



**GÜNCEL İŞ KAZALARI HAM VERİLERİNİN, VERİ MADENCİLİĞİ
SÜRECİ İLE ANALİZİ**

Hasan ÇAKMAK

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

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EKİM 2019

Hasan ÇAKMAK tarafından hazırlanan “GÜNCEL İŞ KAZALARI HAM VERİLERİNİN, VERİ MADENCİLİĞİ SÜRECİ İLE ANALİZİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Endüstri Mühendisliği Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Dr. Öğr. Üyesi Tahsin ÇETİNYOKUŞ

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

.....

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

Başkan: Prof. Dr. Hadi GÖKÇEN

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

.....

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

Üye: Prof. Dr. Ergün ERASLAN

Endüstri Mühendisliği Ana Bilim Dalı, Ankara Yıldırım Beyazıt Üniversitesi

.....

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

Tez Savunma Tarihi: 10/10/2019

Jüri tarafından kabul edilen bu çalışmanın 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 ÇAKMAK
10/10/2019

GÜNCEL İŞ KAZALARI HAM VERİLERİNİN, VERİ MADENCİLİĞİ SÜRECİ İLE ANALİZİ

(Yüksek Lisans Tezi)

Hasan ÇAKMAK

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ekim 2019

ÖZET

İş kazaları neticesinde oluşan gerek maddi gerekse manevi kayıplar, sanayide meydana gelen üretim artışına bağlı olarak gün geçtikçe artmaktadır. Bu durumu örneklendirmek gerekirse Sosyal Güvenlik Kurumu tarafından yayınlanan İstatistik Yıllıklarına göre 2016 yılında 286.068 iş kazası olmuş, bu iş kazaları sonucunda 1.405 çalışan hayatını kaybetmiştir. Bir sonraki yıl olan 2017 yılında ise toplam 359.653 kişi iş kazasına uğramış ve bu kazalarda 1.633 çalışan hayatını kaybetmiştir. Bu nedenle iş sağlığı ve güvenliği kavramının hayatımızdaki önemi her geçen gün katlanarak artmaktadır. Bu durumda yapılması gereken, geçmiş tecrübelerden yararlanarak gelecekte meydana gelebilecek durumları öngörmeye çalışmak ve alınması gereken teknik tedbirleri doğru bir şekilde seçebilmektir. Fakat geçmiş tecrübelerden yararlanması aşamasında çok büyük boyutlardaki veriler arasından doğru ve işe yarar bilgileri tercih etmek her zaman mümkün olmamaktadır. Dolayısıyla, kullanılan veriler arasından önemli bilgilerin çıkarılması için analiz edilmesi kritik önem arz etmektedir. 2016 ve 2017 yılları arasında gerçekleşen iş kazalarına ilişkin veriler kullanılarak, iş kazalarının sebeplerinin tespitine yönelik güncel bir bakış sunan bu çalışmada, iş kazaları ciddi ve hafif kazalar olarak sınıflandırılarak sebepleri üzerinde çeşitli değişkenler sorgulanmış ve veri madenciliği yöntemleri ile analiz edilmiştir.

Bilim Kodu : 90607
Anahtar Kelimeler : İş Kazası, Veri Madenciliği, Birliktelik Kuralları
Sayfa Adedi : 90
Danışman : Dr. Öğr. Üyesi Tahsin ÇETİNYOKUŞ

ANALYSIS OF CURRENT OCCUPATIONAL ACCIDENTS RAW DATA BY DATA
MINING PROCESS

(M. Sc. Thesis)

Hasan ÇAKMAK

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

October 2019

ABSTRACT

As a result of the work accidents, both the material and spiritual losses are increasing day by day with the impact of the increasing in the industrial production. To illustrate this situation, according to the Statistical Annuals published by the Social Security Institution, 286.068 work accidents occurred in 2016 and 1.405 employees lost their lives in these work accidents. In the next year, 359.653 work accidents occurred in 2017 and 1.633 employees lost their lives in these work accidents. Therefore, the importance of the concept of occupational health and safety in our lives increases day by day. In this case, what should be done is to try to predict the future situations by taking advantage of past experiences and to choose the true technical measures to be taken. However, it is not always possible to choose the true and useful information from very large data in the stage of benefiting from past experiences. Therefore, it is critical to analyze the data to extract important information from the big data. This paper presents a current overview of the causes of occupational accidents using data on occupational accidents which occurred in between 2016 and 2017. The occupational accidents which was classified as serious or light accidents were questioned and analyzed by the data mining methods.

Science Code : 90607

Key Words : Occupational Accidents, Data Mining, Association Rules

Page Number : 90

Supervisor : Asst. Prof. Dr. Tahsin ÇETİNYOKUŞ

TEŞEKKÜR

Tez çalışmam süresince tüm bilgilerini benimle paylaşan, her türlü konuda desteklerini esirgemeyen, karamsarlığa düştüğüm anlarda dahi motive eden danışman hocam Sayın Dr. Öğretim Üyesi Tahsin ÇETİNYOKUŞ'a,

Lisans ve lisansüstü eğitimim boyunca tüm imkânlarından faydalandığım Gazi Üniversitesi Endüstri Mühendisliği Bölümü'nde eğitimime katkıda bulunan tüm öğretim üyelerine ve çalışanlara,

Tüm hayatım boyunca beni destekleyen, başarılı olmam için tüm imkânlarını seferber eden, her konuda yanımda olan sevgili aileme,

Çalışanların sağlığını ve yaşam koşullarını konu alan iş sağlığı ve güvenliği alanında faaliyet gösteren tüm meslektaşlarıma,

Sonsuz minnet ve teşekkürü bir borç bilirim...

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER.....	vii
ÇİZELGELERİN LİSTESİ	x
ŞEKİLLERİN LİSTESİ	xi
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ.....	1
2. LİTERATÜR ARAŞTIRMASI	5
3. İŞ SAĞLIĞI VE GÜVENLİĞİ	9
3.1. İş Sağlığı ve Güvenliğinin Tanımı ve Temel İSG Kavramları.....	9
3.2. İş Sağlığı ve Güvenliğinin Önemi	10
3.2.1.Toplum açısından İSG.....	11
3.2.2.Ülke ekonomisi açısından İSG	12
3.2.3. Çalışanlar açısından İSG	12
3.2.4. İşverenler yönünden İSG.....	13
3.3. Dünyada İSG Tarihçesi	13
3.4. Ülkemizde İSG Tarihçesi	14
4. VERİ MADENCİLİĞİ	15
4.1. Veri Madenciliğinin Tarihi.....	16
4.2. Veri Tabanları.....	16
4.3. Veri Ambarları.....	18

Sayfa

4.3.1. Veri ambarının oluşturulması	18
4.3.2. Veri ambarlarının özellikleri.....	19
4.4. Veri Madenciliği Süreci	20
4.4.1. İşi ve iş ortamını anlama (Business understanding).....	20
4.4.2. Veriyi anlama (Data understanding)	21
4.4.3. Veri hazırlama (Data preparation)	21
4.4.4. Modelleme (Modeling).....	23
4.4.5. Değerlendirme	24
4.4.6. Yayma	24
4.5. Veri Madenciliğinde Karşılaşılan Problemler	24
4.6. Veri Madenciliğinin Kullanıldığı Alanlar	25
5. VERİ MADENCİLİĞİ YÖNTEMLERİ	27
5.1. Sınıflama (Classification) ve Regresyon (Regression)	27
5.1.1. Karar ağaçları.....	28
5.1.2. Yapay sinir ağları	30
5.1.3. Mesafeye dayalı sınıflandırma algoritmaları	31
5.1.4. İstatistik tabanlı sınıflandırma algoritmaları.....	31
5.2. Kümeleme	32
5.2.1. Kümeleme yönteminde kullanılan uzaklık ölçümleri	33
5.2.2. Kümeleme yöntemleri	35
5.3. Birliktelik Kuralları (Association Rules)	37
5.3.1. Birliktelik kurallarının belirlenmesinde kullanılan algoritmalar	38
6. UYGULAMA	45
6.1. Apriori Algoritması Analizi	48

Sayfa

6.2. GRI (Generalized Rule Induction) Algoritması Analizi	52
6.3. C5.0 Algoritması Analizi	54
7. SONUÇ VE ÖNERİLER	57
KAYNAKLAR	61
EKLER	67
EK-1.Analizde kullanılan parametre sınıflandırmaları.....	68
EK-2.Apriori Algoritması ile elde edilen birliktelik kuralları örnekleri.....	76
EK-3.GRI Algoritması ile elde edilen birliktelik kuralları örnekleri.....	82
EK-4.C5.0 Algoritması ile elde edilen kural örnekleri.....	86
ÖZGEÇMİŞ.....	90

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 6.1. Kullanılan parametre açıklamaları.....	45

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 4.1. Veri hazırlama adımları	21
Şekil 5.1. Karar ağaçları yapısı	28
Şekil 5.2. Karar ağacı algoritması.....	29
Şekil 5.3. Yapay sinir ağı yapısı	31
Şekil 5.4. Kümeleme örneği	33
Şekil 5.5. Öklid uzaklığı.....	33
Şekil 5.6. Kümeleme yöntemleri	35
Şekil 5.7. Veri paralelleştirme modeli.....	43
Şekil 5.8. Görev paralelleştirme modeli	44
Şekil 6.1. SPSS Clementine programı üzerindeki modelleme gösterimi	47
Şekil 6.2. SPSS Clementine programında veri tablosu görünümü	47
Şekil 6.3. SPSS Clementine programında veri tipleri düzenlemesi	48
Şekil 6.4. SPSS Clementine üzerinde apriori algoritması analiz örnekleri	49
Şekil 6.5. SPSS Clementine üzerinde apriori algoritması analiz örnekleri	50
Şekil 6.6. SPSS Clementine üzerinde GRI algoritması analiz örnekleri.....	52
Şekil 6.7. SPSS Clementine üzerinde GRI algoritması analiz örnekleri.....	53
Şekil 6.8. SPSS Clementine üzerinde C5.0 algoritması analiz örnekleri	55
Şekil 6.9. SPSS Clementine üzerinde C5.0 algoritması analiz örnekleri.....	55

SİMGELER VE KISALTMALAR

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

Kısaltmalar

Açıklamalar

CART

Sınıflama ve Regresyon Ağaçları Algoritması

CRISP-DM

Cross Industry Process for Data Mining

ILO

Uluslararası Çalışma Örgütü

İSG

İş Sağlığı ve Güvenliği

VTBK

Veri Tabanlarında Bilgi Keşfi

WHO

Dünya Sağlık Örgütü

1. GİRİŞ

Çalışma faaliyetinin tarihçesi insanlık tarihi kadar geriye gidebildiği için, çalışma hayatında insanın korunmasına yönelik tedbir alınması da bu derecede geriye götürülebilmektedir. Bu noktada yazılı bir kaynağı temel almak gerekirse, M.Ö. 370 yıllarında Hipokrat tarafından kurşunun zararlı etkilerine yönelik çalışmalar yapılmıştır. Ancak daha yaygın bir kabule göre 16. ve 17. Yüzyıllarda (1633-1714) İtalyan Bernardino Ramazzini tarafından yapılan iş sağlığına yönelik bilimsel çalışmalar bu alanda ilk faaliyetlerdir. Ancak iş sağlığı ve güvenliği konusundaki çalışmalar ne kadar eskiye götürülürse götürülsün, toplumsal olarak Sanayi Devrimi ile birlikte önem kazanmıştır [1].

Sanayi Devriminin gerçekleşmesinden sonra, teknolojinin gelişmesi, imalat sektöründe makineleşmenin artması, üretim hızının değer kazandığı ve sonuç olarak da toplumsal değerlerin önemlerini kaybettiği bir ortamda çalışanların sağlık ve güvenlik yönünden korunmasız kaldığı gerçeği karşısında insanların sağlık ve güvenliklerini korumaya yönelik gerçekleştirilen çalışmalar ve alınan tedbirler bir başlangıç olarak görülmektedir [1].

Hem dünya genelinde hem de kendi ülkemizde sanayileşmenin artması ve teknolojik gelişmeler özellikle işyerlerindeki temel taşlarından birisi olan çalışanların sağlık ve güvenlikleri ile ilgili bazı sorunların ortaya çıkmasına neden olmuştur. Başlangıçta yeterli önemi görmeyen bu sorunlar zaman içerisinde iş verimini düşürmüş ve işletme açısından büyük bir tehlikenin ortaya çıkmasına yol açmıştır. Bu süreçte yapılan çalışmalar neticesinde çalışma ortam ve koşullarına yönelik bir takım kurallar ve kanunlar düzenlenmiştir. Ancak zaman içerisinde yapılan düzenlemelerin yetersiz kaldığı, yeni ihtiyaçların oluştuğu gözlemlenmiş ve farklı bakış açılarıyla sorunlara yaklaşılması gerekliliği ortaya çıkmıştır [2].

Yapılan araştırmalar sonucunda, iş kazalarının %50 oranında basit önlemlerle önlenemediği, %48 oranında önlenmesi için detaylı ve sistemli bir çalışma gerektiği, geriye kalan % 2'sinin ise önlenmesinin mümkün olmadığı ortaya çıkmıştır. Bu veriler ışığında iş kazalarının % 98 oranında önenebileceği sonucuna ulaşılmaktadır. İşyerlerinde kazaya neden olabilecek tehlike kaynaklarını tespit edip bunlardan oluşabilecek riskleri kontrol edebilirsek olası

kazaları azaltmış oluruz. Ancak başarılı sonuçlar için bu çalışmalar uyumlu bir ekip tarafından gerçekleştirilmelidir [3].

Son yıllarda işverenler açısından kalite, insan kaynakları, çevre riskleri gibi konularla ilgili yürütülen faaliyetlerin dışında, iş sağlığı ve güvenliği de yönetimi ve idaresi zor bir alan olarak kendini göstermektedir. İş sağlığı ve güvenliği konusunda başarılı olan işletmeler, iş kazaları ve meslek hastalıkları sonucu karşılaştıkları maliyetlerini asgari seviyelere indirerek karlılık konusunda kendilerine rekabet açısından avantaj sağlamaktadırlar. Aynı zamanda iş sağlığı ve güvenliğine önemli yatırımlar yapan bu işletmeler, pozitif güvenlik kültürü gereği çalışanlarının işe bağlılığını, motivasyonunu ve verimliliklerini arttırmaktadırlar [4].

Çalışma hayatında yer alan en önemli konulardan biri olan iş sağlığı ve güvenliği kavramı, geçen süreç içerisinde pek çok kurum tarafından tanımlanmıştır. Örneğin; Dünya Sağlık Örgütü (WHO) ve Uluslararası Çalışma Örgütü (ILO) tarafından belirlenen ilkelere göre işçi sağlığı ve iş güvenliği, “Tüm çalışanların bedensel, ruhsal ve toplumsal sağlık ve refahlarının en üst düzeye yükseltilmesi ve bu durumun korunması; işyeri koşullarının, çevrenin ve üretilen malların getirdiği sağlığa aykırı sonuçların ortadan kaldırılması” olarak ifade edilmektedir. Bu tanımlama sonrasında yapılan diğer tüm tanımlar; WHO ve ILO örgütlerinin yapmış olduğu açıklama çerçevesinde türetilmiştir.

İş kazası kavramının tanımlanması aşamasına gelindiğinde ise;

- Uluslararası Çalışma Örgütü (ILO); “ Önceden planlanmamış, bilinmeyen ve kontrol altına alınmamış olan, etrafa zarar verecek nitelikteki olaylar”
- Dünya Sağlık Örgütü (WHO); “Önceden planlanmamış, çoğu kişisel yaralanmalara, makinelerin ve araç gereçlerin zarara uğramasına, üretimin bir süre durmasına yol açan bir olay”
- 5510 sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu; “Sigortalının işyerinde bulunduğu sırada; İşveren tarafından yürütülmekte olan iş nedeniyle sigortalı kendi adına ve hesabına bağımsız çalışıyorsa yürütmekte olduğu iş nedeniyle; Bir işverene bağlı olarak çalışan sigortalının, görevli olarak işyeri dışında başka bir yere gönderilmesi nedeniyle asıl işini yapmaksızın geçen zamanlarda; Emziren kadın sigortalının, iş mevzuatı gereğince çocuğuna süt vermek için ayrılan zamanlarda; Sigortalıların, işverence sağlanan bir taşıtla işin yapıldığı yere gidiş geliş sırasında

meydana gelen ve sigortalıyı hemen veya sonradan bedenen ya da ruhen engelli hâle getiren olay.”

- 6331 sayılı İş Sağlığı ve Güvenliği Kanunu; “İşyerinde veya işin yürütümü nedeniyle meydana gelen, ölüme sebebiyet veren veya vücut bütünlüğünü ruhen ya da bedenen engelli hâle getiren olay” şeklinde ifade etmektedirler.

Bilgisayar teknolojisi ve iletişim sektöründeki hızlı gelişmeler, geçmiş dönemlerle kıyaslandığında çok daha fazla verinin toplanma, depolanma ve işlenerek bilgiye dönüştürülüp istenilen yere çabucak iletilmesini sağlamıştır. Ancak günümüzde gelenen noktada, bilgi hacminde inanılmaz derecedeki artışa karşın, toplanan bu verilerden ne şekilde yararlanılacağı ile ilgili bir problem doğmaktadır. Bilinen geleneksel sorgulama ve raporlama yöntemlerinin büyük hacimdeki veriler açısından yetersiz kalması, literatürde Veri Tabanlarında Bilgi Keşfi-VTBK (Knowledge Discovery in Databases) adıyla bilinen yeni arayışlara neden olmuştur. Bu ihtiyaçlar doğrultusunda, modelin kurulması ve modelin değerlendirilmesi adımlarından oluşan Veri Madenciliği (Data Mining) gündeme gelmiştir [5].

Veri madenciliği, karar verme sürecinde kullanılmak üzere veri ambarlarında bulunan farklı çeşitlilikteki verilerin, daha önce bilinmeyen bilgileri ortaya çıkarmak amacıyla kullanılması sürecidir [6].

Veri madenciliği alanında; sınıflandırma, kümeleme ve birliktelik kuralları analizi gibi pek çok konu üzerinde kullanılmak amacıyla çok sayıda algoritma geliştirilmiştir. Bu yöntemlerin çoğu istatistik tabanlı yöntemlerdir. Geliştirilen algoritmalarından bazıları sadece sınıflandırma ya da sadece kümeleme gibi konular kullanılabilmekteyken, bazı örnekleri ise değişik varyasyonlarla birden çok alanda kullanılabilmektedir. Örneklendirmek gerekirse genetik algoritmalar ve yapay sinir ağları gibi yöntemler, hem sınıflandırma hem de kümeleme modellerinde kullanılabılırler. Buna karşın, Apriori algoritması da sadece birliktelik kurallarının belirlenmesinde kullanılan bir algoritma örneğidir [7].

İş kazalarının önlenmesi insan hayatının kutsallığı çerçevesinde hem toplumsal hayat hem de çalışma barışının sağlanması açısından büyük önem arz etmektedir. Sosyal Güvenlik Kurumu Başkanlığınca yayımlanan istatistik yıllıklarına göre 2017 yılında toplam 359.653

kiři iş kazasına uğramış, bu kazalarda ise 1.633 çalışan hayatını kaybetmiştir. Bahsi geçen rakamlar iş kazaları ile mücadele etmenin önemini bir kez daha vurgulamaktadır. İş kazalarının önlenmesi için alınması gereken tedbirlerin belirlenmesi amacıyla uzun yıllar boyunca oluşan verilerin değerlendirilerek, iş kazalarının altında yatan gizli nedenlerin ortaya çıkarılması gerekmektedir. Bu gizli nedenlerin ortaya çıkarılması aşamasında büyük veri yığınlarından bilgi çıkarımı amacıyla veri madenciliğine başvurulması ihtiyacı oluşmuştur.

Bu amaç doğrultusunda, Sosyal Güvenlik Kurumunca tutulan iş kazası istatistiklerinden 2016 ve 2017 yıllarında meydana gelen iş kazalarına ilişkin istatistiki bilgiler alınarak, veri madenciliği yöntemleri aracılığıyla iş kazalarının meydana gelmesinde birbiriyle ilişkili faktörler belirlenmeye çalışılmıştır. Böylece; farklı veri madenciliği yöntemleriyle eldeki veriler analiz edilmiş, sıklıkla pazar araştırmalarında kullanılan birliktelik kuralları analizinin iş kazalarının analizi ile ilgili araştırmalarda da kullanılabileceği gösterilmiş ve iş kazalarının oluşumunda sıklıkla bir araya gelen faktörlerle ilgili anlamlı kurallar tespit edilmeye çalışılmıştır.

2. LİTERATÜR ARAŞTIRMASI

İstatistiksel analiz yöntemleri yüksek miktarlardaki verilerin arasından anlamlı bilgilerin seçilebilmesi için sıklıkla faydalanan yöntemlerdir. Bu yöntemler sadece iş kazaları ile ilgili değil çeşitli kaza türleri üzerinde de farklı tekniklerle birlikte pek çok kez denenmiştir. Bilimsel bir çalışma yapılırken geçmişte yapılan çalışmaların incelenmesi hem çalışmaya değer katmak hem de literatürdeki boşlukları görmek açısından büyük önem arz etmektedir. Bu bakış açısıyla uluslararası bilim dünyasında kabul görmüş aşağıda yer verilen çalışmalar tez çalışmama yön vermemi sağlamıştır.

Bülent Altunkaynak (2018) tarafından yayımlanan çalışmada, Türkiye’de bulunan imalat sektöründeki iş kazaları ile ilgili bir istatistiki çalışma yapılmış olup meydana gelen iş kazalarının yarısından fazlasının (%58) bu sektörde gerçekleştiği belirtilmiştir. Çalışma sonucunda; işletmelerin önlem almaya teşvik edilmesi için önlem almayan işyerlerine daha yüksek prim oranlarının uygulanması ve sosyal güvenlik denetimi uygulamalarının daha sık yapılması önerilmektedir [8].

Aylin Adem, Ali Çolak ve Metin Dağdeviren (2018) tarafından gerçekleştirilen çalışmada ise; Rüzgar türbininin yaşam döngüsündeki iş güvenliği risklerini değerlendirmek için SWOT analizi ve kararsız bulanık dilbilim terimini kullanan bütünleşik bir model önerilmektedir [9].

Çin’de bulunan yer altı madenlerindeki güvensiz davranışların etkileyici faktörlerini analiz etmek için Wanguan Qiao, Quanlong Liu, Xinchun Li, Xixi Luo ve YuLong Wan (2018) tarafından yapılan çalışmada veri madenciliği teknikleri kullanılmıştır. Söz konusu çalışmada gerçekleştirilen analizlere ek olarak apriori algoritması kullanılarak gerçekleştirilen birliktelik kuralları ile 6 güçlü güvensiz davranış kuralı da oluşturulmuştur [10].

Türkiye’de tersanelerde gerçekleşen ölümlü iş kazaları ile ilgili bireysel ve işyeri ile ilgili faktörler üzerinde Barış Barlas ve Fatih Burak İzci (2018) tarafından yapılan çalışmada, ölümlü bir iş kazasının gerçekleşmesi için kaza riskini artıran bireysel ve işyeri ile ilgili faktörlerin tanımlanması ve işyeri ortamında alınması gereken tedbirlerin belirlenmesine

yönelik 2004-2014 yılları arasında gerçekleşen olaylara ilişkin kaza istatistikleri kullanılmıştır. Çalışmanın neticesinde tespit edilen kaza faktörlerinin her biri için ayrı ayrı iyileştirme önerileri sunulmaktadır [11].

AHP Apriori yöntemine dayalı nedensellik analizi yöntemi ile trafik kazalarının ele alındığı Jianfeng Xia, Zhonghao Zhao, Wei Li ve Quan Wang (2016) tarafından yapılan çalışmada, kazaların meydana geliş sebeplerini önemlerine göre sıralamak amacıyla analitik hiyerarşi süreci kullanılmış olup, kazaların etki düzeyini ve kaza derecesini analiz etmek için ise apriori algoritmasından faydalanılmıştır. Yapılan uygulama ile metodun birden fazla faktörün neden olduğu kazaların türünün ve şiddetinin belirlenmesinde etkili bir alternatif olduğu sonucuna ulaşılmıştır [12].

Lluís Sanmiquel, Josep M. Rossell ve Carla Vintó (2015); yayımlanan çalışmalarında, veri madenciliği tekniklerini kullanarak İspanya’da madencilik sektöründe meydana gelen kazaları analiz etmişlerdir. 2003-2012 yılları arasında meydana gelen kazalara ilişkin istatistikler, Bayes sınıflandırma algoritması ve karar ağaçları yöntemlerini kullanarak Weka isimli yazılım programı yardımıyla analiz edilmiştir. Araştırma sonucunda Yer, Boyut, Fiziksel Aktivite, Önleyici Organizasyon, Deneyim ve Yaş gibi değişkenlerin çoğu kazanın oluşumunda etkili faktörler olduğu ortaya çıkmıştır [13].

Tekrarlayan iş kazalarının genel karakteristik özellikleri üzerine Julia Pietilä, Tuula Räsänen, Arto Reiman, Henriikka Ratilainen ve Elina Helander (2018) tarafından yapılan çalışmada, iş kazalarının tekrarlayan nedenlerle yeniden oluşmasının önlenmesinde kazalardan sonra önleme politikalarının yeniden değerlendirilmesinin önemi vurgulanmaktadır [14].

Maryam Sadat Hajakbari ve Behrouz Minaei-Bidgoli (2014) tarafından yapılan ve İran Çalışma Bakanlığı verileri kullanılarak veri madenciliği tekniklerinin kullanıldığı bir vaka çalışmasında iş kazası riskini değerlendirmek için yeni bir puanlama sistemi sunulmaktadır. Yapılan çalışmanın sonucunda önerilen yeni puanlama yönteminin, iş kazalarının oluşmasındaki kritik faktörlerin yüksek doğruluk seviyesinde tahmin edilip değerlendirilebileceğini gösterdiği belirtilmektedir. Yöntemin pratik bir uygulamasının sunulduğu çalışmada, 2010 yılında her bir kategori için risk seviyeleri belirlenmiş ve en fazla risk altındaki işyerleri tanımlanmıştır. Daha sonra 2011 yılında aynı prosedür tekrarlanmış

ve elde edilen veriler ile birlikte 2012 yılında periyodik denetime ihtiyaç duyulan işyerlerinin bir listesi üretilmiştir [15].

