

**İSTATİSTİKSEL VERİ MADENCİLİĞİ TEKNİKLERİ İLE
KAMU ALIMLARI SEKTÖRÜNDE SATIŞ TAHMİNİ**

**SALES FORECASTING IN PUBLIC PROCUREMENT
SECTOR WITH STATISTICAL DATA MINING
TECHNIQUES**

Vildan YILDIRIM

PROF. DR Süleyman GÜNAY

Tez Danışmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2019

Vildan YILDIRIM' ın hazırladığı “İstatistiksel Veri Madenciliği Teknikleri İle Kamu Alımları Sektöründe Satış Tahmini” adlı bu çalışma aşağıdaki jüri tarafından İSTATİSTİK ANABİLİM DALI' nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

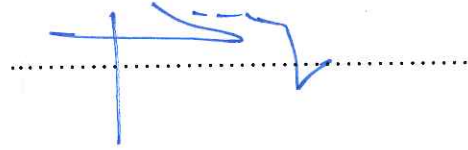
Prof. Dr. Hamza GAMGAM

Başkan



Prof. Dr. Süleyman GÜNAY

Danışman



Doç. Dr. Ayten YİĞİTER

Üye



Doç. Dr. Çağdaş Hakan ALADAĞ

Üye



Dr. Öğr. Üyesi İbrahim ZOR

Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından YÜKSEK LİSANS TEZİ olarak / /..... tarihinde onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU

Fen Bilimleri Enstitüsü Müdürü

Aileme...

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

24 / 09 / 2019


Vildan YILDIRIM

YAYINLANMA FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.

Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ay ertelenmiştir.

Tezim ile ilgili gizlilik kararı verilmiştir.

24/09/2019


(İmza)

Vildan YILDIRIM

ÖZET

İSTATİSTİKSEL VERİ MADENCİLİĞİ TEKNİKLERİ İLE KAMU ALIMLARI SEKTÖRÜNDE SATIŞ TAHMİNİ

Vildan YILDIRIM

Yüksek Lisans, İSTATİSTİK Bölümü

Tez Danışmanı: Prof. Dr. Süleyman GÜNAY

Eylül 2019, 68 sayfa

Günümüz ekonomik koşullarında devletler kamu hizmetlerini sunabilmek için kendi olanaklarının yetmediği durumlarda özel sektör ile ekonomik işbirliği yapmak durumundadır. Devletler ile özel sektörün buluştuğu bu noktada **kamu alımları** kavramı karşımıza çıkmaktadır. Günümüzde kamu alımları, “Kamu İhale Kanunu” çerçevesinde yapılmaktadır. Devletlerin sunduğu hizmetlerin kalitesinin artmasında, dolaylı tasarruf edilmesinde, özel sektörde inovasyonun teşvik edilmesinde, cari açığın kapatılmasında ve yolsuzluğun azaltılmasında önemli rol oynayan **kamu alımları**, ülke ekonomisi için kuşkusuz çok önemli bir paya sahiptir.

Bu çalışma ile kamu alımlarının etkin ve verimli bir şekilde sürdürülebilmesi için “İstisna” kavramı kapsamında merkezi tedarik kurumu olarak yetkilendirilen ve özel sektör ile kamu kurumları arasında aracı rolü üstlenen Devlet Malzeme Ofisi Genel Müdürlüğü satışlarının tahmin edilmesi ve bu tahminlerin duyarlılıklarının araştırılması amaçlanmıştır. İstatistiksel veri madenciliği teknikleri ile elde edilen bu tahminler karar destek sistemine entegre edilerek kurum stratejik hedeflerinin gerçekleşmesi ve izlenmesi sağlanacaktır.

Çalışmanın uygulama kısmında, doğrusal model, genelleştirilmiş doğrusal model ve karar ağaçları algoritmalarından olan random trees ve xgboost trees algoritmaları ile 2007 - 2018 yılları arasındaki veri kümesi için satış tahminleri elde edilmiştir. Böylece bu tez çalışması ile DMO stratejik planında yer alan performans göstergelerinin sağlanmasına ve kamu alımlarının etkinliğinin artırılmasına katkıda bulunulacağı düşünülmektedir.

Anahtar Kelimeler: İstatistiksel Veri Madenciliği, Kamu Alımları, Satış Tahmini, Regresyon Modeli, Doğrusal Model, Genelleştirilmiş Doğrusal Model, Karar Ağaçları, Oracle

ABSTRACT

SALES FORECASTING IN PUBLIC PROCUREMENT SECTOR WITH STATISTICAL DATA MINING TECHNIQUES

Vildan YILDIRIM

Master of Science, Department of Statistics

Supervisor: Prof. Dr. Süleyman GÜNAY

September 2019, 68 pages

In today's economic conditions, states must cooperate with the private sector in economic situations where it is not possible for them to offer public services. At this point where states and private sector meet, the concept of public procurement is emerging. Today public procurement is carried out within the framework of "Public Procurement Law". Public procurement, which plays an important role in improving the quality of services provided by the state, indirect savings, encouraging innovation in the private sector, closing the current account deficit and reducing corruption, undoubtedly has a very important share for the country's economies.

In this study, it was aimed to forecast the sales of State Supply Office, which is authorized as central procurement institution within the scope of "exception" concept and to act as intermediary between private sector and public sector, and to investigate the sensitivities of these forecast in order to sustain public procurement effectively and efficiently with various parameters. These data, forecasted by statistical data mining techniques, will be integrated into the decision support system and the strategic goals of the institution will be realized and followed.

In addition, it was planned to model the sales forecasts with linear model, generalized linear model and random trees and xgboost trees algorithms which are decision trees , using the historical data of 2007-2018 . Thus, it is considered that the key performance indicators included in the institutional strategic plan will be provided and academic contribution will be made to increase the efficiency of public procurement sector.

Keywords: Statistical Data Mining, Public Procurement, Sales Forecasting, Regression Modeling , Linear Model, Generalized Linear Model, Decision Trees, Oracle

TEŐEKKÜR

Yüksek lisans öğrenimimi tamamlayabilmem için beni yönlendiren, motive eden, değerli katkı ve eleştirileri ile bana yol gösteren, zamanını ve desteğini esirgemeyen kıymetli hocam ve danışmanım Sayın Prof. Dr. Süleyman Günay'a;

Bütün hayatım boyunca onların evladı olduğum için şükrettiğim, doğduğum günden bu yana öğrendiğim ve daha öğreneceğim her şeyin mimarı olan annem Gülşen Şenel, babam Halit Şenel ve bana verdikleri en büyük hediye olan kardeşim Cansu Tekin'e;

Hayatımın her anında olduğu gibi tez çalışmam süresince de beni hep destekleyen ve sevgisini esirgemeyerek motive eden yol arkadaşım, sevgili eşim Sertaç Yıldırım'a;

Tezimin her aşamasında bana verdiği destek ve değerli katkılarını unutamayacağım kıymetli arkadaşım Ümit Cengiz Uysal'a;

En büyük motivasyon kaynağım, varlığı yaşama sebebim olan canım oğlum Ömerim ve Keremime;

En içten teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
SİMGELER VE KISALTMALAR.....	xi
1.GİRİŞ	1
1.1.Kamu İşletmeleri İçin Yasal Çerçeve	1
1.2.Türkiye'de ve Dünyada Kamu Alım İstatistikleri	2
1.2.1.Türkiye’de Kamu Alım İstatistikleri.....	2
1.2.2.Dünyada Kamu Alım İstatistikleri	4
2.KAMU ALIMLARI KAVRAMI VE DEVLET MALZEME OFİSİ GENEL MÜDÜRLÜĞÜ.....	6
2.1.Kamu Alımları Kavramı	6
2.2.Devlet Malzeme Ofisi Genel Müdürlüğü	8
3.İSTATİSTİKSEL VERİ MADENCİLİĞİ	10
3.1.İstatistiksel Veri Madenciliğinde Kullanılan Sınıflama ve Regresyon Teknikleri	13
3.1.1.Karar Ağaçları.....	16
3.1.2.Yapay Sinir Ağları	19
3.1.3.K-En Yakın Komşu	20
3.1.4.Destek Vektör Makineleri.....	21
3.1.5.Bayesci Ağlar.....	22
3.1.6.Regresyon Analizi.....	22
3.2.Veri Madenciliği Süreci ve CRISP-DM Döngüsü.....	22
4.UYGULAMA	25
4.1.Veri Kümesinin Analizi ve Değerlendirilmesi	27
4.2.Satış Tahmin Analizleri	30
4.2.1.Araç Satış Tahmin Analizi.....	32
4.2.2.Müteferrik Satış Tahmin Analizi	35

4.2.3.Katalog Satış Tahmin Analizi.....	38
4.2.4.Satış Birimi Bazında Satış Tahmin Analizi	41
4.2.5.İl Bazında Satış Tahmin Analizi.....	43
5.SONUÇ VE TARTIŞMA.....	47
KAYNAKLAR.....	52
EKLER	56
EK 1 – SATIŞ TİPİ SATIŞ TAHMİN ANALİZİ AKIŞI	56
EK 2 – SATIŞ BİRİMİ BAZINDA SATIŞ TAHMİN ANALİZİ AKIŞI.....	56
EK 3 – İL BAZINDA SATIŞ TAHMİN ANALİZİ AKIŞI.....	57
EK 4– GENELLEŞTİRİLMİŞ DOĞRUSAL MODEL ALGORİTMASI İLE MODEL ÇIKTILARI.....	57
EK 5- DOĞRUSAL MODEL ALGORİTMASI İLE MODEL ÇIKTILARI.....	61
EK 6 – XGBOOST TREE ALGORİTMASI İLE MODEL ÇIKTILARI.....	65
EK 7 – RANDOM TREES ALGORİTMASI İLE MODEL ÇIKTILARI.....	66
EK 8- TEZ ÇALIŞMASI ORJİNALLİK RAPORU.....	67
ÖZGEÇMİŞ.....	68

ÇİZELGELER

Çizelge 1.1.	Kamu mal alımları içinde dmo payı	3
Çizelge 4.1.	Yıl bazında kamu alımları istatistikleri	28
Çizelge 4.2.	Kamu alımlarının enflasyon ile korelasyonu.....	29
Çizelge 5.1.	Dört yıllık gerçekleşme ve 2019 yılı tahmin sonuçları	47
Çizelge 5.2.	Araç - katalog - müteferrik için ilk sekiz ay tahmin sonuçları, gerçekleşmeler ve gerçekleşme yüzdeleri	48
Çizelge 5.3.	Dokuz yıllık satış birimi bazında gerçekleşme ve 2019 yılı tahmin sonuçları	49

ŞEKİLLER

Şekil 1.1.	Kamu alımları trendi	2
Şekil 1.2.	Kamu alımlarının GSYİH içindeki payı	3
Şekil 1.3.	Tasarruf tutarı	4
Şekil 1.4.	Kamu harcamalarının yurt dışındaki ülkelerdeki genel durumu	5
Şekil 3.1.	İstatistiksel veri madenciliği	12
Şekil 3.2.	İstatistiksel veri madenciliği teknikleri	14
Şekil 3.3.	Karar ağacı yapısı	16
Şekil 3.4.	Yapay sinir ağları katmanları	20
Şekil 3.5.	CRISP-DM döngüsü	23
Şekil 4.1.	Veri kümesinin yapısı	25
Şekil 4.2.	Kullanılan değişkenler	26
Şekil 4.3.	Satış tipleri bazında satış oranları	27
Şekil 4.4.	Analizde kullanılan süper işlemci	31
Şekil 4.5.	Araç satış tahmin analizi kısım 1	32
Şekil 4.6.	Araç satış tahmin analizi akış kısım 2	33
Şekil 4.7.	Araç satış tahmin analizi akış kısım 3	34
Şekil 4.8.	Araç satış tahmin analizi akış kısım 4	34
Şekil 4.9.	Araç satış tahmini özel durum tanımlama ekranı	35
Şekil 4.10.	Müteferrik satış tahmin analizi kısım 1	35
Şekil 4.11.	Müteferrik satış tahmin analizi kısım 2	36
Şekil 4.12.	Müteferrik satış tahmin analizi kısım 3	37
Şekil 4.13.	Müteferrik satış tahmin analizi kısım 4	37
Şekil 4.14.	Müteferrik satış tahmini özel durum tanımlama ekranı	38
Şekil 4.15.	Katalog satış tahmin analizi kısım 1	38
Şekil 4.16.	Katalog satış tahmin analizi kısım 2	39
Şekil 4.17.	Katalog satış tahmin analizi kısım 3	40
Şekil 4.18.	Katalog satış tahmin analizi kısım 4	40
Şekil 4.19.	Katalog satış tahmini özel durum tanımlama ekranı	41
Şekil 4.20.	Satış birimi bazında satış tahmin analizi kısım 1	41
Şekil 4.21.	Satış birimi bazında satış tahmin analizi kısım 2	42

Şekil 4.22. Satış birimi bazında satış tahmin analizi özel durum tanımlama ekranı	43
Şekil 4.23. İl bazında satış tahmin analizi kısım 1.....	43
Şekil 4.24. Genel akış model aşaması.....	45
Şekil 5.1. Dört yıllık gerçekleşme ve 2019 yılı tahmin sonuçları.....	48
Şekil 5.2. Dokuz yıllık satış birimi bazında gerçekleşme ve 2019 yılı tahmin sonuçları	49
Şekil 5.3. Daha iyi bir tahmin ve daha iyi operasyonel planlamalar	50

SİMGELER VE KISALTMALAR

Kısaltmalar

DMO	Devlet Malzeme Ofisi
CRM	Customer Relationship Management
RFM	Recency, Frequency, Monetary
ERP	Enterprise Resource Planning
DWH	Data Warehouse
ETL	Extract, Transform & Load
OECD	Organization for Economic Cooperation and Development
AB	Avrupa Birliđi
KOBİ	Küçük ve Orta Ölçekli İşletme
EU	European Union
GDP	Gross Domestic Product
SQL	Structured Query Language
CRISP-DM	Cross Industry Standard Process Model for Data Mining
VTBK	Veri Tabanlarında Bilgi Keşfi
KDD	The Knowledge Discovery in Databases

1.GİRİŞ

Kamu alımlarının etkin, verimli ve yerindelik kavramlarına uygun, denetlenebilir bir şekilde sürdürülebilmesi için kesinlikle kamu işletmeleri sistemine ihtiyaç duyulmaktadır. Bu bölümde kamu işletmelerinin yasal çerçevesinin, Türkiye ve dünyadaki diğer örneklerinin incelenmesi yararlı olacaktır.

1.1.Kamu İşletmeleri İçin Yasal Çerçeve

Günümüz ekonomik koşullarında devletler kamu hizmetlerini sunabilmek için kendi olanaklarının yetmediği durumlarda özel sektör ile ekonomik işbirliği yapmak durumundadır. Devletler ile özel sektörün buluştuğu bu noktada kamu alımları kavramı karşımıza çıkmaktadır. Günümüzde kamu alımları, “Kamu İhale Kanunu” çerçevesinde yapılmaktadır [1].

Devletlerin ekonomiye müdahalesinin kaçınılmaz olduğunu savunan ve 1900’ lü yılların ortalarına doğru popülerlik kazanan “Keynesyen Yaklaşım” ile birlikte kamu kaynakları, etkin maliye politika aracı halini almıştır. Bunun bir sonucu olarak da kaynakların etkin dağılımı, ekonomik kalkınma ve istihdam gibi çıktılar üretilmesi beklenmiştir. Kamu harcamaları 1980’li yıllardan günümüze etkisini artıran liberal rüzgârlara karşı ülke ekonomileri içindeki payını sürekli artırmıştır [2].

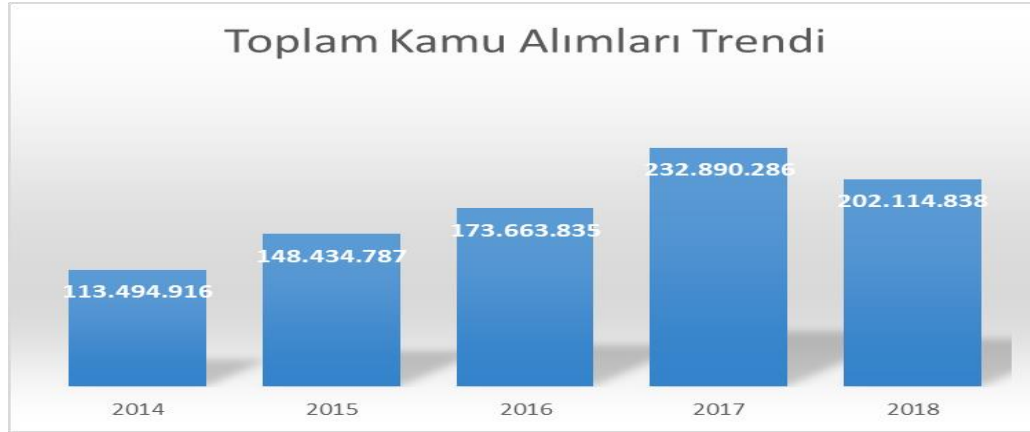
Kamu alımları devletlerin sunduğu hizmetlerin kalitesinin artmasında, kamu kaynaklarından tasarruf edilmesinde, firmalarda inovasyonun teşvik edilmesinde, cari açığın kapatılmasında ve yolsuzlukla mücadelede önemli rol oynamaktadır. Kamu alımları büro işlerinde kullanılacak bir kalemde, otoyol inşasına kadar geniş bir çeşitliliğe sahiptir. Bu nedenle hem alıcı konumundaki devlet hem de ekonomide faaliyet gösteren farklı büyüklüklerdeki tüm özel kesim tarafları için büyük önem taşımaktadır [3].

1.2.Türkiye'de ve Dünyada Kamu Alım İstatistikleri

Kamu yatırımlarının önemli bir kısmının kamu alımları yolu ile yapıldığı günümüzde, kaliteli kamu hizmeti sunabilmek için yurt içinde ve yurt dışında kamu işletmeleri ve tedarik sistemlerinin incelenmesinin en iyi yol olduğu şüphesizdir. Kamu alımlarının orta ve düşük gelirli ülkelerde GSYİH'ye oranı ve nasıl değişim gösterdiği incelenmiştir. Kamu alımlarının rekabeti sağlayarak devlet hizmetlerinin kalitesini arttırmak için bir politika aracı olarak kullanılıp kullanılmayacağı araştırılırken benzer metriklerden yararlanılarak karşılaştırmaların mümkün kılınması sağlanmıştır [4].

1.2.1.Türkiye'de Kamu Alım İstatistikleri

Kamu İhale Kurumu veri kümesine bakıldığında kamu alımları trendi son 5 yılda TL bazında yaklaşık olarak iki kat artış göstererek 113 Milyar TL ' den 202 Milyar TL' ye yükselmiştir. Aynı metrik, dolar (\$) bazında yine son 5 yılda 46,6 Milyar \$ dan 52 Milyar \$ seviyelerine gelmiştir [5].



Şekil 1.1. Kamu alımları trendi

Şekil 1.1.'de kamu harcamalarının bir kısmının yapıldığı yöntemler bütünü olan kamu alımlarının GSYİH içindeki payına bakıldığında son 5 yıl içinde dalgalanmalar gösterdiği gözlemlenmiştir. Şekil 1.2.'de ise sektörün 2017 yılında büyüyerek % 7,5 pazar payına sahip olduğu ancak 2018 de bu payın %5,5 seviyelerine düştüğü görülmektedir [5].



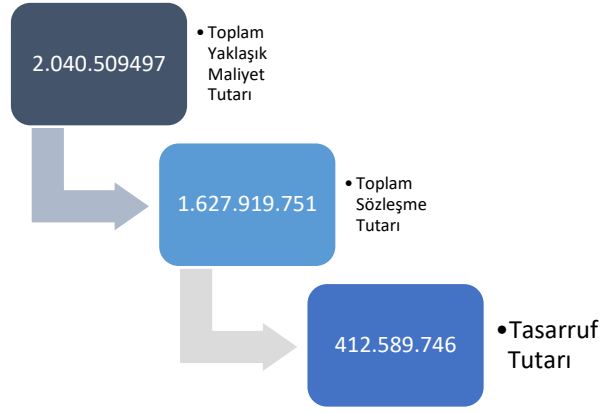
Şekil 1.2. Kamu alımlarının GSYIH içindeki payı

Kamu İhale Kurumu' nun yayınladığı veri kümesi incelendiğinde Devlet Malzeme Ofisi (DMO) 'nin kamu mal alımları içindeki payı 2018 yılı için % 10.33 olarak hesaplanmaktadır. Ancak DMO Ana Statüsü gereği faaliyette bulunulan ürün grubunda bir kısıtlama söz konusudur. Sadece DMO satışlarına konu ürünlere bakıldığında pazar payının daha büyük bir oran olabileceği düşünülmektedir [5].

Çizelge 1.1. Kamu mal alımları içinde DMO payı

YIL	4734 SAYILI KAMU İHALE KANUNU KAPSAMINDA YAPILAN ALIM TUTARI	DOĞRUDAN TEMİN İLE YAPILAN ALIM TUTARI	İSTİSNA KAPSAMINDAKİ TOPLAM MAL ALIM TUTARI	KAMU MAL ALIMLARI TOPLAMI	TOPLAM DMO SATIŞ TUTARI	DMO PAY 1 (İSTİSNA VE DOĞRUDAN TEMİN DAHİL)	DMO PAY 2 (İSTİSNA VE DOĞRUDAN TEMİN HARİÇ)
2014	19.644.253	5.242.381	5.352.893	30.239.527	2.245.833	7,40%	11,40%
2015	24.438.526	5.580.362	7.350.902	37.369.790	3.545.816	9,50%	14,50%
2016	26.400.162	2.249.199	9.090.762	37.740.123	3.738.049	9,90%	14,20%
2017	30.101.939	2.239.287	10.175.571	42.516.797	3.897.240	9,20%	12,90%
2018	33.311.866	3.787.522	15.988.607	51.185.075	3.441.353	6,72%	10,33%

Çizelge 1.1.'de görüldüğü üzere istisna ve doğrudan temin yöntemi ile alımlar hariç DMO 2018 yılı veri kümesine göre % 10.33'lük pazar payına sahiptir. Şekil 1.3.'e bakıldığında ise sadece ihaleye konu ürünler açısından 412.5 Milyon TL tasarruf sağlandığı görülmektedir. Ülke ekonomisine yapılan bu katkının, kurum etkinliğinin artması ile çok daha büyük sayılara ulaşması beklenmektedir [5].



