

Literatür Araştırması

1. Makine öğrenme çalışmaları ile ilgili yapılan genel araştırmalar

Bu bölümde makine öğrenme çalışmalarının sektör bazında kullanımına dair bir inceleme yapılmıştır. Bu kısımdaki literatür taraması ekonomi, eğitim ve sağlık gibi sektör bazındaki alanlara dair yapılan referans çalışmalar incelenerek yapılmıştır. Makine öğrenmesi, esas olarak 1959 yılında bilgisayar biliminin yapay zekada sayısal öğrenme ve model tanıma çalışmalarından geliştirilmiş bir alt dalıdır. Matematiksel ve istatistiksel yöntemler kullanarak mevcut verilerden çıkarımlar yapan ve bu verileri geçmiş deneyimleri kullanarak modeller ve tahminlerde bulunur. Bu bağlamda makine öğrenmesi, insanların geri beslemeli öğrenim şeklinin bilgisayarda taklit edilmiş halidir. [1]

Makine öğrenme metodu Gözetimli ve Gözetimsiz öğrenme olmak üzere iki öğrenme türüne ayrılmaktadır. Gözetimli öğrenme süreci, verilerdeki etiketlenmiş gözlemlerden algoritma gözlemlerinin nasıl etiketlenmesini gerektiğinin öğretilmesidir. Uygulama olarak Sınıflandırma (Classification) ve Regresyon (Regression) problemlerine odaklanmaktadır. Gözetimsiz öğrenme de ise, etiketlenmemiş gözlemlerden bilinmeyen yapı ve ilişkilerin keşfini yapmasını ve gözükmeyen örüntüleri öğrenmesini algoritma sağlamaktadır. Uygulama olarak ise, Kümeleme (Clustering) ve Boyut Azaltımı (Dimensionality Reduction) çalışmalarına odaklanmaktadır.[1,2]

Makine öğrenmesinin bu tanımından hareketle bir çok alanda uygulama çalışmaları yapılmıştır. Bu çalışmalardan ekonomi alanına dair 2018 yılında Kalaycı[3], makine öğrenmesi yöntemleri ile kredi risk analizi konusunu inceleyerek Türkiye'deki KOBİ müşterilerinin ödeme alışkanlıklarına göre ileri bir tarihte müşterinin kredi ödeme durumunun problemli kredi olup olmayacağını tahmin çalışmaları yapılmış ve müşteri risk skoru belirlenmiştir. Bu çalışmada kullanılan veri seti Yapı Kredi Bankası tarafından sağlanmıştır ve kullanılan veriler bankanın KOBİ müşterileri için hazırladıkları kredi paketlerinin, 1 Ocak 2015 - 1 Ekim 2016 tarihleri arasında açılmış olan verileri kullanılarak veri seti hazırlanmıştır. Makine öğrenmesi tekniklerinden, lojistik regresyon (LR), karar ağaçları (KA), destek vektör makineleri (DVM), sinir ağları (SA), rastgele orman algoritması (ROA) ve son olarak kullanılan diğer bir algoritma ise meyilli hızlandırma (MH) yöntemi kullanılmıştır. Yapılan deneysel çalışmalardan en iyi performans başarımı %83,05 başarı ile meyilli hızlandırma(MH) algoritması vermiştir.

Ekonomi çalışmalarına dair yapılan bir diğer güncel çalışma ise 2019 yılında Gülcü ve Çalışkan[4] tarafından, makine öğrenme teknikleri kullanılarak elektrik piyasasında müşteri skorlama için model çalışmaları yapılmıştır. Bu çalışma da kullanılan veri seti, Takasbank platformu kullanılarak günlük elektrik piyasası ile ilgili sözleşmeler yapan müşteri verilerinden oluşturulmuş ve veri seti 1 Ocak 2018 - 31 Aralık 2018 tarihleri arasında Takasbank platformu kullanılarak yapılan sözleşmelerden hazırlanmıştır. Makine öğrenmesi ve veri madenciliği tekniklerinden RFM(Recency,Frequency,Monetar) ve Kümeleme algoritmalarında K-Means, DBSCAN, Agglomerative yöntemleri uygulanmıştır. Çalışmanın sonucunda her bir müşteri için RFM metodu ile müşteri skorları hesaplanılmış ve hesaplanan değerlerle kümeleme teknikleri yapılarak benzer skora sahip müşteri doğru sınıflandırılması yapılmıştır. Başarı ölçütlerini Takasbank kendi skorlama çalışmalarına kıyaslayarak ve kurum işbirliği çerçevesinde iyi bir başarı ile çalışmaları gerçekleştirilmiştir.