Ching-Wu Cheng, Hong-Qing Yao ve Tsung-Chih Wu (2013) tarafından gerçekleştirilen çalışmada 2000 ile 2010 yılları arasında Tayvan’da petrokimya endüstrisinde gerçekleşen 349 büyük kazanın nedenlerinin araştırılması için veri madenciliği yöntemleri kullanılmaktadır. Kazaların nedenlerinin tespit edilebilmesi için veri madenciliğinde sınıflandırma ve regresyon ağaçları olarak CART yöntemi kullanılmıştır. Çalışma sonucunda, petrokimya endüstrisinde faaliyet gösteren firmalarda boru hatları ve kontrol vanaları gibi ekipmanların yüksek kalite standartlarında güvenliğinin sağlanması ve düzenli bakım faaliyetleri sonucunda kazaların önemli ölçüde önlenebileceği ortaya konulmuş olup çalışanların eğitimi ve yüksek güvenlik bilincinin oluşturulmasının önemine dikkat çekilmiştir [16].

Dake Zhang ve Kang Jiang (2012) tarafından; veri madenciliği yöntemleri kullanılarak Çin’de gerçekleşen yangın kazaları üzerinde bir çalışma gerçekleştirilmiştir. Veri analizi ve kaza analizi teorilerinin bir kombinasyonu kullanılmaya çalışılan makalede, akademik çalışmalarda sıklıkla kullanılan birliktelik kuralları, karar ağaçları, zaman serileri ve kümeleme analizi olmak üzere 4 yöntem temel alınmıştır. Çalışma sonucunda, veri madenciliği yöntemlerinin yangın kazalarının ve hatta tüm iş kazalarının analizinde çok etkili yöntemler olduğu ortaya konulmuş olup geleneksel analiz yöntemleri yerine veri madenciliği yöntemlerinin kullanılmasının karar vericiler ve politika belirleyiciler açısından kazaların nedenlerinin daha doğru anlaşılmasına yardımcı olacağı belirtilmiştir [17].

Portekiz’deki maden çıkarma faaliyetleri sırasında gerçekleşen iş kazalarını analiz etmek için Joaquim F.Silva ve Celeste Jacinto (2012) tarafından gerçekleştirilen çalışmada sapma ve temas gibi iki değişken üzerinde çalışılan klasik yöntemler yerine, geliştirilen ve dokuz değişken üzerinde çalışılabilecek veri madenciliği yöntemi ile çalışmanın daha faydalı olacağı belirtilmektedir [18].

Veri madenciliği yöntemlerinin Finlandiya’da gerçekleşen iş kazalarına uygulanması amacıyla Noora Nenonen (2012) tarafından gerçekleştirilen çalışmada, işyerlerinde düşme, tökezleme, kayma gibi kaza nedenleri ile ilgili faktörler analiz edilmiştir. Veri madenciliği yöntemlerinden karar ağaçları ve birliktelik analizi yöntemleri kullanılan çalışma

sonucunda, büyük veri kümeleri üzerinde değişkenler arasındaki ilişkileri kolayca gösterebilme yeteneği nedeniyle, veri madenciliği yöntemlerinin, iş kazası verilerinin analizinde yararlı bir yardımcı yöntem olduğu görülmüştür [19].

Ching-Wu Cheng, Sou-Sen Leu, Ying-Mei Cheng, Tsung-Chih Wu ve Chen-Chung Lin (2012) tarafından yapılan çalışmada Tayvan’da bulunan inşaat sektöründe meydana gelen iş kazaları üzerinde etkileri bulunan faktörlerin tespiti için veri madenciliği tekniklerinin kullanılmıştır. Söz konusu çalışmada inşaat sektöründe meydana gelen kazaların analiz edilmesi için sınıflama ve regresyon ağaçları (CART) yöntemi kullanılmıştır. 2000-2009 yılları arasında gerçekleşen 1542 iş kazası verisi kullanılarak yürütülen çalışmanın sonucunda hem özel sektör hem de kamu işyerlerinde yüksekte düşme ve çökme ile meydana gelen kazaların oluşumunun yaralanmaların oluşumunda kilit rolü olduğu ortaya konulmuştur [20].

Rivas ve diğerleri (2011) tarafından yayınlanan makalede, veri madenciliği teknikleri kullanılarak iş kazalarının açıklanması ve tahmin edilmesi üzerine çalışmalar yapılmıştır. Madencilik ve inşaat sektörlerinde meydana gelen iş kazaları ve olaylara ilişkin veriler ile birlikte ilgili iş kazası veya olayın gerçekleşmesinden kısa bir süre sonra çalışan işçilerle yapılan görüşmelerin modellenmesi için sınıflandırma ağaçları, karar kuralları, destek vektör makineleri, bayes ağları gibi yöntemler kullanılmıştır. Çalışmanın neticesinde kullanılan yöntemlerin, kazaların veya olayların altında yatan gerçek nedenlerin belirlenmesi ve tanımlanmasında karşılaştırmalı olarak etkinliği yorumlanmıştır [21].

Ching-Wu Cheng, Chen-Chung Lin ve Sou-Sen Leu (2010) tarafından gerçekleştirilen çalışmada ise, Tayvan’da inşaat sektöründe gerçekleşen iş kazaları ile ilgili neden-sonuç ilişkisinin ortaya konulması ile ilgili birliktelik kuralları kullanılarak bir çalışma gerçekleştirilmiştir. Yapılan çalışma sonucunda inşaat sektöründeki iş kazalarının genellikle yüksekte çalışma, hareket halindeyken denge kaybı, kişisel koruyucu donanımların kullanılmaması, yetersiz tecrübe gibi nedenlere dayandığı sonucuna ulaşılmış olup kazaların genellikle 10’dan az çalışanı bulunan küçük işletmelerde yaşandığı ortaya çıkmıştır. Çalışma sonucunda, işyerlerindeki yönetimlerin zaaflarına dikkat çekilmiş ve personel eğitimine yönelik etkili güvenlik politikalarının geliştirilmesi gerektiği belirtilmiştir [22].

3. İŞ SAĞLIĞI VE GÜVENLİĞİ

Günümüzde sanayideki ve teknolojiye hızlı gelişmeler, sağladıkları kolaylıkların yanı sıra insan hayatına ve çevreye yönelik farklı tehlikeler yaratmıştır. Yoğun sanayileşme ve makineleşme sonucunda kullanılan yüzlerce kimyasal maddenin neden olduğu endüstriyel kazalar ve meslek hastalıkları çağımızın en büyük sorunları haline gelmiştir [23].

İş sağlığı ve güvenliği kısaca, işyerlerinde mevcut bulunan fiziki çalışma koşulları kaynaklanan mesleki problemlerin ortadan kaldırılması ya da asgari seviyeye düşürülmesi amacıyla yapılan sistematik çalışmalar bütünü olarak tanımlanabilir.

Sağlık ve güvenlik koşullarının sağlanması için dünyanın pek çok ülkesinde ulusal ve uluslararası mevzuat düzenlemeleri yapılmıştır. Ancak geline nokta görülmüştür ki; iş sağlığı ve güvenliğinin sağlanması sadece mevzuat düzenlemeleri ile mümkün olmayıp bilimsel çalışmalar ile birlikte farklı açılardan da ele alınması gereken bir konudur.

3.1. İş Sağlığı ve Güvenliğinin Tanımı ve Temel İSG Kavramları

Sağlık: Dünya Sağlık Örgütü (WHO) tarafından, “Yalnızca hastalık veya sakatlığın olmaması durumu değil, aynı zamanda fiziksel, sosyal ve ruhsal olarak tam bir iyilik hali” biçiminde ifade edilmiştir.

İş Sağlığı ve Güvenliği (İSG): İşyerlerinde gerçekleştirilen faaliyetler sonucunda oluşan ve pek çok farklı nedenden kaynaklanan tehlikeler ile sağlığa zararlı olabilecek çeşitli durumlardan korunmak amacıyla gerçekleştirilen bilimsel çalışmalardır.

İş Sağlığı ve Güvenliğinin temel amaçlarını, çalışanların sağlık ve güvenliklerinin korunması, iş kazaları ve meslek hastalıklarına karşı korunmaları, üretim ve işletme güvenliğinin sağlanması olarak sıralayabiliriz.

İşyeri: 6331 Sayılı İş Sağlığı ve Güvenliği Kanunu’nda yapılan tanıma göre mal veya hizmet üretmek amacıyla maddi olan ve olmayan unsurlar ile çalışanın birlikte örgütlendiği, işverenin işyerinde ürettiği mal veya hizmet ile nitelik yönünden bağlılığı bulunan ve aynı

yönetim altında örgütlenen işyerine bağlı yerler ile dinlenme, çocuk emzirme, yemek, uyku, yıkanma, muayene ve bakım, beden ve mesleki eğitim yerleri ve avlu gibi diğer eklentiler ve araçları da içeren organizasyonu ifade etmektedir.

Tehlike: 6331 Sayılı İş Sağlığı ve Güvenliği Kanunu'nda yapılan tanıma göre işyerinde var olan ya da dışarıdan gelebilecek, çalışanı veya işyerini etkileyebilecek zarar veya hasar verme potansiyelini tanımlamaktadır.

Risk: 6331 Sayılı İş Sağlığı ve Güvenliği Kanunu'nda yapılan tanıma göre tehlikeden kaynaklanacak kayıp, yaralanma ya da başka zararlı sonuç meydana gelme ihtimalini ifade etmektedir.

Risk Değerlendirmesi: 6331 Sayılı İş Sağlığı ve Güvenliği Kanunu'nda yapılan tanıma göre işyerinde var olan ya da dışarıdan gelebilecek tehlikelerin belirlenmesi, bu tehlikelerin riske dönüşmesine yol açan faktörler ile tehlikelerden kaynaklanan risklerin analiz edilerek derecelendirilmesi ve kontrol tedbirlerinin kararlaştırılması amacıyla yapılması gerekli çalışmaları tanımlamaktadır.

İş Kazası: 6331 Sayılı İş Sağlığı ve Güvenliği Kanunu'nda yapılan tanıma göre işyerinde veya işin yürütümü nedeniyle meydana gelen, ölüme sebebiyet veren veya vücut bütünlüğünü ruhen ya da bedenen engelli hâle getiren olay şeklinde tanımlanmıştır.

Meslek Hastalığı: 6331 Sayılı İş Sağlığı ve Güvenliği Kanunu'nda yapılan tanıma göre mesleki risklere maruziyet sonucu ortaya çıkan hastalığı ifade etmektedir.

3.2. İş Sağlığı ve Güvenliğinin Önemi

Bilim ve teknolojiye yaşanan gelişmeler, çalışma hayatında önemli değişiklikler yaratırken aynı zamanda bu değişikliklerin toplumun her noktasında yansımaları yaşanmıştır. İş sağlığı ve güvenliğinin gelişiminde hiç kuşku yok ki en önemli olay sanayi devrimidir. Sanayi devrimi sonrasında yaşanan gelişmeler ile işyerlerinde iş gücü ihtiyacı ortaya çıkmış, bu ihtiyaç sonucunda yoğun bir göç dalgası yaşanmış ve işyerlerinin yoğun olduğu bölgelerde düzensiz ve sağlıklı bir yerleşim meydana gelmiştir.

Sanayi üretiminin yoğun olduğu sektörler gibi tarım ve hayvancılıkta da makineleşme yoğun olarak görülmüş, bunun sonucunda ise gerekli önlemlerin alınmaması nedeniyle çok sayıda iş kazası meydana gelmiş, çalışma saatlerinin uzunluğu, beslenme ve barınma koşullarındaki sağlık sorunları gibi problemler ortaya çıkmıştır.

İş sağlığı ve güvenliğinin önemi farklı bakış açılarına göre ayrı ayrı değerlendirildiğinde aşağıdaki sınıflandırma yapılabilmektedir.

3.2.1.Toplum açısından İSG

Ağır ve korumasız çalışma koşulları, çalışma hayatında görülen iş kazalarının ve meslek hastalıklarının sayısını arttırmış, özellikle çocukların bu faaliyetler sonucunda zarar görmeleri, toplumda ilk tepkilerin oluşmasına neden olmuştur. Çalışan nüfusun artması, çalışma yaşının düşmesi ve çalışma ortamından kaynaklanan problemlerin toplumun genelini etkilemesi iş sağlığı ve güvenliği alanının daha çok dikkat çekmesine ve sosyal bir duyarlılık halini almasını sağlamıştır [24].

Teknolojide yaşanan gelişmeler ve iş sağlığı ve güvenliği konusundaki duyarlılığın artması zaman içinde risklerin önceden belirlenerek tehlikelerin kaynağında yok edilmesi felsefesini temel alan önleyici iş sağlığı ve güvenliği faaliyetlerini ön plana çıkarmış ve daha güvenli çalışma ortamlarını sağlayacak çalışmaların başlamasına imkân vermiştir. Ancak buna karşın karmaşıklaşan üretim süreçleri, kullanılan zararlı maddelerin çeşitliliği ve farklı üretim yöntemlerinin de gelişiyor olması çalışanları etkileyebilecek yeni riskler yaratmaya devam etmekte olup benzeri pek çok faktör iş kazası ve meslek hastalıklarını korkutucu boyutlara taşımaktadır [24].

Çalışanları korumaya yönelik özellikle ILO tarafından 1970'lerden bu yana yürütülen faaliyetlere rağmen iş kazaları ve meslek hastalıklarının sayısı dünya genelinde artmaya devam etmektedir. Bu artışa mevzuat anlamındaki sıkıntılar, çalışma hayatında bulunan aktörlerin güvenlik kültürüne yaklaşımı, eğitim eksiklikleri gibi yerel problemlerin dışında sanayileşme sürecini tamamlamış ülkelerdeki emek yoğun üretimlerin, emeğin ucuz ve sosyal korunma farkındalığının düşük olduğu ülkelere kaydırılmasının da önemli bir etkisi olduğu bilinmektedir.

WHO, dünya üzerinde bulunan 3 milyar çalışanın %80'inden fazlasının temel sağlık ve güvenlik hizmetlerinden yoksun olduğuna ve bu hizmete duyulan ihtiyacın gün geçtikçe daha da büyüdüğüne dikkat çekmektedir [25].

3.2.2. Ülke ekonomisi açısından İSG

Sorunun toplumun sosyal yaşantısının dışında, aileden işyerine ve işyerinden ülke ekonomisine uzanan farklı bir boyutu da değerlendirilmelidir. İş sağlığı ve güvenliği ile ilgili tedbir maliyetlerinin yanı sıra yaşanan olumsuz olaylar sonucunda gerçekleşen maliyetler de hesaplandığında işyerlerinin rekabet gücünün etkilenmesinin yanı sıra ülke ekonomisi açısından da önemli bir kalem oluşturarak ekonomik ve sosyal kalkınmaya ilişkin hedefleri de değiştirebilmektedir [26].

ILO, iş kazaları ve meslek hastalıkları ile mücadele etmek için iş sağlığı ve güvenliğine yatırım yapmak gerektiğini, bu yatırımı yapmamanın ekonomiye büyük bir yük getirdiğini belirtmektedir. Ülkelerin gelişmişlik düzeyi arttıkça iş kazaları ve meslek hastalıkları sonucu ödenen kaybın oranı düşmektedir [26].

Türkiye’de iş sağlığı ve güvenliği ile ilgili ILO’nun maliyet tahminleri göstermektedir ki iş kazalarından kaynaklanan yıllık kayıpların 2012 yılında 57 Milyar TL, 2013 yılında 62 Milyar TL ve 2014 yılında 70 Milyar TL civarında gerçekleştiği tahmin edilmektedir. Pek çok kamu kurumunun bütçesinden fazla olan bu kayıplar, tüm kamu sosyal harcamalarının 3 katını aşarken, Sosyal Güvenlik Kurumu’nun giderlerinin yaklaşık %30’unu oluşturmaktadır. Örnekler halinde açıklanan veriler; iş sağlığı ve güvenliğine yönelik yatırımların sadece insani temelli değil, kar amacı güden işletmeler açısından ekonomik olarak da konuya önem verilmesi gerektiğini göstermektedir [27].

3.2.3. Çalışanlar açısından İSG

İnsan hayatının kutsallığı gereği herhangi bir ayırım gözetmeksizin herkesin yaşama hakkının sağlanması ve güvence altında tutulması gerekmektedir. Oysaki sağlıklı ve güvenli bir çalışma ortamının sağlanamamış olması doğrudan çalışanları etkilemektedir. Çalışanlar ruh ve beden sağlığının bozulması hem kendilerinin hem de bakmakla yükümlü oldukları kişilerin yaşamlarını etkilemektedir [25].

İşyerlerinde sağlık ve güvenlik yönünden bulunan eksiklikler çalışanların yaşamlarını kaybetmesine, ruh ve beden sağlıklarında geçici veya kalıcı bozulmalar yaşamalarına ve en nihayetinde gelir düzeylerinde azalmalara neden olmaktadır. Sosyal güvenlik sistemleri bu açığı kapatmaya yönelik faaliyetler yürütse dahi gelir getirme yeteneğinde meydana gelecek azalmalar kişiler açısından ciddi sorunlar oluşturmaktadır.

İş gücü kaybı ile birlikte kişinin yaşadığı statü kaybı, aileye ve çevresine yük olma duygusu gibi ölçülemeyen dolaylı kayıplar ise iş kazalarının psikolojik boyutunu ortaya koymaktadır [25].

3.2.4. İşverenler yönünden İSG

İş kazaları ve meslek hastalıklarından doğan maliyetler işverenlerin karlılığını etkilemekte ve rekabet avantajlarına zarar vermektedir. Bu maliyetlerin bir kısmı ön görülebilir maliyetlerdir. Bunlar arasında tedavi masrafları, olası ölüm yardımları, mahkeme masrafları, olası tazminatlar vb. sayılabilir. İşverenler sigortacılık marifetiyle bu maliyetleri kontrol altında tutabilmektedir. Ancak daha büyük maliyetlerin öngörülemeyen dolaylı maliyetlerdir. Kaza sırasında kaybedilen zaman, kazanın diğer çalışanlar üzerindeki psikolojik etkisi ve verim düşüklüğü, hasara uğrayan malzeme ve ekipman maliyetleri, olası sipariş cezaları ve pazar kaybı vb. olarak sıralanabilen dolaylı maliyetlerin doğrudan maliyetlerin 4 ila 10 katı arasında değiştiği belirtilmektedir [27].

Güvenli bir çalışma ortamının sağlanması için yapılacak ön harcamalar (koruyucu ekipmanların temini, eğitim, risklerin ortadan kaldırılmasına yönelik faaliyetler vb.) işverenler açısından şüphesiz maliyetleri yükselten unsurlardır. Ancak yapılan çalışmalar göstermiştir ki alınan güvenlik tedbirlerinin maliyeti, uzun dönemde yapılacak kaza sonrası harcamalara göre çok daha düşüktür. Ayrıca alınan tedbirlerin iş kazalarını azaltırken verimliliği artırdığı ve ürün maliyetlerini düşürdüğü gözlenmiştir [28].

3.3. Dünyada İSG Tarihçesi

Dünya üzerinde iş kazaları ile meslek hastalıklarına karşı ilgi Antik Yunanlılardan itibaren başlamıştır. Hipokrat'ın (MÖ 460-370) madenlerde kurşun zehirlenmelerine karşı çalışmaları olmuştur. Yine Romalı Pliny (MS 23-77) ise, ilk kişisel koruyucu donanım

olarak kabul edilebilecek deri maskeleri yapmıştır. Ancak ne var ki her alanda olduğu gibi orta çağda da bir dönem bu araştırmalara ara verilmiştir. Akabinde Rönesans ile birlikte iş sağlığı ve güvenliği alanındaki araştırmalara yeniden başlanılmıştır. Sanayi devrimi sonrasında İngiltere bu konuda dünyaya öncülük etmiştir. Percival Pott'un çalışmaları sonucunda baca temizleyen kişilerin daha yoğun bir şekilde kansere yakalandıklarının tespit edilmesi İngiltere'de Baca Temizleyicileri Yasası 1788'de kabul edilmiştir. Yine 1802'de çıraklık kanunu ile yasal düzenlemeler başlamıştır. Daha sonrasında bu kanunları 1824 yılında sendikaları serbest bırakan kanun ve 1833'deki Fabrikalar yasası takip etmiştir.

Tüm bu yasal düzenlemelerde elbette sendikaların itici etkisinin büyük önemi bulunmaktadır. Ayrıca 1919 yılında kurulan Uluslararası Çalışma Örgütü (ILO) çok büyük katkı sağlanmıştır. ILO, 1946 yılında Birleşmiş Milletler (BM) ile imzaladığı bir anlaşma ile uzmanlık kuruluşu olarak tanınmaya başlamıştır. Ülkemiz de Uluslararası Çalışma Örgütü'nün bir üyesidir ve uluslararası ILO sözleşmelerinin tamamını olmasa da önemli bir bölümünü imzalamıştır.

3.4. Ülkemizde İSG Tarihçesi

Türkiye'de iş sağlığı ve güvenliği kavramında, dünyanın biraz gerisinde kalırsa da, yine de kömür madenciliğini konu alan 1865 tarihli Dilaver Paşa Nizamnamesi ilk yasal düzenleme olarak bilinmektedir. Cumhuriyet Döneminde, 1921 yılında Maden İşçilerinin Hukukuna İlişkin Kanun çıkarılmıştır. 1924'te yürürlüğe giren 394 sayılı Kanun ile hafta tatili getirilmiştir. Türkiye'nin ilk Borçlar Yasasında, (818 s. Kanun - 1926) meslek hastalıkları ve iş kazası ile ilgili konulara da yer verilmiştir.

Konu hakkında daha sonra sırasıyla yapılan yasal düzenlemeler: 1936 yılı 3008 sayılı İş Kanunu; 1945 yılı 4792 sayılı İşçi Sigortaları Kurumu Kanunu; 1964 yılı 506 sayılı Sosyal Sigortalar Kanunu 1971 yılı 1475 sayılı İş Kanunu; 2003 yılı 4857 sayılı İş Kanunu; 2006 yılı 5510 sayılı Sosyal Sigortalar ve Genel Sağlık Sigortası Kanunu; 2013 yılı İş Sağlığı ve Güvenliği Kanunu şeklinde sıralanabilmektedir.

4. VERİ MADENCİLİĞİ

İnsanları diğer tüm canlılardan üstün kılan özellikler sıralandığında başta bilgiyi kullanabilme ve bilimsel çalışmaları takip etmenin geldiği açıktır. Toplumlar ancak sahip oldukları bilgiyi daha etkin kullanarak ve bilimsel gelişmelerin ışığından giderek daha refah yaşam şartlarına sahip olabilirler. Bu sebeplerle bilgi düzeyinin çok önemli bir gösterge olduğu açıktır [29]. Teknolojinin gelişimi ile birlikte, günümüzde her türlü bilgiye hızlı ulaşabilmekle birlikte bu bilgileri harmanlayarak yorumlamak gerekmektedir.

Teknolojik gelişmeler bilgisayar sistemlerinin zaman geçtikçe ucuzlamasına ve bu sistemlerin daha güçlü hale gelmesine imkân sağlamaktadırlar. Bu sistemlerin daha güçlü hale gelmesiyle birlikte daha büyük verilerin saklanabilme ve işlenebilme yolu açılmıştır. Bu gelişmeler ışığında çok büyük verilerin kayıt altında tutulması, bu verilere kolaylıkla erişilmesi ve verilerin güvenli bir şekilde saklanarak doğru ve detaylı bilgilere ulaşmak kolaylaşmış ancak bu kolaylıkların yanında bu verilerin yönetilmesi ve bilgiye dönüştürülmesi amacıyla kullanılabilir hale getirilmesi problemleri ortaya çıkmıştır. Bilinmektedir ki, büyük miktarlardaki veriler işlenerek anlamlı hale getirilmedikçe sadece karmaşık veri yığınları olarak kalacaktır. Buna karşın bu karmaşık veri yığınları işlenerek bilgi haline getirilirse fayda sağlanacak pek çok bilgiye ulaşılacaktır.[30]. Bu sebeple, büyük miktarlardaki veri yığınlarını işleyebilmek için uygun metodolojinin seçimi büyük bir önem arz etmektedir [29].

Kalabalık ve karmaşık veri kümelerinin arasından, uygulamaya dönük ve anlam ifade eden bilgilerin çıkarılması işlemine veri madenciliği denir. Aynı zamanda veri madenciliği bu verilerin analiz tekniklerinin bir bütünü olarak ifade edilebilir [31]. Mevcut verileri kullanarak problemi çözmek, karar alma sürecinde veya ileriye dönük tahminler yapmak için ihtiyaç duyulan bilgileri toplamaya yarayan bir araçtır ve aynı zamanda veri tabanlarında saklanmış ilişkiler ve kuralların ortaya çıkmasını sağlayan bir yöntemdir [32]. Bu avantajlar veri madenciliğini, günümüzde en ilgi çekici ve en çok tercih edilen teknolojiler arasına sokmuştur. Veri madenciliğinin başka yöntemlere göre daha avantajlı görülmesindeki en önemli faktör, büyük miktarlardaki karmaşık veri kümelerinden istenilen bilgiyi elde edebilmesi ve pek çok kimse tarafından fark edilemeyen ilişkileri ortaya çıkarabilmesidir [33].

4.1. Veri Madenciliğinin Tarihi

Veri madenciliğinin geçmişi 1946 yılında II. Dünya savaşında, ABD’li John Mauch’y ve J. Presper Ecbert tarafından ABD ordusu için geliştirilen ve günümüzde kullanılmakta olduğumuz bilgisayarların atası sayılan ilk dijital bilgisayar ENIAC (Elektirical Numerical integrator and Calculator)’a kadar gitmektedir. 170 m²’lik bir alana kurulan ve ağırlığı 30 tonu bulan bu bilgisayarlar önemli bir evrim geçirerek bugünkü halini almıştır. 1950’li yıllarda kullanılan bu bilgisayarlar genelde hesap ve sayım amacıyla kullanılmıştır. 1960’larda ise ilk defa verilerin depolanması fikri ortaya atılmıştır. Bu süreçle birlikte veri tabanı kavramı teknoloji camiasında yerini almış olup 1960’lı yılların sonunda basit öğrenme yeteneğine sahip bilgisayarlar geliştirilmiştir. 1970’lerde basit kuralları temel alan uzman sistemler geliştirilerek basit anlamda makine öğrenimi sağlanmıştır. 1980’li yıllarda veri tabanı yönetim sistemleri pek çok alanda kullanılmaya başlanmış olup veri tabanlarındaki veri miktarları çoğalarak artmıştır. Veri miktarındaki bu artışla birlikte bu hacimdeki verilerin içerisinde yararlı verilerin saptanması problemi ortaya çıkmıştır [34].

Veri tabanlarında yer alan yararlı bilgilerin kullanımı için yapılan araştırmalar sonucunda 1992 yılında bu konu ile ilgili bir yazılım geliştirilmiştir. Bu karmaşık ve büyük hacimdeki verilerden yararlanılabileceğinin anlaşılması üzerine 2000’li yılların başından itibaren veri madenciliği yöntemleri tüm alanlarda sıklıkla kullanılmaya başlanmıştır [35].

