

T.C.
İSTANBUL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
YÜKSEK LİSANS TEZİ

RİSKE MARUZ NAKİT AKIŞI KAVRAMINA YENİ
YAKLAŞIMLAR VE TÜRKİYE UYGULAMASI

Hüseyin KESİM

2501200409

TEZDANIŞMANI

Doç. Dr. Yusuf AYTÜRK

İSTANBUL – 2023

ÖZ

RİSKE MARUZ NAKİT AKIŞI KAVRAMINA YENİ YAKLAŞIMLAR VE TÜRKİYE

UYGULAMASI

Hüseyin KESİM

Finansal Risk Yönetiminde risk göstergeleri önemli bir yer tutar. 1990'lı yıllarda yaygın olarak kullanılan Riske Maruz Değer (RMD) bu göstergelerden bir tanesidir. Ancak pratik ihtiyaçlar gereği RMD göstergesine alternatif olarak 2000'li yıllarda Riske Maruz Nakit Akışı (RMNA) kavramı ortaya çıkmış ve firmaların nakit akış risklerini ölçümlemek amacıyla kullanılmıştır. RMNA göstergesi işletme yöneticisine bir sonraki dönemde gerçekleşebilecek nakit değişimini bir en kötü durum senaryosu olarak verir ve bir öngörü sağlar. Bu çalışmada, BİST-100'de faaliyet gösteren 40 işletme üzerinden RMNA modelleri oluşturulmuş ve hem bir sonraki dönemdeki nakit akışı hem de en kötü durum senaryosu olarak RMNA sınır değerleri tahmin edilmiştir. Tahmin modelleri, zaman serisi olarak düzenlenmiş ve RMNA modellerinde daha yaygın kullanılan Doğrusal Regresyonun yanı sıra Makine Öğrenmesi algoritmaları da kullanılmıştır. Nakit akış tahminlerinin doğruluğunu tespit edebilmek için Kök Ortalama Kare Hatası ve Ortalama Mutlak Hata ölçütleri kullanılmış; RMNA bulgularını test edebilmek için de RMD uygulamalarında kullanılan Kupiec testi ve Süre testi uygulanmıştır. Örneklemimizden elde ettiğimiz bulgulara göre, karmaşık Makine Öğrenmesi temelli modellerin hem tahmin performansı açısından hem de RMNA tahmini açısından daha düşük bir performans sergilemekte olduğu gözlemlenmiş ve gözlemlerle ilgili sebepler ve sonuçlar çalışmamızda tartışılmıştır.

Anahtar Kelimeler: Finansal Risk Göstergeleri, Riske Maruz Nakit Akışı (RMNA), Riske Maruz Değer (VaR), Makine Öğrenmesi, Kupiec Testi, Süre Testi.

ABSTRACT

NEW APPROACHES TO THE CONCEPT OF CASH FLOWS AT RISK AND ITS IMPLEMENTATION IN TURKEY

Hüseyin KESİM

Risk indicators have an important place in Financial Risk Management. Value at Risk (VaR), which was widely used in the 1990s, is one of these indicators. However, due to practical needs, the concept of Cash Flow at Risk (CFaR) emerged in the 2000s as an alternative to the VaR indicator and was used to measure the cash flow risks of companies. The CFaR indicator provides the business manager with a worst-case scenario for the cash change that may occur in the next period and a forecast. In this study, CFaR models were created over 40 businesses operating in BIST-100 and CFaR limit values were estimated as both the cash flow in the next period and as the worst-case scenario. Prediction models are arranged as time series models and Machine Learning algorithms are used as well as Linear Regression, which was widely used in CFaR models. Root Mean Squared Error and Mean Absolute Error criteria were used to determine the accuracy of cash flow forecasts. In order to test the CFaR findings, the Kupiec test and the Duration test, which are used in VaR applications, were applied. According to the findings we obtained from our sample, it was observed that complex Machine Learning-based models showed a lower performance in terms of both cash flow prediction and CFaR prediction, and the reasons and results related to this observation were discussed in our study.

Anahtar Kelimeler: Financial Risk Indicators, Cash Flow at Risk (CFaR), Value at Risk (VaR), Machine Learning, Kupiec Test, Duration Test.

ÖNSÖZ

RMNA göstergesi işletme yöneticisine bir sonraki dönemde gerçekleşebilecek nakit değişimini bir en kötü durum senaryosu olarak verir ve bir öngörü sağlar. Bu öngörü sayesinde yönetici, risk yönetimini daha etkin kılarken aynı zamanda kurum içerisinde açık ve şeffaf bir tartışma ortamı oluşturulmuş olur. Dolayısıyla risk göstergelerini sadece risk ölçümleme işleviyle değil kurumsal risk yönetiminin bir parçası olarak düşünmek gerekir. Açık ve anlaşılır bir tahmin modeli kurabilmek bu yüzden önemlidir. Böylece işletmedeki yöneticilerin takip edebildiği, yenilenmeye müsait, esnek ve işletmenin yapısına uygun bir metot geliştirilebilir. Çalışmamızda, BİST-100’de faaliyet gösteren 40 farklı işletme için 10 farklı tahmin modeli oluşturduk ve modellerin tahmin performanslarını test ettik. Modellerde, özellikle RMNA çalışmalarında yaygın olarak kullanılmadığını düşündüğümüz ve görece daha yeni bir alan olarak düşündüğümüz Makine Öğrenmesi algoritmalarını kullandık. Tüm modelleri zaman serisi olarak düzenledik ve tüm işletmeler için aynı parametreleri kullandık. Sonuç olarak, farklı tahmin modellerini karşılaştırdık ve performans farklılıklarının sebepleri ve sonuçları üzerine tartıştık. Sonuç kısmında da görüleceği üzere, ortaya geliştirilmeye açık bir çalışma ortaya koyduk ve bununla ilgili tavsiyeleri de son bölümde aktardık. Bu açıdan bakıldığında, çalışmamız bir sonuç değil; daha sonraki yapılabilecek potansiyel çalışmaların bir başlangıcı olarak düşünülmelidir.

Çalışmamın hazırlanmasında bana hem destek olan hem de çalışmama ilham veren danışman hocam Doç. Dr. Yusuf AYTÜRK’e, İstanbul Üniversitesi yüksek lisans öğretimimde eğitim aldığım hocalarıma, çalışma sürecimde beni destekleyen eşime ve aileme çok teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZ.....	iii
ABSTRACT.....	İv
ÖNSÖZ.....	v
İÇİNDEKİLER.....	vi
ŞEKİLLER LİSTESİ.....	viii
TABLolar LİSTESİ.....	ix
KISALTMALAR.....	x
GİRİŞ	1
LİTERATÜR TARAMASI.....	10
PROBLEM İFADESİ.....	14
ARAŞTIRMA SORUSU.....	15

BİRİNCİ BÖLÜM MAKİNE ÖĞRENMESİ

1.1. Makine Öğrenmesi Literatürü.....	16
1.2. Makine Öğrenmesi Çeşitleri.....	17
1.3. Makine Öğrenmesi Uygulamalarında Zorluklar.....	19
1.4. Makine Öğrenmesi Modellerinde Kriterler.....	20
1.4.1. Doğru Modelin Seçilmesi.....	20
1.4.2. Doğru Parametrelerin Bulunması.....	21
1.4.3. Aşırı Öğrenme Problemi.....	21
1.4.4. Oto Korelasyon Problemi.....	22

İKİNCİ BÖLÜM MAKİNE ÖĞRENMESİ ALGORİTMALARI

2.1. Doğrusal Regresyon.....	24
2.2. En Yakın Komşuluk.....	25
2.3. Karar Ağaçları.....	26
2.4. Bagging Metodu ve Rastgele Ormanlar.....	27
2.5. Boosting Temelli Modeller.....	30
2.6. Adaboost.....	30
2.7. Gradient Boosting.....	32
2.8. XGBoost.....	33
2.9. Diğer Toplu Öğrenme Modelleri.....	35

ÜÇÜNCÜ BÖLÜM

VERİ SETİNİN İNCELENMESİ VE TAHMİN MODELİ

3.1. Veri Setinin İncelenmesi.....	37
3.2.Tahmin Modeli.....	42

DÖRDÜNCÜ BÖLÜM

UYGULAMA

4.1. Tahmin Sonuçları ve Değerlendirme.....	46
4.2. RMNA Sonuçları ve Değerlendirme.....	49
4.3. Kupiec Testi.....	50
4.4. Süre Testi.....	52
4.5.Test Sonuçları Özeti ve Değerlendirme.....	54
4.54.6.Uygulama Sonuçları ve Değerlendirme.....	55
SONUÇ	56
KAYNAKÇA	61

ŞEKİLLER LİSTESİ

Şekil 1: RMD için Dağılım ve RMD sınır değeri.....	2
Şekil 2: RMNA Dağılımı.....	8
Şekil 3: Makine Öğrenmesi / Derin Öğrenme / Yapay Zekâ.....	16
Şekil 4: Denetimsiz Öğrenme Örneği olarak K-Means Algoritmasının Etkisi.....	18
Şekil 5: Aşırı Öğrenme Durumlarında Varyans ve Yanlılık Dengesi.....	22
Şekil 6: Basit Doğrusal Regresyon.....	24
Şekil 7: En Yakın Komşuluk Değeri 3 için Algoritmanın Tahmini.....	25
Şekil 8: Karar Ağacı Örneği.....	26
Şekil 9: Bootstrapping ile Veri Seçimi.....	28
Şekil 10: Rastgele Ormanlar Algoritması için Ortalama Değerin Alınması.....	29
Şekil 11: TÜFE Oranı (2014-2023)	39
Şekil 12: USD /TL Kuru (2018-2023)	40
Şekil 13 İşletmelerin Zamansal Nakit Değişimleri.....	40
Şekil 14: RMNA Dağılımı.....	44
Şekil 15: İki alogirtmanın Dönemsel KOKH Skorları.....	48
Şekil 16: Modeller ve İhlal Sayıları.....	50
Şekil 17: Kupiec Testi ve Süre Testi Özeti.....	55
Şekil 18: Model Karmaşıklığı ve Hata Oranı İlişkisi.....	57
Şekil 19: Varyansı Yüksek Regresyon Regresyon ve Yanlılık Değeri Yüksek Regresyon.....	57

Şekil 20: Gradient Boosting ve Doğrusal Regresyon (Tahmin ve Gerçekleşen)58



TABLÖLÄR LİSTESİ

Tablo 1: RMD ve RMNA.....	6
Tablo 2: Firma Kısaltmaları ve Regresyon Analizinde Kullanılan Dönemler.....	37
Tablo 3: İşletmeler için Özet İstatistik Değerleri (1.000 USD)	38
Tablo 4: Nakit Dengesi USD / TL Kuru--TÜFE Korelasyonu.....	39
Tablo 5: Kullanılan Bozulma Parametreler ve ilgili Süre Değerleri.....	44
Tablo 6: Regresyon Modelleri ve Kısaltmaları.....	46
Tablo 7: Kök Ortalama Kare Hatası (KOKH) ve Ortalama Mutlak Hata (OMH) Değerleri (Milyon)...	47
Tablo 8: 2022 4.Çeyrek için Kök Ortalama Kare Hatası Değerleri (Milyon).....	48
Tablo 9: İşletme bazında Model İhlal Sayıları.....	49
Tablo 10: Model İhlal Oranları.....	50
Tablo 11: Kupiec Testi Sonuçları.....	51
Tablo 12: Model bazında Kupiec Testi Başarı Sayısı.....	52
Tablo 13: Süre Testi Sonuçları.....	53
Tablo 14: Model bazında Süre Testi Başarı Sayısı.....	54
Tablo 15: Kupiec Testi ve Süre Testi Özeti.....	54

KISALTMALAR LİSTESİ

CFaR	: Cash flow at Risk
ES	: Expected Shortfall
FAVÖK	: Faiz. Amortisman ve Vergi Öncesi Kar
KOKH	: Kök Ortalama Kare Hatası
ML	: Makine Öğrenmesi
OMH	: Ortalama Mutlak Hata
RMD	: Riske Maruz Değer
RMNA	: Riske Maruz Nakit Akışı
TL	: Türk Lirası
USD	: ABD Doları
VaR	: Value at Risk

GİRİŞ

Finansal piyasaların özellikle 1970'li yıllardan sonra gelişmesi ve yaygınlaşmasıyla birlikte firmaların riske duyarlılıkları da aynı ölçüde artış göstermiştir. Buna paralel olarak Risk Yönetimi ayrı bir uzmanlık alanı olarak gelişmektedir. Risk Yönetiminin bir parçası olarak risk göstergelerinin işlevi, risklerin öngörülmesi ve bu risklere karşı kurum ya da firma içerisinde öngörülebilir, şeffaf, pratik ve tutarlı yaklaşımların geliştirilmesini sağlayabilmektir. Risk göstergeleri birçok risk unsurunun tek bir göstergede ifade edilebiliyor olması bakımından pratik bir kullanıma sahiptir. Bunun yanı sıra, matematiksel ve tutarlı bir yönteme dayanıyor olması sayesinde hem öngörülebilirliği arttırırlar hem de açık ve yalın bir bakış açısı geliştirilmesine katkıda bulunurlar.

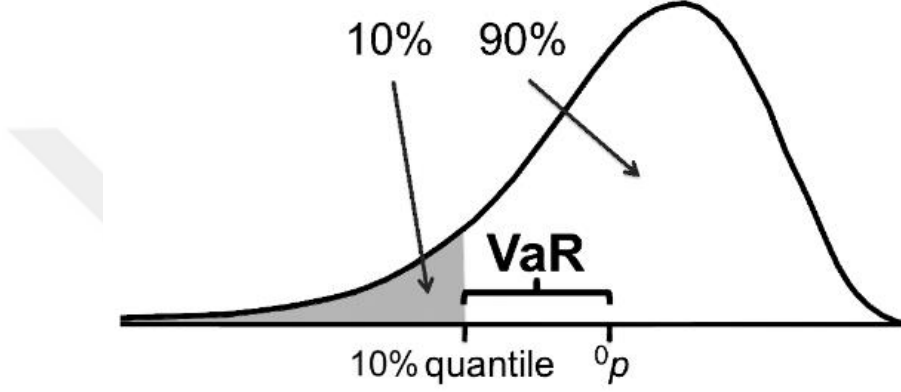
Riske Maruz Nakit Akışı (Cash flow at Risk/ RMNA) firmaların nakit akış risklerini ölçümleyebilmek için yaratılmış bir finansal göstergedir. İlk kez 1990'lı yılların sonlarında JP Morgan içerisinde yer alan RiskMetrics Grubu tarafından bir başka finansal risk göstergesi olan VaR'a (Value at Risk / Riske Maruz Değer) alternatif olarak geliştirilmiştir. Özellikle son 20 yıl içerisinde farklı firmalara ve farklı sektörlerle çeşitli metotlarla uygulanmış ve zamanla yaygın hale gelmiştir. RMD göstergesine alternatif olarak geliştirilmiş olması Riske Maruz Nakit Akışı kavramını tartışırken RMD göstergesine atıfta bulunmayı gerektirmektedir. Bu bakımdan öncelikle bir risk göstergesi olarak RMD'nin temel işlevinin ve uygulamalarının anlaşılması gerekmektedir.

RMD göstergesi 1990'lı yılların başlarında kullanılmaya başlanmış gerek finansal riskleri ölçmedeki başarısı gerek pratik bir kullanımı olması nedeniyle kullanımı yaygınlaşmıştır. RMD finansal enstrümanların dalgalanmalarını varyans / kovaryans hesaplamalarıyla ölçmeye dayanan bir istatistiksel göstergedir ve portföyde tutulan varlıkların / yatırım araçlarının olası değişimini öngörebilmek için bir güven aralığı ortaya koymaktadır. RMD hesaplanırken farklı metotlar uygulansa da temel olarak üç aşamadan söz edebiliriz:

1. Bir portföydeki olası risk değişimleri listelenip dağılımı belirlenir.
2. Güven aralığı (%90, %95 ya da %99 olarak) belirlenir.
3. Bu dağılıma ve güven aralığına göre değişimin belli bir ihtimal dahilinde belli bir aralıkta kalacağı varsayılmaktadır.

Aşağıdaki örneğe bakarak bu durumu daha açık bir şekilde somutlaştırabiliriz. Aşağıdaki dağılıma göre, portföydeki değişimin %90 ihtimalle gri bölgedeki seviyeye düşmeyeceği ifade edilmektedir:

Şekil 1: RMD için Dağılım ve RMD sınır değeri



Kaynak: https://www.glynholton.com/blog/risk-measurement/var_measure/

RMD tahmini yukarıdaki örnekte de görüldüğü üzere portföy değişim tahminlerinin dağılımına göre hesaplanır. Bu dağılıma ulaşmada farklı metotlar kullanılabilir. Bir risk tahmini yapıldığı için en kötü duruma bakılır ve sadece azalış kısmındaki sınır durum incelenir. Örneğin; %95 güven aralığı belirlediyseniz, dağılımın azalış kısmındaki %95’lik sınıra bakılır. Sonrasında %95 ihtimalle nakit akış değişiminin bu sınırın altına inmeyeceği öngörülür. RMD hesaplamasında, yaygın olarak üç farklı yöntemden bahsedebiliriz: Tarihsel veri metodu, Kovaryans matrisi metodu ve Monte Carlo Simülasyonu.¹

Tarihsel veri metodu, verilerin toplanıp bir dağılım oluşturulması yöntemine dayanır. Geçmişte gerçekleşen değerlere dayandığı için belirli bir normal dağılım varsayımı yapılmaz. Tarihsel değişimlerin gelecekte tekrar edeceği varsayımıyla hareket edilir. Ayrıca, tarihsel verilere herhangi bir ağırlık verilmez. Örneğin; günümüze yakın verilerle uzak verilerin etkilerinin aynı ağırlıkta olduğu varsayımı geçerlidir. Dolayısıyla, öncelikle bir portföydeki değişim miktarının geçmiş

¹ Damodaran, Aswath, **Value at Risk**, <https://pages.stern.nyu.edu/~adamodar/pdfiles/papers/VAR.pdf>, s.4

verilere dayanarak ortaya konulması önemlidir. İkinci aşamada, buradan bir sıklık dağılımına ulaşılır. Son aşamada ise belirlenen bir yüzde miktarına göre bir güven aralığı oluşturulur.

