



Python İle Fiyat Tahmini ve Görselleştirme Projesi

Yazılım Mühendisliği Ana Bilim Dalı

Bitirme Projesi

Zeynep BABALI

Y220234062

Tez Danışmanı: Dr. Öğr. Üyesi Serpil Yılmaz

Ağustos 2024

Python İle Fiyat Tahmini ve Grselleřtirme Projesi

Z

Bu alıřmada petrol ve doęal gaz fiyatlarının kresel plastik fiyatlarının belirlenmesinde oynadıęı rol arařtırılmıřtır. Plastik yapımında kullanılan hammaddelerin byk bir kısmı petrokimya endstrisinden gelmektedir. Ham petrol ve doęalgaz fiyatlarındaki dalgalanmalar plastięin retim maliyetini ve dolayısıyla piyasadaki plastięin fiyatını doęrudan etkilemektedir. Plastik fiyatları zerindeki etkiyi analiz etmek iin bir nceki yıllara ait ham petrol ve doęal gaz fiyatı verileri kullanılmıřtır. Elde edilen sonular enerji piyasasındaki deęiřimlerin plastik sektr zerindeki etkisini gstermekte ve gelecekteki fiyat deęiřimlerini tahmin etmeyi mmkn kılmaktadır.

Anahtar Szckler: Plastik Fiyatları, Ham Petrol Fiyatları, Doęalgaz Fiyatları, Enerji Piyasaları, Petrokimya Endstrisi, Fiyat Hareketleri Analizi, Enerji Maliyetlerinin Plastik zerine Etkisi

Price Forecasting and Visualization Project with Python

Abstract

This study investigates the role of crude oil and natural gas prices in determining global plastic prices. A significant portion of the raw materials used in plastic production comes from the petrochemical industry. Fluctuations in crude oil and natural gas prices directly impact the production cost of plastic, and consequently, the market price of plastic. To analyze the impact on plastic prices, historical data on crude oil and natural gas prices from previous years were utilized. The results obtained demonstrate the effect of changes in the energy markets on the plastic sector, enabling predictions of future price fluctuations.

Keywords: Plastic Prices, Crude Oil Prices, Natural Gas Prices, Energy Markets, Petrochemical Industry, Price Movement Analysis, Impact of Energy Costs on Plastics

İçindekiler

Öz.....	i
Abstract	ii
Şekiller Listesi.....	v
Kısaltmalar Listesi	vi
1. Bölüm Giriş	1
2. Bölüm	2
2.1 Veri Analizi ve Python ile Veri Analizinin Önemi.....	2
2.2 Veri Bilimine Katkı Sağlayan Önemli Bilim İnsanları	3
3. Bölüm	5
3.1 Veri Bilimi, Veri Analizi ve Görselleştirme: Verilerden Bilgiye Yolculuk.....	5
3.1.1 Veri Analizi.....	5
3.1.2 Keşifsel Veri Analizi (EDA).....	5
3.1.3 İstatistiksel Analiz.....	5
3.1.4 Veri Görselleştirme	6
3.1.5 Görselleştirme Araçları.....	6
3.1.6 Görselleştirme Teknikleri	6
3.2 Veri Bilimi Ekosisteminin Temel Taşlarını Oluşturan Önemli Araçlar.....	7
3.2.1 Pandas	7
3.2.2 NumPy.....	7
3.2.3 Matplotlib	7
3.2.4 Seaborn.....	8

3.2.5. Scikit-Learn	8
3.2.6. TensorFlow ve Keras.....	9
3.2.7. Tableau ve Power BI	9
3.3 Veri Manipülasyonu	10
3.3.1 Veri Manipülasyonunun Tanımı	10
3.3.2 Veri Manipülasyonunun Aşamaları	10
3.3.3 Veri Manipülasyonu Araçları.....	11
3.3.4 Veri Manipülasyonunun Önemi.....	11
3.3.5 Veri Manipülasyonunun Uygulama Alanları	11
4. Bölüm	12
4.1 Python İle Fiyat Tahmini ve Görselleştirme Projesi	12
4.2 Proje Kod Analizi	13
4.2.1 Veri Yükleme ve Ön İşleme.....	13
4.2.2 Veri Ölçeklendirme.....	13
4.2.3 Zaman Serisi Tahmin Modeli Oluşturma.....	15
4.2.4 Model Eğitimi	16
4.2.5 Tahminlerin Yapılması ve Görselleştirme	16
4.2.6 Sonuçların Değerlendirilmesi	17
5. Bölüm	18
5.1 Sonuç	18
KAYNAKLAR	19

Şekiller Listesi

Şekil 1 PVC Fiyatları - 2023\'den Şu Bugüne Kadar SPVC K 67-68 Değerleri (Eksik Veriler Kaldırıldı).....	14
Şekil 2 Hampetrol Verileri - 2023\'den Şu Bugüne Kadar Şimdi Değerleri.....	14
Şekil 3 Doğalgaz Fiyatları - 2023\'den Şu Bugüne Kadar Doğalgaz Değerleri (Eksik Veriler Kaldırıldı).....	15
Şekil 4 LSTM Modeli İle PVC Fiyat Tahmini	17

Kısaltmalar Listesi

ANOVA	Varyans Analizi (Analysis of Variance):
EDA	Keşifsel Veri Analizi (Exploratory Data Analysis)
FFT	Hızlı Fourier Dönüşümü (Fast Fourier Transform)
LSTM	Uzun Kısa Süreli Bellek (Long Short-Term Memory)

1. Bölüm Giriş

Dünya çapında plastik endüstrisi sürekli değişen ekonomik ortamdan ve pazar trendlerinden etkilenmektedir. Aynı zamanda sektör, ham petrol ve doğal gaz başta olmak üzere temel enerji kaynaklarının fiyatlarından da doğrudan etkilenmektedir. Plastik üretmenin maliyeti bu enerji kaynaklarının maliyetiyle doğrudan bağlantılı olduğundan fiyat dalgalanmaları sektör maliyetlerini önemli ölçüde etkilemektedir. Bu nedenle plastik fiyatlarındaki değişimleri anlamak ve tahmin etmek için veri analizi ve tahmin yöntemleri büyük önem taşımaktadır. Bu teknikler, piyasa koşullarının daha net anlaşılmasını ve gelecekteki eğilimlerin doğruya tahmin edilmesini sağlayarak şirketlerin stratejik karar alma süreçlerini destekler.

