

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE ECZANELER İÇİN İLAÇ
TALEP TAHMİNİ

YÜKSEK LİSANS TEZİ

İlker POYRAZ

Bilgisayar Mühendisliği Ana Bilim Dalı
Bilgisayar Mühendisliği Programı

EYLÜL, 2020

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE ECZANELER İÇİN İLAÇ
TALEP TAHMİNİ

YÜKSEK LİSANS TEZİ

İlker POYRAZ
(Y1713.010081)

Bilgisayar Mühendisliği Ana Bilim Dalı
Bilgisayar Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi Ahmet GÜRHANLI

EYLÜL, 2020

YEMİN METNİ

Yüksek lisans tezi olarak sunduğum “Makine Öğrenmesi Algoritmaları İle Eczaneler İçin İlaç Talep Tahmini” adlı tez çalışmasının, tarafımdan tezin proje safhasından sonuçlanmasına kadar lan bütün süreçlerde bilimsel ahlak ve geleneklere bağlı kalınarak aykırı düşürecek bir yardıma başvurulmaksızın yazıldığını ve yarar sağladığım eserlerin kaynakça’da gösterilen eserlerden oluştuğunu, bu eserlerden atıf yapılarak yararlanmış olduğumu belirtir ve onurumla beyan ederim. (17/09/2020)

İlker POYRAZ

ÖNSÖZ

Günümüzde, eczane ve eczaneler gibi sağlık sistemleri normalde yüksek miktarda veri üretmektedir. Gelişen teknoloji ile birlikte doğru ve güvenilirliği olan talep öngörülere işletmelerin üretim ve verimliliğinin artmasında önem arz etmektedir. Geleceğe yönelik yapılan güvenilir tahminlerle işletmeler; satış ve pazarlama faaliyetlerini, fiyat politikalarını, çevresel değişimlere bağlı etkenleri etkin bir biçimde planlama olanağına sahip olmaktadır. Tüm sektörlerde olduğu gibi sağlık alanındaki sektörlerde insanların ilaç tedarik ihtiyaçlarını karşılayan eczaneler, talep tahminleri tedarik süreci için önemli bir yer oluşturmaktadır. İnsan yaşamında bozulan sağlık davranışları ve dengelerinin iyileştirilmesinde kilit rolleri üstlenen ilaçlar, günümüzde artan bir oranda satın alınmakta ve tüketilmektedir. Bu çalışmanın amacı, bir eczanedeki ilaç satın alımına yönelik veriler kullanılarak sonraki zamanlardaki satış miktarları tahminlenmesini sağlamaktır.

Hayatımın her zamanında maddi ve manevi varlıklarını hissettiğim çok değerli ve kıymetli eşim Reyhan POYRAZ'a, yaşama ve hayat sevinçlerim olan kızlarım Elanaz ve Zeynep Sare'ye varlıklarına her zaman şükrettiğim aileme teşekkür ve şükranlarımı sunarım.

Bu çalışmanın gerçekleştirilme sürecinde, değerli fikir , bilgi ve düşüncelerini benimle paylaşan, her zaman danıştığım değerli ve kıymetli zamanını ayırıp büyük bir sabırla ve ilgiyle bana faydalı bilgiler aktarmak için elinden geleni yapmaya, aktardığı değerli bilgilerden dolayı kıymetli baba'ma şükranlarımı bir borç bilir ve sevgi, saygılarımı sunarım.

Tez çalışmamın planlanmasında, araştırılmasına, yürütülmesine ve oluşumunda her zaman bilgisini, ilgi ve desteğini hiç esirgemeyen, üstün tecrübe ve bilgilerini aktaran, bilgilendirme ve yönlendirmeleri ile çalışmamı bilimsel temeller doğrultusunda şekillendiren gelecekteki hayatında çok daha başarılı olup başarılarına imza atacağına inandığım kıymetli tez danışman hocam **Dr. Öğr. Üyesi Ahmet GÜRHANLI** sonsuz teşekkürlerimi ve saygılarımı sunarım.

Eylül, 2020

İlker POYRAZ

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iv
İÇİNDEKİLER	v
KISALTMALAR	vii
ÇİZELGE LİSTESİ.....	viii
ŞEKİL LİSTESİ.....	ix
ÖZET.....	xi
ABSTRACT	xii
1. GİRİŞ	1
1.1 Veri Madenciliği	2
1.2 Veri Madenciliği Tarihçesi.....	3
1.3 Veri Madenciliği Tanımı	3
1.4 Veri Madenciliği Veri Bilgi Keşfi Süreci.....	4
1.5 Veri Madenciliği Uygulama Alanları.....	5
1.6 Veri Madenciliği Süreçleri	7
1.6.1 İşi ve problemi tanımlama.....	8
1.6.2 Verinin anlama ve tanımlama	8
1.6.3 Verinin hazırlanması	8
1.6.4 Modelleme	9
1.6.5 Değerlendirme & uygulama.....	9
1.7 Veri Madenciliği Yöntemleri	10
1.7.1 Tahmin edici yöntemler	10
1.7.1.1 Sınıflandırma.....	11
1.7.1.2 Regresyon.....	12
1.7.2 Tanımlayıcı yöntemler	13
1.7.2.1 Kümeleme	13
2. LİTERATÜR ÇALIŞMASI	15
3. MATERYAL VE YÖNTEM.....	18
3.1 Veri Madenciliği Yöntemleri	18
3.2 Tahmin Edici Yöntemler	19
3.3 Zaman Serileri	19
4. WEKA VERİ MADENCİLİĞİ UYGULAMASI.....	22
4.1 Weka Zaman Serisi Öngörü Modülü Uygulaması	23
4.2 Weka Uygulaması Zaman Serileri Öngörü Algoritmaları	29
4.2.1 Lineer regresyon algoritması	29
4.2.2 Gaussian process algoritması	30
4.2.3 M5Rules algoritması	31
4.2.4 Multilayer perceptron algoritması.....	31
4.2.5 SMOReg algoritması	31
4.2.6 M5P algoritması	32
4.2.7 Random forest algoritması	32

4.3 Tahmin Yöntemlerinin Hata Ölçüm Teknikleri	32
4.3.1 Ortalama mutlak hata	32
4.3.2 Ortalama mutlak yüzde hatası	33
4.3.3 Kök ortalama kare hatası	33
4.4 Uygulama Ortamı Donanım Özellikleri	34
4.5 Gerçekleştirilen Uygulama Çıktıları ve Bulguları	34
4.5.1 Gerçekleştirilen uygulama çıktıları	34
4.5.1.1 DEVIT-3 AMPUL için linear regresyon algoritması	35
4.5.1.2 DEVIT-3 AMPUL için m5rules algoritması	37
4.5.1.3 DEVIT-3 AMPUL için gaussianprocess regression algoritması	39
4.5.1.4 DEVIT-3 AMPUL için multilayer perceptron algoritması	41
4.5.1.5 DEVIT-3 AMPUL için smoreg algoritması	44
4.5.1.6 DEVIT-3 AMPUL için mp5 algoritması	46
4.5.1.7 DEVIT-3 AMPUL için random forest algoritması	49
4.6 Bulgular	51
5. SONUÇLAR	59
KAYNAKLAR	60
ÖZGEÇMİŞ	62

KISALTMALAR

AI	: Artificial Intelligence
ENIAC	: Electrical Numerical Integrator And Calculator
ABD	: Amerika Birleşik Devletleri
ML	: Machine Learning
ARFF	: Attribute-Relation File Format
CPU	: Central Process Unit
RAM	: Random Access Memory
MLP	: Çok Katmanlı Algılayıcılar

ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 3.1 : Uygulama Ortamı Donanım Özellikleri	34
Çizelge 4.1: Eğitim Verileri için Tahminler: DEVIT-3 AMPUL.....	52
Çizelge 4.2: Test Verileri için Tahminler: DEVIT-3 AMPUL.....	56
Çizelge 4.3: Ortalama Mutlak Yüzde Hata Sonuçları DEVIT-3 AMPUL.....	57

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1: Bilgi Keşif Serüveni	5
Şekil 1.2: Veri Madenciliği Süreci	8
Şekil 1.3 : Veri Madenciliği Yöntemleri	10
Şekil 1.4: Müşteri Maaşı Bazında Risk Sınıflandırması.....	12
Şekil 1.5: Gelir - Yaş Dağılımı Kümelemesi.....	14
Şekil 4.1: WEKA GUI Arayüzü	22
Şekil 4.2: X Eczanesinin 2015-2019 Yılı Satış İlaç Verileri.....	23
Şekil 4.3: WEKA Arff Dosya Formatı	24
Şekil 4.4: WEKA Explorer Uygulama Alanı	25
Şekil 4.5: WEKA Explorer Visualizer Data Viewer	25
Şekil 4.6: WEKA Explorer Visualizer All Attributes Görünümü	26
Şekil 4.7: WEKA Forecast Basic Target Selection Uygulama.....	26
Şekil 4.8: WEKA Forecast Advanced Base Learner Uygulama	27
Şekil 4.9: WEKA Forecast Advanced Lag Creation Uygulama Alanı.....	28
Şekil 4.10: WEKA Forecast Advanced Lag Creation More Options Alanı	28
Şekil 4.11: WEKA Forecast Advanced Evaluation Uygulama Alanı	29
Şekil 4.12: WEKA Forecast Advanced Output Uygulama Alanı.....	29
Şekil 4.13 : Tahminleme Yapılacak İlaç Listesi.....	34
Şekil 4.14: DEVIT-3 AMPUL için Linear Regresyon Çıktıları.....	35
Şekil 4.15: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği	35
Şekil 4.16: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği	36
Şekil 4.17: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği.....	36
Şekil 4.18: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği.....	36
Şekil 4.19 : DEVIT-3 AMPUL için M5Rules Çıktıları.....	37
Şekil 4.20: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği	37
Şekil 4.21: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği	38
Şekil 4.22: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği.....	38
Şekil 4.23: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği.....	39
Şekil 4.24: DEVIT-3 AMPUL için Gaussian Process Regression Çıktıları.....	39
Şekil 4.25: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği	40
Şekil 4.26: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği	40
Şekil 4.27: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği.....	41
Şekil 4.28: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği.....	41
Şekil 4.29: DEVIT-3 AMPUL için Multilayerperceptron Çıktıları	42
Şekil 4.30: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği	42
Şekil 4.31: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği	43
Şekil 4.32: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği.....	43
Şekil 4.33: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği.....	44
Şekil 4.34: DEVIT-3 AMPUL için SMOreg Çıktıları	44
Şekil 4.35: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği	45

Şekil 4.36: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği	45
Şekil 4.37: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği	46
Şekil 4.38: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği	46
Şekil 4.39: DEVIT-3 AMPUL için MP5 Çıktıları	47
Şekil 4.40: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği	47
Şekil 4.41: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği	48
Şekil 4.42: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği	48
Şekil 4.43: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği	49
Şekil 4.44: DEVIT-3 AMPUL için Randomforest Çıktıları.....	49
Şekil 4.45: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği	50
Şekil 4.46: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği	50
Şekil 4.47: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği	51
Şekil 4.48: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği	51
Şekil 4.49: Algoritma MAPE Karşılaştırması	58

MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE ECZANELER İÇİN İLAÇ TALEP TAHMİNİ

ÖZET

Sağlık sektöründe eczane ve eczaneler gibi sağlık sistemlerinin gelişen teknoloji ile birlikte yüksek düzeyde veri sağladığı görülmektedir. Teknolojinin iş süreçlerinin ayrılmaz bir parçası haline geldiği daha bütünleşmiş bir dünyaya ilerledikçe, bilgi aktarımı süreci daha karmaşık hale geldi.

Gelişen teknolojide , ilaç firmalarının stoklarını yönetmelerine ve yeni ürün ve hizmetler geliştirmelerine yardımcı olmak için giderek daha fazla kullanılmaya devam etmektedir.Sağlık sistemlerinde veri madenciliği yöntemlerinin kullanılması geleneksel bir süreç haline gelmiştir. Büyük veri kümelerinden bilgi keşfi olan veri madenciliği, ilaç firmalarının ilaç keşfi ve dağıtım yöntemlerinin kalitesini iyileştirme kalıplarını keşfetmelerine yardımcı olmaktadır.

Bu çalışmanın amacı, bir eczanedeki ilaç satın alımına yönelik veriler kullanılarak sonraki zamanlardaki satış miktarları tahminlenmesini sağlamaktır. Bu tahminlemeler sayesinde eczanede satınalma ile tüketilen ilaçlar üzerinden ilaçların depolanma stok durumları da kontrol altına alınabilir. Bu çalışmada Türkiye’deki bir eczanenin 2015 Ocak ayı ile 2019 Aralık ayı arasındaki 5 yıllık ilaç satış verileri düzenlenmiş ve Weka programı ile zaman serileri uygulanarak haftalık olarak yapılan tahminleme çalışmalarında Makine öğrenme algoritmalarından LinearRegresyon, GaussianProcess, M5Rules, MultilayerPerceptron, SMOreg, M5P, RandomFOREST kullanılmıştır.Bu algoritmaların Ortalama mutlak yüzde hatası(MAPE) karşılaştırılarak en başarılı tahmin modeli bulunmaya çalışılmıştır.

Anahtar Kelimeler: *Makine Öğrenmesi, Tahmin, Zaman Serileri, İlaç Satış Tahmin*

DRUG DEMAND FORECASTING FOR PHARMECIES WITH MACHINE LEARNING ALGORITHMS

ABSTRACT

Today, it is seen that health systems such as pharmacies and pharmacies in the health sector have provided a high level of data with the developing technology. As the technology progresses into a more integrated world where business has become an integral part of business processes, the process of knowledge transfer has become more complex.

In the developing technology, it continues to be used more and more to help pharmaceutical companies manage their stocks and develop new products and services. The use of data mining methods in health systems has become a traditional process. Data mining, which is information discovery from large data sets, helps pharmaceutical companies discover patterns of improving the quality of drug discovery and delivery methods.

The aim of this study is to estimate the sales amounts in the next periods by using the data for drug purchase in a pharmacy. Thanks to these estimates, the storage status of the drugs can be taken under control over the drugs consumed in the pharmacy. In this study, a pharmacy in Turkey, 5 years of pharmaceutical sales data are arranged between the months of December 2019 and January 2015 and the Weka program with time series on a weekly basis to do the forecasting work by applying machine learning algorithms linearregresyo's, gaussianprocess, m5rules, multilayerperceptro's, smoreg, M5P, randomforest is used By comparing the average absolute percentage error (MAPE) of these algorithms, it was tried to find the most successful prediction model.

Keywords: *Machine Learning, Forecasting, Time Series, Drug Sales Estimation*

1. GİRİŞ

Geçmiş yıllarda ilaç endüstrisindeki bilgi akışı nispeten basit ve teknolojinin uygulanması sınırlıydı. Ancak, teknolojinin iş yaşamında ayrılmaz bir parçası haline geldiği daha bütünleşik bir dünyaya gidildikçe, bilginin aktarılma süreci daha karmaşık hale geldi. Günümüzdeki teknoloji ile ilaç endüstrisindeki firmalarının stoklarını yönetmelerine ve yeni ürün ve hizmetler geliştirmelerine yardımcı olmak için giderek daha fazla kullanılmaya başlandı. Sağlık sektörü ivme kazandıkça ve çok çeşitli yazılım teknolojileri ve hizmetleriyle etkileşime girdiğinden, veri madenciliği ön plana çıkmaktadır. Çok sayıda sağlık şirketi, tıbbi hastane ve ilaç üretim birimi, mükemmel verimlilikleri nedeniyle veri madenciliği araçlarını kullanmaktadır.

İlaç sektöründe bilgi teknolojilerinin diğer olası kullanımları arasında fiyatlandırma ve dikey olarak entegre ilaç şirketleri arasında karşılıklı yarar için bilgi alışverişi bulunmaktadır. Bu sayede veri toplama yöntemlerinin geliştirilmiş veri manipülasyon teknikleri henüz bunlara ayak uydurmamasına rağmen zorluk devam etmektedir.

Daha önceki yıllarda ilaç endüstrisindeki bilgi akışı ve süreci değerlendirildiğinde biraz daha basitti ve teknolojinin uygulanması sınırlıydı. Daha sonraları teknolojinin iş süreçlerinin ayrılmaz bir parçası haline geldiği daha entegre ve bütünleşik bir dünyaya ilerledikçe, bilgi aktarımı süreci daha karmaşık hale geldiği gözlemlenmektedir. Veri madenciliği, sağlık ve ilaç endüstrisinin giderek ayrılmaz bir parçası haline gelmektedir. Veri madenciliği, anlamlı kural ve kalıpları keşif için büyük veri kümelerinin araştırılması ve analizidir. Veri madenciliğinde sıklıkla kullanılan istatistiksel analiz tekniklerinden biride zaman serileri analizidir. Zaman serileri denilince, bir değişkenin zaman içerisindeki hareketlerini izleyen, ve elde ettiği gözlem sonuçlarını zamana bağlı olarak dağılım gösterdiği seriler olarak ifade edilmektedir. Zaman serileri frekanslı seriler olup, serilerin frekansları yıllık, dört'er aylık, üçer aylık, aylık, günlük ve haftalık olarak değişim gösteren değerler alabilmektedirler[1]. Zaman

serileri analizi ile geçmiş deęerlerin zaman ierisinde gosterdięi davranış Őeklini dikkate alır. Bundan ama, geçmiş deęerlere bakılarak gelecekle ilgili tahminlerin yapılmasıdır.

Zaman serileri analizi, regresyon analizi gibi neden-sonu iliŐkisinden ok, geçmiş verilerin deęişimleri incelenerek geleceęe yonelik tahminler yapmak iin kullanılır. Zaman serileri analizinde iŐletmelerin ve kuruluşların geçmiş satıŐ deęerleri incelenerek belirlenen bir eęilim olup olmadığına bakılarak, gelecekle ilgili talep tahminlerinin yapılmasıdır[2]. Zaman serileri analizi ile gelecekte gerekleşmesi beklenen olaylar saptanabilir. Zaman serisi ile gelecek'e donük tahmin yurütme iŐlemleri de icraa edilebilmektedir. zellikle nem arz eden finansal deęerler üzerinde ve stratejik nem arz eden yonetim karar surelerinde oldukça nemlidir[3].Bu alıŐmada Turkiye'deki bir eczanenin 2015 Ocak ayı ile 2019 Aralık ayı arasındaki 5 yıllık ila satıŐ verileri duzenlenmiŐ ve Weka programı ile zaman serileri uygulanarak haftalık olarak yapılan tahminleme alıŐmalarında Makine ğrenme algoritmalarından LinearRegresyon, GaussianProcess, M5Rules, MultilayerPerceptron, SMOreg, M5P, RandomFOREST kullanılmıŐtır.Bu algoritmaların Ortalama mutlak yuzde hatası(MAPE) karŐılaŐtırılarak en baŐarılı tahmin modeli bulunmaya alıŐılmıŐtır.

1.1 Veri Madencilięi

Veri madencilięi, anlamlı ve belirgin kalıpları ve bu kalıplar ile birlikte kuralları keŐif etmek iin yuksek hacimli verilerin araŐtırma ve inceleme analizidir. Veri bilimi alıŐma bolgesi altında sıklı duzenli yani disiplinli olarak ifade edilir ve genellikle tahmine ve neriye dayanan analitikler'den farklılık gosterir, unku tarihsel olarak verileri ifade ederken veri madencilięi gelecek zamanlardaki sonuları tahmin ve ngorude bulunmayı amalar. Bunun ile birlikte, veri madencilięi sureleri, arama motorları algoritmaları ile neri ve tahmin sistemleri gibi modern olarak yapay zeka(AI) uygulama ve surelerine gu katan makine ğrenimi(ML) modelleri yaratmak iin kullanılmaktadır. İŐletme ve kuruluşların amalarına bir adım yakın olmalarına ve daha iyi stratejik kararlar almaları iin alıŐmaktadırlar. Veri madencilięi, verilerin etkili bir Őekilde toplanması ve depolanması ile bilgisayar ortamına iŐlemeyi

içermektedir. Mevcut verileri bölümleyerek ve gelecek zamanlardaki olayların öngörülüğünü değerlendirmek için, veri madenciliği kompleks matematiksel algoritmalar kullanılarak sonuç üretmektedir.

1.2 Veri Madenciliği Tarihçesi

Veri madenciliğinin denilince köken olarak ilk ifade edilen sayısal bilgisayar ENIAC'a kadar inilmektedir. 1946 tarihinde geliştirilen ve zamanımızda'da kullandığımız bireysel bilgisayarların babası olarak ifade edilen ENIAC, ABD'li bilim insanları J. Presper Eckert ve John Mauchly eliyle, II. Dünya Savaşı yıllarında Amerika Birleşik Devleti ordusu için tasarlanıp geliştirilmiştir. 170 m²'lik bir alanı içeren ve 30 tonluk ağırlığıyla "ilk" bilgisayarın 60 yıllık geçmişi ile geçirmiş olduğu gelişimin sonucunda boyutlarını günümüzde kullandığımız masa üstündeki bilgisayarlara bakarak görmemiz mümkündür[4].

Veri madenciliğinin ilk uygulamaları pazarlama alanında olmuştur. Veri madenciliği, kavram olarak 1960'lı zamanlarda, bilgisayarların veri süreç problemlerini çözümlenmeye başlamasıyla ilk adımını atmıştır. O zamanlarda bilgisayar sayesinde gereğince büyük bir tarama işlemi yapıldığında, beklenen verileri yakalamın mümkün olacağı gerçeği kabullenilmiştir. Bu çalışma ile süreçte veri madenciliği ile değilde önceleri veri yakalanması, veri taraması gibi adlar verilmişti. 1990'lı zamanlara gelindiğinde ise veri madenciliği adı, bilgisayar mühendislerinin düşüncesi ile ortaya sürüldü. Bilgisayar Mühendislerinin amacı, geleneksel olarak istatistiksel metod ve yöntemler ile değilde, veri analizinde kullanılan algoritmik bilgisayar modülleri olarak değerlendirmesini belirtmekti. Bu durumdan sonra bilim insanları veri madenciliğine çeşitli kavramlar getirmeye başladılar[5].

1.3 Veri Madenciliği Tanımı

Veriden en iyi şekilde faydayı sağlamak için uygulanan yöntemlerin karmaşık olması ve zorluğu nedeni ile bu süreçte gerçekleşecek olan çalışmaların bilgisayar ortamında yapma fikrini ortaya çıkarmış, bu sayede çeşitli istatistiksel ve matematiksel hesaplamalar içeren algoritmaların gelişimi ve "Veri Madenciliği" kavramının çıkmasını sağlamıştır. Basit şekilde tanımlama

yapılırsa veri madenciliği , büyük veritabanlarından önceden bilinmeyen bilgiyi elde etmek ve bilgiyi tutma işidir. Başka bir deyişle, büyük veri kümeleri arasından gelecekle ilgili öngörüyü bulabilmemizi sağlayabilecek ilişkilerin bilgisayar programları aracılığıyla aranama işlemini gerçekleştirebilmektir. Son yıllarda yapılan çalışma ve süreçlerde veri madenciliği kavramı dışında veri ambarlarında bilgi madenciliği, bilgi elde edebilme, veri ve örüntü süreci ile veri arkeolojisi gibi yeni kavramlar da literatüre geçmiştir. Bu tanımlardan genel olarak kullanım veritabanlarında bilgi keşfi olarak ifade edilmektedir. Diğer yönden değerlendirmelerde ise veri madenciliği bakıldığında bilgi keşif sürecinin bir modeli şeklinde ifade edilmektedir. Veri madenciliği, hacim olarak büyük miktarlarda ve hızlı şekilde elde edilen verilerin, çeşitli süreçler doğrultusunda anlamlı bir şekilde bilgiye dönüştürülmesinde gerçekleşen bir süreçtir. Veri madenciliği ifadesine bakıldığında ifadelerde benzer olan unsurların birincisi çok miktarda verinin veri tabanında saklanması, ikincisi olarak bu eldeki verilere bakılarak anlamlı bilgiler çıkartılmasıdır.