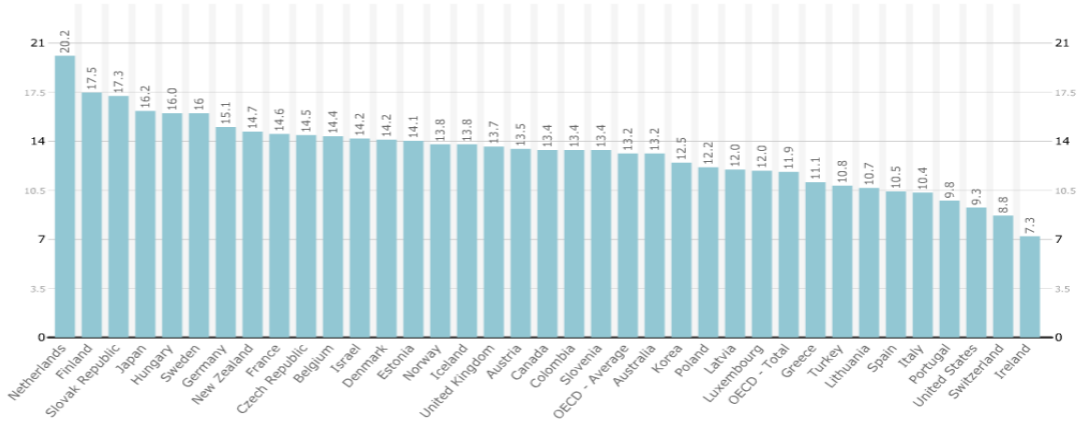
Şekil 1.3. Tasarruf tutarı

1.2.2.Dünyada Kamu Alım İstatistikleri

AB’ de Kamu yatırımlarının büyük çoğunluğu kamu alımları yoluyla yapılmaktadır. Kamu alımları, küresel düzeyde GSYİH’nin %15 - 20’sine, Avrupa Birliği’nde ise %14’üne tekabül etmektedir. Etkin ve kaliteli bir kamu hizmeti sunmak modern ve iyi yönetilen kamu alım süreçleri ile verimli tedarik sistemlerine bağlıdır. Kamu alımlarının iyileştirilmesi büyük tasarruflar sağlayabilir, şöyle ki OECD veri kümesine göre % 1’lik bir verimlilik artışıyla bile yılda 20 milyar € tasarruf sağlanabileceği tahmin edilmektedir [6].

Bu bakış açısı ile AB ve OECD üyesi ülkeler başta olmak üzere küresel düzeyde birçok ülke kamu alım politikalarının etkili, şeffaf, yerindelik kavramına uygun bir şekilde hayata geçirilebilmesi, kamu gereksinimlerinin kaliteli, hızlı ve optimal fiyat yaklaşımı ile karşılanabilmesi amacı ile kamu alım plan ve stratejileri oluşturmaktadır. Benzer şekilde Kamu İhale Kurumu hesaplamalarına göre 2017 yılında 45 Milyar TL tasarruf sağlandığı düşünülmektedir. Bu harcamaların ekonomik faaliyetler içindeki yerini görebilmek açısından 2009-2015 yılları arası OECD veri kümesi incelendiğinde Türkiye’nin de içinde olduğu OECD ülkelerindeki kamu alımlarının kamu harcamalarına oranının %30 seviyesinde olduğu görülmektedir [7,8].

Her yıl Avrupa Birliğinde 250.000’in üzerinde kamu kuruluşu mal, hizmet ve yapım işi alımı yapmaktadır. Eğitim hizmetleri, enerji, güvenlik, bilişim, nakliye ve sağlık hizmeti gibi çeşitli sektörlerde yapılan harcamalar içinde en büyük pay kamu kuruluşlarına aittir. Kamu alımları, vergi mükelleflerinden elde edilen gelirle karşılanan kamu harcamalarının önemli bir kısmını oluşturduğundan hükümetlerin, kaliteli hizmet sunumunu sağlamak ve kamu yararını korumak amacıyla, kamu alımlarını verimli bir şekilde gerçekleştirmesi önem kazanmaktadır. OECD ülkelerinde uygulanan etkili alım yöntemleri neticesinde; ürün ile hizmet sektörlerinde yapılan harcamalardaki %10 oranındaki daralma, kamu harcamaları genelinde %2.9, GSYİH içinde %1.3’lük tasarruf sağlanması demektir. AB üyesi ülkelerde yapılan yıllık kamu harcamalarının neredeyse 2 trilyon € gibi büyük bölümü kamu ihaleleri yoluyla yapılmaktadır [6].



Şekil 1.4. Kamu harcamalarının yurt dışındaki ülkelerdeki genel durumu

Şekil 1.4.’te OECD’nin yayınladığı “Government at a Glance 2017” raporundaki kamu harcamalarının GSYİH’deki payı ile ilgili genel değerlendirmede OECD ülkelerinde 2007’den bu güne kamu harcamalarının GSYİH içindeki payının çok yüksek oranlarla olmasa da artış gösterdiğini görmekteyiz. Bu artış krizin kamu harcamalarını kısmak yerine arttırmak yönünde bir etkisi olduğunu göstermektedir. Türkiye, OECD ortalamasının epeyce altında kalmakla birlikte Yunanistan, Litvanya, İspanya ve İtalya ile ortalama olarak aynı sıralamada yer almaktadır. OECD ülkelerinde, satın alma süreçlerinin %55’inde kamu sözleşmeleri için tek ödül kriteri olarak “en düşük fiyat” kullanılmaktadır. Bu husus, kamu kurumlarının kalite, sürdürülebilirlik, hız ve inovasyona fiyatlar kadar dikkat etmediğine işaret etmektedir [7].

2.KAMU ALIMLARI KAVRAMI VE DEVLET MALZEME OFİSİ GENEL MÜDÜRLÜĞÜ

Önceki bölümde kamu alımları Türkiye’de ve Dünyada sayılarla incelendi. Bu ayrıntıda incelenmesinin nedeni büyük resmi göstererek dünyadaki örnekleriyle karşılaştırma yapabilmektir. Bu bölümde ise kamu alımları kavramı incelendikten sonra Türkiye’de merkezi satın alma kurumu olarak görevini sürdürmekte olan DMO çalışma ilkelerinden söz edilecektir.

2.1.Kamu Alımları Kavramı

Kamu alım kavramı kamu kurumlarının varlığını devam ettirebilmesi için gerekli olan mal, hizmet ve yapım işlerini, ilgili mevzuatta belirtilen usul ve esaslara göre, belli tedarikçiler aracılığıyla temin etmeleri olarak tanımlanabilmektedir. Kamu hizmetlerinin sürdürülebilirliği, yerindeliği gibi temel esasların yanında sektörel, yerel veya bölgesel kalkınma hedeflerinin gerçekleştirilmesi, yerli üretimin teşvik edilmesi, Ar-Ge çalışmalarının güçlendirilmesi ve Küçük ve Orta Ölçekli İşletmelerin (KOBİ) desteklenmesi gibi dolaylı hedeflerin de izlenmesi adına kamu alımları konusunda hazırlanan programlar bulunmaktadır. Ekonomik açıdan ciddi bir oran teşkil eden kamu alımları sadece kamu kaynaklarının etkin, verimli kullanılması, hesap verilebilirliğe uygun, şeffaf bir politika sağlanması değil aynı zamanda ülke olarak siyasi, mali, ekonomik ve sosyal hedeflere ulaşılması gibi amaçlara sahiptir [6].

Bu amaçlar doğrultusunda bazı uluslararası kuruluşlar kamu alımlarında daha etkin politikalar geliştirebilmek adına birtakım düzenlemeler yapmışlardır. *“Bu çerçevede, Dünya Ticaret Örgütü tarafından kabul edilen ve 06.04.2014 tarihinde güncellenerek yürürlüğe giren Kamu Alımları Anlaşması ile daha önce anlaşma kapsamına eklenmemiş kurum ve kuruluşlar kapsama alınarak uluslararası ticaretin geliştirilmesi amaçlanmıştır. Avrupa Birliği Kamu Alımları Direktiflerinde yapılan 2014 yılındaki güncellemelerle birlikte, kamu alımlarında şeffaflığı arttırmak adına elektronik tedarik araçları kullanımının teşvik edilmiş ve 2018 yılına kadar e-tedarik araçlarının kullanımının zorunlu hale getirilmiştir.”* İhale süreleri kısaltılmış ve ihaleler beyan esaslı yapılmaya başlanmıştır. Yeni bir

değerlendirme ölçütü olarak, ihalenin en iyi fiyat-kalite oranı” ya da “maliyet etkinliği yaklaşımına göre en düşük maliyet” esas alınarak sonuçlandırılması ve ihaleler sonuçlandırılırken yaşam döngüsüne dayalı maliyet hesaplamasının yapılabilmesi sağlanmıştır. Kamu alımlarında rekabetin sağlanabilmesi ve ayrımcılığın önlenebilmesi için düzenlemeler yapılmıştır [6].

Avrupa Komisyonunca Ekim 2017'de kabul edilen “Kamu Alım Stratejisi” ile altı stratejik politika önceliği belirlenmiştir. Bu stratejilere göre yenilikçi, yeşil ve sosyal tedarikin geliştirilmesi için ihalelere ilişkin kılavuz belgeler, AB Komisyonunca yenilikçi, yeşil ve sosyal kriterleri içerecek şekilde güncellenerek yayımlanacaktır. Ayrıca sağlık hizmetleri, bilgi teknolojileri ve inşaat gibi stratejik sektörler de dâhil olmak üzere iyi uygulamalar teşvik edilecektir. Kamu kurumlarının kamu alımları için gerekli mesleki ve teknik bilgi ile sürece ilişkin bilgilerindeki yetersizlik ile kurallarının ihlaline ve işletmeler ile vergi mükellefleri üzerinde olumsuz sonuçlara yol açabilmektedir. AB Komisyonu, kamu kurumlarının gerekli bilgi, beceri ve yetkinliğe ulaşılması amacıyla müşteri profesyonelleşmesine yönelik uygulamalar geliştirilmesini teşvik eden bir tavsiye kararı yayımlamıştır. Kamu alımları, KOBİ'ler başta olmak üzere firmalar için daha erişilebilir hale getirilmelidir. Şu anda KOBİ'lerin kazandıkları kamu ihaleleri oranı sadece %45'tir. Ayrıca AB firmaları, AB pazarı dışında kamu ihalelerine katılımları sırasında da birçok engelle karşılaşmaktadır. AB Komisyonu; şeffaflığı, süreçlerin elektronik hale getirilmesini ve stratejik alımları destekleyerek KOBİ'lere yardımcı olmayı, AB firmalarının ticaret anlaşmaları yoluyla AB dışı pazarlara erişimini kolaylaştırmayı amaçlamaktadır. Kamu alımları konusunda daha gelişmiş ve erişilebilir veri kümesi, tedarik politikalarının performansını daha iyi değerlendirebilmeyi, kamu alım sistemleri arasındaki etkileşimi optimize etmeyi ve gelecekteki stratejik kararları şekillendirmeyi mümkün kılacaktır. Şeffaflık ve bütünlüğün sağlanması politikası doğrultusunda Komisyon, kamuya açık sözleşme bilgilerinin raporlanması gerçekleştirilen sözleşmeler ve değişikliklerin yayınlanması ve şikâyet mekanizmasının güçlendirilmesi ile rekabet ve şeffaflığı arttırmayı ve kamusal kaynak tasarrufu sağlamayı amaçlamaktadır. Dijital dönüşümün artırılması kapsamında AB Komisyonu e-Certis, Avrupa Tek Alım Belgesi (ESPD) ve e-Faturalama gibi e-satın alma araçlarının kullanımının geliştirilmesini ve yeni teknolojilerden yararlanılmasını amaçlamaktadır [6].

2.2.Devlet Malzeme Ofisi Genel Müdürlüğü

Ülkemizde kamu alımları alanında farklı görevlere sahip kurumlar olmakla birlikte kamu alım politikasını belirlemekle görevli olan kurum Maliye Bakanlığıdır. Kamu alımları konusunda sorumlu olan ve kamu alımlarını düzenleyip denetleyen kurum ise Kamu İhale Kurumu (KİK) dur. *“Ülkemizin merkezi satın alma kurumu olan Devlet Malzeme Ofisi tüzel kişiliğe sahip, faaliyetlerinde özerk ve sorumluluğu sermayesiyle sınırlı bir İktisadi Devlet Teşekkülüdür.”* DMO'nun temel görevi kamu kurumlarının ihtiyacı olan, ana statüsünde belirtilen mal ve hizmetleri, bu kurumlar adına tedarik etmektir. DMO'nun misyonu *“Kamu kurum ve kuruluşlarının ihtiyaçlarını; kamu yararını gözeterek, fiyat ve kalite dengesini sağlayan, verimli, etkili, profesyonel ve yenilikçi bir yaklaşımla tedarik etmektir.”* Vizyonu ise *“Kamu kurum ve kuruluşlarının ihtiyaçlarını; kamu yararını gözeterek, fiyat ve kalite dengesini sağlayan, verimli, etkili, profesyonel ve yenilikçi bir yaklaşımla tedarik etmektir.”* Bu misyon ve vizyon doğrultusunda faaliyetlerini gerçekleştirmekte olan DMO sürekli gelişen, yenilenen politikaları ile kamu alımları sektöründe önde gelen ve dünya çapında tanınan bir merkezi tedarik kurumu olmayı hedeflemektedir [6].

Yetkilerini ofis ana statüsünden alan DMO kamu hukukuna tabi veya kamu denetiminde olan ya da kamu kaynağı kullanan kamu kurumlarının ihtiyaç duyduğu ürün ile hizmetlerin tedarik edilmesinde ve dağıtılmasında yetkili olan bir kurumdur. Tedarik ve dağıtım hizmetlerini merkezde satış ve satın alma birimleri, taşrada bölge müdürlükleri ve irtibat büroları vasıtasıyla sürdürülmektedir. Basım İşletme Müdürlüğünde ise klasör, toplu iğne, ataş, zarf gibi ürünler üretilmektedir. Bu hizmetlere ek olarak Yüksek Seçim Kurulu seçim faaliyetlerini yürütüyorken seçim malzemelerinin üretimi, tedariki, il ve ilçe seçim kurullarına ulaştırılması, sayısal lotodan kazı kazana, kazı kazandan milli piyango biletlerinin kâğıtlarına kadar Milli Piyango İdaresinin ihtiyaçlarının giderilmesi hizmetlerini de üstlenmektedir [6].

DMO da katalog, müteferrik ve stok olmak üzere üç tip satış yöntemi vardır. Kamu kurumlarının ihtiyaçlarına yönelik alımı toplu halde yapılan ürünler stok ürünleri ve bu ürünlerin depodan yapılan satışları stok satış yöntemi olarak tanımlanmaktadır. Tedarikçi firmalar ile imzalanan dönemsel sözleşmelere göre alınan ürünler katalog ürünleri ve bu

ürünlerin katalogdan yapılan satışı katalog satış yöntemi olarak tanımlanmaktadır. Müşterilerin taleplerini bu iki satış yönteminden de karşılayamadıkları durumlarda marka ya da modelini belirterek talep ettikleri ürünlere ise müteferrik ürünler ve bu ürünlerin satışı da müteferrik satış yöntemi olarak tanımlanmaktadır [6].

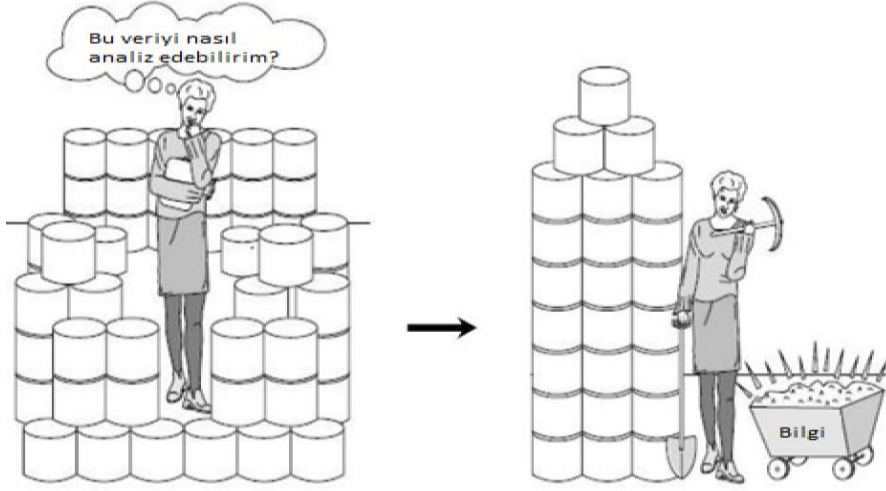
3.İSTATİSTİKSEL VERİ MADENCİLİĞİ

İstatistiksel veri madenciliği istatistik biliminin günümüz gelişen teknolojisi ile yoğrulması sonucunda ortaya çıkmış teknikler bütünüdür. Gelişen bilgi teknolojileri ile yeni programların ortaya konulması, çalışmaları kolaylaştırmıştır. Ancak istatistiksel veri madenciliği ile istenilen sonuçlara ulaşmak için geniş bir iş bilgisine, sorunları tanımlamak ve çözmek için analitik karar verme yetisine ve akademik alanda edinilmiş istatistik bilgisine ihtiyaç vardır. Rekabetin artması ve veri kümesi depolama kapasitesi taşıyan araç teknolojilerindeki gelişme birçok kurum ve firmayı karmaşık veri tabanları ve bu veri tabanlarından çok hızlı bir şekilde anlamlı bilginin elde edilmesi gereksinimi ile karşı karşıya getirmiştir. İşte bu noktada istatistiksel analiz ve modelleme ile istatistiksel veri madenciliği uygulamaları gün yüzüne çıkmıştır. Böylece veriye dayalı kararların kalitesi artacak, güvenilirliği sağlanacak ve bu kararlar ile geleceğine yön veren kurumların değer yaratma potansiyeli de gelişmiş olacaktır [9,10].

Son zamanlarda adını sıkça duyduğumuz istatistiksel veri madenciliğinin tarihi çok kısa bir süre önce başlamış gibi düşünülse de 1700' lere dayanmaktadır. Bayes Teoremi 1700 'lü yıllarda ve 1800 'lü yıllarda Regresyon Analizi gibi veri kümesinin içindeki örüntüleri (pattern) tanımlayan istatistiksel veri madenciliği metotları ile başlayıp günümüze kadar gelişerek gelen bir disiplindir. Basit sayma ve aritmetik işlemler geliştirmek için ilk bilgisayarlar 1950'li yıllarda kullanılmaya başlamıştır. Veri tabanı ve veri kümesi depolanması kavramları 1960'lı yıllarda kullanılmaya başlamıştır. Veri tabanlarının yalnızca veri depolama aracı olmadığı, işlem yapma kapasitelerinin ön plana çıkması ve bu bağlamda kullanılmaya başlanması ise 1960'ların sonunda gerçekleşmiştir. Standart sorgulama dili olan SQL (structured query language) 1980'lerde kullanılmaya başlanmıştır. Bu yıllarda şirketlerin ve kurumların müşterilerini değerlendirdikleri ve müşterilerle ilgili stratejiler oluşturdukları bugün ki müşteri ilişkileri yönetimi projelerinin alt yapısını oluşturan veri tabanları oluşturulmuştur. Veri tabanı yönetim sistemleri 1990'larda sıklıkla kullanılmaya başlanmıştır. Bu gelişmelerle birlikte sonsuz sayıda veri kümesi içeren veri tabanlarından anlamlı bilgi üretmenin sorgu ve raporlamalarla kolay olmayacağı kabul edilmiş ve yeni yöntemler çalışılmaya başlanmıştır. *“Gregory Piatetsky-Shapiro'nun organize ettiği ilk KDD (Knowledge Discovery in Databases) çalışmayı 1989 yılında gerçekleştirmiştir. Bu çalıştay*

1995 yılında, ACM SIGKDD Conference on Knowledge Discovery and Data Mining adı altında düzenli hale getirilmiştir. (IJCAI)-89 Veri Tabanlarında Bilgi Keşfi Çalışma Grubu toplantısı ve 1991, KDD (IJCAI)-89'un sonuç bildirgesi sayılabilecek Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop makalesinin KDD ile ilgili temel tanım ve kavramları ortaya koyması ile süreç daha da hızlanmış ve nihayet 1992 yılında istatistiksel veri madenciliği için ilk yazılım gerçekleştirilmiştir." İstatistiksel veri madenciliği 1990'ların sonunda başlayarak 2000'lerde devamlı gelişerek pek çok alanda kullanılmaya başlanmıştır. İstatistiksel veri madenciliği teknikleri kullanılarak geleceğe yönelik kararlar alabilen şirket ya da kurumlar daha etkili ve tutarlı kararlar alarak sektörde öne çıkabilmişlerdir. Popülarlığı gün geçtikçe artan sosyal medya şirketlerinin istatistiksel veri madenciliği alanında ciddi harcamalar yapmaları alana olan ilginin çok daha artmasına neden olmuştur. Zamanla gelişen teknoloji ve farklı ihtiyaçlar metin madenciliği, görsel madenciliği, grafik madenciliği gibi yeni kaynaklardan madencilik yapılması ihtiyacını doğurmuştur [11,12].

Dijital dönüşüm yaşadığımız günümüzde rekabet koşulları ve gelişen teknoloji ile artan veri kümesi miktarı dikkate alındığında, ham veri kümesinden bilgi üretme sürecinin ne kadar değerli olduğu anlaşılmaktadır. Veri kümelerini anlamlı bilgiye dönüştürme sürecinin en önemli aracı ise ekonomi, istatistik, veri tabanı teknolojileri, makine öğrenmesi, yapay zekâ ve veri görselleştirme gibi değişik disiplinlerden yararlanabilen istatistiksel veri madenciliği yöntemi kullanılmaktadır. Şekil 3.1. de veri yığından anlamlı bilgiye erişim süreci görülmektedir [13].



Şekil 3.1. İstatistiksel veri madenciliği

“İstatistiksel veri madenciliği, veri tabanlarında, veri ambarlarında ve diğer veri depolarında saklanan büyük miktardaki veri kümesinden ilginç bilgi keşfetme sürecidir. Veri tabanından bilgi keşfi olarak da bilinen istatistiksel veri madenciliği terimi, veri tabanlarındaki veri kümelerine ait çok önemli örtük, daha önce bilinmeyen ve potansiyel faydalı bilginin veri tabanından çıkarılmasıdır. Veri kümelerinden modeller keşfetme sürecidir. Belirli bir veri kümesindeki çeşitli modelleri, özetleri ve türetilmiş değerleri keşfetme işlemidir. İstatistiksel veri madenciliği büyük miktarda veri kümesinin çeşitli yöntemler ile analiz edilmesi ve çıkan sonuçların bir uzman gözüyle yorumlanmasıyla geçmiş veri kümesinden gelecek tahminleri yapmaya yarayacak bilgi edinilmesi işlemi şeklinde de tanımlanabilir.” İstatistiksel veri madenciliğine ait literatürde yapılmış tanımlara yukarıda yer verilmiştir ancak özetlemek gerekirse en genel haliyle istatistiksel veri madenciliği, veri tabanlarında daha önce keşfedilmemiş anlamlı bilginin elde edilmesi için kullanılan yöntemler bütünü olarak tanımlanabilir. Burada vurgulanmak istenen anlamlı bilgi veride önceden bilinmeyen veya şüphelenilmeyen örüntü (pattern) ve ilişkileri içermektedir. İstatistiksel veri madenciliği literatürde ‘Veri Tabanlarında Bilgi Keşfi’ (VTBK) ya da ‘The Knowledge Discovery in Databases’ (KDD) olarak da karşımıza çıkmaktadır. *“Dünya ile ilgili en anlaşılmaz şey, her şeyin tamamen anlaşılabilir olmasıdır.” – Albert Einstein. “İstatistiksel veri madenciliği dünyanın anlaşılabilirliğine önemli ölçüde destek olan bir kavramdır”.* İstatistiksel veri

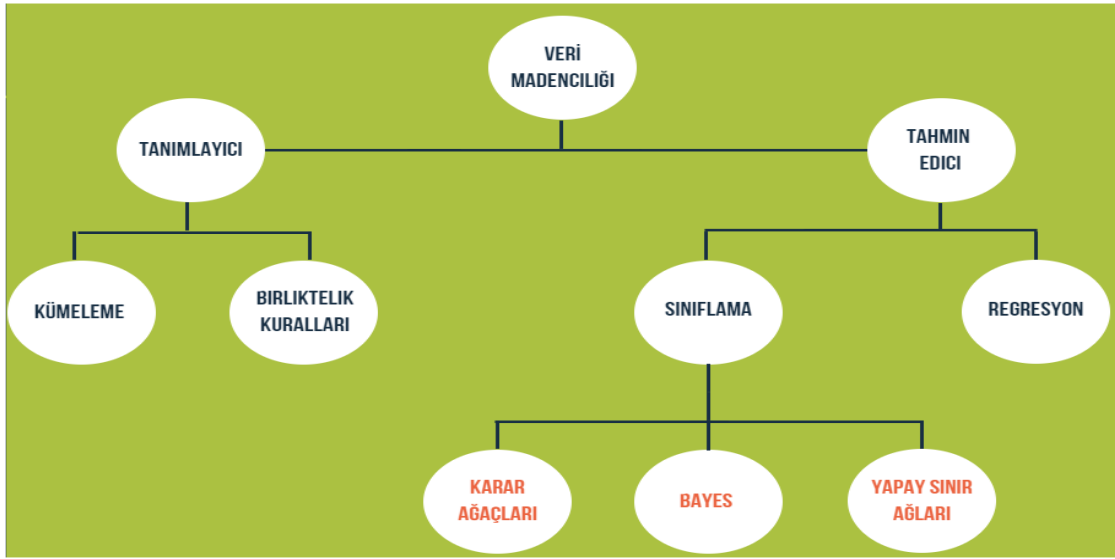
madenciliğinin amacı sonsuz sayıdaki veri kümesini anlayıp yorumlayarak bilgiye ulaşmaktır. Bilgiye ulaşırken asıl amaçlanan daha önce keşfedilmemiş olan bilgiye erişebilmektir. Bir kurumun veri tabanlarında biriken veri kümesi o kurumun geleceğidir. Bu veri kümesi bilgiye dönüştürebilen kurumlar sorunlarını daha hızlı, etkin ve doğru biçimde çözerek piyasada bir adım öne çıkabilmişlerdir [14,15,16,17,18].