Plastik fiyat tahminleri bir şirketin satın alma ve dağıtım stratejisinin belirlenmesinde önemli bir rol oynamaktadır. Bu tahminler pazarda rekabet avantajı elde etmek ve maliyet kontrolünü optimize etmek için kullanılabilir. Üstelik rakiplerle yapılan fiyat analizi, pazar konumunuzu güçlendirmek ve müşteri gereksinimlerini daha iyi karşılamak için önemli bir araçtır. Bu nedenle enerji piyasası fiyat dalgalanmalarının detaylı analizi, plastik üreticileri ve tüketicilerine stratejik bilgi sağlamaktadır. Bu çalışma, petrol ve doğal gaz fiyatları ile plastik fiyatları arasındaki ilişkinin ve bu bilgilerin stratejik planlama süreçlerine nasıl entegre edilebileceğinin araştırılmasını amaçlamaktadır.

2. Bölüm

2.1 Veri Analizi ve Python ile Veri Analizinin Önemi

Veri analitiği, anlamlı sonuçlar çıkarmak ve karar destek süreçlerini desteklemek için büyük miktardaki verileri inceler. Bu süreç özellikle büyük ve karmaşık veri yapılarına sahip sektörlerde stratejik öneme sahiptir. Python, veri analizi için en popüler programlama dillerinden biridir. Kapsamlı kütüphane ekosistemi (Pandas, NumPy, Matplotlib, Seaborn vb.) ve kullanıcı dostu yapısı verilerin işlenmesini, temizlenmesini, işlenmesini ve görselleştirilmesini kolaylaştırır. Veri-analitik düşünme, veriye dayalı karar verme süreçlerinin etkinliğini değerlendirmenizi sağlar. Örneğin, bir çalışan, danışman ya da potansiyel bir yatırım hedefi, veriden bilgi çıkarmayı teklif ediyorsa, bu teklifi sistematik olarak değerlendirebilmeli ve mantıklı olup olmadığını belirleyebilmelisiniz (Provost ve Fawcett, 2013).

Python ile veri analizi, veriler üzerinde ayrıntılı işlemler yapılmasına olanak sağlar. Bu süreçler, veri ön işlemlerinden karmaşık algoritmalar kullanılarak tahmine kadar uzanır. Tahminleme, işletmelerin özellikle değişken piyasa koşullarında gelecekteki eğilimleri tahmin edebilmeleri için çok önemlidir. Doğru tahmine dayalı modeller, pazar değişikliklerine hızla uyum sağlama ve rekabet avantajı elde etme fırsatı sağlar. Python, bilimsel hesaplama ve veri analizi için uygun bir dil olmasının yanı sıra, genel amaçlı yazılım geliştirme için de iyi bir seçimdir. Python'un gelişmiş kütüphaneleri, veri analizi görevleri için güçlü bir platform sunar (McKinney, 2018).

Görselleştirme, karmaşık bilgileri anlamanıza ve verilerinizi anlamlandırmanıza yardımcı olan bir araçtır. Grafikler, çizelgeler ve etkileşimli kontrol panelleri, büyük veri kümelerindeki bilgileri daha etkili bir şekilde görüntülemenize olanak tanır. Bu görsel temsiller, verilerdeki eğilimleri, kalıpları ve ilişkileri hızlı bir şekilde ortaya çıkararak yöneticilerin stratejik kararlar alırken bilinçli kararlar almasına olanak tanır. Yöneticiler, veri analizinden şirket sorunlarını çözmek, fırsatları değerlendirmek ve riskleri yönetmek için kullanılacak eyleme dönüştürülebilir ve stratejik bilgiler beklerler.

Etkin veri analizi sadece mevcut durumu deęerlendirmekle kalmaz, aynı zamanda gelecek senaryolarını tahmin ederek proaktif aksiyon alınmasını saęlar. Bu nedenle veri analizi süreci kapsamında elde edilen sonuçların doęru yorumlanması bir řirketin sürdürülebilir başarısında önemli rol oynamaktadır. Veri bilimini anlamak, onu kendiniz uygulamak zorunda olmasanız bile önemlidir, çünkü veri analizi artık iş stratejisinin çok kritik bir parçası haline gelmiştir. İşletmeler giderek daha fazla veri analitięi tarafından yönlendirildięi için, bu tür işletmelerle yetkin bir şekilde etkileşimde bulunabilmek büyük bir profesyonel avantaj saęlar (Provost ve Fawcett, 2013).

2.2 Veri Bilimine Katkı Saęlayan Önemli Bilim İnsanları

2.2.1 Ronald Fisher

Ronald A. Fisher, modern istatistięin babası olarak kabul edilir.20.yüzyılın başlarında Fisher, varyans analizi (ANOVA) ve maksimum olasılık tahmini gibi istatistiksel yöntemlerin geliştirilmesine öncülük etti.

Bu teknikler halen veri analizinin temelini oluşturmakta ve bilimin çeşitli alanlarında yaygın olarak kullanılmaktadır.

2.2.2 John Tukey

John Tukey, veri analizi ve bilgisayar bilimi üzerinde büyük etkisi olan Amerikalı bir matematikçidir. Keşif amaçlı veri analizi kavramını popüler hale getirdi ve veri kümelerini anlamak için görselleştirme ve grafik tekniklerinin kullanılmasını savundu. Tukey ayrıca Hızlı Fourier Dönüşümü (FFT) algoritmasının geliştirilmesine de katkıda bulundu.