Makine öğrenmesinin bir diğer çalışma alanı bulduğu eğitim konusunda ise birden fazla çalışma bulunmaktadır. Eğitim konusu ile ilgili 2016 yılında Livieris ve arkadaşları[5], makine öğrenme tekniklerini kullanarak eğitimde öğrencilerin performansını tahmin etmek için bir karar destek sistemi yapmışlardır. Bu yapılan çalışma ile öğrencilerin bir öğretim yılının final sınavlarına ilişkin performansını tahmin etmek için kullanıcı dostu olan bir karar destek aracı tasarlanmıştır. Kullanılan veri seti, 2007-2010 yılları arasında Yunanistan'da özel bir okuldaki 14-15 yaşlarındaki öğrencilerin sözlü, vize ve final notları gibi öğrenci performansı hakkında bilgi veren verilerinden hazırlanmıştır. Makine öğrenme algoritmalarında Yapay Sinir Ağları(YSA), K En Yakın Komşu(KNN) ve Destek Vektör Makinesi(SVM) metotları kullanılmıştır. Eğitim alanına dair yapılan bir başka çalışma 2018 yılında Karabıyık[6], akademik yayınlar için makine öğrenmesi tabanlı arama motoru tasarlanması ve uygulanması konusunu incelemiştir. Makine öğrenme teknikleri ile akademik yayınların bulunması için kısıtlayıcı bir ağ örümceği tasarlanmıştır. Tasarlanan akademik arama motoru ile bir kullanıcının yaptığı çalışmasının özetinden yola çıkarak, en uygun yayın seçeneklerini sunacak bir çevrim içi sistemin yapılması sağlanmıştır. Kullanılan veri seti, Scimago Journal & Country Rank veri tabınından alınmış, 14 ana konu başlıklı, 450 adet metin özeti kullanılmıştır. Çalışma da, arama sonuçların doğru şekilde değerlendirilmesi için metin sınıflandırma metotları kullanılmıştır. Bu metotlardan Sade Bayes ve Destek Vektör Makinesi(SVM) kullanılmış ve elde edilen sonuçlardan Sade Bayes sınıflandırıcıda %80 başarı, Destek Vektör Makineleri sınıflandırıcısında ise %70 oranında başarı elde edilmiştir.

2015 yılında Kourou ve arkadaşlarının[7] sağlık çalışmaları ile ilgili yaptıkları, kanser prognoz ve tahmininde makine öğrenme teknikleri konusu incelenmiştir. Kanser tipinin erken teşhisi ve hastalık sürecinin seyri hakkında önceden tahmin yapılmasının, klinik süreçlerin yönetimini kolaylaştıracağı için yapılmış olan kanser tahmin çalışmaları incelenmiştir. Yapılan bu inceleme de makine öğrenme tekniklerin kanser tespitinde kullanımına dair PubMed sonuçlarına göre 7510'dan fazla çalışma yayınlanmıştır. Bu yayınların büyük çoğunluğunun makine öğrenme algoritmaları ile tümörlerin saptanması ve kanser tipinin tahmini/prognoz ile ilgili olduğu belirtilmiştir. Yaygın olarak kullanılan makine öğrenmesi tekniklerinden, Yapay Sinir Ağları (YSA), Bayes Ağları (BN), Destek Vektör Makineleri (SVM) ve Karar Ağaçları (DT) dahil olmak üzere bu teknikler tahmin modellerin geliştirilmesi için kanser araştırmalarında kullanılmıştır.

Sağlık alanında yapılan bir diğer çalışma da, 2017 yılında Çakmak[8], makine öğrenmesi yöntemleriyle tümör kontrol olasılığının hesaplanması konusunu inceleyerek onkoloji hastalarının aldığı radyoterapi tedavisinin tümör de ne kadar değişiklik göstereceğine dair bir tahmin modeli oluşturmuştur. Bu uygulamada kullandıkları veri seti, 2012-2015 yılları arasında, Karadeniz Teknik Üniversitesi Tıp Fakültesi Farabi Hastanesi, Radyasyon Onkoloji servisine gelen 30 hastanın verilerinden oluşturulmuştur. Kullanılmış olan makine öğrenme teknikleri, Destek Vektör Makineleri(SVM) ve Yapay Sinir Ağları(YSA) ile tahmin ve sınıflandırma çalışmaları yapılmıştır. SVM ve YSA başarı oranları ise, SVM modeli için %90 duyarlılık ve YSA için %80 duyarlılık değerleri şeklinde elde edilmiştir. İki model karşılaştırmasında SVM modelinin daha başarılı sonuç verdiği görülmüştür. Sağlık alanına dair son olarak inceleyeceğimiz 2018 yılında Saçlı[9] tarafından yapılan, makine öğrenmesi yardımıyla böbrek taşlarının elektromanyetik özelliklerinin sınıflandırılması konusu çalışılmıştır. Veri seti ise, 20 hastadan elde edilen 105 böbrek taşının elektromanyetik özelliklerini içeren veri kümesinden oluşmaktadır. Makine öğrenme tekniklerinden Yapay Sinir Ağları(YSA) ve En Yakın Komuş(KNN) yöntemleri kullanılmıştır. Deney sonuçlarında ise, YSA başarı ölçütünde %97 ve KNN metodu ile de %99 oranında başarı elde edilmiştir.