1.4 Veri Madenciliği Veri Bilgi Keşfi Süreci

Veri madenciliğinde bilgi keşfi, elimizdeki veri ve verilerden düzenli yararlı, doğru ve anlamlı model ve kalıp'lar elde etmek için kullanılan süreç olarak tanımlayabiliriz. Veri madenciliği, ile ifade edilen bilgi, veri keşfi ve veri ambarlarındaki bilgi, veri keşfi ifadeleri bazı bilim araştırmacıları tarafından karıştırılmaktadır. Günümüzde bir çok araştırmacı ve uygulayıcı veri madenciliği ve bilgi keşif terimleri benzer şekilde ve anlamda kullanılmaktadır. Lakin veri madenciliğinde bilgi ve verinin keşfi sürecinin bir adımıdır. Veri ambarlarında bilgi keşfi sürecini kısaca tanımladığımızda; verideki faydalı yada anlamlı, orjinal olarak ve belli bir anlamı olan ilişkileri ortaya çıkarma sürecidir[6].

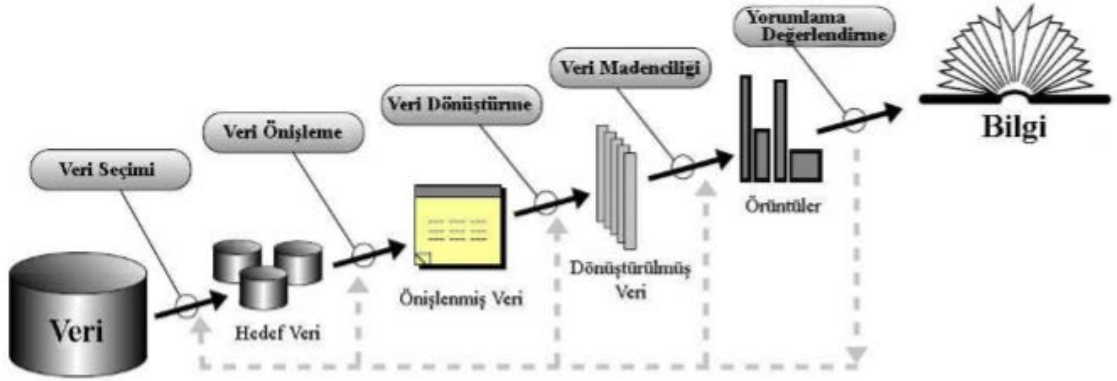
Veri ambarlarında veya veri tabanlarında gizli bilgiyi bulabilmek amacıyla ihtiyaç duyulan yeni nesil süreç hesaplama araçları ve teknikleri, veri ambarlarındaki bilgi keşif sürecinin oluşumunu gerçekleştirmiştir[7]. Basit olarak değerlendirdiğimizde bilgi keşif süreci büyük verilerin değerlendirilmesindeki bilgisayar destekli hesaplamalardır[8].

Bilginin Keşfi süreci aşağıdaki takip eden adımlardan oluşmaktadır[8]:

- Verilerin Temizlenmesi (düzensiz , tutarsız verilerin temizlemek)
- Verilerin Bütünleştirilmesi (bir ve bir çok veri kaynağının birleştirilmesi)
- Verilerin Seçim Süreci (veri ambarlarından araştırılarak elde edilen verileri belirlemek)
- Verilerin Dönüştürülmesi (diğer aşamada kullanılacak verileri verimli biçime dönüştürülmesi)
- Veri Madenciliği (veri tasarımlarını ortaya çıkması için uygulanan mantıklı yöntemlerin bir araya gelmesi)
- Tasarım Değerlendirilmesi (farklı ölçümlere bağlı bilgiyi sunmak için ilginç tasarımların belirlenmesi)
- Bilgi Tanıtımı (bulunan bilgiyi kullanmak için kullanıcılarından gözünde

canlandırma ve bilgiyi tanıtma)

Veriden en iyi şekilde baştan dört etapta veri madenciliği aşamalarında verinin hazır hale getirilmesi için uygun verinin ön işleme adımı olarak değerlendirilmektedir.



Şekil 1.1: Bilgi Keşif Serüveni

1.5 Veri Madenciliği Uygulama Alanları

Veri madenciliği pazarlama , e-ticaret , sigortacılık, bankacılık ve sağlık gibi birçok farklı sektör ve alanlarda kullanıldığı görülmektedir. Geniş ve büyük veri

ambarlarının oluşturulmasına olanak sağlayan yerlerde veri madenciliğinin kullanılması doğru sonuçların üretilmesi ve katkı sağladığı görülmektedir.

Pazarlama

Pazarlama alanı ile ilgili olarak günümüze kadar yapılmakta olan uygulamaların bir kısmı aşağıdaki şekilde ifade edilmiştir.

Müşteri satın alma alışkanlıkları belirlenmesi

Müşterilerin takibinin yapılarak müşteriler bir ürün aldıktan sonra akabinde başka bir ürün alma eğiliminde bulunuyor mu? Örneğin sucuk ve yumurta alan bir müşterinin bunun yanında çay alma eğiliminde bulunuyor mu?

Bu gibi bilgiler işletmeler için büyük önem taşımaktadır. Fakat veri madenciliği tanımına baktığımızda önceden alınan bir malı bilmeyen veya tahmin edilmeyen bilgilerin ortaya çıkartılmasında çalışılıyordu.

Pazar sepeti analizi

Pazar sepeti analizi, her hangi bir yerden belirli bir ürün kümesi satın alırsanız, başka bir ürün kümesi satın alma olasılığınızın daha yüksek olduğu teorisine dayanan bir modelleme tekniğidir. Bu teknik, perakendecinin bir alıcının satın alma davranışını anlamasına izin verebilir. Bu bilgiler, satıcının alıcının ihtiyaçlarını bilmesine ve mağazanın düzenini buna göre değiştirmesine yardımcı olabilir. Farklı analizler kullanılarak farklı mağazalar arasındaki, farklı demografik gruplardaki müşteriler arasındaki sonuçların karşılaştırılması yapılabilir.

Satışların tahminlemesi

Satış hacimlerini doğru bir şekilde tahmin etmek perakende şirketleri için hayati öneme sahiptir. Bununla birlikte, önemli bir istisna, satışların pazarlama kampanyalarından veya promosyonlarından etkilendiği görülmektedir[9].

Sağlık hizmetleri

Veri madenciliği, sağlık sistemlerini iyileştirmek için büyük bir potansiyele sahiptir. Bakımı iyileştiren ve maliyetleri azaltan en iyi uygulamaları tanımlamak için veri ve analiz kullanır. Araştırmacılar, çok boyutlu veri tabanları, makine öğrenimi, yumuşak bilgi işlem, veri görselleştirme ve

istatistik gibi veri madenciliği yaklaşımlarını kullanmaktadır. Hastaların doğru yerde ve zamanda uygun bakım almasını sağlayan süreçler geliştirilmiştir. Veri madenciliği ayrıca sağlık sigortacılarının sahtekarlık ve kötüye kullanımı tespit etmelerine yardımcı olabilir[10].

Finans ve bankacılık hizmetleri

Veri madenciliği finans alanında önemli bir konuma sahiptir bunun yanında bankacılık sektörüne baktığımızda yoğun olarak kullanıldığı gözlenmektedir. Banka kendi müşterilerinin sahip olduğu kart tiplerine ve çektiği kredinin geri ödenebilirliğini kontrol ederek müşterinin profillerini ve segmentasyonlarını belirlemesi sürecinde madencilik teknikleri sayesinde başarılı şekilde sonuç üretilmektedir.

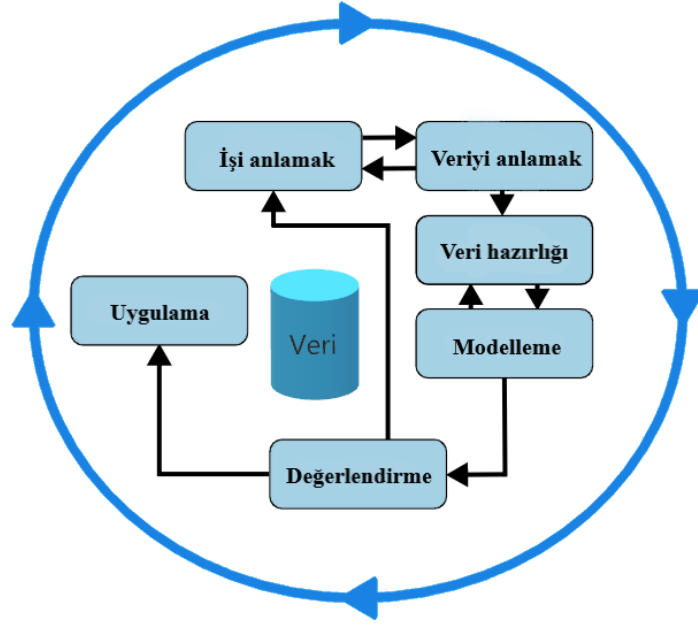
Sigortacılık hizmetleri

Sigorta Şirketleri veri yapısı olarak yapılan işlemlere istinaden büyük bir veriye sahiptirler ve bu verilerin düzenli hale getirilip bir bütün oluşturup bu verilerin analizinde önem arz etmektedir. Sigortacılık sektöründe veri madenciliği , yeni poliçe oluşacak müşterilerin ön görülmesi, Sigorta işlemleri üzerinden sahtecilik tespitlerinde, Şüpheli müşteri yapısının belirlenmesi vb. süreçlerde veri madenciliği uygulanmaktadır.

1.6 Veri Madenciliği Süreçleri

Birçok işletme, büyük verinin farkındalığını kullanmak istiyorlar ise veri yoluyla veriyi toplama, ilgili veriyi işleme çeşitli entegrasyon ve dönüşüm ile resmi politikalar ve uygulamalarla desteklenen bir sistemi analiz etmeye ihtiyacı vardır. Genelde veri madenciliği sistemi gerçek satışlardan ve finansal veritabanlarından, sosyal medyadan , tedarikçi kayıtlarından veri tabanlarına ve daha fazlasına kadar çeşitli kaynakları kullanarak verileri süzer ve toplar.

Veri madenciliği süreçleri ile ilgili genellikle uygulanan model CRISP-DM (Cross Industry Standard Process for Data Mining) modeldir. süreç modeli olarak CRISP-DM, SPSS, NCR ve Daimler Chrysler AG benzeri büyük veri madenciliği tedarikçilerinden ve kullanıcılarından oluşan bir şirketler birliği tarafından geliştirilmiştir[11].



Şekil 1.2: Veri Madenciliği Süreci

1.6.1 İş ve problemi tanımlama

Veri madenciliğinin ilk aşaması, sürecin iş hedeflerinizi nasıl destekleyeceğini tanımlaması gerekmektedir. Hem işletme hem de veri madenciliği hedeflerine ulaşmak için yeni bir veri madenciliği planı oluşturulmalı ve plan mümkün olduğunca ayrıntılı hale getirilmelidir. Sonuç olarak bu adımda, işletmelerin hedefleri belirlenmiş ve hedefe ulaşmada yardımcı olacak önemli faktörler keşfedilmiştir.

1.6.2 Verinin anlama ve tanımlama

Verinin anlama ve Tanımlama aşamasında, veri madenciliği için mevcut olan verileri daha yakından incelemek gerekmektedir. Bu adım, tipik olarak bir proje çalışmasının en uzun kısmı ve sonraki aşamada veri hazırlığı sırasında beklenmedik sorunların önlenmesinde için büyük önem taşımaktadır. Verilerinizi anlamak, hangi veri madenciliği stratejilerinin istediğiniz bilgileri üreteceğini belirlemenize de yardımcı olmaktadır.

1.6.3 Verinin hazırlanması

Veri hazırlama aşamasında, verilerinizi analize hazırlamak için ETL (çıkarma, dönüştürme, yükleme) stratejileri kullanılmaktadır. İlk ham verilerden nihai veri kümesini oluşturmak için gereken tüm etkinlikleri içermektedir.

Farklı kaynaklardan elde edilen, benzer deęişkene ait veri türlerinde ve alan adlarında çakışma olması durumunda gerekli düzenlemelere gidilmesi bütün verileri birlikte tutabilecek model oluşturulması gerekmektedir. Deęişik modellerin incelemesinde ihtiyaçlarını dikkate almak durumundan farklı dönüştürmelere gitmek’de veri hazırlama sürecinde dikkat edilmesi zorunlu hususlar arasındadır. Örnekleme yaptığımızda farklı deęişkenlerdeki tutarlar çok yüksek olduğunda, bu tutarları normalizasyon ederek, uzaklıklar ile birlikte çalışan grup algoritmalarının tanıma ve öğrenme sürecini çabuklaştırarak modelin oluşturulma adımları için basitlik sağlanmalıdır[12].

1.6.4 Modelleme

Modelleme sürecinde İlk olarak, hazırlanan elimizdeki veri seti için kullanmamız gereken modelleme tekniklerini seçmeliyiz. Daha sonra, modelin kalitesini ve geçerliliğini doğrulamak için bir test senaryosu oluşturulması gerekmektedir.

Bu işlemden sonra modelleme araçlarını kullanarak veri kümesinde bir veya daha fazla model hazırlanması sürecini gerçekleştirmek gerekir.

1.6.5 Deęerlendirme & uygulama

İş ve Problemi Tanımlama için uygun bir modelin elde edilmesi ve bulunan modelleme tekniklerinin kurulması kısmıdır. Veri madenciliğinde bulduğunuz veri referanslarından en iyi şekilde verim alabilmek adına bu kısım önem arz etmektedir. İyi bir model iyi kurulum ve analiz sonucunda el edilen sonuçların kalitesini belirleyecektir. Bir modelin uygun olup olmadığını test edilmesi için kullanılan basit hızlı yöntem geçerlik yada geçerlilik test aşamasıdır. Bu durumda öncelikli tüm verileri %5 ila %33 tutarları arasındaki verilerin bir kısım test aşaması için ayrılır ve bu ayrılan kısım model üzerinde denir. Model üzerinde farazi yaparak ortaya çıkan bulgular yorumlanarak başarılı sonuç veren teknikler üzerinde seçim yapılır[13].

Yalnız modelin seçimi ile deęil, model ile beraber kullanılacak tekniklerde neyin uygun olduğuna karar verilmeye çalışılır[14]. Kurulan Modelde doğruluk ve uygunluk derecesinin yüksek olması , gerçek yaşamda doğru olarak modellendiğini garanti etmek olanağı yoktur[15].

1.7 Veri Madenciliği Yöntemleri

Veri madenciliğinde yöntemler makine öğrenmesi açısından, tahmin edici ve tanımlayıcı yöntemler olarak iki kısma ayrılmaktadır. Tahmin edici yöntemlerde sonuç verileri bilinenlerden ilerleyerek bir yöntem geliştirilmesi , elde edilen bu yöntem ile sonuçları bilinmeyen veriler için tahmin edilerek elde edilmesi amaçlanmaktadır.

Tanımlayıcı modellerde ifade edilmek istenen; karar vermeye kılavuzluk etme’de uygulanabilecek eldeki veriler’deki ilişkilerin tanımlanması sağlanmaktadır[16].



Şekil 1.3 : Veri Madenciliği Yöntemleri

Tahmin edici öğrenme yöntemi ile daha önceden bilinen eğitim verileri, bu verilere ait çıktı verileri ile birlikte sisteme dahil edilerek , makinenin tek başına tümevarıma gitmesi sağlanmış olur buna öğrenme süreci denilmektedir.Öğrenme süreci sonuçunda elde edilen yöntem sayesinde gelecekteki veriler için doğru sonuçları elde edecek işlemleri yapılması sağlanacaktır.

Tanımlayıcı öğrenme yöntemleri ise daha önceden belirlenmiş veya belirtilmiş bir değişken söz konusu değildir. Algoritma daha önceki verilerden gizli kalmış bir örnek,biçim veya kural çıkarmaya çalışmaktadır.

1.7.1 Tahmin edici yöntemler

Tahmin edici yöntemler daha önce eldeki verilerden faydalanılarak, meçhul bir durumu tahmin edilmeye çalışmaktadırlar. Tahmin edici öğrenme yöntemi ile daha önceden bilinen eğitim verileri, bu verilere ait çıktı verileri ile birlikte

sisteme dahil edilerek , makinenin tek başına tümevarıma gitmesi sağlanmış olur buna öğrenme süreci denilmektedir.

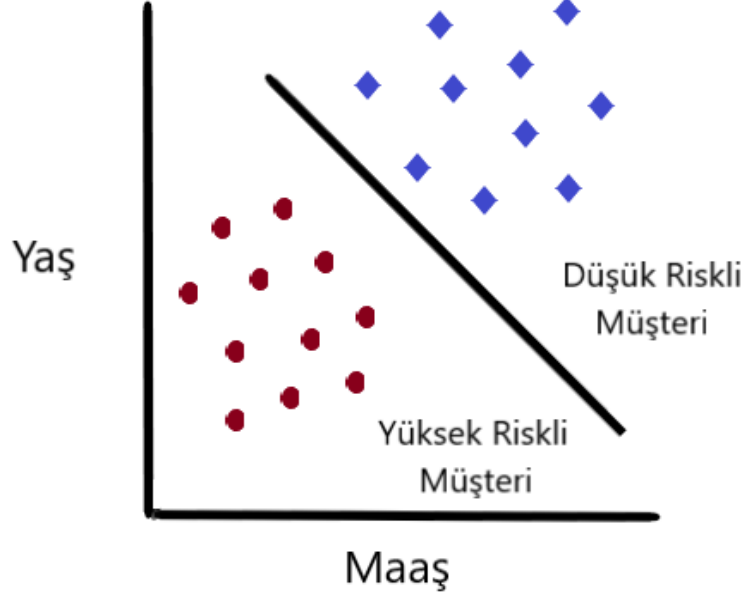
Öğrenme süreci sonucunda elde edilen yöntem sayesinde gelecekteki veriler için doğru sonuçları elde edecek işlemleri yapılması sağlanacaktır.

1.7.1.1 Sınıflandırma

Sınıflandırma, önceden belirlenmiş farklı veri sınıflarını kategorize eden bir model öğrenme sürecidir. Birinci adım olarak öğrenme adımı ve bir sınıflandırma adımından oluşan iki adımlı bir süreçtir. Öğrenme adımı, önceden tanımlanmış bir eğitim seti kullanılarak gerçekleştirilmektedir.

Veri madenciliğinde, sınıflandırma, aday modellerin bir havuzundan “sınıf” veya “kategori” ye yeni gözlemler atamak için istatistiksel modellerin eğitildiği bir görevdir; modeller, önceki örnek gözlemlerin nasıl sınıflandırıldığını gözlemleyerek yeni verileri ayırt edebilmektedir. Örneğin, bir bilgisayar ağındaki aktivite modelinin geçmiş deneyime dayalı olarak kötü amaçlı olup olmadığına karar vermek bir sınıflandırma görevidir[17].

Sınıflandırmada ele alış biçimlerinde normalde bütün nesnelerin zaten malum sınıf etiketleriyle ilişkili eğitim kümesi kullanılır. Sınıflandırma için uygulanan algoritmada eğitim kümesi öğrenir akabinde bir model oluşturulur. Oluşan model, yeni objelerin sınıflandırılmasında kullanılır. Örneğin, bir kredi politikasına başlandıktan sonra banka yöneticileri, müşterilerin kredileri karşısındaki davranışlarını analiz edebilir ve kredi alabilen müşterileri “güvenli”, “riskli” ve “çok riskli” olarak etiketleyebilir. Sınıflandırma analizi, gelecekte kredi taleplerini kabul etmek veya reddetmek için kullanılacak bir model oluşturulabilir.



Şekil 1.4: Müşteri Maaşı Bazında Risk Sınıflandırması

1.7.1.2 Regresyon

Regresyon , bir veya birden çok bağımlı ve bağımsız değişkenler arasındaki, ilişkiyi ve birlikteliği tasarlamak için kullanılır. Bağımsız değişkenler, veri madenciliğinde bilinen özelliklere sahiptir bunun ile birlikte yanıt değişkenleri tahmin edilmek istenilen değerlerden oluşmaktadır. Satış hacimleri, ürün başarısızlık ve hisse senetleri oranlarının tahmin etmek çok zordur, bakıldığında bunlar birden fazla öngörü değişkeninin karmaşık etkilenmesinin zinciri olabilir. Bundan dolayı, gelecek zamandaki değerleri öngöründe bulunmak için daha karışık teknikler (sinir ağları veya karar ağaçları) gerekebilir. Benzer model türleri genel olarak hem sınıflandırma hem de regresyon için kullanılabilir. Örnek verirsek, sınıflandırma ağaçlarını (kategorik yanıt değişkenlerini sınıflandırmak için), hem CART(Regresyon Ağaçları ve Sınıflandırma) karar ağacı algoritması bununla birlikte regresyon ağaçlarını (hep yanıt değişkenlerini öngörüsü için) oluşturmada kullanılabilir. Sinir ağları'da hem regresyon hem'de sınıflandırma'da regresyon modelleri çıkarmaktadır[17].

Regresyon ve Sınıflama yöntemleri arasındaki fark olarak tahmin edilmeye çalışılan bağımlı bir değişkenin kategorik olarak devamlılık arz eden bir değer içermelidir.[18].

1.7.2 Tanımlayıcı yöntemler

Tanımlayıcı yöntemler aslında etiketlenmemiş veri kümelerinden başlar, bu nedenle bir şekilde, doğrudan bilinmeyen özellikleri (örneğin kümeler veya kurallar) bulmakla ilgilenmektedir. Bu yöntemler veri madenciliğinde, tanımlayıcı öğrenme yöntemi sorunu, etiketlenmemiş verilerde gizli yapı bulmaya çalışmaktadır. Tanımlayıcı Öğrenme yöntemi, tahmin edici yöntemde olduğu gibi belirli bir değer belirlemek yerine ilişkileri ve örüntüleri keşfetmeyi amaçlar.

Bu öğrenme yönteminde makinenin görevi, daha önce herhangi bir veri eğitimi almadan sıralanmamış bilgileri , benzerliklere ve farklılıklara göre gruplandırmaktır.

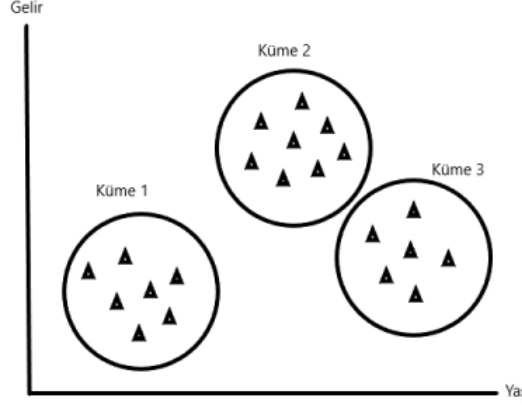
1.7.2.1 Kümeleme

Kümeleme, sınıflandırmadan farklı olarak , tanımlayıcı bir öğrenme yöntemidir.

