

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ZAMAN SERİSİ YÖNTEMLERİ VE XGBOOST ALGORİTMASI
İLE KÖMÜR SATIŞ TAHMİNİ: TÜRKİYE TARIM KREDİ
KOOPERATİFLERİ UYGULAMASI**

YÜKSEK LİSANS TEZİ

Gökhan ACAR

Enstitü Anabilim Dalı : BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ

Tez Danışmanı : Dr. Öğr. Üyesi Tuğrul TAŞCI

Eylül 2022

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ZAMAN SERİSİ YÖNTEMLERİ VE XGBOOST ALGORİTMASI
İLE KÖMÜR SATIŞ TAHMİNİ: TÜRKİYE TARIM KREDİ
KOOPERATİFLERİ UYGULAMASI**

YÜKSEK LİSANS TEZİ

Gökhan ACAR

Enstitü Anabilim Dalı : BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ

Bu tez 28.09.2022 tarihinde aşağıdaki jüri tarafından oybirliği ile kabul edilmiştir.

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Gökhan ACAR

26/09/2022

TEŞEKKÜR

Yüksek lisans eğitimim boyunca bilgi ve deneyimlerinden yararlandığım, araştırmanın tüm aşamalarında yardımlarını esirgemeyen değerli danışman hocam Dr. Öğr.Üyesi Tuğrul TAŞCI'ya, çalışmalarım boyunca beni destekleyen aileme ve varlığı ile motivasyon kaynağım kızım Gökçe'ye teşekkürlerimi sunarım.

İÇİNDEKİLER

TEŞEKKÜR	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ	v
ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ	vii
ÖZET	viii
SUMMARY	ix
BÖLÜM 1.	
GİRİŞ	1
1.1. Araştırmanın Amacı ve Konusu	3
1.2 Tezin Yapısı	4
BÖLÜM 2.	
TEMEL KAVRAMLAR VE YÖNTEMLER	6
2.1. Temel Kavramlar	6
2.1.1. Tarım Kredi Kooperatifleri	6
2.1.2. Tarım Kredi Kooperatifleri Sakarya Bölge Birliği	7
2.1.3. EKOOP (Türkiye Tarım Kredi Kooperatifleri ERP uygulaması)	8
2.1.4. Kömür ürünü	9
2.2. Zaman Serisi Kavramı ve Tahmin Yöntemleri	12
2.2.1. ARİMA (Box-Jenkins yöntemi)	12
2.2.1.1. AR modeli	12
2.2.1.2. MA modeli	13

2.2.1.3. ARMA modeli	13
2.2.1.3. ARİMA modeli	13
2.2.2. SARIMA(SARIMAX)	14
2.2.3. HOLT-WINTERS	14
2.2.4. Facebook Prophet (FB-PROPHET)	14
2.3. Makine Öğrenmesi	14
2.3.1. Makine Öğrenmesi yöntemleri	15
2.3.1.1. Denetimli öğrenme (Supervised learning)	15
2.3.1.2. Denetimsiz öğrenme (Unsupervised learning)	16
2.3.1.3. Takviyeli öğrenme (Reinforcement learning)	16
2.3.2. Karar ağaçları	16
2.3.3. Gradyan artırma ağaçları (Gradient boosting tree)	17
2.3.4. Aşırı gradyan artırma (XGBoost)	20
2.4. Performans Ölçüm Yöntemleri	21
2.4.1. AIC (Akaike bilgi kriteri)	21
2.4.2. MAE (Mean absolute error - Ortalama mutlak hata)	21
2.4.3. MAPE (Mean absolute percentage error - Ortalama mutlak hata yüzdesi)	22
2.4.4. MSE-RMSE (Mean squared error - Root mean squared error)	22
2.4.5. R-Kare (Determinasyon katsayısı)	23
 BÖLÜM 3.	
LİTERATÜR ARAŞTIRMASI VE ÖNCEKİ ÇALIŞMALAR	24
3.1. Boosting yöntemleri ve XGBoost	25
3.2. Zaman Serisi Yöntemleri	26
 BÖLÜM 4.	
GÜNCEL YÖNTEMLER İLE KÖMÜR SATIŞLARININ TAHMİN EDİLMESİ	29
4.1. Tahmin İhtiyacının Tespit Edilmesi	29

4.1. Tahmin İşlemlerinin Aşamaları	31
4.2.1. Veri toplama	32
4.2.2. Veri ön işleme	33
4.2.3. Gerçekleme süreci	34
4.2.3.1. SARIMAX	34
4.2.3.2. Holt-Winters	35
4.2.3.3. FB-Prophet	35
4.2.3.4. XGBoost	36
4.3. Sonuçların Değerlendirilmesi	36
4.4. Çalışmada Kullanılan Teknolojiler	37
 BÖLÜM 5.	
ARAŞTIRMA BULGULARI	38
5.1. Verilerin Toplanması ve Analizi	38
5.2. SARIMAX Uygulanması	38
5.3. Holt-Winters Uygulanması	42
5.4. FB-Prophet Uygulanması	44
5.5. XGBoost Uygulanması	45
5.6. Aylık ve Yıllık Frekans ile Tahmin Çalışması ve Modelleri Karşılaştırılması	49
 BÖLÜM 6.	
SONUÇ VE ÖNERİLER	52
 KAYNAKLAR	 55
ÖZGEÇMİŞ	60

SİMGELER VE KISALTMALAR LİSTESİ

ADF Test	: Arttırılmış Dickey Fuller Testi (Augmented Dickey–Fuller Test)
AIC	: Akaike ölçütü (Akaike Information Criterion)
AR	: Otoregresyon (Autoregression)
ARIMA	: Otoregresif Hareketli Ortalamalar (Autoregressive Integrated Moving Average)
ARMA	: Otoregresif Hareketli Ortalama (Autoregressive Moving Average)
EKOOP	: Türkiye Tarım Kredi Kooperatifleri ERP Uygulaması
ERP	: Kurumsal Kaynak Planlama (Enterprise Resource Planning)
MA	: Hareketli Ortalama (Moving Average)
MAE	: Ortalama Mutlak Hata (Mean Absolute Error)
MAPE	: Ortalama Mutlak Hata Yüzdesi (Mean Absolute Percentage Error)
MSE	: Ortalama Kare Hata (Mean Squared Error)
RMSE	: Kök Ortalama Kare Hatası (Root Mean Squared Error)
SARIMA	: Mevsimsel Otoregresif Entegre Hareketli Ortalama (Seasonal Auto-Regressive Integrated Moving Average)
SARIMAX	: Dışsal Değişken Kaynaklı Mevsimsel Otoregresif Entegre Hareketli Ortalama (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors)
XGBoost	: Aşırı Gradyan Artırma (Extreme Gradient Boosting)

ŞEKİLLER LİSTESİ

Şekil 2.1. Kooperatifler tarafından kullanılan Ekoop uygulaması	8
Şekil 2.2. Kömür ürünü grupları	11
Şekil 2.3. Makine Öğrenmesi metotları	15
Şekil 2.4. Karar ağacı süpermarket örneği	17
Şekil 4.1. Tarım Kredi Kooperatifleri Kömür alım iş süreçleri	31
Şekil 4.2. Tahmin çalışması akış diyagramı	32
Şekil 5.1. 0. ve 1. derece otokorelasyon grafiği	39
Şekil 5.2. SARIMAX tahmin grafiği	40
Şekil 5.3. SARIMAX detaylı tahmin grafiği	41
Şekil 5.4. Holt-Winters tahmin grafiği	42
Şekil 5.5. Holt-Winters detaylı tahmin grafiği	43
Şekil 5.6. FB-Prophet tahmin grafiği	44
Şekil 5.7. FB-Prophet detaylı tahmin grafiği	44
Şekil 5.8. XGBoost tahmin grafiği	46
Şekil 5.9. XGBoost detaylı tahmin grafiği	46
Şekil 5.10. XGBoost parametrelili tahmin grafiği	48
Şekil 5.11. XGBoost parametrelili ve detaylı tahmin grafiği	48
Şekil 5.12. SARIMAX detaylı tahmin grafiği	50

TABLolar LİSTESİ

Tablo 2.1. Seçilen kooperatifler ve ortak sayıları	7
Tablo 2.2. Tarım Kredi Kooperatifleri 2020 yılı ürün satış rakamları	9
Tablo 2.3. Tarım Kredi Kooperatifleri 2016-2020 yılları toplam kömür satışları ..	10
Tablo 2.4. Kömür ürünü alt grupları	11
Tablo 2.5. MAPE değerlerinin sınıflandırılması	25
Tablo 4.1. Veri seti örneği	34
Tablo 5.1. ADF birim kök testi sonuçları	39
Tablo 5.2. SARIMAX yöntemi AIC testi sonuçları	39
Tablo 5.3. SARIMAX tahmin çalışması test sonuçlarının karşılaştırılması	41
Tablo 5.4. Holt-Winters tahmin çalışması test sonuçlarının karşılaştırılması	43
Tablo 5.5. FB-Prophet tahmin çalışması test sonuçlarının karşılaştırılması	45
Tablo 5.6. XGBoost tahmin çalışması test sonuçlarının karşılaştırılması	47
Tablo 5.7. XGBoost parametrelili tahmin çalışması test sonuçlarının karşılaştırılması	49
Tablo 5.8. Modellerin aylık frekans ile sonuçlarının karşılaştırılması	50
Tablo 5.9. Modellerin 1 ve 2 yıllık toplam satış rakamları	51

ÖZET

Anahtar kelimeler: Satış tahmini, Zaman Serisi, XGBoost, SARİMAX, FB-Prophet, Holt-Winters

Şirketlerin ileriye yönelik doğru kararlar verebilmesi için geçmiş verilerden yararlanmalarına birçok durumda ihtiyaç vardır. Günümüzde yöneticiler geçmiş verilere rahatlıkla ulaşabilmektedir. Firmaların iş süreçlerinde kullanılan gelişmiş kurumsal programlar sayesinde geçmiş kayıtları saklayan büyük veri tabanları oluşmuştur. Satış işlemi firmaların en önemli süreçlerinden biridir. Satış kayıtlarına ulaşmak ve bu kayıtları tahmin amaçlı kullanmak mümkündür. Ürün satışı yapan şirketlerde satılan ürünün tahmininin doğru yapılması, şirket kaynaklarının etkin kullanılmasını sağlayarak maliyetleri düşürmekte ve karlılığı artırmaktadır. Kömür ürünü tonlarca satışı gerçekleşen depolama maliyeti yüksek bir üründür. Türkiye’de kış aylarında ısınma amaçlı kullanımı fazladır. Yaz aylarında gereksiz depolanması maliyetleri artırmakta, ayrıca kömür ürünü gramaj kaybederek satıcıyı zarar ettirmektedir. Kömür ürününün satış tahmininin doğru yapılması firmanın zarar etmesini önlemektedir.

İstatistik alanında kullanılan zaman serisi tahmin yöntemleri satış tahmin çalışmalarında yaygın olarak kullanılmaktadır. SARİMAX ve Holt-Winters yöntemleri çalışmalarda tercih edilen popüler zaman serisi tahmin yöntemleridir. FB-Prophet, Facebook tarafından geliştirilmiş yeni bir zaman serisi tahmin yöntemidir. FB-Prophet yıllık, haftalık ve günlük değişen zaman serisi verilerini tahmin etmek için kullanılır. Satış tahmin çalışmalarında makine öğrenmesi algoritmaları da kullanılmaktadır. XGBoost, son zamanlarda yapılan makine öğrenmesi tabanlı çalışmalarda öne çıkan ve tahmin çalışmalarında da kullanılan bir algoritmadır.

Bu tezin amacı Tarım Kredi Kooperatifleri Sakarya Bölge Birliği’ne bağlı 21 kooperatif tarafından yapılan kömür satışlarının tahminini yapabilmek için ideal tahmin yönteminin bulunmasıdır. Çalışmada 2014-2022 yılları arasındaki kömür ürünü satış verileri kullanılarak yıllık ve 2 yıllık toplam kömür satış miktarı ton olarak tahmin edilmiştir. Tahmin yöntemi olarak SARİMAX, Holt-Winters, FB-Prophet zaman serisi yöntemleri ve XGBoost makine öğrenme algoritması kullanılmıştır. Çalışmada kullanılan yöntemlerin performansları MAE, MAPE, MSE, RMSE, R-Kare, ölçüm yöntemleri ile karşılaştırılmış, XGBoost ve SARİMAX yöntemleri ile uzun süreli tahmin sonucunda başarılı sonuçlar elde edilmiştir. XGBoost yönteminin destekleyici parametreler ile kullanılması diğer yöntemlerden daha iyi sonuçlar vermiştir.

SALES PREDICTION USING TIME SERIES AND XGBOOST ALGORITHM: APPLICATION OF AGRICULTURAL CREDIT COOPERATIVES OF TURKEY

SUMMARY

Keywords: Sales Estimation, Time Series, XGBoost, SARIMAX, FB-Prophet, Holt-Winters

In company management, in order to make the right decision, using historical data is a successful methodology. Today, managers can easily access historical data. Thanks to the advanced enterprise programs used in the business processes of the companies, large databases that store historical records have been created. The sales process is one of the most important processes of companies. It is possible to access sales records and use these records for forecasting. Estimating the amount of product sales in retail stores, increases profitability by ensuring the effective use of company resources. Coal is a product with a high storage cost, which is sold in tons. Coal is widely used for heating in Turkey during the winter months. In addition to the cost of storage, the coal loses weight, making sellers losses in the summer. Accurate sales forecast prevents the company from making a loss.

Time series methods used in statistics are widely used in sales forecasting studies. SARIMAX and Holt-Winters methods are popular time series estimation methods, preferred in studies. FB-Prophet is a new time series forecasting method developed by Facebook. FB-Prophet is used to forecast annual, weekly and daily varying time series data. Machine learning algorithms are also used in sales forecasting studies. The XGBoost algorithm is a new algorithm that is preferred in recent machine learning competitions and gives successful results. The XGBoost algorithm is also used in forecasting studies.

The aim of this thesis is to find the ideal method in order to estimate coal sales made from 21 cooperatives affiliated to Sakarya Regional Association of Turkish Agricultural Credit Cooperatives. In this paper, yearly and biennial sales amount of coal are estimated in tons using total sales values between 2014-2022. SARIMAX, Holt-Winters, FB-Prophet time series methods and XGBoost machine learning algorithm are used as estimation methods. The estimation performances of the methods used in the study were compared with MA, MAPE, MSE, RMSE, R-Square performance measurement methods. Successful results have been obtained as a result of long-term prediction with XGBoost and SARIMAX methods. Using the XGBoost method with supporting parameters gave better results than other methods.

BÖLÜM 1. GİRİŞ

Firmaların varlıklarını sürdürebilmeleri için değişen koşullara uyum sağlamaları gerekmektedir. Aksi durumda rakip firmalar zaman içerisinde daha hızlı değişim sağlayarak rekabet avantajı elde edebilir. Günümüzde bilişim sektöründe olan gelişmeler sonucunda işletmelerin bilgi teknolojileri departmanları değişimin sağlanmasında önemli birim durumuna gelmiştir. Firmalarda dijital dönüşüm yaşanmıştır. Yazılımların iş dünyasında kullanımı yaygınlaşmıştır. Özellikle kurumsal kaynak planlama olarak isimlendirilen ve içerisinde üretim, tedarik, maliye, muhasebe, insan kaynakları ve birçok departmanın iş süreçlerini içeren ERP uygulamalarının kullanılması 250 ve üstü personel sayısına sahip büyük ölçekli işletmeler için kaçınılmaz olmuştur. Firmaların faturaları, müşteri bilgileri, satış rakamları, stoktaki ürünleri ve birçok bilgi veri tabanlarında kayıt altına alınmıştır. ERP ve benzeri gelişmiş muhasebe yazılımlarının kullanımının yaygınlaşması ve uzun süreli kullanımı sonucunda firmaların veri tabanlarında büyük veri olarak adlandırılan milyonlarca kayıt içeren tablolar oluşmuştur.

Firmalara ait büyük verilerin ortaya çıkması, veri bilimi yöntemleri ile şirketin yararına kullanılan çalışmalar yapılmasına ve çalışmalar neticesinde verilerden anlamlı sonuçlar çıkarılmasına olanak sağlamıştır. Veri bilimini kullanarak elde edilen raporların firmaların yönetiminde kullanılması yöneticilerin doğru kararlar vermesini sağlayarak firmalara rekabet avantajı sunmuştur. Firmanın sattığı ürünlerin tahminini doğru yapması, firmanın ürün portföyünün yenilenmesi, pazarlama faaliyetlerine yeterli bütçenin ayrılması, üretim faaliyetlerinin arttırılması gibi birçok konuda yöneticilerin doğru kararlar vermesini sağlamaktadır [1].

Firmalar için kilit iş süreçlerinden biri satış işlemleridir. Satış işlemi sonucunda firmalar para kazanmakta ve gelir elde etmektedir. Satış süreçlerinin iyileştirilmesi ve

geliştirilmesi için satış tahmin çalışmalarının başarılı yapılması gerekmektedir. Özellikle perakende sektöründe satış tahminlerinin doğru yapılması, şirketin sattığı ürünleri dağıtıcı ve distribütör firmalarla anlaşmalar yaparken satacağı ürün miktarı ile pazarlık yapma imkânı sağlayarak uygun fiyatlarla ürünü tedarik etmesini sağlamaktadır. Gerekli ürünlerin zamanında tedarik edilmesini sağlayarak stok maliyetlerini düşürmekte ve karlılığı artırmaktadır. Üretim yapan firmalarda üretimin ihtiyaç doğrultusunda yapılmasını sağlamakta, gereksiz stok ve üretim maliyetlerini önlemektedir [2].

Satış tahmin çalışmaları firmalar için finansal faydasından dolayı istatistik, veri analizi, makine öğrenmesi alanlarında araştırılan konulardan biri olmuştur [3]. İstatistik alanında satış tahmin süreçlerinde zaman serisi tahmin yöntemleri yaygın olarak kullanılmaktadır. Zaman serileri geçmiş verilerin gözlem değerlerinden oluşmaktadır. Zaman serileri tahmini farklı disiplinlerde yapılan çalışmalarda tercih edilen ve başarılı sonuçlar veren bir yöntemdir. Satış tahmin süreçlerinde yapılan birçok çalışmada zaman serisi tahmin yöntemleri yaygın olarak kullanılmaktadır [4,5]. Makine öğrenmesi alanında algoritmaların satış tahmin çalışmalarında kullanımı yaygınlaşmakta ve yeni yöntemler bulunmaktadır. Artırma (Boosting) yöntemleri ile farklı modelleri birleştirerek çalışan karma modeller oluşturulmuş ve başarılı sonuçlar alınmaya başlanmıştır. Yüksek işlemci ihtiyacı duyan bu modellerin çalışma performansında günümüz işlemcileri ile başarılı çalışan yöntemler geliştirilmiştir. XGBoost yöntemi de 2016 yılında Gradyan artırma (Gradient boosting) yönteminin iyileştirilmesi ile ortaya çıkmış ve satış tahmin çalışmalarında da kullanılan bir algoritmadır.