Makine öğrenmesin genel uygulama alanlarına dair birden fazla akademik çalışma bulunmaktadır. Bu başlık altında ise literatür taramasında örnek uygulama alanlarında olan ekonomi, eğitim ve sağlık sektörlerinde yapılmış olan akademik çalışmalardan sadece birkaçı üzerinden durulmuştur. Çok geniş bir sektörel bazda makine öğrenme teknikleri kullanılmış ve özellikle Türkiye'de kullanılan akademik çalışmalar ile ilgili 2019 yılında Adar ve Delice[10] tarafından, Türkiye'de

makine öğrenmesi ile ilgili yapılan tez çalışmalarına yönelik bir literatür taramasında, Türkiye’de makine öğrenmesi algoritmalarının hangi alanda daha çok uygulandığının tespit edilmesi ve sektör bazında sınıflandırma işlemleri yapılmaya çalışılmıştır. Bu amaç doğrultusunda 146 tez incelenmiş ve en çok sağlık alanına dair çalışmanın bulunduğu belirtilmiştir. Makine öğrenme tekniklerinden ise, En Yakın Komuşu(KNN), Destek Vektör Makineleri(SVM), Karar Ağaçları ve Naive Bayes algoritmalarının sıklıkla kullanıldığı tespit edilmiştir.

2. Makine öğrenme metotları ile finans sektöründe yapılan fiyat tahmin çalışmaları

Bu bölümde yapılan literatür taramasında makine öğrenme tekniklerinin finans sektöründeki kullanım alanı ve literatür taramasında spesifik olarak ‘fiyat’ tahmini yapılmış olan çalışmaların incelenmesi yer almaktadır.

2004 yılında Tektaş ve Karataş[11] makine öğrenme tekniklerinden Yapay Sinir Ağları’nı kullanarak finans alanında Hisse Senetlerinin fiyat tahminlemesi konusunda yaptıkları çalışma da, İstanbul Menkul Kıymetler Borsası (İMKB)’nda işlem yapan şirketlerden yedisinin, hisse fiyatları için tahmin modeli geliştirmeye çalışılmıştır. Model geliştirme için kullanılan veri seti, 2002-2003 yılları arasında İMKB de işlem gören yedi şirketin haftalık ve günlük verilerinden oluşturulmuştur. Yaptıkları bu uygulamada temel amaçlarının Yapay Sinir Ağlarının finans sektöründe kullanımını yaygınlaştırılmasına katkıda bulunmaktadır. Fiyat tahmininde ise günlük verilerden daha başarılı sonuçlar elde etmişlerdir. Finans sektöründe makine öğrenme teknikleri ile yapılmış olan başka bir çalışma ise, 2005 yılında Karaatlı ve arkadaşları[12], Hisse Senedi Fiyat Hareketlerinin Yapay Sinir Ağları Yönetimi ile Tahmin Edilmesi konusunun incelenmesidir. Uygulamadaki amaçlarının yapay sinir ağları metodunu kullanarak borsa endeksi tahmin modelini geliştirmek ve İstanbul Menkul Kıymetler Borsası (İMKB)’nın verilerini kullanarak uygulama geliştirilmesi olduğunu belirtmişlerdir. Kullanılan veri seti, İMKB 100 endeksin 1990 Ocak - 2002 Aralık tarihleri arasındaki aylık verilerinden oluşmaktadır. Yaptıkları deneysel çalışmada yapay sinir ağları ile elde ettikleri başarı oranının yüksek olduğunu ve başarı ölçütlerinin ise hata karelerin karekökü(RMSE) olduğu bildirilmiştir. Ayrıca, model olarak kullandıkları yapay sinir ağlarının RMSE değerinin regresyon yöntemlerine göre daha düşük değer de çıktığını belirtmişlerdir.