İstatistiksel veri madenciliği uygulamaları hemen hemen her alanda kullanılmaktadır. Bir televizyon kanalının her dakikasının izlenme sıklığından, bir web sayfasının tıklanma sıklığına, bir şehirde herhangi bir ay içinde ne kadar yağmur damlası düşeceğinden, bir markette en çok satılan yumurta markasına kadar birçok veri kümesi kayıt altına alınmaktadır. Bir diğer uygulama ise analitik yöntemlerle müşteri veri kümelerinin sahip olduğu örüntüler keşfedilerek elde edilen müşteri ilişkileri yönetimi ile müşterileri bölümlere ayırarak terk etme eğiliminde olan müşterileri belirlemektir. Gelişmiş müşteri ilişkileri yönetimi programları ile müşteri bölümlerinin belirlenmesi, düşük bölüme geçen müşterinin düşüş nedeninin araştırılması ve aksiyon planlarının uygulanması müşteri veri kümelerinin örüntülerinin keşfedilmesi ile mümkündür. Bunlara ek olarak otomotiv sektöründe, risk analizi uygulamalarında, konuşma tanıma alanında, fen bilimleri ve tıp alanında yapılan çalışmalarda, pazarlama, sigorta, bankacılık ve telekomünikasyon alanlarında istatistiksel veri madenciliği uygulamalarından sıklıkla faydalanılmaktadır. Dijitalleşen dünyada kaydedilen tüm veriler sayılara dönüşmekte ve bu sayılar bilgiye dönüştüğünde kurumlara stratejik kararlar alma aşamasında destek olmakta ve kurumları sektörde bir adım öne çıkartmaktadır [17,19].

3.1.İstatistiksel Veri Madenciliğinde Kullanılan Sınıflama ve Regresyon Teknikleri

İstatistiksel veri madenciliği teknikleri tanımlayıcı teknikler ve tahmin edici olmak üzere iki kısımda incelenmektedir. Tanımlayıcı teknikler kümeleme analizi, birliktelik kuralları gibi teknikleri içerirken; tahmin edici teknikler sınıflama ve regresyon tekniklerini içerir. Tahmin edici istatistiksel veri madenciliği teknikleri ile sonucu bilinen veri kümeleri ile analizler yapılarak sonucu bilinmeyen veri kümeleri için tahminler yapılmaktadır. Örneğin bir şirkette gerçekleşen satış sonuçlarından hareketle bir model kurulması ve geliştirilmesi ile gelecekte satış tutarlarının ne olacağı tahmin edilerek şirketin mali yönetiminin nasıl şekilleneceğine karar verilmektedir. Tanımlayıcı istatistiksel veri

madenciliği teknikleri ise veri kümesindeki örüntüleri keşfederek veri kümesinin genel özelliklerini karakterize etmeyi amaçlanmaktadır. Veri kümesinde saklı örüntülerden yararlanılarak müşteri harcama davranışları incelenerek benzer müşterilerin benzer ürünleri tercih etmelerinin gözlemlenmesi tanımlayıcı modellere örnek olarak verilebilmektedir. Şekil 3.2. de veri madenciliği teknikleri tanımlayıcı ve tahmin edici olarak bölümlere ayrılmıştır. Bu bölümde tahmin edici sınıflama ve regresyon modeli istatistiksel veri madenciliği teknikleri anlatılacaktır [18,19,20,21,22].



Şekil 3.2. İstatistiksel veri madenciliği teknikleri

Sınıflandırma problemi, öğrenme kümesi de denilen nesnelere oluşan bir veri kümesinden ve her nesnenin niteliklerine göre bir sınıf bilgisinden oluşur. Sınıflama ve regresyon gelen yeni veri kümesinin; önceden tanımlanmış sınıflara uygun olan sınıfa atanmasıdır. Burada esas amaç mevcut veri kümesinden yola çıkarak gelecekteki hareketin tahmin edilmesidir. Veri kümesinin dağılımına göre bir model elde edilir. Modelin başarısı test edilir ve model geleceği ya da bilinmeyeni tahmin etmek amacıyla kullanılır. İstatistiksel veri madenciliği uygulamalarında ayrık nitelik değerlerini tahmin etmek sınıflandırma olarak tanımlanır, sürekli nitelik değerini tahmin etmek ise öngörü. Örnek vermek gerekirse, hangi topun hangi sepete konulması gerektiği sınıflama, topun ağırlığı ise öngördür. Sınıflandırma işlemi üç aşamadan oluşur bunlardan ilki model oluşturmaktır. Model oluşturulurken her nesnenin sınıf etiketi olarak tanımlanan ve belli nitelikler özelinde belirlenmiş olan bir sınıfta olduğu varsayılır. Modeli oluşturan

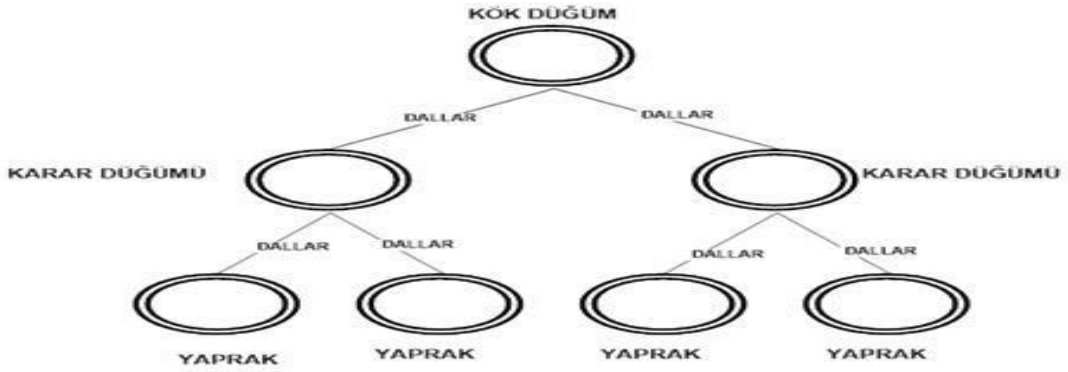
nesneler öğrenme kümesi olarak tanımlanır. Ve model eğer-ise-değilse kuralları, karar ağaçları ve farklı istatistiksel formülleri ile ifade edilebilir. İkinci adım ise model değerlendirme aşamasıdır. Oluşturulan modelin başarısı sınama örnekleri kullanılarak belirlenir. Sınıf etiketi bilinen bir sınama kümesi örneği model kullanılarak belirlenen sınıf etiketi ile karşılaştırılır. Ve nihayetinde modelin doğruluğu, doğru sınıflandırılmış sınama kümesi örneklerinin toplam sınama kümesi örneklerine oranı alınarak belirlenir. Üçüncü adım ise modeli kullanılma aşamasıdır. Model önceden bilinmeyen veri kümesinin sınıflandırılması için kullanılır. Veri kümesinin sınıf etiketini ve niteliğinin değerini tahmin eder. Sınıflandırıcının başarısı ise hız, kararlılık, ölçeklenebilirlik, anlaşılabilir olması ve kurallarının yapısı ile değerlendirilir. Modelin meydana gelebilmesi ve sınıflandırma yapılabilmesi için gereken sürenin minimumda olması hız; veri kümesinin sağlıklı oluşturulması ve eksik olmayan nitelik değerlerine sahip olması kararlılık; büyük miktarda veri kümesi ile çalışıyor olması doğru sonuçlar ortaya koyabilmek ölçeklenebilirlik; kullanıcı tarafından yorumlanabiliyor olması anlaşılabilirlik; birbiriyle örtüşmeyen kurallar olmaması kuralların yapısı olarak nitelendirilmektedir. Sınıflama modelleri istatistiksel veri madenciliği uygulamaları içinde en çok tercih edilen tekniklerdir. Sınıflama modellerinde kullanılan başlıca teknikler Karar Ağaçları (Decision Trees) , Yapay Sinir Ağları (Artificial Neural Networks) , K-En Yakın Komşu (K-Nearest Neighbor) , Destek Vektör Makineleri (Support Vector Machines) , Genetik Algoritmalar (Genetic Algorithms) , Naive-Bayes teknikleridir [22,23].

Bu tez çalışmasında 2007-2018 geçmiş dönem satış veri kümesi kurumsal kaynak planlama veri tabanından SQL sorgularıyla elde edilmiş ve veri kümesi satış birimleri ve aylık bazda derlenerek IBM SPSS Modeler uygulaması ile modellenmiştir. 10 yıllık geçmiş satış veri kümesinden hareketle 2019 yılı satışları tahmin edilmiş ve ilk 7 ay gerçekleşme sonuçları ile karşılaştırması yapılmıştır. Tahmin edici istatistiksel veri madenciliği yöntemlerinden karar ağaçları ve regresyon teknikleri ile çalışmalar yapılmıştır.

3.1.1.Karar Ağaçları

Temel makine öğrenme algoritmalarından olan karar ağaçları, adını gerçek ağaç yapısına benzemesi sebebiyle almıştır. Büyük veri kümelerinde bazı gizli kurallar vardır. Makine öğrenmesi algoritmalarının en önemli ideali büyük veri kümelerinde bulunan bu kuralları yakalamaktır. Bu kurallar ise ancak büyük veri kümelerini birden fazla anlamlı alt gruba bölerek yakalanabilmektedir. Karar ağaçları gerçek ağaç yapısı gibi bir akış şemasına sahiptir. Diğer algoritmalara göre daha kolay yorumlanabilir olması da karar ağaçları algoritmalarını daha popüler bir hale getirmiştir. Karar ağacı algoritmaları tıp, imalat ve üretim, finansal analiz, astronomi ve moleküler biyoloji gibi birçok alanda kullanılmaktadır [20].

Karar ağaçları hiyerarşik yapıyı temsil eden düğümlerden yapılır. Bu yapıdaki üç önemli düğüm kök düğümü, ara düğüm, yaprak veya uç düğüm olarak adlandırılmaktadır. Kararın başlangıç noktası olan kök düğümleri en alt seviyeyi, ağacın yaprak düğümü ile kök düğümü arasındaki bağlantıyı sağlayan düğme karar düğümü, ağacın sonu olan yaprak düğümleri ise sorumlu değerler olarak tanımlanmaktadır. Sınıflandırma modellerinde karar ağaçları temel bir kural kümesi oluşturur. Temel karar ağacı yapısı Şekil 3.3. te görülmektedir [24].



Şekil 3.3. Karar ağacı yapısı

Karar ağacı algoritmaları ağacın kök düğümü, dal ve yapraklarının belirlenme şekline göre gruplara ayrılır. Bunlar regresyon algoritmaları ve bilgi kazancına dayalı

sınıflandırma algoritmalarıdır. Bilgi kazancına dayalı sınıflandırma algoritmalarına örnek ID3 (Iterative Dichotomiser 3) ,C4.5 ve C5.0 algoritmaları olarak verilebilir. Sınıflandırma ve regresyona dayalı algoritmalara örnek ise Twoing, Gini ve CART (Classification And Regression Trees) algoritmalarıdır. İstatistik bazlı algoritmalar Bayesyen sınıflandırması ve CHAID; bellek tabanlı sınıflandırma algoritmalarına örnek olarak ise en yakın k-en yakın komşu algoritması örnek verilebilir. Bilgi kazancı (entropy) ağaç yapısının nasıl oluşturulacağını belirlemede kullanılan bir ölçüttür. Bilgi kazancı ölçüsü ne kadar fazla ise belirlenmeye çalışılan alan da o kadar karmaşıktır. Buna göre bilgi kazancı ölçüsü en az olan alan tercih edilir. Karar Ağaçları teknikleri ile ağacın kök düğümü, dal ve yaprakların belirlenme şekline göre çeşitli algoritmalar çalışılmaktadır. Bunlar ID3 Algoritması, C4.5 Algoritması, CART (Classification and Regression Trees) Algoritması ve CHAID (Chi-Squared Automatic Interaction Detector) algoritmasıdır [24].

ID3 algoritması : ID3 algoritmasında ağaç yapısında kökün hangi özellik olacağı ve kökten sonra bölümlenmenin nasıl gerçekleşeceğine karar verilirken entropi ve bilgi kazanç hesaplamaları kullanılır. Entropi rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Entropi kavramı Shannon tarafından bilgisayar bilimine uyarlanmıştır ve sıklıkla kullanılmaktadır. Bilgisayar biliminde entropi veri kümesinin karmaşıklık ölçümüdür. ID3 algoritmasında veri setindeki her bir özelliğin bilgi kazanç değerleri hesaplanır. Bilgi kazancı en yüksek olan özellik kök düğümü olarak kabul edilir. Ağacın inşa edilmesi sürecindeki sonraki aşamalarda da de yine aynı bilgi kazanç değerlerine göre hareket edilir. Entropi ve bilgi kazanç değerlerine dayalı olmasından dolayı ID3 karar ağacı algoritmasına entropi temelli karar ağacı algoritması adı verilir [25].

C4.5 algoritması ise ID3 algoritmasının devamı niteliğinde ve ID3'ün geliştirilmiş bir versiyonu olarak kabul edilen bir karar ağacı algoritmasıdır. ID3 algoritması veri kümesi içerisinde sayısal veri türüne sahip özelliklerin analizine olanak tanımamaktadır. C4.5 algoritması işleyiş ağaç yapısının oluşturulması gibi yönlerden ID3 ile aynı yolu izler sadece bu algorithmada sayısal veri türündeki özelliklerin de algoritma tarafından analizi olanaklı hale gelmiştir [23].

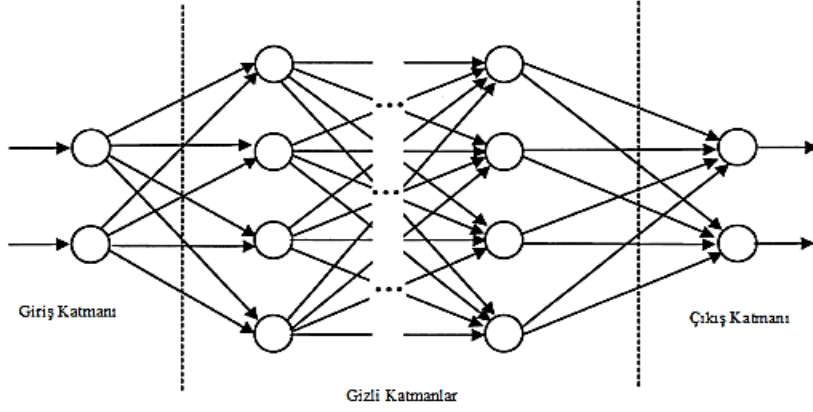
CART (Classification and Regression Trees) Algoritması : CART karar ağacı algoritmaları parametrik olmayan bir tekniktir ve çok sayıda değişken arasından seçim yapılabilmektedir. Burada ki bağımlı değişken kategorik ya da sürekli olabilmektedir. Bağımlı değişken sürekli olduğunda regresyon ağacı; kategorik olduğunda sınıflandırma ağacı oluşturmaktadır. CART karar ağacı algoritmasında her bir düğüm iki dala ayrılarak süreç devam eder ve ağaç yapısında ikili dallanmalar meydana gelir. Ağacın dalları oluşturulurken bölümlenme sürecinde twoing, gini, regresyon ağaçları gibi farklı yöntemler kullanılabilir. CART fonksiyonunu diğer makine öğrenme algoritmalarından ayıran kullandığı budama mekanizmasıdır. Budama mekanizması ağacın büyüklüğü ve en iyi tahmin sonucuna erişebilmek için yapılan sınıflandırmanın doğruluğu konusunda tavsiye edilen bir mekanizmadır. Burada budama kriteri için gini katsayısı gibi parametreler kullanılmaktadır [25,26].

CHAID (Chi-Squared Automatic Interaction Detector) Algoritması: Bir modelin geliştirilmesine ihtiyaç duyulduğunda farklı bir algoritma kullanmak sıklıkla tercih edilen bir yaklaşımdır. Diğer karar ağacı algoritmaları ile karşılaştırıldığında CHAID algoritması yöntem olarak daha kolay ve daha avantajlı olduğu için daha yaygın olarak kullanılmaktadır. Büyük bir örneklem hacmi ile yola çıkmasından dolayı güvenilir tahminler ortaya koyması, bağımsız değişkenlerdeki kayıp gözlemleri tahmin edebilmesi ve modelin gerçek yapısal formunda belirlenen varsayımları dikkate almadığı için ikili (binary) ve multinominal lojistik regresyon modellerine alternatif bir parametrik olmayan ağaç diyagramı olarak kullanılabilmesi CHAID analizinin yaygın olarak kullanılmasının sebeplerindendir. CHAID analizinde en kritik ilişkiler belirlenir ve bu ilişkilere göre ağaç yapısı çizilir. CHAID analizi, bağımlı değişkeni en iyi açıklayabilmek için ayrıntılı homojen alt gruplara böler. Alt gruplar, küçük açıklayıcı gruplardan oluşur. En iyi tahmin sonucunu elde etmek için başlangıç değişkenleri bağımsız olarak yeniden kategorileştirilir. İstatistiksel olarak kategorik değişkenleri analiz etmesi ve bağımsız değişkenler ile kategorik olarak ölçülebilen sonuçlar ortaya koyup bu sonuçlar arasındaki ilişkileri yorumlayabilmek adına önemli sonuçlar ortaya koyabilen CHAID algoritması ki-kare parametrik olmayan istatistik yöntemini kullanır. Bu kategorileştirme işlemi değişkenler arasındaki birleştirme işlemi yapılamayacağına istatistiksel olarak karar verilene kadar sürmektedir. Değişkenlerin bölünmeye uygun olup olmadığına, bonferroni düzeltilmiş p-değeri kullanılarak karar verilir. CHAID analizi gibi, değişkenler arasında

neden-sonuç ilişkisini oldukça detaylandıran ileri düzey istatistiksel yöntemlerin pratikte de kullanılacak olan satış tahmini gibi karar destek sistemini etkileyen önemli alanlar da kullanılmasının kararların doğruluğunu arttırmak için önemli bir çalışma olacaktır. Karar ağacı algoritmalarının bu kadar popüler olmasının sebebi karar ağacı sınıflama algoritmalarının herhangi bir alan bilgisi veya parametre ayarı gerektirmediği için bilgi keşfine çok uygun olması, yüksek boyutta veri kümesi ile çalışabildiği için modelin başarısının da buna paralel olarak yüksek olması, ağaç formunda edindikleri bilginin sezgisel ve genellikle insanlar tarafından asimilasyonun kolay olması, karar ağacının öğrenme ve sınıflandırma adımları kolay ve hızlı olması ve ortaya çıkan modelin başarısı yüksek olmasından dolayı da çıktılarının doğruluğudur [27,28,29,30].

3.1.2.Yapay Sinir Ağları

İnsan beyninin çalışma mantığı ile yola çıkarak nöronların matematiksel olarak modellenmesidir. İnsan beyninin ilişkilendirme, sınıflandırma, genelleme, öğrenme ve optimize etme fonksiyonları gibi tasarlanmaktadır. Yapay Sinir Ağları çok karmaşık ve anlaşılması güç olan yapılardan anlam çıkartıp sonuç ortaya çıkaran bir modeldir. Kurulan model ile bu yöntem kontrol edilmekte ve sürekli öğrenme faaliyeti ile model geliştirilmektedir. Bu süreç sürekli öğrenme yöntemi ile davranış biçimlerini anlamak ve hatayı en aza indirmek üzerine kuruludur. Başka bir deyişle aldığı veri kümesi ile bilgiyi anlamak ve daha sonra her uygulamadan bir ders çıkarmak gibi düşünülebilir. Yapay sinir Ağları birbirine hiyerarşik olarak bağlı ve paralel olarak çalışabilen yapay hücrelerden oluşurlar. Bu hücrelerin her biri birbirine bağlanır ve her bağlantının bir değeri olduğu varsayılır. YSA'lar biyolojik sinir ağlarından esinlenilerek geliştirilmiş, öğrenme yoluyla yeni bilgiler türetebilme ve keşfedebilme gibi yetenekleri hiçbir yardım almadan otomatik olarak gerçekleştirebilmek için tasarlanmış, bilgi işleme sistemleridir. YSA genellikle bir giriş katmanı, gizli katmanlar ve bir çıkış katmanından oluşmaktadır. Basit şekliyle her bir nöron bir önceki katmanlardaki diğer nöronlara, sinaptik ağırlıkları yoluyla bağlanmaktadır. Şekil 3.4. de yapay sinir ağları katmanları görülmektedir [18,19,31,32].



Şekil 3.4. Yapay sinir ağları katmanları

Doğru sonuç vermesi ve doğru bir sınıflandırma sağlaması bakımından avantajlı olan YSA modellerinin öğrenme sürecinin meşakkatli olması ve yorumlanmasının zor olması gibi dezavantajları bulunmaktadır. YSA'lar makina öğrenmesi gerçekleştirmekte, programları çalışma prensibi bilinen programlama tekniklerine benzer ve bilgiyi saklamayı amaçlamaktadır [33].

3.1.3.K-En Yakın Komşu

K-En yakın komşu tekniği 1950'lerin başında çalışılmaya başlanmıştır ancak mevcut bilgi işlem gücü artana kadar yaygın kullanıma geçilmemiştir. Uzayda birbirine yakın olan aynı tür veri kümeleri birbirinin komşusudur. Bu varsayım altında çok kolay ve güçlü bir algoritma olan k-en yakın komşu algoritması geliştirilmiştir. Temel amaç sınıflandırılmak istenen kümeyle en yakın kümeyi bulmaktır. Örneğin belli bir kayıtın özelliği tahmin edilecekse, veri kümesi uzayında ilgili kayıta en yakın 10 kayıtın özelliğine bakılabilir. Bütün örnekler n-boyutlu uzayda bir noktaya karşılık gelmektedir. Nesnelere arasındaki uzaklık Öklid uzaklığı olarak kabul edilir. K-en yakın komşu algoritmasında sınıflandırma işlemi en yakın komşu sayısının en çok sayıda olduğu sınıfa göre yapılmaktadır [34].