Kovaryans matrisi yönteminde ise normal dağılım varsayımıyla hareket edilir. Öncelikle portföydeki enstrümanların değişimlerine göre kovaryans matrisi oluşturulur. Bulunan matris portföydeki enstrümanların ağırlık vektörüyle çarpılır. Elde edilen sonuç bir sayı değeri olarak standart sapma olacaktır. Hesaplamaya geçilmeden önce iki önemli varsayım ortaya konulur: Modele göre ortalama portföy getirisinin 0 olacağı varsayılır. Ayrıca, portföydeki olası değişimlerin normal bir dağılıma sahip olduğu varsayılır. Bu varsayımlardan hareket ederek standart sapma değerine (kovaryans ve ağırlıkların çarpımından oluşan sapma değerine) ve 0 değerli bir ortalamaya göre bir normal dağılım (RMD dağılımı) oluşturulur.

Monte Carlo Simülasyonunda ise tarihsel veri yöntemi ya da kovaryans matrisi yöntemlerinden birisi kullanılabilir. Buradaki ortalama ve standart sapma değeri kullanılarak rastgele örneklemeler yaratılır. Daha sonra, simülasyona tabi tutulmuş bu örneklemelerin ortalamalarının dağılımı üzerinden bir RMD dağılımı oluşturulabilir. Monte Carlo Simülasyonunda alternatif olarak bootstrap yöntemi de kullanılabilir.² Bu yöntem ile hayali bir popülasyon var olan örneklem üzerinden (yerine koyulup tekrar çekilme yöntemiyle) türetilir. Özetle, Monte Carlo simülasyonu metodundaki ana fikir, elimizdeki dağılıma göre yeni rastgele değerler yaratabilmek ve buradan bir RMD dağılımı türetmektir. Simülasyona dahil olan örneklemeler aracılığıyla bir ortalamaya ve bir standart sapma değerine ulaşılır ve olası değişimlerin dağılımı oluşturulur.

Sonuç olarak, RMD hesaplaması değişimler için bir güven aralığı oluşturmaya dayanır. Bu dağılım ise sonraki dönemler için yöneticilere pratik bir öngörü verir. Olası bir portföy riski karşısında yönetici kişi “en kötü durum” senaryosu olarak değişim miktarını tahmin edebilir. Bir başka deyişle, örneğin, “%95 ihtimalle portföy değer kaybı x miktardan fazla olmayacak” şeklinde bir çıkarım elde edilir. RMD bu anlamda oldukça pratik bir kullanıma sahiptir; ancak bazı dezavantajları da mevcuttur.

Birincisi, RMD göstergesi genellikle finans temelli firmalara uygulanabilmekte ve portföy risklerini ortaya koyabilmektedir. Ancak Finans sektörü dışındaki firmaların riskleri ise

² Fook Stephanie, Robin Chong, , “Introduction to Bootstrap”, **Proceedings of Singapore Healthcare**, Volume:20, Number 3 , Singapore, 2011, s.236.

çoğunlukla birden çok deęişkene baęlıdır.³ Bu bakımdan, RMD reel sektörde gerekli performansı göstermeyebilir. Buna ek olarak, RMD hesaplaması sadece sınır deęerine odaklanır ve bu deęerin ne olduęunu belirtir. Ancak bu uç durumun ne ölçüde riskli olduęu üzerinde durulmaz. Ayrıca uç durum gerçekleşirse kayıp riskinin ortalama ne kadar olacaęı yönünde de bir tahmini içermemektedir. İlk dezavantajla ilgili olarak finans sektörü dışındaki firmalar için farklı göstergelere ihtiyaç duyulmaktadır. RMNA göstergesi bunlardan bir tanesidir. Böylece firmalar nakit akış risklerini tahmin etme konusunda öngörü edinebilirler. İkinci dezavantajla ilgili olarak da bir farklı göstergeye bakmamız faydalı olabilir: Beklenen Kayıp (Expected Shortfall) Yöntemi.

Beklenen Kayıp (ES) yöntemi Artzner tarafından 1997 yılında ortaya konulmuş ve RMD'den kaynaklanan zayıflıkları telafi etmek için kullanılmıştır. “Koşullu VaR”, “Kuyruklu VaR” veya “Ortalamayı aşan kayıp” şeklinde de ifade edilmektedir.⁴ RMD göstergesi hesaplanırken, oluşacak kaybın sınır deęeri ortaya konulur; ancak güven aralığı deęerinin ötesindeki uç ihtimal deęerlendirmeye alınmaz. Örneęin, “%95 ihtimalle 100.000 TL'den fazla kayıp yaşanmayacak” şeklinde bir RMD hesaplamamız olsun. Bu kısım RMD ile hesaplanır. ES göstergesi ile “100.000 TL ve daha fazla kaybımızın olması durumunda (sınır durumunun geçilmesi durumunda) ortalama kaybımız ne olur?” sorusu cevaplanmaya çalışılır. Bu hesaplama yapılırken kullanılan yaygın bir yöntem olarak, uç duruma geçildiğinde oluşma ihtimali olan deęerlerin ortalaması alınır.⁵ Bu deęerlerin tamamı 100.000 TL altında oldukları için, ortalamaya göre oluşacak kayıp 100.000 TL'nin üzerinde hesaplanacaktır. Böylece ES göstergesinin hesaplamasında, RMD göstergesine kıyasla daha temkinli bir yaklaşım ortaya konulmuş olur.

RMNA göstergesinin kullanımlarına bakarsak esas olarak RMD deęerinden türetildięini görebiliriz. Örneęin; akademik literatürdeki makalelere baktığımızda göstergenin sadece uç deęerine bakıldığını ancak ES hesaplamasında olduęu gibi uç durumda ne olacaęına dair koşullu bir hesaplamanın yapılmadığını söyleyebiliriz. Finans sektörü dışındaki firmalara baktığımızda bir firmanın nakit akış deęişiminin ya da karlılığın birçok faktöre baęlı olarak deęişebildiğini

³ Håkan Jankensgård, Lars Oxelheim, Niclas Andrén, “Exposure-Based Cash-Flow-at-Risk: An Alternative to VaR for Industrial Companies”, **Lund University**, Volume:17, 2005,pp. 76-86, s.77

⁴ Sezer Bozkuş, “Risk Ölçümünde Alternatif Yaklaşımlar: Riske Maruz Deęer (VaR) ve Beklenen Kayıp (ES) Uygulamaları”, **Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi**, Cilt: 20 Sayı: 2, s. 27-46,2005, s.31.

⁵ Güven Gül Polat, “Sensitivity Analysis of Expected Shortfall by means of a second-order Approximation”, İstanbul Teknik Üniversitesi, Yüksek Lisans Tezi, 2012, s.8.

görmekteyiz. Bu açıdan RMD yeterli olmamaktadır ve RMNA göstergesine olan ihtiyaç da VaR'ın bu sınırlılığından kaynaklanmaktadır. Her ne kadar RMD göstergesi kadar tek tip bir kullanıma sahip olmasa da RMNA'nın oluşturulma mantığı da RMD ile aynıdır: Nakit değişimlerdeki dalgalanmaları öngörmek ve bu değişimleri içeren bir güven aralığı dağılımı bulmak. VaR'dan farklı olarak, finansal enstrüman değerlerindeki değişim yerine nakit değişimi konulmuştur. Ancak RMNA uygulanmasında VaR'a kıyasla bazı zorluklar bulunmaktadır.

Bunlardan bir tanesi RMNA hesaplamasındaki veri miktarının eksikliğidir.⁶ Finansal bir gösterge olarak RMD hesaplamasında böyle bir veri kısıtlılığı sorunu yoktur çünkü portföy değişimi portföydeki finansal enstrümanlara bağlıdır. Finansal araçlarla ilgili de günlük, hatta saatlik veri bulabilmek daha kolaydır. Böylece tümevarımsal bir çıkarım yapılabilir. RMNA için ise firmanın bilançosu, gelir tablosu veya nakit akış tablosu kullanılır. Ancak firmaların finansal tabloları genellikle çeyrek dönem sonlarında yayınlanır. Bu da finansal enstrümanlarla karşılaştırıldığında çok az veri elde edebileceğimiz anlamına gelmektedir. Böylece, RMD hesaplamasındaki gibi bir örüntüye yıllık sadece 4 tane çeyrek dönem nakit raporlaması yaparak ulaşmak mümkün değildir.

Ayrıca, RMD hesaplaması yaparken bir finansal aracın kendi geçmiş verilerine / değişimlerine bakmak bile yeterli olabilir. Buradan genel geçer bir değişim trendine ulaşabiliriz. Ancak nakit değişimlerdeki etkenler biraz daha farklıdır: Karlılık, operasyonel değişimler, yatırım kararları, finansal tercihler ve piyasa / sektör hareketleri gibi farklı unsurlara bağlı olarak nakit / karlılık miktarı değişebilir. Bir başka deyişle, nakit değişimlerini zamansal değişimlerine bakarak kendi kendileriyle açıklamak kolay değildir. Nakit akışı dışsal etkenlere daha çok bağımlıdır ve bilançodaki her bir değişkenin etkisini ölçümlemek kolay değildir.⁷ Kısacası, hem veri setinin görece küçük olması hem de nakit akışını etkileyen faktörlerin çokluğu RMNA hesaplamasını zorlaştırmaktadır. RMNA hesaplamasındaki bu ve benzeri zorluklar uygulamada kendi içinde de farklı farklı metotların kullanılmasına sebep olmuştur. RMNA ve RMD karşılaştırılması Tablo 1'de görülebilir:

⁶ Stein, Jeremy v.d., "A Comparables Approach to Measuring Cashflow-at-Risk for Non-Financial Firms", **Journal of Applied Corporate Finance**, Vol. 13, No. 4, 2001, pp. 100-109, s.8

⁷ Serdar Kuzu, "Endüstri İşletmelerinde Risk Ölçüm ve Yönetiminde Riske Maruz Nakit Akış Yönteminin Kullanımı ve Bir Uygulama Örneği", Marmara Üniversitesi, Doktora Tezi, İstanbul, 2013, s. 164.

Tablo 1: RMD ve RMNA

	RMD	RMNA
<i>Risk Kapsamı</i>	Portföy Riski	Nakit Riski
<i>Firma Tipi</i>	Finansal Şirket	Finansal olmayan Sirket
<i>Veri Hacmi</i>	Büyük	Küçük
<i>Etkileyen Değişkenler</i>	Az sayıda değişken	Çok sayıda değişken
<i>Duyarlılık</i>	Dışsal etkiler daha az	Dışsal etkiler daha çok

RMNA göstergesine ihtiyaç ilk olarak JP Morgan içerisinde yer alan RiskMetrics grubu tarafından görülmüş ve RMNA bir terim olarak kullanılmıştır⁸. Bu yaklaşım nakit akış tahmini için proforma bir nakit akış tablosu oluşturmaya dayanmaktadır. Bir firmanın nakit akışını etkileyen finansal değerler tek tek tahmin edilerek bir “proforma nakit akış tablosu” oluşturulup olası nakit değişim riski belirlenir. Bu daha çok bir “özelden-genele” (bottom-up) olarak adlandırılan bir yöntemdir. Bütün değişkenlerin tek tek hesaplanması gerekmektedir. Bu tür detaylı bir yaklaşımın elbette güvenilir ve gerçekliğe uygun sonuçlar üretmesi beklenir, ancak hesaplanması uzun ve zahmetlidir. Modelin belli bir döneme göre uyarlanmış olması modelin esnekliğini azaltmaktadır. Dolayısıyla bir firma için “özelleştirilmiş” olması bir dezavantaj oluşturmaktadır. Ayrıca birçok değişkenin hesaba katılması hangi değişkenlerin etkili olduğunu belirlemeyi gerektirmektedir ki, bu da RMD gibi pratik bir göstergenin temel özelliğinden uzaklaşmak demektir. Ancak bu metot RMNA kavramının ortaya çıkışında önemli bir role sahiptir ve kullanım amacını anlamak için göz ardı edilmemelidir.

İkinci olarak Stein ve diğerlerinin ortaya koyduğu “genelden-özele” (top-down) olarak adlandırabileceğimiz yaklaşımından söz edebiliriz⁹. Aslında bu çalışmanın RMNA için ilk sistemli yaklaşım olduğu söylenebilir¹⁰. Bir başka deyişle, bu çalışmada RMNA teriminin ilk somut haline rastlamaktayız. Stein ve diğerleri, RiskMetrics yaklaşımını fazla “özelleştirilmiş” olduğu ve sektörel değişimleri ihmal ettiği gerekçesiyle eleştirmişlerdir. Belli bir firmaya özgü olması pratik

⁸ RiskMetrics – **Technical Document**, J.P. Morgan, December 17, 1996.

⁹ Stein, Jeremy v.d., a.g.e.

¹⁰ Kuzu, a.g.e., s. 209.

kullanımını engellemektedir. Ayrıca birçok değişkenin hesaba katılması hesaplamayı hata yapmaya duyarlı hale getirmektedir. Bunların dışında, bir firmanın nakit değişimlerini sadece o firmanın içsel değişkenleriyle açıklanması gerçekçi bir yaklaşım değildir. Bunun yerine sektörel etkilerin de hesaba katılması gerekmektedir.¹¹

Makalede değindikleri ikinci temel konu ise veri eksikliğinin giderilmesidir.¹²Bu durumu aşabilmek için birçok firmanın verisi üzerinden bir genel eğilim belirlemeye çalışmışlardır. Buna göre bir sektör eğilimi belirleyip bunun bireysel firmalara uygulanabilirliğini göstermişlerdir. Yöntem, firmaları farklı farklı gruplara ayırabilirsek benzer firmaların benzer nakit değişim şemalarına sahip olacağı varsayımına dayanmaktadır. Kısacası hem veri eksikliğinin giderilmesi hem de sektörel davranışların etkilerini ölçümleyebilmesi bakımından önemli bir metot geliştirmişlerdir. Özellikle firmaların nakit değişim verilerinin eksik olması durumunda bile modelin işliyor olması oldukça pratik bir avantaj sağlamaktadır.

Üçüncü olarak, bu iki yaklaşımın dışında bir yerde duran çalışmalardan bahsedebiliriz. Örneğin; Andrén / Jankensgård / Oxelheim tarafından yazılan iki makale önemlidir: “Exposure-Based Cash-Flow-at-Risk: An Alternative to VaR for Industrial Companies”¹³ ve “Exposure-Based Cash-Flow-at-Risk for Value-Creating Risk Management under Macroeconomic Uncertainty”¹⁴. Bu çalışmalarda belli bir firmanın nakit değişimlerinin makro ekonomik göstergelerle ilişkisine göre bir regresyon modeli oluşturulmuş ve bu modelin 10.000 kez Monte Carlo simülasyonuna göre bir güven aralığı dağılımı elde edilmiştir. Makro ekonomik göstergelerin (Kur değişimi, enflasyon, güven endeksleri vb.) sürekli olarak takip edildiği ve tahmin edildiği günümüzde bu göstergelerle firmaların nakit değişimleri arasındaki ilişki üzerinden bir tahminde bulunmak hem pratiktir hem de farklı farklı firmalara uygulanabilir olması bakımından esnek bir model sunabilmektedir. Makro ekonomik göstergeler üzerinden RMNA modeli oluşturmanın bazı avantajları vardır. Öncelikle, doğruluk oranı yüksek bir tahmin modeli kurulabilirse, karar alıcı pozisyonundaki kişiler sadece beklenen makro ekonomik göstergelere göre gelecekteki nakit değişim riskini öngörebilir ve buna

¹¹Stein, Jeremy v.d., a.g.e.,s.9.