Çiftçiye perakende satış yaparak faaliyet gösteren Tarım Kredi Kooperatifleri Türkiye’de tarım alanında faaliyet gösteren en büyük kuruluşlardan biridir. 1625 Kooperatif 17 bölge birliği ve 1 merkez birliği ile bir milyona yakın aktif ortağına hizmet vererek faaliyet göstermektedir [6]. Tarım Kredi Kooperatiflerinde en çok satılan ürünlerden biri kömür ürünüdür. Kömür pahalı bir ürün olması, çok fazla yer kaplayarak depo maliyetinin artışına sebep olması, döneme bağlı olarak ihtiyacın değişmesi ve bulunduğu koşullarda ısıya bağlı olarak nem kaybı ile gramaj kaybına

yol açarak satıcıyı zarar ettirmesi sebebiyle satış tahmininin başarılı yapılması gereken bir üründür [7]. Türkiye’de 2018 yılında yerli ve yabancı toplam 122 milyon ton kömür satılmaktadır. Isınma amaçlı satılan kömürün oranı %8,9 (10,85 milyon ton) olarak tespit edilmiştir. Tarım Kredi Kooperatiflerinde 2018 yılında satılan kömürün miktarı 363 bin tondur. Türkiye’de ısınma amaçlı olarak satılan kömür ürününde Tarım Kredi Kooperatiflerinin payı %3,34 oranındadır [8]. Tarım Kredi Kooperatiflerinin satış rakamları Türkiye’nin satış rakamlarını yansıtacak büyüklükte olduğu görülmektedir. Isınma amaçlı olarak kullanılan kömür ürünü kullanımının hava sıcaklığına ve coğrafi konuma göre değişmesinden dolayı bu ürüne yönelik tahmin çalışmanın bölgeye veya ülkeye özgü olması gerekmektedir. Kömürün piyasadaki konumu ve önemine rağmen Türkiye’de kömür ürünü satışlarına yönelik tahmin çalışması bulunmamaktadır.

1.1. Araştırmanın Amacı ve Konusu

Türkiye’de kurumlara ait veri tabanları kullanılarak az sayıda tahmin çalışması yapıldığı görülmektedir. Yapılan araştırmalar çerçevesinde, özellikle kömür ürün satışlarına yönelik tahmin çalışması bulunamamıştır. Kömür ürününü satan firmalar, sattıkları ürünü miktarına göre sezon öncesinde dağıtıcı firmalarla anlaşmalarını yapmak zorundadır. Yıllık kömür miktarının düşük tahmin edilmesi firmanın dağıtıcı firmalarla yapacağı pazarlık gücünü zayıflatacaktır. Satılan ürün miktarına göre uygun fiyatlarla alım yapılması sağlanmaktadır. Satılacak ürün miktarından fazla rakamlarla anlaşma yapılması, kömür ürününün sıcaklık ile ağırlık kaybeden ve stok maliyeti yüksek bir ürün olmasından dolayı firmanın zarar etmesine sebep olacaktır.

Bu çalışmada Türkiye Tarım Kredi Kooperatiflerine bağlı 1622 kooperatifin kömür ürünü satışlarının XGBoost makine öğrenmesi yöntemi ve SARIMAX, Holt-Winters, FB-Prophet zaman serisi tahmin yöntemlerini kullanarak satış rakamlarının tahmin edilmesi amaçlanmıştır. Çalışmada kullanılan tahmin yöntemlerinin araştırılması ve kullanılan yöntemlerin performansları karşılaştırılarak ideal yöntemin bulunması amaçlanmıştır. Çalışmanın neticesinde aşağıdaki konularda sonuç elde edileceği öngörülmüştür.

- 1) Bulunan tahmin yöntemi ile Tarım kredi Kooperatifleri ve diğer kömür satışı yapan firmaların yıllık kömür satış miktarlarını başarılı tahmin ederek tedarik anlaşmalarını satış miktarına yakın rakamlarla yapması sağlanacaktır. Firmaların satacakları ürün miktarından fazla rakam ile anlaşarak stok ve ürün kaybı durumlarında oluşacak zarar önlenecektir. Firmaların satacakları ürün miktarından düşük rakamlar ile anlaşması sonucu uygun olmayan fiyat ile tedarik anlaşması yapması önlenecektir.
- 2) Bulunan tahmin yöntemi ile kooperatiflerin ihtiyacı olan stok miktarı tespit edilerek kömür dağıtımının hangi zaman aralığında hangi miktarda yapılması gerektiği tespit edilecektir.
- 3) İhtiyacı sezona, ısıya ve koşullara göre değişen kömür ürününe benzer ürünleri ve enerji alanındaki ürünleri tahmin etmek amacıyla yapılacak çalışmalarda hangi yöntemlerin tercih edilmesi gerektiği belirlenecektir. Benzer alanlarda faaliyet gösteren farklı firmaların çalışmanın sonuçlarından yararlanarak satış tahmin amaçlı çalışmalar yapması mümkündür.
- 4) Çalışmada bulunan tahmin yönteminin Tarım Kredi Kooperatiflerinin tüm bölgelerinde uygulanması durumunda önümüzdeki yıllarda firmanın yapacağı kömür satışları toplamının tahmin edilmesi mümkündür. Tarım Kredi Kooperatiflerinin kömür ürünü satışlarındaki pazar büyüklüğü düşünüldüğünde (%3,34) yapılan çalışmanın Türkiye'nin önümüzdeki yıllarda ihtiyacı olacak kömür ürünü miktarı tahmin yöntemi belirlenecektir.

1.2. Tezin Yapısı

Çalışma altı bölümden oluşmaktadır. Giriş bölümünde Tarım Kredi Kooperatiflerinden ve kömür ürününün özelliklerinden bahsedilmiştir ve çalışmanın amacı hakkında bilgi verilmiştir. Temel Kavramlar ve Yöntemler bölümünde kullanılan veri kaynağı, zaman serisi yöntemleri, makine öğrenmesi algoritmaları ve yöntemlerin karşılaştırılmasında kullanılan performans ölçüm yöntemleri hakkında

bilgiler verilmiştir. Kaynak Araştırması bölümünde literatür taraması yapılmıştır. Makine öğrenmesi ve zaman serisi yöntemleri kullanılarak yapılan çalışmalar araştırılmış ve özellikle satış tahmin alanında yapılan çalışmalardan bahsedilmiştir. Güncel Yöntemler ile Kömür Satışlarının Tahmin Edilmesi bölümünde çalışmanın süreçleri hakkında bilgi verilmiştir. Araştırma Bulguları bölümünde çalışmada yapılan tahminlerin sonuçları, zaman serisi yöntemlerinin performansları, yöntem bölümünde bahsedilen ölçüm yöntemlerini kullanarak karşılaştırılmıştır. Son bölümde yapılan çalışma sonuçlarının değerlendirilmesi yapılmıştır.



BÖLÜM 2. TEMEL KAVRAMLAR VE YÖNTEMLER

Bu bölümde öncelikle veri setinin ve tahmin çalışmasının yapıldığı satış süreçlerinin anlaşılması için temel kavramlar hakkında bilgi verilmiştir. Ayrıca çalışmada kullanılan tahmin yöntemleri açıklanmış ve bu yöntemlerin tanımları yapılmıştır. Son olarak kullanılan yöntemlerin analizini yapmak ve çalışmanın sonucunu belirlemek için kullanılan analiz yöntemleri anlatılmıştır.

2.1. Temel Kavramlar

Satış tahmin süreçlerinde kullanılan veri setinin oluşturulması için satış yapılan firma ve ürünün incelenmesi gerekmektedir. Tarım Kredi Kooperatifleri, kooperatiflerin organizasyon yapısı ve çalışma yöntemleri anlatılmıştır. Ayrıca çalışmada kullanılan kömür ürünü hakkında bilgi verilmiştir. Kömür ürünü grupları satış amaçlarına göre sınıflandırılmış ve çalışmada kullanılan ürün grupları anlatılmıştır.

2.1.1. Tarım Kredi Kooperatifleri

Tarım kredi Kooperatifleri 10 bin üzeri çalışanı ve 17 Bölge birliği 1625 kooperatifi ile tarım alanında hizmet veren Türkiye'nin en eski kurumlarından biridir. Tarım Kredi Kooperatifleri ilk olarak 1863 yılında Memleket Sandıkları adı ile çiftçinin kredi ihtiyacının ve iş birliğinin sağlanması, köylünün zararına alışveriş yapmasının engellenmesi ve halkın tefecilerden kurtarılması amacıyla kurulmuştur. Tarım Kredi Kooperatifleri ismini 1936 yılında almıştır. Günümüze kadar varlığını sürdürmektedir. İçlerinde Gübretaş, TK Market ve Bereket Sigorta gibi büyük firmaların bulunduğu 17 farklı iştiraki ile TK Holding altında birleşerek faaliyet göstermektedir. Firma çiftçilere gübre, mazot, yem ve birçok tarım sektöründeki ürününü üreterek veya tedarik edilmesini sağlayarak aynı krediler ile satışını gerçekleştirilmekte ve nakdi olarak krediler vermektir [9].

2.1.2. Tarım Kredi Kooperatifleri Sakarya Bölge Birliği

Yapılan çalışmada tahmin işleminin başarılı sonuçlar vermesi için veri kümesinde kullanılan kooperatifler coğrafi olarak yakın konumda olanlar arasından seçilmiştir. Kooperatif seçiminde öncelikle bölgesel anlamda bir seçim yapılmış ve çalışmada Sakarya Bölge Birliğine bağlı kooperatiflerin satış rakamları kullanılmıştır. Sakarya Bölge Birliği'ne bağlı kooperatifler Türkiye'nin kuzey enleminde olmasından dolayı diğer bölgelere göre ısınma amaçlı kömür ürünü satışı fazladır. Sakarya Bölge Birliği'ne bağlı Bolu, Kocaeli, Sakarya, Düzce, Karabük, Bartın, Yalova, Zonguldak ve İstanbul (Anadolu) illerinde toplamda 148 kurum bulunmaktadır. Çalışmada toplam 25431 ortağı bulunan 21 kooperatif seçilmiş, seçilen kooperatiflerin bilgiler ve ortak sayıları Tablo 2.1.'de gösterilmiştir.

Tablo 2.1. Seçilen kooperatifler ve ortak sayıları

Kooperatif Adı	Ortak Sayısı
2845 Sayılı Kozluk TTK	615
2020 Sayılı Hürriyet TTK	1308
275 Sayılı Akyazı TTK	2832
1033 Sayılı Kayalarreşitbey TTK	1498
1091 Sayılı Karasu TTK	1892
2820 Sayılı Kurudere TTK	943
457 Sayılı Sapanca TTK	801
2821 Sayılı Kestanepınarı TTK	981
490 Sayılı Arifiye TTK	915
651 Sayılı Hendek TTK	2968
2868 Sayılı Osmanbey TTK	675
2019 Sayılı Limandere TTK	1478
2842 Sayılı Gökent TTK	1100
1936 Sayılı Ağacık TTK	551
1835 Sayılı Sinanoğlu TTK	594
2720 Sayılı Ferizli TTK	1226
1020 Sayılı Kazımpaşa TTK	1118

Tablo 2.1. Devamı

579 Sayılı Budaklar TKK	1333
2826 Sayılı Kulaklı TKK	1409
2545 Sayılı Darıçayırı TKK	1128
1935 Sayılı Salmanlı TKK	681

2.1.3. EKOOP (Türkiye Tarım Kredi Kooperatifleri ERP uygulaması)

Tarım Kredi Kooperatifleri iş süreçlerinde verimliliği sağlamak amacıyla iş süreçlerinde EKOOP isimli ERP uygulamasını kullanmaktadır. Tüm kooperatif çalışanların gün içerisinde işlemlerini EKOOP uygulaması üzerinde yapmaktadır. Kurumun muhasebe süreçleri, fatura, irsaliye işlemleri, stok giriş ve çıkış işlemleri, ortaklara yapılan satış, kredi ve sigorta işlemleri, personel bordrolarının oluşturulması, çalışan izinler gibi insan kaynakları süreçleri, HAYBIS ve TAKBIS, ÇKS uygulamaları ile entegreli olarak çalışan ortakların kredi limitlerinin tespit işlemleri ve kuruma ait birçok iş süreci EKOOP uygulaması üzerinden yapılmaktadır. EKOOP uygulaması örnek ekran görüntüsü Şekil 2.1. üzerinde gösterilmiştir.

Şekil 2.1. Kooperatifler tarafından kullanılan Ekoop uygulaması.

Çalışmada kullanılan veriler Tarım Kredi Kooperatiflerinin ERP uygulaması EKOOP üzerinden elde edilmiştir. Algoritmalarda çalıştırılmak üzere tüm kömür üst ürününe bağlı satış verileri kullanılmıştır. Zaman serisi algoritmalarının çalıştırılabilmesi için

satılan ürünün ton olarak miktarı ve satış tarihi ile yıl ve ay bilgisini içeren satış tarihi bilgileri ile çıktı elde edilmiştir.

2.1.4. Kömür ürünü

Kömür ürünü %28 kullanım oranı ile dünyada en çok tercih edilen enerji kaynaklarından biridir. Katı fosil yakıt olarak bilinmektedir. Taş kömürü, linyit kömürü ve birçok farklı çeşidi ile perakende satışı olan bir üründür. Tarım Kredi Kooperatiflerinde gübre, mazot, yem ürünlerinin haricinde kooperatifin en fazla sattığı ürünlerden biri kömür ürünüdür. Tablo 2.2.'de 2020 yılında Tarım Kredi Kooperatiflerinde yapılan satış rakamları gösterilmiştir.

Tablo 2.2. Tarım Kredi Kooperatifleri 2020 Yılı Ürün Satış Rakamları [10].

Ürün Grupları	Satış Miktarı	Satış Tutarı (TL)
Kimyevi Gübre (Ton)	1,458,211	3,037,690,484
Motorin (Lt.)	232,500,309	1,146,224,648
Tohumluk (Ton)	190,484	835,364,521
Diğer Yemler (Ton)	160,553	190,084,927
Kömür (Ton)	358,980	321,782,569
Büyükbaş (Adet)	1.026	10,899,435
Bitki Koruma Ürünleri (Adet)	6,776,129	292,198,587
Diğer Bitki Besleme Ürünleri (Adet)	263,057,209	295,549,201
Fideler ve Fidanlar (Adet)	85,201,631	76,909,115
Küçükbaş (Adet)	5,913	10,189,486
Diğer Hayvanlar (Adet)	396,113	9,022,576
Tüm Ürün Grupları		11.938.301.893

Tablo 2.3.'de tüm kömür ürünlerinin 2016-2020 yılları arasında Türkiye Tarım Kredi Kooperatiflerinin toplam 1208400 satış verisinin toplanması ile elde edilmiştir. Görüldüğü gibi 2016-2020 yılları arasında yıllık 300 bin tonun üzerinde satış rakamına ulaşabilmektedir.

Tablo 2.3. Tarım Kredi Kooperatiflerinin 2016-2020 yılları toplam kömür satışları

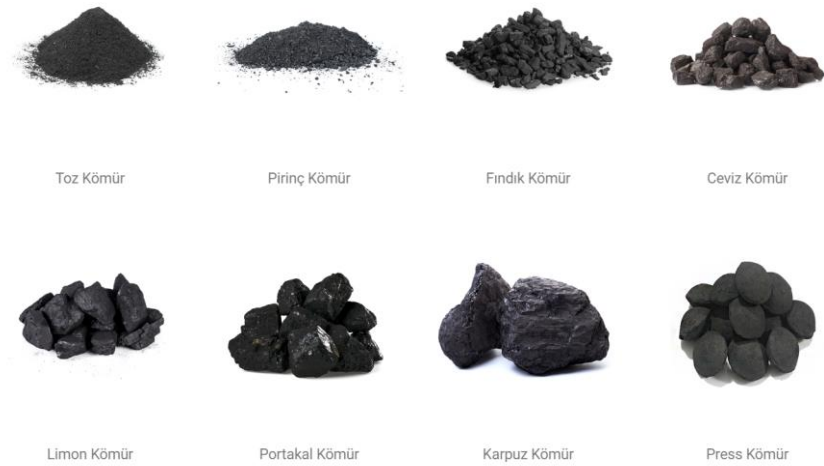
Kömür Satışları	Yıllara Göre Satış Miktarı (TON)				
Çeyrekler	2016	2017	2018	2019	2020
1	64056	113012	76606	75162	65683
2	12453	17442	20049	23689	22059
3	67589	71312	92152	76352	93799
4	277669	300697	174313	176292	197098
Toplam	421758	502463	363120	351495	378639

Tarım Kredi Kooperatiflerinde yerli ve yabancı olarak 16 çeşit kömür ürünü satılmaktadır. Çalışmada kullanılan kömür ürününün satış rakamlarında mevsimsel olarak tutarlı rakamların elde edilmesi için satılan kömür ürünün ısınma amaçlı olarak kullanılması gerekmektedir. Fakat Tarım Kredi Kooperatiflerinde mangal kömürü gibi ısınma amacı dışında kullanılan kömür ürünlerinin de satışı yapılmaktadır. Fabrikalara ve kümeslere satılan kömür ürünlerinin satışı her mevsim mümkündür. Tutarlılığın sağlanması için kooperatiflerden yapılan kömür ürünleri arasından sadece ısınma amacı ile satılan kömür grupları tespit edilmiştir. Genellikle küçük boyutlu kömür ürünlerinin satışı şahıslara yapılmakta ve ısınma amacı ile kullanılmaktadır. Çalışmada fındık, limon, biriket, ceviz, erik, nohut, linyit kömür ürünü grupları seçilmiştir. Kömür ürünleri gruplarının bilgileri ve kullanım amaçları Tablo 2.4.'de özetlenmiştir [11].

Tablo 2.4. Kömür ürünü alt grupları

Kömür Grupları	Kullanım Amacı
İthal fındık kömür	Kat kaloriferi, kümes
İthal limon kömür	Kat kalorifer
İthal portakal kömür	Fabrika, kat kaloriferi
İthal biriket kömür	Soba
İthal ceviz kömür	Soba
İthal karpuz (kelle-parça) kömür	Fabrika
İthal toz kömür	Soba
İthal ayva kömür	Fabrika, kat kaloriferi
İthal erik kömür	Soba
İthal nohut kömür	Soba
İthal linyit kömür	Fabrika
İthal pirinç kömür	Soba
Yerli linyit kömürü	Fabrika
Yerli taş kömürü	Isınma amaçlı kullanılmaz
Yerli toz kömür	Soba

Kömür ürünü gruplarına ait görsel Şekil 2.2.'de gösterilmiştir.



Şekil 2.2. Kömür ürünü grupları [12].

2.2. Zaman Serisi Kavramı ve Tahmin Yöntemleri

Tahmin süreçlerinde çözülmek istenen problemler zamana bağlı sistematik olarak değişebilir. İnsan davranışları, hava değişiminin etkisi ile yılın aynı dönemlerinde benzerlik gösterebilir. Yazın sıcaklık artışına bağlı deniz turizminin hareketlenmesi, tarımsal ürün çeşitliliğinin artması, kışın ısınma ihtiyacına bağlı olarak doğalgaz ve kömür tüketiminin yükselmesi her yıl gerçekleşen durumlardır. Hava sıcaklığında değişimin sürekli ve benzer olması insan davranışlarının değişimine sebep olmakta, insan davranışlarının değişimi ise birçok sistemde veri değişimine sebep olmaktadır. Tekrar eden verileri belirli zaman diliminde gözlemlenmesi ve kayıt altına alınması zaman seriler olarak adlandırılmaktadır [13]. Zaman serilerinin tahminine yönelik çalışmalarda zaman serisi yöntemleri kullanılmaktadır. Arima, SARIMAX, Holt-Winters, FB-Prophet modelleri yaygın olarak kullanılan zaman serisi yöntemleridir. Zaman serisi modellerinin tahmin amaçlı kullanım alanı oldukça geniştir.