3.1.4. Destek Vektör Makineleri

Vladimir Vapnik ve Alexey Chervonenkis tarafından 1963 yılında temelleri atılan “Destek Vektör Makineleri (DVM)” istatistiksel öğrenme teorisine dayalı bir gözetimli öğrenme algoritmasıdır. Her ne kadar temelleri 60'lı yıllara dayansa da 1995 yılında Vladimir Vapnik, Bernhard Boser ve Isabelle Guyon tarafından geliştirilmiştir. Hem kategorik hem de sürekli bağımlı değişkenler için kullanılabilir. Özellikle çok sayıda bağımsız değişken olduğu durumda kullanılması tavsiye edilen bir yöntemdir. Destek Vektör Makineleri, iki sınıfa ait veri kümesinin uygun bir biçimde birbirinden ayrılabilmesi için kullanılır. Ayırma işlemini yapabilmek için karar sınırları veya hiper düzlemler belirlenir. Bunun yanı sıra doğrusal olarak ayrılamayan veya çok sınıflı sınıflama problemlerinin çözümünde de yaygın olarak tercih edilmektedir. Destek Vektör Makineleri, doğrusal olarak ayrılabilen bağımsız değişkenlerde, veri kümesi ayrılacak sonsuz sayıda doğru içerisinde marjini maksimum yapacak olan doğruyu seçmeyi hedeflemektedir. Doğrusal olarak ayrılamayan durumlarda ise, orijinal veri kümesini yüksek boyuta dönüştürmek için doğrusal olmayan haritalama yöntemi kullanılmaktadır. Veri kümesinin taşındığı bu yeni boyutta marjini en büyük ayırıcı düzlemi araştırılmaktadır. Yukarıda açıklanan amaç doğrultusunda, çekirdek fonksiyonu adı verilen fonksiyonlar sayesinde n boyutlu bir veri kümesi daha yüksek boyutlu yeni bir veri kümesine dönüştürülür ve marjini en büyük olan çoklu düzlemler bulunur. Destek Vektör Makineleri günümüzde yüz tanıma sistemlerinden, ses analizine kadar birçok sınıflandırma probleminde kullanılmaktadırlar. Destek Vektör Makineleri tekniğinin avantajları yüksek boyutlu uzaylarda etkili olması, boyut sayısının, örneklem sayısından fazla olduğu durumlarda etkili olması, karar fonksiyonunda bir takım eğitim noktalarının (support vectors) kullanılmasıdır. Dolayısıyla bellek verimli bir şekilde kullanılmış olur. Çok yönlü karar fonksiyonu için çok farklı çekirdek fonksiyonları kullanılabilir. Destek Vektör Makineleri, veri kümesinin doğrusal olarak ayrılabilme ve ayrılamama durumuna göre ikiye ayrılır [34].

3.1.5. Bayesci Ağlar

Bayes sınıflandırıcılar, temeli bayes teoremine dayanan istatistiksel sınıflandırıcılardır. Bayes teoremi adını 18 y.y. matematikçisi Thomas Bayes'den (1763) alır. Bayes teoremi koşullu olasılıkları hesaplayan basit bir matematiksel formüldür. Kanıtlar ışığında olasılıkların nasıl değişeceğini gösterir. Bayes teoremi stokastik süreç esnasında meydana gelen rastgele bir olay olan A ile diğer bir rastgele olay B için, koşullu olasılıkları ve marjinal olasılıkları arasındaki ilişki şeklinde ifade edilebilir. Bayesci analiz yöntemleri, istatistiksel veri madenciliği alanına Bayesci Ağlar ile entegre edilmiştir. Bayesci ağlar, değişkenler arasındaki ilişkilerin grafiksel olarak gösterimini temel alan bir yaklaşımdır. Bu ağlar çok geneldir ve nedensel ilişkileri ifade etmektedirler. Bayesci ağlar, karar ağaçları, yapay sinir ağları, destek vektör makinaları gibi diğer sınıflandırma yöntemlerine alternatif bir yöntemdir. Uygulamada öngörülse bir denklem üretmekten çok skorlama amacıyla kullanılan bir yöntemdir [35].

3.1.6. Regresyon Analizi

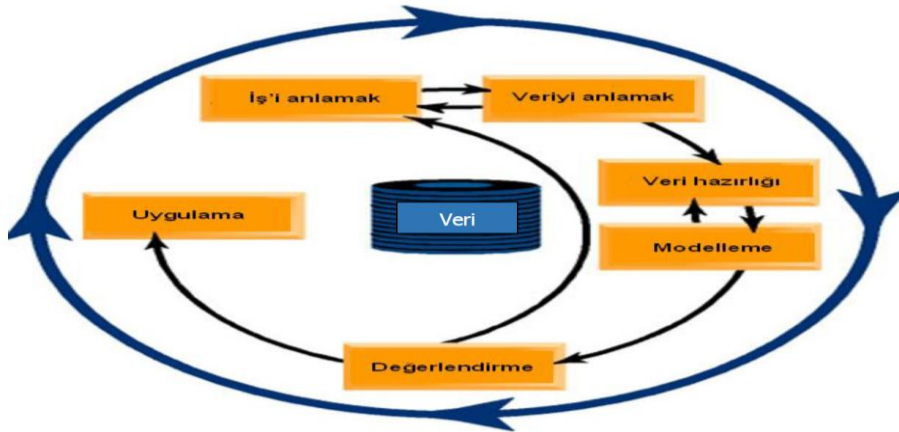
Doğrusal regresyon en çok bilinen ve kullanılan istatistiksel modelleme yöntemidir. Bağımlı değişkenin bağımsız değişkenle ilişkisini istatistiksel bir model ile açıklamak ve ilişkiler kurmak için kullanılmaktadır. Bir değişkenin değerini diğer bir değişkeni kullanarak tahmin etmeye çalışır. Regresyon analizi bağımlı değişkenin ve bağıntının tipine göre doğrusal, doğrusal olmayan, lojistik, cox regresyonu... vb. gibi isimler almaktadır. Tüm regresyon analizlerinin temel amacı kestirim yapmak ve bağımlı değişkeni etkileyen en önemli bağımsız değişkenleri belirleyebilmektir [36].

3.2. Veri Madenciliği Süreci ve CRISP-DM Döngüsü

İstatistiksel veri madenciliği birbirinden farklı disiplinleri içinde bulundurduğu, birbirinden farklı uygulamalardaki görevlerin prosedürlerinin çeşitliliği, büyük ve temizlenmemiş veri kümesinin yapısı ile çalışmanın kolay olmayacağından dolayı standart metodolojiye ihtiyaç duymaktadır. Proje öncesinde soru-cevap şeklinde çalışılması, projenin başarı ile tamamlanabilmesi için etkin bir rol oynamaktadır. Bu sorular; Nasıl bir iş problemi çözülmeye çalışılıyor? , Hangi tür veri kümesi kaynakları

mevcut ve bu iş problemi için ne çeşit veri kümelerine ihtiyaç duyulacaktır? , Analiz öncesi veri kümesi nasıl bir ön işleme ve temizleme işlemlerine ihtiyaç duymaktadır? , Analiz sonuçları nasıl değerlendirilecektir? gibi sorulardır. Ancak bu tür sorulara verilen cevaplarla planlanmış ve kabul görmüş bir metodoloji ile yola çıkılmış istatistiksel veri madenciliği projesi ile süreci başarı ile tamamlanabilir [18,37,38,39].

CRISP-DM Döngüsü : Bu tez çalışmasında istatistiksel veri madenciliğinin uluslararası düzeyde standardı olarak kabul edilmiş CRISP-DM (Cross Industry Standard Process Model for Data Mining) sürecine yer verilerek istatistiksel veri madenciliği metodolojisi aşamalarıyla incelenmiştir. “CRISP-DM (Cross Industry Standard Process Model for Data Mining) metodolojisi, Daimler Chrysler AG, SPSS NCR ve OHRA gibi lider istatistiksel veri madenciliği kullanıcıları ve tedarikçilerinden oluşan bir konsorsiyum tarafından geliştirilmiş ve dünyanın en büyük istatistiksel veri madenciliği projelerinde kullanılan bir süreçtir.” İstatistiksel veri madenciliği projelerinin hızlı, etkin ve daha ekonomik bir şekilde uygulanabilmesi için geliştirilen CRISP-DM süreci altı adımdan oluşan etkileşimli ve yinelemeli bir süreçtir. CRISP-DM süreci aşamaları Şekil 3.7. de görülmektedir [18,37,38,39].



Şekil 3.5. CRISP-DM döngüsü

Proje Amacının Belirlenmesi : Bu adım öncelikle projenin amacının ve iş hedeflerinin anlaşıldığı adımdır. İstatistiksel veri madenciliği teknikleri ile problemi tanımak, amaçlarını belirlemek ve bir proje planı tasarlanması üzerine odaklanılmalıdır. Her işte

olduđu gibi istatistiksel veri madenciliđi projesinde de en önemli ařama bařlangıç ařamasıdır. Hangi teknikler ile nasıl analiz yapılması gerektiđine bu ařamada karar verilmesi gerekmektedir aksi takdirde yanlış kurulmuř bir düzende bařlayan projenin çıktıları da sađlıklı olmayacaktır. Amaç tanımlanırken neyin ölçölmeye çalışıldıđına veya neyin tahmin edileceđine karar vermek gerekmektedir. Bu çalışmanın amacı Devlet Malzeme Ofisi satışlarını tahmin ederek geleceđe yönelik kararlar alabilmek ve bu veri kümesini karar destek sistemine entegre ederek kurum stratejik hedeflerinin gerçekteşmesini ve izlenmesini sađlamaktır.

Veri Kümesinin Deđerlendirilmesi: İleriki ařamalarda modelin yanlış kurulmaması için bu ařamada veri kümesini çok iyi tanımak analiz etmek gerekmektedir. Bu ařama veri kümesinin toplanması, veri kümesinin tanımı, veri kümesinin hangi deđişkenleri içerdiđi ve deđişkenlerin neyi ifade ettiđi ařamalarını içermektedir. Bu çalışmada veri kümesinin yapısı incelenmiř ve tanımlayıcı istatistikler bađlamında deđerlendirilerek uygulamanın Devlet Malzeme Ofisi satışlarına ait dönemseller veri kümesi ile çalışılmasına karar verilmiřtir.

Veri Kümesinin Hazırlanması: 2007-2018 geçmiř dönem satış veri kümesi kurumsal kaynak planlama (ERP) veri tabanından SQL sorgularıyla elde edilmiř ve veri kümesi satış birimleri bazında ve aylık bazda derlenmiřtir.

Modelleme: Veri tabanından çekilip hazırlanan veri kümesi IBM SPSS Modeler uygulaması ile modellenmiřtir.

Deđerlendirme: Bu ařamada modelin beklentileri karşılayıp karşılamadıđına karar verilir ve gerekiyorsa modelleme ařamasına geri dönüp en iyi sonuca ulařana kadar parametreler deđiřtirilerek model yeniden kurgulanır. Çalışmanın bu ařamasında karar ağaçları, yapay sinir ağları ve regresyon yöntemlerini uygulanarak bu yöntemler içerisinden en iyi tahmin sonucuna hangi algoritma ile ulařıldıđını tespit edilmiřtir.

Uygulama: Elde edilen satış tahminleri üst yönetime sunularak satış birimleri performans göstergesi olarak kurum resmi belgelerinde yer almıřtır. Buna ek olarak bu göstergeler iş zekâsı platformuna taşınarak günlük, haftalık ve aylık bazda raporlanan temel performans göstergesi (Key Performance Indicator) olarak takip edilecektir.

4.UYGULAMA

Bu bölümde istatistiksel veri madenciliği tekniklerinden yapay sinir ağları, karar ağaçları ve regresyon analizi yaklaşımı ile Devlet Malzeme Ofisi Genel Müdürlüğüne ait veri kümesine ilişkin satış tahmin değerleri elde edilmiştir. Duyarlılık analizi sonucunda en iyi tahminler değerlendirilirken en yüksek güvenilirlikle doğrusal model, genelleştirilmiş doğrusal model ve karar ağaçları algoritmalarından olan random trees ve xgboost trees algoritmaları ile sonuca ulaşılmıştır. Tahmin analizleri yıl, ay, il ve satış birimi bazında elde edilmiştir. Veri kümesinin yapısı veriden bir kesit alınmak suretiyle Şekil 4.1.’de görülmektedir.

	SATIS_TIPİ	İL	FATURA_TUTAR_TOPLAM	FATURA_TARIHI	SATIS_BİRİMİ	SIPARIS_SAYISI
1	Katalog	NEVŞEHİR	6209.6	2018-02-09	MERSİN İRTİBAT BÜRO MÜDÜRLÜĞÜ	1
2	Müteferrik	ANKARA	6215000.0	2018-04-16	STOK SATIŞLARI VE ÖDEMELER ŞUBE MÜDÜRLÜĞÜ	1
3	Katalog	ŞANLIURFA	18234.0	2018-06-11	DIYARBAKIR İRTİBAT BÜRO MÜDÜRLÜĞÜ	2
4	Katalog	DIYARBAKIR	100152.0	2018-06-12	DIYARBAKIR İRTİBAT BÜRO MÜDÜRLÜĞÜ	1
5	Katalog	KIRŞEHİR	37965.3	2018-02-28	STOK SATIŞLARI VE ÖDEMELER ŞUBE MÜDÜRLÜĞÜ	1
6	Müteferrik	İSTANBUL	148499.6	2019-03-04	İSTANBUL BÖLGE MÜDÜRLÜĞÜ	2
7	Katalog	BARTIN	233179.7	2018-12-12	BURSA BÖLGE MÜDÜRLÜĞÜ	3
8	Katalog	İSTANBUL	1138040.4	2018-08-03	İSTANBUL BÖLGE MÜDÜRLÜĞÜ	5
9	Katalog	AYDIN	136266.0	2018-08-15	İZMİR BÖLGE MÜDÜRLÜĞÜ	2
10	Katalog	KÜTAHYA	2521.5	2018-11-26	ESKİŞEHİR BÖLGE MÜDÜRLÜĞÜ	2
11	Katalog	ADANA	54766.2	2018-09-26	MERSİN İRTİBAT BÜRO MÜDÜRLÜĞÜ	3
12	Araclar	ANKARA	3541555.0	2018-10-19	STOK SATIŞLARI VE ÖDEMELER ŞUBE MÜDÜRLÜĞÜ	2
13	Müteferrik	ANKARA	3538216.0	2018-11-23	STOK SATIŞLARI VE ÖDEMELER ŞUBE MÜDÜRLÜĞÜ	2
14	Müteferrik	İZMİR	95445.0	2019-03-05	STOK SATIŞLARI VE ÖDEMELER ŞUBE MÜDÜRLÜĞÜ	1
15	Katalog	SAMSUN	132355.4	2018-06-20	TRABZON BÖLGE MÜDÜRLÜĞÜ	4
16	Katalog	BİTLİS	116.4	2019-04-01	VAN İRTİBAT BÜRO MÜDÜRLÜĞÜ	1
17	Katalog	İSTANBUL	3062697.9	2019-03-29	İSTANBUL BÖLGE MÜDÜRLÜĞÜ	6
18	Katalog	EDİRNE	138938.4	2019-03-29	İSTANBUL BÖLGE MÜDÜRLÜĞÜ	3
19	Katalog	DIYARBAKIR	64125.7	2018-04-12	DIYARBAKIR İRTİBAT BÜRO MÜDÜRLÜĞÜ	3
20	Katalog	ELAZIĞ	516071.8	2018-07-27	ELAZIĞ BÖLGE MÜDÜRLÜĞÜ	2
21	Katalog	BİTLİS	49024.3	2018-11-12	VAN İRTİBAT BÜRO MÜDÜRLÜĞÜ	1
22	Katalog	AMASYA	1340.0	2018-06-08	TRABZON BÖLGE MÜDÜRLÜĞÜ	1
23	Katalog	BARTIN	78861.7	2018-04-30	BURSA BÖLGE MÜDÜRLÜĞÜ	3
24	Katalog	TRABZON	116728.6	2018-07-27	TRABZON BÖLGE MÜDÜRLÜĞÜ	4
25	Katalog	UŞAK	128141.5	2018-10-09	İZMİR BÖLGE MÜDÜRLÜĞÜ	2
26	Katalog	MANİSA	47303.3	2018-10-24	İZMİR BÖLGE MÜDÜRLÜĞÜ	3
27	Katalog	BURSA	121319.2	2018-11-12	BURSA BÖLGE MÜDÜRLÜĞÜ	3
28	Müteferrik	İSTANBUL	1293094.4	2018-12-25	İSTANBUL BÖLGE MÜDÜRLÜĞÜ	3
29	Katalog	KIRKLARELİ	103840.2	2018-10-17	İSTANBUL BÖLGE MÜDÜRLÜĞÜ	2
30	Katalog	GİRESUN	45515.6	2018-06-29	TRABZON BÖLGE MÜDÜRLÜĞÜ	2

Şekil 4.1. Veri kümesinin yapısı

Hedef değişken başka bir deyişle modele çıktı olacak bağımlı değişken “Tahmin” değişkenidir. Modele girdi olacak bağımsız değişkenler yani tahmin edici değişkenler ise “ay, ağırlıklı_ortalama_1, ağırlıklı_ortalama_2, index, ağırlıklı_birikimli_ortalama_sum, tahmin_1” değişkenleridir. Kullanılan değişkenler Şekil 4.2. de görülmektedir.



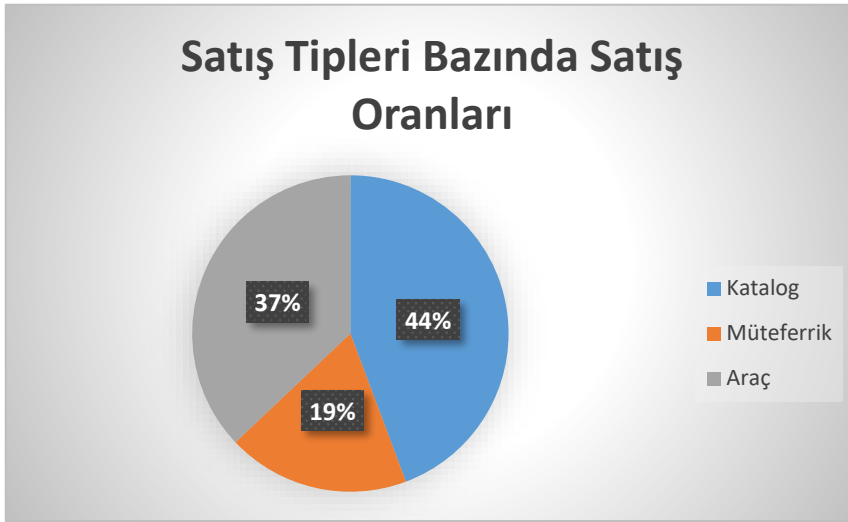
Şekil 4.2. Kullanılan değişkenler

Uygulamada veri kümesi lokal olarak değil IBM SPSS Modeler ürünü ile oracle veri tabanına bağlanıp sistem üzerinden operasyonel veri kümesi ile sağlanmıştır. Bu veri kümesinin sabit değil sürekli güncellenerek artması anlamına gelmektedir. Böylelikle uygulama sonucunda statik bir sonuç elde etmek yerine sonuçları gelen veri kümesi ile güncellenen ve yaşayan bir model üretilmiştir. Böylece Devlet Malzeme Ofisi Genel Müdürlüğü satış tahminlerini elde etmek için her zaman çalışabilen ve operasyonel veri kümesi ile beslenen dinamik bir modele ulaşılmıştır.

Devlet Malzeme Ofisi Genel Müdürlüğünde satışlar satış tipleri bazında ayrılmaktadır. Satış tipleri; Katalog, Müteferrik ve Stok olarak isimlendirilmektedir. Ancak bu çalışmada stok, etkinliği ve satış tutarındaki düşük payı nedeniyle uygulamada dikkate alınmamıştır. Araçlar başlığı altında yeni bir satış tipi kategorisi sürece eklenmiştir. Araç satış tipinin eklenmesinin nedeni bu tipin satış tutarlarındaki payın yüksek olmasıdır.

4.1. Veri Kümesinin Analizi ve Değerlendirilmesi

Ham veri kümesi incelendiğinde Oracle veri tabanından sağlanan satış tablosunda 2018 öncesi 294 bin, 2018 sonrası 134 bin satır olmak üzere yaklaşık olarak 428 bin satır ham veri kümesi ile çalışılmıştır. Bu veri kümesi zaman, il, satış birimi ve satış olarak listelenmiştir. Satış tiplerinin toplam satış tutarı içindeki dağılımına bakıldığında ise katalog satış yönteminin toplam satış tutarı içindeki payı % 44 iken, aynı oran müteferrik için % 19, araçlar için ise % 37 olduğu görülmüştür. Satış tipleri bazında satış oranları Şekil 4.3. de görülmektedir.



Şekil 4.3. Satış tipleri bazında satış oranları

Uygulamanın yapıldığı tüm akışlarda analiz için gerekli olan veri kümesi, veri tabanından alındıktan sonra, veri kümesinin temizleme, tekilleştirme, düzenleme, var olan veri kümesinden yeni değişkenler oluşturma gibi işlemler yapılmaktadır. Bunun yapılmasındaki temel amaç, satış tahminlerinin yüksek güvenilirlikte olabilmesi için, modelleme çalışmalarında girdi teşkil edecek veri kümesinin hem doğru hem de yapılacak analizlere uyumlu olmasını sağlamaktır. Veri kümesi temizleme işlemleri yapılırken, genel olarak operasyonel sistemlerden beslenen analiz veri kümesinin içerisindeki tutarsızlıklar giderilmiş, iş ihtiyaçları doğrultusundaki kısıtlara göre düzenlemeler yapılmış ve anormal kayıtlar analiz dışı bırakılmıştır. Yukarıda genel olarak özetlenen bu işlemler farklı akışlar içerisinde yer almaktadır. Uygulama bölümünde incelenen ve şekillerle gösterilen akışlarda, işlemler ve veri kümesi, ok

yönünü izlemektedir. Akış şekillerinin açıklamalarında işlemcilerin karışmaması için işlemci sırasına göre numaralandırılma yapılmıştır. Eğer ilgili şekilde o işlemciden sadece bir tane varsa numaralandırma yapılmamıştır. Bu akışlar içerisinde sıklıkla kullanılan işlemciler ve süper işlemciler açıklanmıştır. Bu işlemciler ve süper işlemciler akışın farklı noktalarında da olsa görevleri aynı olduğu için şekil detaylarında bu işlemcilerin açıklamalarına tekrar tekrar yer verilmeyecektir.