Özellikle bilgisayar teknolojisindeki gelişmeler hem büyük hacimlerdeki verilerin depolanması için gerekli olan kapasiteyi sağlamakta, hem de bilgisayar hızlarının işlem sürelerini kısaltmasına imkân sağlamıştır. Bu teknolojik altyapı ile birlikte veri madenciliği için farklı yöntem ve algoritmalar geliştirilebilmiş ve hayatın her alanına yaygınlaştırılabilmektedir.

4.2. Veri Tabanları

Günlük hayatımızda yaşadığımız ve çevremizde gelişen her olay sonucunda veri üretmekteyiz. Ürettiğimiz veriler ise ilgililer tarafından kullanılmak üzere saklanmaktadır. Ancak ne var ki verilerimiz kullanıcılar tarafından yapılacak değerlendirmelerden çok daha hızlı bir şekilde artmaktadır. Dolayısı ile bu verilerin düzenli bir şekilde depolanması

gerekmektedir. Düzenli verilerin bir arada tutulduğu bu bilgiler topluluğuna ise “veri tabanı” denilmektedir.

Bilgisayar ve teknoloji dilinde ise veri tabanı kavramı, sistematik ulaşım imkânına sahip olan, yönetilmesi, güncellenmesi, aktarılması mümkün ve içerikleri açısından birbirleriyle ilişkilendirilebilir bilgiler kümesi olarak tanımlanmaktadır. Ancak; bir araya gelmiş her veri kümesini veri tabanı olarak isimlendirmek doğru olmaz. Veri tabanındaki veriler gereksiz yinelemelerden arınmış olarak, düzenli bir biçimde bilgisayar belleklerinde saklanır. Giriş/çıkış verileri ve geçici veriler veri tabanında yer almaz. Veri Tabanında asıl önemli kavram, kayıt yığını ya da bilgi parçalarının tanımlanmasıdır. Bir veri tabanı yönetim sistemi aracılığıyla, veri tabanının bilgisayar belleklerindeki fiziksel yapısı kullanıcılardan gizlenir. Kullanıcılara daha yalın mantıksal yapılar sunulur. Ancak veri tabanındaki veriler üzerinde her zaman merkezi bir denetim vardır.

Veri tabanı yazılımı ise verileri düzenli bir biçimde depolamaya yarayan yazılımlar olarak tanımlanır. Birçok yazılım programı vasıtasıyla veriler depolanabilir ancak bu yazılımın veri tabanı olarak isimlendirilebilmesi için bu verileri verimli ve hızlı bir şekilde yönetip değiştirebilmesi gerekmektedir. Bağıntısal Veri Tabanı Yönetim Sistemleri (Relational Database Management Systems - *RDBMS*) büyük hacimlerdeki verilerin güvenli bir şekilde saklanabildiği, istenilen verilere hızlı bir şekilde erişim imkânının olduğu ve birden fazla kullanıcıya aynı anda istenilen verilere ulaşım imkânının sağlandığı programlar olarak tanımlanmaktadır. Oracle veri tabanı da bağıntısal veri tabanı yönetim sistemlerine verilecek örnekler arasında sayılabilmektedir [36].

Veri tabanlarının sağladığı yararları özetlemek gerekirse:

- ✓ Belgelerin ayrıca saklanmasına gerek olmaması,
- ✓ Bilgilerin çok hızlı bir şekilde güncellenerek zaman tasarrufu sağlanması,
- ✓ Direkt olarak istenilen bilgiye istenilen zamanda erişebilme imkânı,
- ✓ Merkezi idarelerce verilerin kontrollerinin rahatlıkla yapılabilmesi,
- ✓ Veri tekrarlarının engellenerek mükerrerliğin önüne geçilebilmesi,
- ✓ Hatalı bilgi girişlerinin engellenmesi,
- ✓ İstenilen verilerin rahatlıkla paylaşılabilmesi,
- ✓ İstenilen detayda raporlamanın kolaylıkla yapılabilmesi şeklinde sıralanabilir [37].

4.3. Veri Ambarları

Veri ambarı tanımının yaratıcısı olarak isimlendirilen William H. Inmon'un 1992 yılında yapmış olduğu "*Developing the Data Warehouse*" çalışmasına göre veri ambarı, konu odaklı, karar verme sürecini etkinleştiren, bütünleşik ve kalıcı veri topluluğudur. Bir veri ambarı, analiz yapmak ve sorgular oluşturmak için kullanılabilir, bütünleşmiş bilgi deposudur. Veri Ambarları, hemen hemen tüm sektörlerde yönetimin stratejilerini belirlemesinde ve karar vermesini kolaylaştırmakta kullanılmakta olan bir sistemdir. Pahalı bir yatırım maliyeti olsa bile sonuç olarak getirisi (yararı) bu maliyeti rahatlıkla karşılamaktadır. Belirli bir döneme ait, yapılacak çalışmaya göre konu odaklı olarak düzenlenmiş, birleştirilmiş ve sabitlenmiş işletmelere ait veri tabanlarına "veri ambarı" denir [38].

4.3.1. Veri ambarının oluşturulması

Veri ambarının oluşturulması veri madenciliği süreci açısından çok önemli bir aşamadır. Çünkü veri ambarının oluşturulması süreci maliyet ve zaman açısından önemli bir yer tutmaktadır. Veri madenciliği yapılacak veriler tek bir platform üzerinde bulunmayabilir. Bu sebeple bilgi oluşturmak istediğimiz verilerin tek bir çatı altında toplanması gerekmektedir. Ancak bu da tek başına yeterli olmayacaktır. Veri ambarı oluşturmak amacıyla başkaca işlemlerin de veri kümeleri üzerinde yapılması gerekmektedir.

Veri ambarı oluşturma aşamaları:

- i. Toplama: farklı kaynaklarda saklanan verilerin tek bir kaynaktan bir araya getirilmesi işlemidir. Örneğin, hastaların tıbbi verileri yanı sıra bazı demografik özelliklerinin de bilinmesi gerekebilir. Bu tür farklı kaynaklarda saklanan verilerin bir araya getirilerek birleştirilmesi yani toplanması gerekebilir.
- ii. Uyumlandırma: bir araya getirilen veriler arasında uyumsuzluklar ortaya çıkabilir. Veri madenciliği sürecinin başarı ile sonuçlanmasında verilerin uyumlu olması çok önemlidir. Örneğin, cinsiyet verileri bir kaynaktan 1/0 şeklinde tutulurken diğer bir kaynaktan E/K, veya Erkek/Bayan şeklinde tutuluyor olabilir

iii. Birleştirme ve Temizleme: Uyumlandırma işlemi sonrasında verilerin dikkatli bir biçimde taranarak birleştirilmesi ve anlamlılık düzeyini bozan fazlalıkların temizlenmesi gerekir.

iv. Seçme: Kullanılacak model için elverişli verinin seçilmesi işlemidir. Çünkü veri tabanlarındaki işlem hızları artmasına rağmen uygun seçilmemiş veriler ile model denemesi yapılması ancak zaman ve maliyet açısından kayıplara neden olacak, istenilen sonuca ulaşılamayacaktır [39].

4.3.2. Veri ambarlarının özellikleri

Veri ambarlarının taşınması gereken özellikler aşağıdaki şekilde sıralanabilir:

- Konu Odaklı Olma: Veri ambarları, hem kullanıcılar için hem de karar vericiler için yalnızca gerekli olan ve istenilen verilere odaklı olup, o konu hakkında basit ve anlaşılır görüş elde etmeye yönelik olmalıdır.
- Bütünleşik Olma: Veri ambarları, genelde ilişkisel veri tabanları, pek çok farklı yapıdaki kaynakların bütünleştirilmesiyle oluşturulmalıdır.
- Belirli Zaman Dilimine Ait Olma: Veri ambarları zaman açısından belirli bir trend yakalaması amacıyla ihtiyaca göre belirlenecek bir zaman aralığını kapsamalıdır. Her verinin, dolaylı veya doğrudan bir zaman değişkeni ile ilişkilendirilmesi gerekebilir.
- Kalıcılık (Uçucu Olmama): Ambardaki verilerin, direk olarak silinmemesi veya güncellenmemesi gerekmektedir. İhtiyaçlara cevap vermesi açısından veriler yalnızca okunabilir (read-only) yapıda tutulmalıdır. Ambarlarda “veri yükleme (loading)” ve “veriye erişim (access)” olmak üzere yalnızca iki tür işlem uygulanır [7].

Veri ambarlarında özelliklerine göre şu tür veriler bulunur:

- Metaveri: Veri ambarının en temel bileşenlerinden birisi olan metaveri karar destek analizcilerine yardım etmek amacıyla tasarlanmış bir dizindir ve ilgili ambarın içeriği ile ilgili bilgi verir. Bir kılavuz niteliğinde olup verinin konumu, hangi algoritmaya göre oluşturulduğu, içeriği, kapsamı, çerçevesi vb. gibi bilgileri içermektedir.
- Ayrıntı veri: Henüz işlenmemiş olan ve bu sebeple diğerlerine göre daha büyük yer tutan, disk üzerinde saklandığı için erişimi ve yönetimi pahalı olan ve ambardaki en son olayları içeren verilerdir.

- Eski ayrıntı veri: Ayrıntı veriye göre daha eski tarihe ait olan verilerdir. Ayrıntı veriye göre daha düşük bir ayrıntı düzeyine indirgenerek saklanan daha küçük veriler olarak ifade edilmektedir.
- Düşük düzeyde özetlenmiş veri: Ayrıntı veriden elde edilen ve düşük seviyede süzülerek özetlenmiş verilerdir. Hangi verinin özetleme işlemine tabi tutulacağı ve özetleme işleminin ne düzeyde yapılacağı veri ambarının tasarımı esnasında belirlenmelidir.
- Yüksek düzeyde özetlenmiş veri: Ayrıntı veri daha yüksek düzeyde süzülerek özetlenen, kolayca erişilebilir hale getirilmiş ve veri ambarının bir bileşeni olarak yer alan verileri ifade etmektedir.

Kısacası Veri Ambarları, karar verme sürecinde üst yönetimin ihtiyaç duyduğu bilgileri depolayan ve istenilen düzeyde özel sorgulamaların yapılabildiği bir yapı olarak karşımıza çıkmaktadır [40].

4.4. Veri Madenciliği Süreci

Veri madenciliği süreci veri madenciliği uygulamaları açısından uluslararası bir standart olarak kabul edilebilecek yetkinliğe sahip, CRISP-DM (Cross Industry Process for Data Mining) baz alınarak gerçekleştirilmektedir. Veri madenciliği projelerinin daha hızlı, daha verimli ve daha düşük maliyetli gerçekleştirilmesini amaçlayan bu süreçler aşağıda detaylı olarak yer almaktadır [29].

4.4.1. İşi ve iş ortamını anlama (Business understanding)

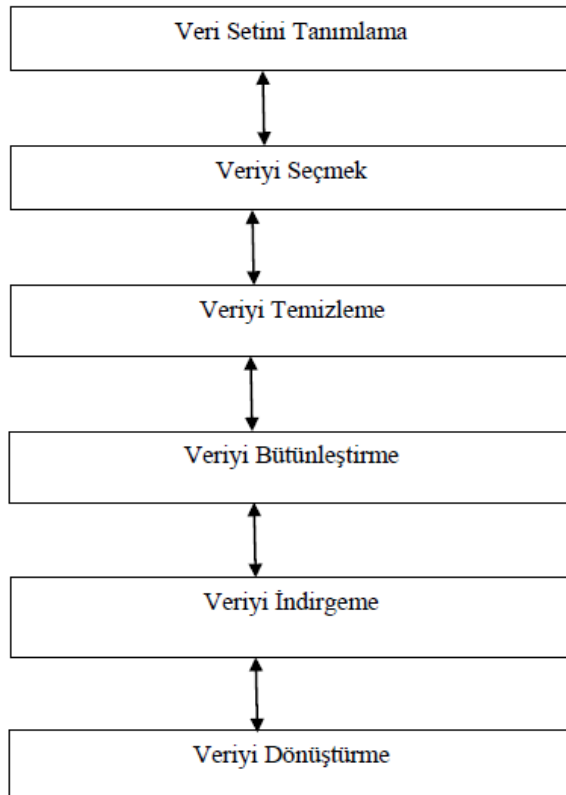
Veri madenciliği sürecinin başlangıç adımı olarak isimlendirilen işi ve iş ortamını anlama aşamasında öncelikle çalışmanın amacı net olarak tanımlanmalıdır. Daha sonra bu çalışmanın hangi zaman dilimi için yapılacağı ve çalışma sonuçlarının hangi kriterlerle değerlendirileceği gibi tanımlamaların bu aşamada planlanması gerekmektedir. Temel amaçlarından biri verimi arttırmak olan veri madenciliğinin bu ilk aşaması gerekli olan kaynaklar, eldeki veriler, riskler, maliyetler gibi kriterlerin hepsinin değerlendirildiği süreçtir [29].

4.4.2. Veriyi anlama (Data understanding)

Esas olarak veriyi anlama süreci ile iş ve iş ortamını anlama süreci iç içe geçmiş durumdadır. Çünkü elde bulunan veriler incelendikçe işle ilgili farklı izlenimler edinebilir ve bu farklı izlenimler ile birlikte verilere farklı bir bakış açısıyla yaklaşılabilir. Bu döngü kendi içerisinde devam ettiği sürece, çalışma süresince kullanılacak olan veriler değişiklik kazanarak netleşecektir. Bu süreçte başlangıç verilerinin ihtiyacı karşılayıp karşılamadığı, keşfedilecek başka verilerin bulunup bulunmadığı sorguları yapılmalıdır [29, 35].

4.4.3. Veri hazırlama (Data preparation)

Sürecin bu aşamasında başlangıç verilerinin, çalışmada kullanılmak üzere hazırlanması amaçlanmaktadır. Bu aşamanın tanımlanmış bir sırası ya da zamanı yoktur. Süreçte oluşan problemler nedeniyle sık sık bu aşamaya dönülmesi ve tekrarlanması gerekebilir. Burada ki aşamalar;



Şekil 4.1. Veri hazırlama adımları

Veri setini tanımlama:

Verilerin toplanacağı kaynaklar belirlenerek elde edilen verilerin yeterli olmaması durumunda başka kaynaklara başvurulması işlemidir.

Veriyi seçmek:

Çalışmada kullanılması düşünülen verilerin belirlenmesidir. Kullanılacak verilerin çalışmanın amacına uygun olması gerekmekte olup aynı zamanda veri sayısının az ya da fazla olmaması gerekmektedir. Veri sayısının az olması durumunda istenilen sonuçların elde edilememesine, fazla olması durumunda ise veri kirliliğine ve sürecin uzamasına yol açmaktadır.

Veri temizleme:

Veri tabanlarında yer alan ve kullanılması planlanan verilerin istenilen kriterlere uymadığı, eksik veya tutarsız olduğu durumlarla sıkça karşılaşılmaktadır. Bu gibi durumlarda veri tabanlarında yer alan eksik veya hatalı veriler gürültü olarak adlandırılmaktadır. Bu gibi gürültülü verilerin olması durumunda, bu sorunun giderilmesi amacıyla veri tabanından bu verilerin silinmesi, yerlerine yenilerinin eklenmesi, bu verilerin yerine sabit bir değer kullanılması, ortalama değerler alınarak gürültülü veriler yerine bu değerlerin kullanılması gibi yöntemlerle bu verilerin temizlenmesi gerekmektedir [41].

Veri bütünleştirme:

Kullanılması planlanan verilerin farklı farklı kaynaklardan toplanması ihtiyacı oluşabilir. Bu durumda bu verilerin tek kaynakta birleştirilmesi aşamasında farklı kaynaklar arasındaki uyumsuzluklar giderilmeye çalışılır. Bu aşamada çok dikkatli olunması gereklidir çünkü olası hatalar ileri aşamalarda yer alan tüm işlemleri etkileyecektir.

Veri indirgeme:

Veri indirgeme aşaması elde edilen verilerin bir kısmının çıkarılması ya da çıkarılmaması sonucu değiştirmeyecekse uygulanmaktadır. Veri indirgeme aşamasında aşağıdaki yöntemler uygulanabilir:

- Birleştirme: Veriler çok boyutlu veri küpleri biçimine dönüştürülür.
- Genelleme: Veriler tek tek değil genel kavramlarla ifade edilir.
- Örnekleme: Tüm veri kümeleri yerine onu temsil eden küme grupları kullanılır.
- Boyut indirgeme: Veriler arasından gereksiz veriler silinir.
- Veri sıkıştırma: Boyut sıkıştırması yapılarak verilerin daha az yer kaplaması sağlanır.

Veri dönüştürme:

Farklı kaynaklardan veri toplanması, verilerin farklı zaman dilimlerine ait olması, güncelleme ile oluşan hatalar, format veya kodlama farklılıkları gibi nedenlerde kullanılacak verilerin uygun ve istenilen forma getirilmesi işlemidir. Bu aşamayı takiben gerekirse tekrar birleştirme işlemi yapılabilir.

4.4.4. Modelleme (Modeling)

Veri madenciliğinde yönteminde benzer problemler için verim alınabilecek birden fazla yöntem bulunmaktadır. Bu sebeple verilerin hazırlanması ve modelin kurulması kademeleri için optimal olduğu düşünülen model tespit edilene kadar bu işlemin tekrarlanması gerekmektedir. Burada ki aşamalar;

- Kullanılacak olan veri madenciliği yöntemi ve algoritmasının belirlenir. (Model Tekniğinin Seçilmesi)
- Veriler hazırlandıktan sonra geçilecek bu aşamada verilerin bir kısmı ile modelin öğrenilmesi, diğer kısmı ile ise kullanılacak modelin geçerliliğinin test edilmesi sağlanır. (Model Test Tasarımı Yapılması)
- Kullanılacak algoritmanın, hazırlanan veri üzerinde çalıştırılması aşamasıdır. (Modelin Kurulması)

- Başarı kriterleri ve test sonuçları kapsamında modelin etkinliğinin değerlendirildiği süreçtir. (Modelin Değerlendirilmesi)

4.4.5. Değerlendirme

Bu aşamada kurulan modelin değerlendirilmesi ve başlangıçta koyulan hedeflerle uyumlu olup olmaması araştırılır. Modelin başlangıç aşamasında koyulan hedefleri karşılayıp karşılamadığı, beklenen verimin alınıp alınmadığı, hangi verilerin kullanıldığı, gelecekte hangi verilerin kullanılmasına ihtiyaç duyulacağı, ek çalışmalara ihtiyaç duyulup duyulmadığı gibi hususlar çerçevesinde değerlendirme yapılır.

4.4.6. Yayma

- Değerlendirme sonucunda alınan sonuçlar için bir yayma stratejisi oluşturulur.
- Projenin takibinin yapılması gerekmekte olup ihtiyaç halinde düzenlemelere gidilebilir.
- Projenin sonuçlarını karar vericilere aktarabilmek için rapor hazırlamak gerekir.
- Çalışmanın gelecekte yenilenmesi gerekip gerekmediği aşaması için gerekli planlama yapılır.

4.5. Veri Madenciliğinde Karşılaşılan Problemler

Pek çok olumlu yönü bahsedilen veri madenciliğinde elbette yüksek miktarda verilerin depolanması nedeniyle bazı sorunlarla da karşılaşmaktadır. Bu sorunlar:

- Artık verilerin çok sayıda olması,
- Boş veri olması (birincil anahtar olup, diğer alanların olmaması),
- Dinamik verilerin mevcut olması ve bu sebeple içeriğin sürekli değişmesi,
- Gürültülü ve kayıp veriler,
- Sınırlı bilgiye sahip olunması,
- Veri tabanı boyutunun çok yüksek olması,
- Tutulan verilerin konuyla uyumsuzluğu,
- Farklı tipteki verilerin olması

gibi sebeplerden kaynaklanmaktadır [35, 41].

4.6. Veri Madenciliğinin Kullanıldığı Alanlar

Veri madenciliğinin yapısı gereği verinin olduğu her alanda veri madenciliğinin kullanılabileceği açıktır. Günümüzde özellikle bilgiye duyulan ihtiyaçtan dolayı birçok alanda veri madenciliği yöntemlerine sıklıkla başvurulmaktadır. Veri madenciliğinin en sık olarak kullanıldığı alanlar; Tıp, Biyoloji, Genetik, Pazarlama, Finans, Bankacılık, Sigortacılık, Veri tabanı Sistemleri, Borsa, Telekomünikasyon, Reklam, Sağlık, Endüstri, İstihbarat, Kamu, Özel sektör, Görselleme, Kriminoloji, Bilim, Mühendislik, Makine öğrenmesi, İstatistik, Matematik ve Astronomi şeklinde sıralanabilir.

5. VERİ MADENCİLİĞİ YÖNTEMLERİ

Karmaşık veri kümelerinin içerisinde yararlı olacak, uygulamaya dönük ve anlamlı bilgilerin çıkarılması veri madenciliği olarak tanımlanmaktadır. Veri madenciliği, analiz yöntemlerinin kullanılması suretiyle mevcut bulunan problemi çözmek, karar almak veya geleceğe yönelik tahmin yapmak için gerekli olan bilgilerin elde edilmesini sağlayan bir araçtır ve aynı zamanda veri tabanlarında gizlenmiş ve fark edilemeyen bağlantı ve kuralları ortaya çıkaran bir yöntemdir [42]. Veri madenciliğinin gelişen teknoloji ile birlikte önem kazanmasıyla pek çok yöntem geliştirilmiştir. Veri madenciliğinde kullanılan pek çok modeli işlevlerine göre üç temel başlık altında kategorize etmek mümkündür [29-43].

- Sınıflama(Classification) ve Regresyon(Regression)
- Kümeleme(Clustering)
- Birliktelik Kuralları(Association Rules)

Sınıflama ve Regresyon modelleri tahmin edici modellerken, kümeleme ve birliktelik kuralları ise tanımlayıcı modellerdir [29-43].

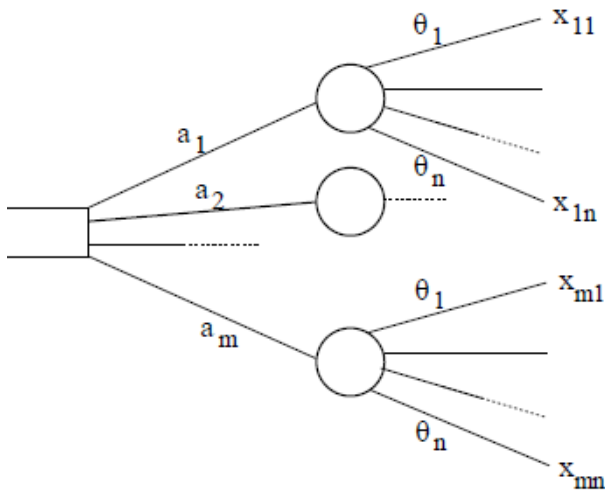
5.1. Sınıflama (Classification) ve Regresyon (Regression)

Sınıflama ve Regresyon yöntemleri veri madenciliğinde tahmin edici modeller olarak bilinmektedir. Çünkü mevcut verileri kullanarak gelecek ile ilgili tahminlerde bulunmak için kullanılırlar. Bu yüzden model kurulurken hangi verilerin kullanılacağına sağlıklı bir şekilde belirlenmesi tahmin edilecek sonuçlar için oldukça önemlidir [31]. Sınıflama ve Regresyon yönteminden elde edilen sonuçların rahatlıkla yorumlanmaya açık olması, kuruluş ve ilk yatırım maliyetlerinin düşüklüğü, veri tabanlarına entegrasyon sürecinin kolay olması, sonuçları itibarıyla yüksek güvenilirlik seviyesi gibi avantajları nedeniyle sınıflama ve regresyon sıklıkla kullanılan yöntemlerdir [42]. Model için kullanılacak veri kümeleri öğrenme ve test verisi olarak ikiye ayrılır. Öğrenme veri kümeleri genellikle modelin oluşturulması sürecinde kullanılırken, test kümesi ise modelin doğruluğunun kontrolü sürecinde kullanılır [29-43]. Sınıflama ve Regresyon yöntemleri için kullanılan temel teknikler ve algoritmalar;

1. Karar ağaçları
 - 1.1. C5.0 Algoritması
 2. Yapay Sinir Ağları
 3. Mesafeye Dayalı Sınıflandırma Algoritmaları
 4. İstatistik Tabanlı Sınıflandırma Algoritmaları
 - 4.1. Bayes Sınıflandırması
 - 4.2. Regresyon
 - 4.3. CHAID Algoritması
- şeklinde sayılabilir.

5.1.1. Karar ağaçları

Veri madenciliği yöntemlerinden birisi olarak kullanılan karar ağaçları en yaygın kullanılan yöntemlerin başında gelmektedir. Karar ağaçları yöntemlerinin temel mantığı verinin farklı gruplara ayrılmasıdır. Karar ağaçları düğüm ve dallardan oluşan ve kolaylıkla yorumlanabilen bir yöntemdir. Karar ağaçlarında kullanılan her bir dalın bir gerçekleşme olasılığı vardır. Bu sayede dallardan köke giderken veya istenilen başka bir noktaya ulaşana kadar bütün olasılıklar hesaplanarak kurallar oluşturulabilir. Akabinde bu kuralların anlamlılık düzeyi üzerinde bir denetleme yapılabilir. Karar ağaçlarını oluşturan düğüm, dal ve yapraklar aşağıdaki gibi gösterilmektedir [7].



Şekil 5.1. Karar ağaçları yapısı

Karar ağaçlarında, karar düğümünde veri üzerinde yapılacak test belirlenir. Yapılacak test sonucunda ağaç dallanmaya başlar. Ancak bu aşamada veri kayıpları yaşamamak adına dallar oluşturulurken eldeki bütün verileri kapsayacak şekilde gerekli sayıda farklı dallar oluşturulmalıdır. Oluşturulan bu dallar testin etkinliğini gösterecektir. Fakat dalın sonunda sınıflandırma işlemi tamamlanamaması durumunda tekrar yeni bir karar düğümü oluşturulması gerekir. Karar düğümünden elde ettiğimiz yeni dalların sonucunda sınıflandırma işleminin tamamlanıp tamamlanmadığını kontrol edilerek ağacın oluşturulmasına devam ettirilir. Eğer dalların oluşumu sonucunda bir sınıflandırma elde edersek yaprağı elde etmiş oluruz. Böylece yaprakların her biri oluşan sınıfların birini temsil edecektir. Karar ağacı algoritmalarında genel olarak aşağıda verilen kodlama kullanır.

```

D:Öğrenme veritabanı
T: Kurulacak ağaç
T=0 //Başlangıçta ağaç boş küme
Dallara ayırma kriterlerini belirle
T= Kök düğümünü belirle
T= dallara ayrılma kurallarına göre kök düğümünü dallara ayır;
Her bir dal için
Do
    Bu düğüme gelecek değişkeni belirle
    If( durma koşuluna ulaşıldı)
        Yaprak ekle ve dur
    Else
Loop

```

Şekil 5.2. Karar ağacı algoritması

Karar ağaçlarının kurulumu aşamasında veri tabanında yer alan verilerin bir kısmı ayrılarak öğrenme işlemi için kullanılır, veri tabanında yer alan diğer veriler ise kurulan ağacın testi için kullanılır. Böylece algoritma ağaç oluşturulurken test edilir ve ağacın istenen düzeyde çalışması durumunda dallanma durdurularak sınıflandırma tamamlanır. Programda tanımladığımız durdurma kriteri ağacımızın hassasiyetini de ortaya koyar. Dallanması geç durdurulan bir ağaç daha fazla dallanacak, bu da ağacımızın daha hassas sonuçlar üretmesine

rağmen programın çalışma ve cevap süresini uzatmak gibi bir olumsuz durumu da beraberinde getirecektir [7].