2.2.1. ARİMA (Box-Jenkins yöntemi)

Bütünleşik Otoregresif Hareketli Ortalama (ARIMA) yaygın olarak kullanılan popüler zaman serisi modelidir. ARIMA modeli değişkenler ile esnek kullanımı AR, MA ve ARMA modellerinin kombinasyonu olarak düşünülebilir. ARIMA modelini anlamak için öncelikle bu modelleri anlamak gerekir.

2.2.1.1. AR modeli

AR modeli otoregresif ismi ile ifade edilir. Geçmiş dönemlerle olan arasındaki ilişkiyi ölçmek için kullanılır. AR modeli denklem 2.1'de gösterilmiştir. Denklem 2.1'de gözüken ϕ sembolü AR modelinin parametresini ifade eder. $X_t = t$ dönemine ait gözlem değeridir. p değeri ϕ parametresinin derecesini ifade eder. Y_t Sembolü t dönemine ait hata terimini ifade etmektedir.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Y_t \quad (2.1)$$

2.2.1.2. MA modeli

MA modeli hareketli ortalamalar ismi ile ifade edilir. Belirli zaman aralığında değerlerin ortalaması ile hesaplanır. MA modelinin gösterildiği denklem 2.2’de $\emptyset$ sembolü modelin parametresini, q değeri θ parametresinin derecesini ifade eder.

$$Y_t = \emptyset_0 Y_t + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \dots + \emptyset_q Y_{t-q} \quad (2.2)$$

2.2.1.3. ARMA modeli

ARMA modeli AR otoregresif ve MA hareketli ortalamalar modellerinin birlikte kullanılmasıdır.

2.2.1.4. ARIMA modeli

Denklem 2.3’ de X’ ifadesi X değerinin daha önceki zaman aralığına ait değeri ile olan farkını ifade etmektedir. Denklem 2.3’ de zaman aralığının ifadesi t ile gösterilmiştir. Fark alma derecesi d ile gösterilmiştir. Örnek olarak aylık tekrar eden döngü içerisinde d ifadesi 3 olarak belirtilmiş ise X’ değeri haziran ayında X değerinin haziran ayı ile mart ayı arasındaki 3 aylık farkı ifade etmektedir.

$$X'_t = X_t - X_{t-d} \quad (2.3)$$

ARIMA modeli içerisinde kullanılan değişkenler ARIMA(p,d,q) olarak gösterilir. Otoregresif AR modeli p sembolü ile ve hareketli ortalama MA modeli q sembolü ile temsil edilir. d Sembolü fark alma değişkenidir. ARIMA yönteminin açıklaması denklem 2.4 ‘deki gibi ifade edilir [14].

$$X'_t = \phi_1 X'_{t-1} + \phi_2 X'_{t-2} + \dots + \phi_p X'_{t-p} + Y_t + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \dots + \emptyset_q Y_{t-q} \quad (2.4)$$

2.2.2. SARIMA(SARIMAX)

SARIMA modeli, ARIMA modelinin mevsimsel etkilere göre sonuç vermesi amacıyla geliştirilmiş bir yöntemdir. Python kütüphanesi içerisinde SARIMAX olarak isimlendirilerek kullanılmaktadır. ARIMA'da kullanılan 3 değişkene ek olarak 4 yeni değişkenlere sahiptir ve SARIMAX(p,d,q)(P,D,Q)s olarak gösterilir. P mevsimsel otoregresif, Q mevsimsel hareketli ortalama olarak kullanılmaktadır [4,15,16].

2.2.3. HOLT-WINTERS

Holt-Winters yöntemi zaman serisi problemlerinin çözümünde kullanılan bir yöntemdir. Yeni verilerin tahmin edilmesinde verilere farklı değerler vererek çalışan Üstel Düzleştirme Yönteminin değiştirilmesi ile üretilmiş özgün bir yöntemdir. Winters Üstel Düzleştirme Yöntemi olarak bilinmektedir. Verinin büyümesi ile eski verilerin tahminlenen değeri etkileme derecesi azalmakta yeni verilerin etkisi artmaktadır. Bu yöntem trend, seviye ve mevsimsellik olarak üç farklı formülün birleşimi olarak türetilmiştir ve her formül için kullanılan parametrelerle 3 farklı parametre almaktadır. Kullanması diğer algoritmalara göre daha basittir [17].

2.2.4. Facebook Prophet (FB-PROPHET)

FB-PROPHET algoritması 2017 yılında Facebook tarafından geliştirilen zaman serisi yöntemidir. Python ve R dilleri ile çalışmakta ve günlük, haftalık ve yıllık olarak tekrar eden değişimlere göre tahmin yapmaktadır. Sezon değişimini ve hafta sonlarını dahil ederek çalışmaktadır. FB-PROPHET algoritması mevsimsel trendi otomatik olarak bulmakta ve istatistik bilgisi olmadan gerekli optimizasyonu otomatik hesaplaması sayesinde kolay çalışmaktadır.

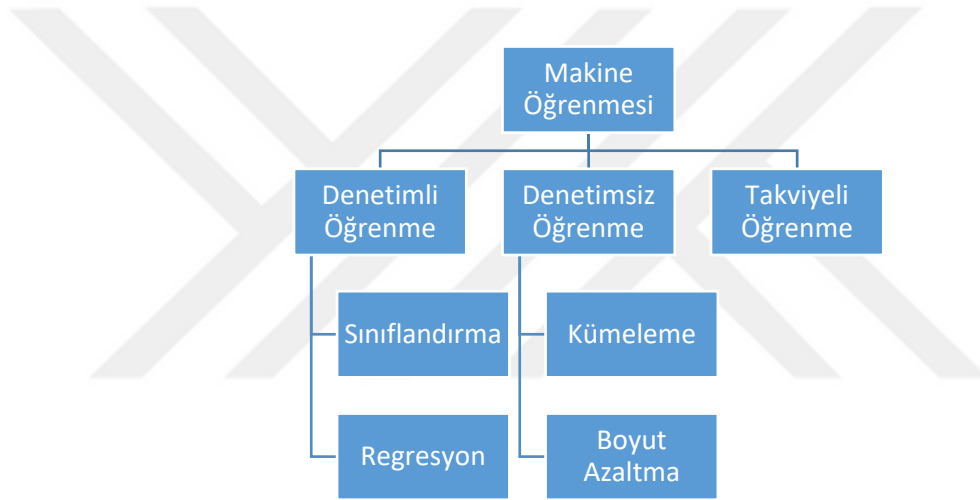
2.3. Makine Öğrenmesi

Öğrenme kavramı geçmiş tecrübelerden ve birikimlerden yararlanarak değişebilme yeteneğidir. Makine Öğrenmesi ise makinelerin değişme yeteneğini arttıran

tekniklerin araştırıldığı mühendislik alanıdır. Makine öğrenmesinin amacı makinenin ilgili verilerden yararlanarak geleceği anlamasını ve başarılı karar vermesini sağlayan matematik modellerinin araştırılması ve geliştirilmesidir [18].

2.3.1. Makine öğrenmesi yöntemleri

Tüm Öğrenme süreçleri genel olarak iki aşamadan oluşur. İlk aşama veri kümesi içerisinde bilinmeyen ilişkilerin tahmin edilmesidir. İkinci aşama tahmin edilen ilişkiler doğrultusunda yeni çıktıların ön görülmesidir. Bu aşamalar Denetimli Öğrenme ve Denetimsiz Öğrenme olarak iki temel öğrenme metodunu oluşturur.



Şekil 2.3. Makine Öğrenmesi metotları [19] .

2.3.1.1. Denetimli öğrenme (Supervised learning)

Denetimli öğrenme belirli veri kümesi içerisinde ilişkisi olduğu düşünülen verilerin ilişkisinin bulunması yöntemidir. İlişki kurulmak istenen veriler içerisinde örnek bir küme ile çalışma yapılır. Örnek kümenin algoritmanın yeterli tahmin performansı ile sonuç üretmesine yetecek seviyede büyük olması beklenir. Örnek küme kullanılarak iki veri grubu arasında ilişki kurulur. Verilerden biri (bağımsız değişken) diğer verinin (bağımlı değişken) değerini belirlediği iddia edilir. Gerçek dünyada denetimli öğrenmeye örnek olarak restoran ve kafelerin sattığı ürünlerden yola çıkarak restoran ve kafe ayrımını yapan bir algoritma düşünülebilir. Elimizde işletmelere ait restoran ve kafe bilgisini ve bu işletmelerde satılan ürünlerin bilgilerini içeren iki liste olduğunu

varsayalım. Denetimli öğrenme yöntemleri ile listede satılan ürünlerin hangisinin restoranda hangisinin kafede satıldığını bilgilerinden yola çıkarak, yeni açılan bir restoranın menüsüne bakarak restoran veya kafe olduğunu anlayan bir algoritma geliştirilebilir. Denetimli Öğrenme yöntemlerinde restoran - kafe ayrımı algorithmadan önce yapılmıştır.

2.3.1.2. Denetimsiz öğrenme (Unsupervised learning)

Denetimsiz öğrenme yönteminde denetimli öğrenme yönteminden farklı olarak öğrenme yöntemini gösteren bir model yoktur. Veri kümesi içerisinde ilişkisi bilinmeyen veriler arasında sonuç bulma çalışmasıdır. Kümeleme çalışmaları denetimsiz öğrenme alanına girmektedir. Veriler arasındaki sınıflar makineye öğretilmez. Algoritmanın veriler arasındaki ilişkiyi kendisinin bulması, verileri kümelerle ayırarak sınıflandırması beklenir [20].

2.3.1.3. Takviyeli öğrenme (Reinforcement learning)

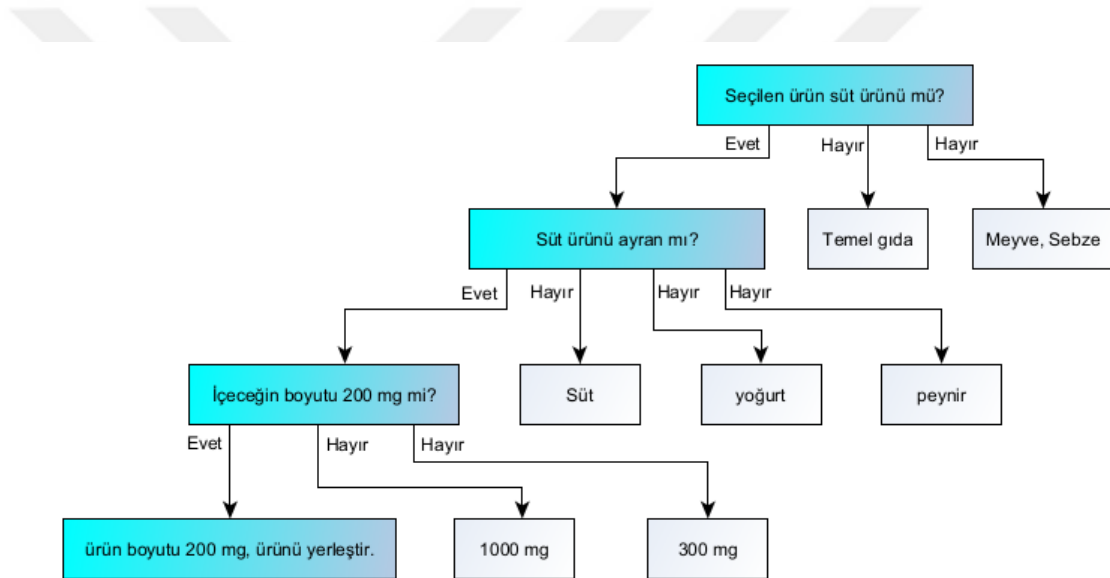
Denetimli ve Denetimsiz öğrenme yöntemlerinden farklı olarak ödül ve ceza sistemine dayanan öğrenme yöntemidir. Makine karşılaştığı problemin çözümüne yönelik kararlar verir. Daha önce verdiği kararlar sonucunda aldığı ödül sayısını en yüksek değere çıkararak bulmaya çalışır. Verilen kararların doğruluğu tek başına hesaplanamaz. Bir satranç oyununda verilen kararın doğruluğu takip eden kararlara bağlıdır.

2.3.2. Karar ağaçları

Problemin çözümü için koşulların oluşturulduğu ve bu koşullar doğrultusunda uygun sonuç kümesi seçilerek çözümün üretildiği makine öğrenmesi yöntemidir. If/else gibi basit koşullu programlama ifadeleri ile oluşturulan fonksiyonlara benzer. Girdi, içerisindeki değerlere göre bir koşuldan geçer. Geçtikleri koşula göre yeni koşullu ifadelere girer. Bu işlem algoritmanın bir sonuç kümesi bulana kadar tekrar edebilir. Süpermarketlerde görevlilerin ürünleri bulunduğu yere yerleştirme işlemi karar

ağaçlarına örnek gösterilebilir. Market çalışanı stoğa giren yeni ürünü raflara yerleştirirken öncelikle ana kategori seçimi yapar. Ürün bir süt ise süt ürünlerinin olduğu alana gider. Ürünün türüne göre (süt, peynir, yoğurt, ayran vb) içeceğin bulunduğu dar alanı bulur. İçeceğin boyutuna (200/500/1000 mg) ve markasına göre ait olduğu alanı bulur ve yerleştirir. Müşteriler de aynı algoritma ile ürünü bulur. Örneğe ait adımlar Şekil 2.4.'de gösterilmiştir.

Küçük veri setlerinde başarılı sonuçlar alınsa da veri boyutu büyüdüğü zaman, veri sınıfları artacaktır. Her veri sınıfı için farklı karar koşulu yazılamayacağı için karar ağaçları büyük verilerde performanslı çalışmamaktadır.



Şekil 2.4. Karar ağacı süpermarket örneği.

2.3.3. Gradyan artırma ağaçları (Gradient boosting tree)

Gradyan Artırma Ağaçları veya Gradyan Artırma Karar Ağaçları (Gradient Boosting Decision Tree) olarak bilinen makine öğrenmesi algoritması günümüzde sınıflandırma ve regresyon işlemlerinde kullanılan ve farklı problemlerde başarılı çalışan bir yöntem olmasından dolayı popüler hale gelen bir algoritmadır [21]. Gradyan artırma ağaçları ilk olarak Friedman tarafından 1999 yılında ortaya atılmıştır. Friedman 2002 yılında modelini sabit boyutlu karar ağaçları kullanarak iyileştirmeler yaparak tekrar düzenlemiştir [22]. Gradyan artırma ağaçları kavramı içerisinde geçen artırma

(Boosting) olarak bilinen yöntem düşük performanslı öğrenme yöntemlerini birleştirerek, daha iyi performanslı öğrenme yöntemi oluşturmayı amaçlar. Alpaydın (2014) Bu yöntem içerisinde ilk modeller sırayla oluşturulur. Tahmin başarısı yeni modeller oluşturulurken iyileştirilir. Her yeni model bir önceki modelin hatalarını düzeltmeyi amaçlar.

Gradyan artırma yöntemini anlamak için algoritma çalışma mantığını kavramak gerekir. Veri seti içerisinde $x(i)$ değerleri ile $y(i)$ değerlerinin tahmin edildiği zayıf bir fonksiyon denklem 2.5' de gösterilmiştir.

$$\text{Tahmin edilen değer} = \hat{y}_i = h(x_i) \quad (2.5)$$

Fonksiyonun çıktısı Tahmin edilen değer olarak düşünülür. Tahmin edilen değer ile gerçek değer arasında residual olarak nitelendirilen bir fark oluşur. Residual (Artık Değer) kayıp fonksiyon olarak kabul edilir ve aşağıdaki gibi ifade edilir.

$$\text{Kayıp fonksiyon} = r_{im} = L(y_i, h(x_i)) \quad (2.6)$$

Y değerini bulmaya çalışan zayıf fonksiyonun y değerini yeterli ifade etmediği düşünülür. Kayıp fonksiyonu kullanarak y değeri ile farkı bulmaya çalışan ikinci bir $h(x)$ fonksiyon oluşturulur. Fonksiyonun ağırlığını belirleyen p değeri kullanılır. Zayıf fonksiyondan daha iyi çalışan yeni bir fonksiyonlar elde edilir.

$$f_{i+1}(x_i) = f_i(x_i) + p_i h_i(x_i) \quad (2.7)$$

Algoritmanın en düşük kayıp fonksiyon değerini kullanarak başlatılır. Eğitim kümemiz içerisinde n adet veri olduğu varsayılır ve artırma yöntemi ilk başta tahmin fonksiyonunu n adet veri için aşağıdaki denklem 2.8 de görüldüğü gibi çalıştırılır.

$$f_0(x) = \arg \min \sum_1^n L(y_i, h(x_i)) \quad (2.8)$$

M adımdan oluşan algoritma için aşağıdaki adımlar sırasıyla uygulanır;

- 1- Resudial değeri hesaplamak için gradyan artırma yöntemi ile denklemin türevi alınır ve yön değiştirilir.

$$r_{im} = - \left| \frac{\partial L(y_i, h(x_i))}{\partial f_0(x)} \right| \quad (2.9)$$

- 2- Resudial değeri ile zayıf fonksiyon $h(x)$ bulunur.

- 3- Bir sonraki adım için tahmin edilen değer $\hat{y}_m$ hesaplanır.

$$\hat{y}_m = \arg \min \sum_1^n L(y_i, f_{m-1}(x_i) + p h_{m-1}(x_i)) \quad (2.10)$$

- 4- Bir sonraki adıma geçmek için model güncellenir.

$$f_m(x_i) = f_{m-1}(x_i) + \hat{y}_m \quad (2.11)$$

Büyük verilerin yaygınlaşması ile yakın zamanda Gradyan artırma yöntemi üzerinde iyileştirmeler yapılmış ve yeni algoritmalar geliştirilmiştir. XGBoost, CatBoost, LightGBM popülerlik kazanmış güncel gradyan artırma yöntemleridir. Gradyan Artırma algoritmasının çalışma presibini kullanarak çalışırlar. Algoritmanın daha hızlı ve başarılı çalışması için geliştirilmiş, performansları veri setine göre değişen güncel yöntemlerdir. XGBoost, Gradyan Artırma yöntemini iyileştiren ilk yöntem olması ve açık kaynak olması sebebi ile kullanımı diğer yöntemlere göre daha yaygındır. XGBoost hakkında daha fazla kaynak bulunması ve algoritmayı kullanan toplulukların fazla olması çalışmalarda avantaj sağlamaktadır.

2.3.4. Aşırı gradyan artırma (XGBoost)

Chen and Guestrin (2016), tarafından geliştirilen Aşırı Gradyan Artırma (XGBoost), Gradyan Artırma algoritmasının temel prensiplerini kullanarak geliştirilmiş denetimli makine öğrenme yöntemidir. XGBoost algoritması sınıflandırma ve regresyon karar ağaçlarından oluşur. Çok sayıda zayıf öğrenme modelini birleştirerek güçlü öğrenme algoritması üretmeyi amaçlar. Zayıf öğrenme modelleri regresyon ağaçlarını oluşturur, algoritmanın başarısı hesaplanır ve residual (hata skoru) iterasyon yöntemi ile bir sonraki ağacın oluşması için kullanılır. Sonuç olarak kullanılan yöntem karmaşık problemlerin çözümünde başarılı sonuçlar vermektedir. XGBoost ve Gradyan Artırma algoritmalarının çalışma yöntemi aynıdır. Fakat XGBoost algoritmasının, Gradyan Artırma ve diğer tahmin algoritmalarına göre performanslı ve başarılı çalışmasını sağlayan özellikleri bulunmaktadır.