¹² Stein, Jeremy v.d., A.e.,s.8

¹³ Håkan Jankensgård, Lars Oxelheim, Niclas Andrén, a.g.e.,

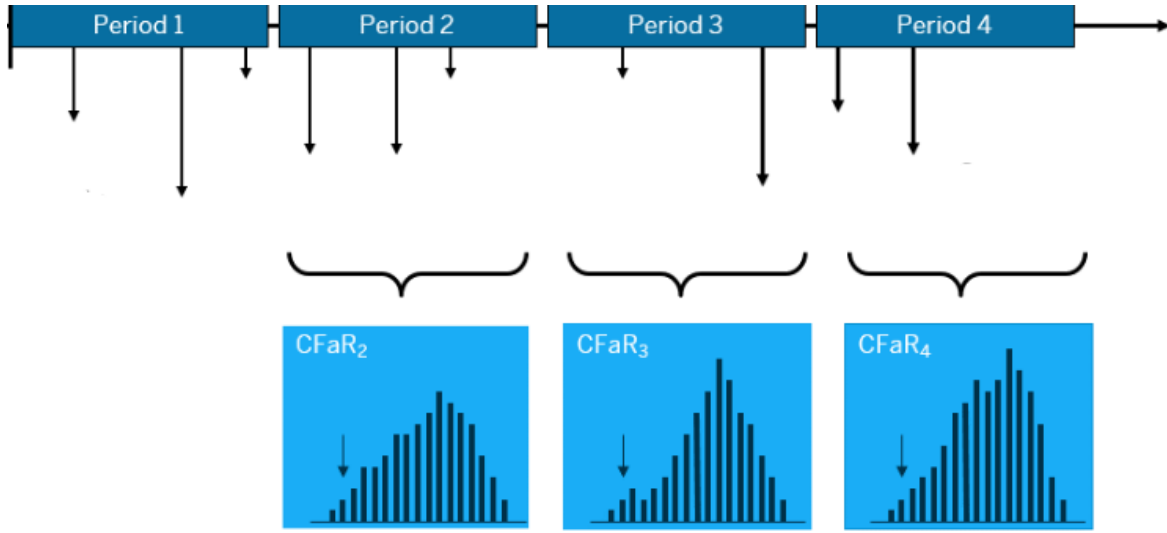
¹⁴Håkan Jankensgård, Lars Oxelheim, Niclas Andrén, “Exposure-Based Cash-Flow-at-Risk for Value-Creating Risk Management under Macroeconomic Uncertainty”, **Working Paper Series 843**, Research Institute of Industrial Economics, 2010.

göre nakit akış politikası oluşturabilir. İkinci olarak, bu değişkenlerdeki beklentilere göre RMNA sürekli olarak güncellenebilir ve her duruma uyarlanabilir, esnek bir model elde edilir.

Özetle, RMNA için gelecekteki olası nakit akış iflas riskini göstermek için bilgi verici bir gösterge olduğunu söyleyebiliriz.¹⁵ RMD göstergesine benzer bir şekilde olası nakit akışların bir dağılımı ortaya konulur ve güven aralığı sınır değeri, RMNA tahmini olarak kabul edilir.

$$RMNA = RMNA (Risk Faktörü1, Risk Faktörü2...)^{16}$$

Şekil 2: RMNA Dağılımı



Kaynak:https://help.sap.com/docs/SAP_ERP/67736291c0b44497b496de907e054745/90325c535b404e3fe10000000a441470.html [19.08.2023]

Son 10 yılda yaşanan bazı gelişmelerin RMNA ile ilgili çalışmaların yapılmasını kolaylaştırdığını düşünebiliriz. Bunlardan birincisi, veriye erişimin son daha kolay gelmesi ve veri saydamlığındaki artıştır. İnternet teknolojisi verinin dijitalleştirilmesini hızlandırmış ve birçok veri tabanı bireylere açık hale gelmiştir. Bu süreçte hem kullanıcıların dijital platformlara erişimi kolaylaşmış hem de

¹⁵Cash flow at Risk (CFaR),

https://help.sap.com/docs/SAP_ERP/67736291c0b44497b496de907e054745/70ee5953c91eff4fe10000000a44176d.html

¹⁶Cash flow at Risk (CFaR), A.e.

firmalar veri sağlayıcı işlevlerini daha saydam ve düzenli olarak yerine getirmektedir. Böylece veri setine erişim sorunu eskiye oranla biraz daha azalmıştır.¹⁷

Bunun dışında, veri analitiğindeki gelişmeleri düşünebiliriz. Yeni bilgisayar teknolojileri (“Büyük Veri”, “İş Zekâsı” vb. gelişmeler) sadece niceliksel olarak büyük verilerin kullanımını değil farklı nitelikteki verilerin de aynı platformlarda kullanılabilmesini ve yönetilebilmesini kolaylaştırmıştır. Bu açıdan düşünüldüğünde, sadece veri büyüklüğünün getirdiği avantajdan değil aynı zamanda bir platformda farklı nitelikteki verilerin tutulabilmesinden ve buradan verilerin yönetilebilmesinden bahsetmemiz daha kolay hale gelmiştir. Bu şekilde bir tahmin modeli kurarken farklı nitelikteki değişkenleri tek bir veri tabanında depolamak, sınıflandırmak ve daha kompleks bir veri seti üzerinden bir model oluşturmak çok daha elverişlidir¹⁸.

Bir diğer unsur da verilerin işlenebilmesindeki teknolojik imkanların artmasıdır. Farklı popülasyonlardan farklı örneklemeler hızlı bir şekilde alınıp incelenebilir ve farklı örneklemeler üzerinden binlerce simülasyon aynı anda yapılabilir¹⁹. Bu şekilde kısa sürede sonuç alabilmek farklı farklı yöntemleri aynı anda uygulayabilmemize yardımcı olmaktadır. Örneğin, İstatistik Biliminde 1900’lü yılların başında oldukça popüler olan t tablosu yerini gelişen programlama tekniklerine ve simülasyonlara bırakmıştır²⁰. Benzer bir örneği Şirket denetiminde de görebiliriz. Muhasebe denetiminde belli bazı örneklemeler seçip onları denetlemek yerine, son yıllarda popülasyon verisi üzerinden denetime geçilmiştir²¹. Gelecekte, tüm bu yeni süreçlerin daha da hızlanması ve yaygınlaşması beklenmektedir.

Sonuç olarak, matematiksel yöntemlerin daha elverişli bir şekilde veri analizine uygulanması günümüzde çok daha kolaydır. Örneğin; daha önceki RMNA çalışmalarında istatistiksel analizler, doğrusal regresyon uygulamaları ve basit zaman serisi algoritmaları kullanılmış ve oldukça başarılı sonuçlar elde edilmiştir. Makine Öğrenmesi (ML) ya da Yapay Zekâ (AI) temelli matematiksel teoriler ve algoritmalar 2000’li yıllardan önce mevcut olsa da bu yöntemlerin pratik uygulanabilirliği geçmiş yıllarda pek de kolay değildi. Ancak bilgisayar programlama dilleri bugün hem kompleks algoritmaları pratik olarak uygulama kabiliyeti kazanmış hem de bunları

¹⁷ Christine Earley, “Data analytics in auditing: Opportunities and challenges”, Business Horizons, Kelly School of Business, Volume 58, Issue 5, pp. 493-500, 2015, s.494.

¹⁸ Earley, A.e. s.494.

¹⁹ Earley, A.e. s.494.

²⁰ Peter Bruce and Andrew Bruce, **Practical Statistics for Data Scientists**,2017, s.77.

²¹ Earley, a.g.e., s.495.

herkesin kullanabileceği şekilde basit bir hale getirmiştir. Ayrıca, günümüzde aynı anda birçok farklı tahmin modeli denenebilir ve test edilebilir. Tam bu noktada, daha önce kullanılan doğrusal regresyon modellerinin yanı sıra farklı algoritmaların da RMNA modellerini geliştirmede bize yardımcı olup olmayacağı sorusu önemli hale gelmektedir.

LİTERATÜR TARAMASI

RMNA literatüründeki çalışmalar 1990'lı yılların sonunda başlamış ve geçtiğimiz yirmi yıl boyunca farklı metotlarla uygulanmıştır. Bu farklılıkların daha iyi anlaşılabilmesi ve tezimizdeki yaklaşımı daha iyi konumlandırmak adına bugüne kadar yapılan başlıca çalışmalara göz gezdireceğiz.

RMNA ile ilgili başlangıç adımını JP Morgan içerisinde yer alan RiskMetrics grubunun 1996 yılında yaptığı çalışma oluşturmaktadır²². RiskMetrics grubu bir firma için olası riskleri ve bu risklerin nasıl ortadan kaldırılacağı konusunda tavsiyeler içermektedir. Nakit akış riski de bir firma için başlıca risklerden birisidir. Çalışmada, nakit akış değişimi proforma nakit akış tablosu yaklaşımı ile modellenmektedir. Bir diğer deyişle, öncelikle bir firmanın nakit akışını etkileyen finansal değerler tek tek tahmin edilir. Sonrasında, bu değerler üzerinden bir “portföy nakit akış tablosu” oluşturulup olası nakit değişim riski belirlenir. Bu daha çok bir “Özelden genele” (bottom-up) olarak adlandırabileceğimiz bir yöntemdir. Çalışmada nakit akış tablosunu oluşturan operasyonel, yatırım ve finansman kaynaklı nakit akış değişimleri ayrı ayrı düşünülmüş ve bu değişimlerin arkasında yatan etkenler hesaba katılmıştır. Oldukça detaylı ve uzun bir çalışma olması bakımından pratik olmadığı gerekçesiyle eleştirilmiş olsa da nakit akış risk modellemesi için oldukça önemli bir başlangıçtır. Ayrıca bir firmanın finansal özellikleri biliniyorsa tek firma üzerinde oldukça somut sonuçlar verebilmektedir. Yine de modelin belirli bir firmaya odaklanmış olması, yani fazlaca “özelleştirilmiş” olması esnek bir gösterge olarak kullanımını engellemektedir. Ayrıca birçok etkenin modele dahil ediliyor olması hesaplamaların uzun zaman almasına yol açmaktadır.

RMNA konusundaki ikinci önemli çalışma Stein ve diğerlerinin 2001 yılında yazdığı makaledir²³. Bu makalede “genelden-özele” (top-down) yaklaşımı benimsenmiştir. Stein ve diğerleri, RiskMetrics yaklaşımını fazla “özelleştirilmiş” olduğu ve sektörel değişimleri ihmal ettiği

²² RiskMetrics a.g.e.

²³Stein, Jeremy v.d., a.g.e.

gerekçesiyle eleştirmiş ve birçok firmanın verisi üzerinden bir genel eğilim belirlemeye çalışmışlardır. Yukarıda da bahsedildiği üzere, nakit akış verileri çeyrek yıllık raporlamalara dayalı olduğu için firmanın yapısı hakkında yeterli bilgi vermemektedir. Bu sorunu aşmak için firmaları hisse dalgalanmaları ve piyasa hacimlerine göre gruplara ayırmışlar ve benzer gruplardaki firmaların benzer nakit akış değişimleri göstereceği varsayılmıştır. Kısacası sektörel eğilimleri hesaba katabilmek, firma riskini sadece içsel verilere indirgememek ve veri sınırlılığı problemini aşabilmek için firmaları gruplandırmışlar ve gruplardaki tüm değişimleri tek bir dağılıma indirmişlerdir. Farklı nakit değişimleri olan firmaları tek bir grupta toplayabilmek için bu verilerin normalize edilmesi gerektiğinden, doğrudan nakit değişimini değil FAVÖK /Toplam Varlıklar oranını kullanmışlardır. Burada, FAVÖK değişim değerinin operasyonel nakit akışını değiştirdiği varsayılarak bir RMNA modeli kurgulanmıştır.

Bu çalışmadaki ikinci önemli unsur, regresyon sonuçları ile gerçek sonuçlar arasındaki hata değerlerine odaklanılmış olmasıdır. FAVÖK / Toplam Varlıklar değişimi oranı her firma için bir doğrusal regresyon ile tahmin edilmiş ve bu tahminlerle gerçek değerler arasındaki sapmalar not edilmiştir. Gruptaki tüm firmaların olası sapmaları bir araya getirildiğinde FAVÖK / Toplam Varlıklar sapmalarının bir dağılımını elde etmişlerdir. Bunun sebebi nakit akış değişiminin beklenen değerden (regresyon sonucundan) ne oranda saptığını görebilmek ve olası şok durumlarını modellemektir. Bu çalışmayla yazarlar hem yetersiz veri problemini aşmışlar hem de her firmaya uygulanabilen (VaR'a benzer bir şekilde) pratik bir yaklaşım ortaya koymuşlardır.

Türkiye'de bu makaleden esinlenen iki farklı teze bakmamız önemli olacaktır. Özvural'ın 2004 yılında yazdığı yüksek lisans tezinde Stein ve diğerlerinin metodu direkt olarak uygulanmıştır²⁴. Yazar, Türkiye'de 216 halka açık firmayı piyasa değerlerine ve hisse değişimlerine göre 15 farklı gruba ayırmış ve gruplara göre FAVÖK / Toplam Varlıklar değişimlerinin regresyon hatalarını listelemişlerdir. Sonrasında, bu hataların oluşturduğu dağılımlara göre her bir grup için bir güven aralığı oluşturulmuştur. Son olarak da bu dağılıma göre bir sonraki çeyrek dönemi değişimi tahmin edilmeye çalışılmıştır.

Bu yöntemle dayanarak yapılan ikinci bir çalışma Kuzu tarafından yazılan 'Endüstri İşletmelerinde Risk Ölçüm ve Yönetiminde Riske Maruz Nakit Akış Yönteminin Kullanımı ve Bir Uygulama

²⁴ Özhan Özvural, "Cash Flow at Risk in Publicly Traded Non-Financial Firms in Turkey: An Application in Defense Companies", İhsan Doğramacı Bilkent Üniversitesi, Yüksek Lisans Tezi, Ankara, 2004.

Örneği' adlı doktora tezidir²⁵. Uygulama örneğinde, Borsa İstanbul'da işlem gören çimento, demir çelik, gıda ve tekstil sektöründe faaliyet gösteren endüstri işletmelerinin 2007/12-2012/9 dönemleri arasındaki çeyrek dönemlik bilanço ve gelir tablosu değerleri kullanılarak bir faaliyet nakit akışı riski modellenmesi yapılmıştır. Gruplama kriteri olarak sektör ve piyasa değerleri kullanılmıştır. Yine yukarıdaki çalışmalara benzer şekilde regresyon hata değerleri bu kriterlere göre gruplandırılmış ve FAVÖK / Toplam Varlıklar oranları tahmin edilmeye çalışılmıştır.

Firmaların gruplandırılması üzerine yapılan bir başka çalışmada Tüfekçi, Türkiye'de çeşitli sektörlerde faaliyet gösteren firmaları nakit akış profillerine göre farklı gruplandırmalara tabi tutmuştur²⁶. Burada RMNA modellemesi yapılmamış olsa da farklı gruplandırma yöntemlerinin kullanılabilirliğini göstermesi açısından bu çalışma önemlidir. Bu tezde, faaliyet nakit akışları yatırım nakit akışları ve finansman nakit akışları durumlarına göre firmalar "Başarılı İşletme", "Büyüyen İşletme" ve "Genç İşletme" vb. olarak gruplandırılmış ve sektörlere göre oluşan kümelerin yapısı analiz edilmiştir.

Yukarıdaki iki perspektifin dışında, Andren, Jankensgard ve Oxelheim 2005²⁷ ve 2010²⁸ yıllarında yazdıkları iki farklı makalede, RMNA tahminini makroekonomik göstergelere dayanarak açıklamaya çalışmışlardır. Yazarlar belli bir firmanın nakit değişimlerinin makro ekonomik göstergelerle ilişkisine göre bir regresyon modeli oluşturmuşlar ve bu modeli 10.000 kez Monte Carlo simülasyonu yaparak bir güven aralığı elde etmişlerdir. Her iki yazıda da önce bir firmanın açık olduğu makro ekonomik risk göstergeleri belirlenmiştir. Sonrasında, bu göstergeleri bağımsız değişken, nakit akış değişimini de bağımlı değişken olarak düşünüp bir regresyon oluşturmuşlardır. Son aşamada, bu regresyon modeline dayanarak 10.000 kez monte Carlo simülasyonu uygulayarak RMNA dağılımı elde edilmiştir.

Gentili ve diğerleri, 2018 yılında yaptığı çalışmada Andren ve diğerleri tarafından ortaya konulan "Exposure-based" yaklaşımdan esinlenmişlerdir.²⁹ Bu bağlamda, belli bir firmanın nakit akış dağılımını bilanço maddelerine (Varlıklar / Yükümlülükler / Alacaklar vb.) formal ve

²⁵ Kuzu, a.g.e., s.

²⁶Tüfekçi, Merve, **Firmaların Nakit Akış Profilleri ile Finansal Performanslarının Karşılaştırılması: BİST'te bir Uygulama**, 2020.

²⁷ Hâkan Jankensgård, Lars Oxelheim, Niclas Andrén, 2005, a.g.e.

²⁸ Hâkan Jankensgård, Lars Oxelheim, Niclas Andrén, 2010, a.g.e.

²⁹ Gentili, Luca, Giacomello, Bruno, Girardi, Dario, Grasselli, Martino: **A Dynamic Model for Cash Flow at Risk**, SSRN, 2018.

matematiksel bir yaklaşım geliştirerek ortaya koymuşlardır. Hem bir firmanın bilanço hassasiyetlerini hem de makro ekonomik risklere olan duyarlılığını tek bir model üzerinden hesaba katıp nakit akış değişimi için güven aralığı oluşturmuşlardır.

Benzer bir “Exposure-based” temelli yaklaşım Yan ve diğerleri tarafından 2011 yılında bankacılık sektörüne uygulanmıştır³⁰. Aslında yaygın olarak RMD yaklaşımı bankalar ve finansal kurumlar için, RMNA yaklaşımı ise finansal sektörü dışındaki şirketler için uygulanmış olsa da bu makalede RMNA kavramı Birleşik Krallık bankacılık sektörüne uyarlanmış ve bankaların dışsal makroekonomik faktörlere duyarlılıklarına göre RMNA modeli ortaya konulmuştur. Bu çalışmada, Andern ve diğerlerinin makalesindeki adımlar birebir uygulanmıştır. Önce regresyon modeli oluşturulmuş ve sonrasında Monte Carlo simülasyonu uygulanmıştır. Ağırlıklı olarak bankaların nakit akış risklerinin kısa vadeli faiz oranları, kur değişimi ve fiyat değişimlerine göre nasıl etkilendiği modellenmektedir.