Ağaç oluştururken yapılacak bir diğer işlem ise budama işlemidir. Budama işlemi ortaya çıkacak sonucu etkilemeyen ve sınıflama işlemine herhangi bir olumlu etkisi olmayan dalların ağaçtan çıkarılması olarak tanımlanabilir. Budama işlemi hem ağacın oluşturulması sırasında hem de ağaç oluşturulduktan sonra yapılabilmektedir. Ağaç budama işlemi için bir anlamlılık düzeyi belirlenmekte ve bu değerin altında kalan dallar budanmaktadır [41].

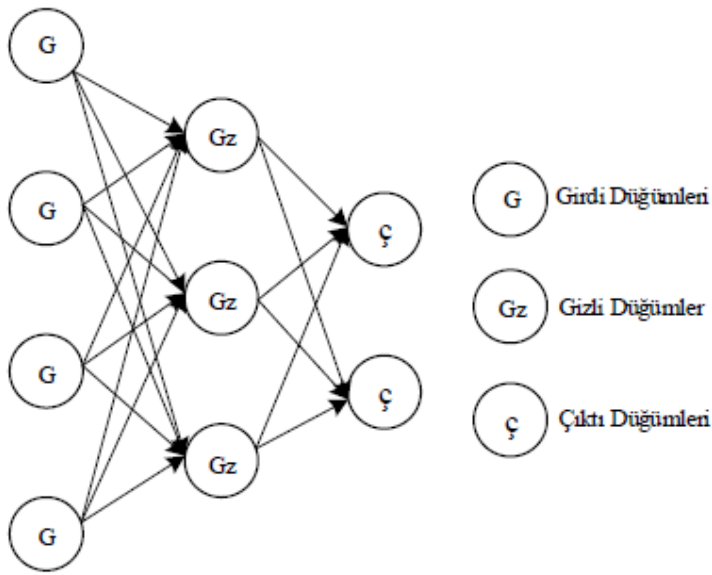
C5.0 algoritması:

C5.0 algoritmasının gelişim süreci incelendiğinde bu alanda kullanılan algoritmalar arasında en bilinen algoritma olan Quinlan'ın ID3 algoritmasına ulaşılmaktadır. ID3 algoritmasının geliştirilmiş hali C4.5 algoritmasıdır [66].

C5.0 algoritması ise C4.5 algoritmasının geliştirilmiş hali olduğu görülmekte ve özellikle büyük veri setlerinin analiz edilmesinde kullanılmaktadır. C5.0 algoritması literatürde, analiz sonuçlarının doğruluğunun artırılması amacıyla boosting algoritmasını kullanılması nedeniyle boosting ağaçları olarak da bilinir. C5.0 algoritmasının C4.5'e göre avantajları arasında çok daha hızlı olması ve hafızayı daha verimli kullanması sayılabilmektedir. Her iki algoritma aynı sonuçları vermiş olsa bile C5.0 algoritması biçim olarak daha düzgün karar ağaçlarının oluşmasına imkan vermektedir [67].

5.1.2. Yapay sinir ağları

Yapay sinir ağları, temeline insan beynindeki sinir sisteminin çalışma mantığını alan elektriksel bir yöntemdir [44]. Yapay Sinir Ağları insan beyni gibi öğrenme yöntemiyle çalışarak, bilgi üretimi, yeni bilgilerin keşfi ve düşünme yeteneği bulunan sistemlerin geliştirilmesi için tasarlanmıştır. Yapay sinir ağları için ayrıca bir metodolojiye ihtiyaç duyulmamaktadır. Çünkü yapay sinir ağları kuralları kendisi üretmekte ve ürettiği kurallardan elde ettiği sonuçları farklı karşılaştırmaktadır. Böylece yapay sinir ağları deneme yanılma yöntemiyle öğrenme faaliyetini gerçekleştirip işlemi nasıl yapacağını belirler [45]. Bu açıdan incelendiğinde bir insan beyni gibi çalışmaktadır. Yapay sinir ağlarının şekli aşağıdaki gibidir.



 ekil 5.3. Yapay sinir ağı yapısı

5.1.3. Mesafeye dayalı sınıflandırma algoritmaları

Mesafeye dayalı sınıflandırma algoritmalarının temelinde mevcut verilerin birbirlerine olan uzaklığı vardır. Burada uzaklık kavramı ile kastedilen verilerin birbirlerine olan benzerliği şeklinde anlaşılmalıdır. Sınıflama işlemi bu uzaklığa ya da benzerliğe g re ger ekle tirilmektedir. Veriler arasındaki uzaklık  l    rken   klid Uzaklığı, Manhattan Uzaklığı, Minkowski Uzaklığı sıklıkla kullanılan y ntemler olarak sıralanabilir. K-en yakın kom u algoritması mesafeye dayalı sınıflandırma algoritmaları arasında en bilinen algoritmadır.

5.1.4. İstatistik tabanlı sınıflandırma algoritmaları

Bayes sınıflandırması:

İstatistik biliminde kullanılan bayes teoremini baz alan bir sınıflandırma y ntemi olan bayes sınıflandırmasının mantığında mevcut verilerdeki kriterlerin sonu   zerindeki etkilerinin olasılık olarak hesaplanması yatmaktadır. Belirsizlikler bulunan durumlar i in karar verme a amasında  ok faydalı bir y ntem olarak g ze  arpmaktadır. Ancak bu sınıflandırma y nteminde de i kenlerin birbirinden bağımsız olarak kabul edilmesi ve dolayısıyla de i kenlerin arasındaki ilişkinin modellenememesi dezavantaj olarak g r lmektedir [46].

Lojistik regresyon:

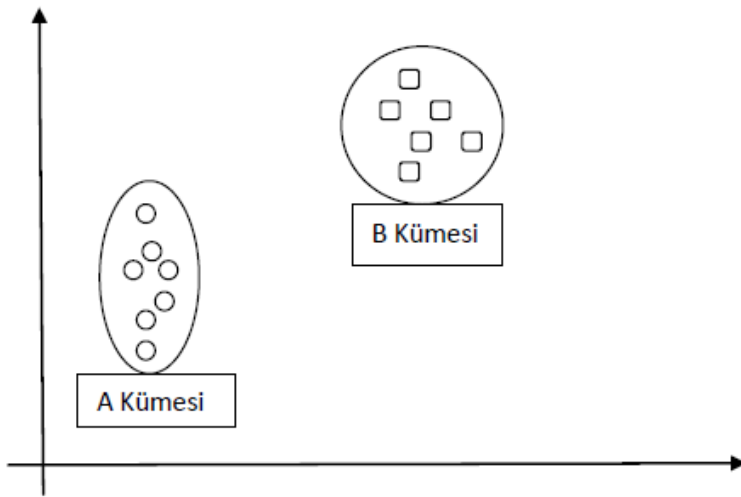
Temelinde bağımlı değişkenlere uygulanan lojit dönüştürme (logit transformation) bulunan lojistik modelinin kullanımı konusundaki ilk çalışmalar 1940 ve 1950’li yıllarda Berkson tarafından yapılmıştır. 1975 yılında ise Koch tarafından eklemeli olasılık modellerinde ortaya çıkan etkileşimin bertaraf edilmesi amacıyla lojistik regresyon modeli önerilmiştir. Lojistik regresyon analizindeki amaçların başında iki ya da daha fazla gruba dair “üyelik” tahmini yapılması gelmektedir. Bu yüzden yöntemin amaçlarından birisinin sınıflandırmanın yapılması iken diğerinin ise bağımlı ve bağımsız değişkenlerin arasında bulunan ilişkilerin tespit edilmesinin olduğu söylenebilir [7, 47].

CHAID algoritması:

İstatistik bilminde kullanılan Ki-kare test yöntemini kullanılması vasıtasıyla bir karar ağacı oluşturan Chaid algoritması bir yandan istatistiğe dayalı bir algoritma iken diğer yandan ise bir karar ağacı algoritması olarak öne çıkmaktadır [7, 41].

5.2. Kümeleme

Veri madenciliğinde tanımlayıcı modeller kategorisinde yer alan kümeleme yönteminde fonksiyonların amacı bir sonuç tahmini yapmak değildir. Bu yöntem ile mevcut verilerimiz arasındaki ilişkilerin gün yüzüne çıkarılması hedeflenmektedir [29]. Elimizde bulunan heterojen bir veri setinden homojen alt veri setleri elde etmek için veriler arasındaki benzerlikler baz alınarak benzer verilerden oluşan kümeler kurulur [31]. Böylece birbirleri ile benzerlik seviyesi yüksek ancak farklı kümelerle benzerlik seviyesi düşük kümeler elde edilmesi amaçlanmaktadır. Sınıflandırma yöntemlerine göre en temel farkı kümeleme yöntemlerinin önceden tanımlanan verileri kullanmamasıdır. Sınıflandırma yöntemlerinde ise tanımlanmış veriler ve bu verilerin aldıkları değerler temel alınarak model oluşturulur [33].



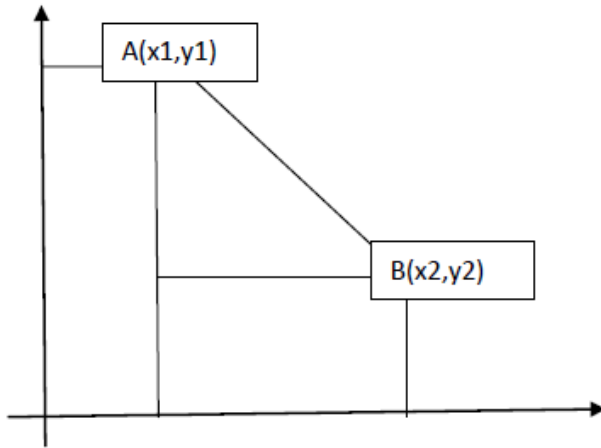
Şekil 5.4. Kümeleme örneği

Kümeleme yöntemlerinde en çok kullanılan uzaklık ölçüm yöntemleri arasında Öklid Uzaklığı, Manhattan Uzaklığı ve Minkowski Uzaklığı gösterilmektedir.

5.2.1. Kümeleme yönteminde kullanılan uzaklık ölçümleri

Öklid uzaklığı:

İki boyutlu uzayda Pisagor teoreminin bir uygulaması şeklinde tanımlanabilir.



Şekil 5.5. Öklid uzaklığı

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (5.1)$$

Manhattan uzaklığı:

Gözlemler arasındaki mutlak uzaklıkların toplanması şeklinde hesaplanmaktadır. Bu denkleme ilişkin hesaplama yöntemi (Eş. 5.2)'de gösterilmektedir.

$$d(i, j) = (|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|) \quad (5.2)$$

Minkowski uzaklığı:

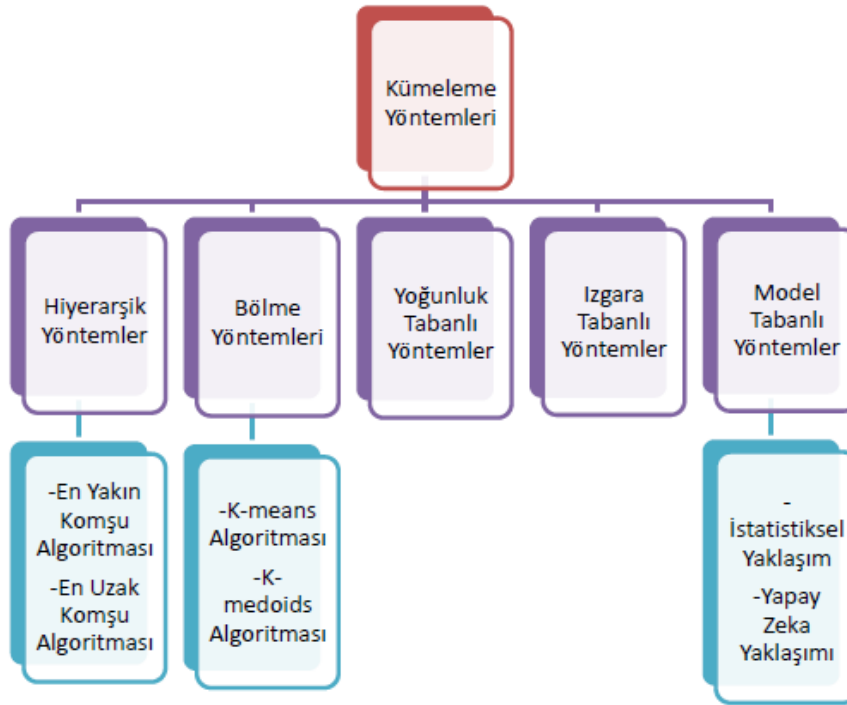
Gözlem değerleri arasındaki uzaklığın hesaplanması mantığına göre çalışan denklem aşağıdaki gibi ifade edilmektedir.

$$d(i, j) = (|x_{i1} - x_{j1}|^m + |x_{i2} - x_{j2}|^m + \dots + |x_{ip} - x_{jp}|^m)^{1/m} \quad (5.3)$$

Kümeleme işlemi için pek çok algoritma kullanılmaktadır. Hangi kümeleme algoritmasının kullanılacağına seçimi noktasında veri tipi ve amacı dikkate alınmalıdır [48]. Kümeleme işlemlerinde kullanılan temel yöntemler aşağıdaki gibi sıralanabilir [29].

1. Hiyerarşik Yöntemler (En yakın komşu algoritması, En uzak komşu algoritması)
2. Bölme Yöntemleri (Partitioning methods)
3. Yoğunluk Tabanlı Yöntemler (Density-based methods)
4. Izgara Tabanlı Yöntemler (Grid-based methods)
5. Model Tabanlı Yöntemler (Model-based methods)

5.2.2. Kümeleme yöntemleri



Şekil 5.6. Kümeleme yöntemleri

Hiyerarşik yöntemler:

Yöntemin mantığında nesnelerin Dendrogram adı verilerin ve ağaç yapısını andıran bir gruplandırma sistemi vardır. Bu yöntemlerde mevcut küme sayısı bilgisine ihtiyaç yoktur. Ama ağaç oluşturulmasında sonlandırmayı sağlayacak bir kritere ihtiyaç duyulur. Ayrı ayrı değerlendirilen kümelerin belirli aşamalardan geçerek bir araya getirilmesini sağlarlar. Hiyerarşik yöntemlerde birden fazla metodoloji vardır. Bu metodolojilerden bazıları aşağıda yer almaktadır.

- En yakın komşu algoritması
- En uzak komşu algoritması

En yakın komşu algoritması:

Başlangıçta tüm veri setleri birer küme olarak kabul edilir. Akabinde bu kümeler bir araya getirilerek yeni kümeler oluşturularak veriler arasındaki uzaklıklar tespit edilir. Minimum

uzaklık seçilerek ve bu uzaklıklara göre ilgili satırlar birleştirilerek yeni bir küme elde edilmesine dayanan yöntemde yeni küme baz alınarak yeniden uzaklıklar hesap edilir.

En uzak komşu algoritması:

Bu yöntemde de en yakın komşu algoritmasına benzeyen bir mantık vardır. Ancak yeni veri kümelerinin elde edilmesinde bu sefer kümeler arasındaki maksimum uzaklıklar esas alınır.

Bölme yöntemleri (Partitioning methods):

Bu yöntemler, n adet nesneden oluşan veri tabanının, giriş parametresi olarak tanımlanan k adet bölüme ($k \leq n$) ayrılmasını esas almaktadır. Veri tabanındaki her bir elemanın fonksiyonlarına göre k adet bölümden birine eklenmesi ile elde edilen bu bölümlerin küme olarak adlandırılması ile çalışmaktadır. Bölme yöntemlerinin kullanıldığı alanlar arasında başlıca K-means ve K-medoids algoritmaları gelmektedir [7].

- K-means (K-ortalamlar) Algoritması
- K-medoids Algoritması

Yoğunluk tabanlı yöntemler (Density-based methods):

Yöntemin kullanılmasında sınır bir yoğunluk değeri belirlenir. Sonrasında bu sınır yoğunluk değerine göre bir yoğunluk fonksiyonu kullanılarak nesnelerin dağılımı yapılır ve sınır yoğunluğun üzerine çıkan bölgeler bir küme olarak adlandırılır [49]. En başarılı kümeleme yöntemlerinden biri olarak gösterilen bu yöntemin avantajları şu şekilde sıralanabilir:

- Kümeleri bulma başarısı çok yüksek olan bir yöntemdir.
- Gürültü ve istisnalardan fazla etkilenmez.
- Tek tarama ile sonuca ulaşır [48].

Izgara tabanlı yöntemler (Grid-based methods):

Yöntemde kare şeklinde hücrelerden meydana gelen ve ızgarayı andıran yapı kullanılarak veri uzayı incelenmektedir. Izgara şeklindeki yapısından dolayı nesne sayısından bağımsız olarak çalışmaktadır [48,50].

Model tabanlı yöntemler (Model-based methods):

Matematiksel bir model ile ifade edilen verilerin kullanıldığı yöntemde iki temel yaklaşım kullanılmaktadır.

- İstatistiksel yaklaşım
- Yapay zekâ yaklaşımı

5.3. Birliktelik Kuralları (Association Rules)

İlk olarak 1993 yılında Agrawal, Imielinski ve Swami tarafından ele alınan birliktelik kuralları analizi problemleri olayların birlikte gerçekleşme durumlarını inceleyerek sonuca ulaşmaya çalışan bir veri madenciliği yöntemidir. Temel olarak büyük veri kümelerinin içerisinde yer alan farklı verilerin arasında sıklıkla tekrar eden ilişkilerin ortaya çıkarılmasını ve bunların özetlenmesini sağlamayı amaçlamaktadır [43]. Veri tabanlarında depolanan veri sayısının gün geçtikçe artması, bu verilerin işlenerek aralarındaki ilişkinin ortaya konulması ve bu bilgilerin faydalı bir hale getirilmesi için işlenmesi ihtiyacını ortaya çıkarmıştır. Bu yöntem hem kolay bir şekilde anlaşılabilmesi hem de yüksek kullanışlı yapısı nedeniyle ekonomi, eğitim, telekomünikasyon, e-ticaret, sağlık ve pazarlama gibi birçok alanda geniş bir yelpazede kullanılmaktadır.

Birliktelik kurallarının oluşturulması aşamasında destek ve güven değerlerinin hesaplanması gerekmektedir. Destek değeri, üzerinde çalışılan bir ilişkinin tespit edilen tüm ilişkiler içerisinde tekrarlanma oranını gösterir. Güven değeri ise üzerinde çalışılan bir ilişkinin başka bir ilişkiyle birlikte gerçekleşme oranını gösterir. Destek ve güven değerlerinin hesaplanmasından sonra yöntemin uygulanabilmesi için bir eşik değere ihtiyaç duyulmaktadır. Destek ve güven değerinin hesaplanan büyüklüğü birliktelik kuralının da gücünü belli edecektir [41]. Birliktelik kurallarının oluşturulması aşamasında 2 temel

faaliyet gerçekleştirilir. Birinci faaliyette sıklıkla tekrar eden öğeler tespit edilir. Tespit edilen öğelerden her birisi daha önceden belirlenmiş minimum destek sayısı kadar tekrar etmelidir. İkinci faaliyette ise sıklıkla tekrar eden öğelerden oluşan birliktelik kuralları oluşturulur. Bu kuralların oluşturulması esnasında tespit edilen minimum destek ve minimum güven değerleri sağlanmalıdır [7,41].

5.3.1. Birliktelik kurallarının belirlenmesinde kullanılan algoritmalar

Birliktelik kurallarının belirlenmesinde kullanılan algoritmalar, sıralı algoritmalar ve paralel algoritmalar olmak üzere 2 ana sınıfta toplanabilmektedir.

Sıralı algoritmalar:

Birliktelik kurallarının belirlenmesi aşamasında pek çok verinin sıralı olarak tanımlandığı ve bu şekilde veri tabanlarında tutulduğu varsayılmaktadır. Bu sıralı yapının veri kümelerinin oluşturulmasında sağladığı kolaylıklardan faydalanan bir yaklaşımdır.

AIS algoritması:

Birliktelik kurallarının belirlenmesinde kullanılan ilk algoritma olan AIS Algoritması Agrawal, Imielinski ve Swami tarafından 1993 yılında geliştirilmiştir. AIS algoritmasının öncelikle anlamlı kuralları ortaya çıkarmayı amaçlamaktır. Veri tabanı tarama işleminin tamamlanmasından sonra, daha önceki taramada rastlanan ve sıklıkla karşılaşılan nesnelerden oluşan ortak nesne kümeleri tespit edilir [51]. Minimum desteği yakalayan yahut aşan nesne kümeleri, yaygın nesne kümeleri olarak kabul edilir [52,53]. Bu işlemler artık yeni nesne kümeleri oluşmayana kadar devam ettirilir.

SETM algoritması:

Nesne kümelerinin üretilmesinde SQL kullanılmasını amaçlayan SETM Algoritması 1995 yılında Houtsma tarafından önerilmiştir. AIS algoritmasında olduğu gibi SETM algoritmasında da veri tabanı üzerinde algoritma adımları birden fazla defa tekrarlanır. İlk aşamada ayrı ayrı her bir nesne için destek sayısı belirlenir ve veri tabanındaki büyüklüklerine göre sıralanır. Sonraki aşamada ise son oluşturulan yaygın nesne

kümelerinin genişletilmesi suretiyle aday nesne kümeleri oluşturulur. Bu aşamalara yeni nesne kümesi oluşmadığında son verilir [54].

Apriori algoritması:

Agrawall ve Srikant tarafından 1994 yılında geliştirilen Apriori algoritması birliktelik kurallarına ilişkin algoritmalar arasında en sık kullanılan yöntemdir. Algoritma temel olarak kullandığı bilgileri bir önceki aşamadan elde etmesi nedeniyle (prior=önceki) ismini bu ifadeden almaktadır [64]. Algoritmanın temelinde tekrar eden bir yapı olması nedeniyle veri tabanında sıklıkla tekrarlanan öğelerin kümelerinin tespit edilmesinde kullanılmaktadır. Çalışma mantığında, eğer k-öge kümesi (k adet elamana sahip öge kümesi) minimum destek kriterini sağlıyorsa, bu kümenin alt kümeleri de minimum destek değerini sağlayacaktır [55].

Birliktelik kurallarında oluşan ilişkilerin gücü destek ve güven değerleri dikkate alınarak hesaplanırlar. Destek(Support) değeri, öğeler arasında bulunan ilişkinin fazlalığını sembolize etmektedir. Örneklendirmek gerekirse A ve B ürünleri birbirinden farklı olmak üzere;

A ürünü için hesaplanacak destek değeri, A ürününün tüm ürünler içerisindeki oranıdır.

$$\text{Destek}(A) = \frac{\text{A ürününün sayısı}}{\text{Toplam alışveriş sayısı}} \quad (5.4)$$

A ve B ürünleri için destek değeri ise, A ve B ürünlerinin tüm alışverişlerde birlikte bulunma olasılığıdır.

$$\text{Destek}(A, B) = \frac{(A,B)\text{ürünlerinin sayısı}}{\text{Toplam alışveriş sayısı}} \quad (5.5)$$

Güven değeri hesaplanacak ise B ürününün hangi olasılıkla A ürünü ile birlikte olacağını ifade eder.

$$\text{Güven}(A, B) = \frac{(A,B)\text{ürünlerini içeren alışveriş sayısı}}{A \text{ ürününü içeren alışveriş sayısı}} \quad (5.6)$$

şeklinde ifade edilebileceği gibi aynı zamanda

$$\text{Güven}(A \Rightarrow B) = \frac{\text{Destek}(A,B)}{\text{Destek}(A)} \quad (5.7)$$

şeklinde de gösterilebilir.

Apriori algoritmasının aşamalarını temel olarak sıralamak gerekirse;

1. Minimum destek sayısı ve minimum güven değerinin belirlenmesi,
2. Öge kümeler içerisindeki öğelerin destek değerlerinin bulunması,
3. Minimum destek değerinden daha küçük destek değerine sahip öğelerin devre dışı bırakılması,
4. Oluşturulan tekli birliktelikler işleme alınarak ikili birlikteliklerin oluşturulması,
5. Minimum destek değerinden daha küçük olan öge kümelerinin çıkartılması,
6. Üçlü birlikteliklerin oluşturulması,
7. Üçlü birlikteliklerden minimum destek değerini aşanların çıkartılması,
8. Üçlü birliktelik gruplarından birliktelik kurallarının çıkarılması,
9. Bu adımların yeni birliktelikler ortaya çıkmayana dek devam ettirilmesi.

Yukarıdaki adımlar takip edilerek oluşturulan apriori algoritmasının performansı destek değeri ve işlem sayısı miktarına göre belirlenmektedir. Destek değerinin düşük olması durumunda oluşacak ilişkilerin küme sayısında artış, destek değerinin yüksek olması durumunda ise oluşacak ilişkilerin küme sayısında azalış gözlenecektir [56]. Apriori algoritmasının tekrarlı yapısından dolayı tüm veriler taranacak ve işlem sayısı arttıkça analiz süresi uzayacaktır.

OCD (Off-line candidate determination) algoritması:

1994 yılında Mannila tarafından geliştirilen OCD Algoritması düşük boyutlu veri tabanlarında daha etkin performans göstermektedir. Veri tabanlarının boyutlarının düşük olması nedeniyle de algoritmanın hızı oldukça yüksektir. Ancak veri tabanlarının boyutu büyüdükçe algoritmanın performansı düşer ve fazlaca zaman kaybı yaşanmaktadır. OCD algoritması, önceki aşamalardan elde edilen bilgileri kullanarak gereksiz aday kümelerin elenmesi mantığıyla işlem yapmaktadır. OCD algoritmasının en önemli farkı önceki aşamalardaki tüm bilgileri koruyarak aday kümelerin budanması için bu bilgileri

kullanmasıdır. Dolayısıyla bu fazla bilginin korunması algoritmanın küçük boyutlu veri tabanlarında daha başarılı olmasını sağlamaktadır [57].