İstanbul Menkul Kıymetler Borsası (İMKB) verilerini kullanarak finans sektöründe fiyat tahmin çalışmalarını devam ettiren bir diğer çalışma 2009 yılında Gür[13], Hisse Senedi Fiyat Hareketlerinin Tahmini için Bir Yapay Sinir Ağı Modeli Önerisi adıyla bir yüksek lisans çalışması olarak yayınlanmıştır. İMKB 30'da yer alan şirketlerin hisse senedi fiyat değişimlerinin seans bazlı öngörüsü için model geliştirmeleri yapılmıştır. Kullanılmış olan veri seti, 04 Ocak 1999 - 30 Kasım 2008 yılları arasındaki İMKB 30'da yer alan 9 şirket verisinden oluşmaktadır. Uygulama çalışmalarında, İMKB 30'da yer alan 9 şirket için YSA başarı tahmini 1.seans için %2,84, 2.seans için %3,5 ortalama mutlak hata oranı olarak elde edilmiştir. 2011 yılında Khan ve arkadaşları[14] tarafından, makine öğrenme tekniklerinden Yapay Sinir Ağlarını kullanılarak Bangladeş Borsası'ndaki hisse senedi fiyat tahmin çalışmaları yapılmıştır. Veri seti Borsada işlem gören ACI ilaç firmasının geçmiş verilerinden oluşmuştur. Deneysel kısmında ise, fiyat tahmini için yapay sinir ağları metotlarından ileri beslemeli sinir ağı ile geri yayılım algoritmaları kullanılmıştır. Fiyat tahmin başarı oranı ise 1.simülasyon için %3,71, 2. simülasyon için % 1,53 ortalamasında elde edilmiştir.

Bir başka çalışma ise, makine öğrenme tekniklerinden Destek Vektör Makineleri(SVM) ve metin madenciliği metotlarını kullanarak 2013 yılında Hagenau ve arkadaşları[15] tarafından, finans haberlerinden yola çıkarak hisse senedi fiyat tahmini yapacak otomatik bir haber okuma modeli geliştirilmesi yapılmıştır. Veri seti, Almanya ve İngiltere'den yayınlanan kurumsal finans kuruluşlarının, 1997-2011 yılları arasındaki borsa hisse senedi fiyat bilgisi olan haber kaynaklarından oluşturulmuştur. Duyurular DGAP ve EuroAdhoc haber kaynaklarından elde edilmiştir. Uygulama kısmında metinlerin sınıflandırılması için metin madenciliği teknikleri ve karar destek sistemi (SVM) ile başarı oranı %76'ya varan doğruluk elde edildiğini ifade etmişlerdir. Fiyat tahmin çalışmalarıyla ilgili 2013 yılında Hegazy ve arkadaşları[16] ise, Borsa'da işlem gören hisselerin fiyat tahmini için bir makine öğrenme modeli önerisinde bulunmuşlardır. Kullanılan veri setinde, S&P 500 borsalarındaki tüm hisse senedi sektörlerini kapsayan birçok şirket verisi işlenmiştir. Bu sektörler, bilgi teknolojisi (Adobe, Hp ve Oracle), finans (American Express ve Bank of New York) ve benzeri verilerden oluşmuştur. Uygulama olarak makine öğrenme algoritmalarından, Parçacık Sürü Optimizasyonu(PSO), En Küçük Kareler Destek Vektör Makineleri (LS-SVM) ve Yapay Sinir Ağları (YSA) metotları kullanılmıştır. Elde edilen sonuçlardan önerilen modelin daha iyi tahmin doğruluğuna, PSO algoritmasının LS-SVM'yi optimize ederek ulaştığını belirtmişlerdir.

Makine öğrenme teknikleriyle fiyat tahmin modeli önerilerinden olan bir diğer çalışma 2014 yılında Leung ve arkadaşları[17] tarafından, hisse senedi fiyat tahmininde makine öğrenmesi yaklaşımıyla bir uygulama önerisi olmuştur. Bu uygulamada kullanılan veri seti, S&P 500'ün Bilgi Teknolojisi sektöründeki şirketlerin verilerinden oluşturulmuştur. Deneysel kısmında, Yapısal Destek Vektör Makinesi (SSVM) ile doğru sınıflandırılmış düğümlerin başarı oranının %78'in üzerinde olmasından dolayı, modelin doğru öğrenildiği ifade edilmiştir. 2015 yılına gelindiğinde Çalışkan ve Deniz[18], makine öğrenme tekniklerinden yapay sinir ağları ile hisse senedi fiyatları ve yönlerin tahmini çalışmalarını yapmışlardır. Veri seti olarak, 14 Aralık 2009 ile 21 Kasım 2014 tarihleri arasında BİST 30'da yer alan 27 şirketin bilgisi kullanılmıştır. Uygulama olarak Yapay Sinir Ağlarıyla kurdukları tahmin modeli, 27 şirket için 25 Kasım 2014 ile 01 Aralık 2014 tarihleri arasındaki verilerle yapılmıştır. Fiyat tahmini için ortalama mutlak hata değerinin %1,80 ve yönlerin artacak, azalacak şeklindeki tahmin değerindeki başarı oranı ise %58 olduğunu tespit etmişlerdir.

