak özel birliktelik kuralları elde edilmesi son aşamada ise elde edilen genel ve özel kuralların nitelik ve nicelik yönünden karşılaştırılması planlanmaktadır.

Resim 5.2’de, SPSS Modeler programının ana ekranı sunulmuştur. Apriori algoritması modeli ile birliktelik kuralları bu ekran üzerinden verilecek komutlarla gerçekleştirilecektir.



Resim 5.2. Modeler boş ara yüzü

5.4.1. Minimum destek ve minimum güven değerinin belirlenmesi

Dördüncü bölümde ayrıntılı olarak değinilen destek değeri, kısaca ilgili verinin analize konu olmaya değecek yeterlilikte olup olmadığının belirlenmesinde kullanılır. Veri setinde yer alan sütun sayısı ile ters orantılı olmalıdır. Yani verinin dahil olacağı sınıf sayısı çoğaldıkça, destek değeri azaltılmalı, sınıf sayısı azaldıkça artırılmalıdır. Böylece dikkate alınması gereken ilişkiler korumaya alınırken, bütün içerisinde önem arz etmeyen ya da nispeten daha az öneme sahip ilişkiler ise analiz dışında bırakılabilmektedir. Minimum

güven değeri ise tamamen analizcinin göze aldığı riskle doğru orantılı olarak belirlenir. Örneğin, analizci hiç risk almak istemiyorsa bu değeri %100 seçebilir. Bu değerler analizin amacına uygun düzeylerde seçilmelidir.

Bu uygulamada bütün veri setinde 180 adet malzeme çeşidi, dolayısıyla 180 adet sütun bulunmaktadır. Veri olarak kullanılan malzeme çeşidi fazla olduğundan minimum destek değeri %2 olarak belirlenmiştir. Bu değer, ilk bakışta düşük olarak algılansa da, kullanılan veri setinin büyüklüğü dikkate alındığında, uygun oranda seçilmiş bir değerdir. Veri setinde 180 farklı malzeme çeşidi bulunduğu ve bir iş emri için ortalama 3,61 adet malzeme çeşidi kullanıldığı bilgilerine dayanarak veri setinde eşit dağılım halinde, spesifik bir malzeme çeşidinin hangi oranda olması gerektiği aşağıda formülize edilmiştir. Söz konusu formül aracılığıyla ulaşılan değer altında bir değere sahip olan birliktelik ilişkileri analiz harici bırakılacağından, ilgili formül aynı zamanda minimum destek değerinin belirlenmesinde kullanılmış olacaktır.

Minimum destek değeri = $100 * \text{Ortalama kullanılan malz. çeşidi} / \text{Toplam malz. çeşidi}$

$$\text{Minimum destek değeri} = 100 * 3,61 / 180 = 2,02 \quad (5.1)$$

değerine ulaşılmıştır. Böylece analizlerde kullanılacak minimum destek değeri %2 olarak belirlenmiştir. Çalışmada kullanılan veri seti boyutunun büyük olması sebebiyle, elde edilecek birliktelik ilişkilerinin de fazla miktarda olacağı düşünülerek, minimum güven değeri %85 olarak, nispeten yüksek bir değer belirlenmiştir. Tüm veri setinden elde edilecek sonuçlarla nicel olarak karşılaştırılacağından, alt veri setleri yani her faaliyet grubunun veri setleri için de bu belirlenen destek ve güven değerleri kullanılacaktır.

5.5. SPSS Modeler Programın Kullanımı ve Veri Seti Üzerinde Çalıştırılması

Birliktelik analizinde kullanılacak olan Modeler programının ve analiz için gerekli olan “Apriori” algoritması modelinin nasıl çalıştığı ekran görüntüleriyle birlikte Ek-2’de sunulmuştur.

5.6. Bulgular

Önceki kısımda da değinildiği üzere analizin üç temel aşamada tamamlanması planlanmıştır. Öncelikle tüm veri setinden elde edilen bulgular değerlendirildikten sonra faaliyet gruplarının verileri değerlendirilecek ve son olarak bu iki değerlendirmenin sonuçları karşılaştırılacaktır. Her ne kadar destek ve güven oranları sayesinde yeterli kullanılabilirlik düzeyinde birliktelik kuralları elde edilmiş olsa da, en faydalı ve güçlü kuralları bulmak için elde edilen kurallar ikinci bir analize tabi tutulacaktır. Bu analiz sonucunda elde edilen kurallara yönelik nitelendirmelerde bulunulacaktır. Diğer bir ifadeyle, güçlü kuralların ortaya çıkarılması sağlanacaktır.

5.6.1. Bütün veri setinden elde edilen bulgular

180 çeşit malzeme sınıfının kullanım bilgilerini içeren tüm veri seti için Apriori algoritmasının %2 minimum destek ve %85 minimum güven kriterleri altında yürütülmesiyle **121** adet birliktelik kuralı elde edilmiştir. Bu birliktelik kuralları, Öncül (Antecedent) ve Ardıl (Consequent) şeklinde belirtilmiştir. Yani, önce gelen malzeme çeşitleri ile takip eden malzeme çeşitleri arasındaki birliktelik kuralları ortaya çıkarılmıştır.

SPSS Modeler programında Apriori algoritmasını yürütülmesiyle elde edilen sonuçların bir kısmının gösterildiği ekran görüntüleri Resim 5.3 ve Resim 5.4'te sunulmuştur.

180 fields

File Generate Preview

Model Settings Summary Annotations

Sort by: Support % 121 of 121

Consequent	Antecedent	Support %	Confidence %
Elektrik irtibat parçaları	Aküler	4.363	100.0
Transistörler	Elektrik tel ve kablolar	3.594	88.561
Elektrik kumanda malz.	Dirençler	3.514	87.17
Elektrik ızalatörleri	Kapasitörler	3.408	89.883
Aküler	Devre kesiciler	3.328	99.602
Elektrik irtibat parçaları	Elektrik ızalatörleri	3.328	99.602
Elektrik irtibat parçaları	Elektrik kumanda malz.	3.315	100.0
Aküler	Şahsi koruyucu malz.	3.315	100.0
Devre kesiciler	Elektrik irtibat parçaları	3.222	95.062
Takat ve tevzi malz.	Elektrik ızalatörleri	3.222	85.185
Devre kesiciler	Elektrik kumanda malz.	3.196	86.722
Elektrik ızalatörleri	Elektrik kumanda malz.	3.196	85.892
Devre kesiciler	Takat ve tevzi malz.	3.143	94.093
Elektrik ızalatörleri	Elektrik ızalatörleri	3.143	85.654
Elektrik kumanda malz.	Devre kesiciler	3.143	87.342
	Sigortalar ve paratonerler		
	Takat ve tevzi malz.		
	Elektrik ızalatörleri		

OK Cancel Apply Reset

Resim 5.3. Destek sıralı birliktelik kuralları

180 fields

File Generate Preview

Model Settings Summary Annotations

Sort by: Confidence % 121 of 121

Consequent	Antecedent	Support %	Confidence %
Elektrik irtibat parçaları	Şahsi koruyucu malz.	3.315	100.0
Aküler	Aküler	3.315	100.0
Elektrik irtibat parçaları	Elektrik irtibat parçaları	4.363	100.0
Elektrik irtibat parçaları	Elektrik tel ve kablolar	2.705	100.0
Aküler	Şahsi koruyucu malz.	2.705	100.0
Aküler	Elektrik irtibat parçaları	3.328	99.602
Elektrik irtibat parçaları	Elektrik tel ve kablolar	3.328	99.602
Aküler	Şahsi koruyucu malz.	2.718	99.512
Elektrik irtibat parçaları	Elektrik tel ve kablolar	2.718	99.512
Elektrik irtibat parçaları	Şahsi koruyucu malz.	2.188	98.788
Devre kesiciler	Yapışkan maddeler	2.055	98.065
Transistörler	Aküler	2.268	97.661
Elektrik ızalatörleri	Elektrik ızalatörleri	2.108	97.484
	Bobinler		
	Dirençler		
	Kapasitörler		
	Mikro elektronik devre parçaları		
	Devre kesiciler		
	Takat ve tevzi malz.		
	Elektrik tel ve kablolar		

OK Cancel Apply Reset

Resim 5.4. Güven sıralı birliktelik kuralları

Belirlenen koşullar altında elde edilen birliktelik kurallarının tamamı (121 adet) ise Çizelge 5.12’de sunulmuştur.

Çizelge 5.12. Bütün veri setinden elde edilen kurallar

Kural No	Öncül	Ardıl	Örnek Adedi	Destek Oranı	Güven Oranı
1	Aküler & Elektrik tel ve kabloları	Elektrik irtibat parçaları	329	4,36	100,00
2	Dirençler & Kapasitörler	Transistörler	271	3,59	88,56
3	Devre kesiciler & Elektrik izalatörleri	Elektrik kumanda malz.	265	3,51	87,17
4	Devre kesiciler & Elektrik kumanda malz.	Elektrik izalatörleri	257	3,41	89,88
5	Şahsi koruyucu malz.	Aküler	251	3,33	99,60
6	Şahsi koruyucu malz.	Elektrik irtibat parçaları	251	3,33	99,60
7	Şahsi koruyucu malz. & Aküler	Elektrik irtibat parçaları	250	3,31	100,00
8	Şahsi koruyucu malz. & Elektrik irtibat parçaları	Aküler	250	3,31	100,00
9	Elektrik kumanda malz. & Elektrik izalatörleri	Devre kesiciler	243	3,22	95,06
10	Elektrik kumanda malz. & Elektrik izalatörleri	Takat ve tevzi malz.	243	3,22	85,19
11	Elektrik kumanda malz. & Takat ve tevzi malz.	Devre kesiciler	241	3,20	86,72
12	Elektrik kumanda malz. & Takat ve tevzi malz.	Elektrik izalatörleri	241	3,20	85,89
13	Takat ve tevzi malz. & Elektrik izalatörleri	Devre kesiciler	237	3,14	94,09
14	Takat ve tevzi malz. & Elektrik izalatörleri	Elektrik kumanda malz.	237	3,14	87,34
15	Devre kesiciler & Sigortalar ve paratonerler	Elektrik izalatörleri	237	3,14	85,65
16	Devre kesiciler & Takat ve tevzi malz.	Elektrik izalatörleri	236	3,13	94,49
17	Devre kesiciler & Takat ve tevzi malz.	Elektrik kumanda malz.	236	3,13	88,56
18	Devre kesiciler & Elektrik kumanda malz. & Elektrik izalatörleri	Takat ve tevzi malz.	231	3,06	86,15
19	Elektrik izalatörleri & Sigortalar ve paratonerler	Devre kesiciler	229	3,04	88,65
20	Devre kesiciler & Takat ve tevzi malz. & Elektrik izalatörleri	Elektrik kumanda malz.	223	2,96	89,24
21	Mikro elektronik devre parçaları & Elektrik anahtarları	Transistörler	220	2,92	89,55
22	Devre kesiciler & Elektrik kumanda malz. & Takat ve tevzi malz.	Elektrik izalatörleri	209	2,77	95,22
23	Elektrik kumanda malz. & Takat ve tevzi malz. & Elektrik izalatörleri	Devre kesiciler	207	2,74	96,14
24	Şahsi koruyucu malz. & Elektrik tel ve kabloları	Aküler	205	2,72	99,51
25	Şahsi koruyucu malz. & Elektrik tel ve kabloları	Elektrik irtibat parçaları	205	2,72	99,51
26	Şahsi koruyucu malz. & Aküler & Elektrik tel ve kabloları	Elektrik irtibat parçaları	204	2,70	100,00
27	Şahsi koruyucu malz. & Elektrik irtibat parçaları & Elektrik tel ve kabloları	Aküler	204	2,70	100,00
28	Devre kesiciler & Elektrik izalatörleri & Sigortalar ve paratonerler	Takat ve tevzi malz.	203	2,69	89,16
29	Devre kesiciler & Elektrik izalatörleri & Sigortalar ve paratonerler	Elektrik kumanda malz.	203	2,69	88,18
30	Devre kesiciler & Elektrik kumanda malz. & Sigortalar ve paratonerler	Elektrik izalatörleri	199	2,64	89,95
31	Devre kesiciler & Elektrik kumanda malz. & Sigortalar ve paratonerler	Takat ve tevzi malz.	199	2,64	86,93
32	Muhtelif kablolar & Mikro elektronik devre parçaları	Transistörler	197	2,61	85,79
33	Devre kesiciler & Takat ve tevzi malz. & Sigortalar ve paratonerler	Elektrik izalatörleri	194	2,57	93,30
34	Takat ve tevzi malz. & Elektrik izalatörleri & Sigortalar ve paratonerler	Devre kesiciler	194	2,57	93,30
35	Devre kesiciler & Takat ve tevzi malz. & Sigortalar ve paratonerler	Elektrik kumanda malz.	194	2,57	89,18

Çizelge 5.12. (devam) Bütün veri setinden elde edilen kurallar

36	Takat ve tevzi malz. & Elektrik izalatörleri & Sigortalar ve paratonerler	Elektrik kumanda malz.	194	2,57	88,14
37	Elektrik aydınlatma malz. & Elektrik kumanda malz.	Takat ve tevzi malz.	192	2,55	88,02
38	Elektrik kumanda malz. & Elektrik izalatörleri & Sigortalar ve paratonerler	Devre kesiciler	189	2,51	94,71
39	Elektrik kumanda malz. & Elektrik izalatörleri & Sigortalar ve paratonerler	Takat ve tevzi malz.	189	2,51	90,48
40	Elektrik aydınlatma malz. & Takat ve tevzi malz.	Elektrik kumanda malz.	185	2,45	91,35
41	Elektrik aydınlatma malz. & Takat ve tevzi malz.	Elektrik izalatörleri	185	2,45	85,95
42	Elektrik aydınlatma malz. & Takat ve tevzi malz.	Devre kesiciler	185	2,45	85,41
43	Devre kesiciler & Elektrik tel ve kabloları	Elektrik izalatörleri	184	2,44	90,22
44	Devre kesiciler & Elektrik tel ve kabloları	Elektrik kumanda malz.	184	2,44	86,41
45	Devre kesiciler & Elektrik tel ve kabloları	Takat ve tevzi malz.	184	2,44	86,41
46	Yapışkan maddeler & Elektrik irtibat parçaları	Aküler	183	2,43	89,07
47	Devre kesiciler & Takat ve tevzi malz. & Elektrik izalatörleri & Sigortalar ve paratonerler	Elektrik kumanda malz.	181	2,40	90,06
48	Elektrik aydınlatma malz. & Devre kesiciler	Elektrik izalatörleri	180	2,39	95,00
49	Elektrik aydınlatma malz. & Sigortalar ve paratonerler	Elektrik kumanda malz.	180	2,39	88,89
50	Elektrik aydınlatma malz. & Devre kesiciler	Takat ve tevzi malz.	180	2,39	87,78
51	Elektrik aydınlatma malz. & Devre kesiciler	Elektrik kumanda malz.	180	2,39	87,22
52	Elektrik aydınlatma malz. & Elektrik izalatörleri	Devre kesiciler	179	2,37	95,53
53	Devre kesiciler & Elektrik kumanda malz. & Elektrik izalatörleri & Sigortalar ve paratonerler	Takat ve tevzi malz.	179	2,37	91,06
54	Elektrik aydınlatma malz. & Elektrik izalatörleri	Takat ve tevzi malz.	179	2,37	88,83
55	Elektrik aydınlatma malz. & Elektrik izalatörleri	Elektrik kumanda malz.	179	2,37	87,15
56	Devre kesiciler & Tesbit kulakları, uçları	Elektrik izalatörleri	175	2,32	90,86
57	Dirençler & Mikro elektronik devre parçaları & Elektrik anahtarları	Transistörler	175	2,32	90,86
58	Devre kesiciler & Tesbit kulakları, uçları	Elektrik kumanda malz.	175	2,32	85,14
59	Devre kesiciler & Elektrik kumanda malz. & Takat ve tevzi malz. & Sigortalar ve paratonerler	Elektrik izalatörleri	173	2,29	94,22
60	Dirençler & Elektrik anahtarları & Transistörler	Mikro elektronik devre parçaları	173	2,29	91,91
61	Dirençler & Kapasitörler & Mikro elektronik devre parçaları	Transistörler	171	2,27	97,66
62	Elektrik kumanda malz. & Takat ve tevzi malz. & Elektrik izalatörleri & Sigortalar ve paratonerler	Devre kesiciler	171	2,27	95,32
63	Elektrik aydınlatma malz. & Devre kesiciler & Elektrik izalatörleri	Takat ve tevzi malz.	171	2,27	89,47
64	Elektrik aydınlatma malz. & Devre kesiciler & Elektrik izalatörleri	Elektrik kumanda malz.	171	2,27	87,72
65	Elektrik kumanda malz. & Elektrik hırdavat malz.	Sigortalar ve paratonerler	170	2,25	92,94
66	Elektrik aydınlatma malz. & Elektrik kumanda malz. & Takat ve tevzi malz.	Elektrik izalatörleri	169	2,24	85,21
67	Devre kesiciler & Elektrik izalatörleri & Elektrik tel ve kabloları	Takat ve tevzi malz.	166	2,20	93,37
68	Devre kesiciler & Elektrik izalatörleri & Elektrik tel ve kabloları	Elektrik kumanda malz.	166	2,20	88,55
69	Yapışkan maddeler & Aküler	Elektrik irtibat parçaları	165	2,19	98,79
70	Elektrik kumanda malz. & Tesbit kulakları, uçları	Devre kesiciler	165	2,19	90,30
71	Elektrik kumanda malz. & Tesbit kulakları, uçları	Elektrik izalatörleri	165	2,19	89,09