Kümeleme sınıflandırmadan farklıdır, çünkü önceden tanımlanmış sınıfları yoktur. Kümelenmede büyük veritabanı küçük farklı alt gruplar veya kümeler biçiminde ayrılmaktadır. Kümeleme veri noktalarını benzerlik ölçüsüne göre bölümlere ayırmıştır[19]. Kümeleme yaklaşımı veri noktaları arasındaki benzerlikleri tanımlamak için kullanılır. Aynı kümedeki her veri noktası diğer kümeyle ait veri noktalarıyla karşılaştırıldığında daha büyük benzerliğe sahiptir. Son birkaç on yılda çeşitli kümeleme teknikleri oluşturulmakta ve kullanılmaktadır. Daha önce de belirtildiği gibi, kümelenmenin verileri analiz etmek için daha az bilgiye ihtiyacı vardır veya hiç bilgi gerektirmemektedir [20]. Kümeleme, belirgin bir doğal gruplaşma olmadığında kullanılır, bu durumda verilerin araştırılması zor olabilir. Verilerin kümelenmesi, daha önce bilmediğiniz grupları ve kategorileri ortaya çıkarabilir.

Bu yeni gruplar, yeni korelasyonlar bulabileceğimiz daha ileri veri madenciliği işlemleri için uygun olabilmektedir. Sınıflandırmaya benzer şekilde, kümeleme de sınıflardaki verilerin organizasyonudur. Ancak, sınıflandırmanın aksine, kümelemelerde, sınıf etiketleri bilinmemektedir ve kabul edilebilir sınıfları bulmak kümeleme algoritmasına bağlıdır. Kümeleme, denetimsiz sınıflandırma olarak da adlandırılır, çünkü sınıflandırma verilen sınıf etiketleri tarafından dikte edilmez. Hepsi aynı sınıftaki nesnelere arasındaki benzerliği en üst düzeye

çıkarma (sınıf içi benzerlik) ve farklı sınıftaki nesnelere arasındaki benzerliği en aza indirme (sınıflar arası benzerlik) ilkesine dayanan birçok kümeleme yaklaşımı vardır.



Şekil 1.5: Gelir - Yaş Dağılımı Kümelemesi

Kümeleme yöntemleri, güncel yaşamda sektörel olarak bir çok amaç için kullanımı gerçekleştirilmektedir. Sık kullanımlarından biri olan pazarlama sektöründe müşterileri segmentlerine ayırarak, farklı segmentler doğrultusunda farklı olarak pazarlama strateji ve öngörülerin geliştirildiği süreçlerdir[21].

Kümeleme yöntemi birkaç süreçten oluşan bir çözüm sürecidir. İlk adımda veri girişleri gerçekleştirilir. Doğal süreçte oluşmuş sınıflamaların ilgili olarak mutlak bilgilerin olmadığı verilerin, araştırılan değişkene ilişkili olarak çıkarılan gözetim sonuç verileri elde edilir. Sonuç olarak veri şemasının oluşması sağlanır. İleriki adımlarda uygun grup yani kümeleme teknikleri seçilerek sürece uygulanır. İlgili tekniğin uygulanması ile veriler kümelere ayrılmış olacaktır. Kümeleme sonuçlarının anlam değerlendirilmesi ve yorumlandığı adım, analizin sonuncu aşaması olarak görülür[22].

2. LİTERATÜR ÇALIŞMASI

Sağlık sektörü ivme kazandıkça ve çok çeşitli yazılım teknolojileri ve hizmetleriyle etkileşime girdiğinden, veri madenciliği ön plana çıkmaktadır. Çok sayıda sağlık şirketi, tıbbi hastane ve ilaç üretim birimi, mükemmel verimlilikleri nedeniyle veri madenciliği araçlarını kullanmaktadır. Veri madenciliğinde bir çok veri kümelerinin değerlendirilmesinde farklı çözümler bulunmaktadır. Karar ağaçları , zaman serileri analizi ,yapay zekâ, yapay sinir ağları, bir çok teknikler bu çözümler arasından yer almaktadır. Bu kısımda zaman serisi veri madenciliğinde yapılmış çeşitli literatür araştırmalarının incelenmesine değineceğiz.

Karasu S., Sarac Z. , Altan A., Hacıoglu R.[23] Çalışmalarında; Makine öğrenmesi yöntemleri arasından destek vektör makinesi ve linear regresyon, zaman serisi kullanılarak 2012 ile 2018 yılları arasındaki bitcoin günlük olarak kapanış değerlerinin tahmini yapılmaktadır. Minimum hata içeren öngörü modeli için linear fonksiyon ve polinom çekirdek(fonksiyon) kapsayan DVM yöntemi gibi farklı parametrik kombinasyonlarla süreç test ediliyor ve sonuçlar çıkarılıyor. Farklı değerlere sahip pencere değerleri için bitcoin fiyat öngörüsü, değişik ağırlık kat sayılarına sahip elemeler kullanılarak gerçekleştirilmektedir. Veri kümesinden bağımsız olarak başarısı yüksek bir model gerçekleştirilebilmesi için eğitim sürecinde 10 kat çapraz validasyon yöntemi kullanılmaktadır. Süreç sonunda modelin başarısı, istatistiksel olarak gösterge sonuçlarından Ortalama Hata Kareleri Karekökü , Ortalama Karesel Hata , Pearson korelasyon , Ortalama Mutlak Hata ile ölçülmektedir. Öngörülen DVM modeli'nin doğrusal regresyon modeline bakarak bitcoin veri kümesi için tahmin sürecinde başarımın yüksek olduğu görülmüştür.

Zeynep Behrin Güven, Turgay Tugay Bilgin[24] Çalışmalarında; Zaman serileri madenciliği ile geçmişte malum bir zaman serisi baz alınarak serinin yeni değeri bilinmeyen elemanların en yakın olasılıklı bir sonuçta tahmin etmeye çalışmaktadır. Bu kullanılan yöntem olarak veri madenciliğinde genellikle tercih

edilen istatikselsel olarak analiz tekniklerinden birisi olarak görülmektedir. Bu çalışma kullanılan WEKA programı kullanılarak Türkiye İstatistik Kurumun'dan elde edilen veri kümesi üzerinde zaman serileri madenciliğindeki kullanılan algoritmalar uygulanmıştır. 2001 ile 2010 yıllarındaki nüfus veri sonuçları kullanımı ve ileriki yılların nüfus öngörüsü yapılmıştır.

Hande Nasuhoğlu[25] Çalışmalarında; İstanbul'da bir eczaneden elde edilen 2015 ile 2018 yılları arasında gerçekleşen satış verileri alınarak 100 Adet ilacın tahmin öngörüsü yapılmıştır. İlgili çalışma ile zaman serileri analizi ve yapay sinir ağı, üssel düzeltme , hareketli ortalama, Holt-Winters ve ikili üssel düzeltme yöntemleri ile talep tahmini uygulanmıştır. Gerçek değerler ve sonuçlar arasındaki çıkan hata değerleri yöntem olarak MSE ile hesaplama yapılmıştır. Her ilaç üzerinde en düşük olarak hata değerini elde edilen yöntem seçimi uygulanmıştır. Seçim sonuçlarında 14 ilaç için hareketli ortalama, 12 ilaç üzerinde ikili üssel düzeltme ,16 ilaç üzerinde üssel düzeltme, 14 ilaç üzerinde Holt-Winters , 44 ilaç üzerinde yapay sinir ağları en iyi öngörü sonuçlarını vermiştir. Uygulamadaki ilaçların genelinde yapay sinir ağları ile çok iyi sonuçlar elde edildiği görülmüştür.

Gianluca BontempiSouhaib Ben TaiebYann-Aël Le Borgne[26] Çalışmalarında; Çok sayıda tarihsel verinin kullanılabilirliğinin artması ve çeşitli bilimsel ve uygulamalı alanlarda gelecekteki davranışların doğru tahmininin yapılması ihtiyacı, geçmiş ve gelecek arasındaki stokastik bağımlılığı gözlemleyebilen sağlam ve verimli tekniklerin tanımlanmasını gerekli olduğunu ifade etmektedir. Tahmin alanı 1960'lardan itibaren ARIMA modelleri gibi doğrusal istatistiksel yöntemlerden etkilenmiştir. Son zamanlarda, makine öğrenme modelleri dikkat çekici olarak ve kendilerini tahmin topluluğundaki klasik istatistiksel modellere ciddi rakipler olarak görmektedir. Bu çalışmada, üç konuya odaklanarak zaman serisi tahmininde makine öğrenimi tekniklerine genel bir bakış sunmaktadır. Tek adımlı tahmin problemlerinin denetimli öğrenme görevleri olarak resmileştirilmesi, yerel öğrenme tekniklerinin geçici verilerle başa çıkmada etkili bir araç olarak tartışılması ve rol bir adımdan çok adımlı öngörmeye geçtiğimizde tahmin stratejisinin ön görülmesini ifade etmektedirler.

Nicholas I. Sapankevych; Ravi Sankar[27] Çalışmalarında; Zaman serisi tahmin teknikleri, finansal piyasa tahmini, elektrikli elektrik yükü tahmini, hava ve

çevresel durum tahmini ve güvenilirlik tahmini gibi birçok gerçek uygulamada kullanılmıştır. Altta yatan sistem modelleri ve zaman serisi veri oluşturma süreçleri genellikle bu uygulamalar için karmaşıktır ve bu sistemlerin modelleri genellikle a priori olarak bilinmemektedir. Bu sistemler tarafından üretilen zaman serisi verilerinin doğru ve tarafsız tahmini her zaman iyi bilinen lineer teknikler kullanılarak elde edilemez ve bu nedenle tahmin süreci daha gelişmiş zaman serisi tahmin algoritmaları gerektirir.

Bu çalışma, yeni bir makine öğrenimi yaklaşımı kullanarak zaman serisi tahmin uygulamalarının bir araştırmasını sunmaktadır: destek vektör makineleri SVM'lerin kullanılmasının altında yatan motivasyon, bu metodolojinin, temeldeki sistem süreçleri tipik olarak doğrusal olmayan, durağan olmayan ve tanımlanmamış a-priori olduğunda zaman serisi verilerini doğru bir şekilde tahmin etme yeteneğidir. SVM'lerin ayrıca çok katmanlı algılayıcılar gibi nöral ağ tabanlı doğrusal olmayan tahmin teknikleri de dahil olmak üzere diğer doğrusal olmayan tekniklerden daha iyi performans gösterdiği kanıtlanmıştır. Nihai hedef, okuyucuya zaman serisi tahmini için SVM kullanan uygulamalar hakkında bilgi sağlamaktır. zaman serisi tahmini için SVM'ler hakkında kısa bir eğitim vermek, zaman serisi tahmini için SVM'leri kullanmanın bazı avantajlarını ve zorluklarını özetlemek ve okuyucunun kitapları, teknik dergileri ve diğer çevrimiçi SVM araştırma kaynaklarını bulması için bir kaynak sağlamak.

Mustafa Can[28] Çalışmalarında; Bilgisayar teknolojisindeki ilerlemeler ve gelişmeler ile birlikte son çeyrek zamanda büyük işler kateden zaman serileri analizi ile ileri zamanlardaki süreçlerin tahmin edilmesi işletme bazında büyük kazanç sağlamaktadır. Bunun tez çalışmasında, Tüpraş işletmesinin istanbul menkul kıymetler borsasına sunduğu üçer aylık gelir tablolarından elde edilerek hazırlanan net satışlar ve ham petrol fiyatları veri seti olarak kullanıldı. İstanbuldaki sanayi odasında yapılan her yıl araştırarak yayınladığı 500 tane büyük ölçekli sanayi işletmesi kuruluşları araştırmasında genelde birinci gelmesi, istanbul menkul kıymetler borsasında işlem görmesinden dolayı mali tablolarının halka açık ve erişimin mümkün olması ve petrol denilince akla ilk gelen Tüpraş olması seçilme nedenleri olarak ifade edilebilir.

3. MATERYAL VE YÖNTEM

İnsan yaşam sürecinde en temel gereksinim olarak sağlık alanında üretim yapan ilaç sektöründeki firmalar dünya genelinde üretim hacmi hemde ticaret kapasite bakımında ekonominin en önemli ve kritik sektörleri arasında yer almaktadır. Günümüzde gelişen teknoloji ile birlikte ilaç firmaları stoklarını yönetmeleri ve yeni ürün ve hizmetler geliştirmelerine yardımcı olmak için teknoloji giderek daha fazla kullanılmaktadır. Bu çalışmamızdaki amac, bir eczanedeki ilaç satın alınma yönelik veriler kullanılarak sonraki zamanlardaki satış miktarları tahminlenmesini sağlamaktır. Bu tahminlemeler sayesinde eczanede satın alınma ile tüketilen ilaçlar üzerinden ilaçların depolanma stok durumları da kontrol altına alınabilir. Bu çalışmamızda Türkiye'deki bir eczanenin 2015 Ocak ayı ile 2019 Aralık ayı arasındaki 5 yıllık ilaç satış verileri düzenlenmiş ve Weka programı ile zaman serileri uygulanarak haftalık olarak yapılan tahminleme çalışmalarında makine öğrenme algoritmalarından LinearRegresyon, GaussianProcess, M5Rules, MultilayerPerceptron, SMOreg, M5P, RandomFOREST kullanılmıştır. Bu algoritmaların ortalama mutlak yüzde hatası(MAPE) karşılaştırılarak en başarılı tahmin modeli bulunmaya çalışılmıştır.

Bu çalışmada kullanılacak olan tanımlar ve açıklamalar , makine öğrenme algoritmaları, gerçekleştirilecek olan ortam ve performans ölçüm sonuçları, Açık kaynaklı Weka uygulaması ile kullanılacak olan veri kümesinin elde edilmesi ve işletilmesine yer verilmiş ve zaman serisi öngörü tahmin süreci gibi konular ele alınmıştır.

3.1 Veri Madenciliği Yöntemleri

Veri madenciliğindeki yöntemler makine öğrenmesi açısından , tahmin edici ve tanımlayıcı yöntemler açısından iki kısma ayrılmaktadır. Tahmin edici yöntemlerde sonuç verileri bilinenlerden ilerleyerek bir yöntem geliştirilmesi , elde edilen bu yöntem ile sonuçları bilinmeyen veriler için tahmin edilerek elde

edilmesini amaç edinmiştir.Tanımlayıcı model olarak ise karar vermede ve rehberlik etmede kullanılan eldeki mevcut verilerden ilişkilerin tanımlanması sağlanmıştır[16].

Tahmin edici öğrenme yöntemi ile daha önceden bilinen eğitim verileri, bu verilere ait çıktı verileri ile birlikte sisteme dahil edilerek , makinenin tek başına tümevarıma gitmesi sağlanmış olur buna öğrenme süreci denilmektedir.Öğrenme süreci sonuçunda elde edilen yöntem sayesinde gelecekteki veriler için doğru sonuçları elde edecek işlemleri yapılması sağlanacaktır.Tanımlayıcı öğrenme yöntemleri ise daha önceden belirlenmiş veya belirtilmiş bir değişken söz konusu değildir. Algoritma daha önceki verilerden gizli kalmış bir örnek,biçim veya kural çıkarmaya çalışmaktadır.

3.2 Tahmin Edici Yöntemler

Tahmin edici yöntemler daha önce bilinen veriler'den faydalanılarak, bilinmeyen bir durumu tahmin edilmesine çalışmaktadırlar. Tahmin edici öğrenme yöntemi ile daha önceden bilinen eğitim verileri, bu verilere ait çıktı verileri ile birlikte sisteme dahil edilerek , makinenin tek başına tümevarıma gitmesi sağlanmış olur buna öğrenme süreci denilmektedir.

Öğrenme süreci sonuçunda elde edilen yöntem sayesinde gelecekteki veriler için doğru sonuçları elde edecek işlemleri yapılması sağlanacaktır.Veri madenciliğinde, yapılan çalışmalarda uygulama alanları ve yaşanan problemlere göre farklılık göstermektedir. Her bir süreç için, karşılaşılan problemleri çözümlenebilmek için farklı olarak algoritmalar üretilmiş ve geliştirilmiştir. Bu alandaki örneklere baktığımızda; web madenciliği, graf veya çizge madenciliği, metin madenciliği, zaman serileri madenciliği gibi alanlar örnek verilebilir.

3.3 Zaman Serileri

Zaman serileri ve Zaman serileri modellemeleri olarak ifade ettiğim değerlerin zamana bağlı olarak değişmesi sürecinde, 1900'lü yılların sonundan sonra bilim sınırlarında büyük adımlar gerçekleştirmiştir. Dünya genelindeki piyasalarda, döviz ve borsa , altın ,petrol gibi kurların gelişimi ile zaman serilerinin'de

modelleme gerçekleştirmek ve öngörü yaklaşımlar geliştirmek büyük bir problem oluşturmaktadır. Bilim adamları bu problemleri çözebilmek için bir çok araştırmalar yaparak ve yeni modeller geliştirmeye çalışmışlardır. Global dünyaya bakıldığında finansla ilişkili literatür araştırıldığında 1980'li yıllara gelindiğinde bilimsel olarak araştırma çalışmalarında artış olduğu görülmektedir. Knight,Feinstein ve Scott ve bu bir kaç çalışmalardan bir kaçını olarak ifade edebiliriz. Türkiyede istanbul m.k. borsası, günümüzdeki adıyla Borsa İstanbul olarak açılışı ile 1980'li yıllarda üçüncü çeyreğinde zaman serileri ve öngörü tahmin modelleri giderek değerini arttırmıştır.Yurtdışından yabancı sermaye'nin Türkiyeye girişi ile yapılacak olan yatırımların belirlemek ve zamanlanması için öngörü tahmini yapmak büyük önem arz etmektedir.Zaman serileri olarak ifade ettiğimiz, bir değişkenin zaman içindeki süreç ve hareketlerini gözlemleyen, elde ettiği gözlem sonuçunun zamana bağlı dağılım gösterdiği seriler olarak ifade edilmektedir. Zaman serileri aynı zamanda frekanslı seriler olup, seriye bağlı frekanslar yıllık olarak, altışar aylık, üçer aylık, aylık, haftalık ve gün olarak değişen değerler olabilmektedir.

Elde edilen her veri grubu, zaman içerisinde seri olarak değerlendirilmez. Bir veri'nin zaman içerisinde seri olarak değerlendirilmesi için veriye içerisindeki değerlerin zamana göre değişmesi gerekmektedir. Örnek olarak verirsek hava yollarındaki yolcu sayılarının devamlı olarak değişmesi zaman serisi olarak değerlendirilir. Bu değerler zamana bağlı olarak değişir. Hava yollarındaki yolcu sayısı zamana bağlı değişmeseydi, daha önceki günün değişim etkisi olmasaydı, bu veri kümesi zaman serisi özellik durumunu kaybedecekti. Diğer bir örnek ile ifade edersek; türkiyeye gelen turist sayıları olarak ifade edebiliriz. Turist sayısı hem mevsimsel hemde zaman serisine bağlıdır olarak değişim göstermektedir[29].

Zaman serilerini dört maddeden oluşmaktadır. **Trend bileşeni, Mevsimsel etkisi, Düzensiz etki ve Konjonktürel etkiler olarak ifade edilmektedir.**

Trend bileşeni

Genellikle borsalarda çok tercih edilen trend bileşeni, finansal olarak piyasaların devam ettiği uzun vade'deki yönü olarak ifade edilmektedir. Örnekleme yaparsak, firmanın bilanço'larında yüksek satış yaparak borçluluğu

azaldığında ilgili firma'nın borsa'da yaşanan hisse fiyatları uzun süreli vade'de yükselen bir trend'e girer bunun ile birlikte her olan dip birbirlerinin üzerinde oluşur, sonuçta fiyat olarak yükselen bir trend'e girmiş olacaktır.

Mevsimsel etkisi

Zaman serileri için mevsimsellik etkisi ifade ettiğimiz, mevsimlere göre mevsimsellik etkisi , zaman serilerinin etki ve değeri'nin değişmesi olarak ifade edilir.Kimi zaman içindeki döngülerde, bazı zamanlara bağlı daha az – çok olabilmektedir. Zaman içerisindeki serilere bakıldığında tren katarları gibi biri diğerini izleyen yılların, benzeri aylarında belirtmiş olduğu aynı veya buna benzer şekildeki dalgalanmalar, mevsimsel dalgalanma olarak ifade edilmektedir[29].

Düzensiz etki(rassal)

Süreçleri belirsiz olan, geçici işlemler ile ortaya çıkan, hata olarak ifade edilebilen düzensiz etki olaylardır. Örnekleme yaptığımızda, savaş, doğal olarak gerçekleşen felaketler, ön görülmeyen hava koşulları gibi beklenmedik nedenlerin meydana getirdiği değişimlerdir. Bunlar önceden öngörü edilemezler[29].

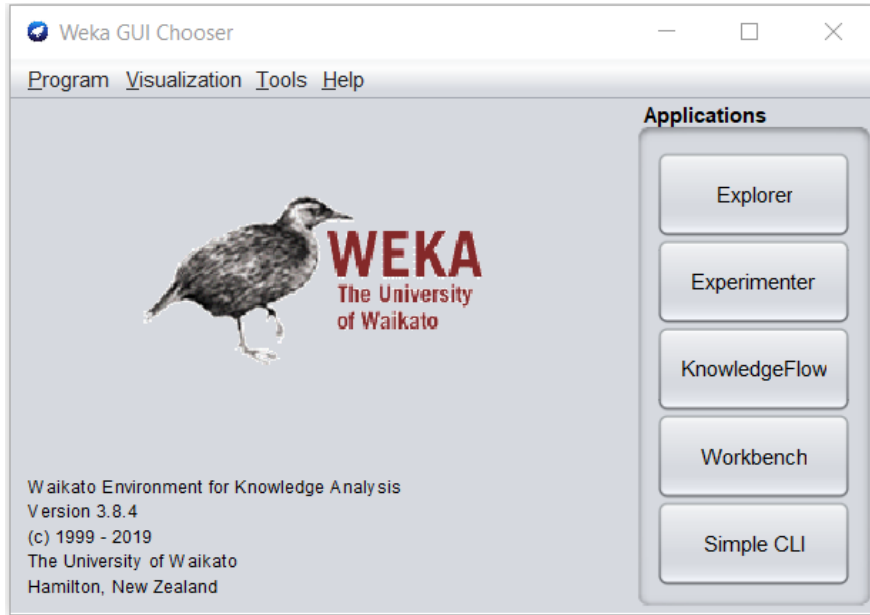
Konjonktürel etki

Zaman serilerinin trend doğrusu olarak ifade edilen veya trend eğrisi çevresindeki uzun period içindeki dalgalanmalarına bağlı konjonktürel çevresel dalgalanmalar veya sansasyon denir. İktisatta ve işletmecilikte ifade edilen yokluk, depresyon ,bolluk, durgunluk ve yükselme evreleri konjonktürel açıdan dalgalanma olarak örneklendirilir[29].

4. WEKA VERİ MADENCİLİĞİ UYGULAMASI

Weka uygulaması Amerika birleşik Devletlerindeki Waikato Üniversitesi tarafından geliştirilmiş ARFF Adında Dosya formatına sahiptir. Dosya içerisinde iki ayrı bileşen mevcuttur bunlar Header ve Data kısımları oluşturmaktadır. Dosya formatındaki Header bileşeni içerisinde veri kümesi hakkında bilgiler , Değişken Tanımlama ve Tipler bunun yanında diğer bilgiler yer almaktadır.