2.2.3 Claude Shannon

Claude Shannon bilgi teorisinin kurucusu olarak kabul edilir. Bu teori veri iletişimi, şifreleme ve veri sıkıştırma tekniklerinin temelini oluşturur.

Shannon'ın arařtırması dijital aęın řekillenmesinde ve veri biliminin temelini oluřturan bilgi iřlem teknolojilerinin geliřtirilmesinde temel bir rol oynadı.

2.2.4 William S. Cleveland

William S. Cleveland, veri bilimi teriminin yaratıcısı olarak kabul edilir ve modern veri analizinin öncülerinden biridir. Cleveland, veri görselleřtirme ve regresyon modelleme alanındaki arařtırmalarıyla tanınmaktadır. Veri analizinde istatistiksel yöntemlerin önemini vurgulayan arařtırmalar da yürütmüřtür.

2.2.5 Edward Tufte

Edward Tufte bilgi tasarımı ve veri görselleřtirmenin öncülerinden biridir. Tufte'nin arařtırması karmařık veri kümelerinin etkili görselleřtirilmesine ve sunumuna odaklanmaktadır. Kitapları ve seminerleri, bilim adamlarına, mühendislere ve iř adamlarına veri sunumunda rehber olarak kabul edilmektedir.

Bu bilim insanları, veri analizi ve bilimde temel kavram ve teknikleri geliřtirerek alanın řekillenmesinde önemli bir rol oynamıřlardır. Veri biliminin günümüzde nasıl uygulandıęını ve geliřtięini anlamak için her birinin katkısı çok önemlidir.

3. Bölüm

3.1 Veri Bilimi, Veri Analizi ve Görselleştirme: Verilerden Bilgiye Yolculuk

Veri bilimi, büyük ve karmaşık veri kümelerinden anlamlı bilgiler çıkarmak için matematik, istatistik, yapay zekâ ve bilgisayar bilimi gibi alanların kesişiminde yer alır. Günümüzde veri bilimi, iş dünyasından tıbbı, hükümet politikasından akademik araştırmaya kadar geniş bir uygulama alanına sahiptir. Veri biliminin temel amacı verileri soruları yanıtlamak, tahminlerde bulunmak ve stratejik kararlar almak için gereken bilgileri sağlamak için kullanmaktır.

3.1.1 Veri Analizi

Veri analizi, veri biliminin temelidir ve büyük miktarda veriyi işleyerek sorunları tanımlamak ve çözüm bulmak için kullanılır. Veri analizi süreci; veri toplama, veri temizleme, keşfedici veri analizi (EDA), istatistiksel test etme ve modelleme gibi aşamalardan oluşur. Etkili veri analizi, verilerden elde edilen bilgilerin doğru ve kullanılabilir olmasını sağlamak için dikkatli ve sistematik bir yaklaşım gerektirir.

3.1.2 Keşifsel Veri Analizi (EDA)

EDA, verileri daha iyi anlamak ve analiz için uygun model ve teknikleri belirlemek için ilk adımdır. Bu süreç, verinin yapısının, eksik değerlerin, aykırı değerlerin ve değişkenler arasındaki ilişkilerin gözlemlenmesini içerir.

3.1.3 İstatistiksel Analiz

Veri analizi için kullanılan istatistiksel yöntemler, verilerin üzerinde çeşitli hipotez testleri gerçekleştirerek sonuçlarını genel geçerliliğini değerlendirilmesine olanak tanır. Regresyon analizi, varyans analizi (ANOVA) ve zaman serisi analizi gibi teknikler sıklıkla kullanılır.

3.1.4 Veri Görselleştirme

Veri görselleştirme, analiz sürecinin açık ve etkileşimli bir temsilini sağlar. İyi tasarlanmış görselleştirmeler, karmaşık veri kümelerinin yorumlanmasını kolaylaştırarak iç görü paylaşımını ve stratejik karar almayı kolaylaştırır.

3.1.5 Görselleştirme Araçları

Matplotlib, Seaborn ve Plotly gibi Python kütüphaneleri, çeşitli grafik türleri oluşturarak veri görselleştirmeyi destekler. Pano oluşturmak için Tableau ve Power BI gibi araçlar önerilir.

3.1.6 Görselleştirme Teknikleri

Veri görselleştirmede kullanılan teknikler arasında çubuk grafikler, çizgi grafikler, dağılım grafikleri, ısı haritaları ve diğer birçok yöntem bulunur. Bu tekniklerin seçimi veri türüne ve analizin amacına bağlıdır.

Veri bilimi, veri analizi ve görselleştirme günümüz dünyasında bilinçli karar vermenin temelidir. Bu disiplinler, verilerin doğru şekilde toplanmasını, işlenmesini, analiz edilmesini ve sonuçların etkili bir şekilde sunulmasını sağlayarak kuruluşların ve bireylerin daha bilinçli kararlar almasına yardımcı olur. İleriye bakıldığında, veri bilimi ve görselleştirmenin yapay zekâ ve makine öğrenimiyle entegrasyonu, bu alanın gelişmeye ve genişlemeye devam edeceğini göstermektedir.

3.2 Veri Bilimi Ekosisteminin Temel Taşlarını Oluşturan Önemli Araçlar

3.2.1 Pandas

Pandas, Python 'da güçlü bir veri analizi ve veri işleme kütüphanesidir. Pandas, özellikle veri kümelerini düzenlemek, temizlemek ve analiz etmek için kullanılan DataFrame ve Seri adı verilen veri yapıları sağlar. Bu yapı, tablo biçiminde düzenlenmiş verilerin yönetilmesini kolaylaştırır. DataFrame, iki boyutlu bir veri yapısıdır. Satır ve sütunları içerir. Verileri Excel'deki tablolara benzer şekilde satırlar ve sütunlar halinde depolar. Seri, tek boyutlu bir veri yapısıdır.