Regülerizasyon, XGBoost algoritmasının karmaşık modelleri cezalandırmak için kullanılan bir özelliğidir. Gama parametresi ile iterasyon adımında kullanılan residual değeri ölçeklendirilir. Bu özellik algoritmanın aşırı öğrenmesini engellemek için kullanılır. XGBoost algoritmasının boş değerler ile çalışabilmesi diğer özelliğidir. Boş değerlerin doldurulması veya kaldırılması veri ön işleme adımından yapılması gereken bir işlemdir. Fakat XGBoost algoritması ağacın boş değerler ile karşılaşılan her adımında yeni bir yön oluşturarak çözüm üretir.

XGBoost algoritmasının tercih edilen bir yöntem olmasının en önemli sebeplerinden biri hızlı olmasıdır. Gradyan Artırma ve diğer tahmin yöntemleri ile sonuç elde etmek zaman almaktadır. XGBoost algoritması, Windows ve Linux işletim sistemlerinde paralel işlemler ile çalışarak gradyan artırma yönteminden 10 kat daha hızlı performans göstermektedir. Hafıza kullanımı daha düşüktür. Daha düşük donanım ihtiyacına gereksinim duyarak hızlı ve başarılı sonuç göstermektedir [23].

2.4. Performans Ölçüm Yöntemleri

Algoritmaların başarılarının hesaplanması için belirli ölçüm yöntemleri geliştirilmiştir. Tahmin süreçlerinde başarı durumunu ölçmek için yaygın olarak kullanılan yöntemler bu bölümde özetlenmiştir. Yöntemlerin anlaşılması için formüller kullanılmıştır. Formüllerde bulunan sembollerin açıklamaları belirtilmiştir.

A: Gerçek değer

B: tahminlenen değer

T: zaman periyodu

N: veri setinin büyüklüğü

Formül içerisine A ve B değerlerinin üst çizgi ile kullanımı, değerler ortalamasıdır.

2.4.1. AIC (Akaike bilgi kriteri)

AIC değeri, zaman serisi yöntemlerini içerdikleri parametre karmaşıklığına göre cezalandıran logaritmik bir fonksiyon ile değer üretir. Bu değer düşük olması gerekmektedir. Özellikle parametre optimizasyonu işleminde kullanılır [24].

2.4.2. MAE (Mean absolute error - Ortalama mutlak hata)

MAE hatanın mutlak ortalamasıdır. Gerçek değer ile tahmin sonucu bulunan değerlerin farklarının ortalaması alınarak hesaplanmaktadır. MAE değerinin formülü denklem 2.12'deki gibidir.

$$MAE = \frac{1}{n} \times \sum_{t=1}^n (A_t - B_t) \quad (2.12)$$

2.4.3. MAPE (Mean absolute percentage error - Ortalama mutlak hata yüzdesi)

MAPE değeri hata değerlerinin gerçek değerlere oranları ortalamasının yüzdesidir. MAPE yönteminde başarı oranı Tablo 2.5.'de gösterilen oranlara bakarak hesaplanmaktadır [25].

Tablo 2.5. MAPE değerlerinin sınıflandırılması

Performans	Başarı Durumu
MAPE < %10	Çok iyi
%10 <MAPE < %20	İyi
%20 <MAPE <%50	Kabul edilebilir
MAPE> %50	Yanlış ve hatalı

MAPE değeri denklem 2.13. içerisinde gösterilen formül ile hesaplanmaktadır.

$$MAPE = 100 \times \frac{1}{n} \times \sum_{t=1}^n \frac{a_t - b_t}{a_t} \quad (2.13)$$

2.4.4. MSE-RMSE (Mean squared error - Root mean squared error)

MSE (Hata kareleri ortalaması) gerçek değer ile tahmin edilen değer arasında farkların toplamı karelerinin ortalaması alınarak hesaplanan performans ölçüm yöntemidir. RMSE MSE değerinin kareköküdür. RMSE değerinin sıfıra yakın olması tahmin başarısının yüksek olduğunu göstermektedir. MAPE RMSE yöntemlerinin formülü denklem 2.14 ve denklem 2.15' de gösterildiği gibi hesaplanmaktadır [26].

$$MSE = \frac{1}{n} \times \sum_{t=1}^n (A_t - B_t)^2 \quad (2.14)$$

$$RMSE = \sqrt{MSE} \quad (2.15)$$

2.4.5 R-Kare (Determinasyon katsayısı)

R-Kare regresyon analizinde verilerin tahmin eğrisine uyumunu gösteren ölçüm yöntemidir. Gerçek değer ile tahmin setinde gözlemlenen değerler arasındaki farklar artık değerleri oluşturur. Gerçek değer setinde gözlemlenen değerlerin, gerçek değerler setinin ortalamasına olan uzaklıklarının toplamı, ortlamaya olan uzaklıkları oluşturur. Artık değerlerin karelerinin ortalamaların karelerine oranı 1 'den çıkarılarak R-Kare hesaplanır. R-Kare, 0 ve 1 arasında değer alır. Hesaplanan değer 1'e yakın olması tahmin çalışmasının başarısını gösterir. R-Kare sonuncun başarılı olduğunu gösteren kesin bir değer yoktur. Kabul edilebilir R-Kare değeri, yapılan çalışmaya göre değişmektedir. Finans alanında yapılan çalışmalarda genellikle R-Kare değerinin 0,40 altında olması çalışmanın başarısız olduğu göstermektedir. R-Kare değerinin 0,40 – 0,75 aralığında olması, çalışmanın kabul edilebilir olduğunu göstermektedir. R-Kare değerinin 0,75 üzerinde olması bulunan modelin iyi olduğunu göstermektedir [27].

$$R^2 = 1 - \frac{\sum (B_t - \bar{B})^2}{\sum (B_t - \bar{A})^2} \quad (2.16)$$

BÖLÜM 3. LİTERATÜR ARAŞTIRMASI VE ÖNCEKİ ÇALIŞMALAR

Günümüzde farklı disiplinlerde yapılan tahmin çalışmaları günden güne artmaktadır. Makine öğrenmesi alanında tahmin çalışmalarında uzun zamandır kullanılan lineer regresyon, polinomal regresyon, karar ağaçları gibi eski algoritmalar ile bu algoritmaların temel çalışma prensipleri iyileştirerek çalışan yeni yöntemler geliştirilmiştir. Adaboost (Adaptive Boosting) ve Gradyan Artırma (Gradient Boosting) algoritmaları karar ağaçları mantığı ile çalışan, fakat her adımda çalışma prensiplerini iyileştirerek karma yöntemler uygulayan ve başarılı sonuçlar veren yöntemler olarak ortaya çıkmıştır. İşlemci hızlarının artması, dağıtık sistemler kullanılarak hızlı hesaplamaların yapılması Artırma (Boosting) algoritmalarının kullanılmasına engelleyen kısıtları azaltmıştır. 2016 yılında Gradyan Artırma Algoritmasını güncel teknolojiler ile verimli çalışmasını sağlayan XGBoost algoritması geliştirilmiştir. Hızlı çalışması ve başarılı sonuçlar vermesi özellikleri ile çalışmalarda tercih edilen bir algoritmadır. XGBoost yöntemi tahmin alanında da tercih edilmekte ve satış tahmin alanında da kullanılmaktadır. Bu bölümde Boosting ve XGBoost yönteminin literatürde gelişimi anlatılmış ve satış tahmin alanında başarılı sonuç alınan örnek çalışmalar hakkında bilgi verilmiştir.

Zaman serisi tahmin yöntemleri istatistik alanında kullanılan eski yöntemler olmasına rağmen güncel çalışmalarda kullanılan yöntemlerdir. Arima yöntemi güncel zaman serisi tahmin çalışmalarında tercih edilen ve başarılı sonuçlar veren popüler bir yöntemdir. Eski bir yöntem olan Holt ve Winters tarafından geliştirilen Holt-Winters yöntemi güncel çalışmalarda kullanılmaktadır. Geleneksel zaman serisi tahmin yöntemlerine ek olarak istatistik bilgisi olmadan ve optimizasyon çalışması yapılmadan kullanılma imkânı sunan FB-Prophet yöntemi 2017 yılında Facebook tarafından geliştirilmiştir. Zaman Serisi Yöntemleri bölümünde literatürde zaman serisi tahmin yöntemleri ile yapılan çalışmalar anlatılmış ve satış tahmin alanında

yapılmış çalışmalar Zaman Serisi Yöntemleri ile Yapılan Satış Tahmin Çalışmaları bölümünde örneklendirilmiştir.

3.1. Boosting Yöntemleri ve XGBoost

Artırma (Boosting) yöntemi ilk olarak Kearns ve Valiant tarafından bir dizi zayıf öğrenme algoritmasının sıralı olarak çalıştırılarak güçlü bir algoritma yaratacağı fikri ile ortaya çıkmıştır [28]. Artırma Yöntemi ile yapılan çalışmalar arasından ilk olarak 1995 yılında Freund ve Schapire tarafından geliştirilen Adaboost yöntemi ile sınıflandırma alanında başarılı sonuçlar alınmıştır. Friedman 99 yaptığı çalışmada Adaboost yöntemini güncelleştirerek Gradyan Artırma (Gradient Boosting) yöntemini geliştirmiştir. 2016 yılında yapılan çalışma sonucundan Gradyan Artırma yöntemi iyileştirilerek daha hızlı çalışan XGBoost yöntemi geliştirilmiştir. XGBoost algoritması Kaggle tarafından düzenlenen ödüllü yarışmalarda XGBoost yöntemi farklı problemlerin çözümünde tercih edilen bir algoritmadır [23].

Güncel bir algoritma olmasına rağmen bilimsel çalışmalarda kullanımı yaygınlaşmıştır. Siber atakların sınıflandırılması, kronik hastalıkların tedavisi, su seviyesinin tahmin edilmesi, kişisel kredi skorlaması, rüzgâr türbinlerinin hata tespiti, gibi farklı alanlarda kullanılmaktadır [29,30,31,32].

XGBoost yöntemi satış tahmin amaçlı farklı çalışmalarda kullanılmıştır. Araba satışlarını tahmini için 2020 yılında yapılan bir çalışmada XGBoost yöntemi ile geliştirilen ForeXGBoost yöntemi kullanılmıştır. Kısa süreli yapılan tahmin çalışmasında yüksek başarı oranının elde edildiği görülmüştür [33].

Kurumsal e-ticaret firmasının ürün satışlarını tahmin etmeye yönelik bir çalışmada ARIMA modeli ve XGBoost modeli özellikleri birleştirilerek C-A-XGBoost isimli yeni bir model oluşturulmuş ve başarılı sonuç elde edilmiştir. Çalışmada yapılan sonuçların karşılaştırılmasında ME, MSE, RMSE ve MAE yöntemleri kullanılmıştır. Çalışmada zaman serisi yöntemlerinin sıralı verilerde başarılı olduğu fakat lineer olmayan kısımlarda XGBoost yönteminin avantaj sağladığından bahsedilmiştir [34].

Güncel bir çalışmada 6 farklı makine öğrenmesi yöntemi ile XGBoost yöntemi altın fiyatı değişim hareketlerinin tahmin edilmesinde kullanılmıştır. Çalışmanın başarısının tespitinde MAE, RMSE, R-Kare yöntemleri kullanılmıştır. XGBoost yöntemi diğer yöntemlere göre başarı sonuç verdiği görülmüştür [35].

Güncel bir çalışmada satış tahmin verilerini içeren zaman serisi veri setlerinde Arima, Fb-Prophet ve XGBoost yöntemleri ile tahmin çalışması yapmıştır. Çalışmada kullanılan veri setleri farklı mağazalardan toplanan süt ve çay markalarının satış bilgilerini içermektedir. Çalışma sonucunda XGBoost yönteminin diğer yöntemlerden daha başarılı olduğundan ve diğer yöntemlerin yakın sonuçlar verdiğinden bahsedilmiştir [36].

Kaggle üzerinden alınan süpermarket satış verilerine yönelik farklı bir çalışmada XGBoost ve LSTM yöntemleri ile geliştirilen bir yöntem kullanılmıştır. Çalışmada 2017 -2019 yıllarına ait 78 farklı mağazadan alınan günlük satış verileri kullanılmış ve başarılı sonuç elde edilmiştir [37].

3.2. Zaman Serisi Yöntemleri

Literatürde zaman serisi yöntemleri ile farklı alanlarda yapılmış çalışmalar bulunmaktadır. Konu ve sektör bağımsız olarak yapılan çalışmalarda başarılı sonuçlar alınmıştır. Finans alanında yapılan bir çalışmada borsa fiyatlarının tahmininde ARIMA yöntemi kullanılmıştır. ARIMA modelleri ile alınan sonuçlar karşılaştırılmış ve uygun model seçilerek kısa dönem fiyat tahmini yapılmıştır [38].

R dilini kullanarak yapılan çalışmada Arima ve FB-Prophet algoritmalarını ile bitcoin fiyatının tahmin etmeye çalışmıştır ve 90 günlük fiyat tahmini sonucunun başarılı olduğu görülmüştür [39].

Cep telefonu trafiğinin tahminini yapan çalışmada LSTM ve ARIMA yöntemleri kullanılmıştır. Çalışmada başarılı sonuçların alınmasının veri setinin büyüklüğü ile ilişkili olduğu tespit edilmiştir [40].

Covid 19 yayılım hızının bulunması amaçlayan bir çalışmada ARIMA yöntemi pandeminin farklı safhalarında kullanılarak tahmin çalışması yapılmıştır [41].

Güncel bir çalışmada 2007-2020 yılları arasında enerji talep tahmini yapan çalışmaları ve çalışmalarda kullanılan yöntemleri araştırmış ve zaman serisi yöntemlerinin kullanıldığı çalışmalar özetlemiştir. Birçok çalışmada Arima, Sarima, Holt-Winters yöntemlerinin enerji fiyatlarının tahmin edilmesinde tercih edilen yöntemler olduğu görülmüştür [42].

3.2.1. Zaman serisi yöntemleri ile yapılan satış tahmin çalışmaları

Zaman serisi yöntemleri satış tahmin çalışmalarında ve satış verilerini ilgilendiren üretim tahmin çalışmalarında yaygın olarak kullanılmaktadır. ARIMA yöntemi ile geliştirilen bir model ile Endonezya’da marketlerde satılan soğan ve patates ihtiyacını belirlemede kullanılmıştır. 2007 ocak ve 2010 şubat tarihleri arasında market verileri toplanmış ve günlük talep tahmini yapılmıştır. MAPE ile yapılan performans ölçümü sonucunda başarılı sonuç elde edilmiştir [43].

Satış tahmin alanında yapılan başka bir çalışmada coğrafi olarak farklı konumlarda bulunan Walmart perakende mağazalarının satış verilerini kullanarak mağazanın performansını etkileyen faktörleri araştırmıştır. Holt-Winters yöntemi ile Walmart mağazalarının geçmiş 3 yıllık verilerine bakılarak gelecek 39 haftaya ait satışların tahmini yapılmıştır. Holt-Winters algoritmasının mevsimsel veri tahmin çalışmalarına yatkın olduğu görülmüştür. Çalışmalarda görüldüğü gibi zaman serisi yöntemleri tahmin süreçlerinde başarılı sonuçlar vermekte ve yaygın olarak kullanılmaktadır [44].

Pirinç ürünü fiyatlarının tahminine yönelik yapılan çalışmada 34 farklı ilde seçilen 101 yerel market içerisinde her gün belirli saatlerde fiyat bilgisi toplanarak çalışma yapılmış ve Arima modeli ile başarılı sonuç elde edilmiştir [45].

Yapılan bir çalışmada ARIMA, SARIMAX, Holt-Winters ve birçok zaman serisi yöntemi satış tahmininde kullanıldığı belirtilmiştir. Tahmin yöntemleri birleştirilerek daha başarılı sonuçların alındığı görülmüştür. Satış tahmin çalışmalarında geçmiş yıllara ait veri bulunduğunda zaman serisi yöntemlerinin başarılı sonuçlar verdiği tespit edilmiştir [46].

Güncel bir çalışmada kayısı üretiminin ARIMA yöntemi ile tahmini yapılmıştır. Türkiye'deki 50 yıllık kayısı üretimi verisi kullanılarak uygun ARIMA modeli geliştirilmiş ve 4 yıllık kayısı üretiminin tol olarak tahminini yapılmıştır [47].

BÖLÜM 4. ZAMAN SERİSİ YÖNTEMLERİ VE XGBOOST ALGORİTMASI İLE KÖMÜR SATIŞ TAHMİNİ

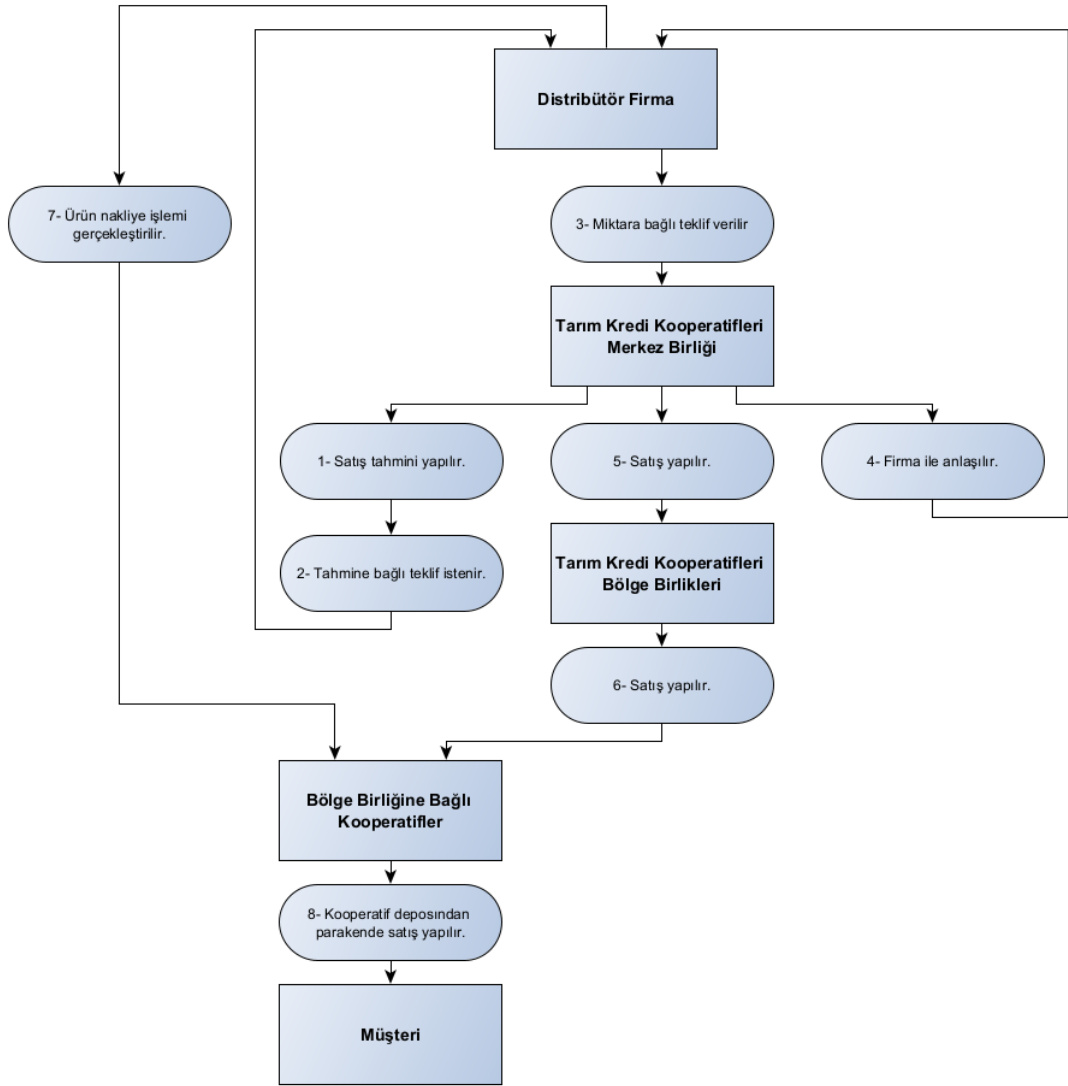
Bu çalışmada Tarım Kredi Kooperatifleri Sakarya Bölge Birliğine bağlı kooperatifler tarafından yapılan kömür satışlarının tahmin edilmesi amaçlanmıştır. Çalışmanın adımları belirlenirken öncelikle mevcut işleyişi incelenmiştir. Mevcut sistemde eksik olan tahmin işleminin hangi adımda hangi aşamada yapılması gerektiği tespit edilmiştir.