WEKA açık kaynak uygulaması, makine öğrenmesi, veri analizi ve veri madenciliği için ticari programlara hitaben olmayıp, daha çok bilimsel araştırma ve çalışmalar için geliştirilmiştir. Weka ağırlıklı olarak makine öğrenmesi yöntemleri üzerine Görselleştirme, sınıflandırma, kümeleme , zaman serileri öngörü tahmin veri ön işleme gibi bir çok çeşitli yöntem ve algoritmalar uygulanmaktadır. Bu çalışmamızda, Weka öngörü tahmin eklentisi olarak WEKA Versiyon 3.8.4 versiyon kullanılmıştır. Weka Uygulamasını çalıştırdığımızda, ilk olarak Şekil 3.2’de gösterilen kullanıcı arayüzü karşımıza çıkmaktadır.



Şekil 4.1: WEKA GUI Arayüzü

Grafik arayüzün, Uygulamalar “Applications” alanından, çalışma yöntemine göre 5 farklı kullanıcı arayüzü seçimi yapılabilmektedir. Ana kullanıcı arayüzü, “Explorer” sekmesidir. Biz bu çalışmamızda yapacaklarımızı Explorer sekmesi altında gerçekleştiriyor olacağız.

Explorer

Gezgin ve Araştırmacı menüleri ile veri kümelerinin yüklenmesi, ön işlenmesi, kümeleme , sınıflandırma , zaman serileri tahmin öngörü gibi pek çok işlem yapmamıza olanak sağlamaktadır. Veri açıldığı zaman tüm veri kümesini ana belleğe aldığı için küçük ve orta büyüklükteki veri kümelerinde çalışmaktadır. Büyük veri kümeleri için uygun görülmemektedir.

Yukarıdaki Şekil 3.2.1’de explorer butonuna basarak açılan ekranda preprocess tabında open file butonuna basarak ilaç Satış veri setimizi seçiyoruz. Seçtiğimiz ilaç Satış veri setine ait arff dosyasında belirttiğimiz Attribute alanları aşağıdaki gibi tahminleme yapılacak ilaç listesi görünmektedir. Her bir attribute değerine tıkladığımızda seçili attributelere ait bilgiler selected attribute alanında detaylı olarak bilgilendirilmektedir.

4.1 Weka Zaman Serisi Öngörü Modülü Uygulaması

Bu bölümde eczanemizden aldığımız İlaç satış verilerinin Weka üzerinde tahminlemesini uygulanması ve ilaç satış verilerinin Weka Arff dosya formatına dönüştürülmüş şeklini ve uygulamasını icrasını gerçekleştiriyor olacağız.

	A	B	C	D	E
1	İşlem No	Tarih	İşlem Tipi	Ürün Adı	Satış
2	11	02/01/2015	P. SATIŞ	MUSCORIL 4 MG 2 ML 6 AMPUL	1
3	12	02/01/2015	P. SATIŞ	SEFSIDAL 50 MG/5ML ORAL SUSPANSİYON 100 ML	1
4	13	02/01/2015	P. SATIŞ	BENEXOL B12 30 FILM KAPLI TABLET	1
5	14	02/01/2015	P. SATIŞ	MARCAINE %0,5 ENJEKSİYONLUK COZELTI	5
6	14	02/01/2015	P. SATIŞ	JETOKAIN 20 MG/ML+0,0125 MG/ML ENJEKSİYONLUK COZELTI	4
7	14	02/01/2015	P. SATIŞ	CITANEST %2 20 ML 1 FLK	5
8	15	02/01/2015	P. SATIŞ	BELOC ZOK 50 MG.20 TB.	1
9	15	02/01/2015	P. SATIŞ	OMEPRAZID 20 MG.14 KAPSUL	1
10	17	02/01/2015	P. SATIŞ	FLUIBRON PED.15 MG 150 ML SURUP	1
11	17	02/01/2015	P. SATIŞ	UREACORT YAGLI KREM	1
12	17	02/01/2015	P. SATIŞ	CONVULEX 150 MG.60 KAPSUL	1
13	17	02/01/2015	P. SATIŞ	CLARITINE 5 MG 200 ML SURUP	1
14	17	02/01/2015	P. SATIŞ	OCULOTECT FLUID GOZ DAMLASI	2
15	17	02/01/2015	P. SATIŞ	AETHOXYSKLEROL %1 30 ML.FLK.	1
16	17	02/01/2015	P. SATIŞ	CARTEOL LP %1 GOZ DAMLASI	2
17	18	02/01/2015	P. SATIŞ	ARVELES 25 MG.20 FILM TABLET	1

Şekil 4.2: X Eczanesinin 2015-2019 Yılı Satış İlaç Verileri

Yukarıda 2015-2019 Yılına ait satış verilerindeki ürün adı bölümündeki ilaç isimlerinde yer alan bozuk veya olmaması gereken karakterlerin düzenlenmesini veya temizliğini yapmak için ilgili satış verilerini oracle database'indeki Satış verileri tablosuna aktarıyoruz.

Weka uygulaması Amerika birleşik Devletlerindeki Waikato Üniversitesi tarafından geliştirilmiş ARFF Adında Dosya formatına sahiptir. Dosya içerisinde iki ayrı bileşen mevcuttur bunlar header ve data kısımları oluşturmaktadır. Dosya formatındaki Header bileşeni içerisinde veri kümesi hakkında bilgiler , Değişken Tanımlama ve Tipler bunun yanında diğer bilgiler yer almaktadır.

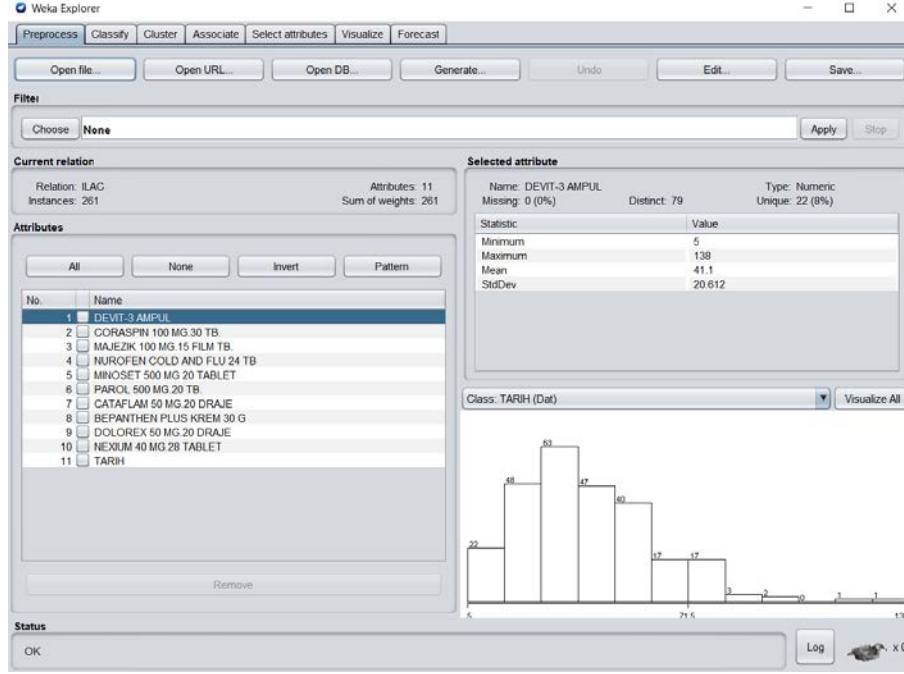
Aşağıdaki şekilde belirtilen oracle database'indeki satış verileri tablosundaki satış verilerini apriori algoritmasını uygulayacağımız weka uygulamasının işleyebileceği arff dosya formatı yapısına dönüştürülmüş hali yer almaktadır.

```
1 @relation ILAC
2
3 @attribute 'DEVIT-3 AMPUL' numeric
4 @attribute 'CORASPIN 100 MG.30 TB.' numeric
5 @attribute 'MAJEZIK 100 MG.15 FILM TB.' numeric
6 @attribute 'NUROFEN COLD AND FLU 24 TB' numeric
7 @attribute 'MINOSET 500 MG 20 TABLET' numeric
8 @attribute 'PAROL 500 MG.20 TB.' numeric
9 @attribute 'CATAFLAM 50 MG.20 DRAJE' numeric
10 @attribute 'BEPANTHEN PLUS KREM 30 G' numeric
11 @attribute 'DOLOREX 50 MG.20 DRAJE' numeric
12 @attribute 'NEXIUM 40 MG.28 TABLET' numeric
13 @attribute TARIH date 'dd/MM/yyyy'
14
15 @data
16
17 12,11,14,11,10,1,5,8,7,3|07/01/2015
18 6,19,16,20,8,7,12,20,12,4,14/01/2015
19 20,25,27,26,21,8,12,12,5,8,21/01/2015
20 19,19,24,17,16,11,15,19,23,10,28/01/2015
21 20,12,38,22,17,26,14,21,7,14,04/02/2015
22 13,10,24,10,12,9,12,17,10,3,11/02/2015
23 15,13,31,20,14,13,6,12,11,9,18/02/2015
```

Şekil 4.3: WEKA Arff Dosya Formatı

Yukarıda üzerinde makine öğrenmesi zaman serileri algoritmalarını uygulayacağımız arff dosyasında yer alan bileşenlere baktığımızda @attribute ile ifade edilen daha öncede belirttiğimiz X eczanesinin 2015-2019 yılına ait satış verilerinde en çok satış yapılan ilaç listesindeki ilaç adlarını tanımlamaya çalıştık.

İlgili veri seti tarih başlangıç olarak 01.01.2015 – bitiş tarihi olarak 31.12.2019 tarihleri arasında 5 yıllık veriyi içermektedir. Bu veriler haftalık toplam satış miktarları dikkate alarak oluşturulmuştur.

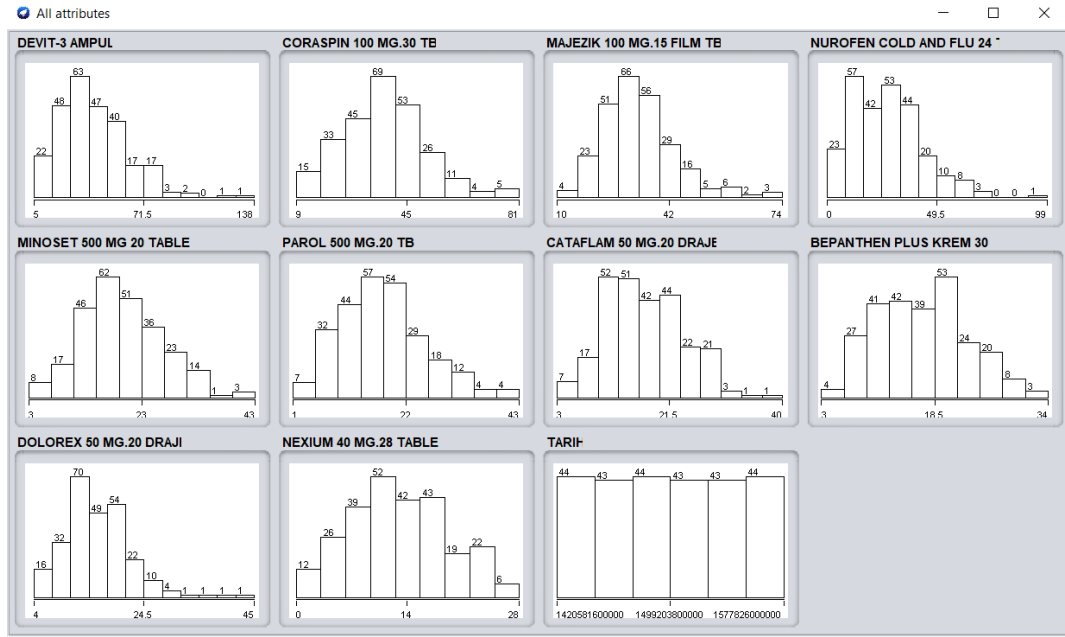


Şekil 4.4: WEKA Explorer Uygulama Alanı

Arff dosya formatının tahminleme yapılacak ilaçların sütun ve haftalık olarak satır değerlerinin gösterimi.

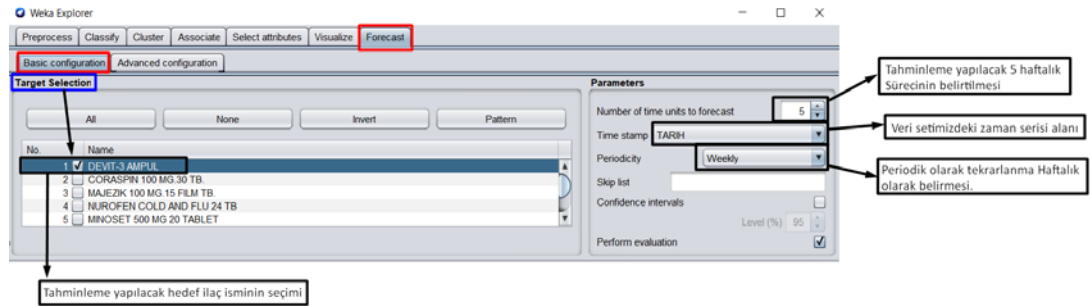
No	1: DEVIT-3 AMPUL	2: CORASPIN 100 MG.30 TB.	3: MAJEZIK 100 MG.15 FILM TB.	4: NUROFEN COLD AND FLU 24 TB	5: MINOSET 500 MG 20 TABLET	6: PAROL 500 MG 20 TB.
1	12.0	11.0	14.0	11.0	10.0	1.0
2	6.0	19.0	18.0	20.0	8.0	7.0
3	20.0	25.0	27.0	26.0	21.0	8.0
4	19.0	19.0	24.0	17.0	16.0	11.0
5	20.0	12.0	38.0	22.0	17.0	26.0
6	13.0	10.0	24.0	10.0	12.0	9.0
7	15.0	13.0	31.0	20.0	14.0	13.0
8	22.0	27.0	19.0	15.0	8.0	13.0
9	22.0	25.0	39.0	32.0	30.0	27.0
10	29.0	19.0	23.0	25.0	14.0	22.0
11	29.0	23.0	27.0	24.0	17.0	13.0
12	57.0	34.0	26.0	23.0	14.0	16.0
13	36.0	32.0	29.0	20.0	15.0	14.0
14	26.0	19.0	21.0	13.0	16.0	11.0
15	46.0	13.0	32.0	15.0	14.0	11.0
16	22.0	21.0	30.0	14.0	20.0	11.0
17	26.0	17.0	19.0	14.0	19.0	2.0
18	23.0	22.0	23.0	10.0	13.0	6.0
19	31.0	29.0	42.0	16.0	21.0	31.0
20	42.0	15.0	35.0	5.0	8.0	7.0
21	21.0	10.0	30.0	13.0	10.0	8.0
22	26.0	23.0	38.0	17.0	13.0	12.0
23	32.0	22.0	28.0	12.0	15.0	13.0
24	18.0	10.0	33.0	7.0	11.0	6.0
25	16.0	27.0	20.0	6.0	15.0	9.0
26	16.0	20.0	26.0	5.0	11.0	6.0
27	13.0	25.0	32.0	8.0	19.0	8.0
28	34.0	11.0	21.0	9.0	7.0	8.0
29	10.0	10.0	14.0	7.0	3.0	5.0

Şekil 4.5: WEKA Explorer Visualizer Data Viewer



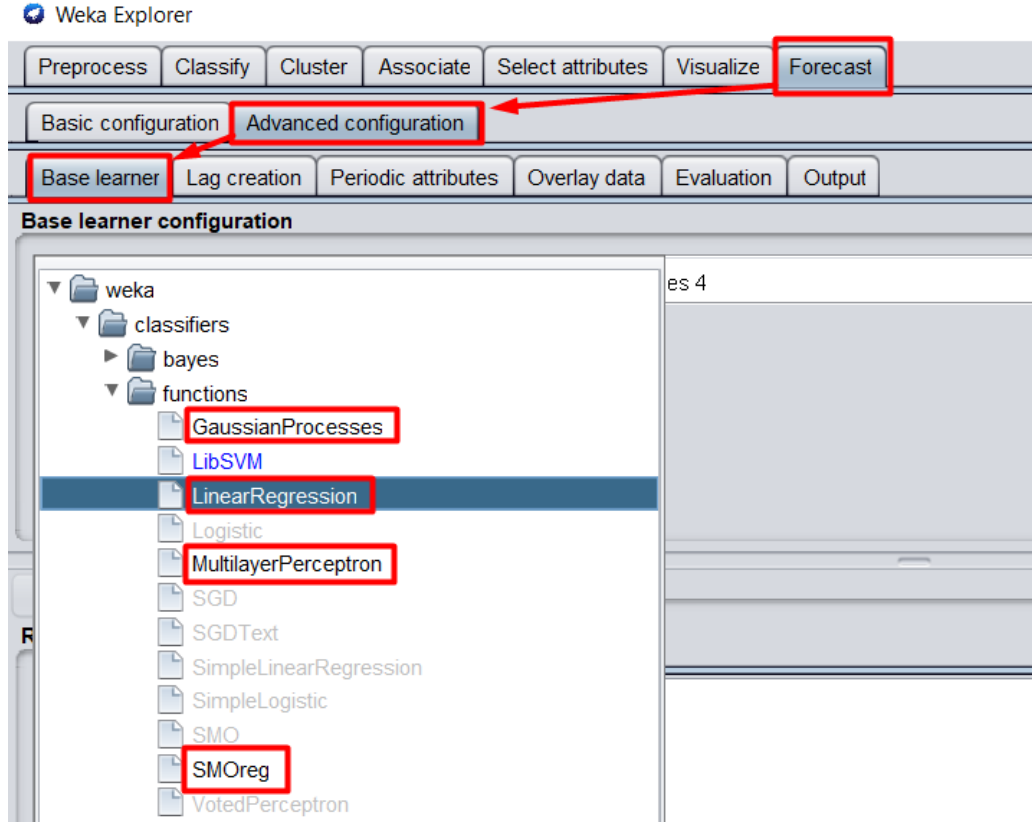
Şekil 4.6: WEKA Explorer Visualizer All Attributes Görünümü

Weka üzerinde elimizdeki satış verilerine bağlı olarak oluşturduğumuz .arff dosya yapısına bağlı olarak dosyayı seçip sonrasında forecast tabındaki tahminlemede kullanılacak parametrelerin belirlenmesi Bunun için weka üzerinde forecast tabındaki bulunan basic configuration tab alanında temel değerlerin belirtilmesini sağlıyoruz.



Şekil 4.7: WEKA Forecast Basic Target Selection Uygulama

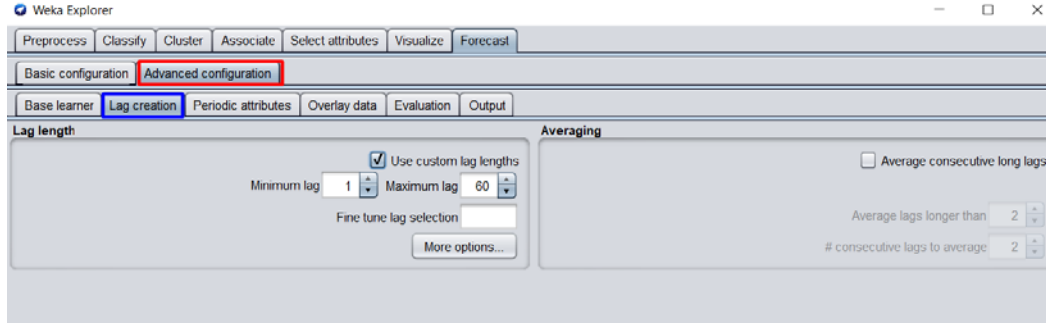
Temel konfigürasyonları belirttikten sonra İleri konfigürasyon değerleri için ilk olarak tahminlemede kullanılacak makine öğrenmesi algoritmalarından birinin seçimini yapılması sağlanıyor.Biz ilk olarak LinearRegression algoritmasını seçerek tahminleme ve performans ölçümlerini yapmaya çalışacağız.



Şekil 4.8: WEKA Forecast Advanced Base Learner Uygulama

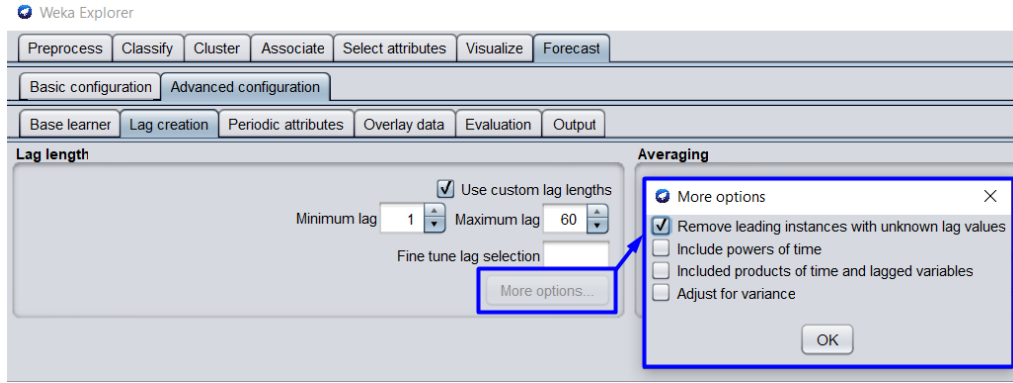
Algoritmalarından LinearRegresyon belirledikten sonra weka da ileri konfigürasyon tabındaki lag creation(geçikme oluşturma) alanında tahminlemede dikkate alması için aşağıdaki parametre ve değerleri kullanıyoruz. Bunları kullanırken daha önce period olarak belirlediğimiz haftalık değerlerin bir yılda 52 hafta üzerinden değerlendirerek en iyi sonucu alana kadar Min ve Max Lag değerine en iyi sonucu alana kadar deneyerek max lag değerini arttırarak ulaşıyoruz.

Lag Creation yani Gecikme oluşturma paneli, kullanıcının gecikmeli değişkenlerin nasıl oluşturulduğunu kontrol etmesini ve değiştirmesini sağlar. Gecikmiş değişkenler, bir serinin(Haftalık) geçmiş ve şimdiki değerleri arasındaki ilişkinin, öneri öğrenme algoritmaları tarafından yakalanabildiği ana mekanizmadır. Bu yüzden Lag Creation alanında Use custom lag Lengths işaretleyerek Min ve max Lag değerlerini belirtiyoruz.



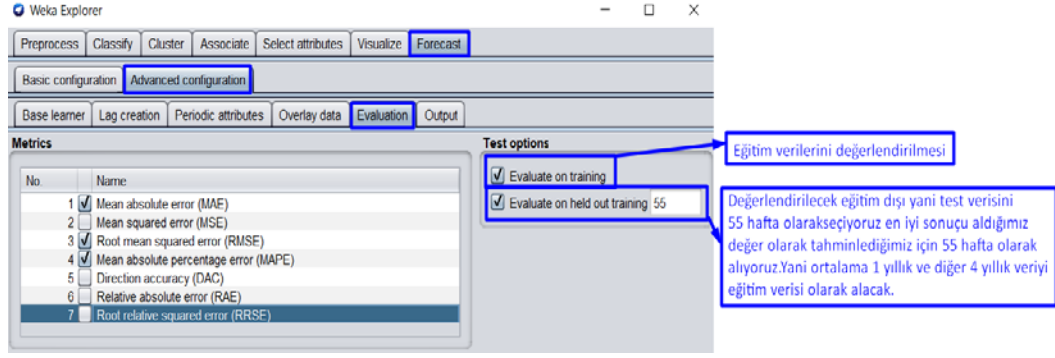
Şekil 4.9: WEKA Forecast Advanced Lag Creation Uygulama Alanı

Lag creation da belirtilen seçeneklere ek olarak More Options butonu üzerinden açılan More Options penceresinde “Remove leading instances with unknown lag values” değerini seçerek lag creation da girilecek değerleri kullandığımız LinearRegresyon algoritması için tamamlamış oluyoruz. Bu alanı işaretlememizin amacı gecikme değerleri bilinmeyen haftalık önde gelen haftalık değerlerin kaldırımını gerçekleştirmiş oluyoruz.