Makine öğrenme teknikleri ile fiyat tahmin çalışmalarında 2016 yılında Yüksel ve Akkoç[19] tarafından, altın fiyatlarının tahmini için yapay sinir ağları kullanılarak bir uygulama model geliştirilmesi yapılmıştır. Uygulamada veri setini, 03 Ocak 2002 ile 31 Ekim 2013 tarihleri arasında yer alan 2885 veri oluşturmuştur. Veri setini hazırlarken altın fiyat tahmininin daha hassas değerlerle yapılması için değişken olarak altın verisini etkileyecek; Gümüş fiyatları, Brent Petrol fiyatları, ABD doları/ EUR paritesi, EuroNext100 endeksi, Amerika Dow Jones Endeksi, 13 Hafta vadeli ABD bonosu faiz oranı ve ABD TÜFE endeksi değerleri de kullanılmıştır. Deneysel kısımda ise yapay sinir ağları ile elde edilen sonuçları R2, RMSE, MAE ve MAPE (%) kriterleri hesaplanarak başarı değerlendirilmesi yapılmıştır. Ayrıca duyarlılık analizi sonuçlarına göre altın fiyatlarının, gümüş ve petrol fiyatlarını önemli derecede belirlediği tespit edilmiştir. 2016 yılında ise Addai[20], finansal tahminler için makine öğrenme tekniklerini kullanarak hisse senedi/endeks getirilerini tahminleme çalışmaları yapmıştır. Veri seti, Yahoo'dan elde edilmiştir. Veri içerikleri; Hisse senetlerinin açılış fiyatları, düşük fiyatları, yüksek fiyatları, işlem hacmi, kapanış fiyatları ve düzeltilmiş kapanış fiyatları hakkındaki bilgilerden oluşturulmuştur. Elde edilen veriler 1 Ocak 1999 ile 31 Aralık 2008 tarihleri arasını kapsamaktadır. Deneysel kısmında ise, hisse senedi endeksindeki günlük getirilerin hareketini tahmin etmek için makine öğrenme tekniklerinden beş farklı teknik uygulanmıştır. Bu teknikler Yapay Sinir Ağı (YSA), Lojistik Regresyon, Doğrusal Diskriminant Analizi (LDA), Karesel Diskriminant Analizi (QDA) ve K-En Yakın Komşu (KNN) algoritmalarından oluşmaktadır. Uygulama sonuçlarından en iyi performansı YSA'nın gösterdiğini

ve hisse senetlerinin/endekslerinin açılış fiyatlarını kullanarak getiri tahmininde yaklaşık başarı değerinin % 61 olduğu ifade edilmiştir. Borsa hisse senedi fiyat tahminiyle ilgili bir başka çalışma 2016 yılında Özçalıcı[21], yapay sinir ağları ile fiyat tahmin uygulamasın BIST30 senetlerini kullanarak yapmıştır. Veri kümesi, Ocak 2010 ile Kasım 2015 tarihleri arasındaki BIST30'da yer alan hisse senetlerinden oluşturulmuştur. Uygulama olarak daha önce yapılmış olan çalışmalardan farklı olarak 1 gün, 2 gün ve 20 gün sonraki hisse senedi fiyatlarının kapanışının tahmin edildiğini ve hisse senetlerinin fiyat hareketlerini %72.88'e oranda tahmin başarısı gösterdiği belirtilmiştir.