Genel olarak bunu bir DataFrame 'in sütunları veya satırları olarak düşünebilirsiniz.

Pandas, verileri hızlı ve verimli bir şekilde okumanıza, yazmanıza, filtrelemenize, gruplandırmanıza ve birleştirmenize olanak tanır. NumPy, Python'da sayısal hesaplamalar için en önemli temel paketlerden biridir. Birçok hesaplama paketi, verileri değiş tokuş etmek için NumPy'nin dizi nesnelerini kullanır (McKinney, 2018).

3.2.2 NumPy

NumPy (Sayısal Python), Python 'da bilimsel hesaplama için temel kütüphanelerden biridir. NumPy, çok boyutlu dizileri ve matrisleri destekler ve bu veri yapıları üzerinde güçlü matematiksel işlemlere olanak tanır. ndarray, NumPy' ın temel veri yapısıdır ve n boyutlu dizilerin oluşturulmasına olanak tanır. Bu diziler matematiksel işlemler için optimize edilmiştir. NumPy, doğrusal cebir, Fourier dönüşümleri ve rastgele sayı üretimi gibi birçok bilimsel hesaplama tekniğini içerir. NumPy, Pandas ve diğer veri bilimi kütüphanelerinin temelini oluşturur ve veri manipülasyonu ve matematiksel hesaplamalar için yaygın olarak kullanılır. NumPy, Python'da sayısal hesaplamalar için en önemli temel paketlerden biridir. Birçok hesaplama paketi, verileri değiş tokuş etmek için NumPy'nin dizi nesnelerini kullanır (McKinney, 2018).

3.2.3 Matplotlib

Matplotlib, Python 'da veri görselleştirme için en çok kullanılan kütüphanelerden biridir. Matplotlib, aşağıdakiler de dâhil olmak üzere farklı grafik türlerini kullanarak verilerin görselleştirilmesine olanak tanır: Örnekler: çizgi grafikler, çubuk grafikler,

histogramlar, dağılım grafikleri vb. Pyplot, Matplotlib 'in en yaygın kullanılan alt modülüdür. Pyplot, hızlı ve kolay bir şekilde grafikler oluşturmak için MATLAB benzeri bir ara yüz sağlar. Matplotlib, grafikleri özelleştirmek için çok sayıda seçenek sunar. Matplotlib, grafikler ve diğer iki boyutlu veri görselleştirmeleri üretmek için en popüler Python kütüphanesidir. Orijinal olarak John D. Hunter tarafından oluşturulmuş olup, şu anda büyük bir geliştirici ekibi tarafından sürdürülmektedir. Yayınlanabilir kalitede grafikler oluşturmak için tasarlanmıştır (McKinney, 2018).

Eksenleri, renkleri, işaretleyicileri, çizgi stillerini ve daha fazlasını özelleştirir. Matplotlib, basit ve karmaşık grafiklerin yanı sıra etkileşimli grafikler oluşturmak için de uygundur.

3.2.4 Seaborn

Seaborn, Matplotlib 'i temel alan bir veri görselleştirme kütüphanesidir. İstatistiksel grafikler oluşturmayı ve verileri görselleştirmeyi kolaylaştıran üst düzey bir ara yüz sağlar. İstatistiksel grafikler, kutu grafikleri, keman grafikleri ve ısı haritaları gibi istatistiksel grafik türlerini kolayca oluşturur. Renk Paletleri, veri görselleştirmelerinde etkili renk paletlerini kullanmayı kolaylaştırarak grafiklerinizi daha okunabilir ve güzel hale getirir. Seaborn, Matplotlib 'den daha yüksek düzeyde soyutlama sağlar ve özellikle istatistiksel analizi görselleştirmek isteyenler için faydalıdır. Seaborn, özellikle karmaşık veri setlerini görselleştirirken veri bilimcileri ve analistler tarafından sıkça tercih edilen bir araçtır (McKinney, 2018).

3.2.5. Scikit-Learn

Scikit-Learn, Python için popüler bir makine öğrenimi kütüphanesidir. Denetimli ve denetimsiz öğrenme algoritmalarını, veri ön işleme araçlarını, model seçimini ve değerlendirme ölçümlerini içerir. Denetimli öğrenme, doğrusal regresyon, lojistik regresyon, karar ağaçları ve destek vektör makineleri gibi regresyon ve sınıflandırma algoritmalarını içerir. Denetimsiz öğrenme, kümeleme, boyutluluk azaltma ve ilişkilendirme algoritmalarını içerir. Model değerlendirmesi, Scikit-Learn çapraz doğrulama ve çeşitli performans ölçümleri (hassasiyet, kesinlik, geri çağırma, F1 puanı) sağlar. Scikit-learn, Python programcıları için makine öğrenimi modelleri oluşturmak amacıyla en çok kullanılan genel amaçlı bir araç setidir. Sınıflandırma,

regresyon, kümeleme, boyut azaltma ve model seçimi gibi birçok model ve araç içerir (McKinney, 2018).

Scikit-Learn, veri bilimi projelerinde yaygın olarak kullanılır ve özellikle makine öğrenimi modellerinin hızla geliştirilmesi ve değerlendirilmesi için uygundur.