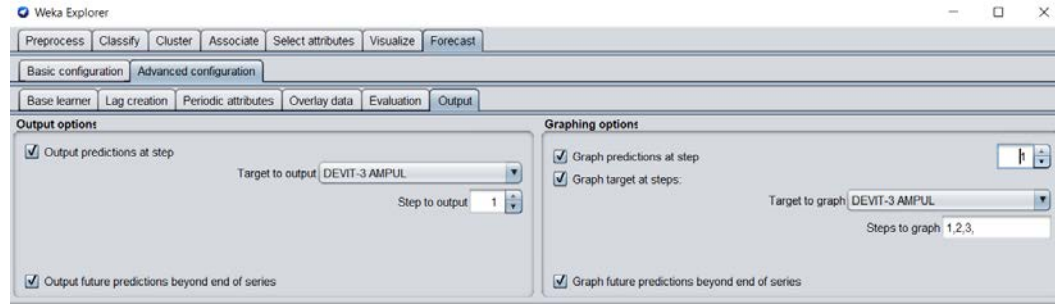
Şekil 4.10: WEKA Forecast Advanced Lag Creation More Options Alanı

Bu kısımda değerlendirme yapılacak eğitim ve test verilerimizin periodlarını belirliyoruz burda en iyi LinearRegresyondaki sonuçları aldığımız test verisi değerlendirmesini 55 hafta olarak belirtiyoruz yaklaşık olarak 1 yıl olmakla beraber geri kalan 4 yılda eğitim verisi olarak belirtmiş oluyoruz. 55 sayısı en iyi sonucu aldığımız değer olarak ifade edebiliriz.



Şekil 4.11: WEKA Forecast Advanced Evaluation Uygulama Alanı

Çıktıların Seçeneklerini belirlediğimiz Output ve Graphing seçenekleri ile Hedef ilaçımızın Grafikselsel sonuçlarını göstereceğimiz seçenekleri ifade etmektedir. Steps to graph ile 1-2-3 haftalık tahminlerin eğitim ve test verilerini grafikselsel olarak gösterimini sağlıyoruz. Aynı zaman gelecek tahminlemesini grafikselsel gösterimini graph future predictions seçeneği ile sağlıyor oluyoruz.



Şekil 4.12: WEKA Forecast Advanced Output Uygulama Alanı

4.2 Weka Uygulaması Zaman Serileri Öngörü Algoritmaları

WEKA uygulaması zaman serileri öngörü eklentisi Weka explorer ekranında Forecast tabında, tahmin öngörü algoritmaları olarak, sayısal çıktı üreten sınıflandırma algoritmalarını kullanmaktadırlar. Sınıflandırmada, sayısal çıktı oluşturmak için genel olarak regresyon analizi kullanılmaktadır.

Regresyon analizinde anlatılmaya çalışılan, verilen zaman serisini en iyi biçimde, yani en küçük hata oranı ile modelleyebilecek ve veri kümesini en iyi şekilde temsil edebilecek bir doğru veya eğri bulma sürecine denir[30].

4.2.1 Lineer regresyon algoritması

İki değişken arasındaki doğrusal ilişkinin bir doğru denklemi olarak tanımlanıp, değişkenin değerlerinden biri bilindiğinde diğeri hakkında tahmin yapılmasını

sağlar. Veriler arasında doğru tahmini yapabilmek için veriler için en iyi doğruyu oluşturmak gerekir. En iyi doğruyu oluştururken bütün noktalara en yakın bölge tercih edilmelidir. Lineer Regresyon'da bir doğru oluşturacağımız için bir bağımlı ve bir bağımsız değişken olmak üzere toplam iki değişken üzerinde çalışılır[3].

Bu model, sınıfı niteliklerin lineer bir kombinasyonu olarak açıklamaktadır. Eğitim verisinden hesaplanan ağırlıklar ile birlikte denklem

$$x = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k \quad (3.1)$$

yukarıda şekilde 3.1de x bir sınıf; a_1, a_2, \dots, a_k nitelik değerleri, $w_0, w_1, w_2, \dots, w_k$ ağırlıklardır. i. eğitim örneği için tahmin öngörü değeri, denklem (3.2)' de gösterilmiştir.

$$w_0 a_0^{(i)} + w_1 a_1^{(i)} + w_2 a_2^{(i)} + \dots + w_k a_k^{(i)} = \sum_{j=0}^k w_j a_j^{(i)} \quad (3.2)$$

a_0 daima 1 dir. Her $i > 0$ için i. eğitim örneği denklemdir. Öngörü değeri fiili değil sınıf değeridir. Lineer regresyon hata kareleri toplamı denklem (3.3)'te gösterilmiştir.

$$\sum_{i=1}^n (x^{(i)} - \sum_{j=0}^k w_j a_j^{(i)})^2 \quad (3.3)$$

Denklem (3.3)' te i. eğitim örneği fiili değer ve öngörü değeri arasındaki farkın parantez karesi olarak ifade edilir[30].

4.2.2 Gaussian process algoritması

Gauss süreçleri istatistiksel bir sınıflandırma-tahmin algoritmasıdır. Gauss süreçleri algoritması öğrenmede, lineer (doğrusal) olmayan regresyon için Bayesyen Gauss süreçleri uygular. Parametrik olmayan Bayesyen Gauss yaklaşımı; girdilerin yakınlığına göre çıktılar arasında yüksek korelasyon sağlayan düzgün ve sürekli bir f fonksiyonu, girdilerin mümkün olan tüm ihtimal dizilimini kullanarak, lineer olmayan fonksiyonların uzayı üzerine, doğrudan bir dağılım yerleştirir ve verileri fonksiyon çıktıları ile ilişkilendirir. Gauss süreçleri öncesi fonksiyonlar üzerinde son çıkarım yoluyla, denetimli öğrenme sağlar[30].

4.2.3 M5Rules algoritması

Algoritma böl ve yönet (sperate and conquer) tekniği kullanmaktadır. Sayısal öngörü için karar listesi kullanarak regresyon modeli ya da regresyon ağacı oluşturur. Regresyon ağacında ağacın yaprakları, yaprağa ulaşan örneklerin ortalamasıdır. Tek bir lineer regresyon denklemine göre daha iyi sonuç verir fakat ağaç ve boyutları büyük ve yorumlanması zahmetli olduğu için model ağaçları geliştirilmiştir. Model ağaçlarında, regresyon denklemi regresyon ağaçları ile birleştirilmiştir. Model ağaçlarının yaprakları tek tahmin edilen değer yerine lineer ifadedi regresyon denklemleridir[30].

4.2.4 Multilayer perceptron algoritması

Perceptronlar doğrusal ayrılabilir problemler için uygun iken Multilayer perceptron (çok katmanlı algılayıcı) doğrusal türde olmayan sınıflandırmalar için de uygundur. MLP , XOR Problemi'ni sonuçlandırmak için yapılan işlemler sonucu oluşmuştur. Gerçek hayatta bir çok problem doğrusal olmayan yapıdadır. MLP ağırları linear olmayan problemlerin çözümü için çok sık kullanılan YSA modelidir.

MLP ağırlarına bağlı en iyi öğrenim yöntemi geriye yayılım(Back Propagation) yöntemidir.Geriye yayılım, 1974 yılında ilk kez Werbos tarafından tavsiye edilmiştir, günümüzde kullanılan yapı 1986 yılında William, Rumelhart ve Hinton üçlüsü geliştirmiştir[31].

4.2.5 SMOreg algoritması

SMOreg algoritmasında, dvm üzerinde çalışarak regresyonu gerçekleştirmektedir. Bu algoritmanın baz olarak dvm olarak tanınan yöntemleri uygular. Dvm'ler yıl olarak 1960 sonlarında Alexey Chervonenkis ve Vladimir Vapnik sayesinde geliştirilmiş baz olarak istatistiksel öğrenme modeline dayanan makine öğrenmesidir. Dvm'nin baz mantığı linear şekilde ayrıştırılan veri süreçleri için çok iyi ayırıcı ortamın belirlenmesidir.

DVM'ler, eğitim sırasında raslanmamış yeni verileri'de eksiksiz olarak sınırlara ayırabilmektedir. Bu durum DVM'leri yaygınlaştırabilme yetisini göstermektedir[32].

4.2.6 M5P algoritması

M5P, karar ağaçları yönteminden türeyen bir metottur. Birçok karar ağacının oluşturulması ve sonuçlarının birleştirilmesi ile sınıflandırma, regresyon gibi işlemler için kullanılmaktadır. Karar ağaçları tahminde bulunma ihtiyaçlarında yoğunlukla kullanılan kullanışlı ve basit bir metottur. Karar ağacı algoritmalarında kesin sonuç ifade edilir ve kuralın oluşturulma şekli açıktır. Ayrıca diğer metotlara üstünlüğü karar ağaçlarının daha anlaşılır yapıda olmasıdır[33].

4.2.7 Random forest algoritması

Random Forest(Rassal orman), yüksek parametre tahmini yapılmadan'da iyi sonuçlar almayı hem regresyonda hemde sınıflandırma problemleri ile de uygulanabilirliği olması açısından, bilinen makine öğrenmesi modelleri arasında gösterilmektedir. Random Forest modelin'de değişik veri kümelerinde eğitim yapıldığı için varyans, diğer bir ifadeyle karar ağaçları'nın en büyük sorunlarından olan overfitting azalmaktadır[9]. Rassal orman algoritması, Karar ağacı gibi hem Sınıflandırma hemde Regresyon problemlerinde tercih edilebilir. Çalışma sürecinde birden çok karar ağacı oluşturmaktadır. Sonuç üreteceğinde bu karar ağaçların'daki ortalama değerleri alınır ve sonuç çıkarılır[34].

4.3 Tahmin Yöntemlerinin Hata Ölçüm Teknikleri

Tahmin modellerinin doğruluğunun denetlenmesi ve ölçülmesi hata teknikleri oldukça önem arz etmektedir.Tahmin Öngörü tekniklerinin doğruluğunu ölçülmesini sağlayan hata ölçümü, tahminleme ile gerçekleşen sonuçlar arasındaki farkı ölçmektedirler.

Bu tez çalışmamızda kullanılan performans değerlendirme hata ölçüm yöntemleri MAE, MAPE , RMSE denklemleri aşağıda açıklanmıştır.

4.3.1 Ortalama mutlak hata

Ortalama Mutlak Hata, Mean Absolute Error(MAE), yönleri dikkate alınmaksızın, bir dizi öngörüsündeki hataların ortalama olarak büyüklüklerini ölçmektedir. Tüm bireysel farklılıkların eşit ağırlığa sahip olduğu tahmin ile

gerçek gözlem arasındaki mutlak farklılıkların test örneği üzerindeki ortalamasıdır. Ortama mutlak hata aşağıdaki formül ile gösterilmektedir[35].

$$mae = \frac{\sum_{i=1}^n abs(y_i - \lambda(x_i))}{n}$$

4.3.2 Ortalama mutlak yüzde hatası

Ortalama Mutlak Yüzde Hatası, Mean Absolute Percentage Error(MAPE) tahmin doğruluğunu kontrol etmek için en yaygın kullanılan ölçüdür. Ölçekten bağımsız olan ve farklı ölçeklerdeki serileri karşılaştırmak için kullanılabilen yüzde hatalarının altında gelir.

$$MAPE = mean(|e_i|/y_i) * 100$$

Yukarıdaki formülde e_i , hata terimidir ve y_i , i zamanındaki gerçek verilerdir. MAPE'nin dezavantajı, verilerdeki herhangi bir gözlem için gerçek şeklin değeri 0 ise sonsuzluğa eğilim göstermesi veya tanımsız hale gelmesidir[36].

4.3.3 Kök ortalama kare hatası

Kök Ortalama Kare Hata (RMSE), tüm hatanın karesinin ortalamasının köküdür. RMSE kullanımı çok yaygındır ve sayısal tahminler için mükemmel bir genel amaçlı hata metriği olarak kabul edilir.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2}$$

Yukarıdaki formülde O_i gözlemler, S_i bir değişkenin değerlerini tahmin etti ve n analiz için mevcut gözlem sayısı. RMSE, doğruluk için iyi bir ölçüttür, ancak yalnızca ölçeğe bağlı olduğu için değişkenler arasında değil, belirli bir değişken için farklı modellerin veya model yapılandırmalarının tahmin hatalarını karşılaştırmak içindir[37].

4.4 Uygulama Ortamı Donanım Özellikleri

Weka üzerinde gerçekleştirdiğimiz zaman serileri çalışmaları için Uygulama ortamı donanım özellikleri ve kullanılan işletim sistemi aşağıdaki Çizelge 3.5 verilmiştir.

Çizelge 3.1 : Uygulama Ortamı Donanım Özellikleri

İşlemci	Intel(R)-Core i7 8550U CPU @1.80 GHZ 1.99 GHz
İşletim Sistemi	Window10
Ram	16 Gb
Sistem Tipi	System 64 bit Operating, x64 based processor

4.5 Gerçekleştirilen Uygulama Çıktıları ve Bulguları

Weka üzerinde gerçekleştireceğimiz ilaç satış verilerine bağlı zaman serileri algoritmalarını uygulayacağımız arff veri formatına uygun olarak oluşturduğumuz dosyada kullanılan en çok satış verilerine sahip ilaç listesi ve ilaçları aşağıdaki şekilde belirtilmektedir.

4.5.1 Gerçekleştirilen uygulama çıktıları

Bu listede Devit-3 Ampul ilacına bağlı yaptığımız tahminleme öngörü çıktı ve grafik arayüzlerini inceleyeceğiz.Bu incelemede 5 yıllık süreçte 4 yılı eğitim verisi , 1 yılı (55 hafta) test verisi olarak incelenmiştir.

```
Relation:      ILAC
Instances:     261
Attributes:    11
              DEVIT-3 AMPUL
              CORASPIN 100 MG.30 TB.
              MAJEZIK 100 MG.15 FILM TB.
              NUROFEN COLD AND FLU 24 TB
              MINOSET 500 MG 20 TABLET
              PAROL 500 MG.20 TB.
              CATAFLAM 50 MG.20 DRAJE
              BEPANTHEN PLUS KREM 30 G
              DOLOREX 50 MG.20 DRAJE
              NEXIUM 40 MG.28 TABLET
              TARIH
```

Şekil 4.13 : Tahminleme Yapılacak İlaç Listesi

4.5.1.1 DEVIT-3 AMPUL için linear regresyon algoritması

Aşağıda linear regresyona bağlı olarak aldığımız üç haftalık tahminleme değerlendirme Ölçüm verileri yer almaktadır. Bu veriler Eğitim ve test verileri olarak sonuçları 3 haftalık olarak alınmıştır.

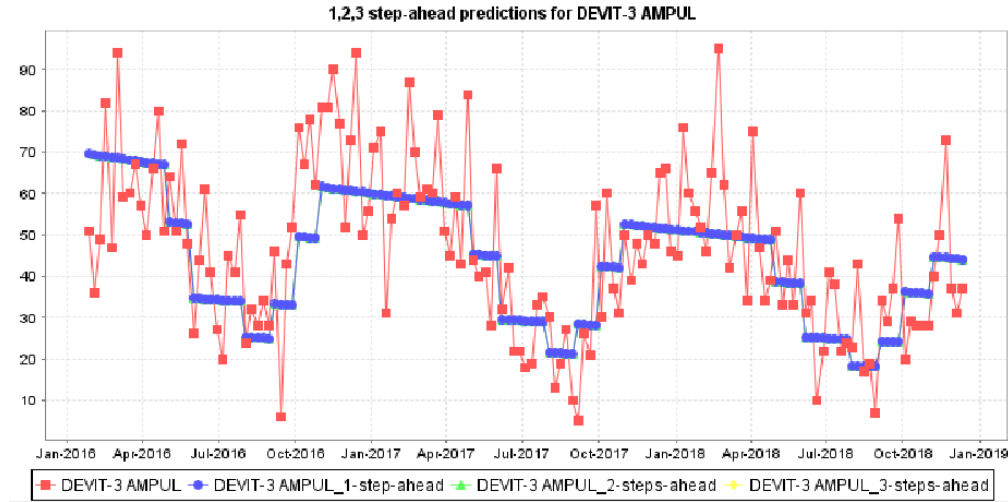
```
=== Evaluation on training data ===
Target          1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                151          150           149
Mean absolute error  11.4914      11.4441      11.2969
Mean absolute percentage error  32.56        32.5341      32.1302
Root mean squared error  14.3158     14.283       14.0676

Total number of instances: 206

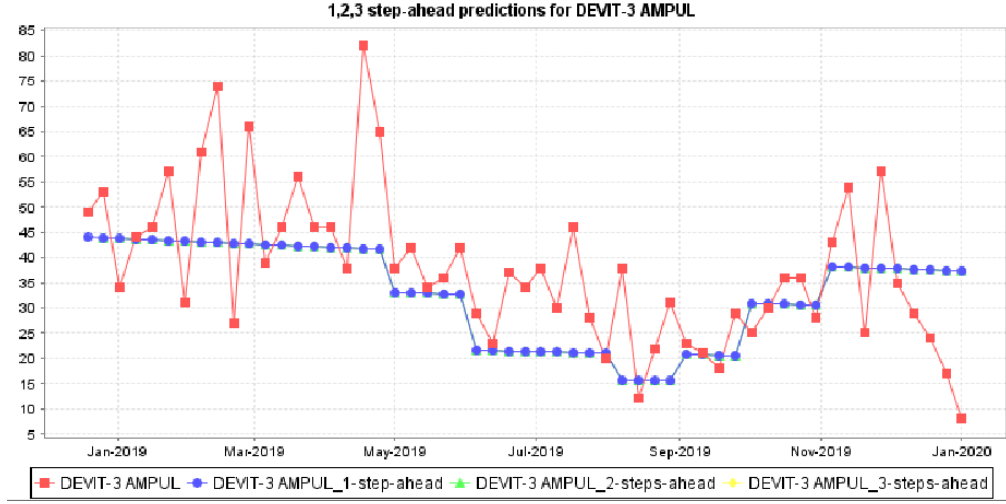
=== Evaluation on test data ===
Target          1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                55           54            53
Mean absolute error  10.3447     10.4433      10.4676
Mean absolute percentage error  32.313      32.7216      33.013
Root mean squared error  13.5266    13.6341      13.7045

Total number of instances: 55
```

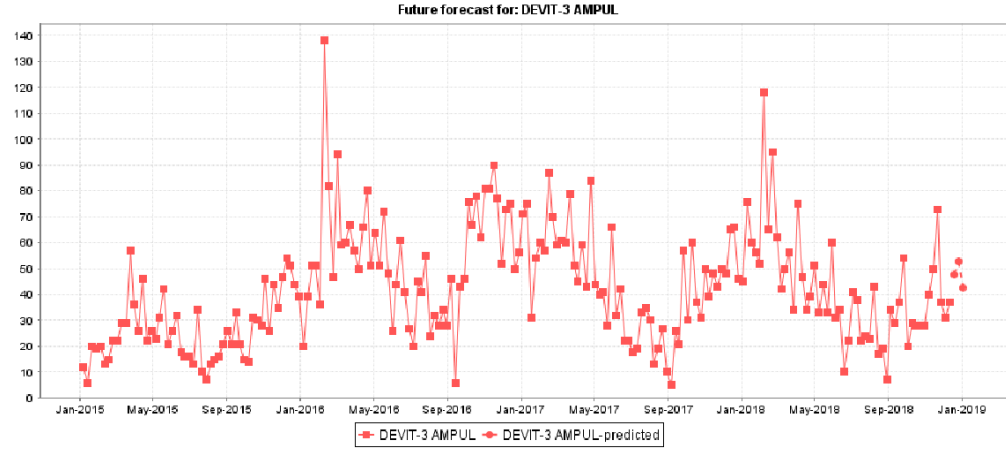
Şekil 4.14: DEVIT-3 AMPUL için Linear Regresyon Çıktıları



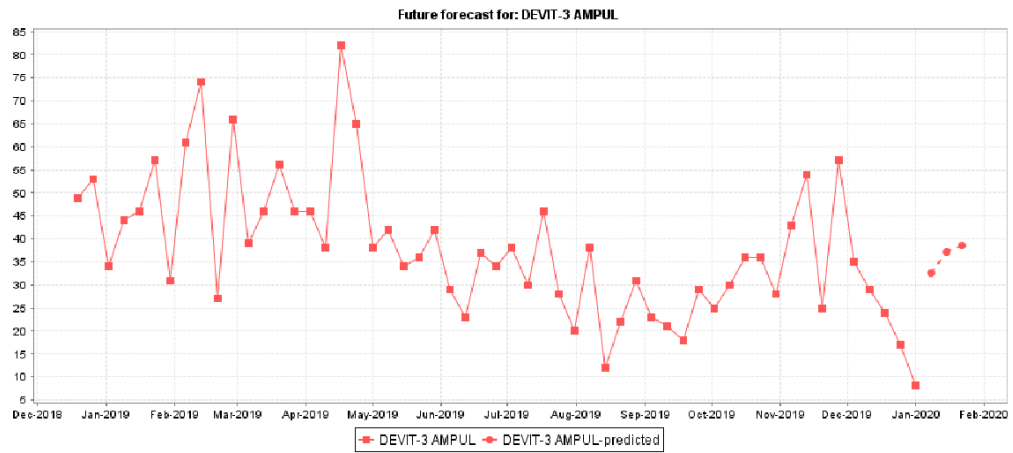
Şekil 4.15: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği



Şekil 4.16: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği



Şekil 4.17: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği



Şekil 4.18: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği

4.5.1.2 DEVIT-3 AMPUL için m5rules algoritması

Aşağıda M5Rules algoritmasına bağlı olarak aldığımız 3 haftalık Tahminleme değerlendirme Ölçüm verileri yer almaktadır. Bu veriler Eğitim ve test verileri olarak sonuçları 3 haftalık olarak alınmıştır.