Çizelge 5.12. (devam) Bütün veri setinden elde edilen kurallar

72	Elektrik kumanda malz. & Tesbit kulakları, uçları	Sigortalar ve paratonerler	165	2,19	85,45
73	PiezoElektrik kristalleri & Transistörler	Mikro elektronik devre parçaları	164	2,17	85,98
74	Takat ve tevzi malz. & Elektrik izalatörleri & Elektrik tel ve kabloları	Devre kesiciler	163	2,16	95,09
75	Takat ve tevzi malz. & Elektrik izalatörleri & Elektrik tel ve kabloları	Elektrik kumanda malz.	163	2,16	87,12
76	Devre kesiciler & Bobinler	Elektrik izalatörleri	160	2,12	95,00
77	Devre kesiciler & Bobinler	Sigortalar ve paratonerler	160	2,12	93,75
78	Takat ve tevzi malz. & Bobinler	Sigortalar ve paratonerler	160	2,12	92,50
79	Devre kesiciler & Bobinler	Takat ve tevzi malz.	160	2,12	88,13
80	Takat ve tevzi malz. & Bobinler	Devre kesiciler	160	2,12	88,13
81	Elektrik aydınlatma malz. & Elektrik kumanda malz. & Sigortalar ve paratonerler	Takat ve tevzi malz.	160	2,12	87,50
82	Devre kesiciler & Bobinler	Elektrik kumanda malz.	160	2,12	86,88
83	Takat ve tevzi malz. & Bobinler	Elektrik izalatörleri	160	2,12	85,63
84	Devre kesiciler & Takat ve tevzi malz. & Elektrik tel ve kabloları	Elektrik izalatörleri	159	2,11	97,48
85	Elektrik aydınlatma malz. & Takat ve tevzi malz. & Elektrik izalatörleri	Devre kesiciler	159	2,11	96,23
86	Devre kesiciler & Elektrik kumanda malz. & Elektrik tel ve kabloları	Elektrik izalatörleri	159	2,11	92,45
87	Elektrik aydınlatma malz. & Takat ve tevzi malz. & Elektrik izalatörleri	Elektrik kumanda malz.	159	2,11	90,57
88	Devre kesiciler & Elektrik kumanda malz. & Elektrik tel ve kabloları	Takat ve tevzi malz.	159	2,11	87,42
89	Devre kesiciler & Takat ve tevzi malz. & Elektrik tel ve kabloları	Elektrik kumanda malz.	159	2,11	87,42
90	Devre kesiciler & Elektrik izalatörleri & Tesbit kulakları, uçları	Elektrik kumanda malz.	159	2,11	86,79
91	Devre kesiciler & Elektrik izalatörleri & Tesbit kulakları, uçları	Takat ve tevzi malz.	159	2,11	86,79
92	Devre kesiciler & Elektrik izalatörleri & Tesbit kulakları, uçları	Sigortalar ve paratonerler	159	2,11	85,53
93	Elektrik aydınlatma malz. & Devre kesiciler & Takat ve tevzi malz.	Elektrik izalatörleri	158	2,09	96,84
94	Elektrik aydınlatma malz. & Devre kesiciler & Takat ve tevzi malz.	Elektrik kumanda malz.	158	2,09	89,87
95	Elektrik aydınlatma malz. & Devre kesiciler & Elektrik kumanda malz.	Elektrik izalatörleri	157	2,08	95,54
96	Mikro elektronik devre parçaları & Elektrik irtibat parçaları	Transistörler	157	2,08	92,36
97	Elektrik aydınlatma malz. & Devre kesiciler & Elektrik kumanda malz.	Takat ve tevzi malz.	157	2,08	90,45
98	Elektrik aydınlatma malz. & Elektrik kumanda malz. & Elektrik izalatörleri	Devre kesiciler	156	2,07	96,15
99	Ölçü aletleri malz. & Elektrik izalatörleri	Devre kesiciler	156	2,07	92,95
100	Elektrik aydınlatma malz. & Elektrik kumanda malz. & Elektrik izalatörleri	Takat ve tevzi malz.	156	2,07	92,31
101	Elektrik izalatörleri & Bobinler	Devre kesiciler	155	2,06	98,06
102	Elektrik kumanda malz. & Elektrik izalatörleri & Elektrik tel ve kabloları	Devre kesiciler	155	2,06	94,84
103	Elektrik izalatörleri & Bobinler	Sigortalar ve paratonerler	155	2,06	94,19
104	Ölçü aletleri malz. & Devre kesiciler	Elektrik izalatörleri	155	2,06	93,55
105	Elektrik kumanda malz. & Takat ve tevzi malz. & Elektrik tel ve kabloları	Elektrik izalatörleri	155	2,06	91,61

Çizelge 5.12. (devam) Bütün veri setinden elde edilen kurallar

106	Elektrik kumanda malz. & Elektrik izalatörleri & Elektrik tel ve kabloları	Takat ve tevzi malz.	155	2,06	91,61
107	Elektrik kumanda malz. & Takat ve tevzi malz. & Elektrik tel ve kabloları	Devre kesiciler	155	2,06	89,68
108	Elektrik izalatörleri & Bobinler	Takat ve tevzi malz.	155	2,06	88,39
109	Devre kesiciler & Takat ve tevzi malz. & Elektrik izalatörleri & Elektrik tel ve kabloları	Elektrik kumanda malz.	155	2,06	87,74
110	Elektrik izalatörleri & Bobinler	Elektrik kumanda malz.	155	2,06	87,10
111	Ölçü aletleri malz. & Devre kesiciler	Sigortalar ve paratonerler	155	2,06	86,45
112	Takat ve tevzi malz. & Elektrik hırdavat malz.	Sigortalar ve paratonerler	153	2,03	94,77
113	Takat ve tevzi malz. & Elektrik hırdavat malz.	Elektrik kumanda malz.	153	2,03	91,50
114	Çeşitli Elektrik ve elektronik parçalar & Mikrofon ve hoparlörler	Hırdavat malzemesi	153	2,03	90,85
115	Elektrik aydınlatma malz. & Devre kesiciler & Takat ve tevzi malz. & Elektrik izalatörleri	Elektrik kumanda malz.	153	2,03	90,20
116	PiezoElektrik kristalleri & Muhtelif kablolar & Hırdavat malzemesi	Mikrofon ve hoparlörler	153	2,03	88,89
117	Çeşitli Elektrik ve elektronik parçalar & Mikrofon ve hoparlörler	Muhtelif kablolar	153	2,03	86,93
118	Devre kesiciler & Elektrik izalatörleri & Bobinler	Sigortalar ve paratonerler	152	2,02	94,08
119	Devre kesiciler & Elektrik izalatörleri & Bobinler	Takat ve tevzi malz.	152	2,02	88,82
120	Devre kesiciler & Elektrik izalatörleri & Bobinler	Elektrik kumanda malz.	152	2,02	87,50
121	Elektrik izalatörleri & Elektrik anahtarları	Devre kesiciler	152	2,02	85,53

Elde edilen 121 kural, önce minimum destek sonra minimum güven değerlerine göre sıralanarak Çizelge 5.12’de gösterilmiştir. Böylelikle değerli ve güçlü kurallar ön sıralarda sunulmuştur. Kuralların tümü incelendiğinde minimum destek değerine sahip kuralın destek oranı %2,02, maksimum destek değerine sahip kuralın destek oranı %4,36 olarak gözlemlenirken, aynı şekilde minimum güven değerine sahip kuralın güven oranı %85,14, maksimum güven değerine sahip kuralın güven oranı %100 olarak gözlemlenmektedir.

Çizelge 5.12’de sunulan, destek ve sonrasında güven sıralı birliktelik kurallarından bazıları aşağıdaki gibi yorumlanabilir.

- 1 no.lu kural: *Aküler & Elektrik tel ve kabloları* → *Elektrik irtibat parçaları* birliktelik kuralı için;

Aküler & Elektrik tel ve kabloları ile Elektrik irtibat parçalarının toplam iş hareketlerinde birlikte görülme olasılıkları % 4,36’dır. Ayrıca Aküler & Elektrik tel ve kablolarının her ikisinin birlikte kullanıldığı bir işte, % 100 olasılıkla Elektrik irtibat parçalarının da kullanılacağı söylenebilir.

- *2 no.lu kural: Dirençler & Kapasitörler → Transistörler birliktelik kuralı için;*
Dirençler & Kapasitörler ile Transistörlerin toplam iş hareketlerinde birlikte görülme olasılıkları % 3,59'dır. Ayrıca Dirençler & Kapasitörlerin her ikisinin birlikte kullanıldığı bir işte, % 88,56 olasılıkla Transistörlerin de kullanılacağı söylenebilir.
- *3 no.lu kural: Devre kesiciler & Elektrik izolatörleri → Elektrik kumanda malzemeleri birliktelik kuralı için;*
Devre kesiciler & Elektrik izolatörleri ile Elektrik kumanda malzemeleri toplam iş hareketlerinde birlikte görülme olasılıkları % 3,51'dir. Ayrıca Devre kesiciler & Elektrik izolatörleri her ikisinin birlikte kullanıldığı bir işte, % 87,17 olasılıkla Elektrik kumanda malzemelerinin de kullanılacağı söylenebilir.
- *4 no.lu kural: Devre kesiciler & Elektrik kumanda malzemeleri → Elektrik izolatörleri birliktelik kuralı için;*
Devre kesiciler & Elektrik kumanda malzemeleri ile Elektrik izolatörleri toplam iş hareketlerinde birlikte görülme olasılıkları % 3,41'dir. Ayrıca Devre kesiciler & Elektrik kumanda malzemelerinin her ikisinin birlikte kullanıldığı bir işte, % 89,88 olasılıkla Elektrik izolatörlerinin de kullanılacağı söylenebilir.
- *5 no.lu kural: Şahsi koruyucu malzemeler → Aküler birliktelik kuralı için;*
Şahsi koruyucu malzemeler ile Akülerin toplam iş hareketlerinde birlikte görülme olasılıkları % 3,33'tür. Ayrıca Şahsi koruyucu malzemelerin kullanıldığı bir işte, % 89,88 olasılıkla Akülerin de kullanılacağı söylenebilir.
- *6 no.lu kural: Şahsi koruyucu malzemeler → Elektrik irtibat parçaları birliktelik kuralı için;*
Şahsi koruyucu malzemeler ile Elektrik irtibat parçalarının toplam iş hareketlerinde birlikte görülme olasılıkları % 3,33'tür. Ayrıca Şahsi koruyucu malzemelerin kullanıldığı bir işte, % 99,60 olasılıkla Elektrik irtibat parçalarının da kullanılacağı söylenebilir.
- *7 no.lu kural: Şahsi koruyucu malzemeler & Aküler → Elektrik irtibat parçaları birliktelik kuralı için;*

Şahsi koruyucu malzemeler & Aküler ile Elektrik irtibat parçalarının toplam iş hareketlerinde birlikte görülme olasılıkları % 3,31'dir. Ayrıca Şahsi koruyucu malzemeler & Akülerin her ikisinin birlikte kullanıldığı bir işte, % 100 olasılıkla Elektrik irtibat parçalarının da kullanılacağı söylenebilir.

- 8 no.lu kural: *Şahsi koruyucu malzemeler & Elektrik irtibat parçaları → Aküler* birliktelik kuralı için;

Şahsi koruyucu malzemeler & Elektrik irtibat parçaları ile Akülerin toplam iş hareketlerinde birlikte görülme olasılıkları % 3,31'dir. Ayrıca Şahsi koruyucu malzemeler & Elektrik irtibat parçalarının her ikisinin birlikte kullanıldığı bir işte, % 100 olasılıkla Akülerin de kullanılacağı söylenebilir.

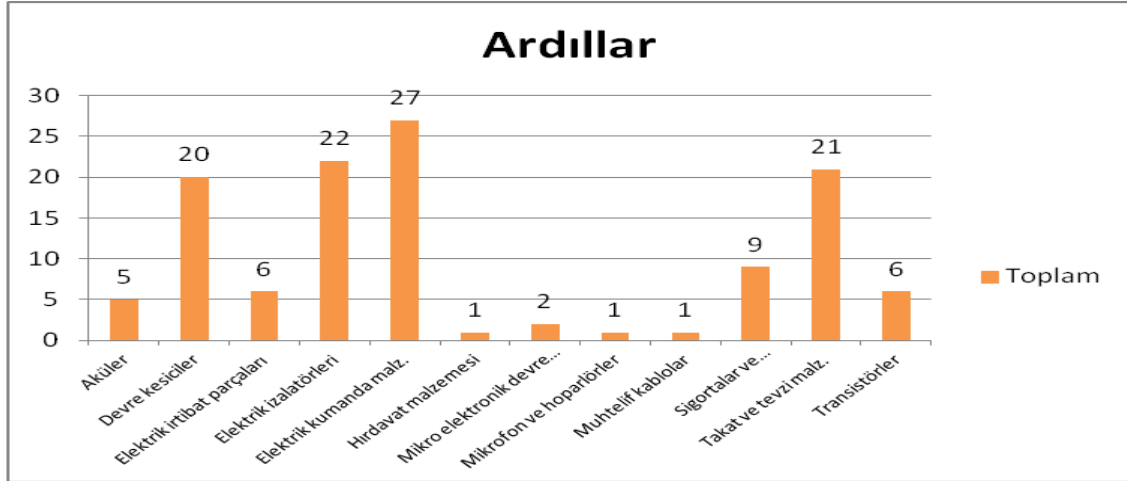
- 9 no.lu kural: *Elektrik kumanda malzemeleri & Elektrik izolatörleri → Devre kesiciler* birliktelik kuralı için;

Elektrik kumanda malzemeleri & Elektrik izolatörleri ile Devre kesicilerin toplam iş hareketlerinde birlikte görülme olasılıkları % 3,22'dir. Ayrıca Elektrik kumanda malzemeleri & Elektrik izolatörlerinin her ikisinin birlikte kullanıldığı bir işte, % 95,06 olasılıkla Devre kesicilerin de kullanılacağı söylenebilir.

- 10 no.lu kural: *Elektrik kumanda malzemeleri & Elektrik izolatörleri → Takat ve tevzi malzemeleri* birliktelik kuralı için;

Elektrik kumanda malzemeleri & Elektrik izolatörleri ile Takat ve tevzi malzemelerinin toplam iş hareketlerinde birlikte görülme olasılıkları % 3,22'dir. Ayrıca Elektrik kumanda malzemeleri & Elektrik izolatörlerinin her ikisinin birlikte kullanıldığı bir işte, % 85,19 olasılıkla Takat ve tevzi malzemelerinin de kullanılacağı söylenebilir.