3.2.6. TensorFlow ve Keras

TensorFlow ve Keras, derin öğrenme modelleri oluşturmak için kullanılan açık kaynak kitaplıklardır. TensorFlow, Google tarafından geliştirilmiştir ve büyük ölçekli sinir ağı modellerinin eğitimi için optimize edilmiştir. Keras ise TensorFlow üzerine kurulu, kullanımı kolay, üst düzey bir API'dir. TensorFlow, büyük, karmaşık modelleri verimli bir şekilde eğitmek için düşük seviyeli işlemler ve esneklik sağlar. Keras daha kullanıcı dostu ve kullanıcı dostu bir ara yüz sağlayarak daha hızlı bir model geliştirme sürecine olanak tanır. Bu kütüphaneler, derin öğrenme algoritmalarıyla çalışmak ve gelişmiş sinir ağı modellerini geliştirmek için yaygın olarak kullanılmaktadır.

3.2.7. Tableau ve Power BI

Tableau ve Power BI, veri görselleştirme ve iş zekâsı (BI) araçlarıdır. Bu araçlar, kullanıcıların veri kümelerini etkileşimli olarak keşfetmesine ve görselleştirmesine olanak tanır. Tableau, büyük miktarda veriyi analiz etmek ve hızla görselleştirmek için güçlü bir araçtır. Kullanıcı dostu ara yüzü sayesinde teknik bilgisi olmayan kullanıcılar bile karmaşık veri görselleştirmeleri oluşturabilir.

Power BI, Microsoft tarafından geliştirilen iş verileri analizi ve raporlama platformudur. Veri analistleri ve iş kullanıcıları için uygundur.

Bu araçlar veri analizi, veri görselleştirme ve makine öğrenimi projelerinde yaygın olarak kullanılan temel araçlardır ve veri bilimi ekosisteminin temelini oluşturur. Her biri belirli bir işlevi gerçekleştirmek üzere optimize edilmiştir ve veri analistlerinin ve iş zekâsı profesyonellerinin verileri etkili bir şekilde görselleştirmesine ve paylaşmasına yardımcı olmak için genellikle veri bilimi projelerinde birlikte kullanılır.

3.3 Veri Manipülasyonu

Veri manipülasyonu, ham verileri belirli bir formata dönüştürmek, düzenlemek veya analize hazır hale getirmek için gerçekleştirilen her türlü işlemi ifade eder. Veri manipülasyonu, veri analizi sürecinde kritik bir aşamadır ve veri bilimi projelerinin verilerden anlamlı iç görüler elde edilmesi için gereklidir. Bu tür bir probleme basit ama verimsiz bir yaklaşım, örnekleme durumunu ve tahmin ediciyi kara kutular olarak ele almaktır. (Andrews, 1972).

3.3.1 Veri Manipülasyonunun Tanımı

Veri manipülasyonu, ham veri üzerinde yapılan işlemlerle ham verinin istenilen formata dönüştürülmesi işlemidir. Veriler analize uygun hale getirilecek şekilde temizlenir, dönüştürülür ve yeniden düzenlenir. Veri manipülasyonunun amacı, veri setini analize uygun hale getirmek ve istediğiniz sorulara cevap verecek şekilde düzenlemektir. Python uzun süredir, kısmen dizgeler ve metin işleme konusunda kullanım kolaylığı nedeniyle, ham veri manipülasyonu dili olarak popüler olmuştur. Veri manipülasyonu, veri bilimcilerin işlenmemiş verileri alarak temiz, yapılandırılmış ve analiz edilebilir bir forma dönüştürme sürecidir. Bu süreç, veri üzerinde çeşitli işlemler yapmayı içerir, örneğin verileri ayıklama, dönüştürme, birleştirme ve yeniden şekillendirme gibi. (McKinney, 2018).

3.3.2 Veri Manipülasyonunun Aşamaları

Veri temizliği, hataları, eksik veriler, aykırı değerler ve ham veri tutarsızlıkları düzeltilir. Örneğin boş hücreler doldurulur, veri türleri ayarlanır ve aykırı değerler kaldırılır. Veri dönüştürme, verisi analizde kullanılan formata dönüştürülür. Bu, veri türlerinin yeniden boyutlandırılmasını, yeniden şekillendirilmesini, normalleştirilmesini, standartlaştırılmasını veya değiştirilmesini içerebilir. Örneğin kategorik verileri sayısal verilere dönüştürülebilir (kodlayabilir) veya birimleri değiştirilebilir. Verileri Birleştirme, farklı veri kaynaklarından gelen veri birleştirilir. Aynı veri kaynağı içerisindeki farklı tablolardan veya tamamen farklı veri kaynaklarından veri toplamak için birleştirme işlemleri gerçekleştirilir. Farklı veri kümeleri birleştirme yöntemleri (iç birleştirme, dış birleştirme, sol birleştirme, sağ birleştirme) kullanılarak birleştirilir. Veri filtreleme, belirli kriterleri karşılayan alt kümeler, veri kümesinden çıkarılır. Örneğin, belirli bir tarih aralığına veya belirli bir kategorideki ürünlere ilişkin verileri filtrelenebilir. Veri Toplama/Gruplama, veri

gruplandırma işlemi, verileri belirli kategorilere veya gruplara gruplandırır. Örneğin belirli bir müşteri segmenti için ortalama satışları hesaplanabilir

3.3.3 Veri Manipülasyonu Araçları

Veri işleme genellikle aşağıdaki araçlar ve kitaplıklar kullanılarak yapılır.

Pandas, Python 'da veri işleme için en yaygın kullanılan kütüphanedir. DataFrame ve seri yapıları, verileri temizleme, dönüştürme, gruplama ve birleştirme gibi işlemler için güçlü araçlar sağlar.

SQL, veri tabanlarından veri almak, filtrelemek ve dönüştürmek için kullanılır. Özellikle büyük miktarda veri için veri işleme amacıyla yaygın olarak kullanılır.

Excel, küçük ve orta ölçekli veri kümelerinin veri düzenlemesine yönelik bir araçtır.

Verileri temizleme, dönüştürme ve filtreleme gibi temel işlemleri destekler.