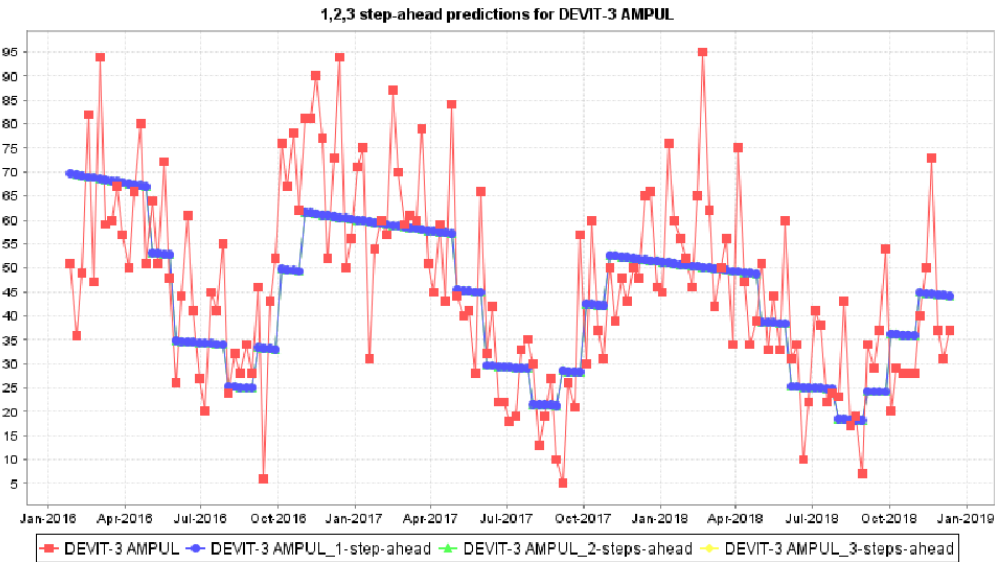
```
=== Evaluation on training data ===
Target          1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                151          150          149
Mean absolute error  11.4914     11.4441     11.2969
Mean absolute percentage error  32.56      32.5341     32.1302
Root mean squared error  14.3158    14.283      14.0676

Total number of instances: 206

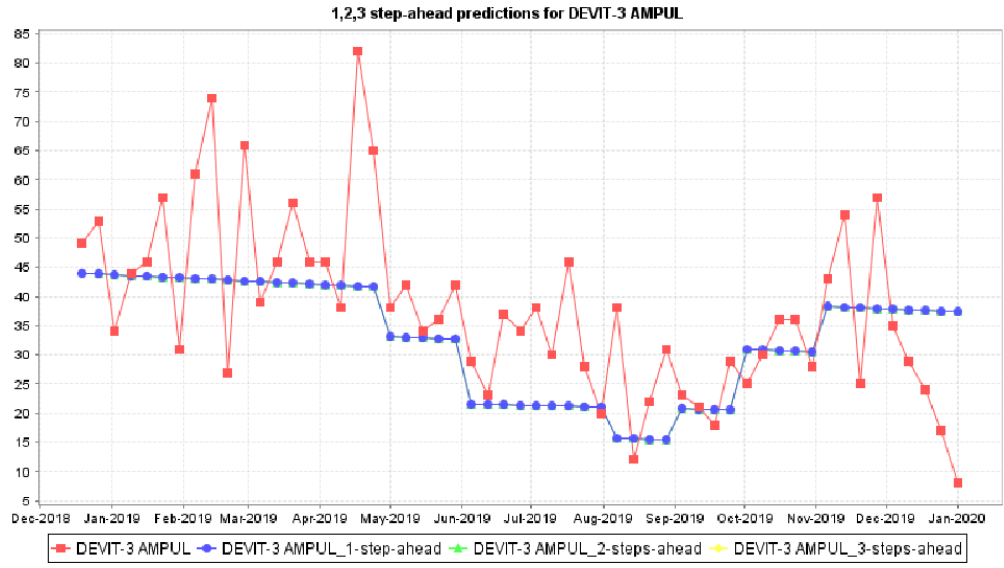
=== Evaluation on test data ===
Target          1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                55          54          53
Mean absolute error  10.3447     10.4433     10.4676
Mean absolute percentage error  32.313     32.7216     33.013
Root mean squared error  13.5266    13.6341     13.7045

Total number of instances: 55
```

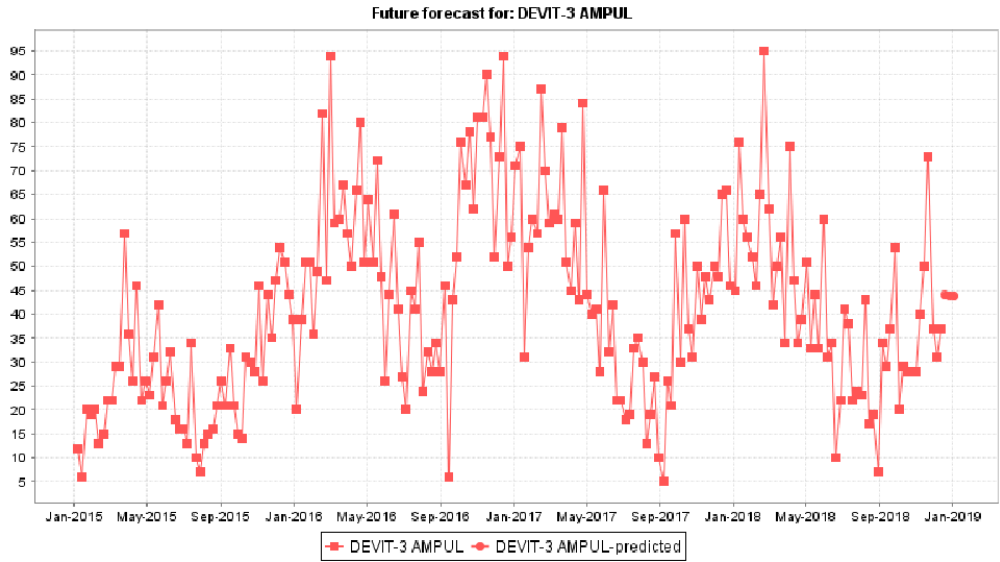
Şekil 4.19 : DEVIT-3 AMPUL için M5Rules Çıktıları



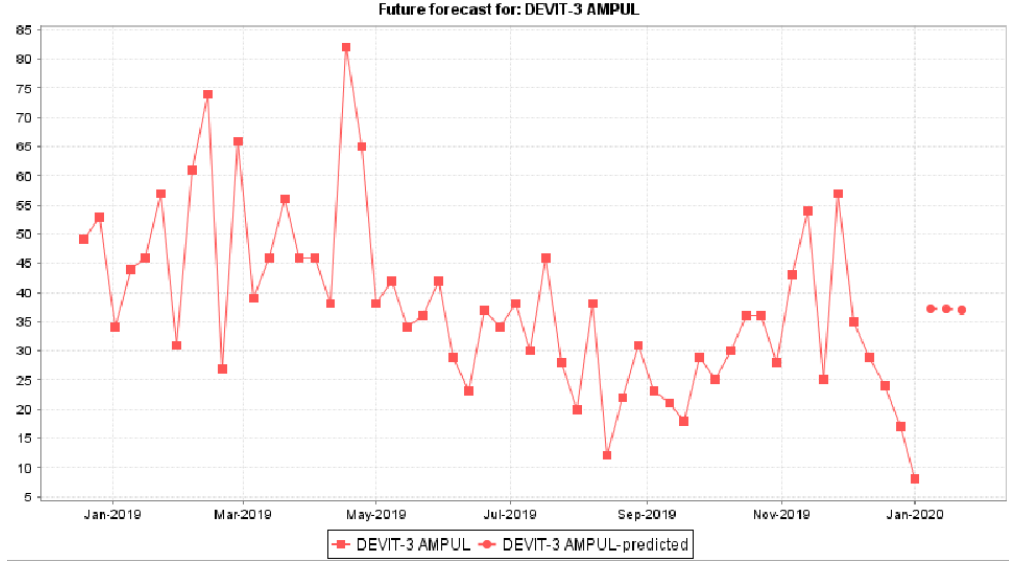
Şekil 4.20: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği



Şekil 4.21: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği



Şekil 4.22: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği



Şekil 4.23: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği

4.5.1.3 DEVIT-3 AMPUL için gaussianprocess regression algoritması

Aşağıda Gaussian process regression algoritmasına bağlı olarak aldığımız 3 haftalık tahminleme değerlendirme ölçüm verileri yer almaktadır. Bu veriler eğitim ve test verileri olarak sonuçları 3 haftalık olarak alınmıştır.

```

=== Evaluation on training data ===
Target                1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                    151           150           149
Mean absolute error  11.5085       11.4801       11.3298
Mean absolute percentage error  32.3265       32.3359       31.9215
Root mean squared error  14.3507       14.3407       14.1187

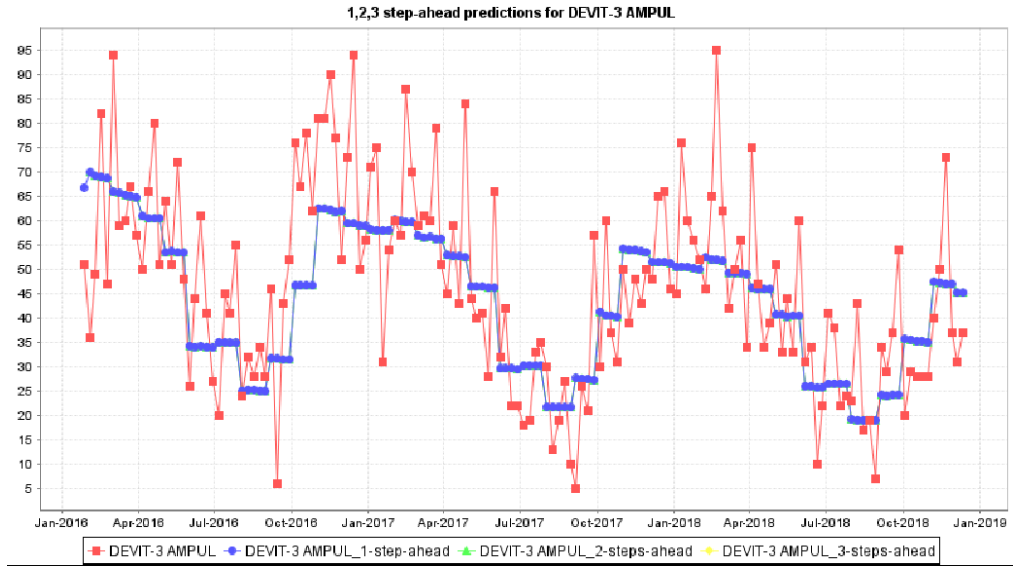
Total number of instances: 206

=== Evaluation on test data ===
Target                1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                    55           54           53
Mean absolute error  9.8825       9.9899       10.024
Mean absolute percentage error  31.947       32.3844       32.704
Root mean squared error  13.252       13.3626       13.4412

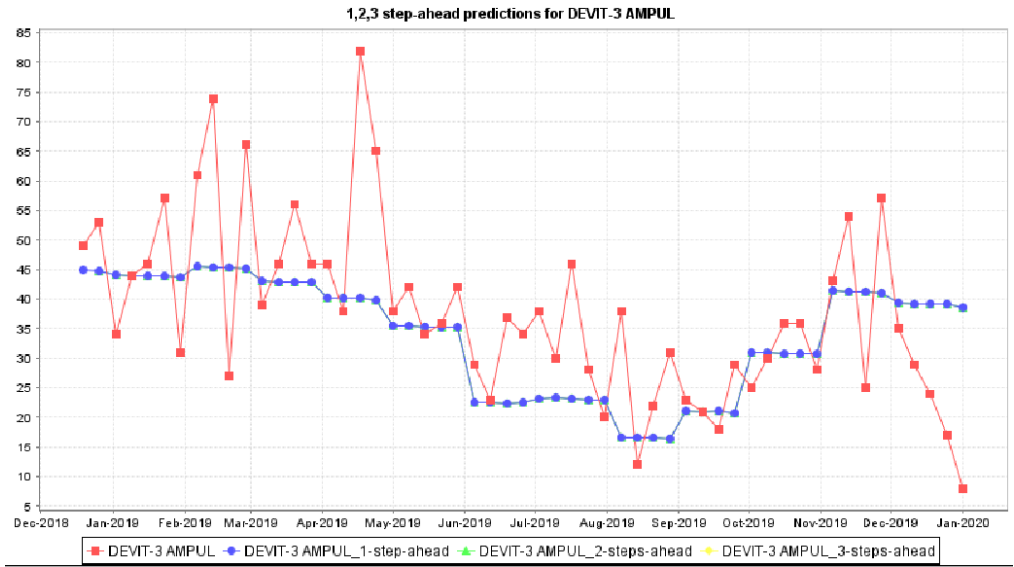
Total number of instances: 55

```

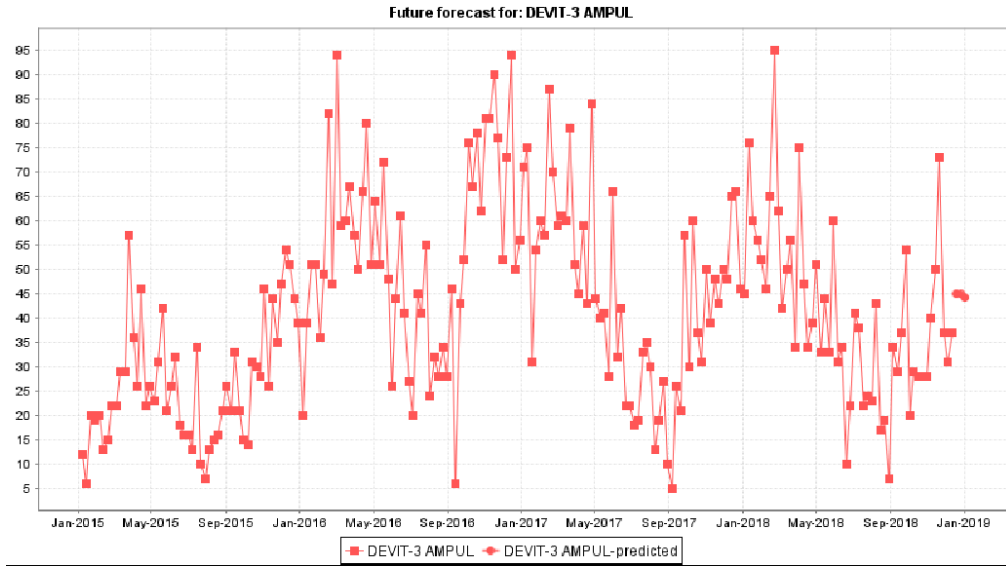
Şekil 4.24: DEVIT-3 AMPUL için Gaussian Process Regression Çıktıları



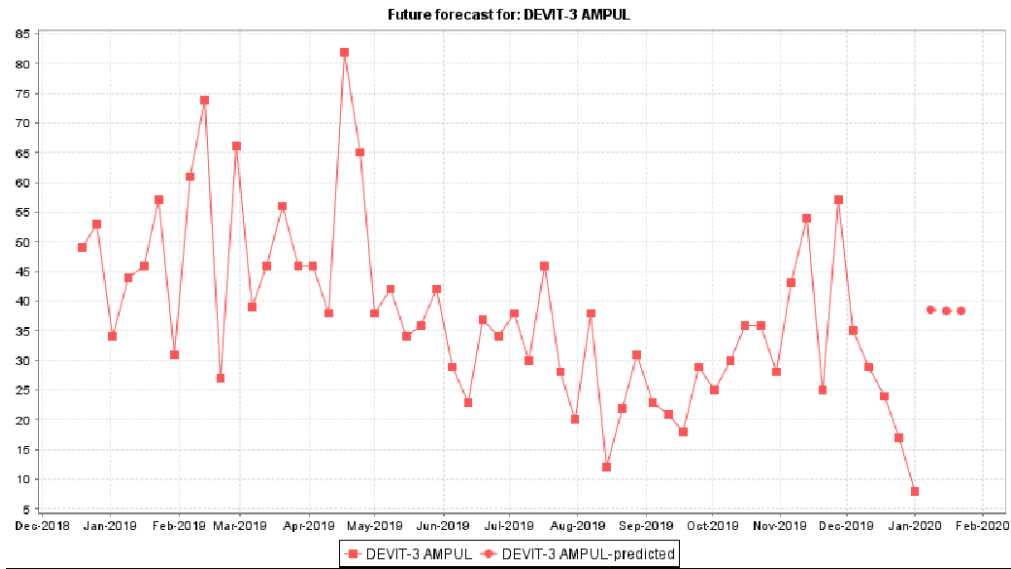
Şekil 4.25: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği



Şekil 4.26: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği



Şekil 4.27: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği



Şekil 4.28: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği

4.5.1.4 DEVIT-3 AMPUL için multilayer perceptron algoritması

Aşağıda MultilayerPerceptron algoritmasına bağlı olarak aldığımız 3 haftalık tahminleme değerlendirme ölçüm verileri yer almaktadır. Bu veriler eğitim ve test verileri olarak sonuçları 3 haftalık olarak alınmıştır.

=== Evaluation on training data ===

Target 1-step-ahead 2-steps-ahead 3-steps-ahead

=====

DEVIT-3 AMPUL

N	151	150	149
Mean absolute error	15.1473	15.213	15.2424
Mean absolute percentage error	33.1484	33.3001	33.3216
Root mean squared error	19.3058	19.3652	19.4098

Total number of instances: 206

=== Evaluation on test data ===

Target 1-step-ahead 2-steps-ahead 3-steps-ahead

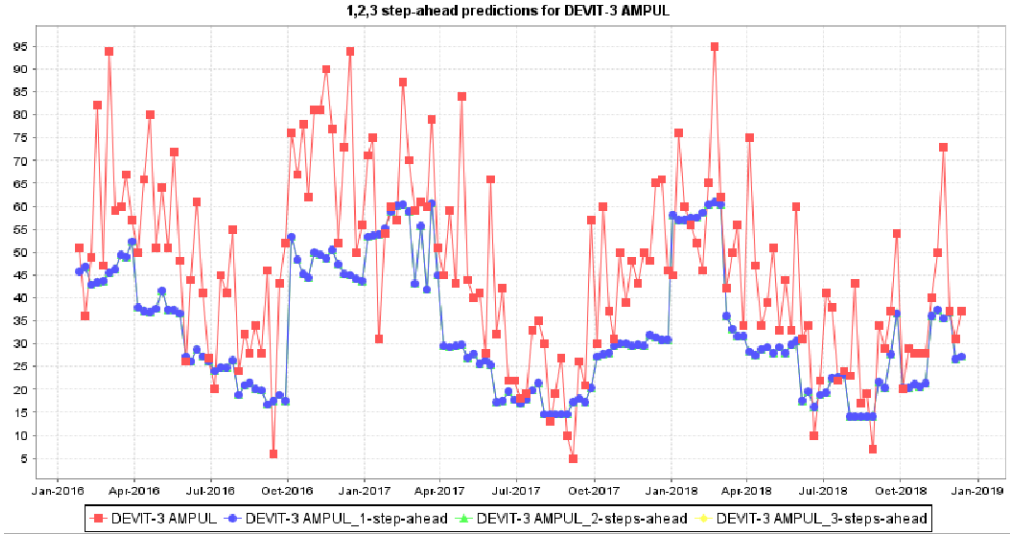
=====

DEVIT-3 AMPUL

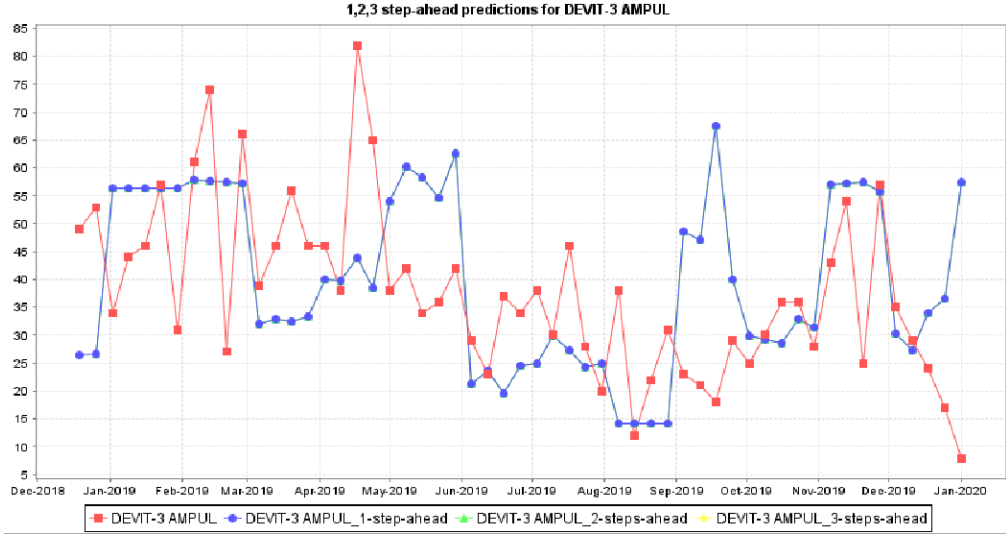
N	55	54	53
Mean absolute error	14.5373	14.3888	14.1617
Mean absolute percentage error	52.5584	52.6793	52.7325
Root mean squared error	18.6401	18.5598	18.3791

Total number of instances: 55

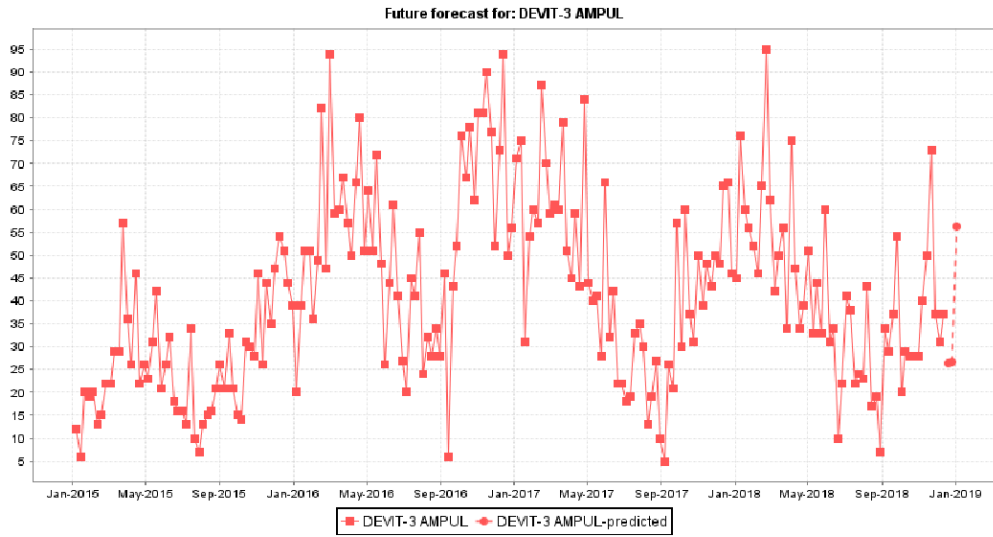
Şekil 4.29: DEVIT-3 AMPUL için Multilayerperceptron Çıktıları



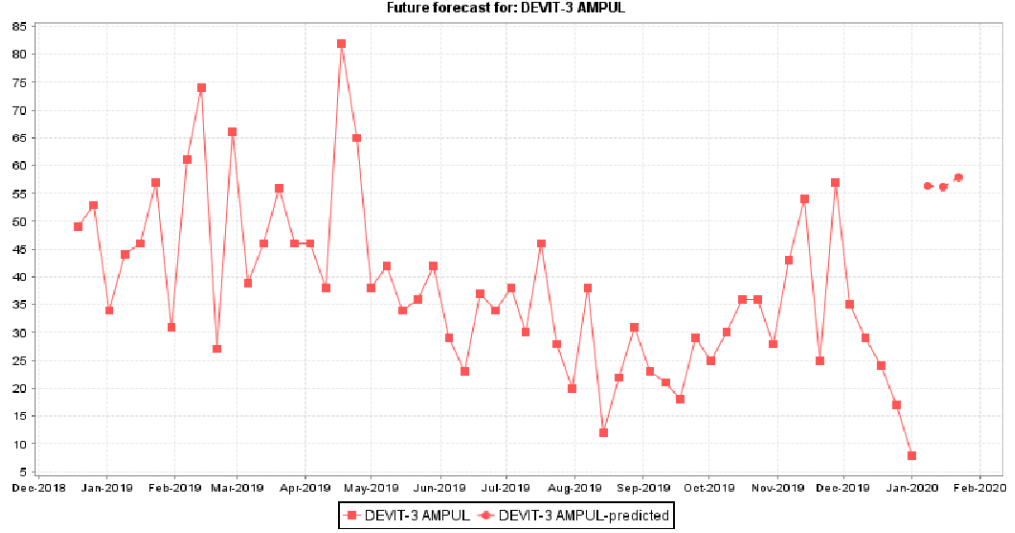
Şekil 4.30: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği



Şekil 4.31: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği



Şekil 4.32: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği



Şekil 4.33: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği

4.5.1.5 DEVIT-3 AMPUL için smoreg algoritması

Aşağıda SMOreg algoritmasına bağlı olarak aldığımız 3 haftalık tahminleme değerlendirme ölçüm verileri yer almaktadır. Bu veriler eğitim ve test verileri olarak sonuçları 3 haftalık olarak alınmıştır.

```

=== Evaluation on training data ===
Target                1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                    151           150           149
Mean absolute error  10.9282       10.8083       10.6326
Mean absolute percentage error  33.2058       33.0493       32.5815
Root mean squared error  15.0063       14.8701       14.6089

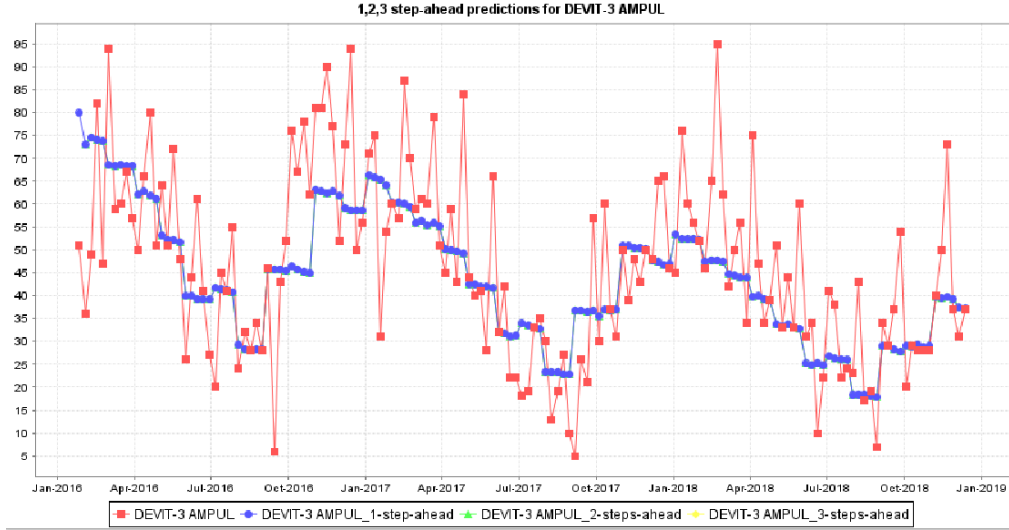
Total number of instances: 206

=== Evaluation on test data ===
Target                1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                    55           54           53
Mean absolute error  12.7115       12.7264       12.6627
Mean absolute percentage error  35.1825       35.3841       35.4785
Root mean squared error  16.0871       16.1544       16.1554

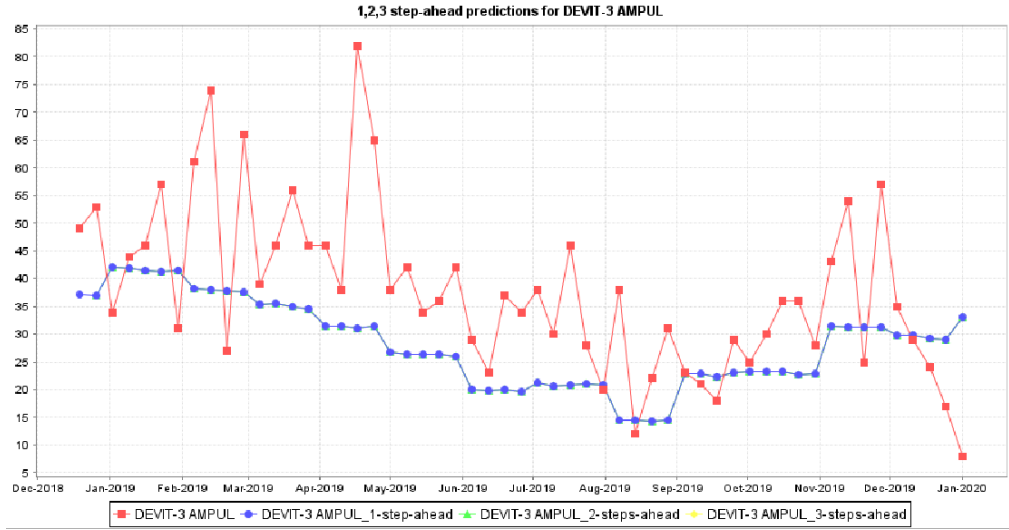
Total number of instances: 55

```

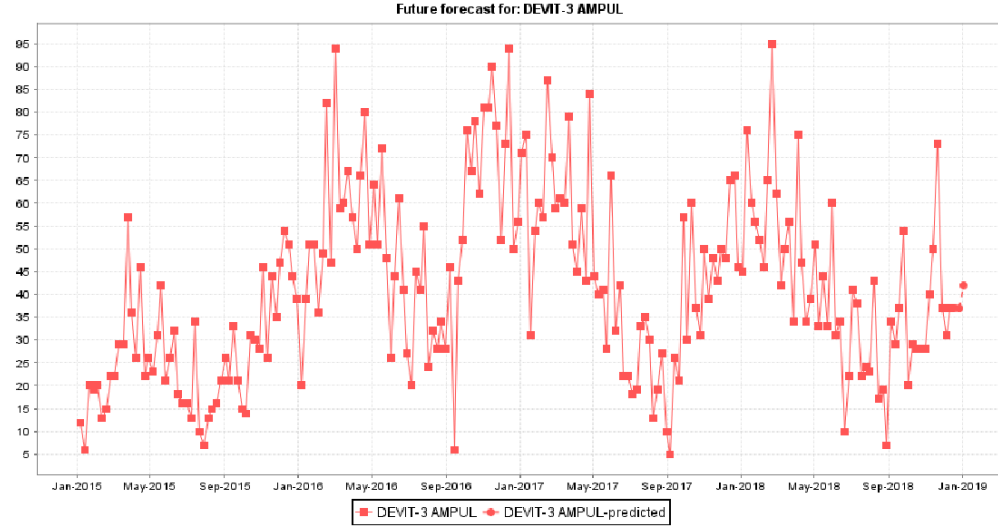
Şekil 4.34: DEVIT-3 AMPUL için SMOreg Çıktıları



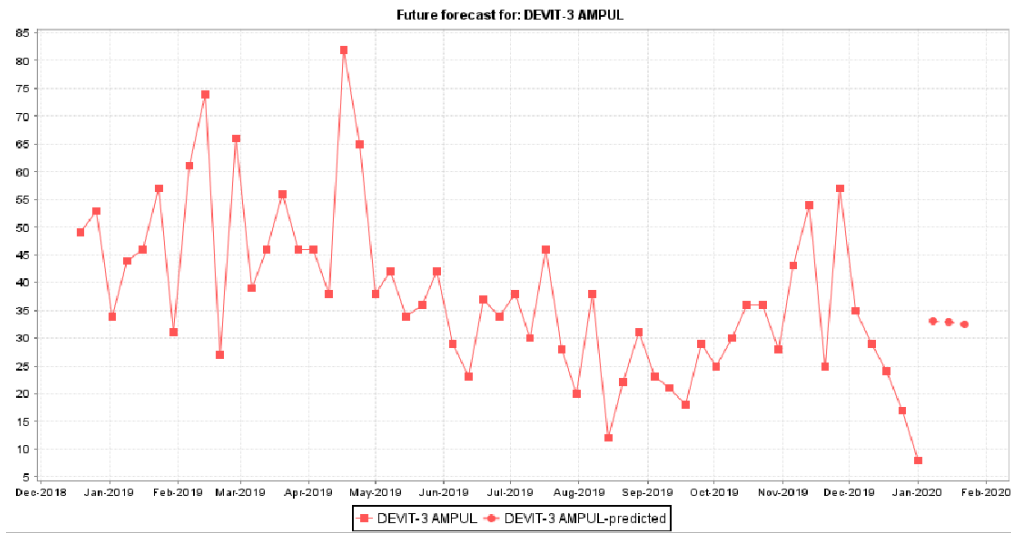
Şekil 4.35: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği



Şekil 4.36: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği



Şekil 4.37: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği



Şekil 4.38: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği

4.5.1.6 DEVIT-3 AMPUL için mp5 algoritması

Aşağıda MP5 algoritmasına bağlı olarak aldığımız 3 haftalık tahminleme değerlendirme ölçüm verileri yer almaktadır. Bu veriler eğitim ve test verileri olarak sonuçları 3 haftalık olarak alınmıştır.

```

=== Evaluation on training data ===
Target          1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N               151          150           149
Mean absolute error  11.4914      11.4441      11.2969
Mean absolute percentage error  32.56       32.5341     32.1302
Root mean squared error  14.3158     14.283       14.0676

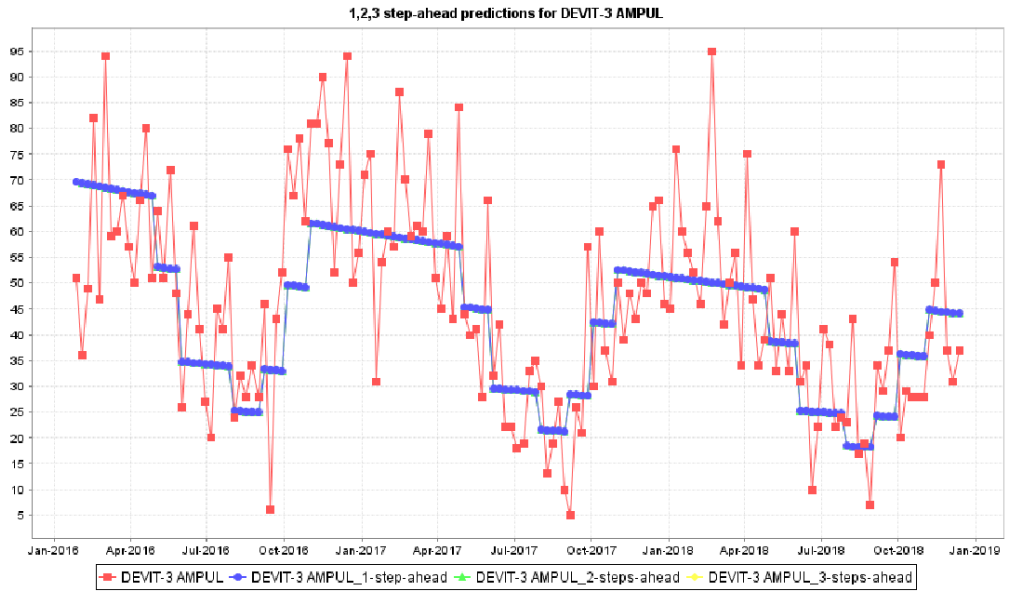
Total number of instances: 206

=== Evaluation on test data ===
Target          1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N               55           54            53
Mean absolute error  10.3447     10.4433     10.4676
Mean absolute percentage error  32.313     32.7216     33.013
Root mean squared error  13.5266    13.6341     13.7045

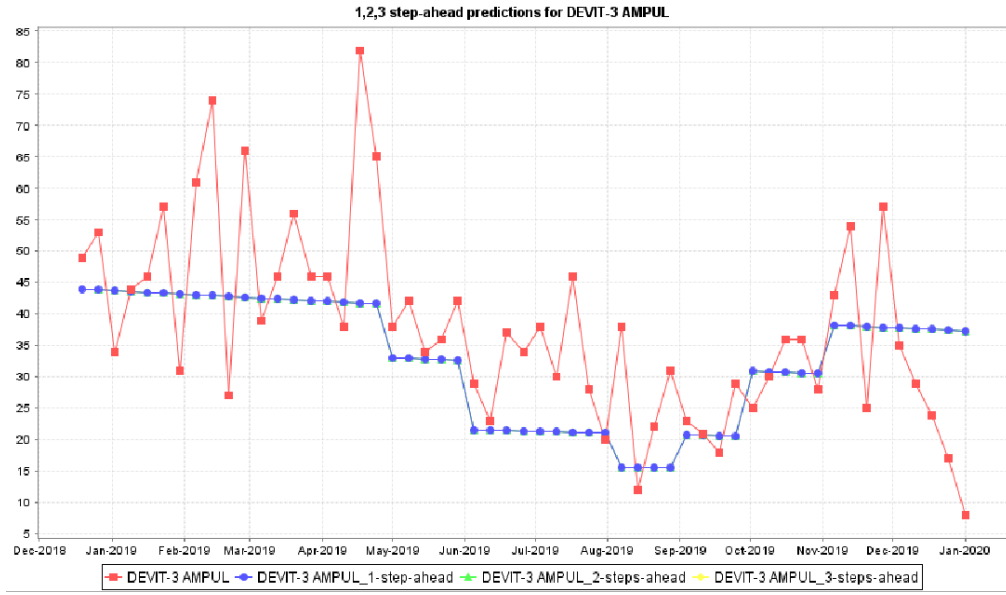
Total number of instances: 55

```

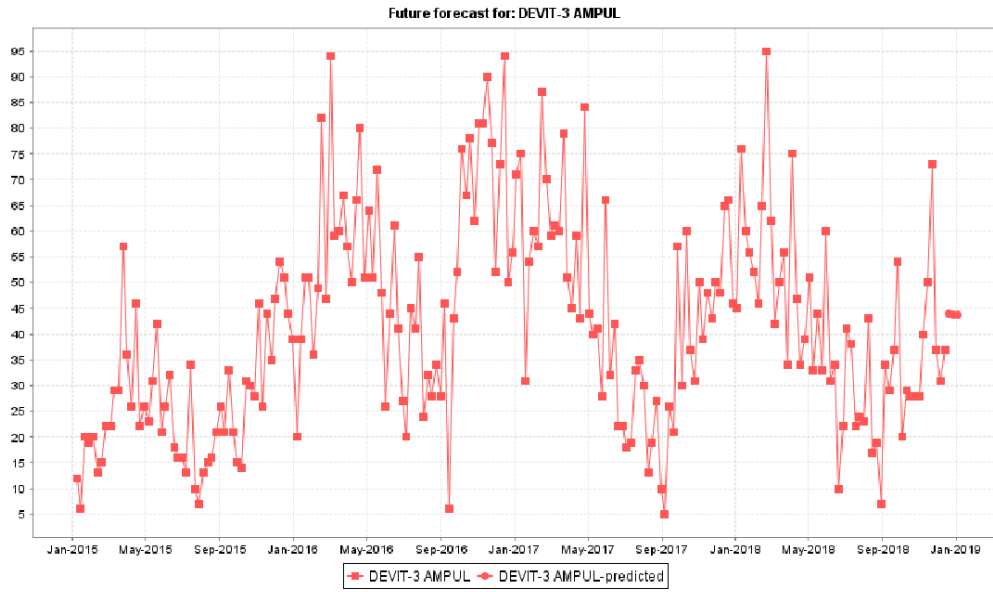
Şekil 4.39: DEVIT-3 AMPUL için MP5 Çıktıları



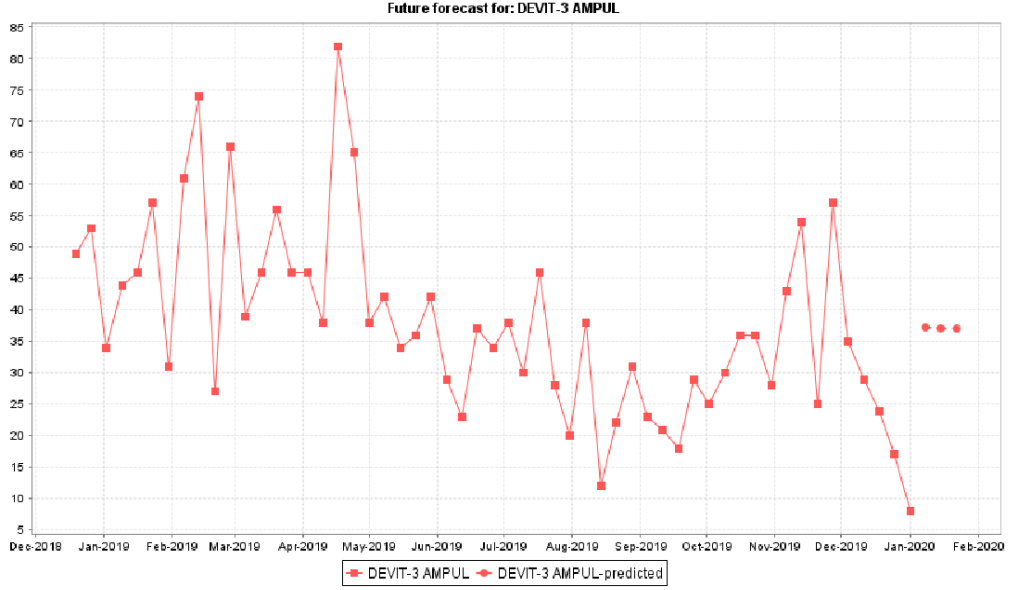
Şekil 4.40: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği



Şekil 4.41: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği



Şekil 4.42: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği



Şekil 4.43: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği

4.5.1.7 DEVIT-3 AMPUL için random forest algoritması

Aşağıda RandomForest algoritmasına bağlı olarak aldığımız 3 haftalık tahminleme değerlendirme ölçüm verileri yer almaktadır. Bu veriler eğitim ve test verileri olarak sonuçları 3 haftalık olarak alınmıştır.

```

=== Evaluation on training data ===
Target                               1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                                     151           150           149
Mean absolute error                   4.9824        5.011         4.9942
Mean absolute percentage error        12.3724       12.4457       12.3893
Root mean squared error               6.6052        6.627         6.6207

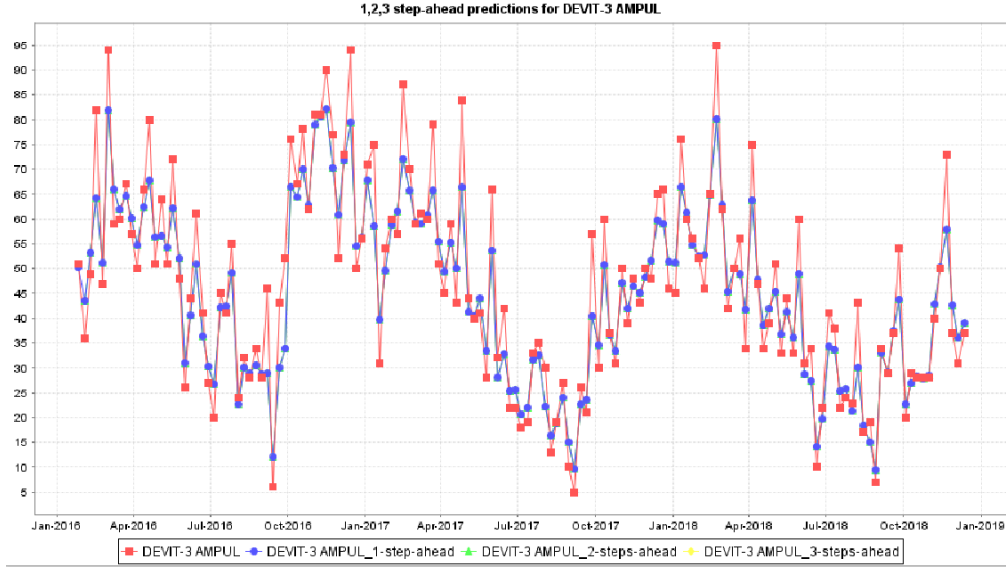
Total number of instances: 206

=== Evaluation on test data ===
Target                               1-step-ahead  2-steps-ahead  3-steps-ahead
=====
DEVIT-3 AMPUL
N                                     55            54            53
Mean absolute error                   11.7021       11.7199       11.6494
Mean absolute percentage error        41.9573       42.3283       42.5768
Root mean squared error               15.294        15.3655       15.3639

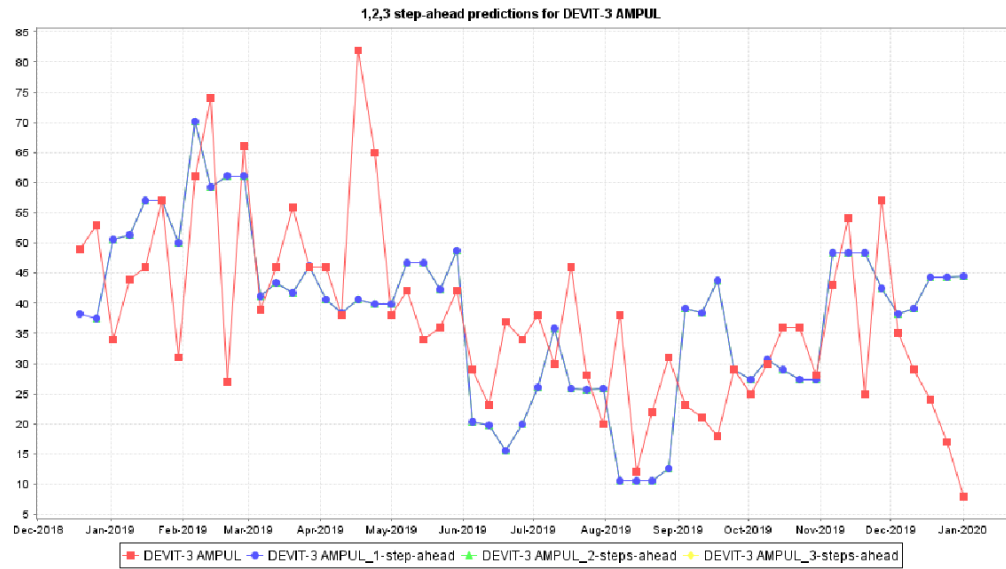
Total number of instances: 55

```

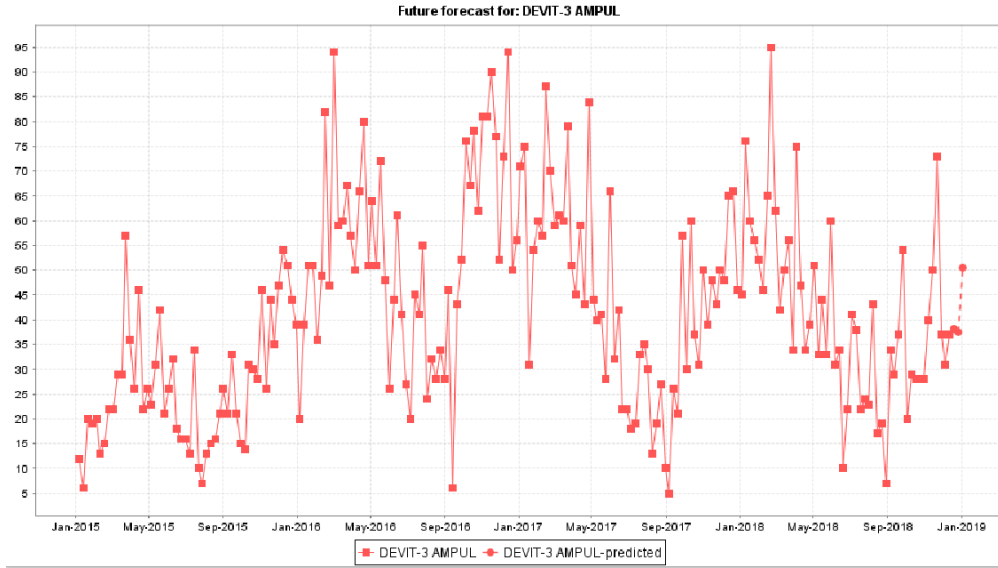
Şekil 4.44: DEVIT-3 AMPUL için Randomforest Çıktıları



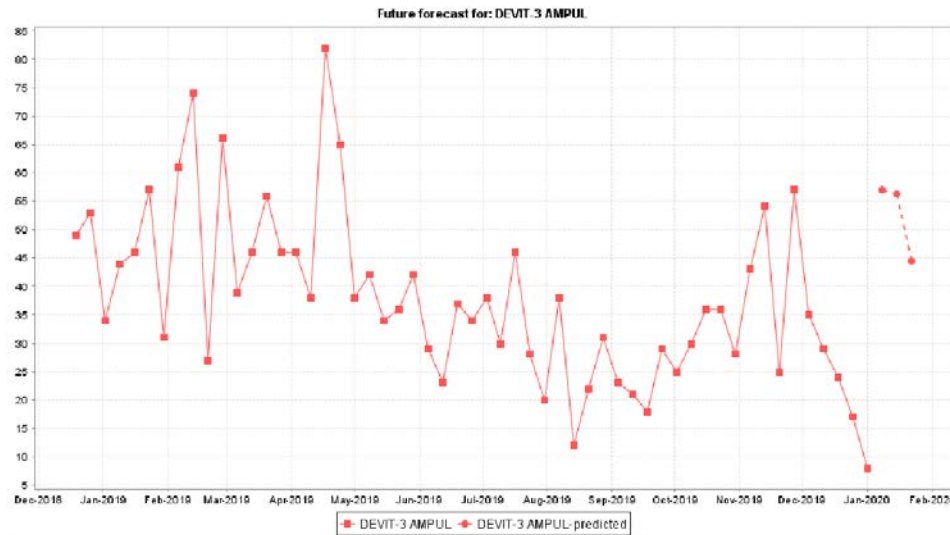
Şekil 4.45: DEVIT-3 AMPUL İlaçın Üç Haftalık Eğitim Grafiği



Şekil 4.46: DEVIT-3 AMPUL İlaçın Üç Haftalık Test Grafiği



Şekil 4.47: DEVIT-3 AMPUL İlaçın Eğitim Gelecek Grafiği



Şekil 4.48: DEVIT-3 AMPUL İlaçın Test Gelecek Grafiği

4.6 Bulgular

Tahminlemesi yapılan 10 ilaç arasından rasgele bir ilaçın zaman serileri analiz sonuçlarını incelediğimizde test veri setinden aldığımız 1 haftalık satış miktarı tahmin sonuçlarında gerçekleşen ve tahmin edilen “DEVIT-3 AMPUL“ isimli ilaçın gerçekleşen ve tahmin edilen satış miktar sonuçları arasındaki en iyi hata oranını sahip “M5Rules” algoritmasında olduğu görülmektedir. M5Rules

algoritmasının çalışma prensibini hatırladığımızda sayısal olarak ifade edilen öngörü için karar listesi kullanarak regresyon modeli veya regresyon ağacı oluşturduğu ifade edebiliriz. Bu algoritma modelleri, ağaç bazlı oluşum gerçekleştirir ve model ağacının yapraklarını çok değişkenli olarak linear regresyon denklemleri ile ifade etmektedir.