Çizelge 4.1. Yıl bazında kamu alımları istatistikleri

	Yıl	Toplam Kamu Alımları	Artış	Toplam Kamu Mal Alımları	Artış	DMO Üzerinden Yapılan Alımları	Artış	GSYİH	Artış	Enflasyon
	2007	65.978.291		21.303.536		727.378		880.460.000		8,8
	2008	83.915.296	27,19%	32.569.812	52,88%	811.999	10,42%	994.782.000	12,98%	10,44
	2009	66.224.836	-21,08%	22.962.172	-29,50%	823.380	1,38%	999.191.848	0,44%	6,25
	2010	69.510.284	4,96%	24.802.366	8,01%	1.131.552	27,23%	1.160.013.978	16,10%	8,57
	2011	91.771.406	32,03%	34.620.921	39,59%	1.700.110	33,44%	1.394.477.166	20,21%	6,47
	2012	94.398.722	2,86%	25.308.860	-26,90%	1.659.361	-2,46%	1.569.672.115	12,56%	8,89
	2013	105.504.100	11,76%	26.556.204	4,93%	1.729.417	4,05%	1.809.713.087	15,29%	7,49
	2014	113.494.916	7,57%	30.239.527	13,87%	2.245.833	22,99%	2.044.465.876	12,97%	8,85
	2015	148.434.787	30,79%	37.369.790	23,58%	3.545.816	36,66%	2.338.647.494	14,39%	7,67
	2016	173.663.835	17,00%	37.740.123	0,99%	3.738.049	5,14%	2.608.525.749	11,54%	7,78
	2017	232.890.286	34,10%	42.516.727	12,66%	3.898.004	4,10%	3.106.536.751	19,09%	11,14
	2018	202.114.838	-13,21%	51.185.075	20,39%	3.441.353	-13,27%	3.700.989.489	19,14%	16,33
Ortalama			12,18%		10,95%		11,79%		0,140653647	9,056666667
Ortanca			11,76%		12,66%		5,14%		0,143891674	8,685

Çizelge 4.1’de 2007 - 2018 yılları arasındaki veri kümesi değerlendirilirken toplam kamu alımları, toplam kamu mal alımları ve özellikle de DMO üzerinden yapılan alımların GSYİH ve enflasyon ile arasındaki korelasyon istatistiksel olarak incelenmiştir. Çizelge 4.1’deki veri kümesi için korelasyon incelemesi sonucunda Çizelge 4.2.’deki verilere ulaşılmıştır.

Çizelge 4.2. Kamu alımlarının enflasyon ile korelasyonu

KORELASYON					
		Toplam Kamu Alımları	Toplam Kamu Mal Alımları	DMO Üzerinden Yapılan Alımları	GSYİH
Toplam Kamu Mal Alımları	r	,877**			
	p	0,000			
	n	12			
DMO Üzerinden Yapılan Alımları	r	,939**	,818**		
	p	0,000	0,001		
	n	12	12		
GSYİH	r	,953**	,888**	,923**	
	p	0,000	0,000	0,000	
	n	12	12	12	
Enflasyon	r	,580*	,690*	0,377	,651*
	p	0,048	0,013	0,226	0,022
	n	12	12	12	12

** .Korelasyon 0.01 düzeyinde önemlidir (2-yönlü).

* . Korelasyon 0.05 düzeyinde önemlidir (2 yönlü).

Çizelge 4.2’ de görüldüğü üzere DMO üzerinden yapılan alımlar ile toplam kamu alımları ve toplam kamu mal alımları arasında anlamlı bir ilişki gözlenmektedir. Ayrıca GSYİH ile DMO üzerinden yapılan alımlar arasında anlamlı bir ilişki olduğu söylenebilir. Buradan da GSYİH’deki büyüme ile doğru orantılı satış hacminin artacağı yorumu yapılabilmektedir. Enflasyon ile DMO üzerinden yapılan alımlar ile arasında ilişki yoktur. Bu sonuçlara istinaden DMO satışlarının tahmini elde edilirken enflasyon dikkate alınmamıştır.

4.2.Satış Tahmin Analizleri

Çalışmanın bu bölümünde IBM SPSS Modeler ürünü ile çalışılan tahmin analizleri daha açıklayıcı olabilmesi için beş başlık altında incelenmiştir. Araç satışları çok yüksek tutarlara sahip olduğu ve tahmin sonuçlarının başarısını düşürdüğü için veri kümesinde ayrı tutularak çalışmalar sürdürülmüştür. Araç satış tahmin analizi, müteferrik satış tahmin analizi, katalog satış tahmin analizi, satış birimleri bazında satış tahmin analizi, il bazında satış tahmin analizi olmak üzere beş başlıkta açıklanmıştır. Burada satış tahmin analizi, araç satış tahmin analizi, müteferrik satış tahmin analizi, katalog satış tahmin analizi için genel akış aynı olmakla birlikte model kurulurken satış tiplerine göre modelin içinde ayrıştırılmıştır. Satış tipi için oluşturulan akış **Ek 1** ' de verilmiştir. Satış birimleri bazında satış tahmin analizi ve il bazında satış tahmin analizi için ise satış birimleri ve il ile başlayarak iki ayrı akış ile model kurulmuştur. Satış birimleri bazında satış tahmin analizi akışı **Ek 2** 'de, il bazında satış tahmin analizi akışı ise **Ek 3** ' de verilmiştir.

Analiz sırasında sık kullanılan işlemciler:

Filter: Değişken için filtreleme ve/veya değişkenin ismini değiştirmek için kullanılır.

Field Reorder: Değişkenlerin, kolonların sıralanması için kullanılır.

Sort: İçerisinde seçilen alanlara, değişkenlere göre veri kümesi satır bazlı sıralar.

Select: Belirli bir durum için kayıtların bir alt kümesini seçmek veya seçmemek için kullanılır.

Derive: Yeni değişkenler oluşturmak, hesaplamalar yapmak için tekli veya çoklu olarak kullanılır.

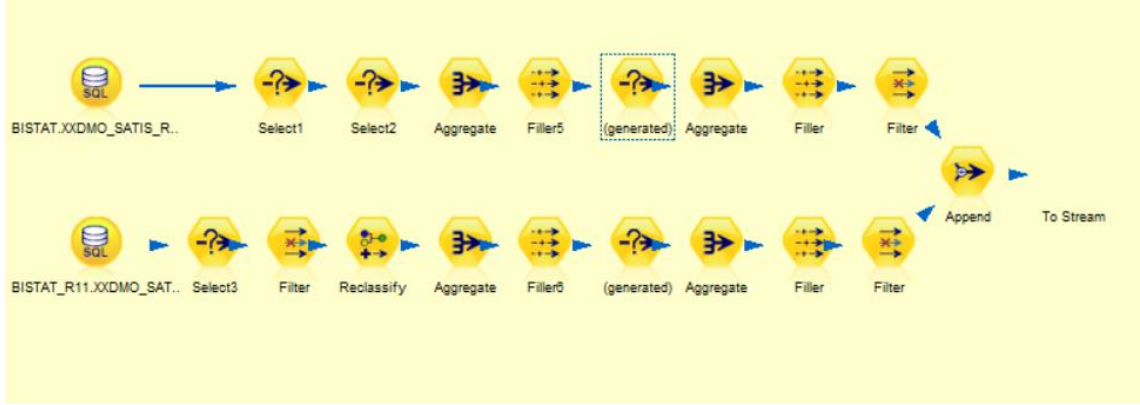
Filler: Değişkenler için tanımlanan kriterler ile doldurma veya üzerine yazma işlemleri için kullanılır.

Append: Veri kümesi satır bazlı birleştirmek için kullanılır.

Merge: Veri kümesi kolon bazlı birleştirmek için kullanılır.

Aggregate: Belirlenen bir ya da birden fazla değişken için hesaplamalar yapmak için kullanılır.

Analiz sırasında sık kullanılan süper işlemci kümesi araç, müteferrik, katalog, satış birimleri ve il bazında yapılan satış tahmin modeline eklenmiş ve satış yöntemine ilişkin etkenler üzerinden yıl, ay, satış birimi ve il bazında satışların durumu özetlenmiş ve Şekil 4.4.'de gösterilmiştir.



Şekil 4.4. Analizde kullanılan süper işlemci

Tanımlayıcı İstatistikler :

Aggregate 1 : Satış tipi, yıl ve sipariş numarasına göre toplam satış tutarı hesaplatılmıştır.

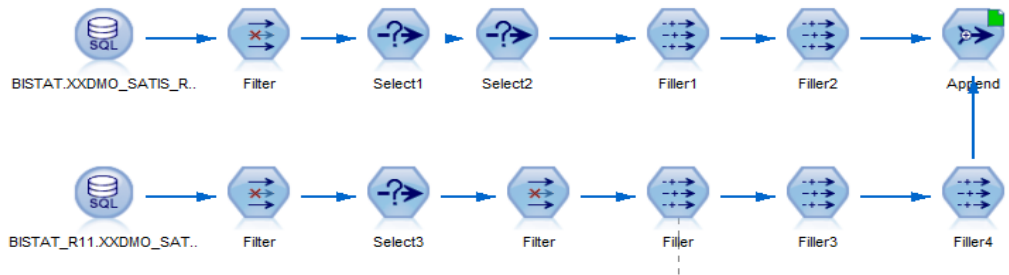
Aggregate 2: Satış tipi ve yıl değişkenlerine göre satış tutarının toplamı, ortalaması, medyanı, maksimum değeri ve record count ile verilen sipariş sayısı hesaplatılarak Sipariş adedi kolonuna yazdırılmış bu şekilde araç satışlarının davranışları gözlenmiştir.

Aggregate 3: Satış tipi, yıl, sipariş kodu ve müşteri numarasına göre satış tutarının toplamı hesaplatılmıştır.

Aggregate 4: Satış tip, yıl ve müşteri numarasına göre satış tutarının toplamı, ortalaması, medyanı, maksimum değeri ve record count ile verilen toplam sipariş sayısı hesaplatılarak Sipariş adedi kolonuna yazdırılmış bu şekilde araç satış tipinin durumu gösterilmiştir.

4.2.1.Araç Satış Tahmin Analizi

Analiz için kullandığımız veri kümesi 2007-2018 yılları arasındaki veri kümesini içermektedir. Devlet Malzeme Ofisi 2007 den itibaren kullandığı oracle kurumsal kaynak planlama programını 2018 yılında güncellediği için analizin bu kısmında 2018 öncesi ve sonrası veri kümesinin eşleştirilebilmesi için 2 ayrı veri tabanından alınan veri kümesinin birleştirme işlemi yapılmıştır. Veri kümesi birleştirme işlemi Şekil 4.5.'de gösterilmiştir.



Şekil 4.5. Araç satış tahmin analizi kısım 1

Database.1: Bistat.xxdmo_satis_raporu1 tablosu okutulmuştur. Tablo 2018 sonrası kayıtları içermektedir.

Database.2: Bistat_r11. xxdmo_satis_raporu1 tablosu okutulmuştur. Tablo 2018 öncesi kayıtları içermektedir.

Select.1: 2018 öncesi kayıtlar için satış sipariş statüsü; “kapalı” ve “alacaklar muhasebesinde arayüz oluşturuldu” olan kayıtlar seçildi.

Select.2: 2018 öncesi kayıtlar için bütçe türü; “Tedarikçi”, “Yurtiçi Gerçek Şahıslar” ve “Yurtiçi Tüzel Şahıslar” durumunu içermeyen kayıtlar seçildi.

Select.3: 2018 sonrasındaki kayıtlar için sipariş tarihi 2009’dan büyük ve büyük eşit koşuluna uygun olan kayıtlar seçildi.

Filler.1: 2018 öncesi kayıtlar için seçim alanı field içerisinde seçilen tüm değişkenler tarih formatına dönüştürüldü.

Filler.2: 2018 öncesi kayıtlar için sipariş_no değişkeni tamsayı formatına dönüştürüldü.

Filler: Kalem kodu değişkenindeki kayıtlar düzenlendi.

Filler.3: 2018 sonrası kayıtlar için seçim alanı field içerisinde seçilen tüm değişkenler tarih formatına dönüştürüldü.

Filler.4: 2018 sonrası kayıtlar için sipariş_no değişkeni tamsayı formatına dönüştürüldü.

Append: Gerekli düzenlemeler yapıldıktan sonra eski ve yeni kayıtlar satır bazlı olarak birleştirildi.



Şekil 4.6. Araç satış tahmin analizi akış kısım 2

Araç satış tahmin analizi akış kısım 2, Şekil 4.6.'da gösterilmiş ve adımlar aşağıda açıklanmıştır.

Reclassify: İl değişkeni için yeniden sınıflandırma işlemi yapıldı.

Yıl (Derive): Sipariş tarihi içerisindeki tarihten yıl kısmı çekilmiştir.

Filler4: Veri kümesi içerisindeki 2021 şeklinde yanlış yazılan kayıt 2012 olarak düzeltilmiştir.

Ay (Derive): Sipariş tarihi içerisindeki tarihten ay kısmı çekilmiştir.



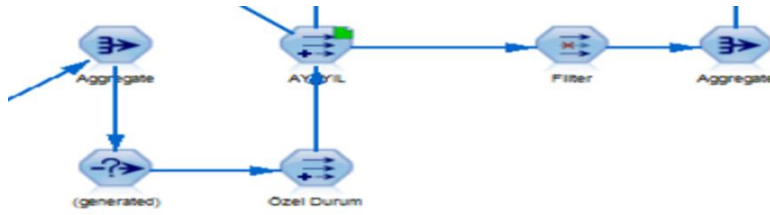
Şekil 4.7. Araç satış tahmin analizi akış kısım 3

Araç satış tahmin analizi akış kısım 3, Şekil 4.7.'de gösterilmiş ve adımlar aşağıda açıklanmıştır.

Filler.5: Satış tipi değişkenindeki istenilen kalem tipleri iş bilgisi dahilinde araçlar olarak tanımlandı.

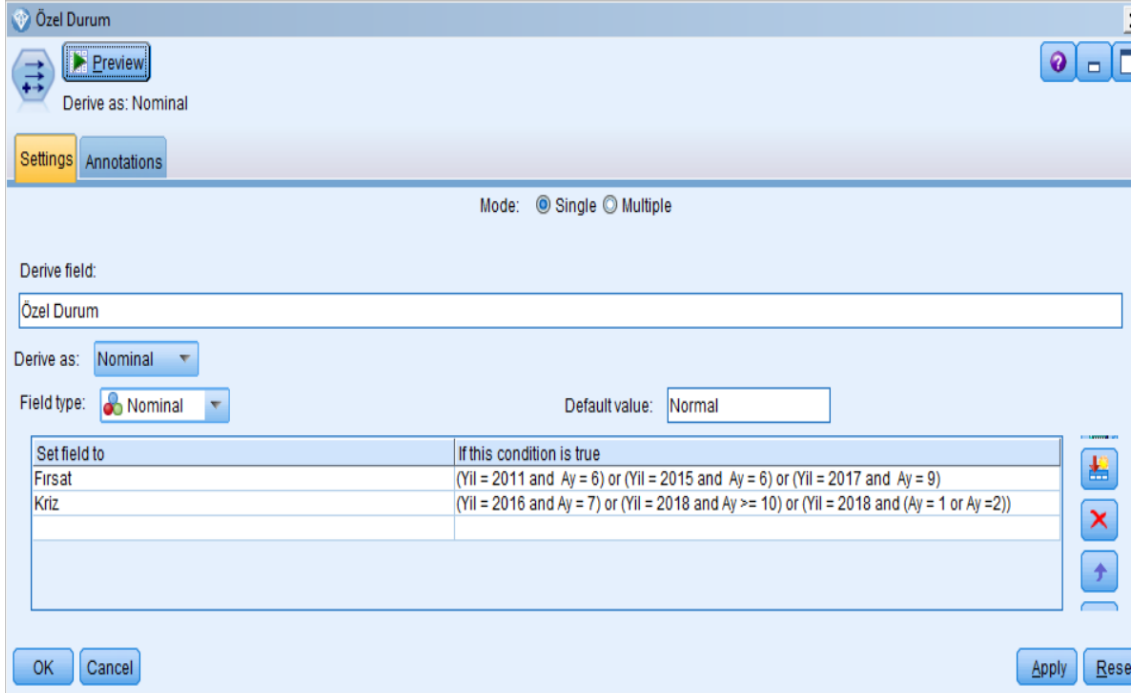
Filler.6: Yılı 2017'den küçük olanlar ile satış tipi nakil vasıta olanlar araçlar olarak tanımlandı.

Select.4: Satış tipi araçlar olanlar ile yılı 2016 ve üstünde olanlar seçildi.



Şekil 4.8. Araç satış tahmin analizi akış kısım 4

Şekil 4.8' de özel durum değişkeni gösterilmiştir. Burada satışları dalgalandırmasından dolayı tahmin sonuçlarını etkileyen; iş deneyimi ile belirlenen özel durumlar Şekil 4.9.'da gösterildiği gibi modele eklenmiştir. Belli yıl ve ay bazında 'Fırsat' , 'Kriz' ve 'Normal' olarak değişkenler tanımlanmıştır. Bu sayede modele ilgili zamanlardaki dalgalanmalardan etkilenmeden tahmin yapılması gerektiği öğretilmiş ve daha doğru tahmin sonucuna ulaşılmış olur.



Şekil 4.9. Araç satış tahmini özel durum tanımlama ekranı

4.2.2. Müteferrik Satış Tahmin Analizi

Akışın bu kısmında 2018 öncesi ve sonrası veri kümesi için düzenleme işlemleri yapıлып veri kümelerinin birleştirilmesi amaçlanmıştır.



Şekil 4.10. Müteferrik satış tahmin analizi kısım 1

Müteferrik satış tahmin analizi kısım 1, Şekil 4.10.'da gösterilmiş ve adımlar aşağıda açıklanmıştır.

Database.1: Bistat.xxdmo_satis_raporu1 tablosu okutulmuştur. 2018 öncesi kayıtları içermektedir.

Database.2: Bistat_r11.xxdmo_satis_raporu1 tablosu okutulmuştur. 2018 sonrası kayıtları içermektedir.

Select.1: 2018 öncesi kayıtlar için satış sipariş statüsü; “kapalı” ve “alacaklar muhasebesinde arayüz oluşturuldu” olan kayıtlar seçildi.

Select.2: 2018 öncesi kayıtlar için bütçe türü; “Tedarikçi”, “Yurtiçi Gerçek Şahıslar” ve “Yurtiçi Tüzel Şahıslar” durumunu içermeyen müşteriler seçildi.

Select.3: 2018 sonrasındaki kayıtlar için sipariş tarihi 2009’dan büyük ve büyük eşit koşuluna uygun olan müşteriler seçildi.

Filler.1: 2018 öncesi kayıtlar için seçim alanı field içerisinde seçilen tüm değişkenler tarih formatına dönüştürüldü.

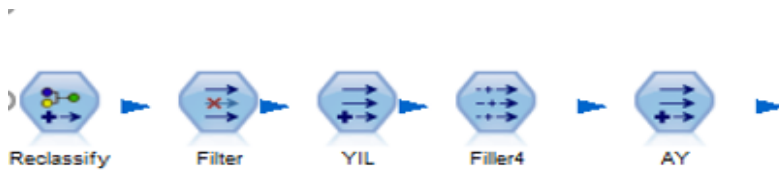
Filler.2: 2018 öncesi kayıtlar için siparis_no değişkeni tamsayı formatına dönüştürüldü.

Filler: Kalem kodu değişkenindeki kayıtlar düzenlendi.

Filler.3: 2018 sonrası kayıtlar için seçim alanı field içerisinde seçilen tüm değişkenler tarih formatına dönüştürüldü.

Filler.4: 2018 sonrası kayıtlar için siparis_no değişkeni tamsayı formatına dönüştürüldü.

Append: Gerekli düzenlemeler yapıldıktan sonra eski ve yeni kayıtlar satır bazlı olarak birleştirildi.



Şekil 4.11. Müteferrik satış tahmin analizi kısım 2

Müteferrik satış tahmin analizi kısım 2, Şekil 4.11.’de gösterilmiş ve adımlar aşağıda açıklanmıştır.

Reclassify: İl değişkeni için yeniden sınıflandırma işlemi yapıldı.

Yıl (Derive): Sipariş tarihi içerisindeki tarihten yıl kısmı çekilmiştir.

Filler4: Veri kümesi içerisindeki 2021 şeklinde yanlış yazılan kayıt 2012 olarak düzeltilmiştir.

Ay (Derive): Sipariş tarihi içerisindeki tarihten ay kısmı çekilmiştir.



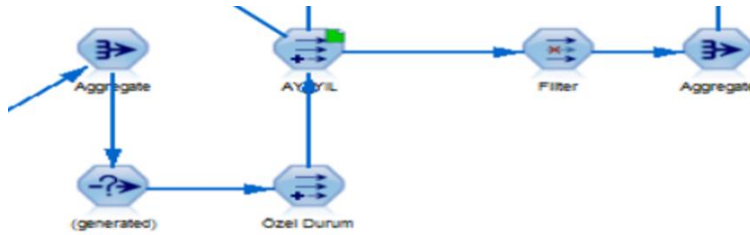
Şekil 4.12. Müteferrik satış tahmin analizi kısım 3

Müteferrik satış tahmin analizi kısım 3, Şekil 4.12.'de gösterilmiş ve adımlar aşağıda açıklanmıştır.

Filler.5: Satış tipi değişkenindeki istenilen kalem tipleri iş bilgisi dahilinde araçlar olarak tanımlandı.

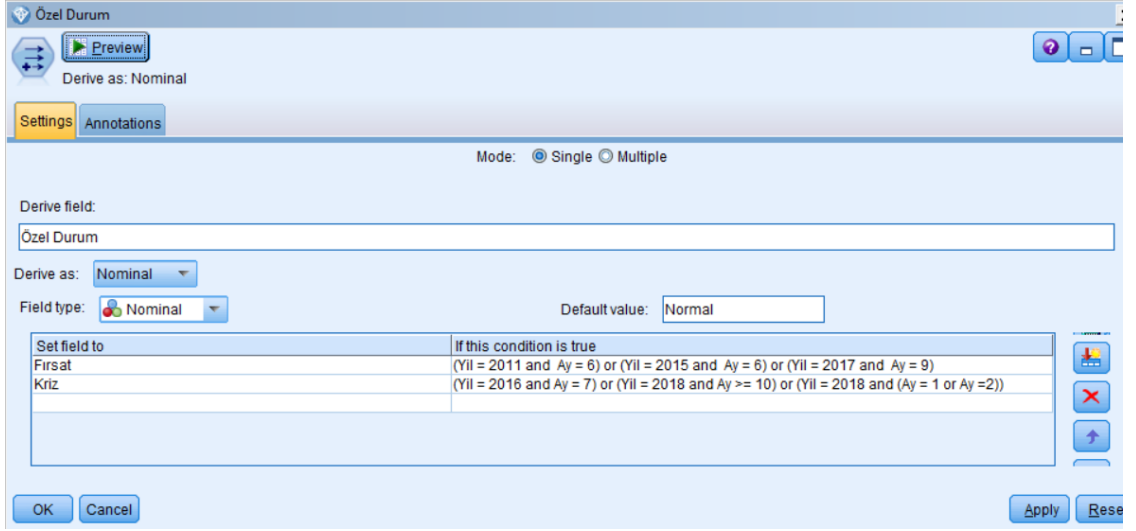
Filler.6: Yılı 2017'den küçük olanlar ile satış tipi nakil vasıta olanlar araçlar olarak tanımlandı.

Select.4: Satış tipi müteferrik olanlar ile yılı 2016 ve üstünde olan siparişler seçildi.



Şekil 4.13. Müteferrik satış tahmin analizi kısım 4

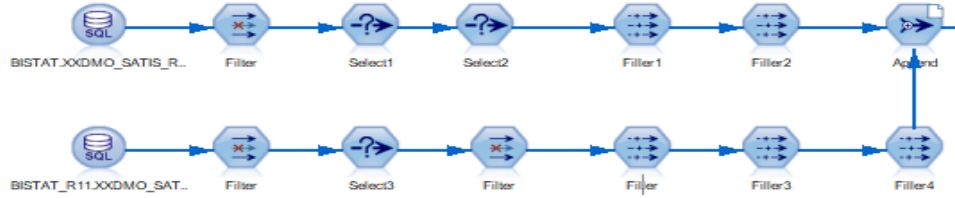
Şekil 4.13'de özel durum değişkeni gösterilmiştir. Burada satışları dalgalandırmasından dolayı tahmin sonuçlarını etkileyen; iş deneyimi ile belirlenen özel durumlar Şekil 4.14.'te gösterildiği gibi modele eklenmiştir. Belli yıl ve ay bazında 'Fırsat' , 'Kriz' ve 'Normal ' olarak değişkenler tanımlanmıştır. Bu sayede modele ilgili zamanlardaki dalgalanmalardan etkilenmeden tahmin yapılması gerektiği öğretilmiş ve daha doğru tahmin sonucuna ulaşılmış olur.