4.1. Tahmin İhtiyacının Tespit Edilmesi

Tarım Kredi Kooperatiflerinde satılan ürün kalemleri Merkez Birliği tarafından belirlenmektedir. Kooperatifler tarafından satılacak ürünlerin bilgileri, ürün grupları merkez birliği tarafından belirlenmiş ve Ekoop ERP uygulamasına kaydedilmiştir. Firma finansal olarak büyük olmanın avantajını kullanmak için satılacak ürünlere ait anlaşmaları ihtiyaç duruma göre toplu olarak yapar. Kömür ürünü anlaşmaları sezon öncesinde Merkez Birliği tarafından yapılır. Merkez Birliğine bağlı ilgili birimler önceki yıllardaki satış raporlarını inceleyerek ihtiyaç duyulan ürünleri tespit eder. Satış raporları şirketin önceki yıllarına ait satış ve satışla ilgili veriler kullanılarak hazırlanır, tahmin içermez. Firmanın ihtiyaçları belirlenirken, satış tahmini yaparak yöneticilerin karar vermelerine yardımcı olan bir program veya yöntem kullanılmaz. Firmanın kömür ihtiyacı yöneticiler tarafından belirlenir ve kömür alımı planlanır. Tarım Kredi Kooperatiflerinde yerli ve ithal kömür ürünlerinin satışı yapılır. İthal kömür ürünleri Türkiye'ye Rus, Ukrayna ve diğer ülkelerden distribütör firmalar tarafından getirilmektedir. Distribütör firmalar, toplu olarak getirdikleri kömür ürününü Türkiye'deki firmalara dağıtımını ve satışını yapar. Tarım Kredi Kooperatifleri tarafından kömür ürünü alımı distribütör ve yerli üretici firmalardan yapılır. Satın alım işlemlerinde öncelikle merkez birliği tarafından ihaleler yapılır. İhaleye yapılan teklifler yönetimin onayına sunulur ve distribütör firma ile anlaşma sağlanır. Merkez Birliği satın aldığı ürünleri bölge birliklerine, bölge birlikleri bölgeye bağlı kooperatiflere fatura keserek satış yapar. Kooperatifler dağıtıcı firmalardan nakliye

işlemlerini gerçekleştirirler ve ürünleri depolarına alırlar. Kooperatifler merkez birliklerinden ve bölge birliklerinden bağımsız özel bir firma statüsündedirler. Kar veya zarar ederek kendi muhasebe işlemlerini gerçekleştirirler. Kooperatiflerin sattıkları ürünler kooperatiflerin kendi depolarında tutulmaktadır. Kooperatiflerde satılan ürünlerin stoklama miktarı kooperatif tarafından belirlenmektedir.

Merkez birliği tarafından yapılan uzun vadeli satış tahmin işleminin başarılı olması tüm kooperatiflerin zarar etmeden uygun fiyatla ürünü almasını sağlar. Kooperatiflere aylık ve haftalık tahminlerin başarılı olması kooperatifin nakliye işlemlerindeki maliyetlerini ve stok maliyetlerini düşürür. Kooperatifin zarar etmesi engellenir. Mevcut iş süreçleri ve tahmin ihtiyacının hangi aşamada yapıldığı Şekil 4.1.'de gösterilmiştir.



Şekil 4.1. Tarım Kredi Kooperatifleri Kömür alım iş süreçleri

4.2. Tahmin İşlemlerinin Aşamaları

Çalışmanın sonucunda bulunan yöntem ile Tarım Kredi Kooperatiflerinden yapılan kömür satışlarının uzun vadeli, aylık ve haftalık olarak tahmin eden yöntemin bulunması amaçlanmıştır. Yapılan çalışmanın aşamaları üç adımda gerçekleştirilmiştir.

1. Adımda Verilerin toplanması ve Analizi için gerekli işlemler gerçekleştirilmiştir. 2. Adımda toplanan veriler tahmin yöntemlerinin çalışması için uygun duruma getirilmesi sağlanmıştır. 3. Adımda tahmin yöntemini belirlemek için farklı tahmin yöntemleri uygulanmış ve bu yöntemler kullanılarak elde edilen sonuçlar oluşturulmuştur. 4. Adımda tahmin sonuçlarına göre performans ölçüm testleri

ortalama sıcaklık bilgileri geçmiş yıllara ait meteoroloji arşivlerinden elde edilerek veri setine eklenmiştir. Toplanan veriler kayıt altına alınarak veri ön işleme kullanılması sağlanmıştır.

4.2.2. Veri ön işleme

Veri Ön işleme adımında toplanan verilerin çalışmanın amacına uygun duruma getirilmesi amaçlanmıştır. Veri ön işleme adımında öncelikle ilgisiz kayıtlar filtrelenerek veri azaltma işlemi gerçekleştirilmiştir. Çalışmada kömür ürününe ait tahmin çalışması yapılacağı için sadece kömür ürünü gruplara ait satış işlemleri kullanılmıştır. Tahmin çalışmasının sezona bağlı olarak artışının etkisini görmek için çalışmada ısınma amaçlı kömür grupları seçilmiştir. Mangal kömürü gibi ilgisiz kömür ürünü grupları ve fabrikalarda kullanılması mümkün olan büyük boyutlu kömür ürünü grupları filtrelenmiş, sadece ısınma amaçlı kömür ürünü gruplarının seçilmesi sağlanmıştır. Boyut azaltma adımında çalışma ile ilgili bilgilerin getirilmesi sağlanmıştır. Satış tarihi ve satış miktarı ve kooperatif bilgileri dışındaki bilgiler çıkarılarak verinin boyutu azaltılmıştır. Örnek evrenin oluşturulması için veri seçimi gerçekleştirilmiştir. Çalışmada kullanılması için 2014-2022 yılları arasındaki veriler seçilmiştir. Çalışmada mevsimsel satışlarının etkisini göstermek için bölge seçimi yapılmıştır. Örnek evren olarak Sakarya Bölge Birliğine ait 21 kömür satış bilgileri seçilmiş ve diğer kayıtlar çıkarılmıştır.

Veri temizleme adımında satış işlemlerinde iadesi yapılan ürünler, kooperatifler tarafından sisteme hatalı olarak yapılmış kayıtlar bulunmaktadır. Verinin hatalı gözükmesine sebep olan kayıtlar, yapılan işlemlerin son durumlarına göre incelenerek çıkarılmıştır. Kooperatiflerin hafta sonları tercihe bağlı olarak açık olma durumundan dolayı cumartesi pazar günlerine alt düşük satış verileri görülmüştür. Özetleme (Aggregation) adımında algoritmaların sağlıklı çalışması için hafta boyunca yapılan günlük satış işlemleri toplanmış ve veri setinin haftalık veri setine dönüştürülmesi sağlanmıştır. Oluşturulan veri seti örneği Tablo 4.1.'de gösterilmiştir. Kooperatiflerden 2014-2022 yılları arasında yapılan 420 haftaya ait satış verileri aşağıdaki kolonlar ile gösterildiği gibi elde edilmiştir.

Tablo 4.1. Veri seti örneği

Tarih	Miktar	Satış Sayısı	Sıcaklık
2014-3-16	82100	112	10
2014-3-23	49275	62	10
2014-3-30	49745	90	10
2014-4-6	65525	82	12
2014-4-13	61975	60	12

4.2.3. Gerçekleme süreci

Çalışmada kullanılan modeller literatürde satış tahmin çalışmaları incelenerek belirlenmiştir. Veri seti göz önünde bulundurularak zaman serisi tahmin yöntemleri ile çalışma yapılması uygun bulunmuştur. Satış tarihi ve satış miktarı verileri zaman serisi olarak kullanılmıştır. Literatürde başarılı çalışmalar yapılmış olan Arima yöntemini temel olarak geliştirilen ve sezon değişimlerine göre başarılı sonuçlar alınan SARIMAX yöntemi çalışmada kullanılmıştır. Literatürde tercih edilen Holt-Winters yönteminin ve facebook tarafından geliştirilen güncel zaman serisi yöntemi FB-Prophet yönteminin çalışmada kullanılması tercih edilmiştir. Tahmin alanında yapılan güncel çalışmalarda tercih edilen XGBoost makine öğrenmesi algoritmasının başarılı sonuçlar verdiği tahmin amaçlı kullanıldığı görüşmüştür. Zaman serisi yöntemlerine alternatif olarak çalışmada tercih edilmiştir. Bu bölümde çalışmada kullanılan yöntemler, çalışma prensiplerinin ve adımlarının farklı olmasından dolayı ayrı başlıklarda incelenmiştir.

4.2.3.1. SARIMAX

SARIMAX yönteminde kullanılan parametreler SARIMAX (p,d,q)(P,D,Q)s sembolü ile gösterilmektedir. Semboldeki d değeri modelin durağan olması için kaçınıcı dereceden farkının alınacağını gösterir. Serinin durağan olmaması serideki değerlerin zaman aralığında çok farklı olmasından kaynaklanır. Zaman serilerinin durağan olmadığı durumlarda, serinin farkı alınarak durağan duruma getirilmektedir [48].

Durağanlığı tespit etmek için ADF (Augmented Dickey-Fuller) birim kök testi uygulanmaktadır. SARIMAX yönteminin diğer parametreleri AIC skoru kullanılarak yapılmaktadır. Farklı parametreler denenerek en düşük AIC değerini içeren parametreler kullanılmaktadır. S değeri ise sezonun frekansını göstermektedir. Yıllık tekrar eden çalışmalarda haftalık çalışmalar için 52, aylık çalışmalarda 12 olarak kullanılmaktadır. Çalışmada SARIMAX yöntemi eğitim amaçlı kullanılan veri setindeki satış tarihi ve satış miktarı ve parametreler kullanılarak oluşturulur. Tahmin edilecek zaman aralığı girilerek model çalıştırılır. SARIMAX yöntemi oto regresyon yapan AR modeli ile MA hareketli ortalamalar modellerini bütünleşik olarak kullanır. Modelin çalıştırılması sonucunda 1 ve 2 yıllık tahmin sonuçları haftalık frekanslarla oluşturulmuştur. Elde edilen sonuçlar performans ölçümü yapılması için kaydedilmiştir.

4.2.3.2. Holt-Winters

Holt-Winter yönteminde tarih ve satış miktarından oluşan zaman serisini kullanmıştır. Yöntemin çalışmasında kullanılan sezon bilgisi yapılan çalışmanın haftalık olduğunu belirtmek için 52 olarak kullanılmıştır. Holt-Winters modeli eğitim verileri ve parametreler kullanılarak oluşturulur ve tahmin edilecek zaman aralığı belirtilerek model çalıştırılır. Holt-Winters kendisine yakın zaman aralığındaki yöntemlere daha fazla önem veren formülü ile tahmin sonuçlarını oluşturur. Çalışmada Holt-Winters yöntemi ile elde edilen sonuçlar performans ölçümü yapılması için kaydedilmiştir.

4.2.3.3. FB-Prophet

FB-Prophet yönteminde parametre optimizasyonuna ihtiyaç duyulmamıştır. FB-Prophet yöntemi modelinin oluşturulması adımı algoritma optimizasyona ihtiyaç duymadan çalışmaktadır. Veri seti olarak tarih ve satış miktarı ile oluşan zaman serisi kullanmıştır. Modelin çalıştırılması sonucunda 1 ve 2 yıllık tahmin sonuçları haftalık frekanslarla oluşturulmuştur. Elde edilen sonuçlar performans ölçümü yapılması için kaydedilmiştir.

4.2.3.4. XGBoost

XGBoost algoritması zaman serisi tahmin yöntemlerinden farklı olarak tarih aralığını indeks olarak almaz. Tarih içerisindeki değerleri eğitim verisi olarak alır. XGBoost ile yapılan zaman serisi çalışmalarında tarih bilgisindeki ay, gün, yıl, mevsim, yılın kaçınıcı günü olduğu ve yılın kaçınıcı haftası olduğu ayrıştırılarak kullanılır. Algoritma çalışırken eğitim parametrelerini deneyerek tahmin formüllerini oluşturur. Çalışmada XGBoost yöntemi kullanılırken iki farklı adım uygulanmıştır. XGBoost yöntemi öncelikle sadece tarih bilgisinden oluşan parametrelerle çalıştırılmıştır. İkinci adımda tarih bilgileri ile satış miktarını ve aylık sıcaklık değerini kullanarak çalıştırılmıştır. XGBoost yönteminin çalışmasında ekstra parametrelerin sonucu etkileme oranı araştırılmıştır. Test verilerinde kullanılan satış miktarını bilgisi ile yapılan çalışmanın aşırı öğrenmeye yol açmasından dolayı XGBoost yöntemi diğer yöntemlerle karşılaştırılırken sadece tarih bilgisi ile oluşan parametreleri kullanarak çalışan adımdaki sonuçlar kullanılmıştır. XGBoost yöntemi parametre olarak öğrenme oranı ve tahmin sayısı bilgilerini almaktadır. Öğrenme oranı, zayıf öğrenme adımlarında oluşan yöntemin derecesini gösterir. Tahmin sayısı ise algoritma çalıştırılırken oluşturulan karar ağacı adımlarını oluşturma ve deneme sayısını gösterir. Algoritma karar ağaçları içerisinde tahmin yöntemleri oluşturarak modelini oluşturur. XGBoost yöntemi modeli test verileri ile çalıştırılmış ve elde edilen sonuçlar kaydedilmiştir.

4.3. Sonuçların değerlendirilmesi

Tahmin çalışması sonucunda elde edilen tahmin verileri ile gerçek veriler MAE, MAPE, MSE ve RMSE, R-Kare performans ölçüm yöntemleri uygulanarak karşılaştırılmıştır. MAE, MAPE, MSE, RMSE değerleri tahminlerin hata oranlarını göstermektedir. Bu testler sonucunda elde edilen değerlerin düşük olması uygulanan yöntemin başarılı olduğunu göstermektedir. MAPE değerinin 2. Bölümde gösterildiği gibi 0,50 değerinden düşük olması gerekmektedir. 0,50'den büyük değerlerin elde edilmesi tahmin yönteminin başarısız olduğunu göstermektedir. R-kare değeri 1-100 aralığında sonuç vermektedir ve sonucun 80 ve üzeri olması beklenmektedir. Tahmin

yöntemleri test sonuçlarındaki beklentilerle karşılaştırılarak ideal yöntem belirlenmiştir.

4.4. Çalışmada Kullanılan Teknolojiler

Çalışmalar 11. nesil Intel i7-11800H @ 2.30 GHz işlemci ve 32 GB ram özelliklerine sahip bilgisayar kullanılarak Windows 10 işletim sisteminde yapılmıştır. Kullanılan veri seti Oracle veri tabanından Oracle SQL Developer programı ile elde edilmiştir. Bilgisayarda çalışma ortamının kurulması için öncelikle Anaconda uygulaması kurulmuştur. Tüm çalışmalar Python 3.8.3 versiyonu ve ide olarak Spyder uygulaması kullanılarak yapılmıştır. Python diline pandas, statsmodels, prophet, matplotlib kütüphaneleri çalışmada kullanılmıştır. Çalışmada kullanılan grafiklerin oluşturulmasında Yed Graph Editor programı kullanılmıştır.

BÖLÜM 5. ARAŞTIRMA BULGULARI

Bu bölümde Tarım Kredi Kooperatiflerine ait kömür satışları verilerinin SARIMAX, Holt-Winters ve FB-Prophet ve XGBoost modelleri kullanarak yapılan tahmin çalışması süreci hakkında bilgi verilmiş ve gözlemler anlatılmıştır. Detaylar örneklerle ve grafiklerle gösterilmiştir. Ayrıca tahmin yöntemlerinin başarılarını karşılaştırmak için kullanılan performans ölçüm testlerinin sonuçları gösterilmiş ve anlatılmıştır.

5.1. Verilerin Toplanması ve Analizi

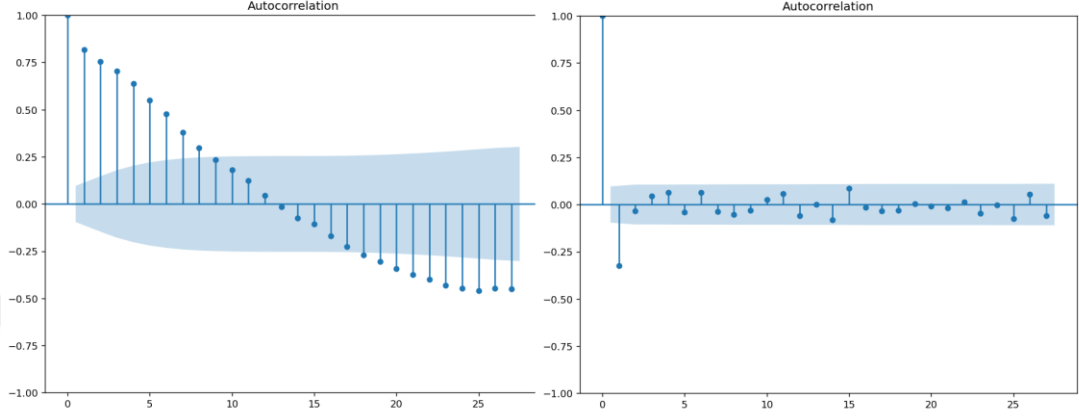
Çalışmada kullanılan veri seti Tarım Kredi Kooperatiflerine bağlı kooperatifler tarafından satılan kömür ürünü ve bu ürüne ait satış bilgileri kullanılarak hazırlanmıştır. Hazırlanan veri seti ile 8 yıllık satış tahmin bilgisi elde edilmiştir. Eğitim ve test verileri 7 yıl -1 yıl ve 6 yıl – 2 yıl oranlarına bölünerek 1 yıllık ve 2 yıllık tahmin sonuçlarının elde edilmesi amaçlanmıştır.