Yukarıda 10 kural için yapılan yorumlar, ortaya çıkan 121 birliktelik kuralından geriye kalanlar için de yapılabilir. Diğer yandan çıkan sonuçların ardılları incelendiğinde, elde edilen 121 kuralın sadece 12 ardıla yöneldiği tespit edilmiştir. Her ne kadar ortaya çıkan kural sayısı 121 olsa da, analiz sonucu olarak, elde edilen kuralların ardılları dikkate alındığından, aslında kümelenmiş 12 adet sonuca odaklanılmalıdır. Söz konusu durum Şekil 5.1'de gösterilmiştir.

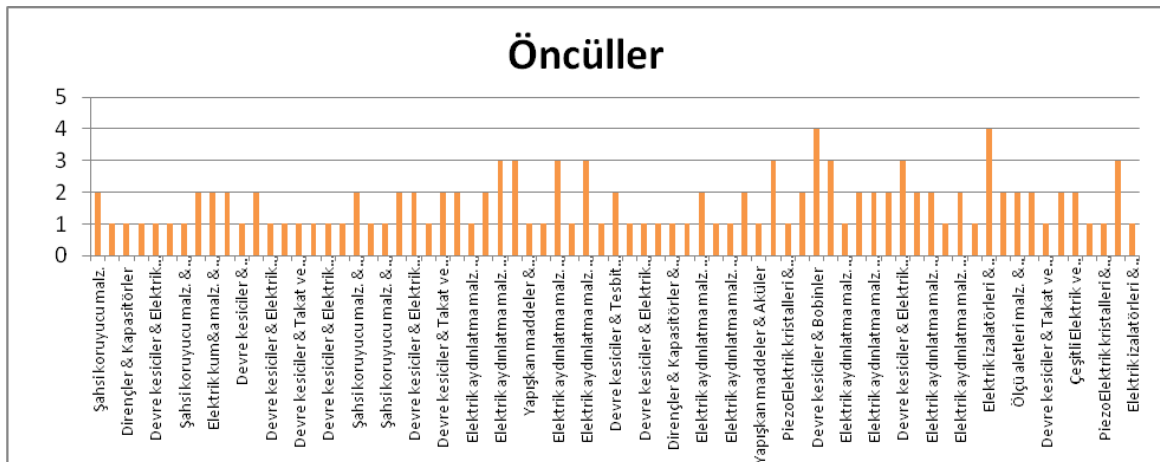


Şekil 5.1. Kuralların ardılları grafiği

Şekil 5.1’de sunulan ardıllardan en sık tekrarlananların;

- Elektrik kumanda malzemeleri (27 ardıl),
- Elektrik izolatörleri (22 ardıl),
- Takat ve tevzi malzemeleri (21 ardıl),
- Devre kesiciler (20 ardıl) olduğu görülmektedir.

Bununla birlikte öncüller incelendiğinde, 121 kuralın oluşumunda toplamda 73 farklı öncül grubunun rol oynadığı tespit edilmiştir. Bir öncül grubundan, tek bir ardılın da, farklı destek ve güven değerlerine sahip farklı ardılların oluşabildiği belirlenmiştir. Söz konusu durum Şekil 5.2’de gösterilmiştir.



Şekil 5.2. Kuralların öncülleri grafiği

Şekil 5.2’den de görüldüğü üzere aynı öncül grubundan 4 farklı ardıl ortaya çıkabilmektedir. Örneğin, Devre kesiciler & Bobinler öncülü ele alındığında yani eğer bir işte devre kesiciler ve bobinler birlikte kullanılırsa;

- %95 olasılıkla Elektrik izolatörleri,
- %86,88 olasılıkla Elektrik kumanda malzemeleri,
- %93,75 olasılıkla Sigorta ve paratonerler ve
- %88,13 olasılıkla Takat ve tevzi malzemeleri kullanılır.

Bununla birlikte tek bir öncülden tek bir ardılın ortaya çıkması, doğal olarak daha muhtemeldir. Örneğin, Piezo elektrik kristalleri ile Transistörlerin birlikte kullanıldığı bir işte, %85,98 olasılıkla Mikro elektronik devre parçaları da kullanılır. Şekil 5.2’de sunulan bilgilere ilişkin özet tablo Çizelge 5.13’te verilmiştir.

Çizelge 5.13. Öncül, ardıl ilişkisi

Öncül Sayısı	Ardıl Sayısı	Kural Sayısı
37	1	37
26	2	52
8	3	24
2	4	8
Toplam		121

Çizelge 5.13’ten de anlaşıldığı üzere 37 öncül grubundan tek bir ardıl meydana gelirken, 26 öncül grubu 2 ardıl birden, 8 öncül grubu 3 ardıl birden ve 2 öncül grubu da 4 ardıl birden meydana getirmişlerdir.

Şekil 5.1 ve Şekil 5.2 birlikte yorumlanırsa, 73 adet öncül grubunun 12 adet ardıl oluşturduğu ve böylece 121 adet birliktelik kuralının elde edildiği söylenebilir.

5.6.2. Faaliyet gruplarının veri setlerinden elde edilen bulgular

Bu bölümde tüm veri setini oluşturan bilgiler 5 ana faaliyet grubuna bölünerek, birliktelik kuralları çıkarılacaktır. Daha önce de belirtildiği üzere, bu beş alt grup işletmenin ana faaliyetleri dikkate alınarak oluşturulmuştur. Bütün veri seti, birbirleriyle ilişkili faaliyet gruplarına ayrıldığından bu kısımda yapılacak analizlerin sonucunda daha fazla birliktelik

kuralı elde edileceği düşünülmektedir. Ayrıca destek oranının bütün veri setindeki değer ile aynı seçilmesi de kural sayısının artmasına etken olacaktır.

Bu aşamada elde edilen kurallardan en güçlü olanlarının seçilmesi sağlanacaktır. Söz konusu seçim iki farklı yöntem kullanılarak yapılacaktır. Yöntem-1; alt veri setinden elde edilen kural, ana veri setinden de elde edilen kurallar arasında mevcut bulunuyorsa, o kural güçlü kural olarak adlandırılıp, sonuç kural listesine dahil edilecektir. Güçlü kural seçiminde Yöntem-2 ise, alt gruplardan yükseltilmiş minimum destek oranını aşan kuralların seçilip, sonuç listesine ekleneceği yöntemdir. Yükseltilmiş minimum destek oranı (YMDO) aşağıdaki formüle göre hesaplanır.

YMDO = Bir iş için kullanılan ortalama malzeme sayısı *100 / Veri setinde kullanılan malzeme çeşidi sayısı

Bu formül, ana veri setinin minimum destek oranı belirlenirken kullanılan formüldür. Yukarıda verilen formüle göre hesaplanan alt veri setlerinin YMDO değerleri Çizelge 5.14'te sunulmuştur.

Çizelge 5.14. Yükseltilmiş minimum destek oranları

Gruplar	Malzeme Çeşidi	Kullanılan Malzeme Ortalaması	YMDO
Tümü	180	3,61	2,02
Elektrik Arz.	98	3,98	4,06
Elektronik Arz.	74	2,78	3,76
Elektronikğin Elektrik Arz.	95	3,42	3,60
Kalibre	41	2,42	5,90
İmalat	129	4,74	3,67

Bu kısımda, önceki kısımda yapılan analiz işlemleri tekrarlanacağından ayrıntılı anlatımda bulunulmayacaktır. Bu kapsamda alt grup veri setlerine ilişkin bilgiler ve analiz sonuçları aşağıda özetlenmiştir.

- *Elektrik arızaları grubu:* Bu grupta, 98 malzeme çeşidinin kullanıldığı 2 599 iş emri üzerinden analizler gerçekleştirilmiştir. Bütün veri setinde olduğu gibi, minimum destek değeri %2, minimum güven değeri de %85 olarak seçilmiştir. Apriori algoritmasının yürütülmesi sonucunda elektrik arıza grubuna ilişkin 6 332 adet birliktelik kuralı elde

edilmiştir. Elde edilen kuralların destek değerleri %2 ile %19,28 arasında sıralanırken, bu kuralların güven oranları ise %85 ile %100 arasında çeşitli değerler almışlardır.

İlk paragrafta değinilen sebeplerden dolayı, bu faaliyet grubunda çok sayıda kural elde edilmiştir. Yöntem-1 dikkate alındığında, bu alt gruptan elde edilen 6 332 kuralın 60 tanesi, bütün veri setinden elde edilen 121 kural ile kesişmektedir. Çizelge 5.12’de verilen listedeki 1, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 18, 24, 25, 26, 27, 30, 31, 37, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 54, 55, 56, 58, 63, 64, 67, 68, 69, 70, 71, 72, 74, 75, 81, 85, 86, 87, 88, 95, 97, 98, 99, 100, 102, 104, 106, 111 numaralı kurallar, elektrik arıza grubu veri setinden de elde edilmiştir. Dolayısıyla sayılan bu kuralların güçlü kurallar oldukları söylenebilir. Yöntem-2’nin gereği olarak elektrik arıza grubunun veri setinden, YDMO (4,06) dikkate alınarak analiz yapıldığında, 550 adet güçlü kural daha çıkarılmaktadır. Yöntem-1’den elde edilen 60 kuralın tamamı bahsi geçen 550 kural içerisinde de mevcuttur. Bu kurallardan, kesin kural nitelendirmesi yapılabilenleri Çizelge 5.15’te sunulmuştur. Böylelikle bu grup için toplamda 550 adet güçlü kural elde edilmiştir. Dolayısıyla bu gruptan sonuç kural listesine 550 adet kural dahil edilecektir.

Çizelge 5.15. Elektrik grubundan elde edilen güçlü ve kesin kurallar

Öncül	Ardıl	Örnek Adedi	Destek Oranı	Güven Oranı
Aküler	Elektrik irtibat parçaları	501	19,28	100,00
Aküler & Elektrik tel ve kabloları	Elektrik irtibat parçaları	300	11,54	100,00
Şahsi koruyucu malz. & Aküler	Elektrik irtibat parçaları	250	9,62	100,00
Şahsi koruyucu malz. & Elektrik irtibat parç.	Aküler	250	9,62	100,00
Şahsi koruyucu malz. & Aküler & Elektrik tel ve kabloları	Elektrik irtibat parçaları	204	7,85	100,00
Şahsi koruyucu malz. & Elektrik irtibat parç. & Elektrik tel ve kabloları	Aküler	204	7,85	100,00
Yapışkan maddeler & Aküler	Elektrik irtibat parçaları	146	5,62	100,00
Aküler & Tesbit kulakları, uçları	Elektrik irtibat parçaları	127	4,89	100,00
Ölçü aletleri malz. & Bobinler & Elektrik ızalatörleri	Devre kesiciler	119	4,58	100,00
Ölçü aletleri malz. & Bobinler & Sigortalar ve paratonerler	Devre kesiciler	114	4,39	100,00
Ölçü aletleri malz. & Bobinler & Sigortalar ve paratonerler & Elektrik ızalatörleri	Devre kesiciler	110	4,23	100,00
Ölçü aletleri malz. & Elektrik anahtarları & Sigortalar ve paratonerler	Devre kesiciler	108	4,16	100,00
Ölçü aletleri malz. & Bobinler & Takat ve tevzi malz.	Devre kesiciler	107	4,12	100,00
Elektrik anahtarları & Bobinler & Elektrik ızalatörleri	Devre kesiciler	107	4,12	100,00
Ölçü aletleri malz. & Elektrik anahtarları & Sigortalar ve paratonerler & Elektrik ızalatörleri	Devre kesiciler	107	4,12	100,00

Elektronik arızaları grubu: Bu grupta, 74 malzeme çeşidinin kullanıldığı 1 335 iş emri üzerinden analizler gerçekleştirilmiştir. Bütün veri setinde olduğu gibi, minimum destek değeri %2, minimum güven değeri de %85 olarak seçilmiştir. Apriori algoritmasının yürütülmesi sonucunda elektronik arıza grubuna ilişkin sadece 3 adet birliktelik kuralı elde edilmiştir. Elde edilen kuralların destek değerleri %2,02 ile %2,25 arasında sıralanırken, bu kuralların güven oranları ise %90 ile %96,30 arasında değer almışlardır.

Bu grubun alt veri setinde kullanılan malzeme çeşidinin nispeten az (74) olması ve özellikle iş emri başına düşen ortalama malzeme sayısının (2,78) düşük olması, elektronik arıza grubundan elde edilen kural sayısının az olmasına neden olmuştur. Yöntem-1 dikkate alındığında, bu alt gruptan elde edilen 3 kuralın hiçbiri, bütün veri setinden elde edilen kurallar arasında bulunmamaktadır. Aynı zamanda Yöntem-2'ye göre de bu 3 kuraldan hiçbir tanesi YMDO (%3,76) kriterinden geçebilmiş değildir. Dolayısıyla bu gruptan sonuç kural listesine herhangi bir kural dahil edilememiştir.

Elektronik parçaların elektrik arızaları grubu: Bu grupta, 95 malzeme çeşidinin kullanıldığı 2 571 iş emri üzerinden analizler gerçekleştirilmiştir. Bütün veri setinde olduğu gibi, minimum destek değeri %2, minimum güven değeri de %85 olarak seçilmiştir. Apriori algoritmasının yürütülmesi sonucunda elektronik parçaların elektrik arıza grubuna ilişkin 2 183 adet birliktelik kuralı elde edilmiştir. Elde edilen kuralların destek değerleri %2,02 ile %8,25 arasında sıralanırken, bu kuralların güven oranları ise %85 ile %100 arasında çeşitli değerler almışlardır.

Bu gruptan çıkan sonuçlar, elektrik arıza grubunun sonuçları ile kıyaslandığında nispeten daha az kuralın oluştuğu görülmektedir. İş emri başına düşen ortalama malzeme miktarının düşük olması, bu durumun bir sebebi olarak nitelendirilebilir. Yöntem-1 dikkate alındığında, bu alt gruptan elde edilen 2 183 kuralın 5 tanesi, bütün veri setinden elde edilen kurallar ile kesişmektedir. Çizelge 5.12'de verilen listedeki 21, 57, 60, 114, 117 numaralı kurallar, elektroniğin elektrik arıza grubu veri setinden de elde edilmiştir. Yöntem-2'ye göre ise, bu alt gruptan elde edilen diğer kuralların 78 tanesi YMDO (%3,60) kriterinden geçebilmiştir. Ayrıca, söz konusu 78 kuralın içerisinde Yöntem-1'den elde edilmiş 5 kural da bulunmaktadır. Dolayısıyla 78 adet kural bu gruptan sonuç kural listesine dahil edilebilecektir. Çizelge 5.16'da bu gruptan elde edilen güçlü ve kesin kurallar gösterilmiştir.

Çizelge 5.16. Elektroniğin elektriği grubundan elde edilen güçlü ve kesin kurallar

Öncül	Ardıl	İş Emri Sayısı	Destek Oranı	Güven Oranı
Çeşitli elektrik ve elektronik parç. & Hırdavat malzemesi & Muhtelif kablolar	Mikrofon ve hoparlörler	124	4,82	100
Mikro elektronik devre parç. & Elektrik anahtarları & Mikrofon ve hoparlörler	Transistorlar	102	3,97	100
Düğme ve göstergeler & Mikro elektronik devre parç.	Transistorlar	98	3,81	100
Düğme ve göstergeler & Mikro elektronik devre parç. & Elektrik anahtarları	Transistorlar	97	3,77	100

Kalibre grubu: Bu grupta, 41 malzeme çeşidinin kullanıldığı 247 iş emri üzerinden analizler gerçekleştirilmiştir. Aynı koşullar altında Apriori algoritmasının yürütülmesi sonucunda kalibre grubuna ilişkin 4 adet birliktelik kuralı elde edilmiştir. Elde edilen kuralların destek değerleri %2,83 ile %4,05 arasında sıralanırken, bu kuralların güven oranları ise %88,89 ile %100 arasında çeşitli değerler almışlardır.