Nakit akış tahmini için bir başka tez çalışması 2021 yılında Vural tarafından yapılmıştır³¹. Bu çalışmada Gıda sektöründeki bir firmanın Operasyonel nakit akışları Sistem Dinamiği modeliyle tahmin edilmeye çalışılmıştır. Doğrudan RMNA modeline dayalı bir çalışma olmasa da disiplinler arası bir yaklaşımın nakit akış tahmininde kullanılması ve sektörel analizin göz önünde bulundurulmuş olması bakımından önem taşımaktadır.

2005 yılında Kısakürek tarafından yapılan bir başka çalışmada³² İMKB üzerinden uygulama ile desteklenen nakit akış tahmin modeli ortaya konulmuştur. Çalışmada, 131 firmanın 11 yıllık verilerine dayanarak gelecekteki nakit akış miktarları tahmin edilmeye çalışılmıştır. Bu uygulamada tek değişkenli ve çok değişkenli regresyon modelleri kullanılmış ve yapılan tahminler t-testi ve Wilcoxon testi ile desteklenmiştir. Sonuç olarak, en iyi tahmin değişkeni olarak Faaliyet (Operasyonel) Nakit Akışı tespit edilmiştir.

³⁰Meilan Yan v.d., “Estimating Liquidity Risk Using The Exposure-Based CashFlow-at-Risk Approach: An Application To the UK Banking Sector”, **Wiley Online Library**, Volume19, Issue3, 2014, pp. 225-238.

³¹Merve Vural, “Operasyonel Nakit Akışlarının Sistem Dinamiği Yaklaşımıyla Tahmin Edilmesi : Gıda Sektöründe bir Uygulama”, İstanbul Ticaret Üniversitesi, yüksek lisans Tezi, 2021.

³² Mustafa Kısakürek, İşletme Faaliyet Nakit Akımlarının Tahmin Edilmesi – İMKB’de bir Uygulama, Hacettepe Üniversitesi, Doktora Tezi, Ankara, 2005.

2019 yılında Demirkapı tarafından yazılan bir başka tez çalışmasında³³ inşaat sektöründe yaşanan ödeme gecikmelerini de göz önüne alarak projelerden kaynaklı nakit akış değişimlerinin tahmini üzerine odaklanmıştır. Bu tür projelerdeki ödeme gecikmelerinden kaynaklanan riskleri de göz önünde tutarak Stokastik bir nakit akış simülasyon modeli ve RMNA modeli geliştirilmiştir. Projeler, ödeme gecikmeleri ve nakit akış arasında ilişki kuruyor olması çalışmayı dikkate değer kılmaktadır.

2019 yılında Avşar tarafından yazılan bir yüksek lisans tezinde ise³⁴ diğer çalışmalardan farklı olarak Rasyo sonuçlarına odaklanılmıştır. Bir firmanın 2015-2017 yılları arasındaki rasyo değerleriyle finansal tabloları dikkate alınmış ve incelemenin sonucunda nakit yönetiminin etkinliği ve doğruluğu konusunda sonuçlara ulaşılmıştır.

PROBLEM İFADESİ

RMNA üzerine yapılan önceki çalışmalarda sıklıkla Doğrusal regresyonun kullanıldığını ve regresyon tahmini üzerinden RMNA değerine ulaşıldığını görmekteyiz. Ayrıca yaygın olarak tahmin performansı üzerinde herhangi bir test yapılmadan direkt olarak RMNA hesaplamasına geçilmektedir. Modelleri biraz daha sofistike hale getirerek, doğrusal regresyon yerine Makine Öğrenmesi algoritmalarının kullanılması hem tahmin performansı açısından hem de RMNA ölçümü açısından fayda sağlayacaktır. Bu araştırmanın amacı, işletmelerde RMNA hesaplamasında, Doğrusal Regresyona kıyasla Makine Öğrenmesi algoritmalarına göre hesaplanmasının etkilerini bir örneklem kullanarak açıklamaktır. Buna ek olarak, önceki çalışmalarda yaygın olarak tahminlere performans testinin çoğu kez uygulanmadığı görülmüştür. Bu çalışmada ise, RMNA tahminlerinin performans ölçütü olarak RMD uygulamalarında kullanılan Kupiec testi ve Süre testi uygulanmıştır.

Hipotez 1: Doğrusal Regresyona ek olarak Makine Öğrenmesi algoritmalarının kullanılması nakit akış tahmin başarısını arttıracaktır.

Hipotez 2: Doğrusal Regresyona ek olarak Makine Öğrenmesi algoritmalarının kullanılması RMNA tahmin başarısını arttıracaktır.

³³ Hande Betül Demirkapı, “A Cash flow at Risk (CFaR) Modle for Managing Payment Delays in Construction Projects”, İzmir Yüksek Teknoloji Enstitüsü, Yüksek Lisans Tezi, 2019.

³⁴Türkan Avşar, “Nakit Akış Tablosu ve Analizi Aracılığıyla İşletmelerde Nakit Yöneimi ve bir Uygulama”, Marmara Üniversitesi Yüksek Lisans Tezi, İstanbul, 2019.

ARAŐTIRMA SORUSU

Finansal olmayan iŐletmeler iin RMNA ve nakit akıŐ deęerleri tahminlerinde doęrusal regresyona ek olarak Makine renmesi algoritmalarının kullanılması nakit akıŐı ve RMNA iin tahmin performanslarını arttırır mı?

Baęımsız deęiŐken: eyreklik dnem sıralaması

Baęımlı DeęiŐken: eyreklik nakit akıŐ deęiŐimi

Tahmin Performans lt: Kk Ortalama Kare Hatası, Ortalama Mutlak Hata

RMNA Performans lt: RMNA sınır deęeri ihlal sayısı, Kupiec Testi ve Sre Testi.

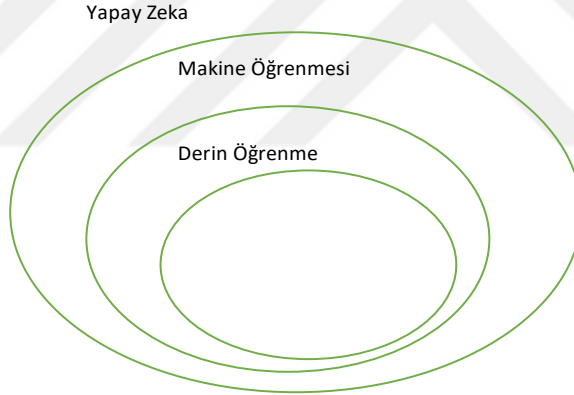
BİRİNCİ BÖLÜM

MAKİNE ÖĞRENMESİ

1.1.Makine Öğrenmesi Literatürü

Makine Öğrenmesi, temel olarak kendi kendine öğrenen ve hatalarını gitgide düzeltebilen algoritmalara (yöntemlere) referansta bulunmaktadır. Kendi kendine ilerleme ve hata düzeltme özelliği bakımından Yapay Zekâ (AI) alanının bir alt dalı gibi düşünülebilir. Bu bakımdan doğrusal regresyondan ağaç temelli çoklu algoritmalara kadar kullanılan bütün yöntemleri Makine Öğrenmesi başlığı altına yerleştirilebilir.

Şekil 3: Makine Öğrenmesi / Derin Öğrenme / Yapay Zekâ



Kaynak: <https://levity.ai/blog/difference-machine-learning-deep-learning> [19.08.2023]

Temel olarak Makine Öğrenmesi algoritmalarının iki temel işlevi olduğunu söyleyebiliriz:³⁵ İnsanların ya da diğer canlıların yapabildiği işleri yapabilirler. Görsel tarama ya da ses tanıma programları bunlara iyi birer örnektirler. Bazı durumlarda ise insan zekasının ötesine geçebilirler. Tıbbi işlemler, e-ticaret, bankacılık veya hava tahmini gibi işlemlerde bir kişinin veremeyeceği kararları verebilirler veya tahminler oluşturabilirler. Banka Kredi başvuruları buna karakteristik bir örnek teşkil eder. Bir kişi kredi başvurusu yaptığında o kişiyle ilgili yüzlerce değişken bir araya getirilir ve bu değişkenlerin sonuç üzerindeki ağırlığına göre bir karar/tahmin oluşturulur.

³⁵ Shai Ben-David and Shai Shalev-Shwartz, **Understanding Machine Learning: From Theory to Algorithms**, Cambridge University Press, 2014, s.21-s.22.

Algoritmalar, bir insanın tek başına veremeyeceği bu tür kararlarda gerek tahmin başarısı olarak gerekse de tahmin hızı olarak insanların yapabileceği işlemlerden çok daha başarılı bir performans gösterebilmektedirler. Makine Öğrenmesi algoritmalarını önemli kılan bir diğer özellik de kolay uyarlanabilmesi ve esnekliğidir. Bir görevin otomatikleştirilmesi süreci kural bazlı bir program yazıldığında sadece o duruma özgü kalabilir; ancak Makine Öğrenmesi temelli programlar ise daha esnekler ve kendi kendine öğrenebilme kabiliyetleri sayesinde kolayca uyarlanabilirler.³⁶

Makine Öğrenmesi literatürü 1960'lı yıllara kadar uzansa da altın çağının olarak 2000'li yıllar olduğunu söyleyebiliriz.³⁷ 2000'li yıllarda Makine Öğrenmesi teknolojisinin gelişmesini kolaylaştıran en önemli etkenlerden birisi Büyük Veri (Big Data) teknolojisidir. Kullanılabilir veri hacminin artması Büyük Veri teknolojisine zemin hazırlamıştır. Bu durum bilim insanlarının merakından ziyade zorunlu bir ihtiyaca, yani büyük hacimli verilerin yönetilmesi ihtiyacına cevap vermektedir.³⁸ Büyük Veri, adında da ifade edildiği üzere, büyük ölçekte verinin depolanması anlamına da gelse de sadece bu tanımla düşünülemez; çünkü sadece niceliksel anlamda büyük ölçekte verinin depolanmasını değil farklı niteliklerdeki verilerin de bir arada yönetilebilmesini mümkün kılmaktadır. Bir başka deyişle, niceliksel bir gelişmeyle birlikte niteliksel bir gelişmeden de söz edebiliriz. Ayrıca, bilgisayar donanımındaki teknolojik yeniliklerin de birçok işlemin çok kısa sürede yapılabilmesi olmasında katkısı büyüktür. Son yıllarda, bilgisayardaki birbirine paralel işlemlerin performans düşürücü etkisi dengelenmiş, Nvidia teknolojisi sayesinde işlemci hızları artmış, RAM maliyetleri ucuzlamıştır. Yazılım teknolojilerinde de bazı sıçramalar olmuştur: NoSQL gibi farklı veri tabanları veri işleme maliyetlerini düşürmüş, bulut tabanlı veri işleme özellikleri geliştirilmiş, 2014 yılında geliştirilen Apache Spark, Makine Öğrenmesi algoritmaları için yapılandırılmış ve zayıf yapılandırılmış verileri Makine Öğrenmesi algoritmaları için hazır hale getirmiştir.³⁹

1.2.Makine Öğrenmesi Çeşitleri

Makine Öğrenmesi algoritmaları genel anlamda denetimli öğrenme ve denetimsiz öğrenme olarak ikiye ayrılabilir.⁴⁰ Denetimsiz öğrenme temelli algoritmalarda (K-Means Kümelendirme vb.)

³⁶ Shai Ben-David and Shai Shalev-Shwartz, a.g.e., s.22.

³⁷ Alexander Fradkov, "Early History of Machine Learning, Institute for Problems in Mechanical Engineering", **Preprints of the 21st IFAC World Congress (Virtual) Berlin**, July 12-17, 2020, s.1377.

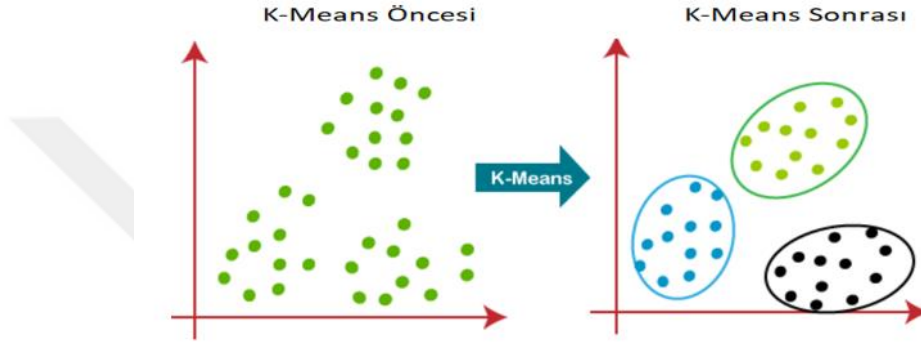
³⁸ Alexander Fradkov, a.g.e.,s.1377.

³⁹ Alexander Fradkov, A.e., s.1377.

⁴⁰ Andrew Bruce, Practical Statistics for Data Scientists, O'Reilly, 2017, s.141, s.283.

algoritmanın eğitilmesi gerekmez. Direkt olarak veriyi kendi kendisine tanır ve buna göre tahmin yürütebilir. Örneğin K-Means kümelendirme algoritması böyledir. Sadece veriler arasındaki ilişkiye bakarak verileri otomatik olarak kümelendirebilir.

Şekil 4: Denetimsiz Öğrenme Örneği olarak K-Means Algoritmasının Etkisi



Kaynak: <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>

Denetimli Öğrenme algoritmalarında ise veri seti, öncelikle eğitim seti ve test (uygulama seti) olarak ikiye ayrılır. Veri önce eğitim setinde eğitilir ve daha sonra gerçek (test) veri setine uygulanır. Yeni veriler eklendikçe algoritma hatalarını düzeltir ve gitgide hata oranını azaltır. Bir bakıma kendi kendini eğitmiş olur. Ancak en başta belli bir veri setine göre eğitilmesi önemlidir. Denetimli Öğrenme algoritmalarını da kendi içerisinde ikiye ayırabiliriz: Regresyon algoritmaları ve Sınıflandırma Algoritmaları. Regresyon algoritmaları sayısal tahmin yaparken kullanılır. Her türlü sayısal tahmin algoritmasını buna örnek gösterebiliriz Sınıflandırma algoritmaları ise sayısal bir tahmin ortaya koymak yerine sonucun hangi gruba ait olduğunu tespit eder; yani onu sınıflandırır. Pozitif/ negatif olarak yapılan COVID-19 testleri, bir işletmede yapılan işlemin hileli olup olmadığı, ya da kredi başvurusu yapan bir kişinin kredi almaya elverişli olup olmadığını tespit eden yazılımlar sınıflandırma temelli denetimli öğrenme grubuna girerler.

1.3. Makine Öğrenmesi Uygulamalarında Zorluklar

Makine Öğrenmesi uygulanmasında bazı dezavantajlar da mevcuttur. Öncelikle, algoritmalar başarılı sonuçlar verebilir ancak nasıl çalıştığı konusunda fikir edinmek zordur. Bu bakımdan bir aksaklık olduğunda, sorunun nereden kaynaklandığının tespit edilmesi kolay değildir. Ayrıca birçok durumda Makine Öğrenmesi temelli programlar önce uygulamaya konulur ancak programın nasıl çalıştığının anlaşılması belli bir süre alır ve gerekli belgelendirme daha sonra yapılır. Bu bakımdan bir gecikme söz konusudur. Bu da kurumsal bir yaklaşımın oluşmasına engel teşkil edebilir. Nitelikli çalışan bulabilmek de en azından günümüzde bir başka problem olarak düşünülebilir. Bu alanda yeterince uzmanlaşmış kişiler bulmak firmalar için daha zordur ve Makine Öğrenmesi uygulamalarının artmasıyla nitelikli çalışan ihtiyacı da buna paralel olarak artmaktadır. Hem doğru ve nitelikli çalışanların bulunması hem de hali hazırdaki çalışanların eğitilmesinin gerekliliği bu noktada öne çıkmaktadır.⁴¹

Son olarak, Makine Öğrenmesi temelli sistemlerin muhasebe risk denetimlerinde ortaya çıkan bazı zorluklardan bahsetmemiz gerekir.⁴²

- Uygulamaların karmaşıklığından ve durumdan duruma farklılık göstermelerinden dolayı Makine Öğrenmesi alanı için doğru bir sınıflandırma literatürünün henüz tam anlamıyla yapılmamış olması.
- İlgili denetim terminolojisinin henüz yeni olması.
- Pratik anlamada bir denetim teamülünün tam olarak kurulmamış olması.
- Akademik alanda Makine Öğrenmesi ile ilgili denetim çalışmalarının yeterli sayıda olmaması.
- Olan akademik çalışmaların fazla teknik (teorik) kalması ve pratik alana uygulanamaması.
- Firmaları kapsayan ortak stratejik öncüllerin belli olmaması nedeniyle denetim sonrasında hangi noktaya varılacağı ve iyileştirmelerin nasıl yapılacağı konusunda ortak bir dilin kurulmaması.
- Denetim firmalarının bu alana karşı esnekliğinin diğer firmalara göre daha az olmasından dolayı denetim firmalarının henüz bu teknolojiye yabancı olmaları.

⁴¹ Earley, a.g.e., s.499.