Partitioning algoritması:

Partitioning algoritması 1995 yılında Savasere tarafından geliştirilmiştir. Algoritmanın diğer algoritmalara göre en temel farklılığı veri tabanlarını 2 defa taraması, dolayısıyla diğer algoritmalara kıyasla daha az tarama gerçekleştirmesidir. Veri tabanının küçük bölümlere ayrılması ve ilk taramada her bir bölüm için yaygın nesne kümelerinin elde edilmesi mantığı üzerine çalışmaktadır [58]. İkinci taramada ise veri tabanındaki her bir bölümde bulunan yaygın nesne kümelerinin birleşimi adaylar olarak kullanılırlar ve bütün nesne kümelerinin tespit edilebilmesi için veri tabanının tamamı tekrar taranmaktadır. Partitioning algoritmasından yüksek bir verim alınabilmesi için mevcut verilerimizin homojen bir şekilde dağılmış olması gerekmektedir. Çünkü bir nesne kümesi her bir bölümde eşit sayıda varsa, ikinci taramada sayılacak çoğu nesne kümesinin sayısı yüksek olacaktır [53].

Sampling (Örnekleme) algoritması:

Toivonen tarafından 1996 yılında geliştirilen Sampling algoritması, veri tabanını en iyi alternatifte bir kez, en kötü alternatifte ise iki kez tarar. Sampling algoritması da seviye mantığı algoritmasına göre işlem yapmaktadır. Algoritmanın ilk adımında ana belleğe işlenmek üzere veri tabanından bir örnek seçilir [59]. Bu seçilen örnek veri tabanında doğrulanabilen adayları elde edebilmek için kullanılır. Adaylar oluşturulduktan sonra veri tabanının tamamı aday sayısının belirlenmesi için bir defa taranır. Eğer elde edilen nesne kümelerinin tamamı, örnek nesne kümesindeki kümelerin bulunduğunu teyit eder ise algoritma sonlandırılır. Örnek nesne kümesindeki kümelerin tamamının bulunmaması durumunda ise veri tabanı tekrar taranır ve tüm nesne kümeleri bulununca algoritma sonlandırılır [53].

DIC (Dynamic itemset counting) algoritması:

DIC algoritmasının temelinde nesne kümelerinin ölçülmesi ve sayılması ilkesi yer almaktadır. Veri tabanı algoritmaya aralıklı olarak gösterilip, oluşturulan aralıklar daha sonra taranmaktadır. İlk aralık taranırken 1. nesne kümeleri elde edilir ve bu küme sayılır.

Birinci aralık sonunda yaygın nesnelerden oluşan 2. nesne kümeleri elde edilir ve akabinde yapılan tarama işleminde 1. ve 2. nesne kümeleri sayılır. İkinci aralık sonunda 3. nesne kümeleri elde edilir ve birlikte sayma işlemine toplanarak devam edilir. Özetle n aralık için, yaygın nesnelerden oluşan $n+1$ tane küme oluşturulur ve birlikte sayma işlemine devam edilir. DIC Algoritmasında da aynen Partition algoritmasında olduğu gibi homojen bir dağılıma ihtiyaç duyulmaktadır.

CARMA (Continuous association rule mining algorithm) algoritması:

Nesne kümelerinin tespiti ve hesaplanması gibi işlemleri çevrimiçi (online) gerçekleştiren CARMA algoritması, veri tabanındaki ilk taramada kullanıcıya minimum destek ve minimum güven parametrelerini değiştirme imkanı sağlayarak, online birliktelik kuralları çıkarmanın yolunu açmaktadır. CARMA algoritmasında da DIC algoritmasına benzer şekilde ilk tarama sırasında nesne kümeleri oluşturulur ve ikinci taramada ise nesne kümelerinin sayılması işlemi gerçekleştirilir [60,61]. Ancak CARMA algoritmasında DIC algoritmasından farklı olarak hareketler gerçekleşirken nesne kümeleri oluşturulur.

FP-Growth(Frequent pattern growth) algoritması:

FP-Growth algoritmasının çalışma mantığına göre yaygın nesne kümelerinin oluşturulması aşamasında adaylar oluşturulmayıp ilk aşamada veri tabanı, yaygın nesneleri tanımlanması için FP-Tree(Frequent Pattern Tree) denilen ağaç yapısına dönüştürülür. Nesne kümelerinin arasında oluşan birlikteliklere ilişkin bilgiler FP-Tree ağacında depolanmakta olup sonraki aşamada ise FP-Tree şeklinde oluşturulan ve her biri yaygın bir nesne ile ilişkilendirilmiş koşullu veri tabanlarına dönüştürülür [62].

FP-Growth algoritmasında veri tabanının ilk taramasında nesne kümeleri tespit edilir ve yaygın nesne kümeleri destek sayısı kriterine göre büyükten küçüğe doğru sıralanır. Sonraki aşamada veri tabanı tekrar taranarak FP-Tree oluşturulur. FP-Tree ağacında yaygın olmayan nesne kümeleri yer almaz ve destek sayısı daha büyük olan nesne kümeleri köke daha yakın şekilde konumlanır [53].

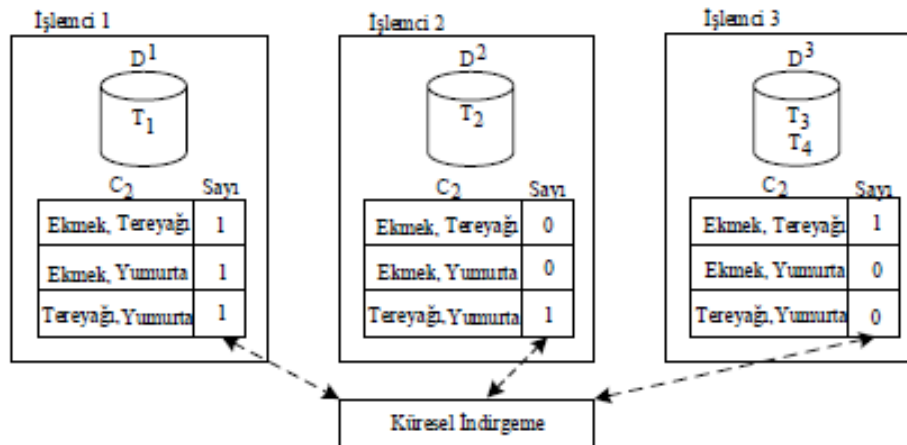
GRI algoritması:

Generalized Rules Induction (GRI) Algoritması, veriler üzerinde düğümler arasındaki ilişki kurallarını keşfeder. GRI algoritması, destek ve güven kurallarını dikkate alarak en kuvvetli bilgi bağımlı seçecek şekilde ilerlemektedir [65].

Paralel ve dağıtılmış (distributed) algoritmalar:

Genel olarak nesne kümelerinin tespiti aşamasında paralelleştirilmesi mantığı üzerine çalışan algoritmalar temeline çoğunlukla apriori algoritmasını alır. Genel anlamda veri paralelleştirilmesi (data parallelism) veya görev paralelleştirilmesi (task parallelism) şeklinde çalışılmaktadır. Veri paralelleştirilme modelinde tüm düğümler aynı aday kümeleri sayar. Görev paralelleştirme modelinde ise aday kümeler bölümlere ayrılarak işlemciler arasında paylaşılır ve her bir düğüm farklı aday kümesini sayar [63].

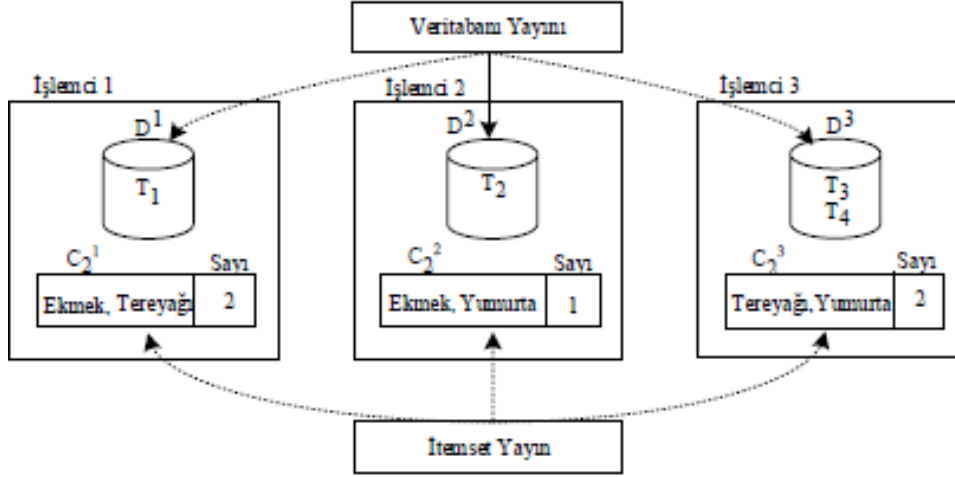
Veri paralelleştirme modeli genel olarak aşağıdaki gibi gösterilmekte olup CD(Count Distribution) Algoritması, PDM(Parallel Data Mining) Algoritması, DMA(Distributed Mining Algorithm) Algoritması, CCPD (Common Candidate Partitioned Database) Algoritması gibi algoritmalar bu modele örnek olarak gösterilebilmektedir.



Şekil 5.7. Veri paralelleştirme modeli

Görev paralelleştirme modeli ise genel olarak aşağıdaki gibi gösterilmekte olup DD(Data Distribution) Algoritması, IDD (Intelligent Data Distribution) Algoritması, HPA (Hash-

based Parallel mining of Association rules) Algoritması, PAR (Parallel Association Rules) Algoritması gibi algoritmalar bu modele örnek olarak gösterilebilmektedir.



Şekil 5.8. Görev paralelleştirme modeli

Bu modeller dışında veri paralelleştirme veya görev paralalleştirme modelleri arasında sınıflandırılmayan ancak her iki modelin karışımı niteliği taşıyan algoritmalar için de Candidate Distribution Algoritması, SH (Skew Handling) Algoritması, HD (Hybrid Distribution) Algoritması gibi örnekler verilebilmektedir.

6. UYGULAMA

Veri madenciliğinde kullanılacak verinin analiz edilebilmesi için uygun formata getirilerek hazırlanması büyük önem taşımaktadır. Verilerde bulunan tüm parametreler ne derece ortak bir forma dönüştürülürse analizden alınacak sonuçlar da o derece doğru olmaktadır. Bu amaçla, SPSS Clementine (IBM Modeler) isimli bilgisayar programı yardımıyla gerekli analiz yöntemleri uygulanmadan önce veriler CRISP-DM adımları izlenerek hazır hale getirilmiş ve en kuvvetli ilişkilere ulaşmak hedeflenmiştir.

İş kazalarının meydana gelmesinde birbiriyle ilişkili faktörlerin belirlenmesi amaçlanan bu çalışmada; 2016 ve 2017 yıllarında meydana gelen 13.079 kaza verisi incelenmiş, her bir kaza için 20 parametre üzerinde analizler gerçekleştirilmiştir. Söz konusu analizlerde kullanılan bu parametreler Çizelge 6.1.'de açıklamaları ile birlikte verilmektedir.

Çizelge 6.1. Kullanılan parametre açıklamaları

Parametre Tanımı	SPSS Açıklama
Sehir	Kazanın meydana geldiği şehir
Calisansayisi	Kazanın meydana geldiği işyerindeki çalışan sayısı
Nace	Kazanın meydana geldiği iş kolu
Genelfaalkodu	Kazanın meydana geldiği işyerinin genel faaliyet alanı
Calismacevrekodu	Kazanın meydana geldiği alanın çevre tanımı
Faalkoduozel	Kaza sırasında yapılan işin detaylı tanımı
Kazaay	Kazanın meydana geldiği ay
Kazasaat	Kazanın meydana geldiği saat dilimi
Saatisbasi	Kazalının kaza tarihinde işbaşı yaptığı saat dilimi
Cinsiyet	Kazalının cinsiyeti
Toptecrube	Kazalının toplam tecrübesi
Sontecrube	Kazalının o işyerindeki tecrübesi
Durumogrenim	Kazalının öğrenim durumu
Yaskodu	Kazalının yaş dilimi
Egitimisg	Kazalının iş sağlığı ve güvenliği eğitimi alma durumu
Egitimmesleki	Kazalının mesleki eğitim alma durumu

Çizelge 6.1. (devam) Kullanılan parametre açıklamaları

Cisguzmani	İşyerinde iş güvenliği uzmanı istihdam edilme durumu
Cisyerihekimi	İşyerinde işyeri hekimi istihdam edilme durumu
Kazasebepkodu	Kaza sebebi kodu
Kazadurum	Kaza sonucunda ortaya çıkan sonuç

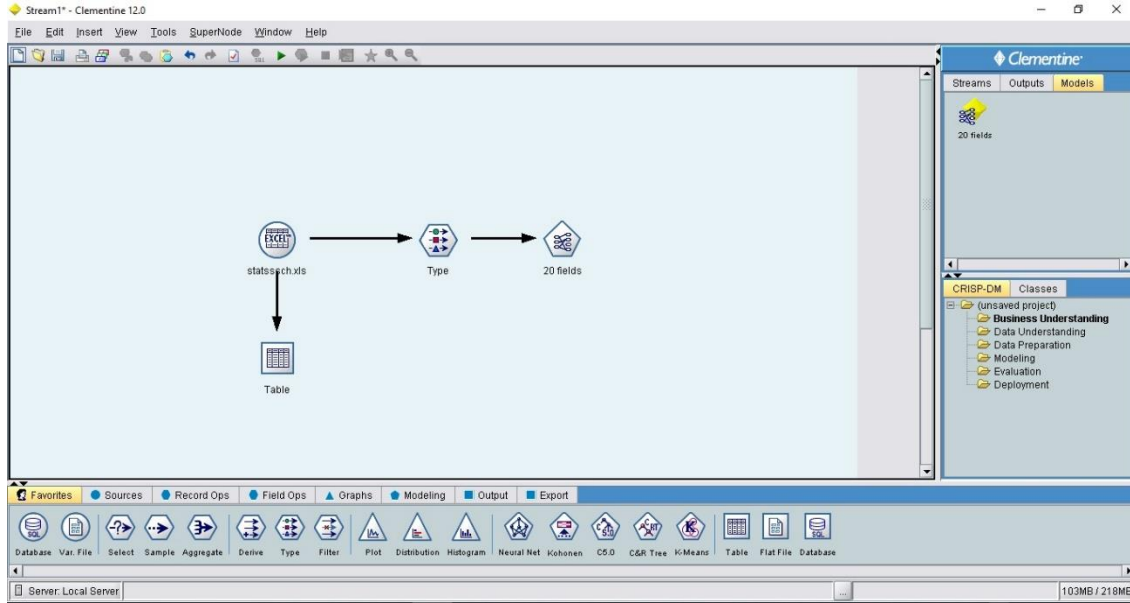
Çalışmanın başında elde bulunan veri yığını, uluslararası bir standart olarak kabul edilebilecek yetkinliğe sahip, CRISP-DM (Cross Industry Process for Data Mining) baz alınarak çeşitli aşamalardan geçirilerek analiz için hazır hale getirilmiştir.

Ham veri halinde bulunan veriler öncelikle veri setini tanımlama, verinin seçilmesi, verinin temizlenmesi, verinin bütünleştirilmesi, verinin indirgenmesi ve verinin dönüştürülmesi adımları izlenerek analiz için hazır hale getirilmiştir. Elde bulunan veri istatistikleri; analizin yapılacağı bilgisayar programında daha efektif bir sonuç alınabilmesi ve kullanılan analiz yöntemlerinin yapısı gereği numerik hale getirilmiş, analizin akışını etkileyebilecek kayıp verilerden temizlenmiş, daha anlamlı sonuçlar elde edilebilmesi için gruplandırılarak son haline erişirilmiştir. Bu aşamada kullanılan parametre gruplandırmasına ilişkin bilgiler EK-1’de sunulmaktadır.

Verilerin analiz edilmesi için pek çok farklı model ve algoritma bulunmakta olup, literatür üzerinde yapılan araştırmalar neticesinde sıklıkla kullanılan ve çözülmeye çalışılan problem ile veri yapısına uygun model ve algoritmalarından 3 tanesi ile analiz gerçekleştirilmesi ve sonuçlarının karşılaştırılması ile amaca daha uygun sonuçlar alınması yoluna gidilmiştir.

Bu noktada birliktelik kuralı analizi yöntemlerinden GRI ve Apriori algoritmaları ile karar ağacı modellerinden C5.0 modeli analiz çalışmalarında kullanılmıştır.

Verilerin analize hazır hale getirilmesi sonrasında, istatistiksel analizler için yaygın olarak kullanılan “SPSS Clementine” programı üzerinde gerekli modelleme kurgusu oluşturulmuştur.



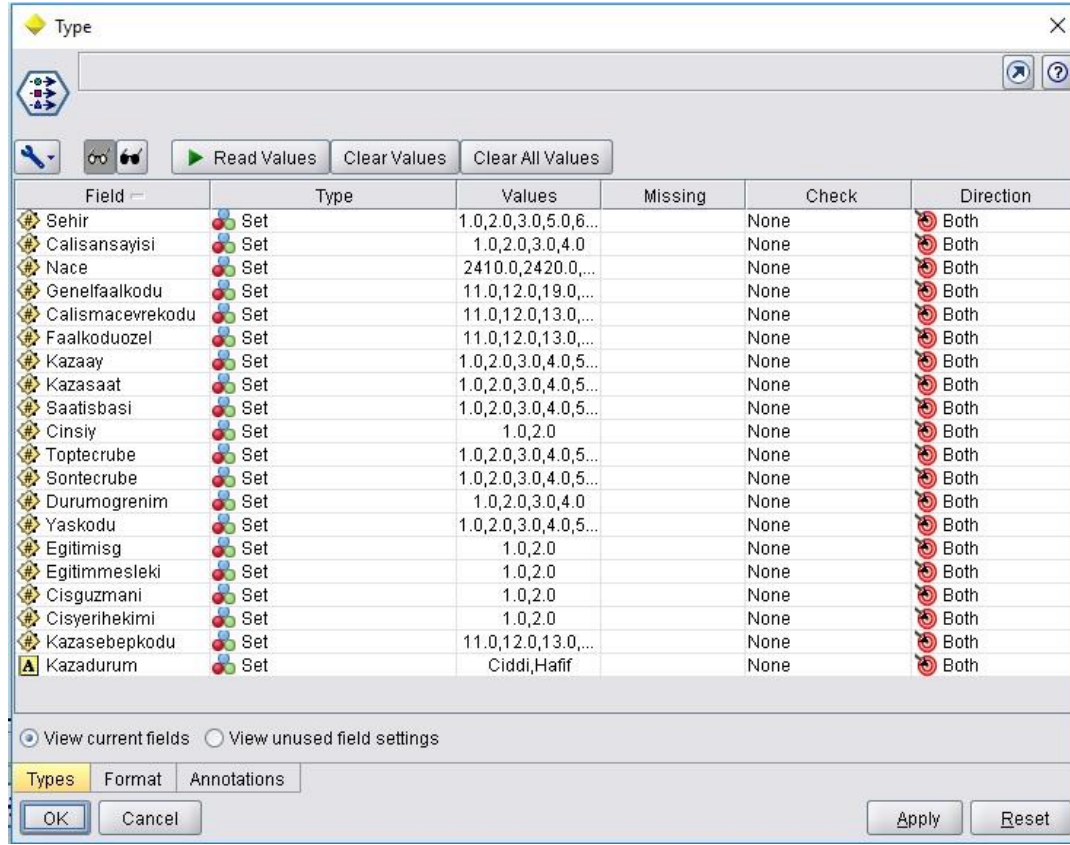
Şekil 6.1. SPSS Clementine programı üzerindeki modelleme gösterimi

Modelleme kurgusu hazırlandıktan sonra programa veriler yüklenerek tüm kayıtlar üzerinden genel tablo hazırlanmıştır.

	Sehir	Calisansayisi	Nace	GenelFaaliyodu	Calisma...	Faalkoduozel	kazaa...	Kazasaat	Saatlisabasi	Cinsiy	Topte...	Sonta...	Durum...	Yaskodu	Egitimsg	Egitim...	Cisguznani	Cisyerih...	Kazasecepkodu	Kazadurum
1	1.030	4000	2451...	11.0C0	11.0C0	59.000	10.000	4.800	3.000	1.0C0	4.000	3.000	2.030	3.030	1.000	1.000	2.000	2.000	53.000	Ciddi
2	1.030	4000	2451...	99.0C0	11.0C0	99.000	1.000	2.800	3.000	1.0C0	4.000	3.000	2.030	3.030	1.000	1.000	1.000	1.000	52.000	Ciddi
3	1.030	4000	2451...	11.0C0	11.0C0	99.000	4.000	3.000	3.000	1.0C0	5.000	1.000	2.030	3.030	1.000	1.000	1.000	2.000	51.000	Ciddi
4	1.030	3000	2410...	11.0C0	11.0C0	53.000	1.000	3.000	3.000	1.0C0	5.000	3.000	1.030	3.030	1.000	1.000	1.000	1.000	62.000	Ciddi
5	1.030	3000	2453...	11.0C0	11.0C0	21.000	6.000	2.000	1.000	1.0C0	5.000	3.000	2.030	3.030	1.000	1.000	2.000	2.000	43.000	Hafr
6	1.030	3000	2453...	11.0C0	11.0C0	51.000	2.000	3.000	3.000	1.0C0	5.000	2.000	3.030	3.030	1.000	1.000	2.000	2.000	22.000	Ciddi
7	1.030	4000	2434...	99.0C0	11.0C0	99.000	8.000	3.000	2.000	1.0C0	4.000	1.000	4.030	3.030	1.000	1.000	2.000	2.000	11.000	Ciddi
8	1.030	4000	2442...	11.0C0	11.0C0	41.000	7.000	4.000	3.000	1.0C0	5.000	1.000	1.030	3.030	1.000	1.000	2.000	1.000	41.000	Ciddi
9	1.030	4000	2442...	11.0C0	11.0C0	70.000	3.000	3.000	3.000	1.0C0	5.000	3.000	2.030	3.030	1.000	1.000	2.000	1.000	22.000	Ciddi
10	1.030	3000	2431...	11.0C0	11.0C0	11.000	10.000	3.000	3.000	1.0C0	3.000	1.000	4.030	2.030	1.000	1.000	2.000	2.000	31.000	Hafr
11	1.030	3000	2431...	11.0C0	11.0C0	32.000	10.000	3.000	3.000	1.0C0	2.000	1.000	3.030	2.030	1.000	1.000	2.000	2.000	31.000	Ciddi
12	1.030	4000	2453...	11.0C0	11.0C0	12.000	2.000	2.000	3.000	1.0C0	1.000	1.000	1.030	3.030	1.000	1.000	1.000	2.000	63.000	Ciddi
13	1.030	4000	2453...	11.0C0	11.0C0	99.000	1.000	3.000	3.000	1.0C0	4.000	1.000	1.030	3.030	1.000	1.000	1.000	2.000	64.000	Hafr
14	1.030	4000	2453...	11.0C0	11.0C0	12.000	9.000	1.000	3.000	1.0C0	2.000	2.000	2.030	3.030	1.000	1.000	1.000	2.000	72.000	Hafr
15	1.030	4000	2453...	11.0C0	11.0C0	99.000	9.000	1.000	3.000	1.0C0	4.000	4.000	1.030	4.030	1.000	1.000	1.000	2.000	72.000	Hafr
16	1.030	4000	2453...	11.0C0	11.0C0	12.000	9.000	3.000	3.000	1.0C0	2.000	1.000	1.030	3.030	1.000	1.000	1.000	2.000	63.000	Hafr
17	1.030	4000	2453...	11.0C0	11.0C0	70.000	2.000	3.000	3.000	1.0C0	3.000	1.000	3.030	3.030	1.000	2.000	2.000	2.000	64.000	Hafr
18	1.030	4000	2453...	11.0C0	11.0C0	12.000	2.000	3.000	3.000	1.0C0	3.000	2.000	3.030	3.030	1.000	1.000	2.000	2.000	64.000	Ciddi
19	1.030	2.000	2453...	12.0C0	13.0C0	53.000	12.000	3.000	3.000	1.0C0	5.000	4.000	1.030	3.030	2.000	1.000	2.000	2.000	51.000	Ciddi
20	1.030	3000	2451...	99.0C0	11.0C0	99.000	10.000	5.000	3.000	1.0C0	5.000	1.000	2.030	4.030	1.000	1.000	2.000	2.000	44.000	Ciddi
21	1.030	2.000	2433...	11.0C0	11.0C0	69.000	4.000	6.000	5.000	1.0C0	5.000	3.000	4.030	3.030	1.000	1.000	2.000	2.000	44.000	Ciddi
22	6.030	2.000	2451...	11.0C0	11.0C0	70.000	6.000	2.000	3.000	1.0C0	4.000	2.000	2.030	3.030	1.000	1.000	1.000	2.000	73.000	Hafr
23	6.030	4000	2451...	11.0C0	11.0C0	42.000	1.000	6.000	3.000	1.0C0	4.000	4.000	1.030	3.030	1.000	1.000	2.000	2.000	52.000	Ciddi
24	6.030	4000	2451...	11.0C0	11.0C0	42.000	12.000	2.000	2.000	1.0C0	4.000	2.000	2.030	3.030	1.000	1.000	2.000	2.000	43.000	Ciddi
25	6.030	4000	2451...	11.0C0	11.0C0	41.000	1.000	3.000	2.000	1.0C0	4.000	1.000	3.030	3.030	1.000	1.000	2.000	2.000	64.000	Ciddi
26	6.030	4000	2451...	11.0C0	11.0C0	46.000	11.000	2.000	3.000	1.0C0	5.000	3.000	1.030	3.030	1.000	1.000	2.000	2.000	53.000	Hafr
27	6.030	4000	2451...	11.0C0	11.0C0	21.000	10.000	3.000	2.000	1.0C0	4.000	2.000	1.030	3.030	1.000	1.000	2.000	2.000	43.000	Hafr
28	6.030	4000	2451...	11.0C0	11.0C0	99.000	8.000	1.000	3.000	1.0C0	5.000	5.000	1.030	3.030	1.000	1.000	1.000	2.000	64.000	Ciddi
29	6.030	4000	2451...	99.0C0	22.0C0	99.000	9.000	3.000	3.000	1.0C0	2.000	1.000	2.030	2.030	1.000	1.000	2.000	2.000	64.000	Hafr
30	6.030	4000	2451...	11.0C0	11.0C0	41.000	10.000	6.000	3.000	1.0C0	3.000	3.000	2.030	3.030	1.000	1.000	2.000	2.000	64.000	Ciddi
31	6.030	4000	2451...	11.0C0	11.0C0	41.000	10.000	2.000	3.000	1.0C0	4.000	2.000	1.030	3.030	1.000	1.000	2.000	2.000	43.000	Hafr
32	6.030	4000	2451...	11.0C0	11.0C0	21.000	9.000	4.000	2.000	1.0C0	5.000	5.000	1.030	3.030	1.000	1.000	2.000	2.000	43.000	Hafr
33	6.030	4000	2451...	11.0C0	11.0C0	13.000	9.000	1.000	3.000	1.0C0	4.000	3.000	3.030	3.030	1.000	1.000	2.000	2.000	43.000	Hafr
34	6.030	4000	2451...	11.0C0	11.0C0	61.000	9.000	3.000	3.000	2.0C0	3.000	1.000	4.030	3.030	1.000	1.000	2.000	2.000	64.000	Hafr
35	6.030	4000	2451...	11.0C0	11.0C0	21.000	9.000	4.000	2.000	1.0C0	4.000	3.000	3.030	3.030	1.000	1.000	2.000	2.000	43.000	Hafr
36	6.030	4000	2451...	11.0C0	11.0C0	61.000	10.000	4.000	2.000	1.0C0	3.000	3.000	1.030	3.030	1.000	1.000	2.000	2.000	64.000	Ciddi
37	6.030	4000	2451...	11.0C0	11.0C0	21.000	10.000	3.000	2.000	1.0C0	4.000	3.000	2.030	3.030	1.000	1.000	2.000	2.000	64.000	Hafr

Şekil 6.2. SPSS Clementine programında veri tablosu görünümü

Verilerin bilgisayar programına yüklenmesini takiben, programın veri yapısını anlayabilmesi amacıyla veri tiplerine ilişkin gerekli düzenlemeler gerçekleştirilmiştir.