Kalibre grubunda daha az malzeme çeşidi ve iş emri kaydı bulunduğundan sadece 4 adet kural oluşturulabilmiştir. Ancak bu 4 kuralın hiçbir tanesi, ne Yöntem-1'e göre bütün veri setinin kuralları ile kesişmiş ne de Yöntem-2'ye göre YMDO kriterine (5,90) uygunluk gösterebilmiştir. Dolayısıyla bu gruptan, sonuç kural listesine hiçbir kural dahil edilemeyecektir.

İmalat grubu: Bu grupta, 129 malzeme çeşidinin kullanıldığı 789 iş emri üzerinden analizler gerçekleştirilmiştir. Yine aynı koşullar altında Apriori algoritmasının yürütülmesi sonucunda imalat grubuna ilişkin 114 226 devasa bir sayıda birliktelik kuralı elde edilmiştir. Elde edilen kuralların destek değerleri %2,03 ile %15,84 arasında sıralanırken, bu kuralların güven oranları ise %85 ile %100 arasında çeşitli değerler almışlardır.

Aslında nispeten düşük sayıda iş emri kaydı olmasına rağmen, imalat grubunun yeterli sayıdaki malzeme çeşidi (129) ve özellikle iş emri başına düşen ortalama malzeme miktarının (4,74) yüksek olması sebebiyle çok sayıda ve güçlü kurallar oluşmuştur. Yöntem-1 dikkate alındığında, bu alt gruptan elde edilen 114 226 kuralın 4 tanesi, ana veri setinden oluşan kurallar ile kesişmektedir. Çizelge 5.12'de verilen listedeki 32, 69, 73, 96 numaralı kurallar, imalat grubu veri setinden de elde edilmiştir. Bununla birlikte Yöntem-2'ye göre de 20 224 adet kural, YMDO (3,67) kriterini aşarak, güçlü kural olarak

seçilmiştir. Yöntem-2'den elde kurallar arasında Yöntem-1'den elde edilen 3 kural da bulunmaktadır. Böylece bu gruptan, sonuç kural listesine 20 225 tane kural dahil edilmiştir. Bu kurallardan 10 tanesi Çizelge 5.17'de verilmiştir.

Çizelge 5.17. İmalat grubundan elde edilen güçlü ve kesin kuralların bir kısmı

Öncül	Ardıl	İş Emri Sayısı	Destek Oranı	Güven Oranı
Kapasitörler	Transistörler	69	8,75	100
Kapasitörler & Dirençler	Transistörler	66	8,37	100
Kapasitörler & Mikro elektronik devre parç.	Transistörler	65	8,24	100
Kapasitörler & Elektrik irtibat parçaları	Transistörler	64	8,11	100
Kapasitörler & Mikro elektronik devre parç. & Dirençler	Transistörler	64	8,11	100
Motorlu vasıta iskeleti imalat parç. & Cıvatalar	Hırdavat malz.	63	7,98	100
Dirençler & Elektrik irtibat parçaları	Transistörler	63	7,98	100
Cıvatalar & Kaynak ve havya malz. & Hırdavat malz.	Tabaka levhalar	63	7,98	100
Elektrik anahtarları & Kapasitörler	Transistörler	62	7,86	100
Motorlu vasıta iskeleti imalat parçaları & Cıvatalar & Tabaka levhalar	Hırdavat malz.	62	7,86	100

Bu kısımda değinilen, Apriori algoritmasının yürütülmesi ve devamında kullanılan yöntemlerle elde edilen birliktelik kural sayılarına ilişkin özet tablo Çizelge 5.18'de sunulmuştur.

5.6.3. Bütün veri setinden elde edilen bulgular ile alt gruplardan elde edilen bulguların karşılaştırılması

Bu çalışmanın ana amacı, bütün veri setinin kullanılmasıyla elde edilen birliktelik kuralları ile veri setinin ilişkili gruplara ayrılmasıyla elde edilen birliktelik kurallarının nicel ve nitel yönden karşılaştırılmasıdır. Analiz sonuçlarının karşılaştırılması, esas olarak veri setinin tamamından çıkan sonuçlar ile alt gruplardan elde edilen sonuçların toplamı arasında yapılacaktır. Bunun yanı sıra bazı noktalarda grupların birbirleriyle mukayesesi de yapılacaktır. Tüm veri setinden elde edilen sonuçlar ile alt gruplardan elde edilen sonuçların karşılaştırılması aşağıda verilen Çizelge 5.18 aracılığıyla yapılacaktır.

Çizelge 5.18. Veri setlerine ait karşılaştırmalı özet bilgiler

Gruplar	Malzeme Çeşidi	Kayıt Sayısı	Ort Malz	Kural Sayısı	Güçlü Kural Sayısı		
					Yöntem-1	Yöntem-2	Toplam
Bütünü	180	7 541	3,61	121	69	-	69
Elektrik	98	2 599	3,98	6 332	60	550	550
Elektronik	74	1 335	2,78	3	0	0	0
Elk.Elkt	95	2 571	3,42	2 183	5	78	78
Kalibre	41	247	2,42	4	0	0	0
İmalat	129	789	4,74	114 226	4	20 224	20 225
Toplam	-	-	-	122 800	69	20 852	20 853

Çizelge 5.18’de verilen tablodaki ilk dört sütun, veri setlerinin içindeki verilerin tasnif edilmesi sonucu ortaya çıkan bilgilerdir. Dolayısıyla bu sütunlar içerisindeki bilgileri anlamlandırmaya yönelik bir analiz ya da karşılaştırma yapılmayacaktır. Bu hususta sadece, tüm veri setinin gruplara ayrılmasının doğal sonucu olarak, gruplardaki mevcut kayıt sayılarının toplamı, tüm veri seti kayıt sayısına (7 541) eşit olacağı söylenebilir. Bunun yanında önceki kısımda anlatılan Yöntem-1’in bütün veri setinden elde edilen kurallarla, alt grup veri setinden elde edilenlerin kesişim kümesini ifade ettiğini, Yöntem-2’nin de alt grup veri setlerine YMDO uygulanarak elde edilmiş kural sayılarını gösterdiğini hatırlatmak yerinde olacaktır.

Mukayeseye öncelikle kural sayısı sütunundan başlanırsa; tüm veri setinden 121, alt grup veri setlerinden ise toplamda 122 748 adet kural meydana geldiği görülmektedir. Aradaki büyük farkın önemli bir sebebi, minimum destek değerinin her iki durumda da aynı değer (%2) seçilmiş olmasıdır. İş emri kayıt sayısı arttıkça, minimum destek kriterini sağlamak ve dolayısıyla kural oluşturmak için gerekli olan kayıt sayısı da fazla olmalıdır. Bütün veri seti için bu kayıt sayısına ulaşamayan, fakat kendi grupları içerisinde bu yeterliliğe ulaşarak birliktelik kuralı oluşturabilen girişler, bu büyük farkı meydana getirmişlerdir.

Kural sayıları arasındaki farkın bir diğer sebebi ise gruplarda bulunan malzeme çeşitliliğidir. Çeşitlilik sayısının ne çok olmasıyla ne de az olmasıyla kural sayısı arasında bir doğrusallık kurulamaz. Yeterli sayıdaki iş emri kayıtlarında kullanılan malzemelerin birbirleriyle ilişkileri arttıkça kural sayısı artacaktır. Ancak bu durum malzeme çeşitliliği sayısıyla ilişkilendirilemez. Bütün veri setinde daha çok çeşit malzeme (180) bulunmasına

rağmen 121 kural oluşmuşken, daha az malzeme çeşitliliğine (129) sahip olan imalat grubundan 114 226 kural ortaya çıkmıştır. Ancak elde edilen kural sayısının malzeme çeşidine bağlı olmadığının en belirgin örneği; elektrik grubuyla elektrik-elektronik grubunun kıyasıyla ortaya çıkar. Benzer çeşit ve kayıt sayısını sahip bu iki grubun birinden 6 332 kural oluşmuşken diğerinden 2 183 kural oluşmuştur.

Ancak, kural sayıları arasındaki bu büyük fark, her bir iş emri kaydı başına düşen ortalama kullanılan malzeme miktarıyla ilişkilendirilebilir. Kayıt sayısının artması tek başına kural sayısını artırmaz. Hatta malzeme kullanımı düşük ortalamada kalırsa, kural sayısının azalmasına sebep olur. Çizelge 5.18'den de görüldüğü üzere ortalama malzeme sütunu değeri fazla olan grupların, çıkardıkları kural sayıları da fazla olmuştur. Öyle ki malzeme ortalama değeri 4,72 olan imalat grubundan 114 226 kural çıkarken, bu değer 2,78 olduğu elektronik grubundan yalnızca 3 kural çıkmıştır. Bu değer artışı, kural sayısının artışında üssel bir etki yaratmaktadır. Bütün veri setinin malzeme ortalaması, grupların ortalamasından oluştuğundan ve doğal olarak bir kısmından düşük olduğundan, üsellik ilişkisinin de etkisiyle kural sayıları arasında büyük bir fark oluşmuştur. Bu noktada, bölüm 5.3.1'de değinilen, Kruskal Wallis ve devamında Mann Whitney U testleriyle kanıtlanan, malzeme kullanımında gruplar arasındaki anlamlı farklılığın, kural sayısına etkisi olup olmadığı hususu menfi olarak değerlendirilecektir. Her ne kadar Çizelge 5.18'de bu hususta müspet bir eğilim mevcut olsa da, Elektronik arıza grubu ile Kalibre grubu arasındaki veriler incelendiğinde böyle bir hipotezin kabul edilemeyeceği anlaşılır. Dolayısıyla ortalama malzeme sayısı ile oluşacak kural sayısı arasında tek başına doğrudan bir ilişki bulunmamaktadır.

Önemle vurgulanmalıdır ki yukarıda değinilen üç faktör de (çeşitlilik, kayıt ve ortalama malzeme sayıları) tek başına bir anlam ifade etmez. Kaldı ki, ortalama malzeme sayısı dışındaki diğer iki faktörün kural sayısı ile ilişkisini doğrusal bir denkleme bağlayamayız. Yani salt bir tanesinin azalmasının ya da artmasının, kural sayısını nasıl etkileyeceği bilinemez. Sonuç olarak, uygun malzeme çeşidi ve kayıt sayısı altında, ortalama kullanılan malzemenin artmasıyla, elde edilen kural sayısı da artacaktır.

Bütün veri seti ile gruplar arasında, hatta grupların kendi aralarında kural sayıları yönünden çok farklı değerlerin mevcudiyetinin söz konusu olduğu ve bunların sebepleri yukarıda özetlenmiştir. Ancak elde edilen bu kuralların hangilerinin bilgi olarak

kullanılacağına, nitel değerlendirmenin sonucunda karar verilecektir. Etkin sonuçlara ulaşmak ve odaklanmayı sağlamak için çıkarımı yapılan 120 bin'den fazla kuralın en güçlü ve en güvenilir olanlarının seçimi yapılmalıdır. Bu noktada öncelikle bütün veri setinden çıkan kurallar ile gruplardan çıkanların kesişiminden faydalanılacaktır.

İlk olarak bütün veri setinden çıkarılan 121 kuralın analizi yapıldığında, bu kurallardan Yöntem-1'e göre 69 tanesinin grupların kuralları ile kesiştiği Çizelge 5.18'de görülmektedir. Böylece hem genel hem de özel veri setlerinden elde edilen söz konusu 69 kural, güçlü kural olarak nitelendirilerek, nihai kural sonuç listesine eklenmelidir. Geri kalan 52 kuralın kesişim kümesinde yer almamasının sebebi, bu kuralların özel gruplarda güven değerini sağlayamamasıdır. Bu durumun sağlanması, Apriori algoritmasının daha düşük güven düzeyinde çalıştırılmasıyla yapılmıştır.

İkinci olarak, gruplardan elde edilen dolayısıyla daha özel ilişkilerden kaynaklanan ve bu itibarla kesişme kümesinde yer almasa da dikkate alınması gereken kuralların mevcut olduğu bir gerçektir. Ancak, bütün veri setinde kullanılan minimum destek oranının burada da kullanılmasıyla, daha önce değinilen sebeplerden ötürü çok sayıda birliktelik kuralı ortaya çıkmıştır. Güçlü olarak nitelendirilen kurallara ulaşmak için bölüm 5.4.1'de ayrıntılı anlatılan, malzemelerin eşit dağılımı mantığına dayanan yükseltilmiş minimum destek oranını (YMDO) kullanılmaktadır. Yöntem-2'nin kullanılmasıyla 120 bin'den fazla olan kural sayısı, Çizelge 5.18'den de görüldüğü üzere 20 852'ye düşmüştür.

Yöntem-1 ve Yöntem-2'nin kriterlerini aşan toplamda 20 853 adet güçlü kural elde edilmiştir. Ancak, kurallara odaklanma açısından bu sayı da oldukça yüksek bir değerdir. Bu aşamada, çok sayıdaki birliktelik kuralından daha nitelikli olanların seçilerek kural sayılarının azaltılması gerekmektedir. Bu kapsamda analizcinin önüne üç farklı seçenek sunulmuştur.

Bu seçeneklerden ilki, kural sayılarını bu haliyle bırakıp, o anki durumun şartları doğrultusunda geniş listeden seçim yapmaktır. Bu seçenek uyumlandırılabilen, ölçülü bir metot olarak değerlendirilebilir. Çünkü, analizci geniş kural listesinden, o an için ilgilendiği malzemelere ilişkin bilgileri kendi tecrübesiyle harmanlayıp karar verme imkanına sahip olacaktır.

İkinci seçenек ise güçlü ve ağırlıklı birliktelik kurallarının tespit edilmesidir. Bu seçenekte, daha önce elde edilen güçlü kurallar korunur ve bu kuralların içinden veritabanında önemli kullanım yüzde oranlarına sahip olan malzemelere ilişkin birliktelik kuralları seçilir. Dolayısıyla kural sayısı azaltılarak, işletme için kritik öneme sahip malzeme gruplarına yönelik analiz yapılır. Ağırlıklı malzemeler, ABC (Pareto) Analizi tekniğı ile tespit edilecektir. Pareto analizi, odaklanması gereken daha önemli konuları bir bütünün içerisinde tespit etmek için kullanılan bir tekniktir. Pareto Analizi, İtalyan ekonomist Wilfredo Pareto tarafından 1897 yılında ortaya atılmıştır. 20-80 ilkesi olarak da bilinen bu çalışmayla, ülke kaynaklarının %80'inin, %20'lik bir azınlığa ait olduğunu ortaya koymuştur. Daha sonraları problem çözme teknikleri içerisinde yer bulmaya başlamış ve sıkça kullanılmıştır. Bu tez için yapılan sarf malzeme kullanım uygulamasında ise toplam kullanılan malzeme miktarının %80'i için, tüm çeşitlerin %17'lik bir kısmının kullanıldığı görülmüştür. Problemin verileri Pareto analizine uygun bir biçimde dağılmıştır. Söz konusu analizde %80'lik kullanımı oluşturan “A Sınıfı” sarf malzemeler Çizelge 5.19'da sunulmuştur. Listenin tamamı Ek-3'te verilmiştir.