3.3.4 Veri Manipülasyonunun Önemi

Veri manipülasyonu, verileri analiz etmede ve anlamlı iç görüler oluşturmada kritik bir adımdır. Ham veriler genellikle değerlendirilebilir bir formatta mevcut olmadığından farklı manipülasyon süreçlerinin uygulanması gerekir. Veri manipülasyonu olmadan veriler eksik, hatalı veya çelişkili olabilir ve bu da hatalı sonuçlara ve yanlış kararlara yol açabilir. Bir veya daha fazla girdi özelliği için eksik değerlere sahip gözlemlerle karşılaşmak oldukça yaygındır. Yaygın yaklaşım, bu eksik değerleri bir şekilde doldurmaktır (Hastie, Tibshirani, & Friedman, 2001).

3.3.5 Veri Manipülasyonunun Uygulama Alanları

Veri manipülasyonu çeşitli alanlarda kullanılmaktadır. İş Analitiği, satış verilerinin analiz edilmesi, müşterilerin kategorilere ayrılması ve pazarlama kampanyalarının etkinliğinin ölçülmesi gibi iş süreçlerinde kullanılır. Finans, risk yönetimi, portföy optimizasyonu, finansal tahmin ve analiz için kullanılır. Sağlık, hasta verilerinin analiz edilmesi, hastalık eğilimlerinin belirlenmesi, klinik araştırmalar için veri setlerinin hazırlanması gibi uygulamalarda kullanılır.

Veri manipülasyonu, veri biliminin önemli bir parçasıdır ve verilerden maksimum fayda sağlamak için yapılması gereken temel işlemlerden biridir.

4. Bölüm

4.1 Python İle Fiyat Tahmini ve Görselleştirme Projesi

Bu projenin amacı enerji piyasasındaki ham petrol ve doğalgaz fiyatlarının plastik sektörü üzerindeki etkisini analiz etmektir. Çalışma Python programlama dili kullanılarak geliştirilmiş olup, zaman serisi analizi ve makine öğrenimi teknikleri kullanılmıştır.

Bu proje, enerji fiyatlarındaki dalgalanmaların plastik fiyatları üzerindeki doğrudan etkisini araştırmaktadır. Petrol ve doğal gaz gibi temel emtia fiyatlarına ilişkin geçmiş verileri toplanmıştır. Özellikle bu veriler, verilerin ön işlenmesi için Python 'un Pandas ve NumPy kitaplıkları kullanılarak işlenmiştir. Fiyat verileri, modeli eğitmeye uygun hale getirmek için MinMaxScaler kullanılarak ölçeklendirilmiştir.

Veri seti üzerinde Uzun Kısa Süreli Bellek (LSTM) ağı kullanılarak bir zaman serisi tahmin modeli geliştirildi. Bu model, geçmiş verilere dayanarak gelecekteki fiyat hareketlerini tahmin etmek için eğitilmiştir. Modelin eğitimi için TensorFlow ve Keras Kütüphanesi kullanıldı. Ayrıca model performansını gerçek verilerle karşılaştırarak değerlendirilmiştir.

Projenin bir diğer önemli parçası olarak veri görselleştirme kullanılmış olup fiyat eğilimleri ve tahmin sonuçları Matplotlib ve Seaborn kütüphaneleri kullanılarak görsel olarak temsil edilmiştir. Bu görselleştirmeler analiz sonuçlarının daha derinlemesine anlaşılmasını sağlar ve stratejik karar alma sürecini desteklemektedir.

Sonuç olarak bu proje, enerji piyasasındaki fiyat dalgalanmalarının plastik sektörü üzerindeki etkisinin detaylı bir analizini sunmakta ve bu bilgilere dayanarak şirketlere stratejik önerilerde bulunmaktadır.

Bu proje aynı zamanda veri bilimi tekniklerinin ve araçlarının sektör analizi için nasıl etkili bir şekilde kullanılabileceğini de göstermektedir.

4.2 Proje Kod Analizi

4.2.1 Veri Yükleme ve Ön İşleme

Kodun ilk aşaması olan bu bölümü, enerji fiyatlarına dair verileri yüklemek ve ön işleme tabi tutmak için oluşturulmuştur. Pandas kütüphanesi kullanılarak Excel'den veri çekilmiş, veri tarihlerine göre sıralanmış ve eksik veriler temizlenmiştir. Bu işlem, analiz için veri setinin temiz ve işlenmiş bir formatta olmasını sağlar. Temel bileşen analizinin bir diğer uygulaması veri ön işlemedir. Bu durumda, amaç boyutluluğu azaltmak değil, veri kümesinin belirli özelliklerini standartlaştırmaktır. Bu, sonrasında uygulanacak desen tanıma algoritmalarının veri kümesine başarıyla uygulanabilmesini sağlamak açısından önemli olabilir (Bishop, 2006).

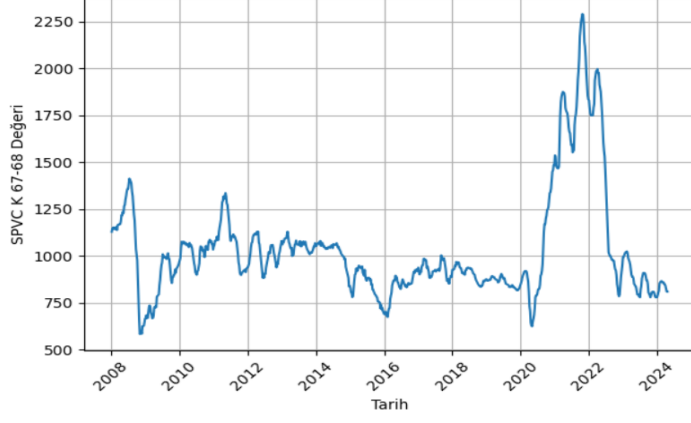
```
data = pd.read_excel(filepath)
data['Tarih'] = pd.to_datetime(data['Tarih'])
data.dropna(inplace=True)
data.set_index('Tarih', inplace=True)
```