Şekil 6.3. SPSS Clementine programında veri tipleri düzenlemesi

6.1. Apriori Algoritması Analizi

Apriori algoritması, Agrawal ve Srikant tarafından 1994 yılında geliştirilmiş olup; birliktelik kurallarına ilişkin algoritmalar arasında en sık kullanılan yöntemdir. Algoritma temel olarak kullandığı bilgileri bir önceki aşamadan elde etmesi nedeniyle (prior=önceki) ismini bu ifadeden almaktadır [64].

“SPSS Clementine” programı üzerinde apriori algoritması vasıtasıyla gerçekleştirilen birliktelik kuralları analizi sonucunda “Ciddi” ve “Hafif” sonuçlu kazalar için en kuvvetli birliktelikleri oluşturan analiz sonuçları Şekil 6.4. ve Şekil 6.5’de gösterilmektedir. Daha kapsamlı örnek analiz sonuçları Ek-2’de sunulmuştur.

20 fields

File Generate

Sort by: Confidence % 4696 of 4696

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Sehir = 38.0 Saatisbasi = 2.0 Nace = 2562.0 Cisyerihekimi = 1.0 Cinsiy = 1.0	1,002	99,237
Kazadurum = Hafif	Sehir = 38.0 Saatisbasi = 2.0 Nace = 2562.0 Cisyerihekimi = 1.0 Calisansayisi = 4.0 Cinsiy = 1.0	1,002	99,237
Kazadurum = Hafif	Sehir = 38.0 Saatisbasi = 2.0 Nace = 2562.0 Cisyerihekimi = 1.0 Egitimmesleki = 1.0 Cinsiy = 1.0	1,002	99,237
Kazadurum = Hafif	Sehir = 38.0 Saatisbasi = 2.0 Nace = 2562.0 Cisyerihekimi = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,002	99,237
Kazadurum = Hafif	Sehir = 38.0 Saatisbasi = 2.0 Nace = 2562.0 Cisyerihekimi = 1.0 Calisansayisi = 4.0	1,002	99,237

Model Summary Annotations

OK

Şekil 6.4. SPSS Clementine üzerinde apriori algoritması analiz örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 80.0 Nace = 2420.0 Egitimmesleki = 1.0 Cinsiy = 1.0	1,002	81,679
Kazadurum = Ciddi	Sehir = 80.0 Nace = 2420.0 Egitimmesleki = 1.0 Egitimisg = 1.0	1,002	81,679
Kazadurum = Ciddi	Sehir = 80.0 Nace = 2420.0 Cinsiy = 1.0 Egitimisg = 1.0	1,002	81,679
Kazadurum = Ciddi	Sehir = 80.0 Nace = 2420.0 Egitimmesleki = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,002	81,679
Kazadurum = Ciddi	Sehir = 78.0 Nace = 2410.0 Cisyerihekimi = 1.0 Yaskodu = 3.0	1,185	78,71
Kazadurum = Ciddi	Sehir = 78.0 Nace = 2410.0 Cisyerihekimi = 1.0 Yaskodu = 3.0 Cinsiy = 1.0	1,185	78,71
Kazadurum = Ciddi	Sehir = 78.0 Nace = 2410.0		

Şekil 6.5. SPSS Clementine üzerinde apriori algoritması analiz örnekleri

Apriori algoritması kullanılarak yapılan analizler sonucu oluşan birliktelik kuralları örnek olarak incelendiğinde;

- 38 plaka kodlu Kayseri ilinde, 2562 Nace kodlu “Metallerin makinede işlenmesi ve şekil verilmesi” işinde, cinsiyeti erkek olan çalışanların işyerlerinde işyeri hekimi istihdam edilmesi durumunda işbaşı yaptıktan 2 saat sonra % 99,237 güven ve %1,002 destek oranıyla Hafif kazalar geçirdiği,
- 80 plaka kodlu Osmaniye ilinde, 2420 Nace kodlu “Çelikten tüpler, borular, içi boş profiller ve benzeri bağlantı parçalarının imalatı” işinde çalışanların, iş sağlığı ve

güvenliği eğitimi ile mesleki eğitim almış olmalarına rağmen % 81,679 güven ve %1,002 destek oranıyla Ciddi kazalar geçirdiği,

- 2420 Nace kodlu “Çelikten tüpler, borular, içi boş profiller ve benzeri bağlantı parçalarının imalatı” işinde çalışanların, iş sağlığı ve güvenliği eğitimi ile mesleki eğitim almış olmalarına rağmen 44 kaza sebep kodlu “Denetim kaybı (tam veya kısmi) - nesnenin (taşınan, oynatılan, kullanılan, vb.)” sebebiyle kazalanmaları durumunda % 73,239 güven ve %1,086 destek oranıyla Ciddi kazalar geçirdiği,
- 35 plaka kodlu İzmir ilinde, iş güvenliği uzmanı bulunan işyerlerinde iş sağlığı ve güvenliği eğitimi ile mesleki eğitim verilmiş erkek çalışanların dahi 11 çalışma çevresi kodlu üretim alanı, fabrika veya atölyede kaza geçirmeleri durumunda bu kazaların % 73,881 güven ve %1,025 destek oranıyla Ciddi sonuçlar doğurduğu,
- 78 plaka kodlu Karabük ilinde, 2410 Nace kodlu “Ana demir ve çelik ürünleri ile demir alaşımları imalatı” işinde orta yaş grubunda çalışanların % 78,71 güven ve %1,185 destek oranıyla Ciddi kazalar geçirdiği,
- 80 plaka kodlu Osmaniye ilinde, 11 çalışma çevresi kodlu üretim alanı, fabrika veya atölyede üretim, imalat işleri gerçekleştiren lise mezunu çalışanların % 76,978 güven ve %1,063 destek oranıyla Ciddi kazalar geçirdiği,
- 35 plaka kodlu İzmir ilinde, 100 ve daha fazla çalışanı bulunan büyük ölçekli işletmelerde 2562 Nace kodlu “Metallerin makinede işlenmesi ve şekil verilmesi” işinde iş sağlığı ve güvenliği eğitimi alan çalışanların % 71,206 güven ve %1,965 destek oranıyla Hafif kazalar geçirdiği,
- 16 plaka kodlu Bursa ilinde, 2451 Nace kodlu “Demir döküm” işinde üretim alanı, fabrika veya atölyede üretim, imalat işleri gerçekleştiren çalışanların iş sağlığı ve güvenliği eğitimi ile mesleki eğitim almış olmaları durumunda % 92,857 güven ve %1,927 destek oranıyla Hafif kazalar geçirdiği,

gibi sonuçlara ulaşılmıştır. Bu örneklerle de açıkça anlatıldığı üzere apriori algoritması ile yapılan analizde iş kazalarının sonuçlarına etki eden faktörler arasında yüksek oranda benzerlikler gözlemlenebilmektedir.

6.2. GRI (Generalized Rule Induction) Algoritması Analizi

Generalized Rules Induction (GRI) Algoritması, veriler üzerinde düğümler arasındaki ilişki kurallarını keşfeder. GRI algoritması, destek ve güven kurallarını dikkate alarak en kuvvetli bilgi bağıını seçecek şekilde ilerlemektedir [65].

“SPSS Clementine” programı üzerinde GRI algoritması vasıtasıyla gerçekleştirilen birliktelik kuralları analizi sonucunda “Ciddi” ve “Hafif” sonuçlu kazalar için en kuvvetli birliktelikleri oluşturan analiz sonuçları Şekil 6.6. ve Şekil 6.7’de gösterilmektedir.

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Cinsiy = 1,000 Egitimmesleki = 1,000 Cisguzmani = 1,000	48,08	51,85
Kazadurum = Ciddi	Saatisbasi = 3,000 Cinsiy = 1,000 Yaskodu = 3,000	44,94	51,85
Kazadurum = Ciddi	Cinsiy = 1,000 Yaskodu = 3,000 Cisguzmani = 1,000	44,28	52,15
Kazadurum = Ciddi	Yaskodu = 3,000 Egitimmesleki = 1,000 Cisguzmani = 1,000	40,21	52,5
Kazadurum = Ciddi	Calismacevrekodu = ... Toptecrube = 5,000	22,03	53,18
Kazadurum = Ciddi	Toptecrube = 5,000 Egitimmesleki = 1,000	21,91	53,37
Kazadurum = Ciddi	Toptecrube = 5,000 Egitimisg = 1,000 Egitimmesleki = 1,000	21,83	53,27
Kazadurum = Ciddi	Calismacevrekodu = ... Toptecrube = 5,000 Egitimisg = 1,000	21,77	53,32
Kazadurum = Ciddi	Calismacevrekodu = ... Cinsiy = 1,000 Toptecrube = 5,000	21,59	53,29
Kazadurum = Ciddi	Cinsiy = 1,000 Toptecrube = 5,000 Egitimmesleki = 1,000	21,5	53,49

Şekil 6.6. SPSS Clementine üzerinde GRI algoritması analiz örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Sehir = 16,000	11,12	55,23
Kazadurum = Hafif	Sehir = 16,000	9,59	57,97
Kazadurum = Hafif	Sehir = 16,000 Genelfaalkodu = 11,...	9,34	58,07
Kazadurum = Hafif	Sehir = 16,000 Calisansayisi = 4,000	9,08	58,25
Kazadurum = Hafif	Sehir = 16,000 Genelfaalkodu = 11,...	8,66	58,96
Kazadurum = Hafif	Sehir = 16,000 Genelfaalkodu = 11,...	8,17	61,09
Kazadurum = Hafif	Sehir = 38,000	5,18	65,78

Şekil 6.7. SPSS Clementine üzerinde GRI algoritması analiz örnekleri

GRI algoritması kullanılarak yapılan analizler sonucu oluşan birliktelik kuralları örnek olarak incelendiğinde;

- İş güvenliği uzmanının istihdam edildiği işyerlerinde mesleki eğitim almış olan erkek çalışanların % 51,85 güven ve % 48,08 destek oranıyla Ciddi kazalar geçirdiği,
- 25 ila 50 yaş aralığındaki erkek çalışanların iş başı yaptıktan 8 saat sonra geçirdikleri kazaların % 51,85 güven ve % 44,94 destek oranıyla Ciddi sonuçlar doğurduğu,
- 20 ve üzeri yıl tecrübeye sahip çalışanların 11 çalışma çevresi kodlu üretim alanı, fabrika veya atölyede kaza geçirmeleri durumunda bu kazaların % 53,18 güven ve %22,03 destek oranıyla Ciddi sonuçlar doğurduğu,
- 16 plaka kodlu Bursa ilinde, 11 genel faaliyet kodlu “üretim, imalat, işleme” işlerinde faaliyet gösteren erkek çalışanların % 58,25 güven ve %9,08 destek oranıyla Hafif kazalar geçirdiği,

gibi sonuçlara ulaşılmıştır.

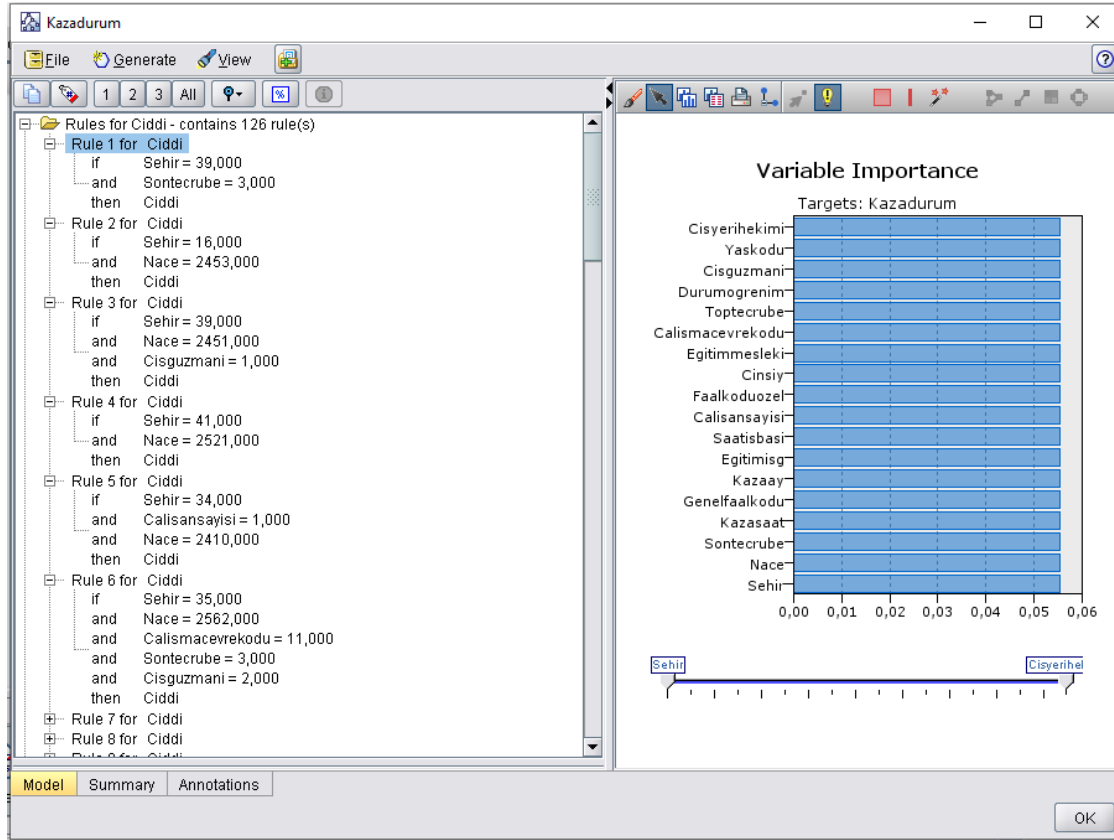
6.3. C5.0 Algoritması Analizi

Karar ağaçlarının oluşturulmasında faydalanılan pek çok algoritma mevcuttur. Bu algoritmalara örnek olarak ID3, C4.5, C5.0, CART, CHAID ve QUEST gösterilebilir.

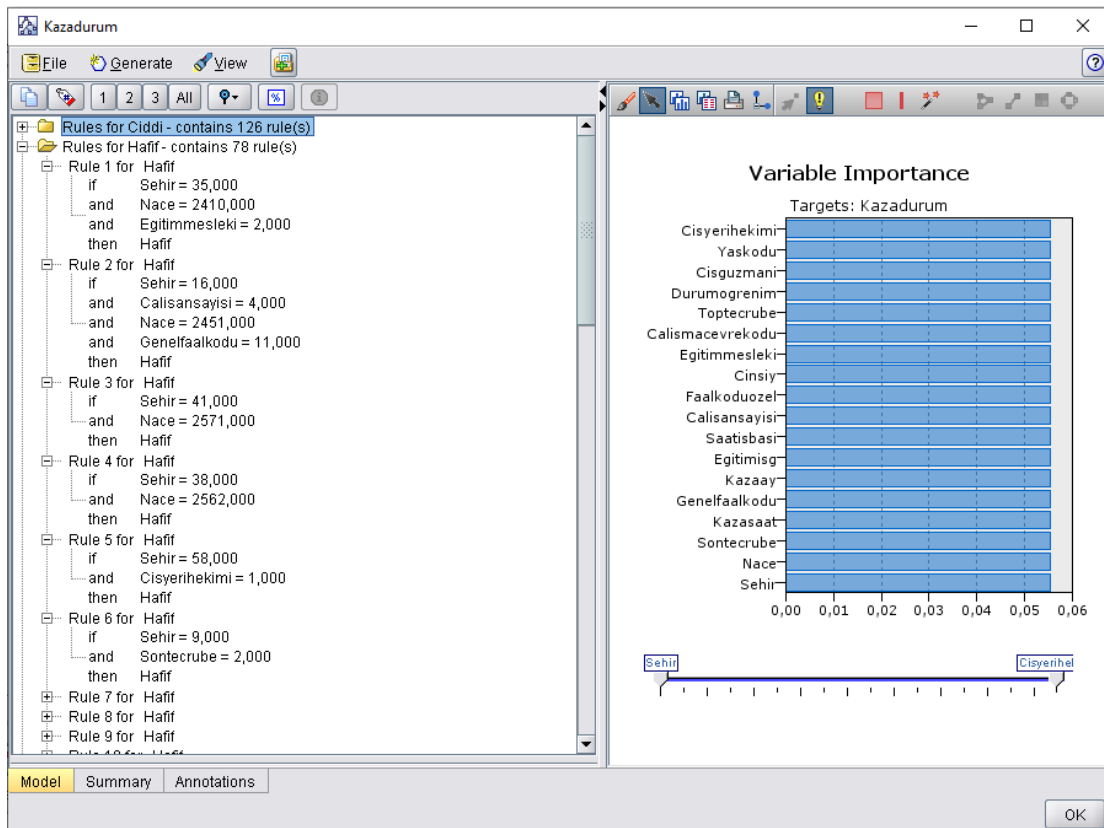
Bu alanda kullanılan algoritmalar arasında en bilineni ise Quinlan'ın ID3 algoritmasının geliştirilmiş hali olan C4.5 algoritmasıdır [66].

C5.0 algoritması incelendiğinde ise C4.5 algoritmasının geliştirilmiş hali olduğu görülmekte ve özellikle büyük veri setlerinin analiz edilmesinde kullanılmaktadır. C5.0 algoritması, analiz sonuçlarının doğruluğunun artırılması amacıyla boosting algoritmasını kullandığından, literatürde boosting ağaçları olarak da bilinir. C5.0 algoritmasının C4.5'e göre avantajları arasında çok daha hızlı olması ve hafızayı daha verimli kullanması sayılabilmektedir. Her iki algoritma aynı sonuçları vermiş olsa bile C5.0 algoritması biçim olarak daha düzgün karar ağaçlarının oluşmasına imkân vermektedir [67].

“SPSS Clementine” programı üzerinde C5.0 algoritması vasıtasıyla analiz sonucunda “Ciddi” ve “Hafif” sonuçlu kazalar için en kuvvetli kuralları oluşturan analiz sonuçları Şekil 6.8 ve Şekil 6.9'da gösterilmektedir.



Şekil 6.8. SPSS Clementine üzerinde C5.0 algoritması analiz örnekleri



Şekil 6.9. SPSS Clementine üzerinde C5.0 algoritması analiz örnekleri

C5.0 algoritması kullanılarak yapılan analizler sonucu oluşan kurallar örnek olarak incelendiğinde;

- 16 plaka kodlu Bursa ilinde, 2453 Nace kodlu “Hafif Metallerin Dökümü” işinde Ciddi kazalar gerçekleştiği,
- 39 plaka kodlu Kırklareli ilinde, 2451 Nace kodlu “Demir Döküm” işinde iş güvenliği uzmanının istihdam edildiği işyerlerinde Ciddi kazalar gerçekleştiği,
- 41 plaka kodlu Kocaeli ilinde, 2521 Nace kodlu “Merkezi ısıtma radyatörleri (elektrikli radyatörler hariç) ve sıcak su kazanları (boylerleri) imalatı” işinde Ciddi kazalar gerçekleştiği,
- 35 plaka kodlu İzmir ilinde, 2562 Nace kodlu “Metallerin makinede işlenmesi ve şekil verilmesi” işinde, iş güvenliği uzmanı istihdam edilmeyen, 11 çalışma çevresi kodlu üretim alanı, fabrika veya atölyede meydana gelen ve o işyerinde 5 ila 10 yıl arasında tecrübeye sahip çalışanların uğradığı kazaların Ciddi sonuçlar doğurduğu,
- 58 plaka kodlu Sivas ilinde işyeri hekimi istihdam edilen işyerlerinde Hafif kazalar gerçekleştiği,
- 41 plaka kodlu Kocaeli ilinde, 2571 Nace kodlu “Çatal-bıçak takımları ve diğer kesici aletlerin imalatı” işinde Hafif kazalar gerçekleştiği,

gibi sonuçlara ulaşılmıştır.

7. SONUÇ VE ÖNERİLER

Çalışanların sağlıklı ve güvenli yaşam şartlarının korunması hem yasal bir görev hem de sosyal yaşantının bir gereğidir. Çalışma hayatında iş kazaları ve meslek hastalıkları istenmeyen sonuçlar olarak ortaya çıkmaktadır. Ancak ülkemiz ve dünya gerçekleri incelendiğinde pek çok sosyal ve ekonomik sebepler ile her ne kadar istenmese de iş kazaları ve meslek hastalıkları ile sıklıkla karşılaşmaktadır.

Sağlıklı ve güvenli çalışma ortamlarının oluşturulması, sürekliliğinin sağlanması ve geliştirilmesi işverenin sorumluluğundadır. Nitekim; 6331 Sayılı İş Sağlığı ve Güvenliği Kanunu'nun 4'üncü maddesinde açıkça belirtildiği üzere işveren, çalışanların işle ilgili sağlık ve güvenliğini sağlamakla yükümlü olup, mesleki risklerin önlenmesi, eğitim ve bilgi verilmesi dâhil her türlü tedbirin alınması, organizasyonun yapılması, gerekli araç ve gereçlerin sağlanması, sağlık ve güvenlik tedbirlerinin değişen şartlara uygun hale getirilmesi ve mevcut durumun iyileştirilmesi için çalışmalar yapmak, işyerinde alınan iş sağlığı ve güvenliği tedbirlerine uyulup uyulmadığını izlemek, denetlemekle ve uygunsuzlukların giderilmesini sağlamak, risk değerlendirmesi yapmak veya yaptırmak, çalışana görev verirken çalışanın sağlık ve güvenlik yönünden işe uygunluğunu göz önünde bulundurmak, yeterli bilgi ve talimat verilenler dışındaki çalışanların hayati ve özel tehlike bulunan yerlere girmemesi için gerekli tedbirleri almak zorundadır.

Ancak; çalışma hayatındaki istenmeyen durumların önlenmesi, önlenemese bile asgari seviyelere çekilmesi topyekûn bir mücadele gerektirmektedir. Çalışanlar, iş sağlığı ve güvenliği profesyonelleri ve işverenlerden oluşan bu üçlü sacayağının bir parçası eksik kalırsa, iş güvenliği konusunda verilen mücadelenin başarıya ulaşması mümkün olmayacaktır.

Bu amaçla iş kazaları ile mücadele edilirken kaza sebeplerinin, kazaya neden olacak faktörleri etkileyen durumların net bir biçimde analiz edilmesi, bu faktörlerin ortadan kaldırılması için büyük önem arz etmektedir.

Bu çalışmada literatürde veri madenciliği yöntemleri arasında sıklıkla başvuru alan birliktelik kuralları analizlerinden 2 farklı algoritma (Apriori ve GRI) ile karar ağacı yöntemlerinden

C5.0 algoritması olmak üzere 3 farklı algoritma ile aynı veriler analiz edilmiş, iş kazalarına neden olan sebepler arasındaki gizli ilişkiler ortaya çıkarılmaya çalışılmış ve kaza sonuçlarına göre farklı kurallar belirlenmiştir.

Çalışmanın sonuçlarını iki farklı açıdan değerlendirmek mümkündür. Birinci bakış açımız iş kazalarının meydana gelmesinde etkili olan parametreleri yorumlamak, ikinci bakış açısı ise iş kazasında kullanılan algoritmaların birbirleri ile kıyaslanması olmalıdır.

Analiz sürecinde her algoritmanın farklı avantaj ve dezavantajları olmasına rağmen bazı ortak iş kazası kuralları oluşturduğu gözlemlenmiştir. Bununla birlikte, algoritmaların kendi çalışma mantıkları ile çok farklı kuralların da oluşumuna imkân sağladıkları tespit edilmiştir.

Çalışma sonucunda elde edilen kurallara göre iş kazalarının oluşumunda tüm analiz yöntemlerinde ortak olarak; Bursa, Kayseri ve Kocaeli illerinde metal sektöründeki tüm alt kollar dâhil olmak üzere ciddi kazalar meydana geldiği, İzmir ilinde ise metallerin makine işlenmesi ve şekil verilmesi işinde ciddi kazalar meydana geldiği görülmüştür. Ayrıca iş güvenliği uzmanı ve işyeri hekimi çalıştırmak ile iş güvenliği eğitimleri ve mesleki eğitimlerin kazaların meydana gelmesinde etkin parametreler olduğu dikkat çekmektedir.

Analiz sonuçları algoritmaların farkları açısından incelendiğinde;

- C5.0 Algoritmasının; diğer algoritmaların oluşturduğu kurallar arasında sıklıkla yer almayan Bilecik, Çorum, Eskişehir, Samsun ve Aksaray gibi iller için de kural oluşturduğu,
- C5.0 Algoritmasının sonuç verilerinin daha düzenli ve kontrol edilebilir olduğu,
- GRI Algoritmasının; %48'e varan destek oranıyla diğer yöntemlere göre büyük fark yarattığı,
- GRI Algoritmasının; yüksek destek oranına rağmen güven oranında diğer yöntemlerin gerisinde kaldığı,
- GRI Algoritmasının; ortak parametreler(Şehir, Faaliyet Kodu, Cinsiyet vb.) dışında, mesleki eğitim, iş sağlığı ve güvenliği eğitimi, iş güvenliği uzmanı istihdamı gibi iş güvenliğinin organizasyonu olarak isimlendirilebilecek alanlarda diğer yöntemlere oranla daha fazla kural oluşturduğu,

- Apriori Algoritmasının; destek oranının düşük olmasının yanında %82'ye varan güven oranı ile diğer yöntemlere göre daha başarılı olduğu,
- Apriori Algoritmasının; diğer algoritmaların oluşturduğu kurallar arasında sıklıkla yer almayan Karabük ve Osmaniye gibi iller için de kural oluşturduğu,
- Apriori Algoritmasının diğer iki yöntemle oranla çok daha hızlı sonuç verdiği ve hızının yanı sıra; elde edilen kuralların çeşitliliği yönünden de diğer algoritmalarından daha verimli olduğu,

gözlemleri yapılmıştır.