Çizelge 5.19. A Sınıfı'na giren sarf malzemeler

S.Nu.	Malzeme İsmi	Kullanılan Adet	Kümülatif Yüzde	Sınıf Yüzdesi	Sınıfı
1	Transistörler	5829	10,15	0,56	A
2	Mikro elektronik devre parçaları	5164	19,15	1,11	A
3	Dirençler	3034	24,43	1,67	A
4	Kapasitörler	2788	29,29	2,22	A
5	Elektrik tel ve kabloları	2426	33,51	2,78	A
6	Muhtelif kablolar	2277	37,48	3,33	A
7	Monte edilmemiş antifrikasyon yatakları	2024	41,01	3,89	A
8	Sigortalar ve paratonerler	2019	44,52	4,44	A
9	Elektrik irtibat parçaları	1913	47,85	5,00	A
10	Hırdavat malzemesi	1443	50,37	5,56	A
11	Elektrik anahtarları	1386	52,78	6,11	A
12	Aküler	1297	55,04	6,67	A
13	Radio lambaları	1273	57,26	7,22	A
14	Bataryalar	1188	59,33	7,78	A
15	Bobinler	1144	61,32	8,33	A
16	Devre kesiciler	942	62,96	8,89	A
17	Tesbit kulakları, uçları	914	64,55	9,44	A
18	Takat ve tevzi malz.	873	66,07	10,00	A
19	Elektrik hırdavat malz.	797	67,46	10,56	A
20	Elektrik izalatörleri	782	68,82	11,11	A
21	Muhabere cihaz malz.	732	70,10	11,67	A
22	Mikrofon ve hoparlörler	729	71,37	12,22	A
23	Boyalar	647	72,50	12,78	A
24	Elektrotlar	640	73,61	13,33	A
25	Roleler, kontaktörler	608	74,67	13,89	A
26	Elektrik kumanda malz.	603	75,72	14,44	A
27	Piezoelektrik kristalleri	600	76,76	15,00	A
28	Tabaka levhalar	586	77,79	15,56	A
29	Ölçü aletleri malz.	564	78,77	16,11	A
30	Güdümlü füze uzaktan kontrol sistem malz.	521	79,67	16,67	A
31	Plastik malz.	513	80,57	17,22	A

Apriori algoritması bütün malzeme listesiyle çalıştırılıp, kurallar elde edildikten sonra Pareto analizinde ilk %17'lik A sınıfına girmiş olan malzemelere ait kurallar işaretlenecektir. 180 çeşit malzemeden 31 çeşit malzeme seçilmiştir. Eğer en baştan A sınıfı belirlenerek, algoritma sadece bu sınıf için çalıştırılırsa, genele ait kriterlere uymayan kurallar da elde edilebilirdi. Genel listeden elde edilen kurallara ABC analizinin uygulanmasıyla bu olumsuzluğun önüne geçilmesi amaçlanmıştır. Bu çalışmada Pareto analizi kullanılarak yapılan ağırlıklandırma sonucunda öncelik verilmesi gereken 3 498 adet birliktelik kuralı kalmıştır. Bu yöntemden elde edilen kurallar, taslak plan hazırladıkları dönemlerde malzeme plancılarına analizlerinde kullanılmaya uygun bilgiler sağlar.

Üçüncü seçenekte daha önce elde edilmiş olan güçlü ve ağırlıklı kurallar korunarak, ilave kriter olarak güven faktörü işleme alınır. Bu yöntemde, elde edilen 3 498 birliktelik kuralından güven oranı %100 olanların seçilmesi işlemi yapılmıştır. Böylece elde kalan kural sayısı toplamda 1 489'a düşmüştür. Bu yöntem, işletme açısından risk alınamayacak seviyedeki maliyetli malzeme alımlarında kullanılmaya uygundur. Hiçbir şekilde risk barındırmadığından diğer faaliyetler (depo düzeni, malzeme planlama, arıza tespiti öncesi malzemenin nakli) için de kullanılmaya uygun bir yöntemdir.

Yukarda anlatılan seçenekler doğrultusunda üç alternatifli nihai sonuç kural listesindeki kural sayıları Çizelge 5.20'de sunulmuştur.

Çizelge 5.20. Alternatif yöntemlerden elde edilen kural sayıları

Gruplar	Klasik Apriori Kural Sayısı	Güçlü Kural Sayısı	Güçlü ve Ağırlıklı Kural Sayısı	Güçlü, Ağırlıklı ve Kesin Kural Sayısı
Bütünü	121	69	36	1
Elektrik	6 332	550	459	10
Elektronik	3	0	0	0
Elk.Elkt.	2 183	78	27	1
Kalibre	4	0	0	0
İmalat	114 226	20 225	3 012	1 478
Toplam	122 800	20 853	3 498	1 489

Analizci, amacına uygun yöntemi belirledikten sonra Çizelge 5.19'da verilen sayıda kuraldan faydalanabilir. Bu tabloda bir sonraki sütun bir önceki sütundan türetilmiştir. “Güçlü Kurallar” “Klasik Apriori”den, “Güçlü ve Ağırlıklı Kurallar” “Güçlü Kurallar”dan “Güçlü, Ağırlıklı ve Kesin Kurallar” da “Güçlü ve Ağırlıklı Kurallar”dan, türetilmiştir. Bu itibarla en fazla kural sayısına sahip seçenek (Klasik Apriori hariç bırakılırsa) doğal olarak güçlü kural alternatifi, en az kurala sahip seçenek ise “Güçlü, Ağırlıklı ve Kesin Kural” seçeneğidir. Bütün veri setinden elde edilen kurallar, diğer alt grupların içinde de bulunduğundan Çizelge 5.19'un toplam satırı, kural sayılarının toplamına eşit değildir.

6. SONUÇ VE ÖNERİLER

Bu çalışmada veri madenciliğinin uygulamalarından biri olan birliktelik analizinin açıklanması ve bir uygulamasının yapılması amaçlanmıştır. Veri madenciliği, anlaşılması ve kullanılması zor olan büyük miktardaki veri yığınlarının işlenerek, anlamlı bilgiler haline getirilmesidir. Birliktelik kuralları ise veri setlerinde bağımsız gibi görünen bilgilerin birbirleriyle bağıntısını ortaya koyar. Günümüzde artık veri madenciliği ve bu alanda kullanılan teknikler çok yaygınlaşmıştır. Neredeyse tüm alanlarda faydalanılan veri madenciliği, özellikle iş sektöründe kaynakların etkin kullanılması açısından kritik öneme sahiptir.

Bu tezin uygulama safhası için kamuya ait bir fabrika seçilmiştir. Çeşitli faaliyetlerin yürütüldüğü fabrikada ağırlıklı olarak bakım ve onarım faaliyetleri yapılmaktadır. Kuruluş bu faaliyetleri; elektrikli cihazların arıza onarımı, elektronik cihazların arıza onarımı, elektronik cihazların elektrik arızası onarımı, kalibrasyon ve imalat olarak beş gruba ayırmıştır. İncelemeye alınan beş yıllık dönemde bu faaliyetler için veriler analize uygun hale getirildiğinde 7 541 adet iş emri açılmış ve 180 çeşit sarf malzeme kullanılmıştır.

Birliktelik kurallarının çıkarımı; yaygın kullanılması, tercih edilen bir program olması, kullanım kolaylığı, farklı kaynaklardan elde edilen veri setlerinin kolay entegre edilebilmesi ve orta ölçek büyüklüğündeki veri setine uygunluğundan dolayı SPSS Modeler programı ile gerçekleştirilmiştir.

Birliktelik kurallarının gerçek ve büyük bir veri setine uygulanmasında karşılaşılan en büyük sorunlardan bir tanesi, çok sayıda kuralın elde edilmesi ve dolayısıyla bu kuralların seçiminin zorlaşmasıdır. Bu soruna ilişkin bir çözüm üretmek amacıyla, bu tez çalışmasında elde edilen kuralların niteliklendirilmesi, sayılarının azaltılması ve tasniflendirilmesi için kuralların analizi yapılmıştır. Bu analizden elde edilen bulgular açıklanmış ve yorumlanmıştır.

Bulgulardan elde ettiğimiz kuralların dışında, kullanılan veri setinin önemine dikkat çekilmiştir. Veri setinin birbiriyle ilişkili gruplar halinde ele alınmasının önemi anlatılmıştır. Alt grup oluşturmada yapılan analizlerden çıkan sonuçların, amaca hizmet

noktasında yeterli olamayacağı belirtilmiştir. Nitekim bütün veri setinden elde edilen 121 kuralın 69 tanesinin kullanım amacına uygun olduğu, geri kalan 52 tanesinin bir takım rastlantı sonucunda olduğu ya da gerçekte istenen güven değerini sağlamadığı gösterilmiştir. Bunun yanında alt grupların veri setlerinden oldukça nitelikli ilave kuralların varlığı tespit edilmiştir. Diğer çalışmalardan farklı olarak, analizler sonucunda elde edilen birliktelik kuralları, ikinci bir analize tabi tutularak niteliklerinin artırılması ve “Güçlü Kural”, “Güçlü ve Ağırlıklı Kural”, “Güçlü, Ağırlıklı ve Kesin Kural” olarak üç farklı kategoride sınıflandırılması sağlanmıştır. İkinci analizle birlikte klasik Apriori algoritması ile elde edilen 122 800 kural ilk önce iki farklı yöntemin uygulanmasıyla 20 853 güçlü kurala, devamında Pareto analizinin uygulanmasıyla 3 498 güçlü ve ağırlıklı kurala ve en sonunda güven kriterinin %100 seçilmesiyle 1 489 güçlü, ağırlıklı ve kesin kurala indirgenmiştir. Böylece işletmenin talepleri doğrultusunda farklı kuralların seçilebilmesi sağlanmıştır.

Elde edilen birliktelik kurallarına ilişkin yapılan bir başka analizin sonucunda, elde edilen kural sayısı çok olsa bile, kuralların ardıllarının belli değerlerde kümелendiği görülmüştür. Örneğin, bütün veri setinden elde edilen 121 kuralın sadece 12 sonuca yöneldiği tespit edilmiştir. Konuya ilişkin çarpıcı bir örnek de “İmalat” alt grubundan elde edilmiş 20 225 adet güçlü kuralın sadece 22 çeşit ardıla ulaşmasıdır. Bu bağlamda üretim planlaması yapılırken ürün ağacı ve malzeme temin süreleri dikkate alınarak ortaya çıkan 22 çeşit ardılın dikkate alınması birçok sorunun oluşmasını engelleyecektir.

Bir diğer farklılık, minimum destek oranının belirlenmesi işleminde yapılmıştır. Genel itibarıyla analizcinin sezgisel olarak seçtiği söz konusu değer için eşit kullanım mantığına dayanan bir formül geliştirilerek, uygulamada bu formül aracılığıyla elde edilen değer kullanılmıştır.

Ayrıca, veri setinden elde edilen verilerle, ortaya çıkan kural sayısı arasındaki ilişki araştırılmıştır. Bu noktada, tek bir veri türüne dayanarak, kural sayısına ilişkin bir yorum yapılamayacağı belirtilmiştir. Öyle ki iş başına düşen ortalama malzeme miktarı dışındaki veri başlıklarının, tek başlarına kural sayısını ne yönde etkileyecekleri dahi yorumlanamamaktadır.

Çalışma sonunda elde edilen bilgiler ile, etkin ve hızlı bir planlama süreci, standart zamanda azalma, uygun depo düzeni, malzeme tedarik maliyetinde azalma ve iş yükü ve iş gücü konularında iyileştirmeler olacaktır.

Diğer yandan kullanılan veri setinin döneminin mümkün olduğunca geniş tutulması, çalışmanın daha net sonuçlar vermesi açısından önemlidir. Ayrıca bu anlamda, birliktelik kurallarının dönemsel olarak incelenmesi farklı sonuçların elde edilmesini sağlayabilir. Yine bu şekilde gerçekleştirilecek araştırmalar, yapılacak bir arıza analizine girdi olarak kullanılabilir. Bu tarz birliktelik kuralı çıkarımı yapılan çalışmalara daha garanti sonuç vermesi için dönem kısıtı dahil edilebilir.

Gerçekleştirilen uygulamada tüm iş emirleri dikkate alınmış ve bunların gruplandırılması suretiyle birliktelik kuralları elde edilmiştir. Ancak veritabanını daha da spesifik hale getirmek mümkündür. Örneğin ürün bazında da birliktelik kurallarına ulaşılabilir. Böylece daha özel kurallar elde edilebilir. Bu durumun dezavantajı ise ürün sayısı kadar algoritmanın çalıştırılmasıdır.

Konuya ilişkin yapılabilecek bir diğer çalışma da Pareto analizi ile minimum destek oranı arasında bir ilişkinin var olup olmadığının tespit edilmesi yönünde olabilir.

Ayrıca, veri muhafazasının artmasıyla birlikte çok fazla kural sayısı ile karşılaşma ihtimali de artmıştır. Bu kurallardan en faydalı olanlara odaklanılmasını sağlayacak yeni yöntemler geliştirilebilir.

KAYNAKLAR

1. Karabatak, M., ve İnce, M .C. (2004, Aralık). *Apriori Algoritması İle Öğrenci Başarısı Analizi*. Eleco Elektrik-Elektronik ve Bilgisayar Mühendisleri Sempozyumunda sunuldu, Bursa.
2. İnternet: Flank, A. (September, 2004). Multirelational association rule mining. Web: [https:// pdfs.semanticscholar.org/ f810/ d62c60afef5dc4ffbca10d14c60ee3b472b4.pdf](https://pdfs.semanticscholar.org/f810/d62c60afef5dc4ffbca10d14c60ee3b472b4.pdf) adresinden 10 Ağustos 2019’da alınmıştır.
3. Witten, I. H., and Frank, E. (2005). *Data mining practical machine learning tools and techniques* (Second Edition). San Francisco, USA: Morgan Kaufmann Publishers.
4. Cavique, L. (2007). A scalable algorithm for the market basket analysis. *Journal of Retailing and Consumer Services*, 14, 400-407.
5. Özçakır, F.C., ve Çamurcu, Y. (2007). Birliktelik kuralı yönetimi İçin bir veri madenciliği yazılımı tasarımı ve uygulaması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(12), 21-37.
6. Chen, Q., He, S.B., Gao, W.W., Sun, Y.K., and Wang, Y.J. (2008, March). *The Research of Multidimensional Association Rule in Traffic Accidents*. Paper presented at the Fourth International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, China.
7. Ay, D., ve Çil, İ. (2008). Migros Türk. a.ş. de birliktelik kurallarının yerleşim düzeni planlamada kullanılması. *Endüstri Mühendisliği Dergisi YA/EM 2008 Özel Sayısı*, 21(2), 14-29.
8. Erdem, S., ve Özdağoğlu, G. (2008). Ege bölgesi’ndeki bir araştırma ve uygulama hastanesinin acil hasta verilerinin veri madenciliği ile analiz edilmesi. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 9(2),261-270.
9. Oktay, H.M. (2010). *Web Kullanım Madenciliğinde Birliktelik Kurallarının Uygulanması*. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
10. Wang, H., and Liu, X. (2011, July). *The Research of Improved Association Rules Mining Apriori Algorithm*. Paper presented at the Eighth International Conference on Fuzzy Systems and Knowledge Discovery, Shanghai, China.
11. Ekim, U. (2011). *Veri Madenciliği Algoritmalarını Kullanarak Öğrenci Verilerinden Birliktelik Kurallarının Çıkarılması*. Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
12. Alaeddinoğlu, M.F. (2012). *Birliktelik Kuralları ile Van Gölü İçin Mekânsal-Zamansal Veri Madenciliği*. Yüksek Lisans Tezi, Atatürk Üniversitesi Fen Bilimleri Enstitüsü, Erzurum.