4.2.2 Veri Ölçeklendirme

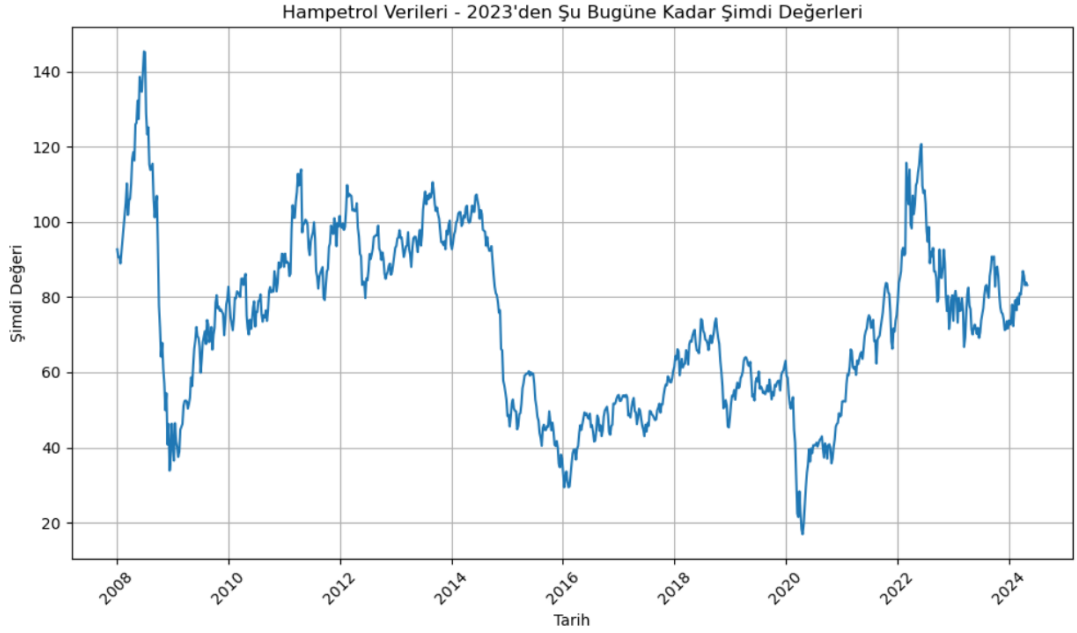
MinMaxScaler kullanılarak fiyat verileri ölçeklendirilmiştir çünkü bu adım, LSTM modelinin daha etkili öğrenmesi için gereklidir. Ölçeklendirilmiş veriler, algoritmanın daha hızlı ve daha doğru çalışmasına yardımcı olur. Veri ön işleme yönelik bir diğer yaygın uygulama, veri kümesini sıfır ortalama ve birim kovaryansa sahip olacak şekilde dönüştürmektir. Bu işlem, verilerin farklı değişkenlerini dekorrele etmek amacıyla kullanılır (Bishop, 2006).

```
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(prices)
```

PVC Fiyatları - 2023'den Şu Bugüne Kadar SPVC K 67-68 Değerleri (Eksik Veriler Kaldırıldı)



Şekil 1 PVC Fiyatları - 2023'den Şu Bugüne Kadar SPVC K 67-68 Değerleri (Eksik Veriler Kaldırıldı)



Şekil 2 Hampetrol Verileri - 2023'den Şu Bugüne Kadar Şimdi Değerleri



Şekil 3 Doğalgaz Fiyatları - 2023'den Şu Bugüne Kadar Doğalgaz Değerleri (Eksik Veriler Kaldırıldı)

4.2.3 Zaman Serisi Tahmin Modeli Oluşturma

TensorFlow ve Keras kullanılarak bir LSTM modeli kurulmuştur. LSTM, zaman serileri verileri için uygundur bunun sebebi geçmiş bilgileri hatırlama yeteneğinin olması ve bu bilgileri gelecek tahminleri yapmak için kullanabilir yeteneğe sahip olmasıdır. Özellikle finansal tahminler gibi uygulamalarda, bir zaman serisindeki önceki değerlerin gözlemlerine dayanarak bir sonraki değeri tahmin etmek önemlidir. Ardışık gözlemler genellikle yüksek derecede korelasyonludur ve bu nedenle gelecekteki değerleri tahmin etmede daha eski gözlemlere göre daha bilgilendiricidir. (Bishop, 2006).

```
model = Sequential()  
model.add(LSTM(units=50, return_sequences=True, input_shape=(X_train.shape[1],  
1)))  
model.add(LSTM(units=50))  
model.add(Dense(1))
```

4.2.4 Model Eğitimi

Model, önceden belirlenen epoch ve batch size değerleri kullanılarak eğitilmiştir. Model eğitimi sırasında, hem eğitim hem de test verileri üzerindeki kayıp azaltılmaya çalışılmıştır. Bu oran talebe, istenilen sonuca göre belirlenebilir farklı oranlarda denenme şansına sahiptir. Eğitim hatası, model karmaşıklığını artırdığımızda, yani veriye daha sıkı uyum sağladığımızda azalma eğilimindedir. Ancak, aşırı uyum sağlandığında model, eğitim verilerine çok sıkı bir şekilde adapte olur ve genelleme konusunda başarısız olur (Hastie, Tibshirani, & Friedman, 2001).