Bunun yanı sıra, algoritmalarından elde edilen sonuçlar tek tek incelendiğinde, farklı yaş grupları ve tecrübedeki çalışanlar için gerçekleştirdikleri faaliyetler bazında özel politikalar geliştirilmesi gerektiği ortaya çıkmaktadır.

Ana amacın, kazalara neden olabilecek ilişkilerin tespit edilmesi olan bu çalışmada, her algoritmanın verdiği sonuçlar kıymetli ve dikkate değerdir. Kazaya neden olan faktörler bilindiği sürece, önleme kültürünün bir yansıması olarak alınabilecek tedbirler daha kolay, daha hızlı ve daha verimli olarak belirlenebilecektir. Bu çalışmada elde edilen kuralların göz önünde bulundurulması ile olası kazaların engellenmesi için alınacak tedbirlere yön verilebilir.

Tez yazımı sürecinde bilimsel literatürde kabul görmüş pek çok veri madenciliği yöntemi bulunmasına rağmen iş kazaları ile ilgili fazla sayıda çalışma yapılmadığı görülmüş olup, bu tez çalışmasından elde edilecek özgün bulguların farklı analiz yöntemleri ile çalışılarak geliştirilebileceği değerlendirilmektedir. Bu çalışmadan elde edilecek sonuçların konuya ilgi duyan araştırmacılara yol göstermesi, insan hayatının kutsallığı göz önünde bulundurulduğunda, büyük önem arz etmektedir.

Unutulmamalıdır ki;

Önlemek, ödemekten ucuzdur!

KAYNAKLAR

1. Gençler, A. (2007). İş Sağlığı ve İş Güvenliği Alanında Mevzuatımızda Bulunan Düzenlemelerden Doğan Yükümlülükler. *İş Sağlığı ve Güvenliği Dergisi*, 35, 16.
2. Akyüz, N. (1980). *İş Güvenliği Herkes İçin*. İstanbul: Arpaz Matbaacılık, 2.
3. İnternet: İyibozkurt, A. (2006). İş Sağlığı ve Güvenliğinin Önemi Web: <http://blog.milliyet.com.tr/isguvenligi> 03.05.2019 tarihinde alınmıştır.
4. Şardan, S. (2007). İş Sağlığı ve Güvenliğini Yönetmek. *İş Sağlığı ve Güvenliği Dergisi*, 35, 12.
5. Akpınar, H. (2000). Veri tabanlarında Bilgi Keşfi ve Veri Madenciliği, *İstanbul Üniversitesi, İşletme Fakültesi Dergisi*, 29 (1), 1-22.
6. Özmen, Ş. (2001). *İş Hayatı Veri Madenciliği Uygulamaları ile İstatistik Uygulamalarını Yeniden Keşfediyor*. V. Ulusal Ekonometri ve İstatistik Sempozyumu, Çukurova Üniversitesi, Adana, 1- 6.
7. Silahtaroglu, G. (2008). *Kavram ve Algoritmalarıyla Temel Veri Madenciliği*. İstanbul: Papatya Yayıncılık, 83-98
8. Altunkaynak, B., (2018). A statistical study of occupational accidents in the manufacturing industry in Turkey. *International Journal of Industrial Ergonomics*, 66, 101-106.
9. Adem, A., Çolak, A., Dağdeviren, M. (2018). An integrated model using SWOT analysis and Hesitant fuzzy linguistic term set for evaluation occupational safety risks in life cycle of wind turbine. *Safety Science*, 106, 184-190.
10. Wanguan, Q., Quanlong, L., Xinchun, L., Xixi, L. ve YuLong, W. (2018). Using data mining techniques to analyze the influencing factor of unsafe behaviors in Chinese underground coal mines. *Resources Policy*, 59, 210-216.
11. Barlas, B. ve İzci, F.B. (2018). Individual and workplace factors related to fatal occupational accidents among shipyard workers in Turkey. *Safety Science*, 101, 173–179.
12. Xia, J., Zhaoa, Z., Lib, W. Ve Wanga, Q. (2016). A Traffic Accident Causation Analysis Method Based on AHP Apriori. *Procedia Engineering*, 137, 680 – 687.
13. Sanmiquel, L., Rossell, J.M. ve Vintro, C. (2015). Study of Spanish mining accidents using data mining techniques. *Safety Science*, 75, 49-55.
14. Pietilä, J., Räsänen, T., Reiman, A., Ratilainen, H. ve Helander, E. (2018). Characteristics and determinants of recurrent occupational accidents. *Safety Science*, 108, 269-277.

15. Hajakbari, M.S. ve Minaei-Bidgoli, B. (2014). A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data. *Journal of Loss Prevention in the Process Industries*, 32, 443-453.
16. Cheng, C., Yao, H. ve Wu, T. (2013). Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. *Journal of Loss Prevention in the Process Industries*, 26(6), 1269-1278.
17. Zhang, D. ve Jiang, K. (2012). Application of Data Mining Techniques in the Analysis of Fire Incidents. *Procedia Engineering*, 43, 250-256.
18. Silva, J.F. ve Jacinto, C. (2012). Finding occupational accident patterns in the extractive industry using a systematic data mining approach. *Reliability Engineering and System Safety*, 108, 108–122.
19. Nenonen, N. (2012). Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database. *Applied Ergonomics*, 44(2), 215-224.
20. Cheng, C., Leu, S., Cheng, Y., Wu, T., ve Lin, C. (2012). Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. *Accident Analysis and Prevention*, 48, 214– 222.
21. Rivas, T., Paz, M., Martin, J. E., Matias, J. M., Garcia, J. F., Taboada, J. (2011). Explaining and predicting workplace accidents using data-mining techniques. *Reliability Engineering and System Safety*, 96(7), 739–747.
22. Cheng, C., Lin, C. ve Leu, S. (2010). Use of association rules to explore cause–effect relationships in occupational accidents in the Taiwan construction industry. *Safety Science*, 48(4), 436-444.
23. Bayram F. (2008). *Türk İş Hukukunda İş Sağlığı ve Güvenliği Denetimi*. İstanbul: Beta Yayınları.
24. Kılıkış, İ. (2016). *İş Sağlığı ve Güvenliği*. Bursa: Dora Yayınları.
25. Çelik, E. (2016). *Sağlık çalışanlarının iş sağlığı ve güvenliğine dair farkındalıklarının incelenmesine yönelik bir alan araştırması*, Yüksek Lisans Tezi, İstanbul Gelişim Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
26. Çelikçapa, S. (2015). *İş Sağlığı ve Güvenliği Tekstil Sektöründe Risk Değerlendirmesi*. Bursa: Ekin Basım.
27. Alper, Y. (1992). *Bazı Ülkelerde İşçi Sağlığı – İş Güvenliği Uygulamaları ve Türkiye'deki Uygulama ile Karşılaştırılması*, Sosyal Siyaset Konferansları, 37-38., İstanbul Üniversitesi Yayını, İstanbul.
28. Akgün, S. (2016). *PCB atölyelerinde iş sağlığı ve güvenliği*, Yüksek Lisans Tezi, Gedik Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.

29. Argüden, Y., Erşahin, B. (2008). Veri Madenciliği: *Veriden bilgiye, masraftan değere*. İstanbul: ARGE Danışmanlık Yayınları, 10.
30. Baykal, A. (2006). Veri madenciliği uygulama alanları. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 7, 95-107.
31. Sever, H., Oğuz, B. (2002). Veritabanlarında bilgi keşfine formel bir yaklaşım: Kısım I: Eşleştirme sorguları ve algoritmalar. *Bilgi Dünyası*, 3(2), 173-204.
32. Alan, M. A. (2012). Veri madenciliği ve lisansüstü öğrenci verileri üzerine bir uygulama. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 33, 165-174.
33. Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Commerce University Journal of Science*, 3(3), 65-82.
34. Ögüt, S., 2007. *Veri madenciliği kavramı ve gelişim süreci*. Yüksek Lisans Tezi, Yeditepe Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
35. Savaş, S., Topaloğlu, N., Yılmaz, M. (2012). Veri madenciliği ve Türkiye’deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23
36. Usgurlu, B., Özcan, Ö. ve Demirörs, O. (2010), *A Clustering Based Functional Similarity Measurement Approach*, IEEE, Conference Proceedings of 36th EUROMICRO Conference on Software Engineering and Advanced Applications, 371-375
37. İnternet: Çıngı, H., 2007, “Veri Madenciliğine Giriş”, Ders notları, 1-35, Web: <http://yunus.hacettepe.edu.tr/~hcingi/ist376a/6Bolum.doc> 03/05/2019 tarihinde alınmıştır.
38. İnternet: Küçüksille, E. U., 2010, “Veri Madenciliği ve Uygulama Programları”, Veri Tabanı Yönetim Sistemleri Ders Notları, Web: <http://ab.org.tr/ab06/ozet/61.html> 03/05/2019 tarihinde alınmıştır.
39. İnternet: Kaya, E., Bulun, M. ve Arslan, A., 2010, “Tıpta Veri Ambarları Oluşturma ve Veri Madenciliği Uygulamaları”, <http://ab.org.tr/ab03/tammetin/96.doc> 03/05/2019 tarihinde alınmıştır.
40. Atılğan, E. (2011). *Karayollarında meydana gelen trafik kazalarının karar ağaçları ve birliktelik kuralı ile analiz edilmesi*. Yayımlanmamış Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
41. Özkan, Y. (2008). *Veri madenciliği yöntemleri*. İstanbul: Papatya Yayıncılık Eğitim.
42. Koyuncugil, A. S., Özgülbaş, N., 2009. Veri Madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. *International Journal Of Informatics Technologies*, 2(2), 21-32.
43. Güngör, E., Yalçın, N., Yurtay, N. (2008). Apriori Algoritması ile teknik seçmeli ders seçim analizi. *Endüstri Mühendisliği Dergisi*, 21(2), 14-29.

44. Wilkinson, L. (1992). *Tree structured data analysis: AID, CHAID and CART*. Chicago: SPSS Inc.
45. Özdemir, M. E., Yıldırım, E., Yıldırım, S. (2015). *Classification of emotional valence dimension using artificial neural networks*. Signal Processing and Communications Applications Conference (SIU), 23, 2549-2552.
46. Çalışkan, S. K., Soğukpınar, İ. (2008). *K. Knn: K-Means ve K-En Yakın Komşu Yöntemleri ile ağlarda nüfuz tespiti*. 2. Ağ ve Bilgi Güvenliği Sempozyumu, 120-124, EMO Yayınları, Girne.
47. Yu, S., Tranchevent, L. C., Liu, X., Glänzel, W., Suykens, J. A., De Moor, B., Moreau, Y. (2012). Optimized data fusion for kernel k-means clustering. *Pattern Analysis and Machine Intelligence, IEEE*, 34(5), 1031-1039.
48. Kaya, H, Köymen, K. (2008). Veri madenciliği kavramı ve uygulama alanları. *Doğu Anadolu Bölgesi Araştırmaları*, 6(2), 159-164.
49. Park, H. S., Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2), 3336-3341.
50. Kim, M. S., Han, J. (2009). A particle-and-density based evolutionary clustering method for dynamic networks. *Proceedings of the VLDB Endowment*, 2(1), 622-633.
51. Tao, L., McCurdy, C. W., Rescigno, T. N. (2009). Grid-based methods for diatomic quantum scattering problems: A finite-element discrete-variable representation in prolate spheroidal coordinates. *Physical Review A*, 79(1), 012719.
52. Hart, E., Timmis, J. (2008). Application areas of AIS: The past, the present and the future. *Applied Soft Computing*, 8(1), 191-201.
53. Döşlü, A. (2008). *Veri madenciliğinde market sepet analizi ve birliktelik kurallarının belirlenmesi*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
54. Blahut, R. E. (2010). *Fast algorithms for signal processing*. Cambridge, İngiltere: Cambridge University Press.
55. Moore, A. W., Zuev, D. (2005). Internet traffic classification using bayesian analysis techniques. *In ACM Sigmetrics Performance Evaluation Review*, 33(1), 50-60.
56. Guo, Z., Chi, D., Wu, J., Zhang, W. (2014). A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm. *Energy Conversion and Management*, 84, 140-151.
57. Aflori, C., Craus, M. (2007). Grid implementation of the Apriori algorithm. *Advances in engineering software*, 38(5), 295-300.

58. Morales-Esteban, A., Martínez-Álvarez, F., Scitovski, S., Scitovski, R. (2014). A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning. *Computers and Geosciences*, 73, 132-141.
59. Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., Renard, P. (2013). A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. *Computers and Geosciences*, 52, 307-324.
60. Tille, Y. (2011). *Sampling algorithms*. Berlin, Almany: Springer Berlin Heidelberg, 1273-1274.
61. Ummel, K. (2012). CARMA revisited: an updated database of carbon dioxide emissions from power plants worldwide. *Center for Global Development*, 304, 1-30.
62. Jie, D., Yang, S., Feng, D. (2008). *Convergence properties of multi-innovation ESG algorithms for multi-input multi-output CARMA models*. In Control Conference, Chinese, 270-274.
63. Sidhu, S., Meena, U. K., Nawani, A., Gupta, H., Thakur, N. (2014). FP Growth Algorithm Implementation. *International Journal of Computer Applications*, 93(8), 6-10.
64. Singh, J., Ram, H., Sodhi, D. J. (2013), Improving efficiency of apriori algorithm using transaction reduction. *International Journal of Scientific and Research Publications*, 3(1), 1-4.
65. SPSS Inc. (2007). *Clementine® 11.0 User's Guide*. Chicago, USA: SPSS Inc.,15-44.
66. Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning* (First Edition). California, USA: Morgan Kaufmann Publishers, 21-57.
67. Çalış, A., Kayapınar, S., Çetinyokuş, T. (2014). *Endüstri Mühendisliği Dergisi*, 3-4, 2-19.

EKLER

EK-1. Analizde kullanılan parametre sınıflandırmaları

Çalışan Sayısı Sınıflandırması

Çalışan Sayısı	Sınıf Kodu
0-9	1
10 ve 49	2
50 ve 99	3
100 ve üzeri	4

Çalışma Çevresi Sınıflandırması

Çalışma Çevresi Kodu	Çalışma Çevresi Tanımı
11	Üretim alanı, fabrika, atölye
12	Bakım alanı, onarım atölyesi
13	Temelde depo, yükleme, boşaltma için kullanılan alan
19	Yukarıda listelenmemiş diğer başka 010 tür çalışılan çevre
21	İnşaat mevki - yeni inşa edilen bina
22	İnşaat mevki - yıkılan, onarılan, bakımı yapılan bina
23	Açık hava taşocağı, Açık hava madeni, kazı, çukur (Açık hava madenciliği ve çalışan tas ocakları dahil)
24	İnşaat mevki - yeraltı
25	İnşaat mevki - su üstü veya üzerinde
29	Yukarıda listelenmemiş diğer başka 020 tür çalışılan çevre
31	Yetiştirme alanı
32	Çiftlik alanı - toprak ürünleri
34	Ormancılık alanı
39	Yukarıda listelenmemiş diğer başka 030 tür çalışılan çevre
41	Büro, toplantı salonu, kütüphane, vb.
43	Küçük veya büyük satış alanı (sokak satışları dahil)

EK-1. (devam) Analizde kullanılan parametre sınıflandırmaları

44	Lokanta, dinlenme alanı, geçici konaklama (müze, oditoryum, stadyum, fuar, vb. dahil)
49	Yukarıda listelenmemiş diğer başka 040 tür çalışılan çevre
51	Sağlık Kurumu, özel hastane, hastane, bakım evi
59	Yukarıda listelenmemiş diğer başka 050 tür çalışılan çevre
61	Sürekli olarak kamu geçişine Açık alan (karayolları, yan yollar, park alanları, istasyon veya havaalanı bekleme salonları, vb.)
62	Ulaşım araçları - kara veya demiryolu ile- özel veya kamu (her tür: tren, otobüs, araba, vb.)
63	Kamu alanlarına bağlı alanlar olmakla birlikte sadece yetkili kimselerin erişimine izin verilen alanlar: demiryolu hattı, havaalanı pisti, karayolu banketi
69	Yukarıda listelenmemiş diğer başka 060 tür çalışılan çevre
71	Kişisel ev
72	Bir binanın ortak alanı, ekleri, kişisel aile bahçesi
79	Yukarıda listelenmemiş diğer başka 070 tür çalışılan çevre
81	Kapalı spor alanı - spor salonu, jimnastik salonu, kapalı yüzme havuzu
82	Açık spor alanı - spor sahası, Açık yüzme havuzu, kayak parkuru
89	Yukarıda listelenmemiş diğer başka 080 tür çalışılan çevre
91	Yükseltme esnasında - sabit bir düzeyde (çatı, teras, vb.)
101	Yeraltında - tünel (yol, ten, tüp)
111	Deniz veya okyanus - her tür tekne, platform, gemi, sandal, mavnada
112	Göl, nehir, liman - her tür tekne, platform, gemi, sandal, mavnada
129	Yukarıda listelenmemiş başka 120 türü çalışılan çevre, inşaat şantiyesi dışında
999	Sınıflandırmada listelenmemiş başka çalışma ortamları

EK-1. (devam) Analizde kullanılan parametre sınıflandırmaları

Genel Faaliyet Sınıflandırması

Genel Faaliyet Kodu	Genel Faaliyet Tanımı
11	Üretim, imalat, işleme - tüm türler
12	Depolama - tüm türler
19	Yukarıda listelenmemiş diğer başka 10 tür Kaza anında kazazedenin yürütmekte olduğu genel faaliyet
21	Kazı
22	Yeni inşaat – bina
23	Yeni inşaat - inşaat mühendisliği, altyapı, yol, köprü, baraj ve limanlar
24	Yeniden modelleme, onarım, genişletme, bina bakımı - her tür inşaat
25	Yıkım - her tür inşaat
29	Yukarıda listelenmemiş diğer başka 20 çeşit kaza anında kazazedenin yaptığı faaliyet
31	Tarımla ilgili meslek türü - toprağın islenmesi
33	Tarımla ilgili meslek türü - canlı hayvanlarla
39	Yukarıda listelenmemiş diğer başka 30 çeşit kaza anında kazazedenin yaptığı faaliyet
41	Kamu hizmeti, bakım, yardım, servis
42	Zihinsel çalışma - öğretmenlik, eğitim, bilgi işlem, büro işi, organizasyon, yönetim
49	Yukarıda listelenmemiş başka 40 çeşit kaza anında kazazedenin yaptığı faaliyet
51	Kurulum, hazırlama, montaj, yükleme, sökme, parçalama
52	Bakım, onarım, ayar, akort
53	İs alanının temizlenmesi, makine sanayi veya elle
54	Atık yönetimi, kontrol, her tür atık işleme
55	İzleme, donanımları olsun veya olmasın imalat süreçlerinin, çalışma alanlarının, taşıt araçlarının izlenmesi, teftişi
59	Yukarıda listelenmemiş başka 50 çeşit kaza anında kazazedenin yaptığı faaliyet
61	Taşıt araçlarıyla olanı da dahil olmak üzere hareket;
62	Spor, sanatsal faaliyet
69	Yukarıda listelenmemiş başka 60 çeşit kaza anında kazazedenin yaptığı faaliyet
99	Sınıflandırmada listelenmemiş başka kaza anında kazazedenin yaptığı faaliyet

EK-1. (devam) Analizde kullanılan parametre sınıflandırmaları

Kaza Sebep Kodu Sınıflandırması

Kaza Sebep Kodu	Kaza Sebebi Tanımı
11	Elektrik arızası nedeniyle elektrik sorunu - doğrudan temasa yol açan
12	Elektrik sorunu - doğrudan temasa yol açan
13	Patlama
14	Yangın, tutuşma
19	Yukarıda listelenmemiş başka 10 çeşit Sapma
21	Kati durumunda - tasma, devrilme
22	Sıvı durumunda - sızma, kaçırma, akma, sıçrama, püskürme
23	Gaz durumunda - buharlaşma, aerosol oluşum, gaz oluşumu
24	Toz halindeki madde - duman oluşumu, havadaki/yayılmış toz ve zerrecikleri
29	Yukarıda listelenmemiş başka 20 çeşit Sapma
31	Maddenin kırılması - eklemlerde, birleşme noktalarında
32	Kırılma, patlama - kıymık oluşumu (tahta, cam, metal, tas, plastik, diğerleri)
33	Maddi aracın kayma, düşme, çökmesi - yukarıdan (kazazedenin üstüne düşerek)
34	Maddi aracın kayma, düşme, çökmesi - aşağıdan (kazazedeyi aşağı çekerek)

EK-1. (devam) Analizde kullanılan parametre sınıflandırmaları

35	Maddi aracın kayma, düşme, çökmesi - aynı düzeyde
39	Yukarıda listelenmemiş başka 30 türü-çeşit Sapma
41	Denetim kaybı (tam veya kısmi) - makine (istenmeyen başlama da dahil olmak üzere) veya maddi araç
42	Denetim kaybı (tam veya kısmi) - taşıt aracı veya ekipman kullanımının (motorlu olsun veya olmasın)
43	Denetim kaybı (tam veya kısmi) - el aleti (motorlu olsun veya olmasın) veya alet tarafından kullanılan maddi aracın
44	Denetim kaybı (tam veya kısmi) - nesnenin (taşınan, oynatılan, kullanılan, vb.)
45	Denetim kaybı (tam veya kısmi) - hayvanın
49	Yukarıda listelenmemiş başka 40 çeşit Sapma
51	Kişinin düşmesi - alt düzeye
52	Kayma - tökezleme ve düşme - Kişinin düşmesi - aynı düzeyde
59	Yukarıda listelenmemiş başka 50 çeşit Sapma
61	Keskin bir nesne üzerinde yürüme
62	Diz çökme, oturma, yaslanma
63	Kapılma veya götürülme - bir şey veya ivme tarafından
64	İşbirliği olmaksızın yapılan hareket, gereksiz veya zamansız eylemler
69	Yukarıda listelenmemiş başka 60 çeşit Sapma
71	Kaldırma, taşıma, ayakta durma
72	İtme, çekme
73	Aşağı bırakma, eğilme
74	Bükülme, dönme
75	Çiğnenmek, bacak veya bilek bükülmesi, düşmeden kaymak
79	Yukarıda listelenmemiş başka 70 çeşit Sapma
81	Sok, korku
82	Vahşet, saldırı, tehdit - işverenin otoritesine tabi olan şirket çalışanları arasında
84	Saldırı, dürtülme - hayvan tarafından
85	Üçüncü bir kimsenin veya kişinin kendisinin, kendine ve başkasına da tehlike yaratması
89	Yukarıda listelenmemiş başka 80 türü çeşit Sapma

EK-1. (devam) Analizde kullanılan parametre sınıflandırmaları

Özel Faaliyet Sınıflandırması

Özel Faaliyet Kodu	Özel Faaliyet Tanımı
11	Makinenin çalıştırılması, makinenin durdurulması
12	Makinenin beslenmesi, makinenin boşaltılması
13	Makinenin izlenmesi, makinenin işletilmesi veya sürülmesi
19	Yukarıda listelenmemiş başka 10 çeşit kaza anında kazazedenin yaptığı faaliyet
21	El makineleriyle çalışma - elle
22	El makineleriyle çalışma - motorlu
29	Yukarıda listelenmemiş başka 20 çeşit kaza anında kazazedenin yaptığı faaliyet
31	Bir taşıt aracını sürmek veya donanımı kullanmak - seyyar ve motorlu
32	Bir taşıt aracını sürmek veya donanımı kullanmak - seyyar ve motorsuz
33	Bir taşıt aracının yolcusu olmak
39	Yukarıda listelenmemiş başka 30 çeşit kaza anında kazazedenin yaptığı faaliyet
41	Ele almak, tutmak, kavramak, yakalamak, yerleştirmek - yatay bir düzeyde
42	Bağlamak, yapıştırmak, ayırmak, sökmek, sıkmak, vidasını açmak, vidalamak, çevirmek
43	Sıkılamak, asmak, yükseltmek, koymak - düşey bir düzeyde
44	Atmak, savurmak
45	Açmak, kapatmak (kutu, paket, parsel)
46	Akıtmak, dökmek, doldurmak, sulamak, püskürtmek, boşaltmak
47	Açmak (çekmeceyi), itmek (depo/büro/dolap kapağını)
49	Yukarıda listelenmemiş başka 40 çeşit kaza anında kazazedenin yaptığı faaliyet
51	Düşey olarak taşıma - bir nesneyi kaldırma, yükseltme, alçaltma
52	Yatay olarak taşıma - bir nesneyi çekme, itme, yuvarlama
53	Bir yükü taşıma - bir kişi tarafından taşınması
59	Yukarıda listelenmemiş başka 50 çeşit kaza anında kazazedenin yaptığı faaliyet
61	Yürüme, koşma, çıkma, inme, vb.
62	Çıkma veya girme

EK-1. (devam) Analizde kullanılan parametre sınıflandırmaları

63	Atlama, zıplama, vb.
64	Sürünme, tırmanma, vb.
65	Kalkma, oturma
67	Yerdeki hareketler
69	Yukarıda listelenmemiş başka 60 çeşit kaza anında kazazedenin yaptığı faaliyet
70	Bulunma - Belirtilmemiş
99	Bu sınıflandırma listelenmemiş başka kazadan az önceki zamanda kazazedenin yürüttüğü özel faaliyet

Cinsiyet Sınıflandırması

Cinsiyet	Kod
Erkek	1
Kadın	2

Öğrenim Durumu Sınıflandırması

Öğrenim Durumu	Kod
Okuryazar değil ya da ilkokul	1
İlköğretim ya da ortaokul	2
Lise	3
Üniversite, yüksek lisans, doktora	4

Saat Aralığı Sınıflandırması

Saat Aralığı	Kod
00:00-03:59	1
04:00-07:59	2
08:00-11:59	3
12:00-15:59	4
16:00-19:59	5
20:00-23:59	6

EK-1. (devam) Analizde kullanılan parametre sınıflandırmaları

Çalışan Tecrübesi Sınıflandırması

Çalışan Tecrübesi (Yıl)	Kod
0 ile 1 arası	1
2 ve 4 arası	2
5 ile 10 arası	3
11 ve 20 arası	4
20 ve üzeri	5

Çalışan Yaşı Sınıflandırması

Çalışan Yaşı	Kod
18e kadar	1
18 ile 24	2
25 ile 50	3
51 ile 60	4
61 üzeri	5

Kısa Durum Sınıflandırmaları

Durum	Kod
Evet/Var	1
Hayır/Yok	2

EK-2. Apriori Algoritması ile elde edilen birliktelik kuralları örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 80.0 Nace = 2420.0 Egitimmesleki = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,002	81,679

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 78.0 Nace = 2410.0 Cisyerih ekimi = 1.0 Yaskodu = 3.0 Cinsiy = 1.0 Egitimisg = 1.0	1,185	78,71