13. Erpolat, S. (2012). Otomobil yetkili servislerinde birliktelik kurallarının belirlenmesinde Apriori ve FP-Growth algoritmalarının karşılaştırılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 12(1), 37-146.
14. Gosain, A., and Maneela, B. (2013, April). *A Comprehensive Survey of Association Rules on Quantitative Data in Data Mining*. Paper presented at the Proceedings of 2013 IEEE Conference on Information and Communication Technologies, India.
15. Shweta, and Garg, K. (2013). Mining efficient association rules through Apriori algorithm using attributes and comparative analysis of various association rule algorithms. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3, 6.
16. Aytaç, M.B., ve Bilge, H.Ş. (2013). Tele pazarlama verilerinin birliktelik kurallarıyla ve crisp-dm yöntemiyle analiz edilmesi. *Aksaray Üniversitesi İİBF Dergisi*, 5(2), 25-40.
17. Doğan, B., Erol, B., ve Buldu, A. (2014). Sigortacılık sektöründe müşteri ilişkileri yönetimi için birliktelik kurallarının kullanılması. *Marmara Fen Bilimleri Dergisi*, 3, 105-114.
18. Keleş, A.E., ve Kaya, M. (2014, Şubat). *Duvar İnşa Edilmesinde Verimliliği Etkileyen Faktörlerin Apriori Veri Madenciliği Yöntemi Kullanılarak Analizi*. Onaltıncı Akademik Bilişim Konferanslarında sunuldu, Mersin.
19. Doğrul, G., Akay, D., ve Kurt, M. (2015). Trafik kazalarının birliktelik kuralları ile analizi. *Gazi Mühendislik Bilimleri Dergisi*, 1(2), 265-284.
20. Kurt, P.M., ve Duru, N. (2015). Veri madenciliği teknikleri kullanılarak ortaokul öğrencilerinin sosyal ağ kullanım analizi: Kocaeli ili örneği. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 3, 508-517.
21. Chen, C.H., Hong, T.P., Lan, G.C., and Lin, S.B. (2016). Mining fuzzy temporal association rules by item lifespans. *Applied Soft Computing*, 41(C).
22. Kaur, M., and Kang, S. (2016). Market basket analysis, identify the changing trends of market data using association rule mining. *Procedia Computer Science*, 85, 78-85.
23. Yuan, X. (2017). An improved Apriori algorithm for mining association Rules. *American Institute of Physics Conference Proceedings*, 1820.
24. Ateş, Y., (2017). *Nicel Birliktelik Kuralları İçin İlginçlik Ölçütü ve Çoklu Destek Değeri*. Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ.
25. Özçalıcı, M. (2017). Veri madenciliğinde birliktelik kuralları ve ikinci el otomobil piyasası üzerine bir uygulama. *Ordu Üniversitesi Sosyal Bilimler Araştırmaları Dergisi*, 7(1), 45-58.
26. Sağın, A.N. (2018). *Veri Madenciliği Algoritmaları ile Birliktelik Kurallarının Belirlenmesi: Perakende Sektöründe Bir Uygulama*. Yüksek Lisans Tezi, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

27. Karamaşa, Ç. (2018). *Bayramlarda Gerçekleşen Trafik Kazalarının Birliktelik Kuralları ile Analiz Edilmesi*. Doktora Tezi, Eskişehir Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Eskişehir.
28. Dümrek, N.C. (2018). *Hızlı Tüketim Sektöründe Tüketici Davranışlarının Birliktelik Kurallarına Göre İncelenmesi ve Sepet Analizi*. Yüksek Lisans Tezi, T.C. Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
29. Hand, D.J. (1998). Data mining: statistics and more?, *The American Statistician*, 52: 112-118.
30. Jacobs, P. (1999). Data Mining: What general managers need to know, *Harvard Management Update*, 4 (10).
31. Chen, L. and Sycara, K. (1998, May). WebMate: a personal agent for browsing and searching. Paper presentet at the second International Conference on Autonomous Agents, ACM, New York, USA.
32. Kumar, A.V.S., and Wahidabanu, R.S.D. (2008, December). *A frequent item graph approach for discovering frequent itemsets*. Paper presented at the ICACTE '08 International Conference, Phuket, Thailand.
33. Alpaydın, E. (2000). Zeki veri madenciliği: ham veriden altın bilgiye ulaşma yöntemleri. Bilişim 2000 Eğitim Seminerinde sunuldu, İstanbul.
34. Han, J., and Kamber, M. (2001). *Data mining: concepts and techniques*. USA: Morgan Kaufmann Publishers.
35. Döşlü, A. (2008). *Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi*. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
36. Ergün, E. (2008). *Ürün Kategorileri Arasındaki Satış İlişkisinin Birliktelik Kuralları ve Kümeleme Analizi ile Belirlenmesi ve Perakende Sektöründe Bir Uygulama*. Doktora Tezi, Afyon Kocatepe Üniversitesi Sosyal Bilimler Enstitüsü, Afyon.
37. Tantuğ, A. C. (2002). *Veri Madenciliği ve Demetleme*. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
38. Özcan, C. (2014). *Veri Madenciliğinin Güvenlik Uygulama Alanları ve Veri Madenciliği ile Sahtekârlık Analizi*. Yüksek Lisans Tezi, İstanbul Bilgi Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
39. Çelikay, E.K. (2010). *Metin Madenciliği Yöntemiyle Türkçede En Sık Kullanılan ve Birbirini Takip Eden Harflerin Analizi ve Birliktelik Kuralları*. Yüksek Lisans Tezi, Beykent Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
40. Akpınar, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29(1), 1-22.

41. Dinçer, E. (2006). *Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması*. Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli.
42. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1994). *Advances in data mining and knowledge discovery*. USA: MIT Pres.
43. Şen, F. (2008). *Veri Madenciliği ile Birlikte Kurallarının Bulunması*. Yüksek Lisans Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
44. Ganti, V., Gehrke, J., and Ramakrishnan, R. (1999). Mining very large databases. *IEEE Computer* 40, 32(8), 38-45.
45. Berry, M.J.A., and Linoff, G.S. (2004). *Data mining techniques: for marketing, sales and customer relationship management*. New York, USA: John Wiley & Sons Inc.
46. Dolgun, M. Ö. (2006). *Büyük Alışveriş Merkezleri için Veri Madenciliği Uygulamaları*. Yüksek Lisans Tezi, Hacettepe Üniversitesi İstatistik Anabilim Dalı, Ankara.
47. Hui, S., and Jha, G. (2000). Application data mining for customer service support. *Information and Management*, 38, 1-13.
48. Ramkumar, G. D., and Swami, A. (1998). Clustering data without distance functions. *Bulletin of the Technical Committee on Data Engineering*, 21, 9-14.
49. Anand, A. (2003). *A Study and Comparison of Data Clustering Techniques*. Master Thesis, Faculty of The Graduate School of The University of Texas, El Paso, Texas, USA.
50. Şimşek, U.T. (2007). *Veri Madenciliği ve Müşteri İlişkileri Yönetiminde Bir Uygulama*. Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
51. Agrawal, R., and Srikant, R. (1994, September). *Fast algorithms for mining association rules*. Paper presented at the Twentieth VLDB Conference, Santiago, Chile.
52. Agrawal, R., Imielinski, T., and Swami, A. (1993, May). *Mining association rules between sets of items in large databases*. Paper presented at the In Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, USA.
53. Tang, P.N., Steinbach, M., and Kumar, V. (2006). *Introduction to data mining*. USA: Addison Wesley Pres.
54. Houtsma, M.A.W., and Swami, A. (1995, March). *Set-Oriented mining for association rules in relational databases*. Paper presented at the Eleventh International Conference on Data Engineering, Taipei, Taiwan.
55. Gürgen, G. (2008). *Birlikte Kuralları İle Sepet Analizi ve Uygulanması*. Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

56. Silahtaroglu, G. (2008). *Kavram ve algoritmalarıyla veri madenciliği* (Birinci Basım). İstanbul: Papatya Yayınları.
57. Dunham, M. H., Xiao, Y., Gruenwald L., and Hossain, Z. (2000). *A survey of association rules*. Southern Methodist University, Department of Computer Science.

EKLER

EK-1. (devam) Uygulamada kullanılan veri seti şablonu ve malzeme listesi

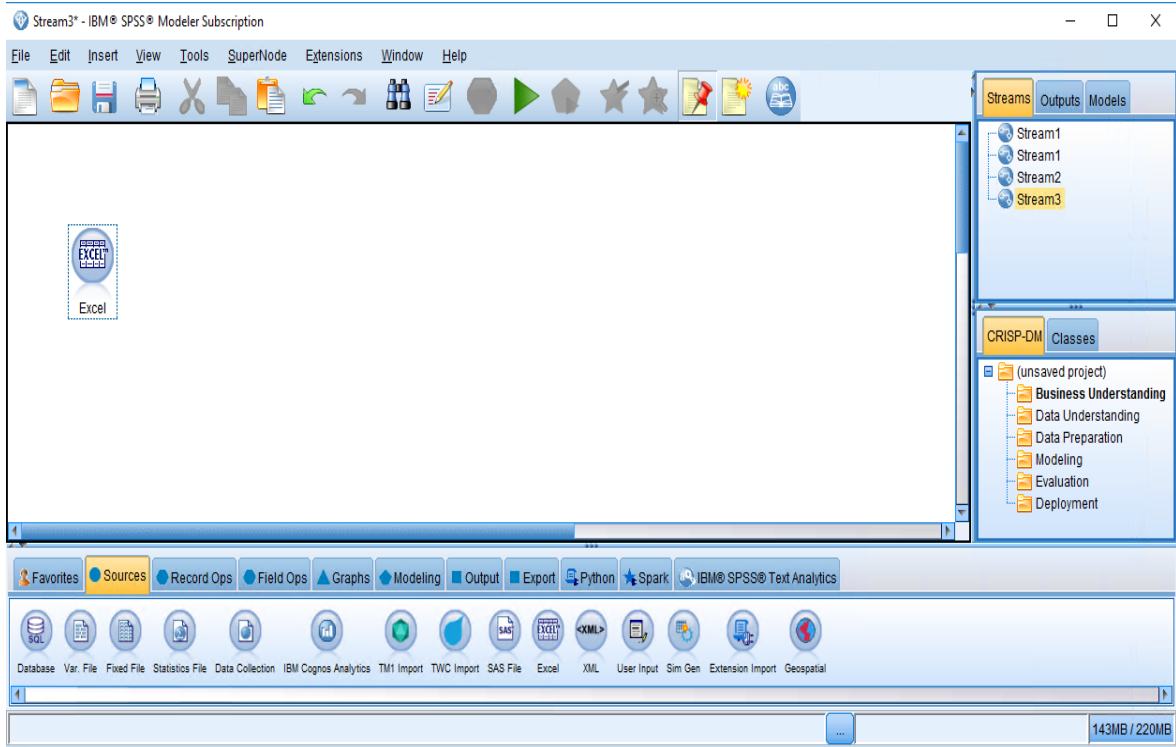
1	Transistörler	50	Kimyevi maddeler
2	Mikro elektronik devre parçaları	51	Demir çubuklar
3	Dirençler	52	Tel yaylar
4	Kapasitörler	53	Filtreler ve şebekeler
5	Elektrik tel ve kabloları	54	Teksir malz.
6	Muhtelif kablolar	55	Yanmaz boyalar
7	Monte edilmemiş antifikasyon yatakları	56	Köşebentler
8	Sigortalar ve paratonerler	57	Elektrik jeneratörleri
9	Elektrik irtibat parçaları	58	Yer radar malz.
10	Hırdavat malzemesi	59	Eğitim malz.
11	Elektrik anahtarları	60	Mantarlar
12	Aküler	61	Ofis malz.
13	Radıo lambaları	62	Tapalar ve kapsüller
14	Bataryalar	63	Vantilatörler ve hava devrettiren cihazlar
15	Bobinler	64	Yüzük
16	Devre kesiciler	65	Kimya maddeleri
17	Tesbit kulakları, uçları	66	Madeni borular
18	Takat ve tevzi malz.	67	Araç elektrik aydınlatma malz.
19	Elektrik hırdavat malz.	68	Motorlu vasıta iskeleti imalat parçaları
20	Elektrik ızalatörleri	69	Motora ait hava ve yağ filtreleri
21	Muhabere cihaz malz.	70	Matkap malz.
22	Mikrofon ve hoparlörler	71	Kontaplak ve kaplamalar
23	Boyalar	72	Bilgi işlem malz.
24	Elektrotlar	73	Ahşap malzemeler
25	Roleler, kontaktörler	74	Tahtalar
26	Elektrik kumanda malz.	75	Elektrik motorları
27	Piezoelektrik kristalleri	76	Bükülebilir hortum ve borular
28	Tabaka levhalar	77	Dahili muhabere ilan sistem malz.
29	Ölçü aletleri malz.	78	Bombalar ve parçaları
30	Güdümlü füze uzaktan kont. sistem malz.	79	Aşındırıcı malzemeler
31	Plastik malz.	80	Elektronik modüller
32	Vidalar	81	Laboratuar malz.
33	Yapışkan maddeler	82	Güdümlü füze malz.
34	Çubuk ve barlar	83	Fren, direksiyon, aks parçaları
35	Çeşitli elektrik ve elektronik parçalar	84	Ofis aletleri
36	Düğme ve göstergeler	85	Tahrip malzemeleri
37	Elektrik redresör teçhizatı	86	Monte edilmemiş düz yatakları
38	Radıo ve tv malzemeleri	87	Yağlar
39	Paketleme ve conta malzemeleri	88	Macunlar
40	Cıvatalar	89	Fotoğraf malz.
41	Antenler	90	Kayışlar
42	Elektrik lambaları	91	Güdümlü füze lançerleri
43	Tesbit malzemeleri	92	Tehlike arama cihaz malz.
44	İnşaat profilleri	93	Elektrik konvertörleri
45	Çiviler, kama ve pimler	94	Muhtelif takat transmisyon malz.
46	Kaynak ve havıya malzemeleri	95	Zincir ve tel halatlar
47	Elektrik aydınlatma malz.	96	Motorlu ve el pompaları
48	Şahsi koruyucu malz.	97	Optik nişangahlar
49	Somun ve rondelalar	98	Variller, tenekeler

EK-1. (devam) Uygulamada kullanılan veri seti şablonu ve malzeme listesi

99	Radyo seyrüsefer malz.	148	Uçak yer hizmet malz.
100	Seyrüsefer malz.	149	Bilgisayar aksesuarları
101	Gece görüş malz.	150	Süpürgeler
102	Sıkıştırılmış gazlar	151	Uçak alarm malz.
103	Kesici el malz.	152	Bilgisayar cihazları
104	Motor elektrik sistemi parçaları	153	Kurutucular
105	Takat santral transformatörleri	154	Pistonlu benzin motor parçaları
106	Gemi alarm sis.	155	Torna tezgah malz.
107	Özel elbiseler	156	Atış kumanda radar malz.
108	Tayyare radar malz.	157	Bezler
109	Otomatik pilot mekanizması malz.	158	El avadanlık malz.
110	Radyasyon temizleyici malz.	159	Tezgah aksesuarları
111	Rutubet ölçüm malz.	160	Muhtelif inşaat malz.
112	Ölçü aletleri malzemeleri	161	Sanayi kazanları
113	Ambalaj	162	Diğer malz.
114	Kırtasiye malz.	163	Kauçuk malz.
115	Güdümlü füze aksamaları	164	Mintikavi ısıtma cihazları
116	Sandık	165	Müteferrik atış kumanda malz.
117	Sinkrolar ve rezolverler	166	Silah malz.
118	Kompresör ve vakum pompaları	167	Soğutma ve havalandırma tesis parçaları
119	Metal kafesler	168	Yağ ve akaryakıt dağıtım malz.
120	Profiller	169	Yangın söndürme malz.
121	Tel	170	Akaryakıt sistemi parçaları
122	Zaman ölçü aletleri	171	Atış kumanda hesaplayıcı nişangahlar
123	Dişliler, kasnaklar	172	Cam
124	Dizel motor parçaları	173	Cam çubuklar
125	Hortum ve boruların rekor ve tertibatları	174	Deriler
126	Telefon ve telgraf malzemeleri	175	Duvar tahtaları
127	Sübaplar	176	Gaz tüpleri
128	Çeşitli alarm malz.	177	Motor aksesuarları
129	Lamba tutucular	178	Pafta ve klavuzlar
130	Monte edilmiş düz yatakları	179	Saplamalar
131	Motorlu vasıtaların kuvvet nakli parçaları	180	Torbalar
132	Muhtelif uçak aksesuarları		
133	Müteferrik silah malz.		
134	Aşındırıcı disk ve taşlar		
135	Kablo bağlantı parçaları		
136	Tayyare radyo seyrüsefer malz.		
137	Gaz ölçüm malz.		
138	Kesici olmayan el malz.		
139	Uçuş aletleri malz.		
140	Devir ve sürat değiştiriciler		
141	Motor soğutma sistemi parçaları		
142	Motorlu vasıta tamir malz.		
143	Portatif elektrik aydınlatma malz.		
144	Separötörler ve vakum filtreleri		
145	Ses kayıt cihaz malz.		
146	Takat kaynaksız çalışan sübaplar		
147	Uçak makinalı tüfek atış kumanda tertibatı		