⁴² Earley, A.e.,s. 497-s.498

Yukarıda saydığımız denetim problemleriyle de ilişkili olarak hem Makine Öğrenmesi uygulamalarının çalışma prensiplerinin birçok kurum için henüz yeni olduğunu söyleyebiliriz. Firmalar, en iyi tahmin yapan modelini değil de basit ama anlaşılır bir yöntemi kullanmayı seçebilir. Örneğin; geleceğe dair toplam maliyet tahminini geçmiş verilere dayanarak yapmak isteyen bir firma bir algoritmayı geçmiş verilere göre eğitip geleceğe doğru bir tahmin modeli ortaya koyabilir. Ancak alternatif olarak, geçmiş değerlerin ortalamasını alıp aynı koşullar altında gelecekte de aynı değerlerin (ortalama değerlerin) gerçekleşeceğini varsayabilir. İlk model (Makine Öğrenmesi modeli) daha karmaşıktır. Bu model daha doğru bir tahmin üretse de ikinci model hem firma içerisinde iletişim kurabilme, tahmin üzerine tartışabilme ve maliyet değişiminin arkasındaki etkenleri anlayabilme açısından daha pratik bir kullanıma sahip olacaktır. Kısacası, bir firma daha iyi tahmin yapan bir modeli değil, daha anlaşılır ve daha basit bir modeli pratik nedenlerden dolayı tercih edebilir çünkü Makine Öğrenmesi temelli bir model zaman zaman hala bir “kara kutu” olarak kalmaktadır.⁴³ Makine Öğrenmesi uygulamalarının “kara kutu” olarak kalmaması, şeffaf bir denetim ve belgelendirme temellerinin oluşturulmasıyla mümkündür. Kısacası Makine Öğrenmesi alanı yeni bir teknoloji olarak yetkin bir alan olsa da uygulamada pratik olmayabilir.

1.4. Makine Öğrenmesi Modellerinde Kriterler

RMNA tahmin modellerinde temel olarak denetimli öğrenmenin bir parçası olarak regresyon modellerini kullanacağız. Elde edeceğimiz tahmin önünde sonunda bir sayı olacağı için (Evet / Hayır şeklinde olmadığı için) sınıflandırma algoritmalarına ihtiyaç duymayacağız. Öncelikle, doğru bir regresyon oluşturmak için aşağıda bahsedeceğimiz belli konulara dikkat edilmesi gerekmektedir.⁴⁴

1.4.1. Doğru Modelin Seçilmesi:

Bir regresyon oluştururken doğru algoritmanın ve doğru parametrelerin seçilmesi önemlidir. Örneğin; bir algoritma belli bir veri setinde diğerlerine göre daha iyi bir tahmin performansı gösterebilir. Bunu belirlemek için çalışmamızda belli bir firma için birçok algoritmayı deneyip hangisinin daha iyi sonuç vereceğini bulacağız. Çalışmamızda aşağıdaki algoritmalar kullanılacaktır:

⁴³ Bruce, a.g.e.,s.258.

⁴⁴ Marc Peter Deisenroth., Aldo Faisal, Cheng Soon Ong, **Mathematics for Machine Learning**, Cambridge University Press, 2021, s.249.

- Doğrusal Regresyon
- En Yakın Komşuluk
- Karar Ağaçları
- Rastgele Ormanlar
- Adaboost
- Gradient Boosting
- XGBoost

1.4.2. Doğru Parametrelerin Bulunması:

Algoritmalar farklı durumlarda farklı veri setlerine göre farklı parametrelerin kullanılmasını gerektirir. Bu parametrelerin değiştirilmesi tahmin gücünü arttıracaktır. Örneğin Karar Ağacı modellerinde bir ağacın boyunun uzunluğu ya da kısalığı bir parametre olarak ayarlanır. Rastgele Ormanlar algoritmasında da kullanılacak ağaç sayısı bir parametre olarak ayarlanabilir. Bu ve benzeri parametrelerin farklı farklı değerlerle denenmesi pratik açıdan doğru bir tahmin üretilmesinde önemlidir. Örneğin; Python programında scikit-learn kütüphanesinin bir parçası olan GridSearchCV bize farklı parametrelerin farklı kombinasyonlarını bir kerede deneyip en uygun parametreleri bize gösterebilir.⁴⁵ Ancak biz bu çalışmamızda, esnek ve karşılaştırılabilir modeller ortaya koyabilmek adına, Python scikit-learn kütüphanesi tarafından verili parametreleri kullanacağız ve firmalar arasında herhangi bir parametre ayırımına gitmeyeceğiz.

1.4.3. Aşırı Öğrenme Problemi:

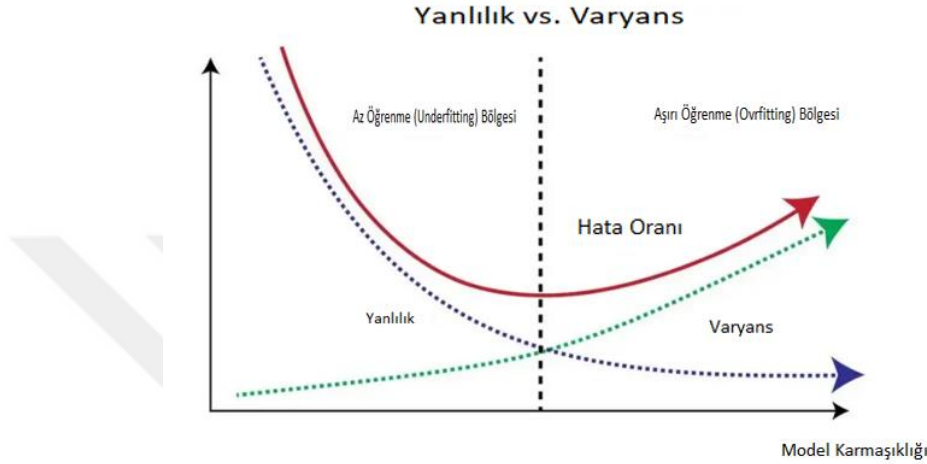
Aşırı öğrenme (overfitting) bir algoritmanın bir veri seti üzerinde çok iyi bir şekilde eğitilmesi sonucu esnekliğini kaybetmesidir. Algoritma bir veri setini o kadar iyi bir şekilde yansıtır ki başka bir veri setine (örneğin; test veri setine) uygulanamayacak derecede esnekliğini kaybeder.⁴⁶ Böyle bir problemin olup olmadığını anlamak bazen kolay olmayabilir. Bu tip modellerde, algoritma eğitim seti üzerinde çok iyi sonuçlar verirken aynı başarıyı test veri setinde gösteremeyebilir. g. Örneğin; Karar Ağaçları algoritmasında, tüm bağımsız değişkenlerin kullanılması ya da ağaç derinliğinin kontrol edilmemesi bize çok detaylı ve kompleks bir model verebilir; ya da sadece bir duruma özgü oldukça sofistike bir model sağlayabilir. Ancak aynı model, bir başka veri setinde

⁴⁵ GridSearchCV, https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

⁴⁶ Bruce, Peter, Bruce, Andrew: **Practical Statistics for Data Scientists**, O'Reilly, 2017, s.113.

ise aynı başarıyı gösteremeyebilir. Bu tip modellerde, ağaç derinliği aşırı derecede artar, eğitim veri setini çok iyi yansıtır, ancak model test veri setinde başarısız olur.

Şekil 5: Aşırı Öğrenme Durumlarında Varyans ve Yanlılık Dengesi



Kaynak: <https://towardsdatascience.com/bias-and-variance-but-what-are-they-really-ac539817e171> [19.08.2023]

Yukarıdaki grafikte modelin karmaşıklığı ile hata fonksiyonu tarafından ortaya konulan hata değeri (sapma değeri) arasındaki ilişkiyi görmekteyiz. Sezgisel olarak düşünüldüğünde, modelin varyansının (karmaşıklığı) yüksek olması yanlılık (bias) miktarının da düşük olması beklenir. Ancak modelin karmaşıklığının artması bir noktadan sonra modelin test veri seti üzerindeki tahmin başarısını düşürür. Bir başka deyişle standart hata oranını arttırmış olur. Bunun en önemli sebebi yukarıda da bahsedildiği üzere modelin eğitim setini aşırı öğrenmesi ve bu yüzden esnekliğini kaybediyor olmasıdır. Dolayısıyla yanlılığı mümkün olduğunca düşürüp karmaşıklığın çok fazla arttırılmadığı bir optimum nokta vardır.⁴⁷ Bu nokta optimum noktadır.

1.4.4. Otokorelasyon Problemi:⁴⁸

Otokorelasyon problemi bir verinin geçmişteki verilerle ilişkili olarak açıklanabilmesi durumudur. Böyle durumlarda kurulan model çok iyi bir tahmin modeli gibi görünebilir, ancak aslında tahmin başarısı modelin kendisinden değil verinin zaten geçmişteki veriler üzerinden çıkarsanıyor

⁴⁷ Bruce, a.g.e.,s.247.

⁴⁸ Chris Brooks, **Introductory Econometrics for Finance**, Cambridge University Press, 2019, s.265.

olmasından kaynaklanabilir. Bu sorunu aza indirmek için doğrudan tahmin etmek istediğimiz değerler (Nakit Dengesi) yerine dönemsel değişim değerlerini (Nakit Değişimini) ele aldık.

Bu çalışmada Makine Öğrenmesi algoritmalarını RMNA modellerinde test edeceğiz. Aşağıda bu çalışmamızda kullanacağımız Makine Öğrenmesi algoritmalarından bahsedeceğiz ve temel çalışma prensiplerini açık hale getireceğiz. Algoritmaların sadece temel çıkış noktaları değil aynı zamanda bu metotlarda kullanılan parametreler de anlatılacaktır.



İKİNCİ BÖLÜM

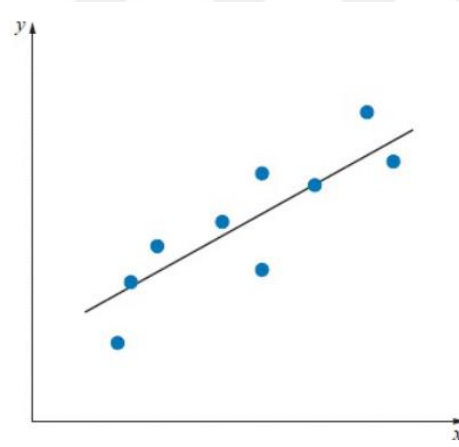
MAKİNE ÖĞRENMESİ ALGORİTMALARI

2.1. Doğrusal Regresyon

Doğrusal regresyon en yaygın tahmin yöntemlerinden birisidir. Basit olarak bağımlı ve bağımsız değişkenlerden oluşan bir model oluşturulur:

$$y_t = \alpha + \beta x_t + u_t$$

Şekil 6: Basit Doğrusal Regresyon



Kaynak: Brooks, Chris, Introductory Econometrics for Finance, 2019, Cambridge University Press.

Çoklu doğrusal regresyonda ise bağımsız değişken sayısı artmaktadır:⁴⁹

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p + e$$

Doğrusal regresyon da Makine Öğrenmesi alanının bir parçası olarak düşünülebilir çünkü yukarıda da bahsedilen kendi kendini düzeltme eğilimine sahiptir.

⁴⁹ Brooks, a.g.e., s.151.

2.2. En Yakın Komşuluk

Hem sınıflandırma tahminlerinde hem de regresyon tahminlerinde kullanılır. Biz modelimizde regresyon versiyonunu kullanacağız. Temel olarak üç aşamadan bahsedebiliriz:

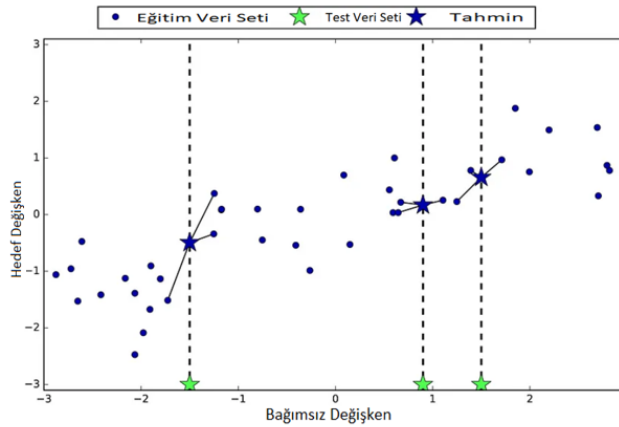
1-Komşuluk sayısına bakılacak nokta sayısı belirlenir. (k)

2-Eğitim setine göre en yakın k tane değer bulunur.

3-En yakın k farklı değerlerin ortalaması alınır.⁵⁰

Örnek olarak tek bir bağımsız değişkeni olan (X) ve bir bağımlı değişkeni olan (Y) bir veri seti düşünelim. En yakın komşuluk parametresini 3 olarak alalım. Bu durumda aşağıdaki gibi bir grafik elde ederiz.

Şekil 7: En Yakın Komşuluk Değeri 3 olan Algoritmanın Tahmini



Kaynak: <https://laptrinhx.com/machine-learning-k-nearest-neighbors-3214844670> [19.08.2023]

Belli bir X değeri için daha önce gerçekleşmiş değerlerden oluşan eğitim veri setini ele alırsak, X değerinin oluştuğu değere en yakın üç noktanın (k=3) ortalama değeri bizim tahmin değerimiz olacaktır. En yakın bu üç değerlerin ortalaması alınır. Mesafe değeri için genellikle öklitsel uzaklık kullanılır:⁵¹

$$\sqrt{\sum_{j=1}^n (x_{1j} - x_{2j})^2}$$

⁵⁰ Bruce, a.g.e.s.238.

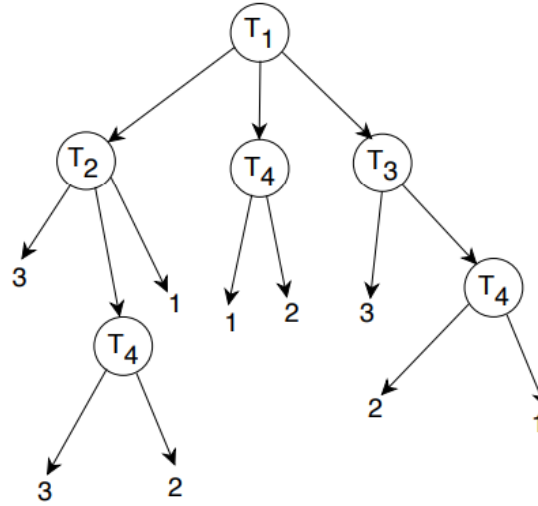
⁵¹ Nils J. Nilsson, **Introduction to Machine Learning**, Stanford University, 1998, s.71.

Parametre değeri olarak sadece k değerine (en yakın komşuluk sayısına) bakılmaktadır.

2.3. Karar Ağaçları

Leo Breiman ve diğerleri tarafından ilk kez 1984'te geliştirilen ve farklı yöntemlerle (Gini ya da Entropy), uygulanan karar ağaçları algoritması farklı bağımsız değişkenleri ağaç şeklinde bir tahmin modeline dönüştürür. Bu yöntem hem sınıflandırma kararlarında hem de regresyon tahminlerinde kullanılır. Sınıflandırma modellerine örnek olarak hile denetimi uygulamaları ya da banka kredi başvuru kabul süreçleri verilebilir.

Şekil 8: Karar Ağacı Örneği



Kaynak: Nilsson, Nils, Introduction to Machine Learning, Stanford University, 1998.

Karar ağaçları modeli “impurity” (kirlilik / belirsizlik) ölçütüne göre oluşturulur. Impurity hesaplaması için Entropy ya da Gini metodu yaygın olarak kullanılır⁵². Algoritma en düşük belirsizlik miktarına göre bağımsız değişkenleri sınıflandırmaya tabi tutar ve bu değerdeki artışın olup olmamasına göre yeni bir dal (sınıflandırma) ekler ya da bulunduğu noktada durur.