Şekil 4.14. Müteferrik satış tahmini özel durum tanımlama ekranı

4.2.3.Katalog Satış Tahmin Analizi

Akışın bu kısmında 2018 öncesi ve sonrası katalog satış yöntemi veri kümesi için düzenleme işlemleri yapıлып veri kümelerinin birleştirilmesi amaçlanmıştır.



Şekil 4.15. Katalog satış tahmin analizi kısım 1

Katalog satış tahmin analizi kısım 1, Şekil 4.15.'de gösterilmiş ve adımlar aşağıda açıklanmıştır.

Database.1: Bistat.xxdmo_satis_raporu1 tablosu okutulmuştur. 2018 öncesi kayıtları içermektedir.

Database.2: Bistat_r11.xxdmo_satis_raporu1 tablosu okutulmuştur. 2018 sonrası kayıtları içermektedir.

Select.1: 2018 öncesi kayıtlar için satış sipariş statüsü; “kapalı” ve “alacaklar muhasebesinde arayüz oluşturuldu” olan müşteriler seçildi.

Select.2: 2018 öncesi kayıtlar için bütçe türü; “tedarikçi”, “yurtiçi gerçek şahıslar” ve “yurtiçi tüzel şahıslar” durumunu içermeyen müşteriler seçildi.

Select.3: 2018 sonrasındaki kayıtlar için sipariş tarihi 2009’dan büyük ve büyük eşit koşuluna uygun olan kayıtlar seçildi.

Filler.1: 2018 öncesi kayıtlar için seçim alanı field içerisinde seçilen tüm değişkenler tarih formatına dönüştürüldü.

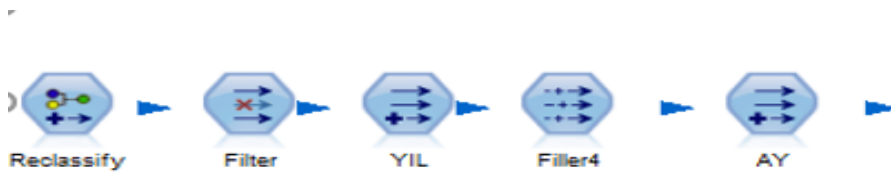
Filler.2: 2018 öncesi kayıtlar için siparis_no değişkeni tamsayı formatına dönüştürüldü.

Filler: Kalem kodu değişkeninin veri kümesinin yapısı düzenlendi.

Filler.3: 2018 sonrası kayıtlar için seçim alanı field içerisinde seçilen tüm değişkenler tarih formatına dönüştürüldü.

Filler.4: 2018 sonrası kayıtlar için siparis_no değişkeni integer (tamsayı) formatına dönüştürüldü.

Append: Gerekli düzenlemeler yapıldıktan sonra eski ve yeni kayıtlar satır bazlı olarak birleştirildi.



Şekil 4.16. Katalog satış tahmin analizi kısım 2

Katalog satış tahmin analizi kısım 2, Şekil 4.16.’da gösterilmiş ve adımlar aşağıda açıklanmıştır.

Reclassify: İl değişkeni için yeniden sınıflandırma işlemi yapıldı.

Yıl (Derive): Sipariş tarihi içerisindeki tarihten yıl kısmı çekilmiştir.

Filler4: Veri kümesi içerisindeki 2021 şeklinde yanlış yazılan kayıt 2012 olarak düzeltilmiştir.

Ay (Derive): Sipariş tarihi içerisindeki tarihten ay kısmı çekilmiştir.



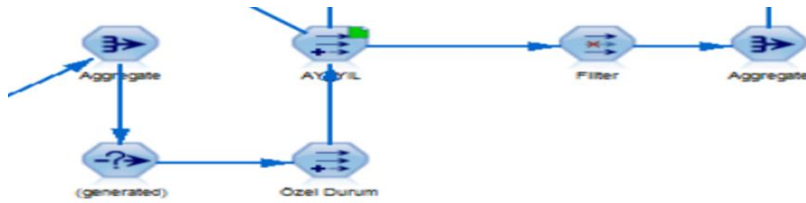
Şekil 4.17. Katalog satış tahmin analizi kısım 3

Katalog satış tahmin analizi kısım 3, Şekil 4.17.'de gösterilmiş ve adımlar aşağıda açıklanmıştır.

Filler.5: Satış tipi değişkenindeki istenilen kalem tipleri iş bilgisi dahilinde araçlar olarak tanımlandı.

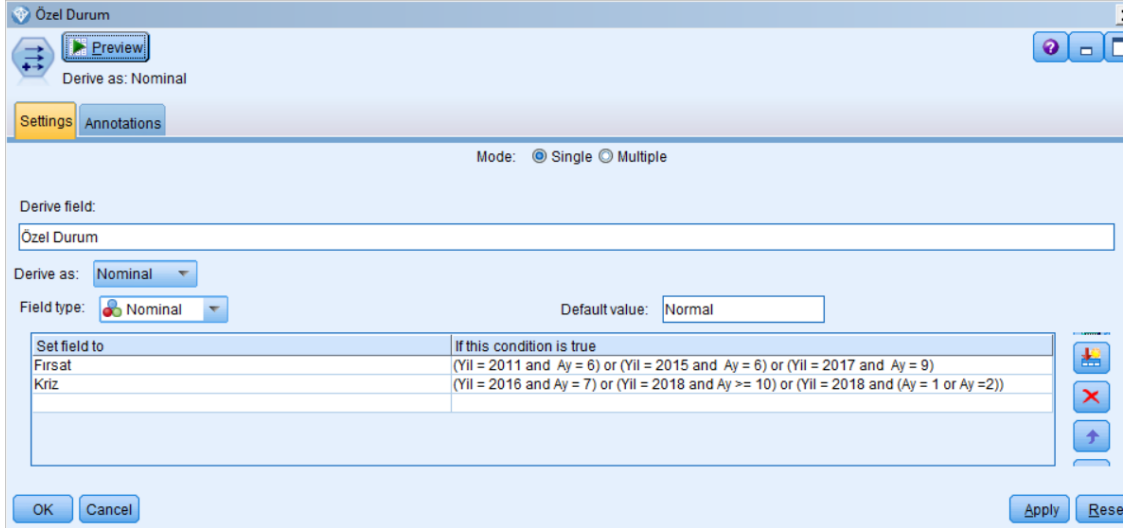
Filler.6: Yılı 2017'den küçük olanlar ile satış tipi nakil vasıta olanlar araçlar olarak tanımlandı.

Select.4: 2016 yılı ve üstünde satış yöntemi katalog olan siparişler seçildi.



Şekil 4.18. Katalog satış tahmin analizi kısım 4

Şekil 4.18'de özel durum değişkeni gösterilmiştir. Burada satışları dalgalandırmasından dolayı tahmin sonuçlarını etkileyen; iş deneyimi ile belirlenen özel durumlar Şekil 4.19.'da gösterildiği gibi modele tanımlanmıştır. Belli yıl ve ay bazında 'Fırsat' , 'Kriz' ve 'Normal' olarak değişkenler eklenmiştir. Bu sayede modele ilgili zamanlardaki dalgalanmalardan etkilenmeden tahmin yapılması gerektiği öğretilmiş ve daha doğru tahmin sonucuna ulaşılmış olur.



Şekil 4.19. Katalog satış tahmini özel durum tanımlama ekranı

4.2.4. Satış Birimi Bazında Satış Tahmin Analizi

Satış birimi bazında yapılan satış tahmin analizi genel akış görüntüsüne **Ek 2**'de yer verilmiştir. Daha sonraki adımlarda genel akış kısımlara bölünerek uygulama anlatılmaya çalışılmıştır.



Şekil 4.20. Satış birimi bazında satış tahmin analizi kısım 1

Satış birimi bazında satış tahmin analizi kısım 1, Şekil 4.20.'de gösterilmiş ve adımlar aşağıda açıklanmıştır.

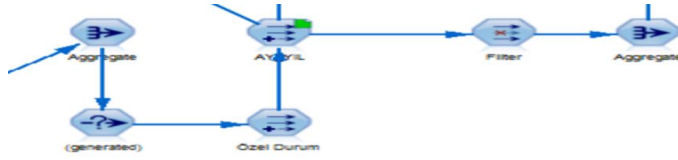
Sort : Satış birimleri satış tutarına göre sıralanmıştır.

SB_LISTE : Sıralanan satış birimleri listesi yazdırılmıştır.

SB_SAYI: Satış birimleri sayısı yazdırılmıştır.

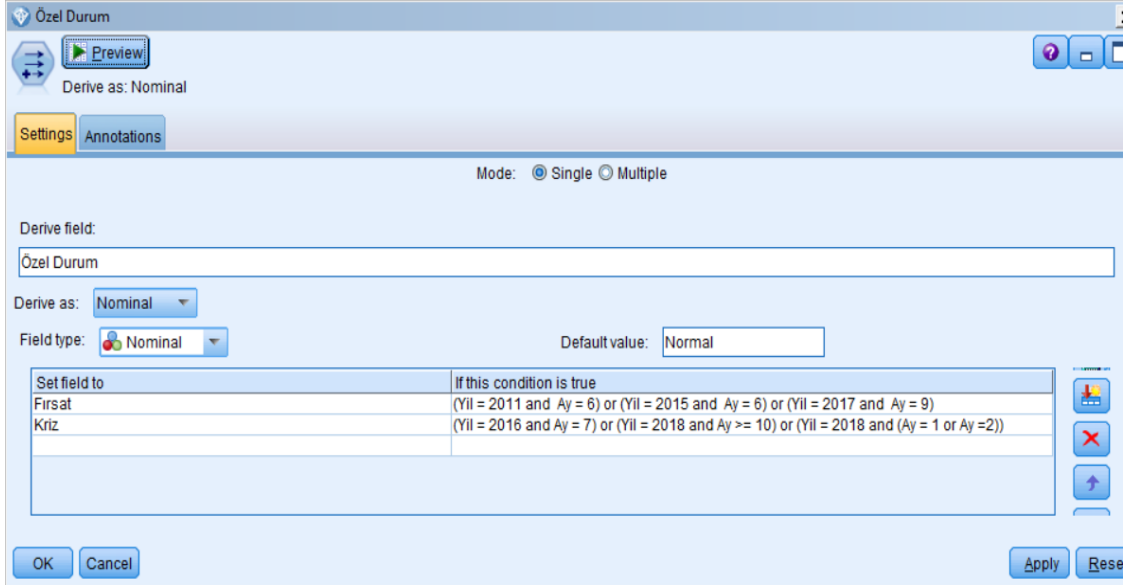
Aggregate 1 : Satış tipine, sipariş koduna, yılına ve dönemine göre satış tutarının toplamı yazdırılmıştır.

Aggregate 2 : Toplam satış tutarına göre satış birimleri yazdırılmıştır.



Şekil 4.21. Satış birimi bazında satış tahmin analizi kısım 2

Şekil 4.21’de özel durum değişkeni gösterilmiştir. Burada satışları dalgalandırmasından dolayı tahmin sonuçlarını etkileyen; iş deneyimi ile belirlenen özel durumlar Şekil 4.22.’de gösterildiği gibi modele tanımlanmıştır. Belli yıl ve ay bazında ‘Fırsat’ , ‘Kriz’ ve ‘Normal’ olarak değişkenler eklenmiştir. Bu sayede modele ilgili zamanlardaki dalgalanmalardan etkilenmeden tahmin yapılması gerektiği öğretilmiş ve daha doğru tahmin sonucuna ulaşılmış olur.



Şekil 4.22. Satış birimi bazında satış tahmin analizi özel durum tanımlama ekranı

4.2.5. İl Bazında Satış Tahmin Analizi

İl bazında yapılan satış tahmin analizi genel akış görüntüsüne Ek 3’de yer verilmiştir. Daha sonraki adımlarda genel akış kısımlara bölünerek uygulama anlatılmaya çalışılmıştır.



Şekil 4.23. İl bazında satış tahmin analizi kısım 1

İl bazında satış tahmin analizi kısım 1, Şekil 4.23.’de gösterilmiş ve adımlar aşağıda açıklanmıştır.

Sort : İller satış tutarına göre sıralanmıştır.

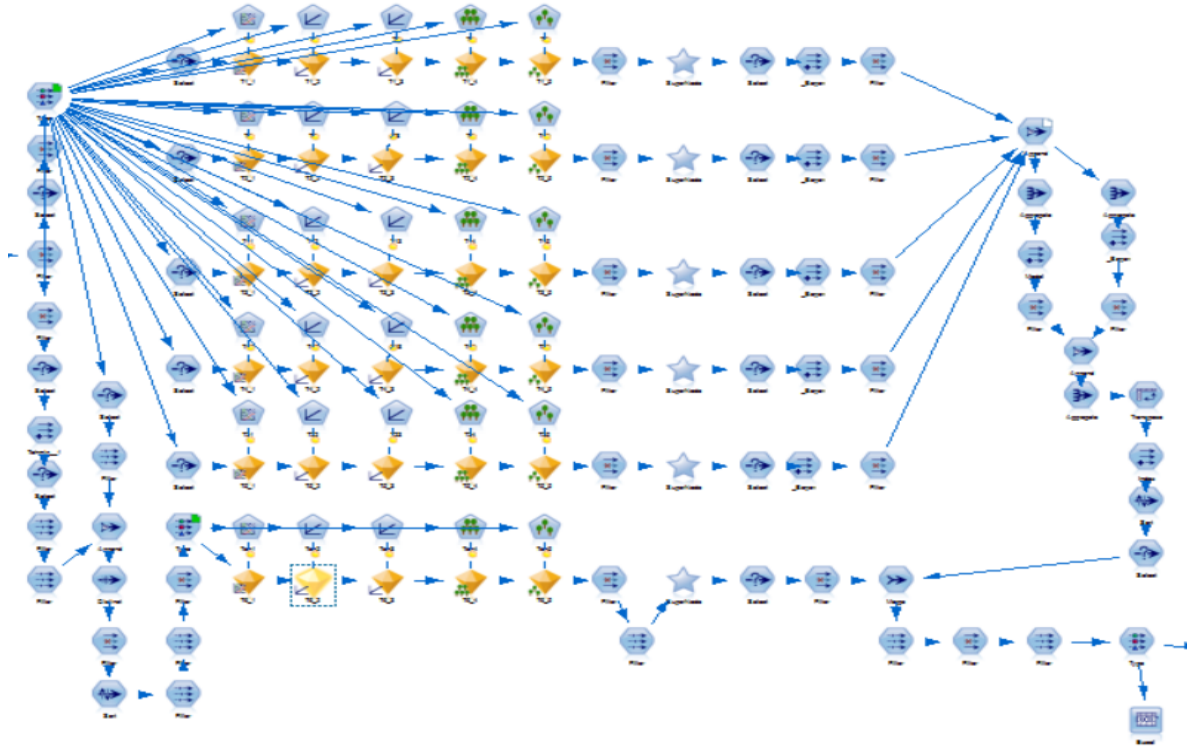
IL_LISTE : Sıralanan iller listesi yazdırılmıştır.

IL_SAYI: İllerin sayısı yazdırılmıştır.

Aggregate 1 : Satış tipine, sipariş koduna, yılına ve dönemine göre satış tutarının toplamı yazdırılmıştır.

Aggregate 2 : Toplam satış tutarına göre iller yazdırılmıştır.

Bu kısma kadar araç, katalog, müteferrik, satış birimi bazında ve il bazında satış tahmin analizi uygulaması için IBM SPSS Modeler ürünü ile veri kümesinin nasıl düzenlediği ayrıntılı olarak gösterilmiştir. Bu bölümde artık istatistiksel veri madenciliği tekniklerinden hangilerini kullandığımız ve ulaştığımız tahmin sonuçlarına yer verilecektir. Aynı zamanda tahmin-gerçekleşme karşılaştırması yapılarak da modelin başarısı test edilecektir. Burada satış tahmin analizi, araç satış tahmin analizi, müteferrik satış tahmin analizi, katalog satış tahmin analizi için genel akış aynı olmakla birlikte model kurulurken satış tiplerine göre modelin içinde ayrıştırılmıştır. Satış tipi için oluşturulan akış **Ek 1** ' de verilmiştir. Satış birimleri bazında satış tahmin analizi ve il bazında satış tahmin analizi için ise satış birimleri ve il ile başlayarak 2 ayrı akış ile model kurulmuştur. Satış birimleri bazında satış tahmin analizi akışı **Ek 2** 'de il bazında satış tahmin analizi akışı ise **Ek 3** ' de verilmiştir.



Şekil 4.24. Genel akış model aşaması

Bu çalışmada tahmin edici istatistiksel veri madenciliği tekniklerinden sınıflama ve regresyon teknikleri incelenmiştir. Yapay Sinir Ağları, Karar Ağaçları ve Regresyon analizi yaklaşımı ile Devlet Malzeme Ofisi Genel Müdürlüğüne ait veri kümesine ilişkin satış tahmin değerleri elde edilmiştir. Duyarlılık analizi sonucunda en iyi tahmin sonuçları değerlendirilirken en yüksek güvenilirlikle doğrusal model, geliştirilmiş doğrusal model ve karar ağaçları algoritmalarından olan random trees ve XGBoost trees algoritmaları ile sonuca ulaşılmıştır. En iyi tahmin sonucuna hangi algoritmayla gidileceğine karar verirken model kendi geçerliliğini sağlamak için geçmişte gerçekleşen satışları kapatılarak tahmin yapıyor ve modelin başarısına bakılıyor. Bu işlemin hangi sıklıkta yapılacağına yani aylık, çeyreklik ya da yıllık bazda yapılacağına risk algısına göre karar verilmektedir. Bu uygulamada Şekil 4.24’ de görüldüğü üzere test kısmını sağlayabilmek için her bir algoritma ile öncelikle beş kez tahmin yapılmıştır. Bu beş tahmin sırasıyla mevcut veri kümesinde rastgele beş ay, dört ay, üç ay, iki ay ve bir ay gerçekleşme değeri kapatılarak tahmin edilip sonuçlar doğrultusunda model başarısı elde edilmiştir. Geçmiş veri kümesini kapatılarak yapılan tahminlerdeki teknikler olan, doğrusal model, random trees, XGBoost trees ve geliştirilmiş doğrusal model ile elde edilen

sonuların ortalamaları ile meta model elde edilmiř ve en iyi tahmin sonucuna ulařılmıřtır. Bu iřlem modele bir sper iřlemci eklenerek yapılmıřtır. En iyi tahmin sonularını veren modellerin ıktıları **Ek 3**, **Ek 4**, **Ek 5**, **Ek 6** ve **Ek 7** de verilmiřtir.

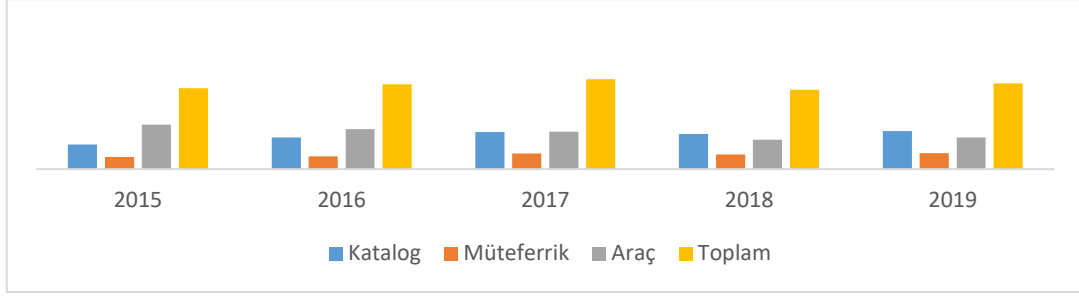
5.SONUÇ VE TARTIŞMA

Uygulama sonucunda IBM SPSS Modeller ürünü ile çeşitli istatistiksel veri madenciliği teknikleri kullanılarak elde edilen modellerin başarısı önceki bölümde istatistiksel olarak değerlendirilmiş ve en başarılı modeller ile bir çıktı elde edilmişti. Bu bölümde ise elde edilen tahmin analiz sonuçları gerçekleşen verilerle karşılaştırılacaktır. Yıllık gerçekleşen satışlar ve tahmin analizi sonucunda elde edilen 2019 yılı satış tahminleri Çizelge 5.1’deki gibidir.

Çizelge 5.1. Dört yıllık gerçekleşme ve 2019 yılı tahmin sonuçları

YIL	Katalog	Müteferrik	Araç	Toplam
2015	1.048.173.055	511.520.022	1.894.225.438	3.453.918.515
2016	1.355.852.142	548.099.729	1.710.143.617	3.614.095.488
2017	1.576.662.504	661.352.117	1.594.038.354	3.832.052.975
2018	1.500.861.534	632.046.890	1.256.157.794	3.389.066.218
2019	1.621.389.457	682.803.937	1.357.034.581	3.661.227.975

Çizelge 5.1’de görüldüğü gibi son dört yıllık değişimler incelendiğinde üç yıl boyunca artış gösteren satışlarda 2018 yılında düşüş olduğu görülmüştür. 2019 yılı tahmin sonuçları incelendiğinde 2018 yılına göre bir miktar artış olacağı ancak bu artışla birlikte bile 2019 yılı satış tutarının 2017 yılı satış tutarının altında kalacağı tahmin edilmektedir.



Şekil 5.1. Dört yıllık gerçekleştirme ve 2019 yılı tahmin sonuçları

Şekil 5.1' de ise Çizelge 5.1.'de verilen verilerin grafiği yapılarak satış tutarlarındaki dalgalanmaların yıllara göre karşılaştırılabilmesi amaçlanmıştır. Dört yıllık gerçekleştirme sonuçları ile 2019 yılı tahmin sonuçlarının birbirlerine yakın seyrettiği görülmektedir.