Fiyat tahmin çalışmalarının kripto para piyasasındaki uygulanmasına dair ise 2016 yılında McNally[22], makine öğrenme tekniklerini kullanarak Bitcoin fiyatını tahmin etme modeli önerisinde bulunmuştur. Uygulama amacının, Bitcoin fiyatının USD cinsinden yönünün ne kadar doğru olabileceğinin tahmin edilmesi olduğunu ve fiyat verilerinin Bitcoin fiyat endeksinden elde edildiği belirtilmiştir. Kullanılan veri seti, 19 Ağustos 2013 ile 19 Temmuz 2016 yılları arasında oluşturulmuştur. Deneyisel kısmında, Yinelenen Sinir Ağı (RNN) ve Uzun Kısa Süreli Bellek (LSTM) tekniklerin uygulanmasıyla fiyat tahmin sonuçları elde edilmiştir. Tahmini başarı oranı olarak, LSTM algoritması ile %52'lik en yüksek sınıflandırma doğruluğu ve %8'lik bir ortalama hata karesi (RMSE) elde edilmiştir. Kripto para piyasası fiyat tahminiyle ilgili yapılmış olan bir başka çalışma 2018 yılında Sakız ve Gencer[23] tarafından, makine öğrenme tekniklerinden yapay sinir ağlarını kullanarak Bitcoin fiyatını tahminleme konusunda bir sunum şeklinde gerçekleştirilmiştir. Kullanılan veri setini, Ocak 2015 ile Nisan 2018 tarihleri arasındaki Bitcoin fiyatlarının günlük kapanış fiyatları alınarak oluşturulmuştur. Uygulama kısmında yaptıkları tahmin sonucu, 2018 Mayıs ayındaki Bitcoin fiyatının 80,955 USD olarak bulunmasıdır. Bulunan sonucun tahmin doğruluğunun performans ölçümü ortalama hata karesi(RMSE) değerine göre bulduklarını belirtmişlerdir. Ama gerçekleşen Mayıs 2018 ortalama fiyatı ise 7,487 USD olmuştur. Tahmin edilen değeri ile gerçekleşen değer arasındaki büyük farklılığın sebebinin Bitcoinin fiyat değişkenin çok olması ve kullandıkları verinin azlığından kaynaklandığını belirtmişlerdir. 2018 yılında kripto para fiyat tahminiyle ilgili yapılmış olan bir diğer çalışma da ise Aktepe[24], makine öğrenmesi tekniklerinden yararlanarak kripto para piyasasında fiyat tahminleme ve kar getirebilecek algoritmalar üzerinde çalışmıştır. Veri kümesi, 1 Eylül 2017 ile 1 Mayıs 2018 tarihleri arasında Binance Coin (BNB)'den aldığı kripto paraların, Açılış-Yüksek-Düşük Kapanış fiyat verilerinden oluşmaktadır. Uygulama kısmında, makine öğrenme tekniklerinden hem sınıflandırma hem de regresyon kullanılmıştır. Daha sonra geliştirilen her modelin fiyat tahmin hedeflenmesinde farklı eşik ayarları kullanılarak özellik eşlemesini kullanan bir topluluk öğrenmesi önerilmiştir.

Devamında seçili olan kripto para birimlerini birlikte göz önünde bulundurarak, modellerin birbirleriyle ve satın alma / tutma stratejisinin karşılaştırılması ile basit portföy alım satım durumu için başarılı model geliştirilmesi yapılmıştır.

Borsa’da işlem gören hisse senedi fiyat verileri üzerine 2018 yılında Kanmaz[25], Borsa İstanbul’da farklı sektörlerden oluşan bir çok şirkete ait fiyat verileriyle bu şirketlerin ekonomi haberlerinde kullanılması arasındaki bağlantıyı inceleme konusu seçmiştir. Çalışmasında, makine öğrenme teknikleri, doğal dil işleme ve metin madenciliği yöntemlerini kullanarak ekonomi haberlerinin hisse senetleri üzerindeki etkisini incelemiştir. Deneysel kısımda geliştirilen model başarı tahmin oranı, ekonomi haberlerindeki sınıflandırılmışlığına göre %70 oranında bir doğruluk payı vermiştir. Ayrıca, yapılan çalışmanın sonucunda ekonomi haberlerinin olumlu/olumsuz etkisinin ilgili hisse senedi fiyatları üzerinde de olumlu/ olumsuz bir etki yarattığının tespit edildiği belirtilmiştir.