5.2. SARIMAX Uygulanması

SARIMAX yöntemi ile çalışma yapılırken öncelikle serinin durağanlığı araştırılmıştır. Serinin durağanlık durumu SARIMAX yöntemi parametrelerinden diferansiyel katsayısı (d) değerini etkilemektedir. Zaman serilerinde durağanlığın tespit edilmesi için ADF (Augmented Dickey-Fuller) birim kök testi uygulanmıştır. ADF birim kök testi sonucunda ADF değerinin kritik değerlerden yüksek olduğu görülmektedir. Serinin durağan olmadığı anlaşılmıştır. Serinin birinci dereceden farkı alınarak serinin durağan olması sağlanmıştır. 1. Dereceden farkı alınmış ADF testi sonuçları Tablo 5.1.'de gösterilmiştir. Serinin farkı alınmadan önce ve 1. Dereceden farkı alınarak oluşturulmuş otokorelasyon (ACF) grafikleri ise Şekil 5.1.'de gösterilmiştir.

Tablo 5.1. ADF birim kök testi sonuçları

Test	ADF istatistik	Kritik değerler		
		%1	%5	%10
1. derece ADF	-19.41	-3.446	-2.869	-2.570



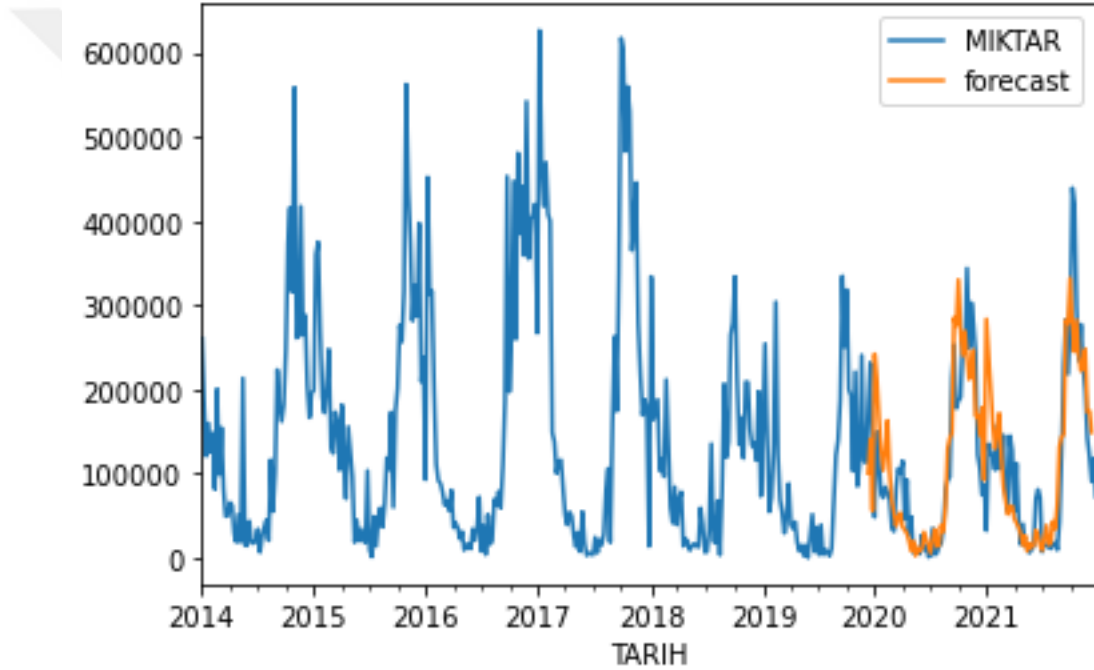
Şekil 5.1. 0. ve 1. derece otokorelasyon grafiği

SARIMAX algoritması ile çalışma yapılırken öncelikle autoArima ile ideal parametreler bulunmuş, sonrasında bu değerlere yakın değişkenler denenerek parametre optimizasyonu yapılmıştır. İdeal parametreler bulunurken Akaike değerin yararlanılmıştır. En düşük Akaike değerini içeren 10 modele ait sonuçlar Tablo 5.2.'de gösterilmektedir.

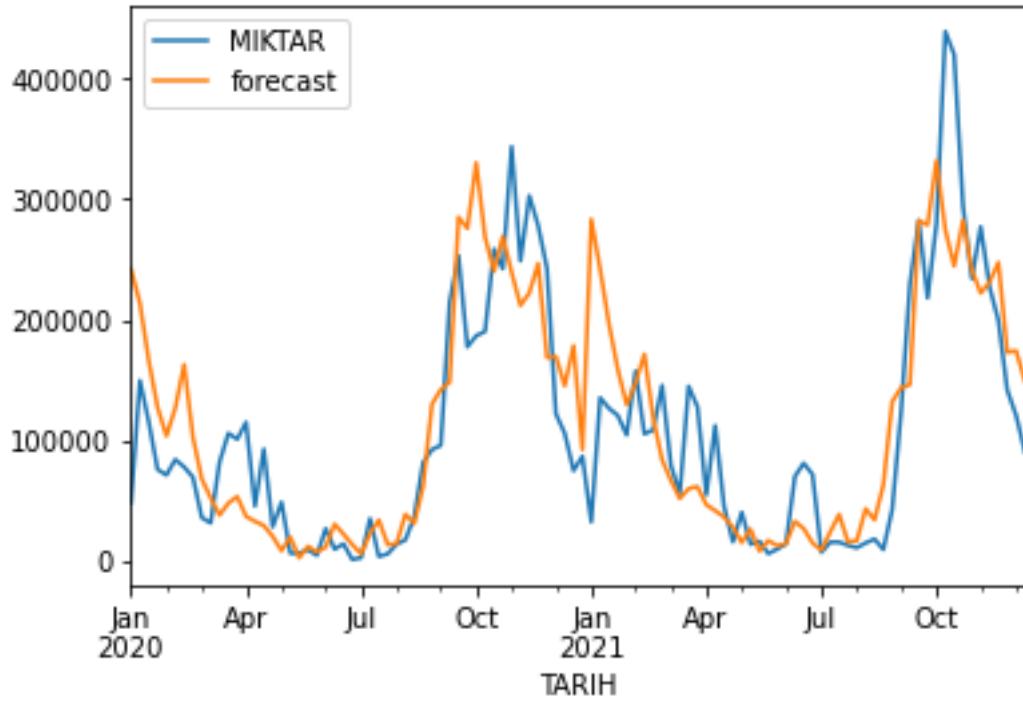
Tablo 5.2. Sarimax yöntemi AIC Testi Sonuçları

Sıra	Model	AIC
1	SARIMAX (1,0,1) (1,1,1,52)	6564
2	SARIMAX (2,0,1) (1,1,1,52)	6565
3	SARIMAX (1,0,1) (1,1,2,52)	6567
4	SARIMAX (2,0,2) (1,1,1,52)	6568
5	SARIMAX (3,0,3) (1,1,1,52)	6569
6	SARIMAX (3,0,3) (2,1,2,52)	6570
7	SARIMAX (2,0,2) (2,1,2,52)	6571
8	SARIMAX (2,0,2) (2,1,2,52)	6572
9	SARIMAX (2,0,2) (3,1,3,52)	6575
10	SARIMAX (1,0,1) (1,1,0,52)	6580

Parametre optimizasyonu çalışması sonucunda en düşük AIC skoru elde edilen SARIMAX(p,d,q)(P,D,Q)s modeli (1,0,1)(1,1,1,52) parametreleri ile elde edilmiştir. Modelin p,d,q parametreleri mevsimsel olmayan bölümü, P,D,Q parametreleri mevsimsel bölümü temsil etmektedir. Otoregresif modellerin gözlem sayıları p ve P parametrelerine göre belirlenmiştir. MA modellerin boyutları q ve Q parametreleri ile belirlenmiştir. İdeal parametrelerin belirlenmesinden sonra tahmin çalışması yapılmıştır. 1 yıllık ve 2 yıllık haftalık frekanslı tahmin değerleri elde edilmiştir. 2 yıllık tahmin çalışmasının gerçek değerler ile karşılaştırmalı görseli Şekil 5.2. ve Şekil 5.3.'de gösterilmiştir.



Şekil 5.2. SARIMAX tahmin grafiği



Şekil 5.3. SARIMAX detaylı tahmin grafiği.

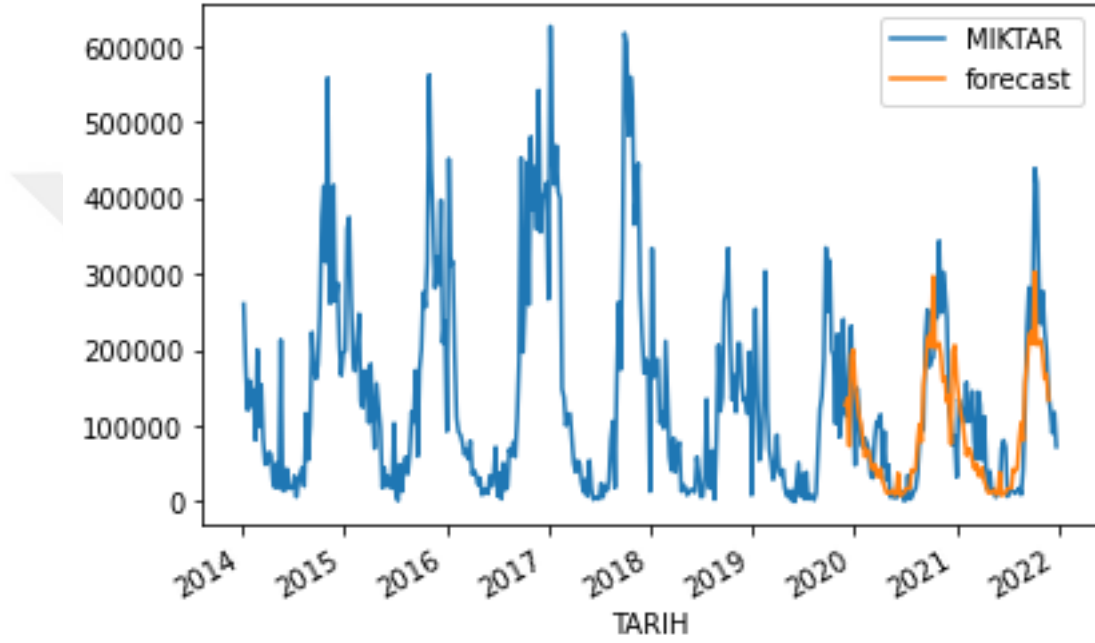
Sonuçlara R-Kare, MAE, MAPE, MSE ve RMSE testleri uygulanmıştır. Yapılan testlerin sonuçları Tablo 5.3.'de gösterilmiştir. Sonuçlar incelendiğinde, SARIMAX yöntemi ile yapılan haftalık verilerin 1 ve 2 yıllık tahmin çalışmalarının R-Kare başarı puanına göre kabul edilir seviyede olduğu fakat iyi olarak değerlendirilen seviyede olmadığı görülmektedir. MAPE sonuçları da kabul edilebilir seviyede değildir. SARIMAX yönteminin MAE ve MSE hata sonuçları Holt-Winters, FB-Prophet ve XGBoost yöntemlerinin sonuçlarından düşüktür. Diğer yöntemlerin sonuçlarından az hata alınmıştır. SARIMAX yönteminin tahmin sonucunun diğer yöntemlerden başarılı olduğu görülmektedir.

Tablo 5.3. SARIMAX tahmin çalışması test sonuçlarının karşılaştırılması

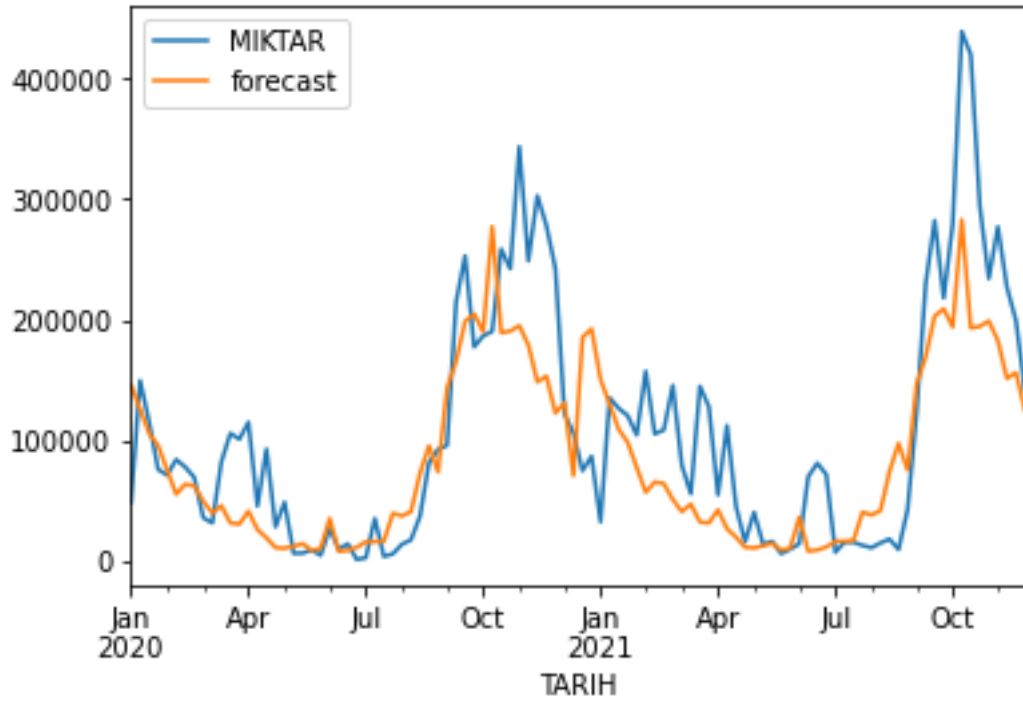
	1 yıllık	2 yıllık
MAE	41963	46763
MAPE	0.57	0.85
MSE	$35 * 10^7$	$38 * 10^7$
RMSE	59366	64039
R-Kare	0.69	0.60

5.3. Holt-Winters Uygulanması

Haftalık frekanslı veriler ile Holt-Winters yöntemi kullanılarak tahmin çalışması yapılmıştır. Çalışmada seasonal_periods parametresi olarak kullanılmıştır. 2 yıllık tahmin çalışmasının gerçek değerler ile karşılaştırmalı görseli Şekil 5.4. ve Şekil 5.5.' de gösterilmiştir.



Şekil 5.4. Holt-Winters tahmin grafiği



Şekil 5.5. Holt-Winters detaylı tahmin grafiği

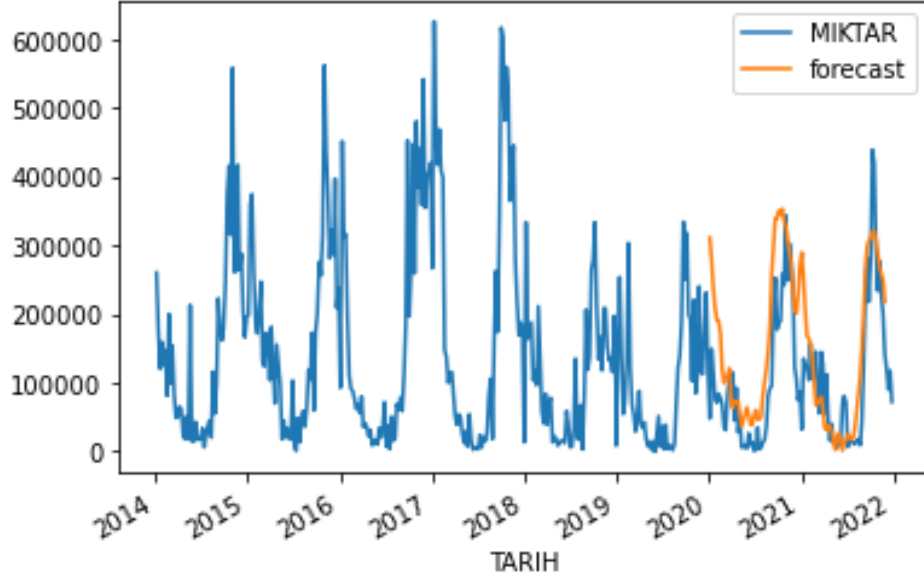
Holt-Winter yönteminin sonuçlarına R-Kare, MAE, MAPE, MSE ve RMSE testleri uygulanmıştır. Yapılan testlerin sonuçları Tablo 5.4.'de gösterilmiştir. Sonuçlar incelendiğinde, Holt-Winters yöntemi ile yapılan haftalık verilerin 1 ve 2 yıllık tahmin çalışmalarının R-Kare başarı puanına göre kabul edilir seviyede olduğu fakat iyi olarak değerlendirilen seviyede olmadığı görülmektedir. MAPE sonuçları da kabul edilebilir seviyede değildir. Holt-Winters yönteminin MAE ve MSE hata sonuçları, SARIMAX yönteminin sonuçlarından daha yüksek, FB-Prophet ve XGBoost yöntemlerinin sonuçlarından düşüktür. Holt-Winters yönteminin tahmin sonucunun SARIMAX yönteminden başarısız, diğer yöntemlerden başarılı olduğu görülmektedir.

Tablo 5.4. Holt-Winters tahmin çalışması test sonuçlarının karşılaştırılması

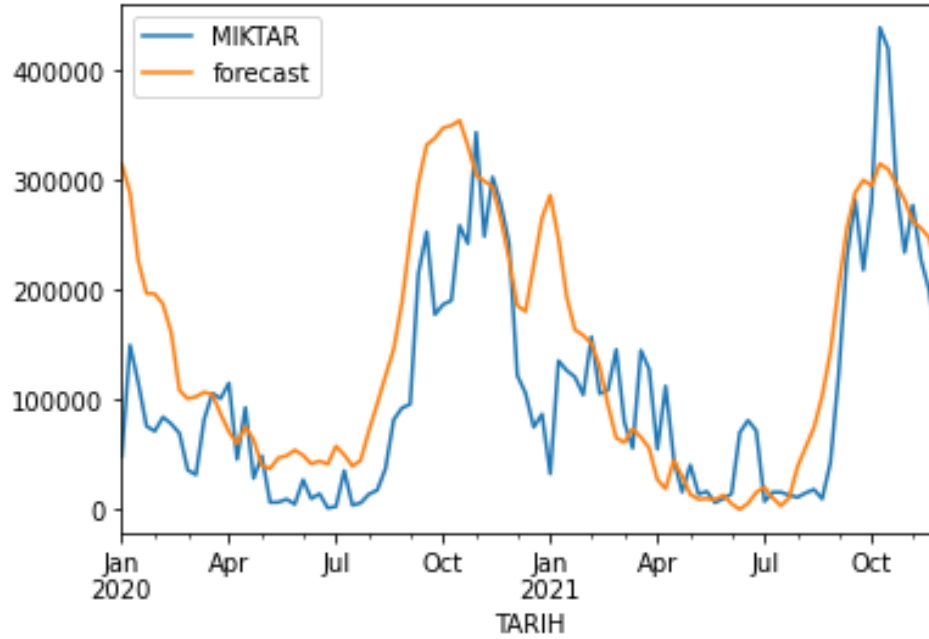
	1 yıllık	2 yıllık
MAE	45682	44295
MAPE	0.97	0.93
MSE	$39 * 10^7$	$36 * 10^7$
RMSE	62861	60625
R-Kare	0.65	0.62

5.4. FB-Prophet Uygulanması

FB-Prophet yönteminin istatistik bilgisi ve optimizasyonuna gerek olmadan çalışan bir yöntem olması sebebiyle parametre optimizasyonu yapılmadan tahmin çalışması çalıştırılmıştır. Şekil 5.6. ve Şekil 5.7.'de 2 yıllık tahmin çalışması karşılaştırılarak gösterilmiştir.