Çizelge 5.2. Araç – Katalog - Müteferrik İçin İlk Sekiz Ay Tahmin Sonuçları, Gerçekleşmeler ve Gerçekleşme Yüzdeleri

AY	ARAÇ			KATALOG			MÜTEFERRİK		
	TAHMİN	GERÇEKLEŞME	GERÇEKLEŞME ORANI	TAHMİN	GERÇEKLEŞME	GERÇEKLEŞME ORANI	TAHMİN	GERÇEKLEŞME	GERÇEKLEŞME ORANI
1	19.147.348	19.147.348	100%	10.304.519	10.304.519	100%	8.146.808	8.146.808	100%
2	46.905.929	47.852.136	102%	42.707.032	45.632.127	107%	24.233.525	24.233.525	100%
3	45.468.712	44.685.869	98%	110.167.890	77.860.449	71%	20.356.276	14.524.797	71%
4	45.468.712	89.445.145	197%	105.873.078	96.163.964	91%	24.240.530	37.114.512	153%
5	77.077.888	86.544.518	112%	129.346.583	122.001.574	94%	27.697.113	47.997.279	173%
6	102.009.616	86.014.608	84%	136.067.661	136.088.812	100%	32.814.158	25.276.349	77%
7	111.783.904	65.158.717	58%	128.928.553	83.172.513	65%	25.403.655	65.840.847	259%
8	68.642.216	44.572.618	65%	126.964.577	75.693.340	60%	27.440.279	27.065.725	99%
9	103.270.544			169.797.337			53.624.475		
10	96.702.200			137.776.435			47.293.848		
11	101.988.080			162.724.315			44.064.513		
12	322.461.568			318.686.020			239.583.449		

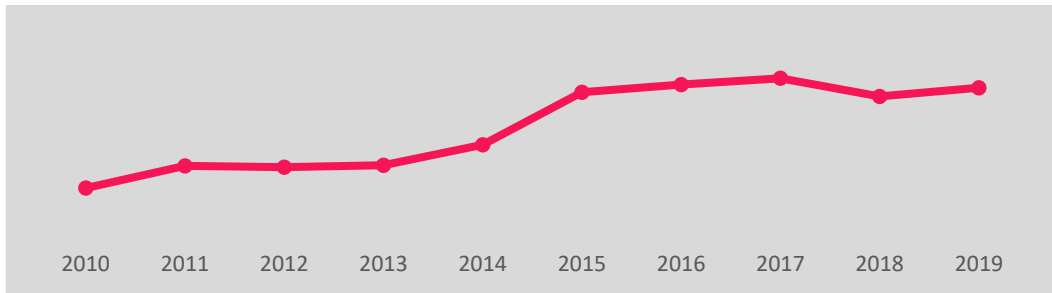
Tahmin sonuçları ve gerçekleştirmeler Çizelge 5.2.'de gösterildiği gibi sürekli olarak izlenmekte ve gerçekleştirme oranı düştüğü noktalarda düşüş sebepleri araştırılarak model geliştirilmektedir. İş hafızası ile modele ne kadar çok özel durum eklenirse modelin başarısı o kadar artacaktır. Daha öncede bahsedildiği gibi veri kümesinin canlı veri

tabanından sağlanması ve güncel veri kümesi ile çalışılabilme imkânı sağlanması sebebiyle tahmin sonuçları anlık olarak yenilenerek alınabilmektedir.

Çizelge 5.3. Dokuz yıllık satış birimi bazında gerçekleşme ve 2019 yılı tahmin sonuçları

SATIŞ BİRİMİ	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Değişim
Pazarlama Daire Başkanlığı	610.077.520	922.019.487	935.201.148	913.866.121	1.264.776.225	2.008.668.853	2.031.793.376	2.010.635.745	1.836.909.628	1.919.913.621	5%
İstanbul Bölge Müdürlüğü	123.086.485	184.854.296	181.881.301	190.705.573	229.257.187	438.572.214	384.210.831	423.593.017	325.892.762	372.142.536	14%
İzmir Bölge Müdürlüğü	92.925.594	142.842.585	130.764.224	106.484.351	117.004.427	278.613.330	289.765.688	348.637.909	257.913.664	281.483.253	9%
Bursa Bölge Müdürlüğü	45.038.785	65.405.055	60.060.084	84.843.740	116.551.008	134.824.062	217.875.730	249.636.289	195.131.864	206.126.192	6%
Mersin İrtibat Bürosu	54.462.052	87.082.150	64.735.151	73.999.487	75.566.313	144.449.845	147.818.675	164.484.443	126.594.393	146.151.531	15%
Eskişehir Bölge Müdürlüğü	32.349.993	47.582.940	55.422.416	55.078.212	70.432.055	105.512.626	185.551.678	172.958.593	159.056.828	183.038.005	15%
Gaziantep Bölge Müdürlüğü	31.303.426	55.034.614	48.932.818	73.430.541	74.295.592	131.311.434	129.109.567	130.707.424	129.544.778	142.503.710	10%
Trabzon Bölge Müdürlüğü	37.338.752	42.651.517	42.897.744	56.165.163	79.814.529	88.477.251	109.735.022	119.981.395	116.781.891	132.835.420	14%
Diyarbakır İrtibat Bürosu	22.992.635	42.160.699	44.452.678	33.585.786	40.113.247	46.486.605	52.381.145	61.692.277	73.739.969	69.579.151	-6%
Elazığ Bölge Müdürlüğü	19.095.890	25.687.470	26.146.514	40.738.913	37.478.283	53.889.417	60.249.027	67.534.427	68.035.175	67.183.769	-1%
Erzurum İrtibat Bürosu	24.622.947	33.825.526	29.495.108	30.026.040	36.300.836	45.990.779	55.499.396	75.257.592	67.625.743	68.415.263	1%
Van İrtibat Bürosu	13.445.240	18.190.981	15.768.034	24.853.236	32.671.178	39.036.314	40.286.286	50.003.669	40.387.088	39.827.437	-1%
Basım İşletme Müdürlüğü	21.095.996	25.029.124	19.233.195	24.404.087	50.978.488	31.025.837	34.289.197	22.118.109	41.484.281	32.028.087	-23%
Genel Toplam	1.127.835.315	1.692.366.445	1.654.990.416	1.708.181.251	2.225.239.368	3.546.858.566	3.738.565.615	3.897.240.890	3.439.098.064	3.661.227.975	6%

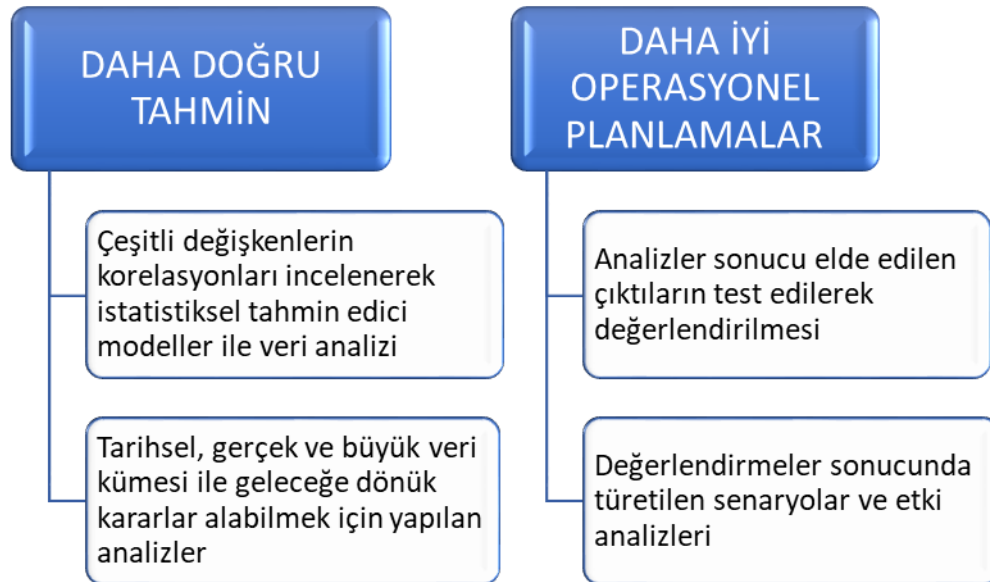
Satış birimlerinin satış tahminleri ay bazında ayrı ayrı yapılmış ve Çizelge 5.3.'de gösterilmiştir. Dokuz Yıllık Satış Birimi Bazında Gerçekleşme ve 2019 Yılı Tahmin Sonuçlarının artışı ve dalgalanmaları Şekil 5.2.'de gösterilmektedir.



Şekil 5.2. Dokuz yıllık satış birimi bazında gerçekleşme ve 2019 yılı tahmin sonuçları

Burada amaç tahminler baz alınarak kurumun risk iřtahi ile birlikte satıř birimlerine satıř hedefi verebilmektir. Verilen satıř hedefleri dođrultusunda satıř birimlerinin performansları ölçülerek bir performans göstergesi olarak kullanılabilir.

Satıř tahminlerinin tutarlı bir řekilde yapılması řirket ya da kurumların bütçeleme ve planlama süreçlerinde hayati öneme sahiptir. Kuřkusuz birçok durumdan etkilenen stratejik kararların, analitik yöntemlerle elde edilen satıř tahminleri ile bütünleřtirilerek alınması daha sađlıklı olacaktır. Büyük ve gerçek bir veri kümesi ile çalıřarak elde edilen tahmin sonuçları dođrultusunda planlamalar yapılacaktır. Daha dođru tahmin ile daha iyi operasyonel planlamalar yapılabileceđi řekil 5.3.'de gösterilmiřtir. Yöneticiler tarafından izlenip deđerlendirilen planlar ile kurumun geleceđi, güçlü ve zayıf yönleri belirlenip performansın artırılabilmesi için kararlar alınacaktır. İstatistiksel veri madenciliđi teknikleri kullanılarak geleceđe yönelik kararlar alabilen řirket ya da kurumlar daha etkili ve tutarlı kararlar alarak sektörde bir adım öne çıkabileceklerdir. Her kurum ve řirket için çok önemli bir performans göstergesi olan 'satıř' bilgisi, tasarruf, yerindelik ve kalkınma teması ile bir dönüşüm hikâyesi bařlatan Devlet Malzeme Ofisi Genel Müdürlüđu için günümüzde ve gelecekte kamu alımları açısından yeniden yorumlanması gereken bir performans göstergesi olarak karřımıza çıkmaktadır [40] .



řekil 5.3. Daha iyi bir tahmin ve daha iyi operasyonel planlamalar

Bu çalışma ile dijitalleşen dünyada var oluş amacı açıklanmaya çalışılan Devlet Malzeme Ofisi ve kamu alımları sektörünün en önemli çıktılarından olan satışların tahmin edilmesi amaçlanmıştır. Böylece kamu alımlarını, tasarruf, yerlilik ve kalkınma açısından değerlendirmenin daha doğru olacağı kanaatine varılmıştır. İstatistiksel veri madenciliği teknikleri kullanılarak elde edilen satış tahmin sonuçları ve kurumun risk iştahı dikkate alınarak belirlenen satış hedeflerinin gerçekleştirilmesi amaçlanmıştır. Bu hedeflerle birlikte satışların artması, az maliyetli ve etkin aksiyon planları uygulanması beklenmektedir. Bu bağlamda ilk aşama olarak istatistiksel veri madenciliği süreci CRISP-DM metodolojisine göre iş bilgisi ile harmanlanmıştır. Sonrasında kurumsal kaynak planlama otomasyon sistemi ile kurum veri tabanında tutulan veri kümesi SQL sorgulama dili ile 2007 - 2019 tarih aralığında çekilmiştir. Mevcut veri kümesinin tanımlayıcı istatistikleri ortaya çıkarılıp bu özet bilgilerden uygulamanın nasıl yapılacağına karar verilmiştir. DMO geçmiş satış veri kümesi aylık ve yıllık bazda derlenerek IBM SPSS Modeler uygulaması ile modellenmiştir. Uygulama sonucunda 2019 yılı satışları tahmin edilmiş ve ilk yedi ay gerçekleşme sonuçları ile karşılaştırması yapılmıştır. Yapılan ampirik uygulama modeli, bir datamart üzerine canlı veri kümesinden beslenecek şekilde sisteme entegre edilmiştir. Bu entegrasyon sayesinde kaydedilen yeni veri kümesi ile istenilen periyotlarda raporlama yapılabilmiştir. Böylece kurum stratejik planında yer alan performans göstergelerinin takibi, kamu alımlarının etkinliğinin artırılması ve kamu alımları ile ilgili politika belirleyicilerin hızlı ve etkili kararlar alabilmesi sağlanmıştır.

KAYNAKLAR

- [1] Arrowsmith, S., & Kunzlik, P. (2009) .Social And Environmental Policies In EC Procurement Law : New Directives And New Directions. Cambridge University Press,. ProQuest Ebook Central, <http://ebookcentral.proquest.com/lib/uludag-ebooks/detail.action?docID=424537>. (Eriřim tarihi : **02 Aralık 2017**)
- [2] Edizdođan, N., etinkaya . ve Gmř, E., Kamu maliyesi(6.Baskı). Bursa:Ekin Yayınevi, **2015**.
- [3] Wiehen, M., Olaya, J.(2006) How To Reduce Corruption İn Public Procurement. K. Kostyo(Ed.).Handbook for curbing corruption in public procurement (13-108). Retrieved December 02, **2017**, http://issuu.com/transparencyinternational/docs/2006_curbingcorruptionpublicprocurement (Eriřim tarihi : **23 Ađustos 2018**)
- [4] Djankov, S., Is public procurement a rich country’s policy? , **2016**, <http://blogs.worldbank.org/governance/public-procurement-rich-country-s-policy> (Eriřim tarihi : **09.10.2017**)
- [5] Kamu Alımları İzleme Raporu, Kamu İhale Kurumu , **2018** https://dosyalar.kik.gov.tr/genel/Raporlar/Kamu_al%C4%B1mlar%C4%B1_izleme_raporu_2018_y%C4%B1lsonu_rev8.pdf (Eriřim tarihi : **15 Mart 2019**)
- [6] Sektör Raporu ,Devlet Malzeme Ofisi ,**2017** <https://www.dmo.gov.tr/Home/Icerik/32> (Eriřim tarihi : **23 Ađustos 2018**)
- [7] Organization of Economic Cooperation and Development., Government At A Glance 2017 stats.oecd.org/# (Eriřim tarihi : **01.12.2017**)

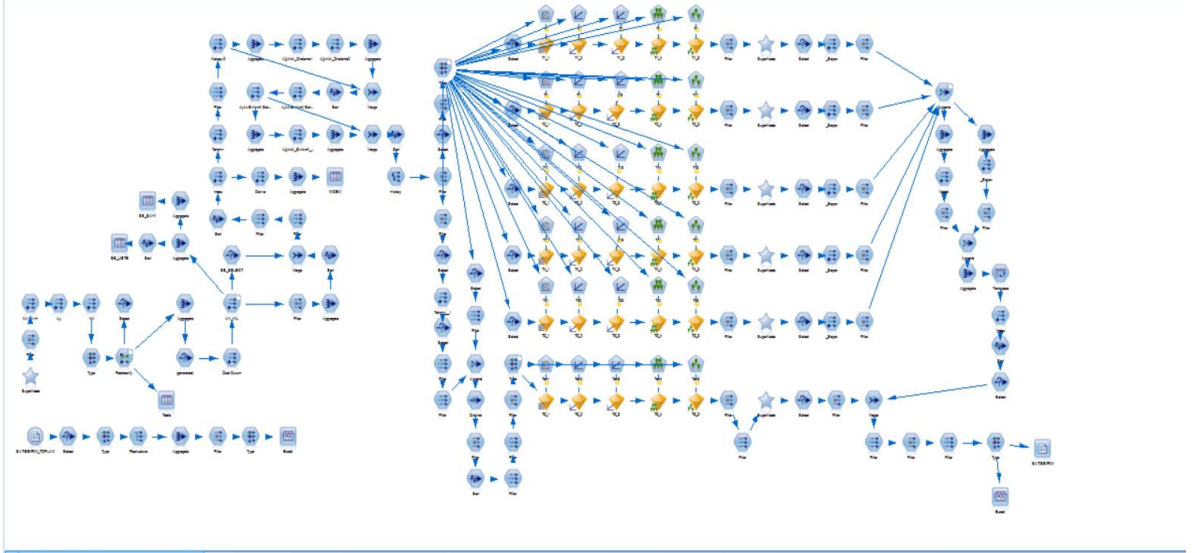
- [8] PricewaterhouseCoopers (PwC), London Economics and Ecorys ,Public procurement in Europe Cost and effectiveness., **2011**, <https://ukmin.lrv.lt/uploads/ukmin/documents/files/Studija%20d%C4%97%20kainos%20ir%20efektyvumo%20vie%C5%A1uosiuose%20pirkimuose.pdf> (Eriřim tarihi : **07.11.2018**)
- [9] Larose, D.T., “Data Mining Methods and Models”, Wiley-İnterscience, United States of America, **2005**.
- [10] Larose, D.T., “Discovering Knowledge in Data”, Wiley-İnterscience, United States of America, **2005**.
- [11] Veri Madencilięi , <http://ozgurdogun.com/?p=19> (Eriřim tarihi : **03.11.2018**)
- [12] Veri Madencilięinin Tarihi , <https://www.exastax.com.tr/veri-analitigi/veri-madenciliginin-tarihi/> (Eriřim tarihi : **03.11.2018**)
- [13] Odabař, Ö., Veri Madencilięi Teknikleri İle Telekom Sektöründe Ayriilan Müřteri Analizi , Yüksek Lisans Tezi, Bilgisayar Mühendislięi ABD, İstanbul Ticaret Üniversitesi, İstanbul, **2017**.
- [14] Wahbeh, A., ve ark., **2011**. A comparison Study between Data Mining Tools over some Classification Methods. International Journal of Advanced Computer Science and Applications, 18-25.
- [15] Han, J., Kamber, M. , “Data Mining Concept and Techniques”, Morgan Kauffman, San Francisco, **2011**.
- [16] Witten, I., Frank, E., Hall, M., **2011**. Data Mining Practical Machine Learning Tools and techniques. Morgan Kaufmann, 558.
- [17] Gorunescu, F., **2011**. Data Mining Concepts, Models, Methods and Algorithms. Springer, 361.
- [18] Argüden , Y., Erřahin, B. , “Veri Madencilięi; Veriden Bilgiye Masraftan Deęere”, Arge Danıřmanlık, **2008**.

- [19] Ahi, L., Veri Madenciliği Yöntemleri İle Ana Harcama Gruplarının Paylarının Tahmini, Yüksek Lisans Tezi, İstatistik ABD, Hacettepe Üniversitesi, Ankara, **2015**.
- [20] Ersöyleyen, F.,Ö., Credit Cardholders Churn Analys is using Data Mining Methods, Master's Thesis, Department Of Master Of Business Administration, Yeditepe Üniversitesi, Adana, **2017**.
- [21] <https://acikerisim.istanbul.edu.tr/bitstream/handle/123456789/28001/45671.pdf?sequence=1&isAllowed=y> (Erişim tarihi: 27.12.2018)
- [22] Erduran, G.,Y., Online Müşteri Şikayetlerinin Veri Madenciliği İle İncelenmesi, Doktora Tezi, İşletme ABD, Trakya Üniversitesi, Edirne, **2017**.
- [23] Berry, M,J.A. Linoff , G. S., Data Mining Techniques ,For Marketing, Sales and Customer Relationship Management, Wiley Publishing, Inc.
- [24] Çığşar, B., Kredi Risklerinde Veri Madenciliği Sınıflandırma Algoritmaları, Yüksek Lisans Tezi, İstatistik ABD, Çukurova Üniversitesi, Adana, **2017**.
- [25] Güldal, H., Karar ağacı algoritmalarının eğitsel veriler üzerindeki performanslarının İncelenmesi , 13.İnternational Balkan Education and Science Congress , **2018**.
- [26] Kılıçalan, M.B., Hanehalki İşgücü Araştırma Verileri İle Veri Madenciliği Yöntemlerinin Uygulanması Ve Modellerin Karşılaştırılması, Yüksek Lisans Tezi, İstatistik ABD, Hacettepe Üniversitesi, Ankara, **2018**.
- [27] Doğan N., Özdamar K., CHAID Analizi ve Aile Planlaması ile ilgili Bir Uygulama, T Klin Tıp Bilimleri, 23, 392-397, **2003**.
- [28] Kayri, M., & Boysan, M. (2007). Araştırmalarda Chaid Analizinin kullanımı ve baş etme stratejileri ile ilgili bir uygulama. Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 40(2), 133-149.
- [29] Kass GV., An Exploratory Technique For Investigating Large Quantities Of Categorical Data, Applied Statistics, 29-2, 119-127 , **1980**.

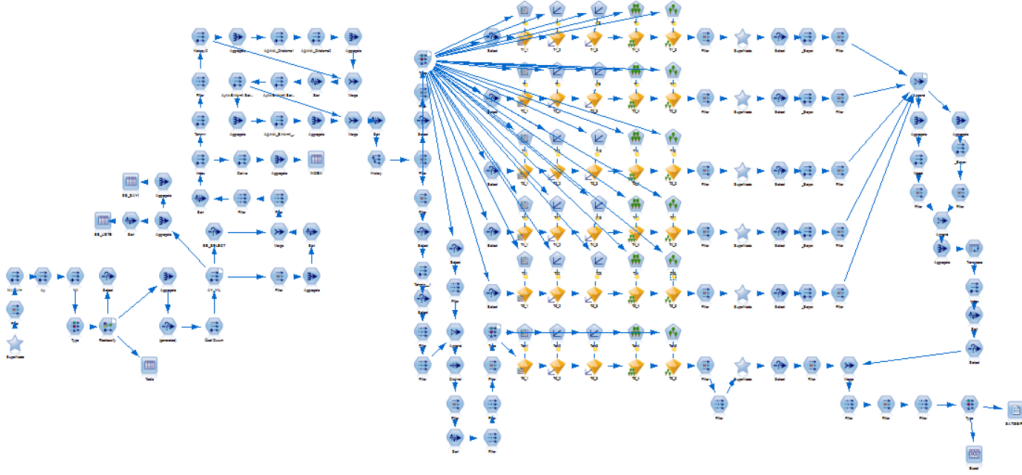
- [30] Şata M., Çakan M., CHAID Analizi ve Lojistik Regresyon Analizi Sonuçlarının Karşılaştırılması*, Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi, **2018**.
- [31] Dursun, G.Ö., İç Kontrol Sisteminde Bir Araç Olarak Veri Madenciliğinin Kullanılmasının İşletmenin Bilançosuna Etkisi, Yüksek Lisans Tezi, İstatistik ABD, Marmara Üniversitesi, İstanbul, **2016**.
- [32] Silahtaroglu, G., .Veri Madenciliği Kavram ve Algoritmaları (3.Baskı). Papatya Yayınevi, **2016**.
- [33] Deveci, M.A., Müşteri İlişkileri Yönetiminde Veri Madenciliği ve İş Zekası Uygulamaları, Yüksek Lisans Tezi, Yönetim Bilişim Sistemleri ABD, Cumhuriyet Üniversitesi, Sivas, **2018**.
- [34] Yakupoğlu, Y., Eğitimsel Veri Madenciliği Ve Bir Uygulaması, Yüksek Lisans Tezi, Endüstri Mühendisliği Programı, İstanbul Teknik Üniversitesi, İstanbul, **2018**.
- [35] Adalı, G.K., Veri Madenciliğinde Birliktelik Yöntemleri ve Müşteri İlişkileri Yönetimine İlişkin Bir Uygulama, İstanbul Üniversitesi, Doktora Tezi, Enformatik ABD, İstanbul, **2017**.
- [36] Khalau, J; Doğrusal Regresyonda Bozulma Noktasına Sahip Tahminler, Yüksek Lisans Tezi, İstatistik ABD, Hacettepe Üniversitesi, Ankara, **2017**.
- [37] Baykal, A., Veri Madenciliği Uygulama Alanları, D.Ü.Ziya Gökalp Eğitim Fakültesi Dergisi 7, 95-107 , **2006**.
- [38] Akkçük , U., “Veri Madenciliği; Kümeleme ve Sınıflama Algoritmaları”, Yalın Yayıncılık, **2011**.
- [39] Yiğit, D., Müşteri İlişkileri Yönetimi, Veri Madenciliğinin Müşteri İlişkileri Yönetiminde Kullanımı Ve Bir Uygulama, Yüksek Lisans Tezi, İşletme ABD, Galatasaray Üniversitesi, İstanbul, **2017**.
- [40] <https://www.ibm.com/downloads/cas/8VQGMN1R> (Erişim tarihi: **09.03.2019**)