EK-2. SPSS Modeler programının kullanımı ve veri seti üzerinde çalıştırılması

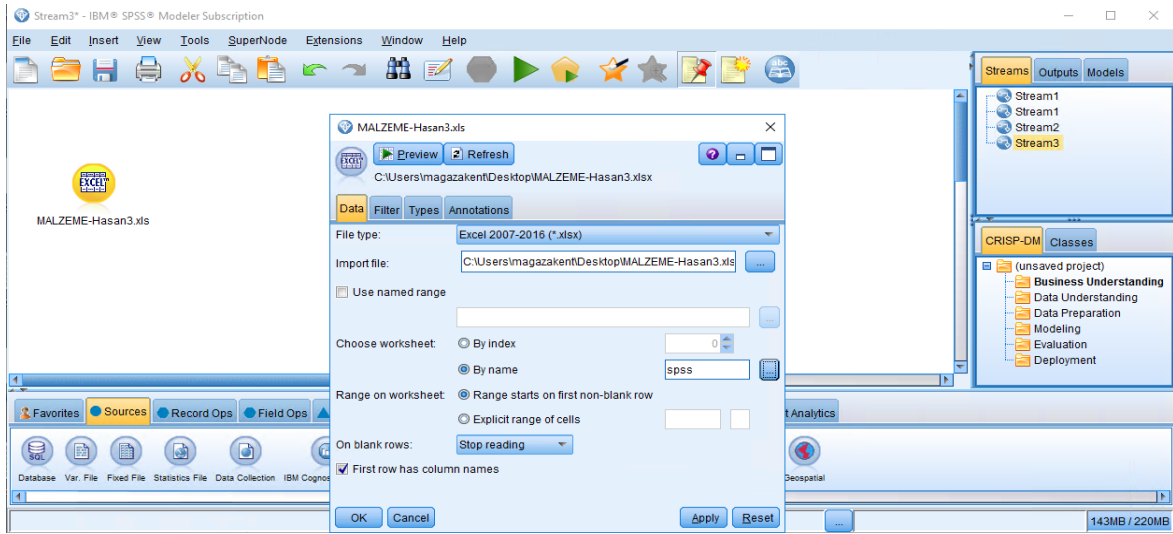
SPSS Modeler programı başlatıldıktan sonra boş bir ara yüz, “Stream” ana ekranı açılır. İlk olarak, aşağıda bulunan “Sources” sekmesinden kullanılacak veri seti hangi formatta ise ana ekrana, gerekli format sürüklenir. Bu çalışmada kullanılan veri seti Excel formatında olduğundan burada “Excel” formatı seçilmiştir. İlgili işlem Resim 2.1’de görülmektedir.



Resim 2.1. Veri seti formatının seçimi

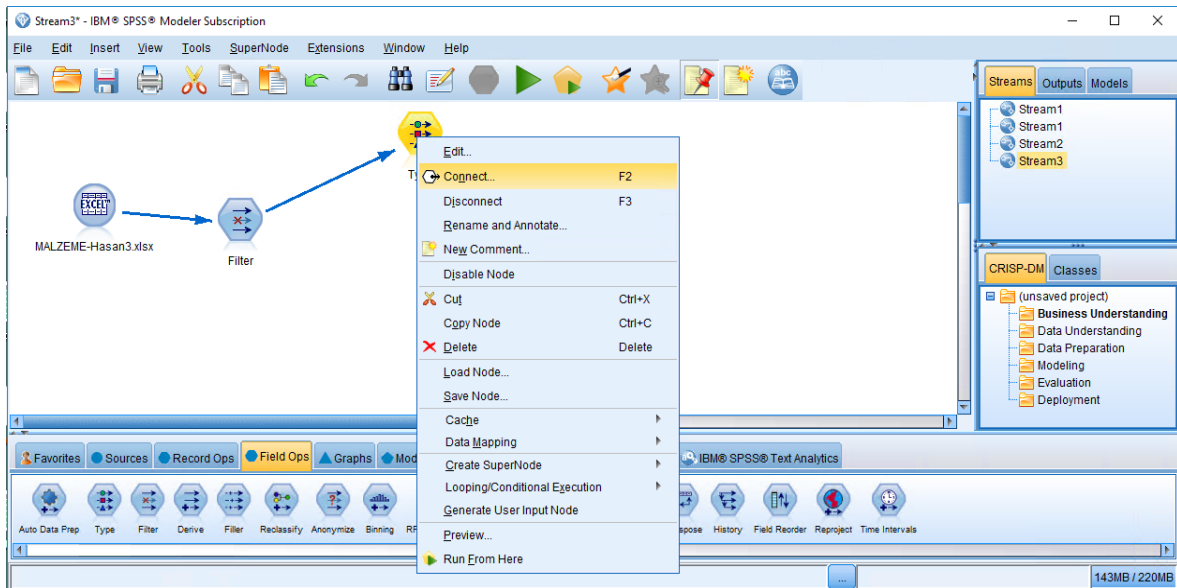
İkinci aşamada “Excel” formatına sağ tıklanır. “Edit” tıklanır ve buradaki “Import File” kısmından kullanılacak veri seti seçilip “OK” alanına basılarak veri seti modeler programına tanımlanır. İlgili işlem Resim 2.2’de görülmektedir.

EK-2. (devam) SPSS Modeler programının kullanımı ve veri seti üzerinde çalıştırılması



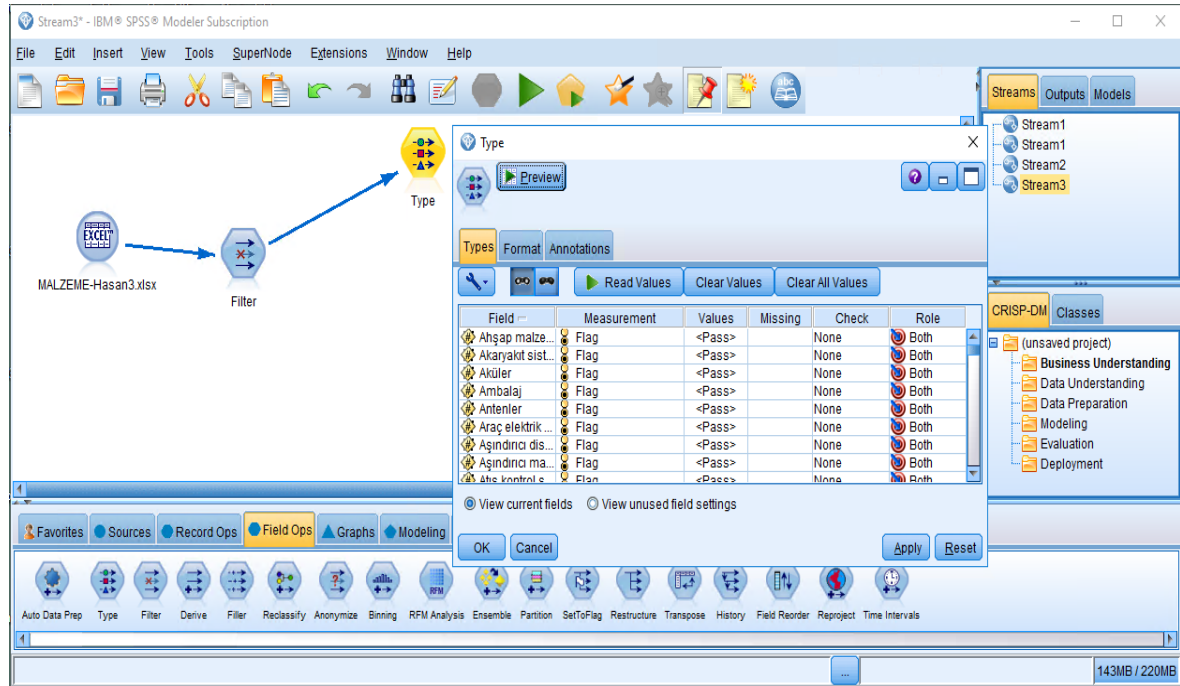
Resim 2.2. Veri setini tanımlama

Üçüncü aşamada, “Field Ops” kısmından “Filter” ve “Type” ikonları ana ekrana taşınır. Modeli oluşturabilmek için ana ekrana taşınan bu üç fonksiyondan, veri ikonu “Filter” ikonuna, “Filter” ikonu da “Type” ikonuna her birinin üzerine sağ tıklanıp, “Connect” kısmına basılarak birbirleri ile bağlantıları kurulur. Bağlantı kurulurken “Connect” özelliği bir sonraki aşamada hangi özellik ile bağlantı kurulacaksa o özellik seçilir, bu sayede en sonuncu bağlantıda her özellik birbiri ile bağlanmış olacaktır. İlgili işlem Resim 2.3’te görülmektedir.



Resim 2.3. Filter ve type fonksiyonlarını ekleme ve bağlantı kurma

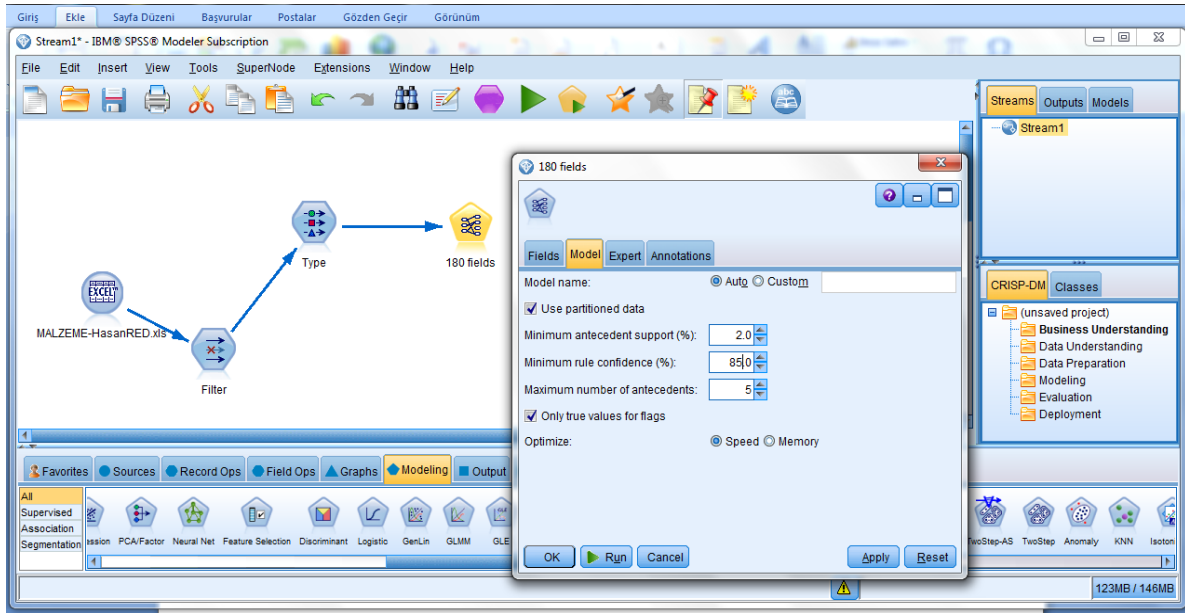
EK-2. (devam) SPSS Modeler programının kullanımı ve veri seti üzerinde çalıştırılması



Resim 2.5. Type fonksiyonu

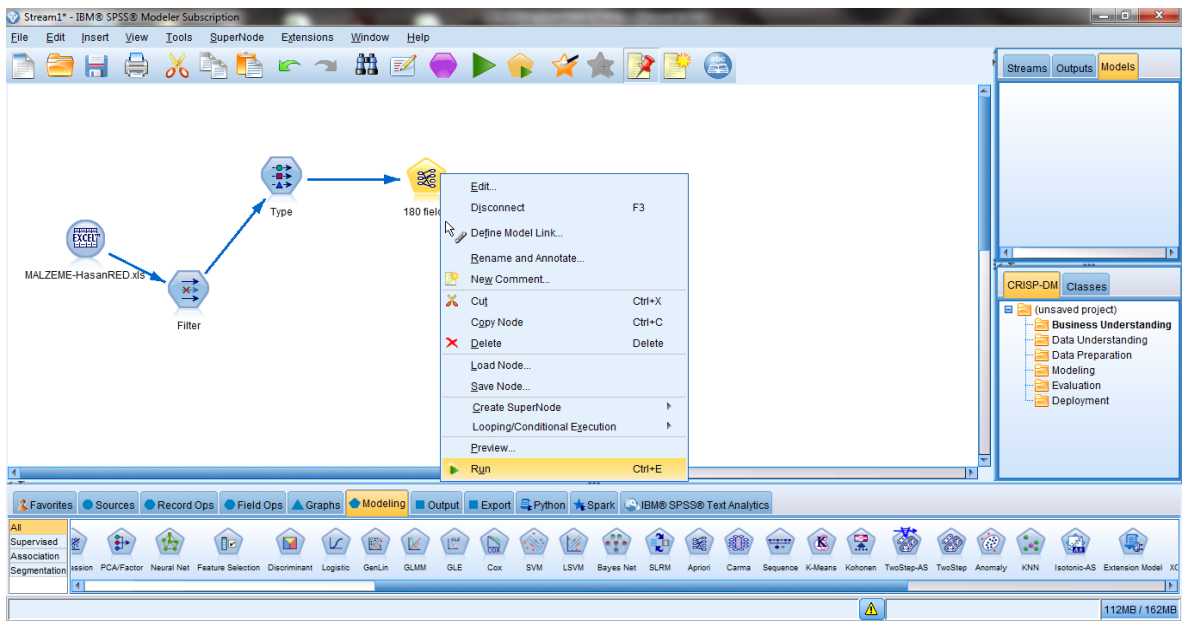
Son aşamada modelin kurulması tamamlanıp çalıştırılmaya hazır hale getirilecektir. Öncelikle, alt kısımda bulunan “Modelling” sekmesinden “Apriori” modelinin ikonu sürüklenerek ekrana taşınır ve “Type” özelliği ile bağlantısı kurulur. “Apriori” modelinde kaç malzeme çeşidi ile işlem yapılıyorsa, isim olarak ekranda otomatik belirecektir. Daha sonra, ekrandaki “Apriori” modeline tıklanıp özellik ekranının altındaki “Model” alanına tıklanarak oradaki belirli kısımlara minimum destek ve minimum güven değerleri ve maksimum öncül sayısı değeri girilir. Öncül sayısı analizin ayrıntısı göz önünde bulundurularak kullanıcı tarafından belirlenir. Bu uygulamada, önceki kısımda bahsedilen sebeplerden dolayı, minimum destek değeri %2, minimum güven değeri ise %85 olarak belirlenmiştir. Her hangi bir kuralın gözden kaçırılmaması için maksimum öncül sayısı default değer de olan, 5 olarak belirlendikten sonra “OK” alanına basılarak model çalıştırma aşamasına geçilmiş olur. İlgili işlem Resim 2.6’da görülmektedir.