Şekil 5.6. FB-Prophet Tahmin Grafiği



Şekil 5.7. FB-Prophet detaylı tahmin grafiği

FB-Prophet yönteminin sonuçlarına R-Kare, MAE, MAPE, MSE ve RMSE testleri uygulanmıştır. Yapılan testlerin sonuçları Tablo 5.5.'de gösterilmiştir. Sonuçlar incelendiğinde, FB-Prophet yöntemi ile yapılan haftalık verilerin 1 ve 2 yıllık tahmin çalışmalarının R-Kare başarı puanına göre kabul edilir seviyenin altında olduğu görülmektedir. MAPE sonuçları da kabul edilebilir seviyenin üstündedir. FB-Prophet yönteminin MAE ve MSE hata sonuçları, diğer yöntemlerin sonuçlarından daha yüksektir. Diğer yöntemlerin sonuçlarından fazla hata alınmıştır. FB-Prophet yönteminin tahmin sonucunun diğer yöntemlerden başarısız olduğu görülmektedir.

Tablo 5.5. FB-Prophet tahmin çalışması test sonuçlarının karşılaştırılması

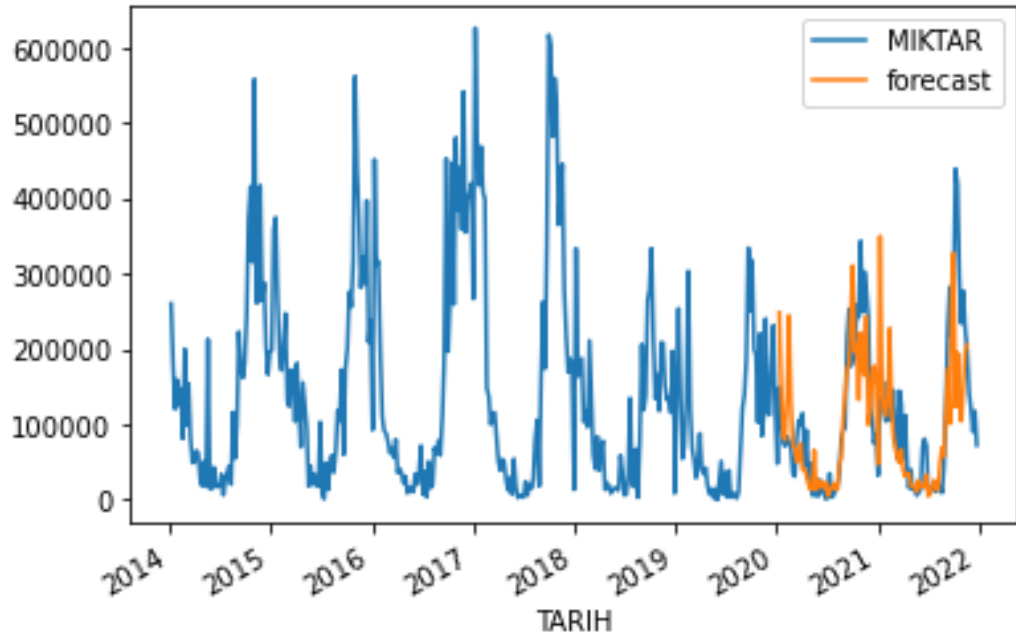
	1 yıllık	2 yıllık
MAE	67623	71088
MAPE	2.04	2.29
MSE	$70 * 10^7$	$64 * 10^8$
RMSE	83952	80480
R-Kare	0.38	0.35

5.5. XGBoost Uygulanması

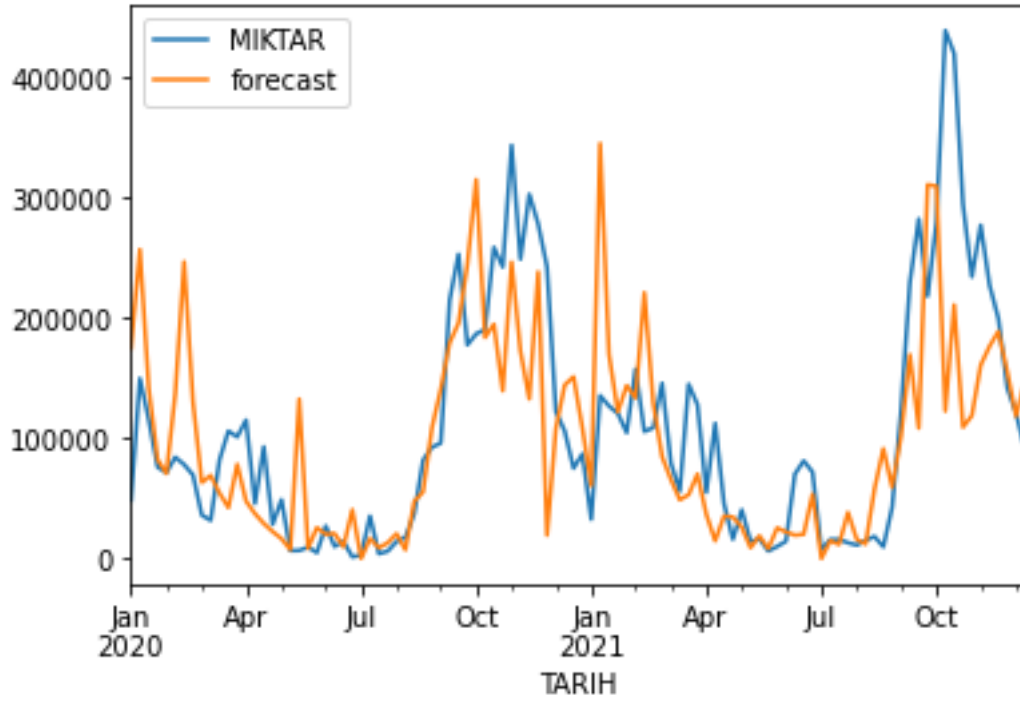
XGBoost algoritması parametrelerle daha verimli çalışan bir yöntem olmasından dolayı iki farklı senaryo uygulanmıştır. Öncelikle zaman serisi yöntemleri ile performansın sağlıklı karşılaştırılması için sadece satış miktarı ve satış tarihi parametreleri ile çalıştırılmıştır. İkinci uygulamada XGBoost yönteminin performansını iyileştirmek için ekstra parametrelerle birlikte çalıştırılmıştır.

Senaryo 1: XGBoost Yönteminin Parametresiz Uygulanması

XGBoost algoritması sadece zaman parametreleri ile çalıştırılmıştır. Şekil 5.8. ve Şekil 5.9.'da 2 yıllık tahmin çalışması karşılaştırılarak gösterilmiştir.



Şekil 5.8. XGBoost tahmin grafiği.



Şekil 5.9. XGBoost detaylı tahmin grafiği.

Çalışmanın sonuçları Tablo 5.6.'de gösterilmiştir. Sonuçlar incelendiğinde, XGBoost yöntemi ile yapılan haftalık verilerin 1 ve 2 yıllık tahmin çalışmalarının R-Kare başarı

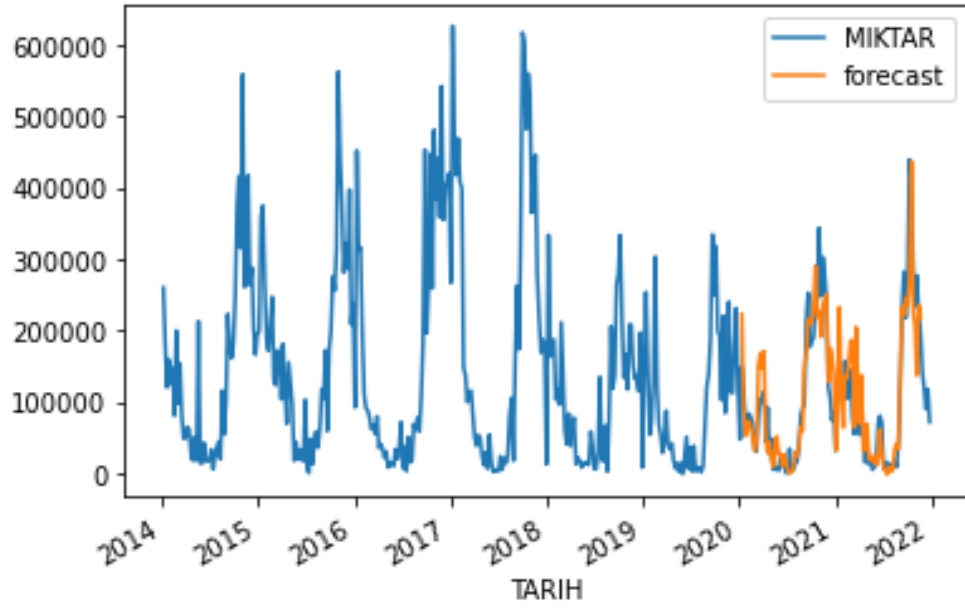
puanına göre kabul edilir seviyede olduğu fakat iyi olarak değerlendirilen seviyede olmadığı görülmektedir. MAPE sonuçları da kabul edilebilir seviyede değildir. XGBoost yönteminin MAE ve MSE hata sonuçları, SARIMAX ve Holt-Winters yöntemlerinin sonuçlarından daha yüksek, FB-Prophet ve yönteminin sonuçlarından düşüktür. Holt-Winters yönteminin tahmin sonucunun FB-Prophet yönteminden başarılı, diğer yöntemlerden başarısız olduğu görülmektedir.

Tablo 5.6. XGBoost tahmin çalışması test sonuçlarının karşılaştırılması

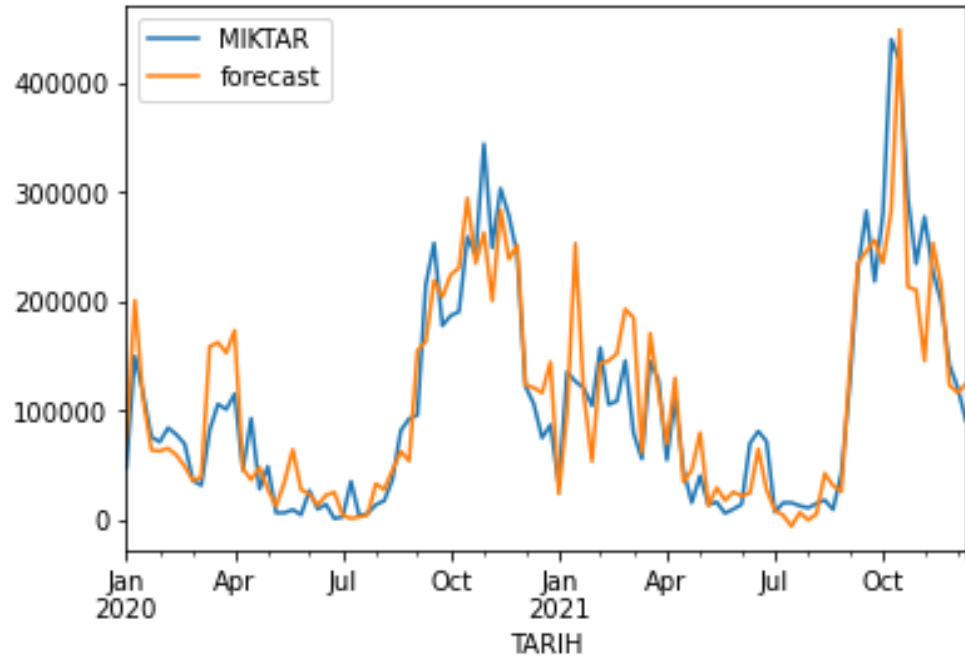
	1 yıllık	2 yıllık
MAE	48246	49453
MAPE	0.62	0.97
MSE	$51 * 10^7$	$55 * 10^7$
RMSE	71392	74644
R-KARE	0.53	0.45

Senaryo 2: XGBoost Yönteminin Parametrelerle Uygulanması

XGBoost algoritmasının performansını artırmak için parameter sayısı arttırılarak çalıştırılmıştır. Parametrelere satış sayısı ve sıcaklık bilgisi eklenerek algoritmanın tahmin başarısı yükseltilmiştir. 2 yıllık tahmin çalışmasının gerçek değerler ile karşılaştırmalı görseli Şekil 5.10. ve Şekil 5.11.'de gösterilmiştir.



Şekil 5.10. XGBoost parametrelili tahmin grafiği



Şekil 5.11. XGBoost parametrelili ve detaylı tahmin grafiği.

Çalışmada elde edilen değerler Tablo 5.7.'de gösterilmiştir. XGBoost yöntemi ile parametreler kullanılarak yapılan haftalık verilerin 1 ve 2 yıllık tahmin çalışmalarının R-Kare başarı puanına göre iyi olarak değerlendirilen seviyede olmadığı

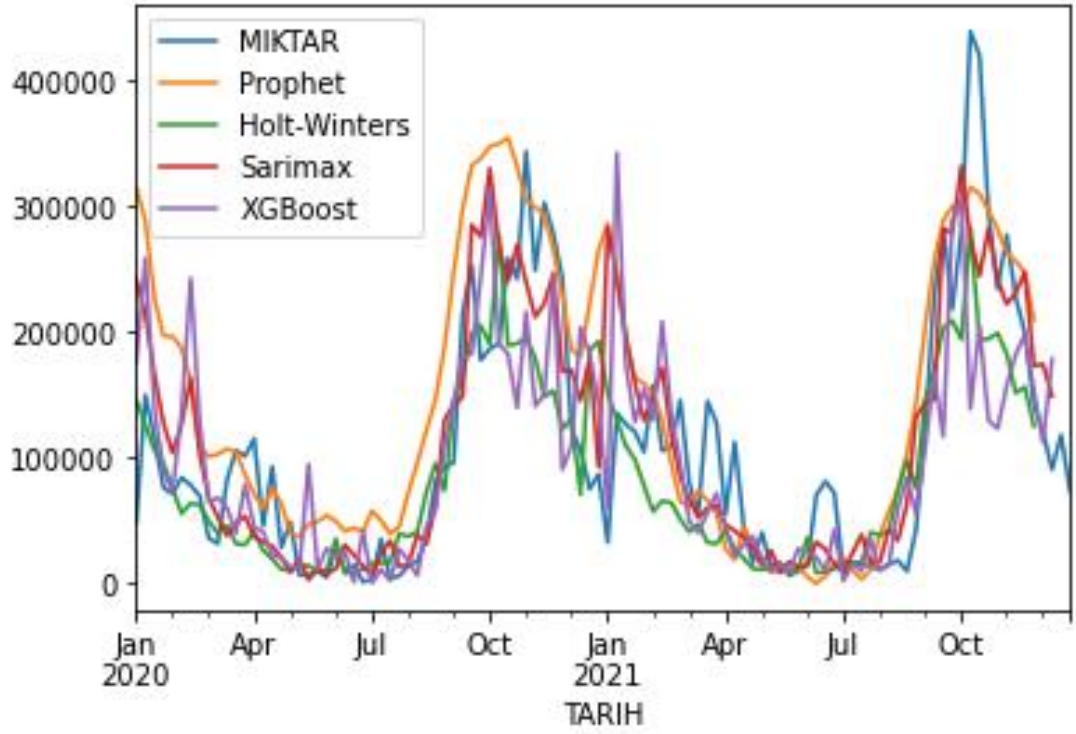
görülmektedir. Özellikle 1 yıllık tahmin çalışmasının MAPE testinin sonucunda 0.42 değeri elde edilmiştir. Bu değer 0.50 hata oranı değerinin altında olmasından dolayı kabul edilir seviyededir. XGBoost yönteminin MAE ve MSE testleri sonucundan diğer yöntemlerin sonuçlarından daha az hata alınmıştır ve daha düşük değerler elde edilmiştir. XGBoost yönteminin tahmin sonucunun diğer yöntemlerden başarılı olduğu görülmektedir.

Tablo 5.7. XGBoost tahmin çalışması test sonuçlarının karşılaştırılması

	1 yıllık	2 yıllık
MAE	27854	31047
MAPE	0.42	0.73
MSE	$15 * 10^7$	$19 * 10^7$
RMSE	71392	43737
R-Kare	0.87	0.81

5.6. Aylık ve Yıllık Frekans ile Tahmin Çalışması ve Modelleri Karşılaştırılması

Elde edilen haftalık tahmin sonuçları incelendiğinde haftalık frekans ile başarılı sonuçlar elde edilememiştir. 1 ve 2 yıllık tahmin çalışması sonucunda yöntemlerin MAPE testi sonucunda bulunan değerlerin 0.50 değerinin altında olması gerekmektedir. Fakat MAPE değerleri beklenen seviyenin üstündedir. Başarılı çalışmalarda 90 değerinin üstünde sonuç alınan R-kare testi sonuçları da çalışmada yeterli seviyede değildir. Fakat tahmin sonuçlarına göre yöntemlerin sezona bağlı olarak satış rakamlarını azalış ve artış yönlerinin başarılı tespit edildiği görülmüştür. Bu durum modelleri karşılaştırılmasının görselleştirildiği Şekil 5.12.'de grafiklerinde de görülmektedir.



Şekil 5.12. Modellerin karşılaştırılması

Sonuçların tahmin aralığı değiştirilerek tekrar incelenmiştir. MAE, MAPE, MSE ve RMSE, R-kare testleri uygulanmıştır. Modellerin aylık ve yıllık frekanslarla alınan sonuçları gerçek değerler ile karşılaştırılmıştır. Tablo 5.8.'de modellerin aylık frekanslarla yapılan performans ölçüm testi sonuçları gösterilmiştir. Alınan hata değerlerinin azaldığı ve R-kare tahmin performansının yükseldiği görülmüştür.

Tablo 5.8. Modellerin aylık frekans ile sonuçlarının karşılaştırılması

	SARIMAX	Holt-Winters	FB-Propet	XGBoost
MAE	99528	66002	202978	132068
MAPE	0.39	0.36	0.69	0.37
MSE	$188 * 10^8$	$84 * 10^8$	$678 * 10^8$	$363 * 10^8$
RMSE	$137 * 10^3$	$92 * 10^3$	$259 * 10^3$	$190 * 10^3$
R-Kare	0.85	0.86	0.46	0.71

Tablo 5.9.'da 2020 yılına ait 1 yıllık ve 2020-2021 yıllarına ait 2 yıllık toplam satış rakamları ve yöntemlerin tahmin sonuçları gösterilmiştir. Toplam satış rakamları

karşılaştırıldığında 2 yıllık 11317 ton toplam satış rakamı SARIMAX yöntemi ile yapılan tahmin sonucunda 11769 ton olarak tahmin edilmiştir. XGBoost yöntemi ile parametreler kullanılarak daha başarılı sonuçlar elde edilmiştir, fakat sadece tarih ve satış bilgisi girdiler ile yöntemlerin performansları karşılaştırıldığında SARIMAX yönteminin başarılı sonuçlar verdiği görülmüştür. XGBoost yönteminde kullanılan parametrelerden satış sayısı parametresi satış sonrası elde edilen bir parametredir. XGBoost yönteminde aşırı öğrenme ile sonucu en başarılı sonuçları vermiştir. Satış rakamlarını destekleyen parametreler kullanıldığında XGBoost yöntemi ile başarılı sonuç veren yöntemin bulunmasının mümkün olduğunu göstermek için kullanılmıştır. SARIMAX yöntemi zaman ve miktar sonuçları ile karşılaştırıldığında en başarılı sonucu veren yöntemdir. Elde edilen sonuçlara uygulanan test sonuçlarına göre MAPE hata oranı yüzde 3,48 hata sonucu alınmıştır. Yıllık tahmin sonucu çok iyi durumdadır.