EKLER

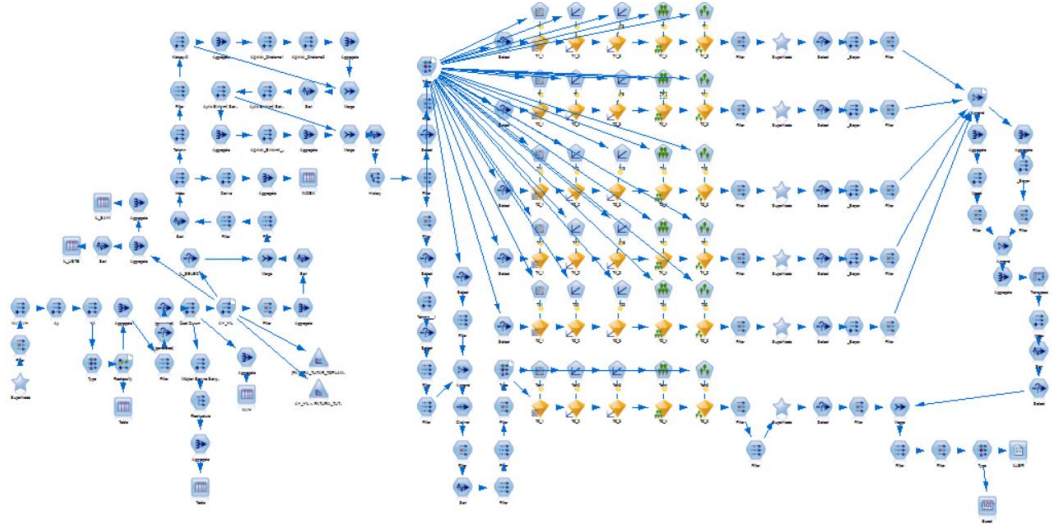
EK 1 – SATIŞ TİPİ SATIŞ TAHMİN ANALİZİ AKIŞI



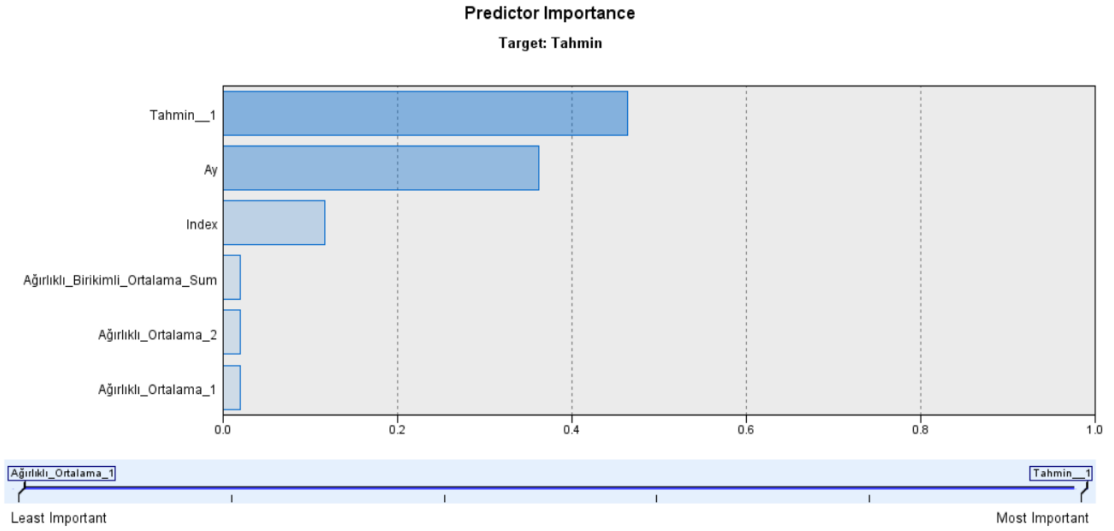
EK 2 – SATIŞ BİRİMİ BAZINDA SATIŞ TAHMİN ANALİZİ AKIŞI



EK 3 – İL BAZINDA SATIŞ TAHMİN ANALİZİ AKIŞI



EK 4– GENELLEŞTİRİLMİŞ DOĞRUSAL MODEL ALGORİTMASI İLE MODEL ÇIKTILARI



Model Information	
Dependent Variable	Tahmin
Probability Distribution	Normal
Link Function	Identity

Case Processing Summary

	N	Percent
Included	98	81.7%
Excluded	22	18.3%
Total	120	100.0%

Categorical Variable Information

			N	Percent
Factor	Ay	1	9	9.2%
		2	9	9.2%
		3	8	8.2%
		4	8	8.2%
		5	8	8.2%
		6	8	8.2%
		7	8	8.2%
		8	8	8.2%
		9	8	8.2%
		10	8	8.2%
		11	8	8.2%
		12	8	8.2%
		Total	98	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Tahmin	98	3143403	454372856	83270587.53	78105327.96
Covariate	Ağırlıklı_Ortalama_1	98	5164638	356402430	106187157.7	84501776.69
	Ağırlıklı_Ortalama_2	98	4984408	336514810	99947193.01	79685036.86
	Index	98	13	110	61.50	28.434
	Ağırlıklı_Birikimli_Ortalama_Sum	98	15	100	40.45	28.357
	Tahmin__1	98	2423870	454372856	72101315.23	74238124.38

Goodness of Fit			
	Value	df	Value/df
Deviance	6.865E+16	84	8.173E+14
Scaled Deviance	98.000	84	
Pearson Chi-Square	6.865E+16	84	8.173E+14
Scaled Pearson Chi-Square	98.000	84	
Log Likelihood ^b	-1814.016		
Akaike's Information Criterion (AIC)	3658.032		
Finite Sample Corrected AIC (AICC)	3663.885		
Bayesian Information Criterion (BIC)	3696.806		
Consistent AIC (CAIC)	3711.806		

Dependent Variable: Tahmin
 Model: (Intercept), Ay, Ağırlıklı_Ortalama_1, Ağırlıklı_Ortalama_2, Index, Ağırlıklı_Birikimli_Ortalama_Sum, Tahmin__1^a

a. Information criteria are in smaller-is-better form.
 b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test

Likelihood Ratio Chi-Square	df	Sig.
211.095	13	.000

Dependent Variable: Tahmin
 Model: (Intercept), Ay, Ağırlıklı_Ortalama_1, Ağırlıklı_Ortalama_2, Index, Ağırlıklı_Birikimli_Ortalama_Sum, Tahmin__1^a

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	4.510	1	.034
Ay	.356	8	1.000
Ağırlıklı_Ortalama_1	. ^a	.	.
Ağırlıklı_Ortalama_2	. ^a	.	.
Index	17.970	1	.000
Ağırlıklı_Birikimli_Ortalama_Sum	. ^a	.	.
Tahmin__1	39.321	1	.000

Dependent Variable: Tahmin

Model: (Intercept), Ay, Ağırlıklı_Ortalama_1, Ağırlıklı_Ortalama_2, Index, Ağırlıklı_Birikimli_Ortalama_Sum, Tahmin__1

a. Unable to compute due to numerical problems

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	108527449.4	18760508.64	71757528.14	145297370.7	33.465	1	.000
[Ay=1]	-138885982	23755894.81	-185446680	-92325283.8	34.180	1	.000
[Ay=2]	-128589103	22724716.86	-173128729	-84049476.1	32.019	1	.000
[Ay=3]	-109598609	21213497.26	-151176300	-68020918.5	26.692	1	.000
[Ay=4]	-106912507	20706626.07	-147496748	-66328265.6	26.659	1	.000
[Ay=5]	-97861436.3	20093604.49	-137244177	-58478695.2	23.720	1	.000
[Ay=6]	-95509743.2	19834635.82	-134384915	-56634571.3	23.187	1	.000
[Ay=7]	-100302584	20287285.45	-140064933	-60540235.6	24.444	1	.000
[Ay=8]	-101924270	20025209.58	-141172960	-62675580.4	25.906	1	.000
[Ay=9]	-86260192.9	18736760.49	-122983569	-49536817.1	21.195	1	.000
[Ay=10]	-94020642.4	19500036.70	-132240012	-55801272.8	23.247	1	.000
[Ay=11]	-84099886.3	18486041.48	-120331862	-47867910.8	20.697	1	.000
[Ay=12]	0 ^a
Ağırlıklı_Ortalama_1	0 ^a
Ağırlıklı_Ortalama_2	0 ^a
Index	546453.681	128907.7400	293799.153	799108.208	17.970	1	.000
Ağırlıklı_Birikimli_Ortalama_Sum	0 ^a
Tahmin__1	.517	.0824	.355	.678	39.321	1	.000
(Scale)	7.005E+14 ^b	1.0008E+14	5.295E+14	9.269E+14			

Dependent Variable: Tahmin

Model: (Intercept), Ay, Ağırlıklı_Ortalama_1, Ağırlıklı_Ortalama_2, Index, Ağırlıklı_Birikimli_Ortalama_Sum, Tahmin__1

a. Set to zero because this parameter is redundant.

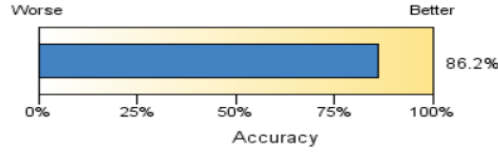
b. Maximum likelihood estimate.

EK 5- DOĞRUSAL MODEL ALGORİTMASI İLE MODEL ÇIKTILARI

Model Summary

Target	Tahmin
Automatic Data Preparation	On
Model Selection Method	Forward Stepwise
Information Criterion	3,372.171

The information criterion is used to compare to models. Models with smaller information criterion values fit better.



Automatic Data Preparation

Target: Tahmin

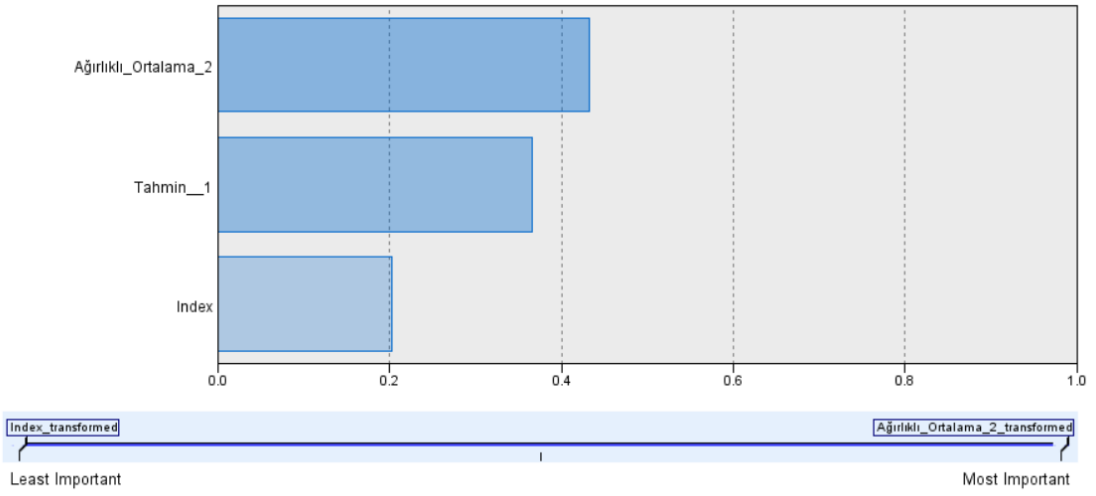
Field	Role	Actions Taken
(Ağırlıklı Birikimli Ortalama_Sum_transformed)	Predictor	Trim outliers
(Ağırlıklı Ortalama_1_transformed)	Predictor	Trim outliers
(Ağırlıklı Ortalama_2_transformed)	Predictor	Trim outliers
(Ay_transformed)	Predictor	Merge categories to maximize association with target
(Index_transformed)	Predictor	Trim outliers
(Tahmin_1_transformed)	Predictor	Trim outliers

If the original field name is X, then the transformed field is displayed as (X_transformed). The original field is excluded from the analysis and the transformed field is included instead.

One or more records were excluded because of a predictor or target that is missing, a frequency weight that is missing or less than one after rounding, or a regression weight that is missing, negative, or zero.

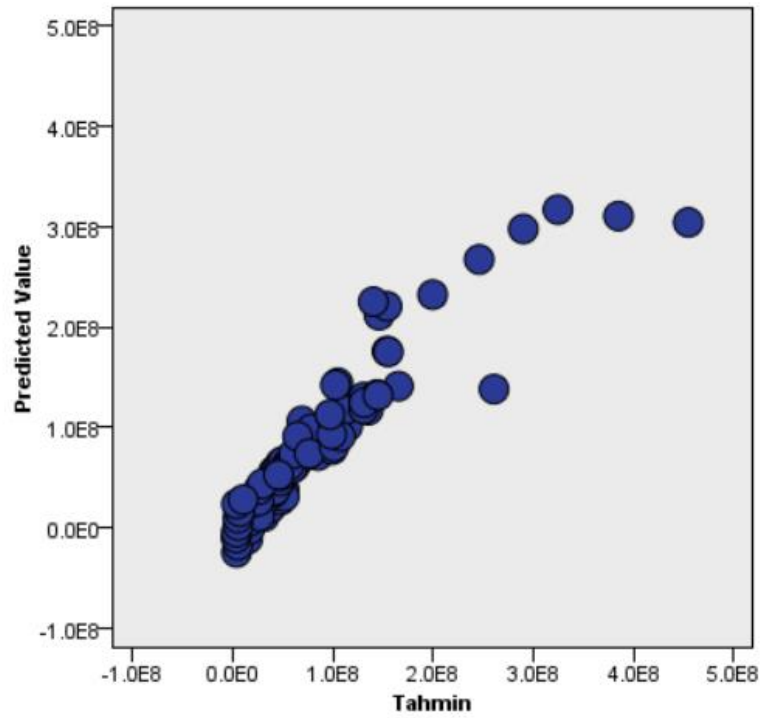
Predictor Importance

Target: Tahmin

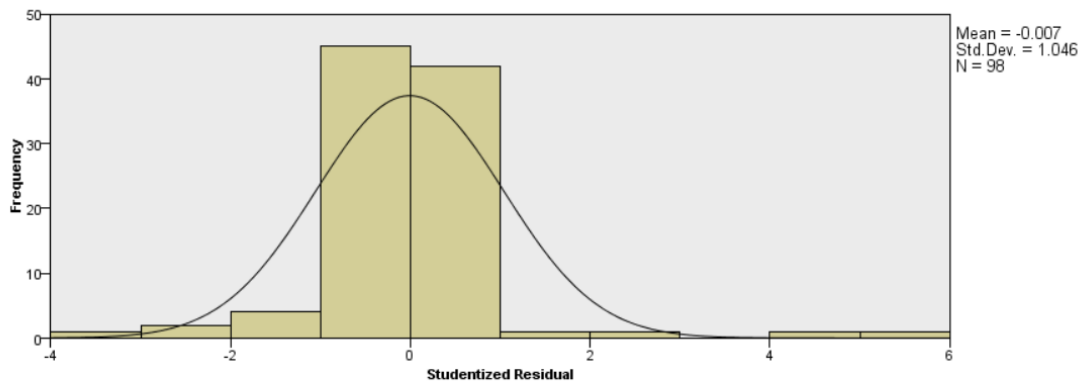


Predicted by Observed

Target: Tahmin



Residuals
Target: Tahmin

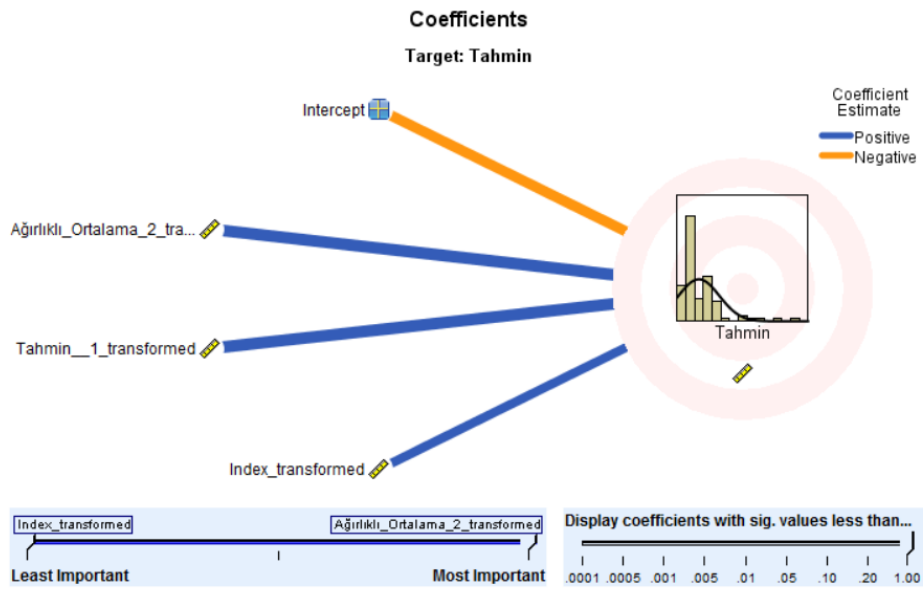
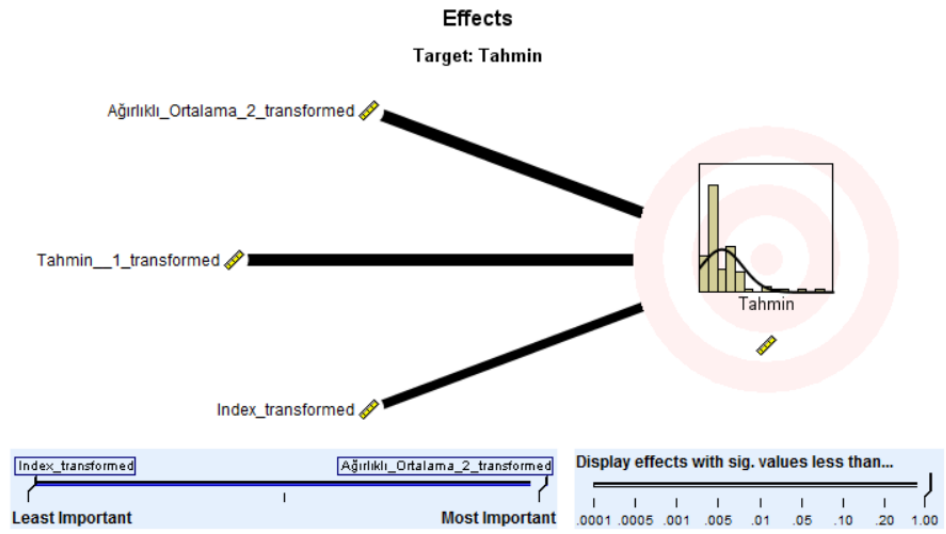


The histogram of Studentized residuals compares the distribution of the residuals to a normal distribution. The smooth line represents the normal distribution. The closer the frequencies of the residuals are to this line, the closer the distribution of the residuals is to the normal distribution.

Outliers
Target: Tahmin

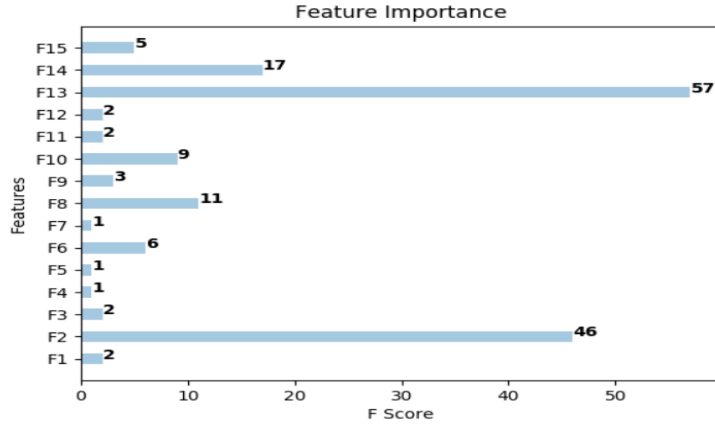
Tahmin	Cook's Distance
140061326.2	1.738
454372855.7	0.944
153966812.3	0.244
145998205.1	0.242
384742965.6	0.234
260555598.5	0.110
199445611.0	0.057

Records with large Cook's distance values are highly influential in the model computations. Such records may distort the model accuracy.



EK 6 – XGBOOST TREE ALGORİTMASI İLE MODEL ÇIKTILARI

XGBoost Tree Predictor Importance



Features names for short

Original field name	Field name on graphic
Ay_7	F1
Index	F2
Ay_5	F3
Ay_4	F4
Ay_3	F5
Ay_1	F6
Ay_6	F7
Ay_9	F8
Ay_8	F9
Ay_12	F10
Ay_11	F11
Ay_10	F12
Tahmin_1	F13
Ağırlıklı_Ortalama_1	F14
Ağırlıklı_Birikimli_Ortalama_Sum	F15

EK 7 – RANDOM TREES ALGORİTMASI İLE MODEL ÇIKTILARI

Random Trees

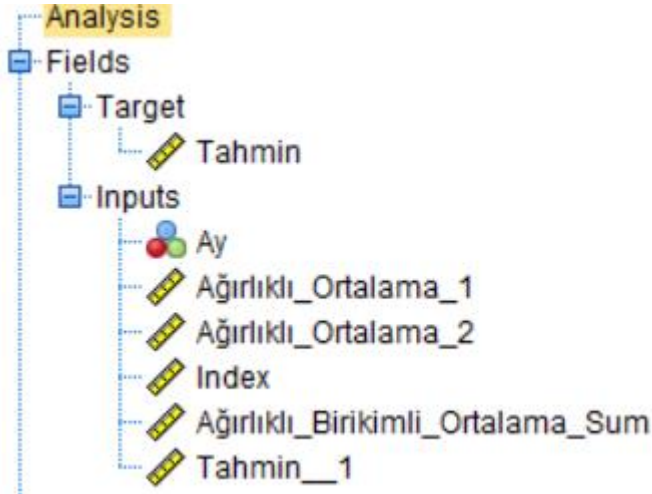
Model Information

Target Field	Tahmin
Model Building Method	Random Trees Regression
Number of Predictors Input	6
Root Mean Squared Error	13,027,401.399
Relative Error	0.358
Variance Explained	0.642

Records Summary

Records	Number	Percent
Included	75	62.50
Excluded ^a	45	37.50
Total	120	100.00

a. User-missing values are excluded.





HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/~~DOKTORA~~ TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 27/08/2019

Tez Başlığı / Konusu: İstatistiksel Veri Madenciliği Teknikleri ile Kamu Alımları Sektöründe Satış Tahmini

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 63 sayfalık kısmına ilişkin, 27/08/2019 tarihinde tez danışmanım tarafından *Turnitin* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 8'dir.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/~~çeviri~~
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

27/08/2019



Tarih ve İmza

Adı Soyadı: Vildan Yıldırım

Öğrenci No: N10229460

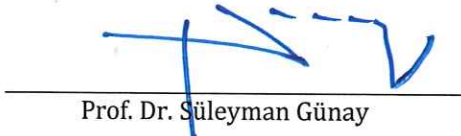
Anabilim Dalı: İstatistik

Programı: İstatistik

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.



Prof. Dr. Süleyman Günay

ÖZGEÇMİŞ

Adı Soyadı : Vildan YILDIRIM
Doğum yeri : Trabzon
Doğum tarihi : 15.05.1985
Medeni hali : Evli
Yazışma adresi : Basınevleri Mah. Selçuklu Cad. 60/7 Keçiören/Ankara
Telefon : 05052344951
Elektronik posta adresi : vildansenel_@hotmail.com
Yabancı dili : İngilizce

EĞİTİM DURUMU

Lisans : Samsun Ondokuz Mayıs Üniversitesi İstatistik Bölümü
Yüksek Lisans : -
Doktora : -

İş Tecrübesi

2009 – Devlet Malzeme Ofisi Genel Müdürlüğü – Merkezi Satınalma Uzmanı