Model karmaşıklığının, veri setinin büyüklüğüne göre ayarlanmasının önemine değinir. Özellikle, modelin aşırı öğrenim (overfitting) sorununu önlemek için model karmaşıklığının uygun şekilde ayarlanması gerektiği vurgulanır. Ayrıca, Bayes yaklaşımı benimsenerek aşırı öğrenim probleminin önlenilebileceği belirtilir. Bu yaklaşım, parametre sayısının veri seti büyüklüğüne otomatik olarak ayarlanmasına olanak tanır (Bishop, 2006).

```
model.compile(optimizer='adam', loss='mean_squared_error')
model.fit(X_train, y_train, epochs=100, batch_size=32)
```

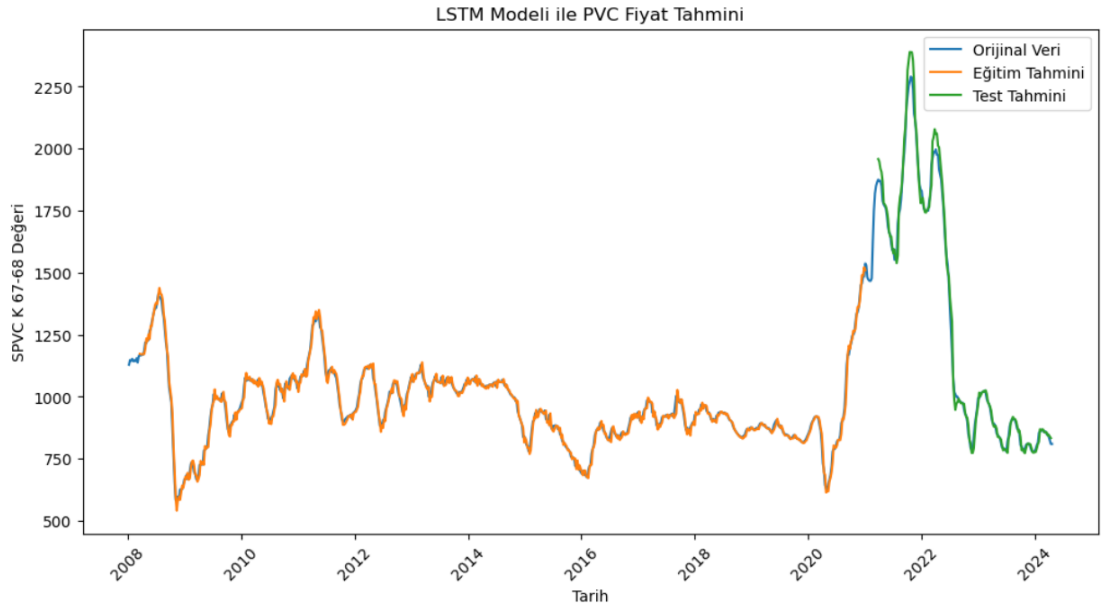
4.2.5 Tahminlerin Yapılması ve Görselleştirme

Modelin eğitilmesinin ardından, eğitim ve test verileri üzerinde tahminler yapılmış ve gerçek verilerle karşılaştırılması ile sonuçlar görselleştirilmiştir. Matplotlib kütüphanesi, sonuçların grafik üzerinde sunulması için kullanılmıştır.

```
train_predict = model.predict(X_train)
test_predict = model.predict(X_test)
# Gerçek ve tahmin edilen değerlerin grafik üzerinde gösterilmesi
plt.plot(history.history['loss'], label='Train Loss')
plt.plot(history.history['val_loss'], label='Test Loss')
plt.title('Model Loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend()
plt.show()
```

4.2.6 Sonuçların Değerlendirilmesi

Projenin sonunda, modelin performansı değerlendirilmiştir. Ardından elde edilen tahminler üzerinden stratejik öneriler geliştirilmiştir. Bu adım, projenin iş dünyası uygulamaları açısından sonuçlarının değerlendirilmesinde kritik bir öneme sahiptir.



Şekil 4 LSTM Modeli İle PVC Fiyat Tahmini

5. Bölüm

5.1 Sonuç

Bu projede petrol ve doğal gaz gibi temel enerji kaynaklarının plastik sektörü üzerindeki etkisi daha yakından incelendi. Python 'da geliştirilen bir zaman serisi tahmin modeli, enerji fiyatlarındaki değişikliklerin plastik fiyatları üzerindeki doğrudan etkisini yüksek bir başarı oranıyla tahmin etti. Bu başarı modelin daha sık eğitimi ve testiyle artış gösterebilmektedir.

LSTM ağı kullanılarak eğitilen model, geçmiş verilere dayanarak gelecekteki fiyat hareketlerini tahmin etmiş ve bu tahminler, model doğrulama sürecinde gerçek verilerle karşılaştırılarak değerlendirilmiştir.

Veri görselleştirme teknikleri, tahmin sonuçlarını ve trend analizini açık ve etkili bir şekilde görüntülenmiş olup, bu görselleştirmeler yalnızca model performansını göstermekle kalmamakla birlikte, aynı zamanda analiz süreci sırasında sonuçları destekleyecek bilgiler de sağlamıştır. Model eğitimi sırasında gözlemlenen kayıp oranı ve öğrenme eğrisi, algoritmanın optimizasyon sürecini ortaya koymuştur.

Sona yaklaşırken bu çalışma, enerji fiyatlarındaki dalgalanmaların plastik sektörü üzerindeki etkisini niceliksel olarak ölçmüş ve stratejik karar alma süreçlerinde kullanılmak üzere bilgi sağlamıştır.

Bu projeden elde edilen bilgiler, şirketlerin satış ve satın alma stratejilerini optimize etmelerine yardımcı olacak, ancak aynı zamanda pazarda rekabet avantajı elde etmek ve maliyet kontrolünü geliştirmek için de kritik önem taşımaktadır.

Bu çalışma, veri bilimi ve makine öğrenimi tekniklerinin endüstri analizi için ve stratejik iş kararlarına rehberlik etmek için nasıl etkili bir şekilde kullanılabileceğini başarıyla göstermiştir.

KAYNAKLAR

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). *Robust estimates of location: survey and advances*. Princeton University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

McKinney, W. (2017). *Python for data analysis: Data wrangling with pandas, numpy, and ipython* (2nd ed.). O'Reilly Media.