Algoritmanın belli bir noktada durması ya da dal sayıları gibi parametreler tahmin performansı açısından önemlidir. Çok sayıda bağımsız değişkene sahip olan bir veri setinde ağaç çok fazla büyüyebilir ve aşırı detaylı olabilir. Bu da yukarıda bahsettiğimiz aşırı öğrenme (overfitting)

⁵² Bruce, a.g.e, s.254.

problemini ortaya çıkarır. Bu ve benzeri sorunları aşabilmek için ve daha doğru bir tahmin modeline ulaşılabilmesi için genel olarak aşağıdaki parametreler önemlidir:

- Maksimum Derinlik (Python scikit-learn: max_depth): Ağacın maksimum derinliği ayarlanır. Çok fazla dal bölünmeleri önlenirse aşırı öğrenme probleminin önüne geçilebilir.
- Maksimum özellik sayısı (Python scikit-learn: max_features): Ağaçlarda kullanılacak bağımsız değişken sayısı kontrol edilir. Genel olarak toplam özellik sayısının karekökü olarak hesaplanır.
- Rastgele-durum seçimi (Python scikit-learn: random_state)
- Minimum Örneklem Sayısı (Python Scikit-learn: minimum sample split): Bir düğümde kullanılan en az örneklem sayısını ifade eder. Eğer elimizde kalan örneklem sayısı bu sayıdan az ise daha fazla bölünmeye izin verilmez.
- Bitim Noktası Örneklem Sayısı (Python scikit-learn: Minimum_samples_leaf): Bir bitim noktasında kullanılan minimum örneklem sayısını ifade eder. Eğer elimizde kalan örneklem sayısı bu sayıdan az ise daha fazla bitim noktasına izin verilmez.⁵³

Son olarak, Karar Ağaçları algoritmasının önemli bir avantajından bahsetmemiz gerekir. Birçok algoritma (doğrusal regresyon da dahil) bir “kara kutu” olma özelliği taşır. Yani sadece parametreleri kontrol edebiliriz ve sonuçları değerlendiririz. Ancak hangi değişkenlerin daha önemli olduğu ya da hangi değişkenlerin birbirini nasıl etkilediği hakkında sezgisel ve açık bir fikir edinmemiz çoğunlukla zordur. Bu durum, algoritmanın kontrolünü zorlaştırır. Ancak Karar Ağaçları algoritmasında, işlem sonucunda bir ağaç şekline ulaşabilmek uzman olmayan kişilere bile daha görsel ve anlaşılır sonuçlar sunabilmektedir. Bu da yaygın olarak kullanılmasını kolaylaştırmaktadır.⁵⁴

2.4. Bagging Metodu ve Rastgele Ormanlar

Rastgele Ormanlar algoritması “Toplu Öğrenme” (ensemble learning) algoritmalarından birisidir. Fikir olarak “The Wisdom of the Crowds” (kalabalıkların bilgeliği) fikrine atfedilebilir.⁵⁵ Bir diğer deyişle, çok daha basit tahmin yöntemlerinden birçoğunu bir araya getirdiğimizde, bu küçük

⁵³ DecisionTreeRegressor, <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>, 21 Mayıs 2023.

⁵⁴ Bruce, a.g.e.,s.258.

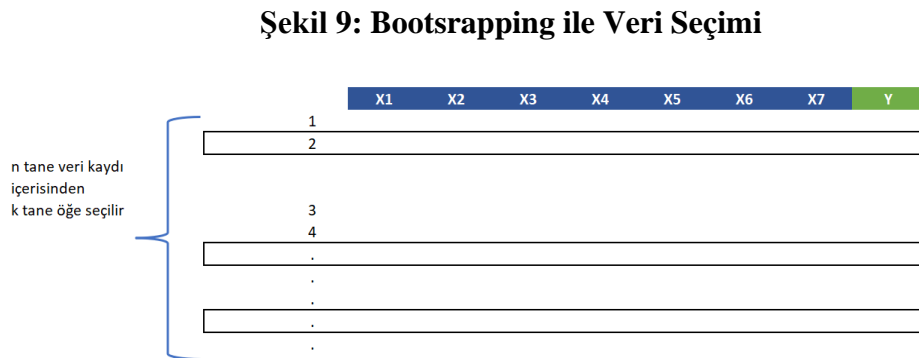
⁵⁵ Bruce, A.e.,s.259.

modeller tekil olarak kompleks ve başarılı tahminler üretmeseler de toplu olarak beklenen potansiyellerinden çok daha iyi sonuçlar üretebilirler. Rastgele Ormanlar'daki temel fikir de budur: Tek bir büyük karar ağacı kullanmak yerine, rastgele küçük ağaçlar seçilir ve bunların ortalaması alınır.⁵⁶ Böylece, büyük bir karar ağacındaki aşırı öğrenme problemini azaltır çünkü tek bir model kullanmak yerine çok fazla küçük ağacı bir arada kullanır. Bu da daha esnek bir modele ulaşmamızı sağlar. Algoritmada Bootstrapping / Bagging metodu kullanılır.

Bootstrapping bir örneklem alma metodudur. Merkezi Limit Teoremine bir alternatif olarak düşünülebilir.⁵⁷ Bir büyük popülasyondan çeşitli örneklem almak yerine, bir örneklem üzerinden daha büyük bir popülasyona ulaşılabilir. Bunun için, küçük bir örneklem yerine koymak şartıyla belli büyüklükte ve sonsuz sayıda örneklem alınarak sanal bir popülasyon elde edilir. Bu yöntemde, örneklemdeki bir değeri birden çok kez kullanabiliriz çünkü "yerine geri koymak şartıyla" bu metot uygulanır.

Bagging, Bootstrapping yönteminin bir versiyonudur. Toplu öğrenme algoritmalarının Rastgele Ormanlar metodundaki uygulamaları bagging örneği olarak düşünülmelidir. Rastgele Ormanlar'da birden çok ağaç seçilir ve bu ağaçların ortalaması (ya da bir farklı bir şekilde bir araya getirilmesi) olarak uygulanır. Ağaçların seçimi ise Bootstrapping metoduna göre yapılır. Yani her seferinde farklı farklı kayıtlar (veriler) yerine koymak şartıyla tekrar tekrar seçilirler ve rastgele küçük ağaçları meydana getirirler. Bu türden Bootstrapping seçimlerinin bir araya toplanması bagging olarak adlandırılır. Aşağıda Rastgele Ormanlar algoritmasının aşamalarını göreceğiz:

1. N tane veri kaydı içerisinde Bootstrapping ile k tane veri seçilir.

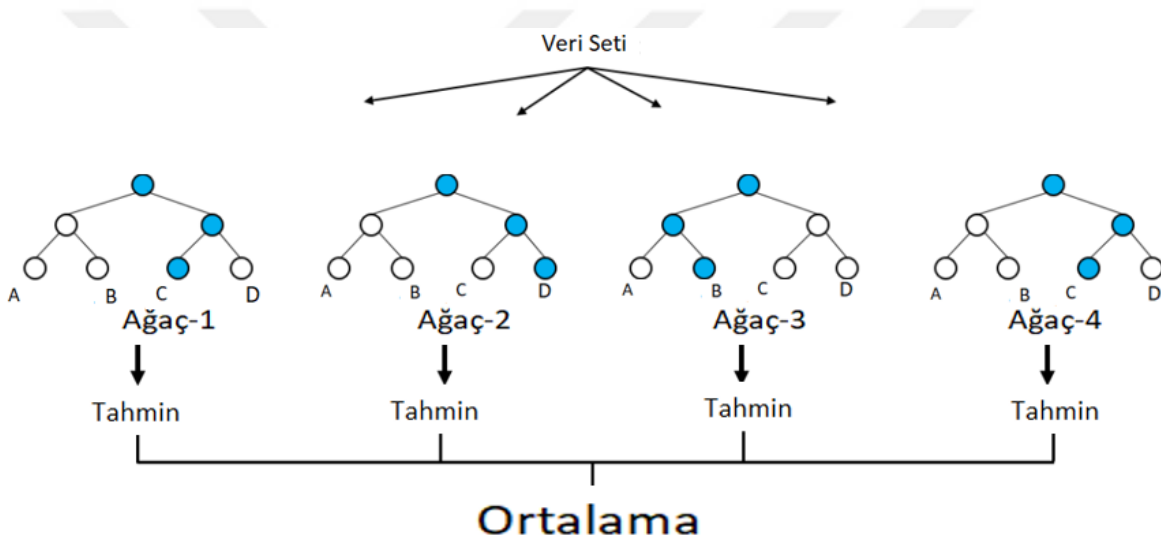


⁵⁶ Bruce, A.e.,s.260.

⁵⁷ Bruce, A.e., s.60.

2. Daha sonra bağımlı değişken ve bağımsız değişkenler arasından bazıları Bootstrapping yöntemine göre seçilir.
3. Seçilen X (bağımsız değişkenler) değerlerine göre küçük ağaç oluşturulur.
4. Bu işlem n kere devam eder ve n tane ağaca ulaşılır.
5. Bu ağaçlarda, Karar Ağaçları algoritmasındaki aşamalar uygulanır.
6. Ağaçlardan elde edilen sonuçlar kaydedilir ve ortalamaları alınır.⁵⁸

Şekil 10: Rastgele Ormanlar Algoritması için Ortalama Değerin Alınması



Kaynak: <https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/>
[19.08.2023]

Algoritma uygulanırken aşağıdaki parametreler yaygın olarak kullanılır ve daha iyi bir tahmine ulaşmak için bu parametrelerde ayarlamalar yapılır. Aşağıda parametre adlarını ve Python programındaki karşılıklarını görmekteyiz:⁵⁹

- Ağaç Sayısı (Python Scikit-learn: `n_estimators`)
- Maksimum Derinlik (Python Scikit-learn: `max_depth`): Kullanılan küçük ağaçların maksimum derinliği ayarlanır. Çok fazla dal bölünmelerini önlersek aşırı öğrenme probleminden kurtulmuş oluruz.

⁵⁸ Bruce, A.e.,s.264.

⁵⁹**Random Forest Regressor**, <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>.20 Mayıs 2023.

- Maksimum özellik sayısı (Python Scikit-learn: max_features): Ağaçlarda kullanılacak bağımsız değişken sayısı kontrol edilir. Genel olarak toplam özellik sayısının karekökü olarak hesaplanır.
- Rastgele-durum seçimi (Python Scikit-learn: random_state)
- Minimum Örneklem Sayısı (Python Scikit-learn: minimum sample split): Bir düğümde kullanılan en az örneklem sayısı. Eğer elimizde kalan örneklem sayısı bu sayıdan az ise daha fazla bölünmeye izin verilmez.
- Bitim Noktası Örneklem Sayısı (Python Scikit-learn: Minimum_samples_leaf): Bir bitim noktasında kullanılan minimum örneklem sayısını ifade eder. Eğer elimizde kalan örneklem sayısı bu sayıdan az ise daha fazla bitim noktasına izin verilmez.
- Bootstrap: “True” olarak ayarlandığında bootstrap yöntemine göre veri setinden (bagging temelli) seçimler yapılır. Aksi takdirde tüm veri seti kullanılır.

2.5. Boosting Temelli modeller

Rastgele Ormanlarda kullanılan Bagging metoduna benzer şekilde Boosting modelleri de ağaç temelli zayıf ve küçük ağaçların bir araya gelmesiyle çalışmaktadır. Ancak Bagging’e göre daha fazla dikkat ve özen gerektirmektedir çünkü kullanılan parametreler çok daha fazladır. Öncelikle, Boosting modellerinde de birçok ağaç kullanılır ancak ağaçların boyutu Bagging modellerine göre farklı olabilir. Örneğin, Gradient Boosting algoritmasında daha büyük ağaçlar kullanılırken Adaboost algoritmasında Bagging yöntemine kıyasla daha küçük ağaçlar kullanılır. İkinci olarak, Bagging temelli modellerde ağaçlar eş zamanlı olarak kullanılır. Yani birçok ağacın tahmininin ortalaması kullanılır. Ancak Boosting modellerinde kullanılan ağaçlar sırasıyla gelir ve sonra gelen ağaç bir önceki tahmin hatasını düzeltir. Böylece Boosting algoritmaları bir önce gelenleri düzelterek öğrenirler ve tahminlerini iyileştirirler.⁶⁰

2.6. Adaboost

Temel olarak aşağıdaki adımlar uygulanır:

1. N tane veri kaydı bir tabloya yerleştirilir ve her bir kayda 1/N oranında ağırlık atanır.
2. Standart hatalar bulunur ve bu hata oranlarını minimize edecek bir model oluşturulur.

⁶⁰ Bruce, a.g.e. s.272.

3. Bu model önceki modele eklenir ve eski model güncellenmiş olur:⁶¹

$$\hat{F}_m = \hat{F}_{m-1} + \alpha_m \hat{f}_m$$

$$\alpha_m = \frac{\log 1 - e_m}{e_m}.$$

Güncellenen modelde ağırlık oranları tekrar hesaplanır; yani 1/N olan ağırlıklar yeni ağırlıklarla güncellenir.

5. Bu süreç birçok kez tekrarlanır ve aşağıdaki gibi bir modele ulaşılır.⁶²

$$\hat{F} = \alpha_1 \hat{f}_1 + \alpha_2 \hat{f}_2 + \dots + \alpha_M \hat{f}_M$$

α değeri (öğrenme oranı) bu algoritmada oldukça kritiktir; çünkü modelin hangi hızda güncelleneceğini belirler.

Önemli parametreler:⁶³

- Öğrenme Oranı (Python Scikit-learn: Learning Rate): Ağırlıkları düzelten yeni ağaç modellerinin düzeltilme oranını ifade eder. Yüksek bir öğrenme oranı sonra gelen modellerin yapacağı katkıyı da arttıracaktır. Ancak daha küçük bir öğrenme oranı yavaş fakat daha dirençli (robust) bir model sağlayacaktır.

-Ağaç Sayısı (Python Scikit-learn: n_estimators)

-Rastgele-durum seçimi (Random_State)

⁶¹ Bruce, A.e.,s.272.

⁶² Bruce, A.e.,s.272.

2.7. Gradient Boosting

Adaboost algoritmasından farklı olarak ağırlıkların yerine hata oranları güncellenir ve düzeltilir. Basit olarak aşağıdaki aşamalar izlenir.⁶⁴

1. Öncelikle sabit bir başlangıç tahmini oluşturulur ve bu sabit modele göre tahmin edilen değerle gerçekleşen değer arasındaki hata tespit edilir.
2. Rastgele bir ağaç oluşturulur ve hata payları bu ağaçtaki bağımsız değişkenlerin konumuna göre dağıtılır. Böylece her bir veri kaydı için yeni bir hata değeri elde edilir
3. Yeni elde edilen hata değerleri α (öğrenme oranı) ile çarpılır ve uygun verilere eklenir.
4. Bu eklemeyen sonra ilk baştaki hata payının (gerçekleşen değer ile ilk tahmin edilen değer arasındaki farkın) azalması beklenir.
5. Hata oranları azalana kadar sırasıyla ağaçlar eklenir.

Temel olarak aşağıdaki parametreler ayarlanır.⁶⁵

-Öğrenme Oranı (Python Scikit-learn: Learning Rate): Ağırlıkları düzelten yeni ağaç modellerinin düzeltilme oranını ifade eder. Yüksek bir öğrenme oranı sonra gelen modellerin yapacağı katkıyı da arttıracaktır. Ancak daha küçük bir öğrenme oranı yavaş fakat daha dirençli (robust) bir model oluşturulmasına katkı sağlayacaktır.

- Ağaç Sayısı (Python Scikit-learn: n_estimators)

- Rastgele-durum seçimi (Python Scikit-learn: random_state)

- Maksimum Derinlik (Python Scikit-learn: max_depth): Kullanılan küçük ağaçların maksimum derinliği ayarlanır. Çok fazla dala bölünmeleri önlersek aşırı öğrenme probleminden kurtulmuş oluruz.

- Maksimum özellik sayısı (Python Scikit-learn: max_features): Ağaçlarda kullanılacak bağımsız değişken sayısı kontrol edilir. Genel olarak toplam özellik sayısının karekökü olarak hesaplanır.

-Rastgele-durum seçimi (random_state)

⁶⁴ Bruce, A.e.,s.272.

⁶⁵ **Gradient Boosting Regressor**, https://scikit-learn.org/stable/auto_examples/ensemble/plot_gradient_boosting_regression.html, 23Mayıs 2023

- Minimum Örneklem Sayısı (Python Scikit-learn: minimum sample split): Bir düğümde kullanılan en az örneklem sayısı. Eğer elimizde kalan örneklem sayısı bu sayıdan az ise daha fazla bölünmeye izin verilmez.

-Bitim Noktası Örneklem Sayısı (Python Scikit-learn: Minimum_samples_leaf): Bir bitim noktasında kullanılan minimum örneklem sayısını ifade eder. Eğer elimizde kalan örneklem sayısı bu sayıdan az ise daha fazla bitim noktasına izin verilmez.

-Alt Örneklem (Python Scikit-learn: Subsample): İşlem esnasında kullanılacak örneklem büyüklüğünü belirler. Bir başka deyişle, her bir aşamada kullanılacak satır sayısını belirler. Çok fazla veri kullanmak aşırı öğrenme problemine sebep olacağı için örneklem büyüklüğü bu parametre sayesinde ayarlanır. 0 ile 1 arasında bir değer alır. Yüksekse varyans (model kompleksliği ya da aşırı öğrenme) artar düşükse yanlılık (bias) artmaktadır.