2019 yılında Demirel [26], makine öğrenme teknikleri ve derin öğrenme yöntemleri ile hisse senetlerinin açılış ve kapanış fiyatlarının tahmini konusunu incelemiştir. Veri seti, BIST 100’de işlem gören 42 adet şirketin 1 Ocak 2010 ile 1 Ocak 2019 tarihleri arasındaki verilerinden oluşturulmuştur. Uygulama bölümünde, makine öğrenme tekniklerinden Çok Katmanlı Algılayıcılar (ÇKA), Destek Vektör Makineleri (DVM) ve Uzun Kısa Dönemli Hafıza (UKVH) gibi derin öğrenme metotları kullanılmıştır. Çalışmanın sonucu olarak, ÇKA ve UKVH metotların da DVM’e yöntemine göre daha tutarlı tahminler tespit edilmiştir. Ayrıca, ÇKA ve UKVH metotlarından elde edilen tahmin başarı oranı % 95 güven aralığında gerçekleşmiş, gerçek oranlar ile öngörülen oranlar arasında anlamlı bir farkın olmadığı tespit edilmiştir. Bir başka çalışma ise 2019 yılında Pabuçcu[27] tarafından, makine öğrenme algoritmaları ile borsa endeks hareketinin yönüne dair tahmin çalışmaları yapılmıştır. Veri kümesi, BİST 100 endeksin 2009 ile 2018 tarihleri arasındaki endeks günlük kapanış fiyatlarından oluşturulmuştur. Uygulama bölümünde makine öğrenme tekniklerinden yapay sinir ağları (YSA), destek vektör makineleri (SVM) ve naive Bayes metotları kullanılmıştır. Sonuç olarak ise, her üç modelinde borsa endeks hareketinin pozitif/negatif yön tahmininde başarılı olduğunu ve yapay sinir ağı (YSA) algoritmasından diğer iki modelden daha performanslı sonuçlar elde edildiği ifade edilmiştir. Bu bölümde inceleyeceğimiz son bir diğer çalışma ise 2019 yılında Akşehir ve Kılıç[28]’a ait, makine öğrenme tekniklerini kullanarak banka senetlerinin fiyat tahmin çalışmalarının yapılmasıdır. Uygulama alanında, banka hisse senetlerinde bir sonraki kapanış fiyatlarının tahmininin yapılmasında bir çok değişkenin var olduğunu ve bu durumun zor bir problem teşkil ettiği ifade edilmiştir. Veri seti, 1 Ocak 2016 ile 9 Mayıs 2019

tarihleri arasındaki Borsa İstanbul BİST 100 endeksinde yer alan 5 büyük bankanın, hisse senetlerinin açılış, kapanış, yüksek/düşük işlem hacim bilgilerinden oluşturulmuştur. Deneysel kısmında, karar ağaçları (DT), çoklu regresyon (MLR) ve rassal olmayan orman (RF) yöntemleri kullanılmıştır. Model başarı ölçütünü, R^2 (R Square) yöntemi ile yaptıklarını ve fiyat tahmin için kullanılan makine öğrenme algoritmalarından başarılı sonuçlar elde edildiği belirtilmiştir.

Bu bölümde makine öğrenme teknikleri ile finans sektöründe yapılmış olan fiyat tahmin çalışmaları genel olarak incelendiğinde, sıklıkla borsa hisse senetlerinde fiyat tahmini yapılması ciddi bir çalışma sahası oluşturmuştur. Makine öğrenme algoritmalarından en çok kullanılanların ise, Yapay Sinir Ağları, Karar Ağaçları ve Çoklu Regresyon yöntemleri olduğu görülmüştür.

3. Metot bölümünde kullanılan algoritmalar ile yapılan çalışmalar

Bu bölümde yapılacak olan literatür taramasında, uygulama bölümünde fiyat tahmin model geliştirmesi için kullanılacak olan PLS, Ridge, SVR ve YSA algoritmaları üzerine yapılan çalışmalar incelenecektir.

3.1 Kısmi En Küçük Kareler(PLS) Regresyon algoritması

PLS algoritmasını kullanarak 2007 yılında Janik ve arkadaşları[29] tarafında, toprakta mevcut olan organik karbon fraksiyonlarının konsantrasyonunu tahmin etmek için model geliştirmeleri yapılmış.

Devam edilecek model geliştirilmesi bittikten sonra...:)

Kaynakça

- 1.Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi(Social Sciences Research Journal, Volume 7, Issue 2, 41-53 (June 2018), ISSN: 2147-5237)
2. Alpaydın, E. (2004). Introduction to machine learning (Adaptive Computation and Machine Learning). Cambridge: The MIT press, 11-26.
- 3.Kalaycı, S. (2018). Makine öğrenmesi yöntemleri ile kredi risk analizi, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- 4.Clustering Electricity Market Participants, Turk J Elec Eng & Comp Sci, Tübitak.(Yayın aşamasında)
- 5.A decision support system for predicting students performance, Themes in Science & Technology Education, 9(1), 43-57, 2016.
- 6.Karabıyık, M.A. (2018). Akademik yayınlar için makine öğrenmesi tabanlı arama motoru tasarlanması ve uygulanması, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.
- 7.Machine learning applications in cancer prognosis and prediction, Computational and Structural Biotechnology Journal 13(2015) 8-17.
- 8.Çakmak, I.(2017). Makine öğrenmesi yöntemleriyle tümör kontrol olasılığının hesaplanması, Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi, Sağlık Bilimler Enstitüsü, Trabzon.
9. Saçlı, B.(2018). Machine Learning Aided Kidney Stone Classification With Electromagnetic Properties, Istanbul Technical University, Department of Electronics and Communication Engineering, Istanbul.