Aşağıda incelediğimiz DEVIT-3 AMPUL ilaç'ı için en iyi miktarsal tahmin sonucunu veren M5Rules algoritmasının eğitim ve test verilerine ait haftalık olarak gerçekleşen ile tahmin edilen miktarların ve bunlara ait hata miktarı sonuçları yer almaktadır.

Çizelge 4.1: Eğitim Verileri için Tahminler: DEVIT-3 AMPUL

Hafta	Gerçekleşen	Tahmin	Hata
51	44	52,596	8,596
52	39	59,4261	20,4261
53	20	47,9283	27,9283
54	39	44,3035	5,3035
55	51	45,1868	-5,8132
56	51	47,5001	-3,4999
57	36	57,9316	21,9316
58	49	58,4441	9,4441
59	82	61,8068	-20,1932
60	47	64,6794	17,6794
61	94	65,897	-28,103
62	59	63,5093	4,5093
63	60	51,7321	-8,2679
64	67	58,746	-8,254
65	57	59,588	2,588
66	50	60,5838	10,5838
67	66	58,2316	-7,7684
68	80	66,9696	-13,0304
69	51	68,9869	17,9869
70	64	52,0372	-11,9628
71	51	59,3927	8,3927
72	72	49,6736	-22,3264
73	48	55,8178	7,8178
74	26	38,9035	12,9035
75	44	34,3384	-9,6616
76	61	32,3218	-28,6782
77	41	38,6252	-2,3748
78	27	35,2334	8,2334
79	20	38,9673	18,9673
80	45	40,0684	-4,9316
81	41	45,0766	4,0766

Çizelge 4.1: (Devam) Eğitim Verileri için Tahminler: DEVIT-3 AMPUL

Hafta	Gerçekleşen	Tahmin	Hata
82	55	47,3834	-7,6166
83	24	36,659	12,659
84	32	22,8942	-9,1058
85	28	34,7801	6,7801
86	34	28,0195	-5,9805
87	28	21,4325	-6,5675
88	46	37,8522	-8,1478
89	6	29,3708	23,3708
90	43	32,136	-10,864
91	52	40,9165	-11,0835
92	76	57,9446	-18,0554
93	67	50,6223	-16,3777
94	78	58,7706	-19,2294
95	62	51,4261	-10,5739
96	81	58,4454	-22,5546
97	81	64,548	-16,452
98	90	61,2188	-28,7812
99	77	59,2927	-17,7073
100	52	63,616	11,616
101	73	64,4652	-8,5348
102	94	66,932	-27,068
103	50	68,395	18,395
104	56	63,0502	7,0502
105	71	58,5793	-12,4207
106	75	62,9275	-12,0725
107	31	58,3709	27,3709
108	54	56,0458	2,0458
109	60	65,3926	5,3926
110	57	62,2658	5,2658
111	87	63,9807	-23,0193
112	70	71,891	1,891
113	59	61,9099	2,9099
114	61	63,4213	2,4213
115	60	69,8525	9,8525
116	79	67,1345	-11,8655
117	51	71,4042	20,4042
118	45	61,9586	16,9586
119	59	60,4523	1,4523
120	43	57,9169	14,9169
121	84	67,0211	-16,9789
122	44	51,7106	7,7106
123	40	48,3258	8,3258
124	41	44,4351	3,4351
125	28	52,7879	24,7879

Çizelge 4.1: (Devam) Eğitim Verileri için Tahminler: DEVIT-3 AMPUL

Hafta	Gerçekleşen	Tahmin	Hata
126	66	51,5441	-14,4559
127	32	35,9907	3,9907
128	42	31,8274	-10,1726
129	22	39,4658	17,4658
130	22	30,0478	8,0478
131	18	46,7911	28,7911
132	19	36,2254	17,2254
133	33	30,658	-2,342
134	35	34,5774	-0,4226
135	30	37,4096	7,4096
136	13	28,7564	15,7564
137	19	20,7775	1,7775
138	27	22,1221	-4,8779
139	10	18,9818	8,9818
140	5	27,2354	22,2354
141	26	30,0467	4,0467
142	21	32,0797	11,0797
143	57	36,3231	-20,6769
144	30	47,2176	17,2176
145	60	52,2351	-7,7649
146	37	47,7416	10,7416
147	31	43,3533	12,3533
148	50	54,136	4,136
149	39	45,2564	6,2564
150	48	52,9269	4,9269
151	43	54,9949	11,9949
152	50	52,0041	2,0041
153	48	56,5987	8,5987
154	65	52,9606	-12,0394
155	66	60,0699	-5,9301
156	46	49,3424	3,3424
157	45	44,4686	-0,5314
158	76	46,2059	-29,7941
159	60	54,6711	-5,3289
160	56	48,9772	-7,0228
161	52	45,5029	-6,4971
162	46	51,3643	5,3643
163	65	48,8107	-16,1893
164	95	56,0133	-38,9867
165	62	64,5917	2,5917
166	42	53,2987	11,2987
167	50	57,6796	7,6796
168	56	61,7879	5,7879
169	34	55,9058	21,9058

Çizelge 4.1: (Devam) Eğitim Verileri için Tahminler: DEVIT-3 AMPUL

Hafta	Gerçekleşen	Tahmin	Hata
170	75	54,8954	-20,1046
171	47	50,0029	3,0029
172	34	49,0516	15,0516
173	39	42,5111	3,5111
174	51	46,3784	-4,6216
175	33	46,3383	13,3383
176	44	39,7926	-4,2074
177	33	41,1961	8,1961
178	60	43,7045	-16,2955
179	31	28,3489	-2,6511
180	34	25,2247	-8,7753
181	10	24,2935	14,2935
182	22	19,3788	-2,6212
183	41	26,4664	-14,5336
184	38	32,9406	-5,0594
185	22	28,343	6,343
186	24	21,5462	-2,4538
187	23	17,7377	-5,2623
188	43	25,091	-17,909
189	17	15,7643	-1,2357
190	19	12,6524	-6,3476
191	7	10,9218	3,9218
192	34	18,9393	-15,0607
193	29	25,4543	-3,5457
194	37	35,903	-1,097
195	54	25,6991	-28,3009
196	20	35,8155	15,8155
197	29	41,8092	12,8092
198	28	34,2406	6,2406
199	28	30,1108	2,1108
200	28	34,9101	6,9101
201	40	41,1232	1,1232
202	50	45,1319	-4,8681
203	73	53,2529	-19,7471
204	37	52,0708	15,0708
205	31	43,8313	12,8313
206	37	34,694	-2,306

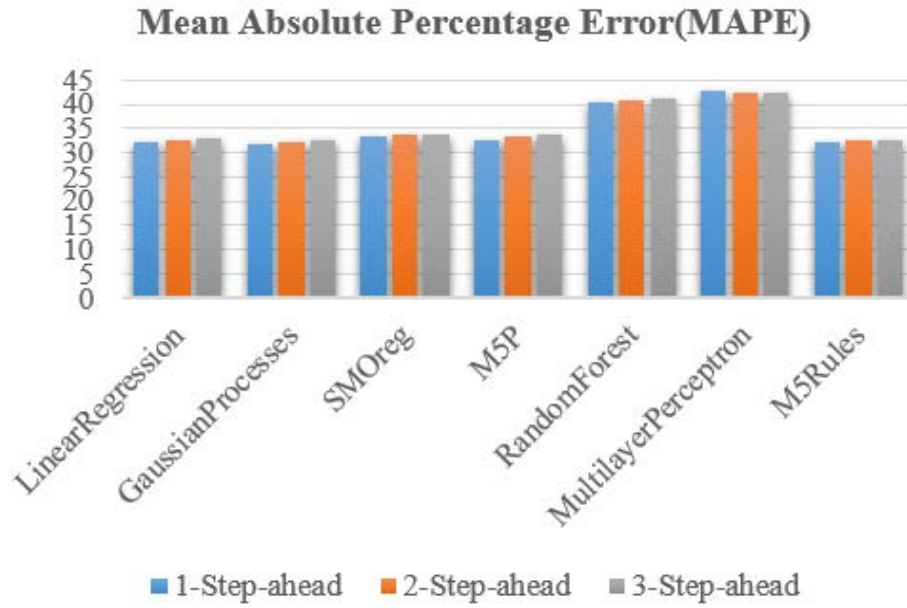
Çizelge 4.2: Test Verileri için Tahminler: DEVIT-3 AMPUL

Hafta	Gerçekleşen	Tahmin	Hata
207	49	50,0064	1,0064
208	53	46,059	-6,941
209	34	40,9904	6,9904
210	44	33,9697	-10,0303
211	46	37,3923	-8,6077
212	57	39,985	-17,015
213	31	48,3202	17,3202
214	61	39,9912	-21,0088
215	74	41,7206	-32,2794
216	27	48,506	21,506
217	66	47,4418	-18,5582
218	39	49,7567	10,7567
219	46	51,6412	5,6412
220	56	41,5626	-14,4374
221	46	47,1893	1,1893
222	46	42,2481	-3,7519
223	38	36,4068	-1,5932
224	82	42,339	-39,661
225	65	56,4642	-8,5358
226	38	36,6553	-1,3447
227	42	46,1743	4,1743
228	34	36,8805	2,8805
229	36	28,7063	-7,2937
230	42	28,4181	-13,5819
231	29	13,0644	-15,9356
232	23	20,0634	-2,9366
233	37	15,8959	-21,1041
234	34	25,602	-8,398
235	38	27,4507	-10,5493
236	30	27,8134	-2,1866
237	46	23,4948	-22,5052
238	28	23,0816	-4,9184
239	20	23,1653	3,1653
240	38	3,8032	-34,1968
241	12	17,8295	5,8295
242	22	0,1821	-21,8179
243	31	4,9335	-26,0665
244	23	28,8859	5,8859
245	21	17,9276	-3,0724
246	18	17,9465	-0,0535
247	29	22,2714	-6,7286
248	25	34,9841	9,9841
249	30	25,0426	-4,9574
250	36	32,3673	-3,6327
251	36	27,7118	-8,2882
252	28	23,7838	-4,2162
253	43	34,0176	-8,9824
254	54	38,3408	-15,6592
255	25	37,7097	12,7097
256	57	33,3751	-23,6249
257	35	48,8196	13,8196
258	29	38,5379	9,5379
259	24	38,0626	14,0626
260	17	39,0169	22,0169
261	8	28,438	20,438

Çizelge 4.3: Ortalama Mutlak Yüzde Hata Sonuçları DEVIT-3 AMPUL

Algoritmalar	1-hafta	2-hafta	3-hafta
Linear Regression	32.313	32.7216	33.013
Gaussian Processes	31.947	32.3844	32.704
SMOreg	35.1825	35.3841	35.4785
M5P	32.313	32.7216	33.013
Random Forest	41.9573	42.3283	42.5768
Multilayer Perceptron	52.5584	52.6793	52.7325
M5Rules	32.313	32.7216	33.013

Yukarıda Çizel 4.2.3’de Mean Absolute Percentage error sonuçlarına göre incelediğimizde 1.haftalık öngörülerde en iyi tahmin 31.947 sonucu olarak sıfıra en yakın sonuç olan Gaussian Processes algoritmasının ürettiği sonucu 1.hafta için en iyi sonuç olarak , 2.haftalık öngörülerde en iyi tahmin 32.3844 sonucu olarak sıfıra en yakın sonuç olan Gaussian Processes algoritmasının ürettiği sonucu 2.hafta için en iyi sonuç olarak , 3.haftalık öngörülerde en iyi tahmin 32.704 sonucu olarak sıfıra en yakın sonuç olan Gaussian Processes algoritmasının ürettiği sonucu 3.hafta için en iyi sonuç olarak ifade edebiliriz. Gaussian Processes algoritması 3 haftalık serilerde en iyi sonuçları veren algoritma olarak görülmektedir.



Şekil 4.49: Algoritma MAPE Karşılaştırması

Mean absolute percentage error değerleri farklı kaynaklarda sayısal aralıklara göre farklı şekilde değerlendirilmektedir. Bu değerlendirilmeleri aşağıdaki şekilde ifade edebiliriz.

Mean absolute percentage error değerlendirme

- Değer aralığı $MAPE < 10 \Rightarrow$ 0-10 arasında olan modeller “çok iyi”, olarak ifade edilmektedir.
- Değer aralığı $10 < MAPE < 20 \Rightarrow$ 10-20 arasında olan modelleri “iyi” olarak öngörülmektedir.
- Değer aralığı $20 < MAPE < 50 \Rightarrow$ 20-50 arasında olan modelleri “kabul görülebilir”
- Son olarak $50 < MAPE \Rightarrow$ 50 üstünde olan modeller ise “hatalı ve yanlış” ifade edilir.

5. SONUÇLAR

Eczane ilaç satış verilerinden elde ettiğimiz 5 yıllık zaman sürecinde en çok satış miktarına sahip "DEVIT-3 AMPUL" ilacının 7 algoritma üzerinde yaptığımız 3 haftalık tahminleme sonuçlarında "Mean absolute percentage error" değerlerinde "GaussianProcesses" algoritmasının en iyi sonuç olarak karşımıza çıkmaktadır. 3 haftalık tahminlemede "GaussianProcesses" algoritmasına en yakın değerleri 1, 2. ve 3. hafta tahminlemede "M5Rules" algoritması vermektedir.

"M5Rules" algoritmasını 1.haftada gerçekleşen ve tahmin sonuçları olarak miktarsal bazda en iyi sonucu veren algoritma olarak Çizelge 4.2.1'de incelemiştik."M5Rules" algoritması performans olarak zaman serilerinde başarılı sonuçlar ürettiği görülmektedir.

KAYNAKLAR

- [1] **Zaman Serileri Analizi** **ErişimAdresi :**
<https://ekonometrice.blogspot.com/2015/04/zaman-serisi-bolum1.html?m=1> Nisan 04, 2015
- [2] **Dr. Aysel Çetindere Filiz İktisadi ve idari bilimler fakültesi** Erişim Adresi :
<https://avys.omu.edu.tr/storage/app/public/ayssel.cetindere/131686/11.%20Hafta%20Talep%20Tahmin%20Y%C3%B6ntemleri.pdf>
- [3] **Şeker, Sadi Evren(2015). Zaman Serisi Analizi**, YBS Ansiklopedisi, Cilt 2, Sayı 4.
- [4] **Öğüt, Sertaç (2009)**. Veri madenciliği kavramı ve gelişim süreci.
- [5] **Köktürk Firuzan, Ankaralı Handan, SümbülOğlu Vildan (2008)**. Veri Madenciliği Yöntemlerine Genel Bakış.
- [6] **Özdemir Abdulkadir, Yalçın Fulya Aslay, Çam Handan (2010)**.Veri tabanında bilgi keşfi süreci. SÜ İİBF Sosyal ve Ekonomik Araştırmalar Dergisi S.349
- [7] **Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth (1996)**. Knowledge Discovery and Data Mining: Towards a Unifying Framework. s.82.
- [8] **Ahmet AKÇAY (2014)**. Bilgi ve Belge Yönetiminde Veri Madenciliği.
- [9] **Ryder Braddock (2019)**. What are the different fields where data mining is used?. Alındığı Tarih : 20.02.2020, Adres :
<https://www.quora.com/What-are-the-different-fields-where-data-mining-is-used>.
- [10] **Mohsen Ghorbian (2019)**. 14 areas where data mining is widely used. Alındığı Tarih: 11.02.2020, Adres :
https://www.researchgate.net/post/14_areas_where_data_mining_is_widely_used
- [11] **Songül Şekeroğlu (2010)**. Hizmet Sektöründe bir veri madenciliği uygulaması
- [12] **Mine Çelik (2009)**.Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama.
- [13] **Haldun Akpınar (2000)**. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. s.7
- [14] **Arlene Zaima, James Kashner (2013)**. Data Mining Primer for the Data Warehouse Professional.
- [15] **Haldun Akpınar (2000)** , a.g.e., s.13.
- [16] **Haldun Akpınar (2000)** , a.g.e., s.1-22.
- [17] **Bharati M. (2010)**. Data mining techniques and applications.
- [18] **Ahmet AKÇAY (2014)**. S:67.
- [19] **M. Silver, T. Sakara, H. C. Su, C. Herman, S. B. Dolins and M. J. O'shea(2001)**. How to apply data mining techniques in a healthcare data warehouse.

- [20] **J. J. Tapia, E. Morett and E. E. Vallejo(2009)**. A Clustering Genetic Algorithm for Genomic Data Mining.
- [21] **Mine Çelik (2009)**. S.29.
- [22] **Bryan F.J. Mainly (1994)**. Multivariate Statistical Methods.
- [23] **Karasu S., Altan A., Sarac Z. , Hacıoglu R. (2018)**. Zaman Serisi Verilerini Kullanarak Makine Öğrenmesi Yöntemleri ile Bitcoin Fiyat Tahmini
- [24] **Zeynep Behrin Güven, Turgay Tugay Bilgin (2014)**. Zaman Serileri Madenciliği Kullanılarak Nüfus Artışı Tahmin Uygulaması
- [25] **Hande Nasuhoğlu (2019)**. Eczacılık Sektöründe Yapay Sinir Ağları ve Zaman Serileri Analizi ile Talep Tahmini
- [26] **Gianluca Bontempi, Souhaib Ben Taieb, Yann-Aël Le Borgne (2012)**. Machine Learning Strategies for Time Series Forecasting
- [27] **Nicholas I. Sapankevych ; Ravi Sankar (2009)**. Time Series Prediction Using Support Vector Machines: A Survey
- [28] **Mustafa Can (2009)**. İşletmelerde zaman serileri analizi ile tahmin
- [29] **Utku Kubilay ÇINAR**. R ile Derinlemesine Zaman Serisi **Erişim Adresi** : <https://www.veribilimiokulu.com/zaman-serileri-geleneksel-yontemler-ile-yapay-sinir-aglarinin-karsilastirilmasi/>
- [30] **Esin ERGUVAN ETGİN (2017)**. Zaman Serilerinde veri madenciliği öngörü algoritmalarının etkinlik ve verimliliğinin bist100 hisse senetleri üzerinde gerçekleşmesi.
- [31] **Dr. Gökçen UYSAL(2019)**. Yapay Sinir Ağları ile Hidrolojik Modelleme.
- [32] **Zeynep Behrin Güven**. Zaman Serileri Madenciliği Kullanılarak Nüfus Artışı Tahmin Uygulaması.
- [33] **Evin GARİP (2017)**. OECD Ülkelerindeki CO2 Emisyonunun Makine Öğrenmesi ile tahmin edilmesi.
- [34] **Ekrem Hatipoglu(2018)**. Machine Learning — Classification — Decision Tree — Random Forest **Erişim Adresi**: <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-decision-tree-random-forest-part-12-8c9515d811b9>
- [35] **MAE and RMSE — Which Metric is Better?(2016)** **Erişim Adresi**: <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d>
- [36] **Ridhima Kumar(2020) MAD over MAPE?** **Erişim Adresi**: <https://towardsdatascience.com/mad-over-mape-a86a8d831447>
- [37] **Simon P. Neill M. Reza Hashemi(2020). Ocean Modelling for Resource Characterization.** **Erişim Adresi**: <https://www.sciencedirect.com/science/article/pii/B9780128104484000082>

ÖZGEÇMİŞ

Ad -Soyad: İlker Poyraz

Doğum Tarihi ve Yeri: 1981, İstanbul

E-Posta: poyrazilker@hotmail.com



ÖĞRENİM DURUMU:

- **Ön Lisans:** 2001, Süleyman Demirel Üniversitesi, Bilgisayar Programcılığı
- **Lisans:** 2008, Anadolu Üniversitesi, İşletme Fakültesi
- **Yüksek Lisans:** Devam etmektedir, İstanbul Aydın Üniversitesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİMLER:

- **09.2001 – 10.2001** : KoçSistem Bilgi Teknolojileri – Stajer
- **10.2001 – 10.2004** : aMVG – Yazılım Geliştirici
- **01.2005 – 07.2008** : Doğan Yayın Holding – Kıdemli Yazılım Geliştirici
- **07.2008 – 07.2009** : Innovaktif Bilgi Teknolojileri – Kıdemli Yazılım Geliştirici
- **02.2010 – 10.2010** : IMS Health – Kıdemli Yazılım Geliştirici
- **02.2011 - Devam** : Uyumsoft Bilgi Sistemleri - Kıdemli Yazılım Geliştirici