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Saatibasi = 2.0 Sehir = 41.0 Cisyerih ekimi = 1.0 Calisansayisi = 4.0 Genelfaalkodu = 11.0 Cinsiy = 1.0	1,002	78,626

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 78.0 Nace = 2410.0 Cisyerih ekimi = 1.0 Yaskodu = 3.0 Calisansayisi = 4.0 Cinsiy = 1.0 Egitimisg = 1.0	1,177	78,571

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Nace = 2420.0 Durumogrenim = 1.0 Cisyerih ekimi = 1.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Egitimisg = 1.0	1,002	77,863

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 80.0 Durumogrenim = 3.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0 Egitimisg = 1.0	1,063	76,978

EK-2. (devam) Apriori Algoritması ile elde edilen birliktelik kuralları örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Nace = 2420.0 Durumogrenim = 1.0 Calisansayisi = 4.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0	1,086	76,761

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 35.0 Nace = 2410.0 Cisguzmani = 1.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0 Egitimisg = 1.0	1,025	73,881

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Nace = 2420.0 Durumogrenim = 1.0 Calisansayisi = 4.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Egitimisg = 1.0	1,139	73,826

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Nace = 2420.0 Kazasebepkodu = 44.0 Egitimmesleki = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,078	73,759

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Nace = 2420.0 Faalkoduozel = 12.0 Cisguzmani = 1.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,032	71,852

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 41.0 Toptecrube = 5.0 Durumogrenim = 1.0 Yaskodu = 3.0 Calisansayisi = 4.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Egitimisg = 1.0	1,047	71,533

EK-2. (devam) Apriori Algoritması ile elde edilen birliktelik kuralları örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Kazasebepkodu = 44.0 Durumogrenim = 1.0 Cisyerihekimi = 1.0 Yaskodu = 3.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,04	70,588

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Nace = 2420.0 Durumogrenim = 1.0 Yaskodu = 3.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,132	73,649

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 80.0 Yaskodu = 3.0 Calisansayisi = 4.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0 Egitimisg = 1.0	1,07	73,571

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Nace = 2420.0 Kazasebepkodu = 44.0 Egitimmesleki = 1.0 Egitimisg = 1.0	1,086	73,239

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 35.0 Durumogrenim = 3.0 Cisguzmani = 1.0 Yaskodu = 3.0 Calisansayisi = 4.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0 Egitimisg = 1.0	1,162	73,026

EK-2. (devam) Apriori Algoritması ile elde edilen birliktelik kuralları örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 41.0 Toptecrube = 5.0 Durumogrenim = 1.0 Yaskodu = 3.0 Calisansayisi = 4.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0	1,047	72,263

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Sehir = 38.0 Saatisbasi = 2.0 Nace = 2562.0 Cisyerih ekimi = 1.0 Calisansayisi = 4.0 Egitimmesleki = 1.0 Cinsiy = 1.0 Egitimisg = 1.0	1,002	99,237

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Kazasebepkodu = 64.0 Nace = 2562.0 Cisguzmani = 2.0 Calisansayisi = 4.0 Egitimisg = 1.0	1,2	81,529

Kazadurum = Hafif	Sehir = 38.0 Cisyerih ekimi = 1.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Egitimisg = 1.0	3,41	71,076
-------------------	---	------	--------

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Sehir = 38.0 Cisyerih ekimi = 1.0 Calisansayisi = 4.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Egitimisg = 1.0	3,127	73,35

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Nace = 2562.0 Cisguzmani = 2.0 Cisyerih ekimi = 1.0 Egitimmesleki = 1.0 Egitimisg = 1.0	3,104	71,182

EK-2. (devam) Apriori Algoritması ile elde edilen birliktelik kuralları örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Nace = 2451.0 Durumogrenim = 1.0 Calisansayisi = 4.0 Cinsiy = 1.0 Egitimisi = 1.0	2,753	70,278

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Saatibasi = 2.0 Nace = 2562.0 Cisyerihedimi = 1.0 Genelfaalkodu = 11.0 Calismacevrekodu = 11.0 Egitimisi = 1.0	2,286	71,572

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Sehir = 38.0 Cisyerihedimi = 1.0 Yaskodu = 3.0 Calisansayisi = 4.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Egitimisi = 1.0	2,24	73,038

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Nace = 2562.0 Cisguzmani = 2.0 Cisyerihedimi = 1.0 Calisansayisi = 4.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0 Egitimisi = 1.0	1,988	80,385

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Kazasebepkodu = 35.0 Nace = 2562.0 Calisansayisi = 4.0 Egitimisi = 1.0	1,965	71,206

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Kazasebepkodu = 39.0 Calisansayisi = 4.0	1,934	71,146

EK-2. (devam) Apriori Algoritması ile elde edilen birliktelik kuralları örnekleri

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Nace = 2451.0 Sehir = 16.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0 Cinsiy = 1.0 Egitimisg = 1.0	1,927	92,857

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Kazasebepkodu = 39.0 Calisansayisi = 4.0 Cinsiy = 1.0	1,911	70,8

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Kazasebepkodu = 35.0 Nace = 2562.0 Calisansayisi = 4.0 Calismacevrekodu = 11.0	1,873	71,837

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Kazasebepkodu = 35.0 Nace = 2562.0 Genelfaalkodu = 11.0 Egitimmesleki = 1.0 Calismacevrekodu = 11.0	1,873	70,204

Consequent	Antecedent	Support %	Confidence %
Kazadurum = Hafif	Faalkoduozel = 49.0 Sontecrube = 1.0 Genelfaalkodu = 11.0	1,208	70,886

EK-3. GRI Algoritması ile elde edilen birliktelik kuralları örnekleri

Kazasebepkodu & Kazadurum			
File Generate			
Sort by: Support % 100 of 100			
Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Cinsiy = 1,000 Egitimmesleki = 1,000 Cisguzmani = 1,000	48,08	51,85
Kazadurum = Ciddi	Saatisbasi = 3,000 Cinsiy = 1,000 Yaskodu = 3,000	44,94	51,85
Kazadurum = Ciddi	Cinsiy = 1,000 Yaskodu = 3,000 Cisguzmani = 1,000	44,28	52,15
Kazadurum = Ciddi	Yaskodu = 3,000 Egitimmesleki = 1,000 Cisguzmani = 1,000	40,21	52,5
Kazadurum = Ciddi	Calismacevrekodu = 11,000 Toptecrube = 5,000	22,03	53,18
Kazadurum = Ciddi	Toptecrube = 5,000 Egitimmesleki = 1,000	21,91	53,37
Kazadurum = Ciddi	Toptecrube = 5,000 Egitimisg = 1,000 Egitimmesleki = 1,000	21,83	53,27
Kazadurum = Ciddi	Calismacevrekodu = 11,000 Toptecrube = 5,000 Egitimisg = 1,000	21,77	53,32
Kazadurum = Ciddi	Calismacevrekodu = 11,000 Cinsiy = 1,000 Toptecrube = 5,000	21,59	53,29
Kazadurum = Ciddi	Cinsiy = 1,000 Toptecrube = 5,000 Egitimmesleki = 1,000	21,5	53,49
Model	Summary	Annotations	OK

EK-3. (devam) GRI Algoritması ile elde edilen birliktelik kuralları örnekleri

Kazasebepkodu & Kazadurum			
File Generate			
Sort by: Support % 100 of 100			
Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Cinsiy = 1,000 Toptecrube = 5,000 Egitimmesleki = 1,000	21,5	53,49
Kazadurum = Ciddi	Cinsiy = 1,000 Durumogrenim = 3,000 Cisguzmani = 1,000	21,47	53,31
Kazadurum = Ciddi	Calismacevrekodu = 11,000 Toptecrube = 5,000 Egitimmesleki = 1,000	19,36	53,91
Kazadurum = Ciddi	Toptecrube = 5,000 Yaskodu = 3,000 Egitimmesleki = 1,000	18,59	53,29
Kazadurum = Ciddi	Calisansayisi = 4,000 Durumogrenim = 3,000 Cisguzmani = 1,000	18,45	53,42
Kazadurum = Ciddi	Durumogrenim = 3,000 Yaskodu = 3,000 Cisguzmani = 1,000	17,87	54,34
Kazadurum = Ciddi	Calisansayisi = 4,000 Cisguzmani = 1,000 Cisyerihekimi = 2,000	14,71	53,33
Kazadurum = Ciddi	Sontecrube = 2,000 Egitimmesleki = 1,000 Cisguzmani = 1,000	14,7	53,93
Kazadurum = Ciddi	Calismacevrekodu = 11,000 Saatisbasi = 3,000 Toptecrube = 5,000	14,24	53,78
Kazadurum = Ciddi	Saatisbasi = 3,000 Toptecrube = 5,000 Egitimmesleki = 1,000	14,08	54,13
Model Summary Annotations			
OK			

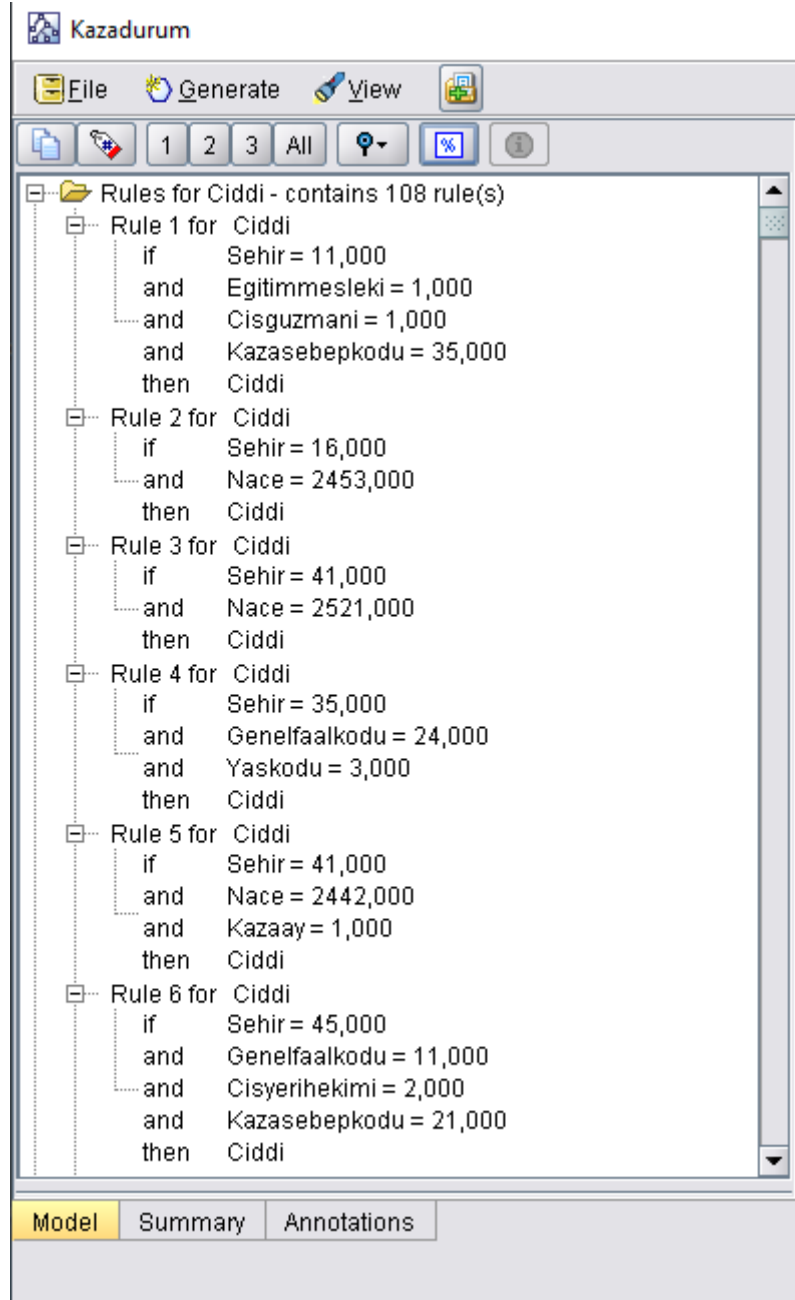
EK-3. (devam) GRI Algoritması ile elde edilen birliktelik kuralları örnekleri

Kazasebepkodu & Kazadurum			
File Generate			
Sort by: Support % 100 of 100			
Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sontecrube = 2,000 Yaskodu = 3,000 Cisguzmani = 1,000	13,98	54,05
Kazadurum = Ciddi	Sehir = 41,000 Genelfaalkodu = 11,000	13,67	53,58
Kazadurum = Ciddi	Saatisbasi = 3,000 Toptecrube = 5,000 Yaskodu = 3,000	13,59	53,82
Kazadurum = Ciddi	Sehir = 41,000 Genelfaalkodu = 11,000 Cinsiy = 1,000	13,19	54,26
Kazadurum = Ciddi	Toptecrube = 5,000 Egitimmesleki = 1,000 Cisguzmani = 1,000	12,65	55,41
Kazadurum = Ciddi	Nace = 2410,000 Egitimmesleki = 1,000	11,32	54,73
Kazadurum = Ciddi	Nace = 2410,000 Egitimisg = 1,000 Egitimmesleki = 1,000	11,29	54,74
Kazadurum = Ciddi	Nace = 2410,000 Cinsiy = 1,000 Egitimmesleki = 1,000	11,27	54,75
Kazadurum = Hafif	Sehir = 16,000	11,12	55,23
Kazadurum = Ciddi	Kazasaat = 3,000 Yaskodu = 3,000 Cisyerihekimi = 2,000	11,09	54,9
Kazadurum = Ciddi	Sehir = 34,000 Yaskodu = 3,000	10,57	54,16
Model Summary Annotations			
OK			

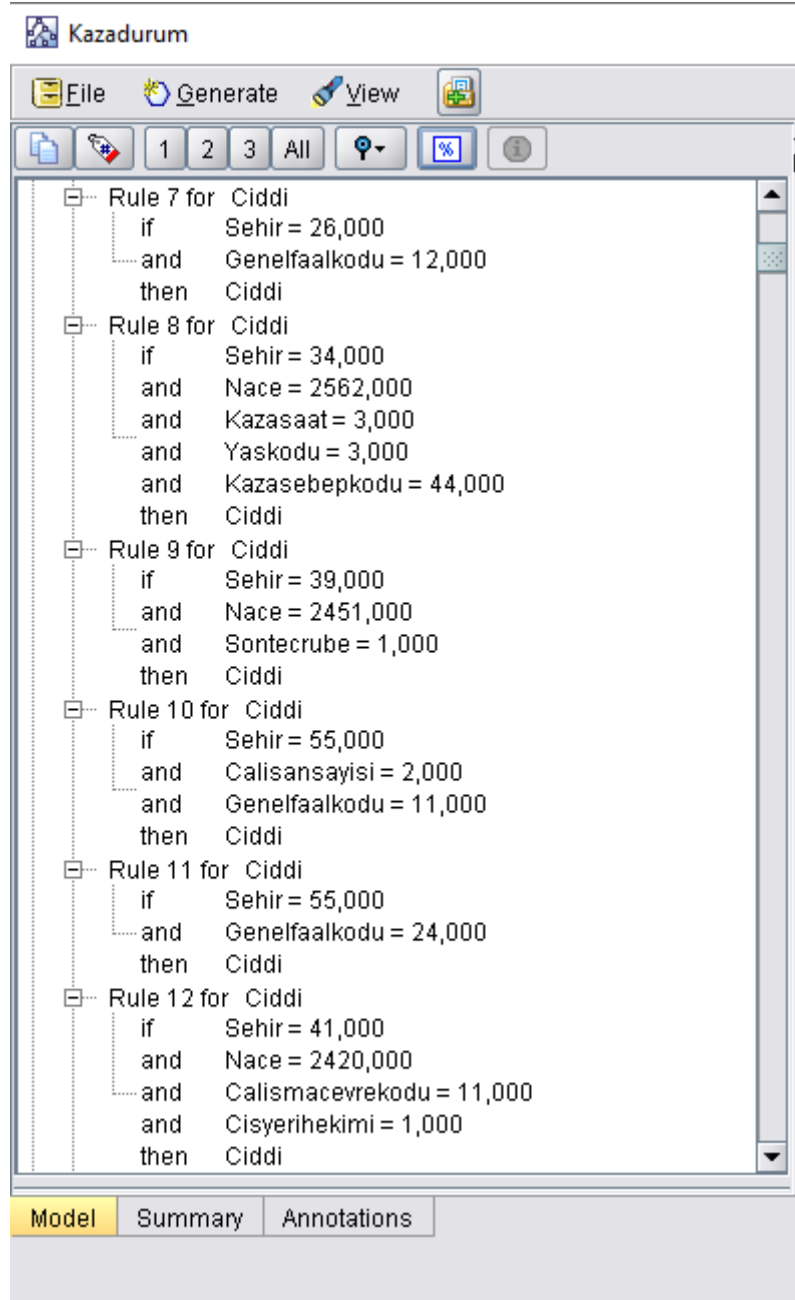
EK-3. (devam) GRI Algoritması ile elde edilen birliktelik kuralları örnekleri

Kazasebepkodu & Kazadurum			
File Generate			
Sort by: Support % 100 of 100			
Consequent	Antecedent	Support %	Confidence %
Kazadurum = Ciddi	Sehir = 34,000 Cinsiy = 1,000 Yaskodu = 3,000	10,01	54,32
Kazadurum = Ciddi	Calisansayisi = 4,000 Nace = 2410,000 Egitimmesleki = 1,000	9,8	54,99
Kazadurum = Ciddi	Nace = 2410,000 Yaskodu = 3,000 Egitimmesleki = 1,000	9,66	55,5
Kazadurum = Hafif	Sehir = 16,000 Genelfaalkodu = 11,000	9,59	57,97
Kazadurum = Ciddi	Toptecrube = 5,000 Durumogrenim = 1,000 Egitimmesleki = 1,000	9,51	56,11
Kazadurum = Hafif	Sehir = 16,000 Calisansayisi = 4,000	9,34	58,07
Kazadurum = Ciddi	Kazasebepkodu = 44,000	9,28	55,27
Kazadurum = Ciddi	Egitimisg = 1,000 Kazasebepkodu = 44,000	9,11	55,2
Kazadurum = Hafif	Sehir = 16,000 Genelfaalkodu = 11,000 Cinsiy = 1,000	9,08	58,25
Kazadurum = Ciddi	Toptecrube = 4,000 Durumogrenim = 3,000 Cisguzmani = 1,000	9,06	55,53
Kazadurum = Ciddi	Sehir = 34,000 Calismacevrekodu = 11,000 Yaskodu = 3,000	9,06	54,68
Model Summary Annotations			
OK			

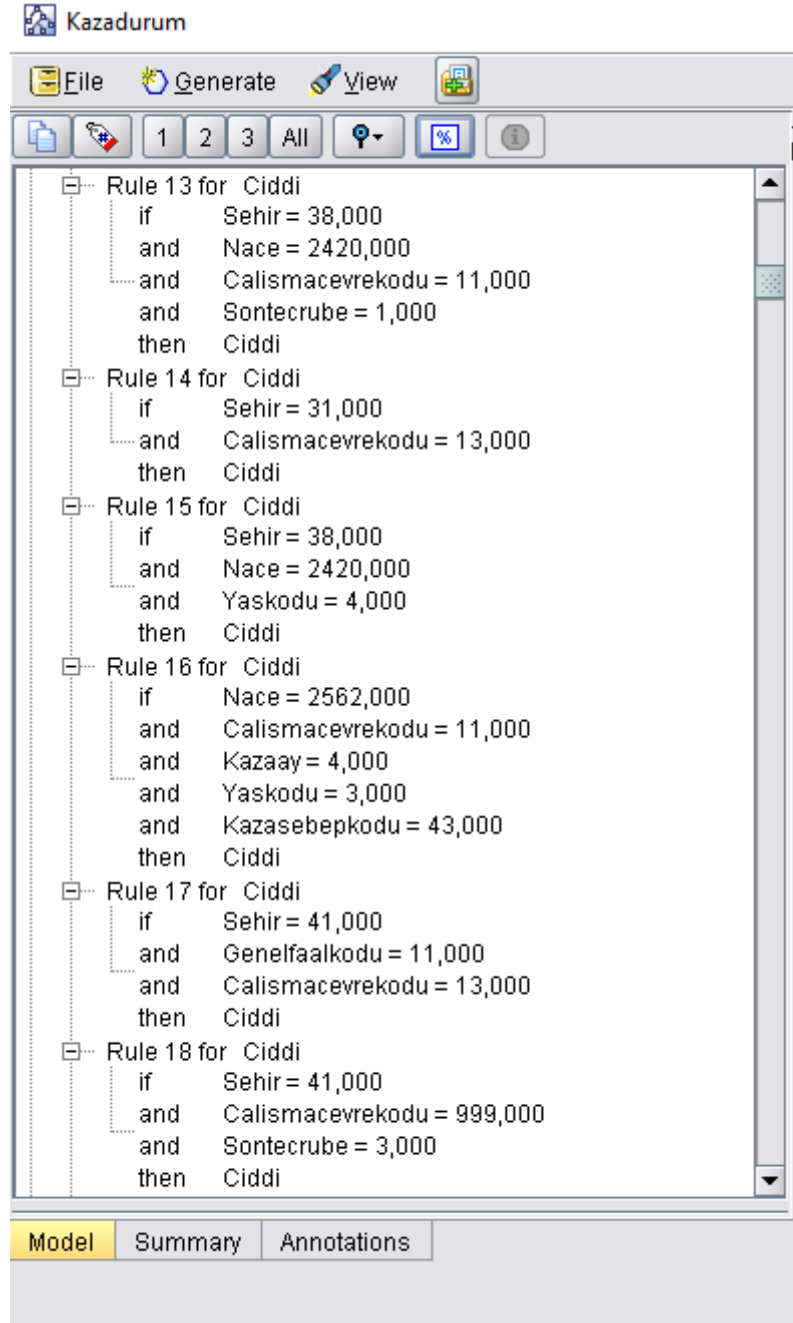
EK-4. C5.0 Algoritması ile elde edilen kural örnekleri



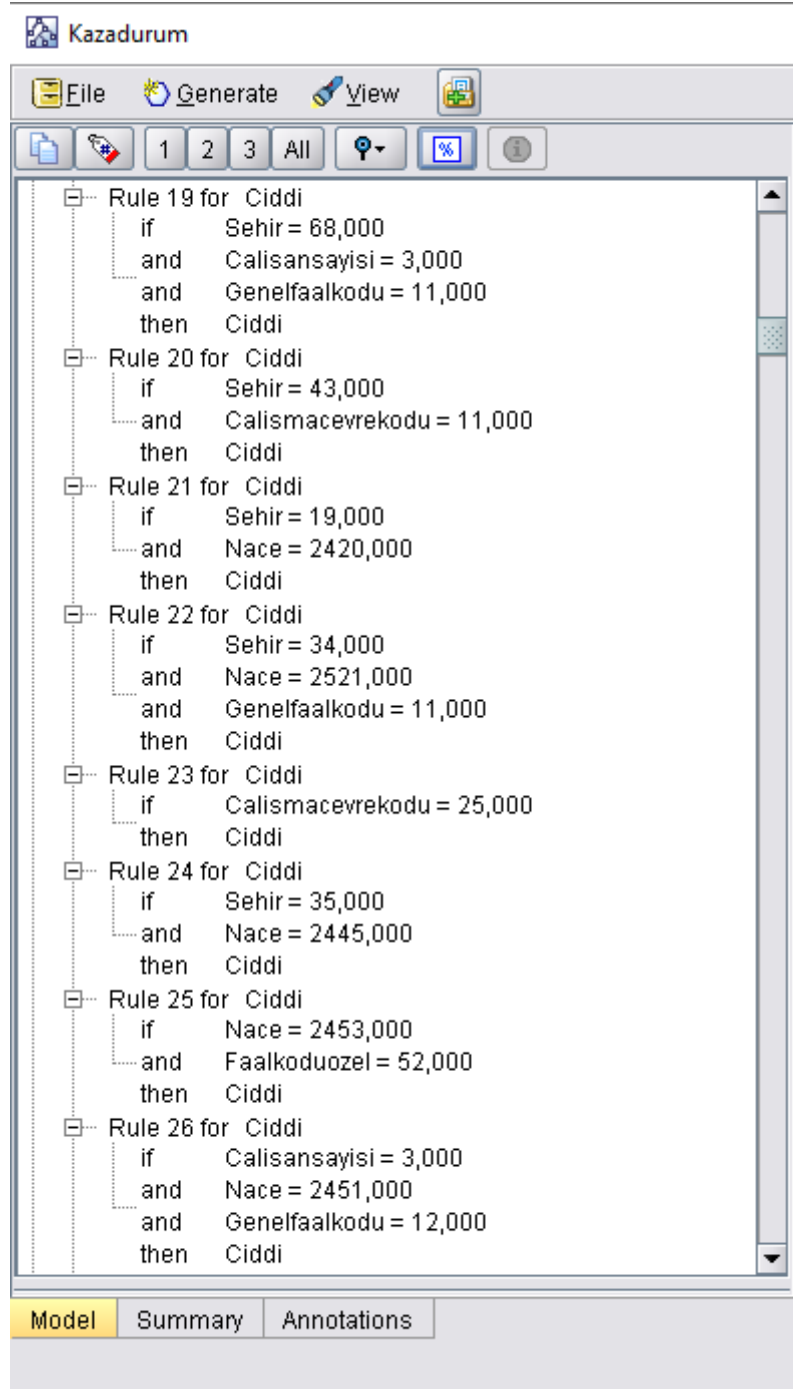
EK-4. (devam) C5.0 Algoritması ile elde edilen kural örnekleri



EK-4. (devam) C5.0 Algoritması ile elde edilen kural örnekleri



EK-4. (devam) C5.0 Algoritması ile elde edilen kural örnekleri



ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : ÇAKMAK, Hasan
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 31.08.1986, Ankara
 Medeni hali : Evli
 Telefon : 0 (555) 624 92 95
 e-mail : hasan2cakmak@yahoo.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / Endüstri Mühendisliği	Devam ediyor
Lisans	Gazi Üniversitesi / Endüstri Mühendisliği	2008
Lise	Mehmet Emin Resulzade Anadolu Lisesi	2004

İş Deneyimi

Yıl	Yer	Görev
2014-Halen	AÇSHB Rehberlik ve Teftiş Başkanlığı	İş Müfettişi
2010-2014	AÇSHB Rehberlik ve Teftiş Başkanlığı	İş Müfettişi Yardımcısı
2008-2010	Algorizma İş Yönetim Sistemleri Ltd.Şti.	İş ve Süreç Analisti

Yabancı Dil

İngilizce, İtalyanca

Hobiler

Bilgisayar Teknolojileri, İş Kazası İncelemeleri, Yüzme, Futbol, Sinema



GAZİ GELECEKTİR..