EK-2. (devam) SPSS Modeler programının kullanımı ve veri seti üzerinde çalıştırılması



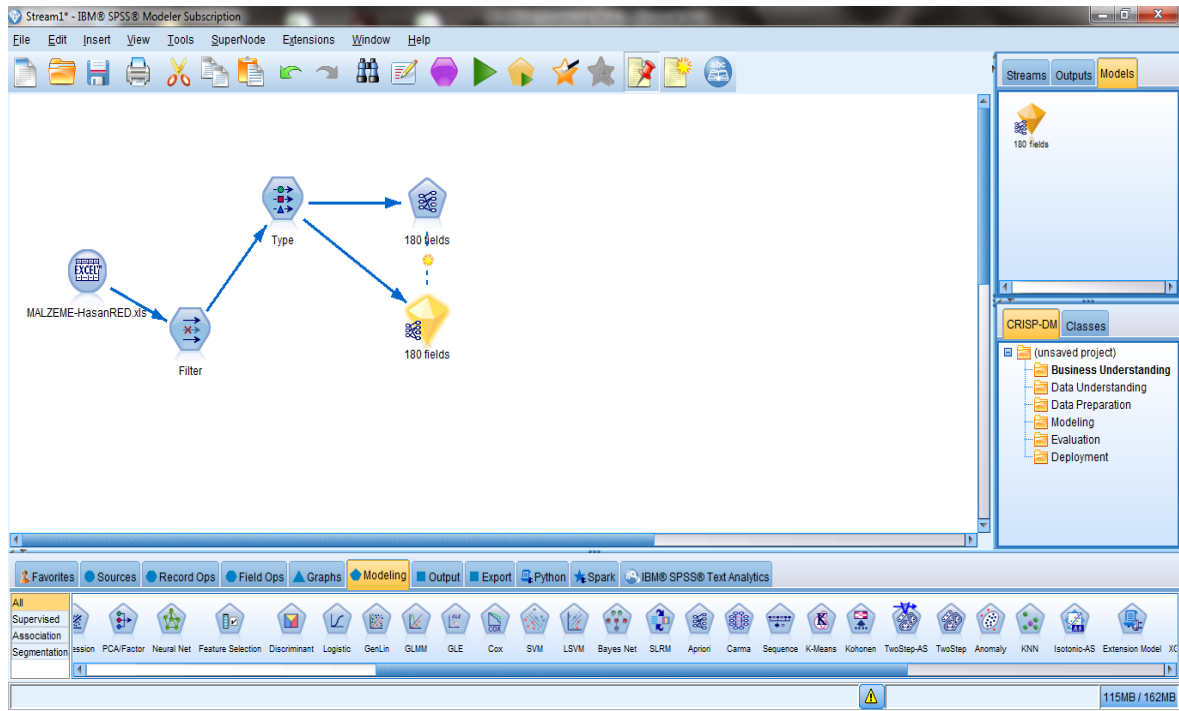
Resim 2.6. Apriori modeli görüntüsü

Son olarak, Resim 2.7’de bulunan “Apriori” modeline sağ tıklanarak “Execute” ya da yukarıda bulunan araç çubuğundaki “Run” alanlarına basılarak analiz gerçekleştirilir. Daha sonra üst kısımda bulunan “Streams” bölümünde analizin sonuçları ortaya çıkacaktır. Ortaya çıkan ikon ana ekrana sürüklenerek ve üzerine tıklanarak analizin sonuçları ayrıntılı olarak incelenebilir. İlgili işlem Resim 2.8’de görülmektedir.



Resim 2.7. Modeli çalıştırma

EK-2. (devam) SPSS Modeler programının kullanımı ve veri seti üzerinde çalıştırılması



Resim 2.8. Analiz sonuçlarının ara yüze aktarılışı

Resim 2.8'deki elmas görüntüsüne tıklanarak hesaplanan birliktelik kuralları ve bu kuralların istatistikî tanımlayıcı bilgilerine ulaşılır.

EK-3. Sarf malzeme kullanımına ait pareto analizinin tüm listesi

S.Nu.	Malzeme İsmi	Kullanılan Adet	Kümülatif Yüzde	Sınıf Yüzdesi	Sınıfı
1	Transistörler	5829	10,15	0,56	A
2	Mikro elektronik devre parçaları	5164	19,15	1,11	A
3	Dirençler	3034	24,43	1,67	A
4	Kapasitörler	2788	29,29	2,22	A
5	Elektrik tel ve kabloları	2426	33,51	2,78	A
6	Muhtelif kablolar	2277	37,48	3,33	A
7	Monte edilmemiş antifriksiyon yatakları	2024	41,01	3,89	A
8	Sigortalar ve paratonerler	2019	44,52	4,44	A
9	Elektrik irtibat parçaları	1913	47,85	5,00	A
10	Hırdavat malzemesi	1443	50,37	5,56	A
11	Elektrik anahtarları	1386	52,78	6,11	A
12	Aküler	1297	55,04	6,67	A
13	Radyo lambaları	1273	57,26	7,22	A
14	Bataryalar	1188	59,33	7,78	A
15	Bobinler	1144	61,32	8,33	A
16	Devre kesiciler	942	62,96	8,89	A
17	Tesbit kulakları, uçları	914	64,55	9,44	A
18	Takat ve tevzi malz.	873	66,07	10,00	A
19	Elektrik hırdavat malz.	797	67,46	10,56	A
20	Elektrik izalatörleri	782	68,82	11,11	A
21	Muhabere cihaz malz.	732	70,10	11,67	A
22	Mikrofon ve hoparlörler	729	71,37	12,22	A
23	Boyalar	647	72,50	12,78	A
24	Elektrotlar	640	73,61	13,33	A
25	Roleler, kontaktörler	608	74,67	13,89	A
26	Elektrik kumanda malz.	603	75,72	14,44	A
27	Piezoelektrik kristalleri	600	76,76	15,00	A
28	Tabaka levhalar	586	77,79	15,56	A
29	Ölçü aletleri malz.	564	78,77	16,11	A
30	Güdümlü füze uzaktan kontrol sistem malz.	521	79,67	16,67	A
31	Plastik malz.	513	80,57	17,22	A
32	Vidalar	486	81,42	17,78	B
33	Yapışkan maddeler	479	82,25	18,33	B
34	Çubuk ve barlar	445	83,02	18,89	B
35	Çeşitli elektrik ve elektronik parçalar	409	83,74	19,44	B
36	Düğme ve göstergeler	406	84,44	20,00	B
37	Elektrik redresör teçhizatı	372	85,09	20,56	B
38	Radyo ve tv malzemeleri	343	85,69	21,11	B
39	Paketleme ve conta malzemeleri	334	86,27	21,67	B
40	Cıvatalar	330	86,85	22,22	B
41	Antenler	329	87,42	22,78	B
42	Elektrik lambaları	321	87,98	23,33	B
43	Tesbit malzemeleri	312	88,52	23,89	B
44	İnşaat profilleri	300	89,04	24,44	B
45	Çiviler, kama ve pimler	297	89,56	25,00	B
46	Kaynak ve havya malzemeleri	282	90,05	25,56	B
47	Elektrik aydınlatma malz.	281	90,54	26,11	B

EK-3. (devam) Sarf malzeme kullanımına ait pareto analizinin tüm listesi

48	Şahsi koruyucu malz.	252	90,98	26,67	B
49	Somun ve rondelalar	244	91,41	27,22	B
50	Kimyevi maddeler	204	91,76	27,78	B
51	Demir çubuklar	202	92,11	28,33	B
52	Tel yaylar	195	92,45	28,89	B
53	Filtreler ve şebekeler	194	92,79	29,44	B
54	Teksir malz.	189	93,12	30,00	B
55	Yanmaz boyalar	155	93,39	30,56	B
56	Köşebentler	150	93,65	31,11	B
57	Elektrik jeneratörleri	141	93,90	31,67	B
58	Yer radar malz.	130	94,12	32,22	B
59	Eğitim malz.	110	94,31	32,78	B
60	Mantarlar	109	94,50	33,33	B
61	Ofis malz.	109	94,69	33,89	B
62	Tapalar ve kapsüller	101	94,87	34,44	B
63	Vantilatörler ve hava devrettiren cihazlar	99	95,04	35,00	B
64	Yüzük	99	95,22	35,56	C
65	Kimya maddeleri	90	95,37	36,11	C
66	Madeni borular	90	95,53	36,67	C
67	Araç elektrik aydınlatma malz.	85	95,68	37,22	C
68	Motorlu vasıta iskeleti imalat parçaları	84	95,82	37,78	C
69	Motora ait hava ve yağ filtreleri	79	95,96	38,33	C
70	Matkap malz.	78	96,10	38,89	C
71	Kontaplak ve kaplamalar	75	96,23	39,44	C
72	Bilgi işlem malz.	74	96,36	40,00	C
73	Ahşap malzemeler	73	96,48	40,56	C
74	Tahtalar	73	96,61	41,11	C
75	Elektrik motorları	72	96,74	41,67	C
76	Bükülebilir hortum ve borular	68	96,85	42,22	C
77	Dahili muhabere ilan sistem malz.	64	96,97	42,78	C
78	Bombalar ve parçaları	61	97,07	43,33	C
79	Aşındırıcı malzemeler	58	97,17	43,89	C
80	Elektronik modüller	51	97,26	44,44	C
81	Laboratuar malz.	49	97,35	45,00	C
82	Güdümlü füze malz.	46	97,43	45,56	C
83	Motorlu vasıtaların fren, direksiyon, aks parçaları	44	97,50	46,11	C
84	Ofis aletleri	44	97,58	46,67	C
85	Tahrip malzemeleri	44	97,66	47,22	C
86	Monte edilmemiş düz yatakları	42	97,73	47,78	C
87	Yağlar	42	97,80	48,33	C
88	Macunlar	41	97,88	48,89	C
89	Fotoğraf malz.	39	97,94	49,44	C
90	Kayışlar	39	98,01	50,00	C
91	Güdümlü füze lançerleri	36	98,07	50,56	C
92	Tehlike arama cihaz malz.	35	98,13	51,11	C
93	Elektrik konvertörleri	33	98,19	51,67	C
94	Muhtelif takat transmisyon malz.	32	98,25	52,22	C
95	Zincir ve tel halatlar	32	98,30	52,78	C
96	Motorlu ve el pompaları	31	98,36	53,33	C

EK-3. (devam) Sarf malzeme kullanımına ait pareto analizinin tüm listesi

97	Optik nişangahlar	31	98,41	53,89	C
98	Variller, tenekeler	31	98,47	54,44	C
99	Radyo seyrüsefer malz.	29	98,52	55,00	C
100	Seyrüsefer malz.	29	98,57	55,56	C
101	Gece görüş malz.	28	98,62	56,11	C
102	Sıkıştırılmış gazlar	28	98,66	56,67	C
103	Kesici el malz.	27	98,71	57,22	C
104	Motor elektrik sistemi parçaları	27	98,76	57,78	C
105	Takat santral transformatörleri	27	98,81	58,33	C
106	Gemi alarm sis.	26	98,85	58,89	C
107	Özel elbiseler	26	98,90	59,44	C
108	Tayyare radar malz.	26	98,94	60,00	C
109	Otomatik pilot mekanizması malz.	25	98,98	60,56	C
110	Radyosyon temizleyici malz.	25	99,03	61,11	C
111	Rutubet ölçüm malz.	24	99,07	61,67	C
112	Ölçü aletleri malzemeleri	22	99,11	62,22	C
113	Ambalaj	21	99,14	62,78	C
114	Kırtasiye malz.	21	99,18	63,33	C
115	Güdümlü füze aksamaları	20	99,22	63,89	C
116	Sandık	19	99,25	64,44	C
117	Sinkrolar ve rezolverler	18	99,28	65,00	C
118	Kompresör ve vakum pompaları	16	99,31	65,56	C
119	Metal kafesler	16	99,34	66,11	C
120	Profiller	16	99,36	66,67	C
121	Tel	16	99,39	67,22	C
122	Zaman ölçü aletleri	16	99,42	67,78	C
123	Dişliler, kasnaklar	14	99,44	68,33	C
124	Dizel motor parçaları	14	99,47	68,89	C
125	Hortum ve boruların rekor ve tertibatları	14	99,49	69,44	C
126	Telefon ve telgraf malzemeleri	14	99,52	70,00	C
127	Sübaplar	13	99,54	70,56	C
128	Çeşitli alarm malz.	12	99,56	71,11	C
129	Lamba tutucular	11	99,58	71,67	C
130	Monte edilmiş düz yatakları	10	99,60	72,22	C
131	Motorlu vasıtaların kuvvet nakli parçaları	10	99,62	72,78	C
132	Muhtelif uçak aksesuarları	10	99,63	73,33	C
133	Müteferrik silah malz.	10	99,65	73,89	C
134	Aşındırıcı disk ve taşlar	9	99,67	74,44	C
135	Kablo bağlantı parçaları	9	99,68	75,00	C
136	Tayyare radyo seyrüsefer malz.	9	99,70	75,56	C
137	Gaz ölçüm malz.	8	99,71	76,11	C
138	Kesici olmayan el malz.	8	99,72	76,67	C
139	Uçuş aletleri malz.	8	99,74	77,22	C
140	Devir ve sürat değiştiriciler	7	99,75	77,78	C
141	Motor soğutma sistemi parçaları	7	99,76	78,33	C
142	Motorlu vasıta tamir malz.	7	99,78	78,89	C
143	Portatif elektrik aydınlatma malz.	7	99,79	79,44	C
144	Separötörler ve vakum filtreleri	7	99,80	80,00	C
145	Ses kayıt cihaz malz.	7	99,81	80,56	C

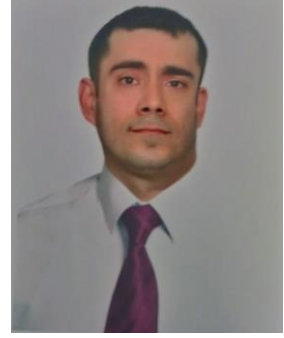
EK-3. (devam) Sarf malzeme kullanımına ait pareto analizinin tüm listesi

146	Takat kaynaklı çalışan sübaplar	7	99,82	81,11	C
147	Uçak makinalı tüfek atış kumanda tertibatı	7	99,84	81,67	C
148	Uçak yer hizmet malz.	7	99,85	82,22	C
149	Bilgisayar aksesuarları	6	99,86	82,78	C
150	Süpürgeler	6	99,87	83,33	C
151	Uçak alarm malz.	6	99,88	83,89	C
152	Bilgisayar cihazları	5	99,89	84,44	C
153	Kurutucular	5	99,90	85,00	C
154	Pistonlu benzin motor parçaları	5	99,91	85,56	C
155	Torna tezgah malz.	5	99,91	86,11	C
156	Atış kumanda radar malz.	4	99,92	86,67	C
157	Bezler	4	99,93	87,22	C
158	El avadanlık malz.	4	99,94	87,78	C
159	Tezgah aksesuarları	4	99,94	88,33	C
160	Muhtelif inşaat malz.	3	99,95	88,89	C
161	Sanayi kazanları	3	99,95	89,44	C
162	Diğer malz.	2	99,96	90,00	C
163	Kauçuk malz.	2	99,96	90,56	C
164	Mıntıkavi ısıtma cihazları	2	99,96	91,11	C
165	Müteferrik atış kumanda malz.	2	99,97	91,67	C
166	Silah malz.	2	99,97	92,22	C
167	Soğutma ve havalandırma tesis parçaları	2	99,97	92,78	C
168	Yağ ve akaryakıt dağıtım malz.	2	99,98	93,33	C
169	Yangın söndürme malz.	2	99,98	93,89	C
170	Akaryakıt sistemi parçaları	1	99,98	94,44	C
171	Atış kumanda hesaplayıcı nişangahlar	1	99,98	95,00	C
172	Cam	1	99,99	95,56	C
173	Cam çubuklar	1	99,99	96,11	C
174	Deriler	1	99,99	96,67	C
175	Duvar tahtaları	1	99,99	97,22	C
176	Gaz tüpleri	1	99,99	97,78	C
177	Motor aksesuarları	1	99,99	98,33	C
178	Pafta ve klavuzlar	1	100,00	98,89	C
179	Saplamalar	1	100,00	99,44	C
180	Torbalar	1	100,00	100,00	C

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : TONTUŞ, Hasan
 Uyuğu : T.C.
 Doğum tarihi ve yeri : 22.05.1980, Mersin
 Medeni hali : Evli
 Telefon : 0 (505) 893 23 19
 e-mail : hasantontus@gmail.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / Endüstri Mühendisliği	Devam ediyor
Lisans	Gazi Üniversitesi / Endüstri Mühendisliği	2003
Lise	Elbistan Anadolu Lisesi	1998

İş Deneyimi

Yıl	Yer	Görev
2005-Halen	3. HBF.Md.lüğü	Endüstri Mühendisi

Yabancı Dil

İngilizce

Yayınlar

Tontuş, H. ve Güner, E. (2019, 12-14 Haziran). Sarf malzeme kullanımına yönelik elde edilen birliktelik kurallarının analizi ve sınıflandırılması. 39. Yöneylem Araştırması ve Endüstri Mühendisliği Ulusal Kongresi'nde sunuldu, Ankara.

Hobiler

Ülkeler ve dinler tarihi



GAZİ GELECEKTİR