Tablo 5.9. Modellerin 1 ve 2 yıllık toplam satış rakamları

	2020	2020 – 2021
Satış Rakamları	5626	11317
XGBoost (Parametre ile)	5371	11278
SARIMAX	5614	11769
XGBoost	5070	10095
FB-Prophet	4849	9880
Holt-Winters	4470	9049

BÖLÜM 6. SONUÇ VE ÖNERİLER

Günümüzde firmaların süreçlerini yöneten ERP ve benzeri uygulamaların yaygınlaşması ve uzun süreli kullanılması sonucunda firmalara ait veri tabanlarında büyük veriler içeren tablolar ortaya çıkmıştır. İstatistik ve Veri Bilimi alanındaki gelişmeler büyük verilerin tahmin amaçlı kullanılmasını sağlamıştır. Tarım Kredi Kooperatiflerinin kullandığı Ekoop ERP uygulamasının uzun süredir satış süreçlerinde kullanılması sonucu satış işlemlerine ait büyük kayıtlar ortaya çıkmıştır. Şirketin en çok sattığı ürünlerden biri olan kömür ürünü, stok maliyetlerinin fazla olması, satın alım süreçlerinin başarılı tahmin edilerek iyileştirilmesi gereken bir ürün olması sebebiyle satış tahmin çalışmasına ihtiyaç duyulan bir üründür.

Çalışmada Türkiye Tarım Kredi Kooperatiflerinin Sakarya Bölge Birliğine bağlı 21 kooperatif tarafından yapılan 2014-2022 yılları arasındaki ısınma amaçlı kömür ürünü satış verilerini kullanarak tahmin çalışmasının yapılması amaçlanmıştır. SARIMAX, Holt-Winters, FB-Prophet zaman serisi yöntemlerini ve XGBoost makine öğrenmesi algoritmasını kullanarak uzun vadeli tahmin çalışması yapılmıştır. Haftalık satış miktarını ve satış tarihini kullanarak zaman serisi oluşturulmuştur. Zaman serisine satış sayısı ve aylık sıcaklık verileri eklenerek veri seti büyütülmüştür. Toplamda 420 haftaya ait satış rakamlarını içeren veri seti 7-1 ve 6-2 oranlarında bölünmüştür. Veri setini kullanarak 2021 yılının ve 2020, 2021 yıllarının satış tahmin çalışması yapılması anmıştır.

Çalışmada öncelikle haftalık veri setlerini kullanarak haftalık satış rakamları tahmin edilmiştir. Tahmin çalışmasında Haftalık frekans ile yapılan tahmin çalışmasının sonuçları MAPE, MAE, MSE, RMSE, R-Kare performans ölçüm yöntemleri kullanılarak incelenmiş ve yeterli sonuç alınamadığı görülmüştür. Performans ölçüm testleri sonucunda yöntemlerin hata oranları beklenen seviyenin altındadır. Tahmin

sonucunda MAPE değeri 0.50 değerinin altında olması gerekmektedir. Fakat yapılan çalışma sonucunda yöntemlerin MAPE değerleri beklenen seviyenin üstündedir. Başarılı çalışmalarda 75 değerinin üstünde sonuç alınan R-kare testi sonuçları da çalışmada yeterli seviyede değildir. Yöntemlerin sonuçları incelendiğinde tahminlerin gerçek değerlere göre azalış ve artış yönlerini başarılı yaptığı görülmüştür. Sonuçların performansını iyileştirmek için haftalık tahmin sonuçları aylık ve yıllık aralıklarla toplanarak gerçek değerler ile tekrar karşılaştırılmıştır. Aylık tahmin çalışmasında satış tarihini ve satış miktarını kullanarak tahmin çalışması yapılmış, Holt-Winters, SARIMAX ve XGBoost ile başarılı sonuç elde edilmiştir. Holt-Winters yöntemi ile en başarılı sonuç alınmıştır. Fakat SARIMAX, Holt-Winters ve XGBoost yöntemlerinin hata ölçüm yöntemleri ve R-Kare testi sonucunda yakın değerler elde edilmiştir. Yıllık tahmin sonuçları performans ölçüm testleri yapılarak karşılaştırılmış ve en başarılı sonucu veren SARIMAX yönteminin MAPE hata değeri sonucu 0.35 (yüzde 3,5) elde edilerek tahmin sonucunun başarılı olduğu görülmüştür.

Çalışma sonucunda elde edilen bulgular çalışmanın amaçlarına göre karşılaştırıldığında XGBoost ve SARIMAX yöntemlerinin yıllık kömür satış miktarını başarılı tahmin ettiği görülmüştür. Tahmin yöntemleri ile Tarım Kredi Kooperatifleri Sakarya Bölge Birliğine bağlı 21 kooperatifin 1 ve 2 yıllık kömür satışlarının aylık ve yıllık aralıklarla tahmin edilmesi mümkündür. Satış tahminlerinin başarılı yapılması ile kooperatiflerin anlaşmalar ve ihaleler için gerekli kömür ihtiyaçları doğru tahmin edilecektir. Kooperatiflerin satış tahmin rakamlarını kullanarak kömürü ithal eden ve toptan satışını yapan distribütör firmalar ile sezon başında kârlı anlaşmalar yapması mümkündür. Kömür ürününün gereğinden az alınarak düşük fiyatlarla tedarik edilmesi veya gereğinden fazla satın alınarak satış maliyeti oluşturulması ve stok maliyetlerinin artırılması önlenecektir. Aylık kömür satışlarının başarılı yapılması kooperatifin deposunda aylık olarak saklanması gereken kömür miktarının doğru tespit edilmesini sağlayacaktır. Tedarik edilmesi gereken kömür ürünü miktarı belirlenecektir. Stok ve nakiye maliyetleri düşürülerek şirketin kar etmesi sağlanacaktır. Enerji veya farklı sezona ve ısıya bağlı değişen ürünleri olan farklı sektörlerdeki firmaların SARIMAX veya XGBoost yöntemleri ile çalışmalar yaparak uzun vadeli tahmin çalışması yapılması mümkündür. Yöntemlerin tüm bölgelerin kooperatiflerinde kullanılarak ve

sonuçların toplanması ile Tarım Kredi Kooperatiflerden uzun vadeli satılacak ısınma amaçlı kömür ürünü tahmin edilebilir. Bu çalışma %3-%4 aralığında Türkiye'nin ihtiyacı olan kömür ürününü yansıtacağından dolayı ülkenin uzun vadeli kömür ihtiyacı tespit edilebilir. Çalışma sonucu firmanın stok ihtiyacının haftalık olarak belirlenmesi için amacına ulaşamamıştır. Fakat çalışmada kullanılan SARIMAX ve XGBoost yönteminin tahmin süresi azaldıkça ve eğitim için kullanılan veri setinin arttırılması ile başarının yükseldiği görülmüştür. Kısa vadeli tahmin çalışması ile veri seti arttırılarak kısa vadeli stok ihtiyacı tahmini başarılı sonuç elde edilebilir.

Çalışmada SARIMAX yönteminin zaman ve miktar parametrelerinden oluşan zaman serilerinde başarılı sonuç verdiği görülmüştür. XGBoost yöntemi ile sonuçları etkileyen verilerin arttırılması ile başarılı sonuçlar vermesinin mümkün olduğu tahmin edilmiştir. Çalışmada kullanılan satış sayısı ve sıcaklık parametreleri arttırılabilir. Sıcaklık verisi günlük olarak gece ve gündüz sıcaklık değerleri kullanılarak sonuçlara olan etkisi araştırılabilir. Satış Sayısı parametresi satış sonrası elde edilen bir parametre olmasından dolayı XGBoost yönteminin parametreler ile SARIMAX yönteminden başarılı sonuçlar vermesi yeterli değildir. XGBoost yönteminin ileriki çalışmalarda zaman serisi tahmin yöntemlerine göre başarılı olan özelliği gösterilmiştir. Kooperatiflerin ortak sayıları değişimleri ortakların yaş aralıkları, ortaklara ait kredi skorları, satış yapılan bölgelere ait nüfus tahminleri ile çalışmanın genişletilerek XGBoost yöntemi ile yapılan çalışmaların başarılarının yükseltilmesi mümkündür. Kooperatif depolarında saklanan kömür ürünü miktarının depoya konulan ve çıkan miktarı kıyaslanarak ürünün hangi koşullarda daha az kayıp ile saklanması gerektiğinin araştırılması başka bir çalışma konusu olabilir. İleriye yönelik çalışmalarda satış yapılan yerin koordinat bilgileri ile ısı miktarı tahmin edilerek bölgeye göre değişen ısı miktarının satış miktarına olan etkisi araştırılabilir. Kömür harici diğer ürünlerin de satış rakamlarının tahmininde benzer yöntemler kullanılarak satış rakamlarına yönelik tahmin çalışması yapılabilir.

KAYNAKLAR

- [1] Meeran, S., Dyussekenewa, K. Goodwin, P., Forecasting using combination of diffusion model and forecast market – an adaption of prediction/preference markets, IFAC Proceedings Volumes, pp. 87-92, 2013.
- [2] Nguyen, H. D., Tran, K. P., Thomassey, S., Hamad, M., Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management, International Journal of Information Management, p. Volume 57, 2021.
- [3] Patnaik, S., Yang, X. S., Sethi, I. K., Advances in Machine Learning and Computational Intelligence, Proceedings of ICMLCI, 2019.
- [4] Arunraj, N. S., Ahrens, D., Fernandes, M., Application of Sarimax Model to Forecast Daily Sales in Food Retail Industry, International Journal of Operations Research and Information Systems, p. 21, 2016.
- [5] Ensafi, Y., Amin, S. H., Zhang G., Shah, B., Time-series forecasting of seasonal items sales using machine learning – A comparative analysis, International Journal of Information Management Data Insights, pp. V2-I1, 2022.
- [6] Uysal, M. C., Tarım Kredi Kooperatifleri Faizsiz Kredi Sistemi Projesi ve Alternatif Bir Model Önerisi, Uluslararası Yönetim Akademisi Dergisi, pp. Cilt: 3, Sayı: 2, ss.444-463, 2020.
- [7] Doğan, K., Özdeniz, A. H., Kömür Yığınlarında Kendiliğinden Yanma Olayına Güneş Işınlığının Etkisi, S.Ü. Müh. Bilim ve Tekn. Derg, pp. c3,s2, 2015.
- [8] Tamzok, N., Dünyada ve Türkiye'd Kömür 2019, TMMOB makina mühendisleri odası, Ankara, 2019.
- [9] <https://www.tarimkredi.org.tr/kurumsal/koklu-gecmisimiz/>. Erişim Tarihi: 14.04.2022

- [10] <https://www.tarimkredi.org.tr/media/i55eqzxx/2020-yılı-faaliyet-raporu.pdf>
Erişim Tarihi: 15.06.2022
- [11] İ. Kavaz., Yerli ve Milli Enerji, Seta Analiz, p. 265, 2019.
- [12] <https://isikomur.com/markalarimiz/efor-komur/>. Erişim Tarihi: 25.07.2022
- [13] De Gooijer, J. G., Hyndman, R. J., 25 years of time series forecasting, International Journal of Forecasting 22, pp. 443-473, 2006.
- [14] Duru, Ö., Zaman Serileri Analizinde ARIMA Modelleri ve Bir Uygulama, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, 2007.
- [15] Zhou, L., Zhao, P., Wu, D., Cheng C., Huang, H., Time series model for forecasting the number of new admission inpatients, MC Med Inform Decis Mak 18, p. 39, 2018.
- [16] <https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>. Erişim Tarihi: 22.04.2022
- [17] Tüzemen, A., Yıldız, Ç., Holt-Winters Tahminleme Yöntemlerinin Karşılaştırılmalı Analizi: Türkiye İşsizlik Oranları Uygulaması, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 2018.
- [18] Bonaccorso, G., Machine Learning Algorithms, Birmingham: Pacht Publishing Ltd, 2017.
- [19] ŞAHİN, D., Comparative analysis of XGBoost and LightGBM methods for day ahead spot natural gas price forecasting, Lisansüstü Eğitim Enstitüsü, İstanbul Teknik Üniversitesi, 2022.
- [20] Mahesh, B., Machine Learning Algorithms -A Review, International Journal of Science and Research (IJSR), 2019.
- [21] Ke, G., Meng, Q., Finley, T., Wang T., Chen, W., Ma, W., Ye, Q., Liu, T. Y., LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017.
- [22] Friedman, J., Stochastic gradient boosting, Computational Statistics & Data Analysis, pp. 367-378, 2002.

- [23] Chen, T., Guestrin, C., XGBoost: A Scalable Tree Boosting System, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining., 2016.
- [24] Lord, D., Qin, X., Geedipally, S. R., Chapter 2 Fundamentals and data collection, %1 içinde Highway Safety Analytics and Modeling, Elsevier, 2021, pp. 17-57.
- [25] Önen, V., ARIMA yöntemiyle Türkiye'nin Hava Yolu Kargo Talep Tahmin Modellemesi ve Öngörüsü, cilt 18, no. 29-53, 2020.
- [26] <https://veribilimcisi.com/2017/07/14/mse-rmse-mae-mape-metrikleri-nedir/>. Erişim Tarihi: 10.04.2022.
- [27] Yeşilyurt, S. N., Dalkılıç, H. Y., Daily River Flow Forecasting with Xgboost and Gradient Boost Machine, 3 rd International Symposium of Engineering Applications on Civil Engineering and Earth Sciences 2021, pp. 36-44, 2021.
- [28] Schapire, R., A Brief Introduction to Boosting, Proceedings of the Sixteenth International Joint, 1999.
- [29] Ogunleye, A., Wang Q. G., XGBoost Model for Chronic Kidney Disease Diagnosis, IEEE/ACM Trans Comput Biol Bioinform., pp. 2131-2140, 2020.
- [30] Bhattacharya, S., Maddikunta, P., Kaluri, R., Singh, S., Gadekallu T., Alazab, M., Tariq, U., A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU, Electronics, pp. 9, 219, 2020.
- [31] Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F., El-Shafie, A., Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia, Ain Shams Engineering Journal Volume 12, Issue 2, pp. 1545-1556, 2021.
- [32] Qin, C., Zhang, Y., Bao, F., Zhang, C., Liu, P., XGBoost Optimized by Adaptive Particle Swarm Optimization for Credit Scoring, Mathematical Problems in Engineering, pp. 1-18, 2021.
- [33] Zhenchang, X., Xue, S., Wu, L., Sun, J., ForeXGBoost: passenger car sales prediction based on XGBoost, Distributed and Parallel Databases 38, 2020.
- [34] Ji. S., Wang, X., Zhao, W., Guo, D., An Application of a Three-Stage XGBoost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise, Mathematical Problems in Engineering, pp. 1-15, 2019.

- [35] Wali, S. M., Viviani J. L., Jabeur S. B., Forecasting gold price with the XGBoost algorithm and SHAP interaction values, *Annals of Operations Research*, 2021.
- [36] Zhang, L., Bian W., Qu, W., Tuo L., Time series forecast of sales volume based on XGBoost, *Journal of Physics: Conference Series*, p. 012067, 2021.
- [37] Wei, H., Zeng, Q., Research on sales Forecast based on XGBoost-LSTM algorithm Model, *Journal of Physics: Conference Series*, p. 012191, 2021.
- [38] Adebiyi, A., Adewumi, A., Ayo. C., Stock Price Prediction Using the ARIMA Model, *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 2014.
- [39] Yenidoğan, I., Çayır, A., Kozan, O., Dağ, T., Arslan, Ç., Bitcoin Forecasting Using ARIMA and PROPHET, %1 içinde 3rd International Conference on Computer Science and Engineering UBMK'18, Bosnia, 2018.
- [40] Azari, A., Papapetrou, P., Denic, S., Peters, G., Cellular Traffic Prediction and Classification: A Comparative Evaluation of LSTM and ARIMA, *Discovery Science*, pp. 129-144, 2019.
- [41] Kufel, T., ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries, *Equilibrium. Quarterly Journal of Economics and Economic Policy*, pp. 181-204, 2020.
- [42] Nebati, E. E., Taş, M., Ertaş, G., Türkiye’de Elektrik Tüketiminde Talep Tahmini: Zaman Serisi ve Regresyon Analizi ile Karşılaştırma, *European Journal of Science and Technology*, pp. 348-357, 2021.
- [43] Shukla, M., ve Jharkharia, S., Applicability of ARIMA Models in Wholesale Vegetable Market: An Investigation, *International Journal of Information Systems and Supply Chain Management*, pp. 105-119, 2013.
- [44] Harsoor, A. S., ve Patil A., Forecast of Sales od Walmart Store Using Big Data, *IJRET: International Journal of Research in Engineering and Technology*, 2015.
- [45] Ohyvera, M., Pudjihastuti, H., Arima Model for Forecasting the Price of Medium Quality Rice to Anticipate Price Fluctuations, *3rd International Conference on Computer Science and Computational Intelligence*, 2018.
- [46] Pavlyshenko, B. M., Machine-Learning Models for Sales Time Series Forecasting, *IEEE*, 2019.

- [47] Uçar, K., Güler, D., Engindeniz, S., 2021. Türkiye'de Kayısı Üretimin ARIMA Model ile Tahmini, Tarım Ekonomisi Dergisi, pp. 55-62.
- [48] Palabıçak, M. A., Türkiye'de Kırmızı Et Sektörü ve Geleceğe Yönelik Üretim ve Tüketim Dengesinin Analizi, Harran Üniversitesi Fen Bilimleri Enstitüsü Tarım Ekonomisi Anabilim Dalı, 2019.



ÖZGEÇMİŞ

Adı Soyadı : Gökhan ACAR

ÖĞRENİM DURUMU

Derece	Eğitim Birimi	Mezuniyet Yılı
Yüksek Lisans	Sakarya Üniversitesi / Fen Bilimleri Enstitüsü / Bilişim Sistemleri Mühendisliği	Devam ediyor
Lisans	Ege Üniversitesi / Mühendislik Fakültesi / Bilgisayar Mühendisliği	2012
Lise	İzzet Baysal Anadolu Lisesi	2004

İŞ DENEYİMİ

Yıl	Yer	Görev
2022-Halen	Ericsson Inc.	Software Developer
2019-2022	Tarnet A.Ş.	Kıdemli Yazılım Geliştirici
2016-2019	Acr Yazılım	Kurucu/Yazılım Geliştirici
2013-2016	Türk Hava Yolları A.Ş.	Yazılım Geliştirici

YABANCI DİL

İngilizce