10. Türkiye’de makine öğrenmesi ile ilgili yapılan tez çalışmalarına yönelik bir literatür taraması, UEMK 2019 Proceedings Book 24/25 October 2019 Gaziantep University, Gaziantep.
11. Tektaş, A., Karataş, A., Yapay Sinir Ağları ve Finans Alanına Uygulanması: Hisse Senedi Fiyat Tahminlemesi, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi. 2004.
12. Karaatlı, M., Güngör, İ., Demir, Y., Kalaycı, Ş. Hisse Senedi Fiyat Hareketlerinin Yapay Sinir Ağları Yönetimi ile Tahmin Edilmesi. Yönetim ve Ekonomi Araştırmaları Dergisi. 2005; 3(3): 48-38.
13. Gür, N. (2009). Hisse Senedi Fiyat Hareketlerinin Tahmini için Bir Yapay Sinir Ağı Modeli Önerisi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
14. Price Prediction of Share Market using Artificial Neural Network (ANN), International Journal of Computer Applications (0975 – 8887) Volume 22– No.2, May 2011.
15. Michael Hagenau Michael Liebmann Dirk Neumann, *Automated news reading: Stock price prediction based on financial news using context-capturing features*, Decision Support Systems Volume 55, Issue 3, June 2013, Pages 685-697
16. Osman Hegazy , Omar S. Soliman , Mustafa Abdul Salam., *A Machine Learning Model for Stock Market Prediction*, International Journal of Computer Science and Telecommunications [Volume 4, Issue 12, December 2013].
17. Carson Kai-Sang Leung, Richard Kyle MacKinnon , Yang Wang, *A Machine Learning Approach for Stock Price Prediction*, IDEAS '14: Proceedings of the 18th International Database Engineering & Applications Symposium, July 2014.
18. Yapay Sinir Ağlarıyla Hisse Senedi Fiyatları ve Yönlerinin Tahmini, Eskişehir Osmangazi Üniversitesi, İİBF Dergisi, Aralık 2015, 10(3), 177- 194.

- 19.Yüksel, R., Akkoç, S., Altın Fiyatlarının Yapay Sinir Ağları ile Tahmini ve Bir Uygulama, Doğu Üniversitesi Dergisi, 17 (1) 2016, 39-50
20. Addai, S. Financial Forecasting Using Machine Learning, Masters Degree, African Institute for Mathematical Science, South Africa, May 2016.
- 21.Özçalıcı, M. Yapay Sinir Ağları ile Çok Aşamalı Fiyat Tahmini: BIST30 Senetleri Üzerine Bir Araştırma, Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt:31, Sayı:2, Yıl:2016, ss. 209-227.
22. McNally, S., Predicting the price of Bitcoin using Machine Learning, School of Computing National College of Ireland, MSc Research Project in Data Analytics, September 2016.
- 23.Sakız, B., Gencer, A.H, Forecasting the Bitcoin Price via Artificial Neural Networks, International Conference of Eurasian Economies 2018, pp.438-444, Tashkent, UZBEKISTAN.
- 24.Aktepe, Ç. (2018). Algorithmic trading on cryptocurrency markets using machine learning techniques, M.Sc. Thesis, Boğaziçi University, Department of Industrial Engineering, İstanbul.
- 25.Kanmaz, M. (2018). The effect of financial news on bist stock prices: A machine learning approach, M.Sc. Thesis, Middle East Technical University, Department of Economics, Ankara.
- 26.Demirel, U. (2019). Hisse senedi fiyatlarının makine öğrenmesi yöntemleri ve derin öğrenme algoritmaları ile tahmini, Yüksek Lisans Tezi, Gümüşhane Üniversitesi, Sosyal Bilimler Enstitüsü, Gümüşhane.
- 27.Pabuçcu,H., Borsa Endeksi Hareketlerinin Makine Öğrenme Algoritmaları ile Tahmini. Uluslararası İktisadi ve İdari İncelemeler Dergisi. 2019; (23): 190-179.
28. Akşehir, D.Z, Kılıç, E., Makine Öğrenme Teknikleri ile Banka Hisse Senetlerinin Fiyat Tahmini. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi. 2019; 12(2): 39-30.