2.8. XGBoost Algoritması

XGBoost algoritması boosting temelinde geliştirilmiş bir algoritmadır. Gradient Boosting algoritmasını daha iyi bir donanım performansına göre uygular. Diğer algoritmalara kıyasla çok daha fazla sayıda parametre kontrol edilir ve paralel olarak birçok işlem aynı anda uygulanır. Genel olarak Gradient Boosting algoritmasına benzer şekilde çalıştığı söylenebilir. Yani birçok ağaç ya da fonksiyon sırasıyla yaratılır ve her biri bir öncekinde yapılan hataları düzelterek ilerler. Gradient Boosting Algoritmasına kıyasla bazı farklılıklar ve onunla bazı benzerlikler içermektedir.⁶⁶

Öncelikle XGBoost algoritmasında Düzenleştirme (Regularization) tekniği kullanılırken diğerlerinde bu tekniğe rastlamayız. Düzenleştirme aşırı öğrenmeyi zorlaştıran bir tekniktir. Hata fonksiyonuna normal hata oranından daha fazla eklemeler yapılır ve hata olduğundan daha fazla gösterilir. Bu, özellikle Lasso ve Ridge fonksiyonlarında kullanılan bir tekniktir. Hatayı biraz daha büyütürsek algoritma daha fazla öğrenmeye ağırlık vermeyecek ve bir noktada duracaktır. Bir başka deyişle, XGBoost algoritmasında hata oranına ceza puanı eklenerek bu durum kontrol edilebilmekte ve aşırı öğrenmeye karşı önlem alınabilmektedir.⁶⁷ XGBoost diğer algoritmalara göre daha esnektir. Kullanıcı, birçok parametreyi kendisine uygun olarak değiştirebilir. Üçüncü olarak, birçok veri analizi uygulamasındaki en büyük sorun bazı verilerin eksik olmasıdır. Bu

⁶⁶ Bruce, A.e., 274

⁶⁷ Bruce, A.e. 274.

verilerin nasıl ve hangi yöntemle doldurulacağı ve algoritma başarısını etkileyebilecek düzeyde bu eksikliğin etki etmemesini sağlamak başlı başına önemlidir. XGBoost algoritması ise eksik verilere nasıl davranılması gerektiğini de öğrenebilmektedir. Çapraz doğrulama yöntemini her aşamada uygulayabilmektedir. Çapraz doğrulama diğer algoritmalarda kullanıcıya bağlı olarak sadece belli parametreler seçilip onlara uygulanabilirken, XGBoost algoritmasında her aşamada ve çok fazla değişkene uygulanabilir. Son olarak, XGBoost bilgisayarların donanımsal özelliklerine göre de optimize edilmiştir. Diğer algoritmalarda bir işlem yapılırken tüm işlemci çekirdekleri o işlemi bitirmek için çalışır ve daha sonra da diğer işleme geçerler. Ancak XGBoost çalışırken bir işlemde kullanılacak işlemci çekirdeği sayısı da ayarlanabilmektedir. Böylece paralel hesaplamalar yapılabilen ve belli bir sürede yapılacak işlem sayısı ve performansı artmaktadır.⁶⁸

En yaygın kullanılan parametreler aşağıdaki gibidir:⁶⁹

-Booster: sırasıyla yaratılan tahmin modellerinin ağaç temelli mi olduğu ya da doğrusal (linear) bir fonksiyon olup olmadığını belirler.

Gbtree: Ağaç temelli modeller oluşturulur.

Gblinear: Doğrusal fonksiyon temelli modeller oluşturulur.

-Nthread: Kullanılacak işlemci çekirdeği sayısı ayarlanır.

-Lambda: Boosting yönteminde ağırlık optimizasyonu yaparken L2 (Ridge) düzenleştirmesi uygulamaktadır.

-Alpha: Boosting yönteminde ağırlık optimizasyonu yaparken L1 (Lasso) düzenleştirmesi uygulamaktadır.

- Maksimum Derinlik (XGBoost Library: max_depth): Kullanılan küçük ağaçların maksimum derinliği ayarlanır. Çok fazla dala bölünmeler önlenirse aşırı öğrenme problemi engellenmiş olur.

⁶⁸ Jain, Aarshay, **Mastering XGBoost Parameter Tuning: A Complete Guide with Python Codes**”<https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>

⁶⁹ **XGBoost**, https://xgboost.readthedocs.io/en/stable/python/python_api.html, 23 Mayıs 2023

- Öğrenme Oranı (Eta / Learning Rate): Ağırlıkları düzelten yeni ağaç modellerinin düzeltilme oranını ifade eder. Yüksek bir öğrenme oranı sonra gelen modellerin yapacağı katkıyı da arttıracaktır. Ancak daha küçük bir öğrenme oranı yavaş fakat daha dirençli (robust) bir model sağlayacaktır. 0 ile 1 arasında bir değer alır.

-Alt Örneklem (Python Scikit-learn: Subsample): İşlem esnasında kullanılacak örneklem büyüklüğünü belirler. Bir başka deyişle, her bir aşamada kullanılacak satır sayısını belirler. Çok fazla veri kullanmak aşırı öğrenme problemine sebep olacağı için örneklem büyüklüğü bu parametre sayesinde ayarlanır. 0 ile 1 arasında bir değer alır. Yüksekse varyans (model kompleksliği ya da aşırı öğrenme) artar düşükse yanlılık (bias) artmaktadır.

- Ağaç Sayısı (Python Scikit-learn: n_estimators)

- Rastgele-durum seçimi (Python Scikit-learn: random_state)

- Maksimum Derinlik (Python Scikit-learn: max_depth): Kullanılan küçük ağaçların maksimum derinliği ayarlanır. Çok fazla dala bölünmelerini önlersek aşırı öğrenme probleminden kurtulmuş oluruz.

- Maksimum özellik sayısı (Python Scikit-learn: max_features): Ağaçlarda kullanılacak bağımsız değişken sayısı kontrol edilir. Genel olarak toplam özellik sayısının karekökü olarak hesaplanır.

-Rastgele-durum seçimi (random_state)

-Eksiklik (Python Scikit-learn: Missing): Eksik verinin olduğu değişken tanımlanıp kontrol edilebilir.⁷⁰

2.9. Diğer Toplu Öğrenme Modelleri

Yukarıdakilere ek olarak bazı algoritmaları aynı anda kullandığımız modeller de mümkündür. Çalışmamızda üç adet toplu öğrenme modeli oluşturduk: Toplu Öğrenme 1 modelinde, boosting temelli üç algoritmayı (Adaboost, GradientBoosting ve XGBoost) kullandık. Bu üç algoritma daha karmaşık işlemlere sahip olduğu için her üçünü birden denemek faydalı olacaktır. Toplu Öğrenme Modeli 2’de ise iki karmaşık algoritmaya (GradientBoosting ve XGBoost) ek olarak Karar Ağaçları algoritmasını kullandık. Toplu Öğrenme 3’te ise görece daha basit iki algoritmayı

⁷⁰ XGBoost, https://xgboost.readthedocs.io/en/stable/python/python_api.html. 23 Mayıs 2023

(Doğrusal Regresyon ve En Yakın Komşuluk) kullandık. Böylece basit metotların birlikte kullanıldıklarında verdikleri sonuçları görebileceğiz.



ÜÇÜNCÜ BÖLÜM

VERİ SETİNİN İNCELENMESİ VE TAHMİN MODELİ

3.1. Veri Setinin İncelenmesi

Veri setimizde BİST 100’de faaliyet gösteren ve ekonomik döngülerden aşırı etkilendiğini ve daha az etkilendiğini düşündüğümüz “Metal Eşya Makine Elektrikli Cihazlar ve Ulaşım Araçları” sektöründen 22 adet, “Gıda, İçecek ve Tütün” sektöründen 18 adet firmanın (en erken dönem 2008 2. Çeyrek olmak üzere) geçmiş dönemlerden 2022 4. Çeyrek dönemine kadar Nakit Akışı verileri üzerinde çalıştık. Tablo 2’de veri setini oluşturan firmalar, firmaların BİST 100’deki kısaltmaları ve her firma için kullanılan tarihsel dönemler gösterilmiştir.

Tablo 2: Firma Kısaltmaları ve Regresyon Analizinde Kullanılan Dönemler

Firma Kısaltma	Firma Adı	İlk Dönem	Son Dönem
ALCAR	ALARKO CARRIER SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
BFREN	BOSCH FREN SİSTEMLERİ SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
BNTAS	BANTAŞ BANDIRMA AMBALAJ SANAYİ TİCARET A.Ş.	2014 4. Çeyrek	2022 4. Çeyrek
DITAS	DİTAŞ DOĞAN YEDEK PARÇA İMALAT VE TEKNİK A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
EGEEN	EGE ENDÜSTRİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
EMKEL	EMEK ELEKTRİK ENDÜSTRİSİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
GEREL	GERSAN ELEKTRİK TİCARET VE SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
IHEVA	İHLAS EV ALETLERİ İMALAT SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
JANTS	JANTSA JANT SANAYİ VE TİCARET A.Ş.	2010 4. Çeyrek	2022 4. Çeyrek
KARSN	KARSAN OTOMOTİV SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
KATMR	KATMERCİLER ARAÇ ÜSTÜ EKİPMAN SANAYİ VE TİCARET A.Ş.	2010 2. Çeyrek	2022 4. Çeyrek
KLMSN	KLİMASAN KLİMA SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
MAKTK	MAKİNA TAKIM ENDÜSTRİSİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
OTKAR	OTOKAR OTOMOTİV VE SAVUNMA SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
PRKAB	TÜRK PRYSMIAN KABLO VE SİSTEMLERİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
SAYAS	SAY YENİLENEBİLİR ENERJİ EKİPMANLARI SANAYİ VE TİCARET A.Ş.	2013 2. Çeyrek	2022 4. Çeyrek
SILVR	SİLVRLİNE ENDÜSTRİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
TMSN	TÜMOSAN MOTOR VE TRAKTÖR SANAYİ A.Ş.	2012 2. Çeyrek	2022 4. Çeyrek
TOASO	TOFAŞ TÜRK OTOMOBİL FABRİKASI A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
TTRAK	TÜRK TRAKTÖR VE ZİRAAT MAKİNELERİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
VESBE	VESTEL BEYAZ EŞYA SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
VESTL	VESTEL ELEKTRONİK SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
AVOD	A.V.O.D. KURUTULMUŞ GIDA VE TARIM ÜRÜNLERİ SANAYİ TİCARET A.Ş.	2010 4. Çeyrek	2022 4. Çeyrek
BANVT	BANVİT BANDIRMA VİTAMİNLİ YEM SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
EKİZ	EKİZ KİMYA SANAYİ VE TİCARET A.Ş.	2009 3. Çeyrek	2022 4. Çeyrek
ERSU	ERSU MEYVE VE GIDA SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
FRIGO	FRİGO-PAK GIDA MADDELERİ SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
KENT	KENT GIDA MADDELERİ SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
KERTV	KEREVİTAŞ GIDA SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
KNFRT	KONFRUT GIDA SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
KRSTL	KRİSTAL KOLA VE MEŞRUBAT SANAYİ TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
PETUN	PINAR ENTEGRE ET VE UN SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
PINSU	PINAR SU VE İÇECEK SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
PNSUT	PINAR SÜT MAMULLERİ SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
SELGD	SELÇUK GIDA ENDÜSTRİ İHRACAT İTHALAT A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
TATGD	TAT GIDA SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
TBORG	TÜRK TUBORG BİRA VE MALT SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
TUKAS	TUKAŞ GIDA SANAYİ VE TİCARET A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek
OYLUM	OYLUM SİNAİ YATIRIMLAR A.Ş.	2011 2. Çeyrek	2022 4. Çeyrek
DARDL	DARDANEL ÖNENTAŞ GIDA SANAYİ A.Ş.	2008 1. Çeyrek	2022 4. Çeyrek

Veriler bir zaman serisi olarak modele dahil edilmişlerdir. Çeyreklik dönemler bağımsız değişken olarak düşünülmüş; bağımlı değişken olarak firmaların çeyreklik nakit akışı modellere dahil edilmiştir. Veri kaynağı olarak <https://www.isyatirim.com.tr/> web sitesi kullanılmıştır.⁷¹

Özet istatistik değerleri Tablo 3'teki gibidir:

Tablo 3: İşletmeler için Özet İstatistik Değerleri (1.000 USD)

Firm	Mean	Median	STD	Max	Min
ALCAR	55,4	282,2	7.170,1	17.025,1	(20.571,2)
BFREN	(2,3)	108,7	2.241,5	8.504,1	(5.209,7)
BNTAS	298,1	111,5	2.055,2	6.200,1	(3.946,4)
DITAS	(88,4)	(160,0)	985,5	1.851,8	(4.276,5)
EGEEN	908,7	911,6	9.798,1	47.210,2	(26.931,7)
EMKEL	(7,2)	(44,7)	653,7	1.908,9	(1.422,6)
GEREL	22,4	189,4	2.179,0	6.458,2	(8.523,4)
IHEVA	(156,3)	(1,5)	5.722,7	24.599,3	(16.650,5)
JANTS	363,0	185,3	2.861,9	6.602,4	(6.341,4)
KARSN	39,1	270,2	7.665,4	21.352,1	(27.120,4)
KATMR	32,4	8,0	3.570,4	11.698,4	(7.993,1)
KLMSN	110,3	(117,1)	12.564,9	29.549,0	(29.162,2)
MAKTK	(5,4)	0,2	1.040,1	6.162,3	(3.366,4)
OTKAR	1.492,6	540,0	26.821,3	69.863,9	(66.891,7)
PRKAB	213,0	(919,8)	17.237,0	37.696,3	(34.974,6)
SAYAS	74,9	(13,8)	2.277,5	8.850,1	(7.408,7)
SILVR	(16,8)	(10,1)	1.637,1	6.306,9	(6.947,8)
TMSN	495,2	77,7	5.022,4	18.208,6	(12.120,8)
TOASO	5.884,6	19.318,5	110.684,4	268.765,7	(227.694,5)
TTRAK	3.287,6	10.132,4	44.383,7	86.185,9	(125.362,3)
VESBE	(755,9)	(1.372,9)	26.718,6	75.955,1	(86.190,0)
VESTL	(6.373,2)	(1.311,1)	79.493,5	134.297,7	(212.286,8)
AVOD	51,5	10,2	860,2	3.557,5	(1.769,8)
BANVT	261,0	712,1	13.966,9	60.679,8	(64.264,5)
DARKL	(51,6)	8,2	2.811,7	14.292,3	(14.825,5)
EKIZ	(10,2)	0,0	223,0	660,0	(842,0)
ERSU	(12,4)	0,9	243,5	673,1	(925,7)
FRIGO	29,1	(12,2)	1.316,0	5.048,7	(5.800,5)
KENT	151,7	212,8	7.383,4	25.311,3	(34.555,1)
KERV	238,9	(36,4)	22.385,0	115.641,2	(108.728,4)
KNFRT	(58,1)	29,0	1.452,6	5.375,3	(6.508,7)
KRSTL	(26,5)	(62,7)	2.691,8	11.654,7	(10.752,0)
OYLUM	15,7	5,0	560,3	1.866,2	(1.581,2)
PETUN	(25,8)	(185,0)	2.659,0	6.817,8	(5.494,5)
PINSU	(18,9)	(120,1)	1.056,2	4.352,9	(2.883,0)
PNSUT	17,5	(102,5)	2.599,9	11.004,2	(7.048,7)
SELGD	(0,4)	(5,3)	169,8	667,3	(477,3)
TATGD	258,1	(143,6)	10.356,3	37.300,3	(33.414,2)
TBORG	1.912,8	1.975,2	22.476,6	61.532,2	(65.891,3)
TUKAS	(238,8)	(0,0)	5.681,4	27.522,7	(26.700,2)

⁷¹ İş Yatırım, <https://www.isyatirim.com.tr/tr-tr/Sayfalar/default.aspx> , 23 Mayıs 2023

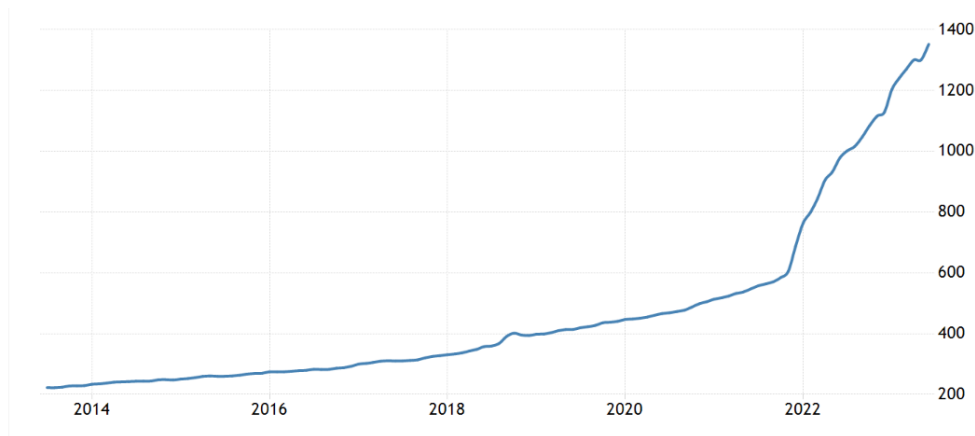
Her firmanın verilerini genel eğilim bakımından incelediğimizde, Türkiye’de 2021 yılı son çeyreğinden itibaren TL’nin yabancı kurlar karşısında değerinin düşmesi ve yüksek enflasyon sebebiyle, firmaların nakit akış değişimleri nominal bazda artmıştır. Tablo 4’te örnekleminizde yer alan tüm firmaların dönemsel toplam nakit toplamının enflasyon ve kur artışı ile korelasyonu görülmektedir.

Tablo 4: Nakit Dengesi – USD/TL Kuru—TÜFE Korelasyonu

	Nakit Dengesi	Dolar / TL Kuru	TÜFE
Nakit Dengesi	1,00	0,96	0,88
Dolar / TL Kuru	0,96	1,00	0,93
TÜFE	0,88	0,93	1,00

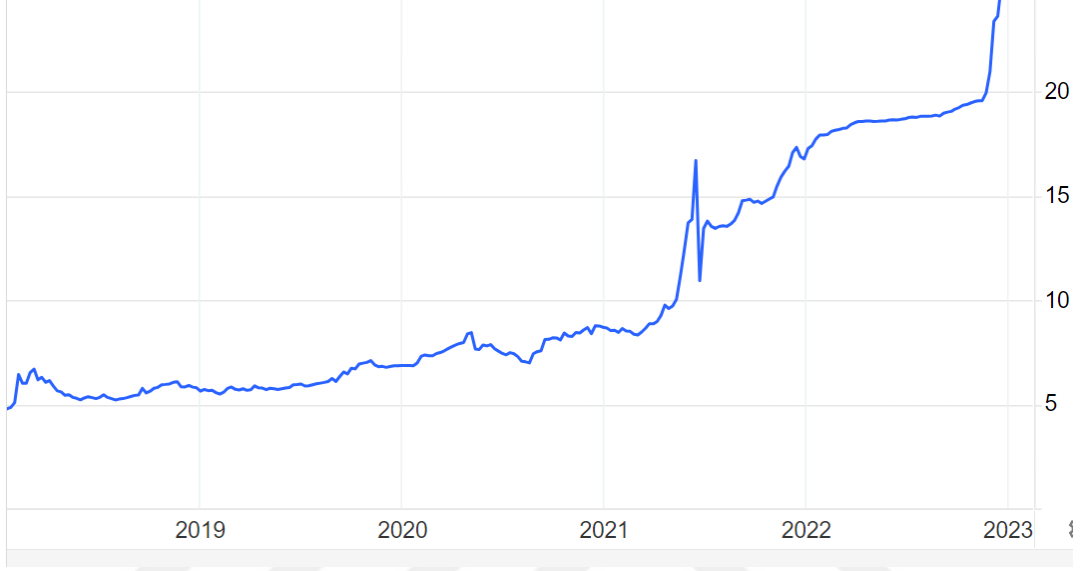
Son dönemleri düşündüğümüzde, nakit akış değişimlerinin dışsal etkenler sebebiyle geçmişteki değişimlere göre daha çok arttığı ve düzensizlikler gösterdiği söylenebilir. Bu durum doğal olarak son dönem tahminlerini zorlaştırmaktadır. Türkiye’deki yüksek enflasyon sebebiyle Bilanço Nakit Dengesi değerleri enflasyon oranı görece daha düşük ve öngörülebilir bir para birimi olan USD para birimine çevrilmiştir. Çevrim için çeyrek dönemler sonundaki kur değerleri kullanılmıştır. Her çeyreklik dönemdeki TL Nakit dengesi ilgili dönemdeki kur değerine bölünmüş ve daha sonra regresyona tabi tutulmuştur.

Şekil 11: TÜFE Oranı (2014-2023)



Kaynak : <https://www.investing.com/economic-calendar/turkish-cpi-871> [19.07.2023]

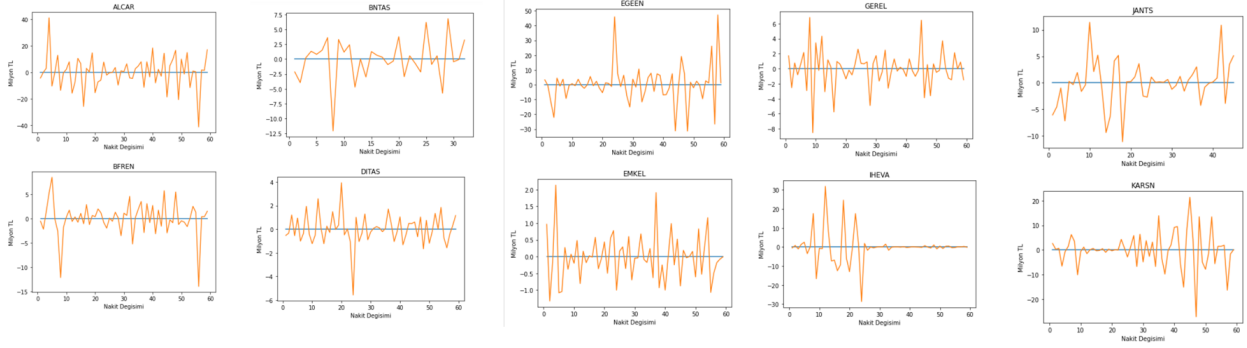
Şekil 12: USD /TL Kuru (2018-2023)

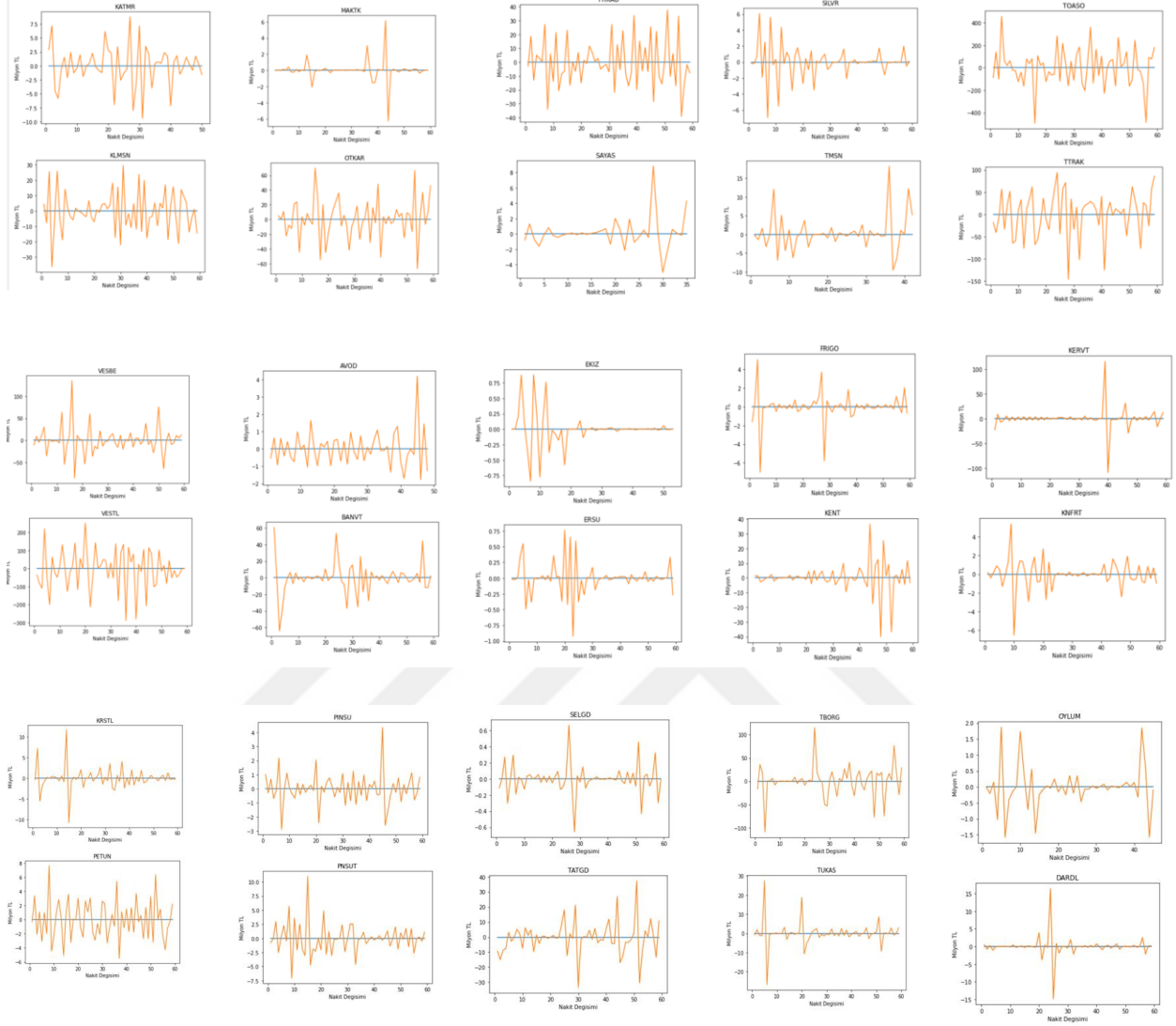


Kaynak : <https://tradingeconomics.com/turkey/currency> [19.07.2023]

Firmaların nakit akış değişimlerinin USD bazında çeyreklik değişimleri aşağıdaki gibidir:

Şekil 13: İşletmelerin Zamansal Nakit Değişimleri





RMNA ile ilgili literatür taraması kısmında bahsedilen çalışmaların bir kısmında nakit akışındaki aşırı dalgalanmaların etkisini kırabilmek için nakit akışı yerine yakın değişkenler kullanılmıştır. Örneğin, nakit akış değerleri yerine FAVÖK değerinin yakın değişken olarak kullanılması mümkündür. Bunun sebebi FAVÖK değerinin daha dirençli (robust) bir değişken olmasıdır. Alternatif olarak Operasyonel Nakit Akış değişimi de iyi bir gösterge olarak kullanılabilir.⁷² Çalışmamızda, yakın değişkenler yerine doğrudan gerçek değişken (nakit akış değerleri) üzerinden modellerimizi oluşturduk.

⁷² Kısakürek a.g.e., s.123.

Veri setinden aykırı değerlerin ayıklanması ve geçmişteki bu değerler üzerinden bazı düzeltmelerin yapılması, bir modelin gerçekçi ve kararlı bir tahminde bulunmasına olanak sağlayacaktır. Dolayısıyla, aykırı değerler veri setinden ayıklanmış; yerine düzleştirme (doğrusal interpolasyon) metoduyla yeni değerler yerleştirilmiştir. Aykırı değerlerin atılması için çeyrekler açıklığı formülünden faydalandık:

$$\text{Çeyrekler Açıklığı} = Q3 - Q1$$

Q3: Veri setindeki Üçüncü Dörtte birlik

Q1: Veri setindeki Birinci Dörtte birlik

Bir değer $Q3+1.5 \times \text{ÇA}$ değerinden büyük ya da $Q1-1.5 \times \text{ÇA}$ değerinden küçük ise aykırı değer olarak kabul edilmiştir.⁷³ Aykırı değerler modelden ayıklandıktan sonra bu değerlere doğrusal interpolasyon uygulanmıştır.

3.2. Tahmin Modeli

Tahmin yönetimimizi aşağıdaki gibi özetleyebiliriz:

1. Her bir firma için dönemsel nakit değişimleri bağımlı değişken ve çeyreklik dönem sıralaması bağımsız değişken olarak modellere dahil edilmişlerdir.
2. Her bir firma için, ilk 30 dönemden son döneme (2022 4. Çeyrek) olan veriler tahmin edilmiştir. Veriler kümülatif olarak regresyon eğitim setinde yer almışlardır. Örneğin; 31. Dönem tahmini için ilk 30 dönem, 32. Dönem tahmini için ilk 31 dönem veya 40. Dönem tahmini için de ilk 39. Dönem eğitim veri seti olarak regresyona tabi tutulmuştur. Bir başka deyişle, sonraki dönemi bilmediğimiz varsayılarak regresyon modelleri tahmin (hedef) dönemine kadar olan dönemler üzerinden eğitilmiş ve bir sonraki dönemlerde test edilmişlerdir.

Bu şekilde ilerleyerek her bir firmanın otuzuncu döneminden itibaren tahminler oluşturulmuş ve gerçekleşen değerlerle tahmin değerleri karşılaştırılmıştır. Bu şekilde düşünüldüğünde, 59 dönemlik verisine sahip olduğumuz bir firma için, 31.dönemden 59.döneme kadar (son 29 çeyrek dönemi) tahmin etmiş oluyoruz.

⁷³ Brooks, a.g.e.,s.98.

3. RMNA modelinin dağılımını ortaya koyabilmek için bir de standart hata değeri oluşturmamız gerekiyor. Standart hata değeri hesaplaması için Üssel Ağırlıklandırılmış Hareketli Ortalama üzerinden standart sapma değerini hesapladık. Son dönemlerdeki nakit akış dalgalanmalarını daha gerçekçi biçimde hesaba katabileceğimizi düşünerek bu yöntemi seçtik. Süre (span) parametre değeri olarak eğitim veri seti dönem sayısını kullandık. Üssel ağırlıklandırma formülü şu şekilde ifade edilebilir:⁷⁴

$$\sigma_t^2 = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j (r_{t-j} - \bar{r})^2$$

λ bozulma (decay) oranını ifade eder. Bu oran arttıkça ağırlıklandırma miktarı, ilk dönemlere göre daha fazla olurken, bu parametrenin düşmesi durumunda geçmiş dönemlere kıyasla daha az olacaktır.⁷⁵ Bozulma değeri için RiskMetrics'in volatilité tahmin tavsiyesi günlük işlemler için 0.94'tür.⁷⁶ Ancak bu oran günlük işlemler için tavsiye edilmiştir. Çalışmamızda bozulma parametresi yerine süre parametresi uyarlanarak volatilité hesaplanmıştır. Süre değeri olarak her eğitim veri setindeki öge sayısı baz alınmıştır. Örneğin, 31. Dönem tahmini için eğitim veri setinde 30 adet dönem vardır ve süre değeri 30 alınmıştır. Tahmin dönemlerimiz 30. Dönemden ileriye doğru gittikçe süre parametremizi bir birim arttırdık. Süre ve bozulma oranı arasındaki formülü şu şekilde özetleyebiliriz:

$$\lambda \text{ Bozulma Oranı} = 2/(1+\text{süre})^{77}$$

Bu formüle göre süre değerimizi güncellediğimizde bozulma oranı da güncellenmiş olur. Kullandığımız süre değerleri ve ilgili bozulma oranları Tablo 5'te görülebilir.

⁷⁴ Brooks, A.e. s.505.

⁷⁵ Brooks, A.e.,s.505.

⁷⁶ RiskMetrics – The RiskMetrics 2006 Methodology, J.P. Morgan, 2007.

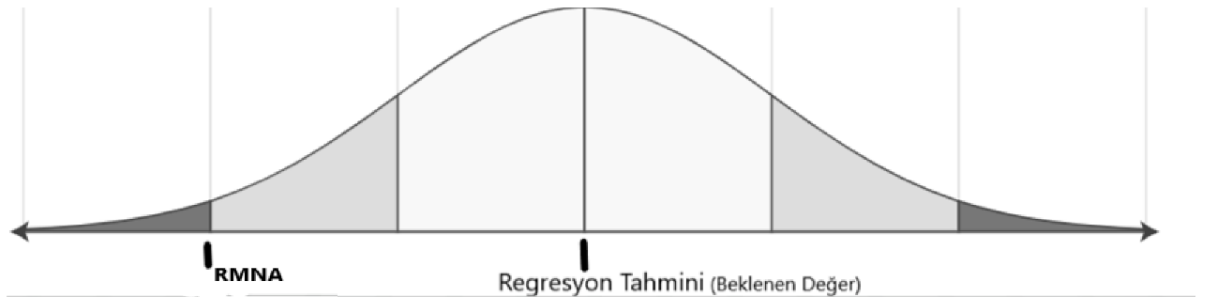
⁷⁷ EWMSTD <https://pandas.pydata.org/pandas-docs/version/0.17.0/generated/pandas.ewmstd.html>,. 19 Temmuz 2023

Tablo 5: Kullanılan Bozulma Parametreler ve ilgili Süre Değerleri

Süre	Bozulma
30	0.06
40	0.05
50	0.04
59	0.03

RMNA değeri de tahmin değerlerimiz gibi firmaların otuzuncu dönemlerinden itibaren kümülatif olarak hesaplanmıştır. RMNA, yukarıda da tanımlandığı üzere nakit akış değişimlerinin beklenen nakit akış değişiminden sapma derecesiyle ilgilidir. RMNA dağılımını ve sapma oranını hesaplayabilmek için Python Numpy kütüphanesini kullanarak 10.000 veriden oluşacak şekilde sanal bir normal dağılım oluşturduk. Ortalama değer yerine regresyon tahminini, standart sapma değeri olarak da madde 3'te bahsedilen üssel olarak ağırlıklandırılmış standart sapma değerini dahil ettik. Sonuç olarak, normal dağılıma göre alt ve üst sınır oluşturduk. Oluşturulan %95 güven aralığındaki alt sınır RMNA değerimizi ifade etmektedir.

Şekil 14: RMNA Dağılımı



Kaynak: <https://wumbo.net/concepts/normal-distribution/>[19.08.2023]

μ : Regresyon Tahmini

σ : Üssel Ağırlıklandırılmış Standart Sapma

4. Modellerimizin tahmin başarısını test etmek için Kk Ortalama Kare Hatası ve Ortalama Mutlak Hata ltlerini kullanacađız. RMNA başarısı iin RMD testlerinde kullanılan Kupiec testine ve Sre testine bařvuracađız.



DÖRDÜNCÜ BÖLÜM

UYGULAMA

4.1.Tahmin Sonuçları ve Değerlendirme

40 farklı firma için Tablo 6’da bahsedilen regresyon algoritmaları kullanılmıştır. Bunlar, yukarıda bahsedilen algoritmaları içerdiği gibi, bunlara ek olarak üç adet toplu öğrenme algoritmasını da kapsamaktadır. Makine Öğrenmesi modelleri oluşturulurken Bölüm 4’te bahsettiğimiz parametre ayarlamaları yapılabilir. Bu çalışmadaki uygulamada ise, pratik bir karşılaştırma ortaya koyabilmek adına, Python programı scikit-learn kütüphanesinde yer alan verili parametreler kullanılmıştır.⁷⁸

Tablo 6’de kullanılan algoritmalar ve ilgili kısaltmalar görülmektedir.

Tablo 6: Regresyon Modelleri ve Kısaltmaları

Kullanılan Metot	Kısaltma
AdaBoost	AB
Gradient Boosting	GB
En Yakın Komşuluk	KNN
Doğrusal Regresyon	LR
Rastgele Ormanlar	RF
XGBoost	XGB
Toplu Öğrenme - 1	T1
Toplu Öğrenme - 2	T2
Toplu Öğrenme - 3	T3
Karar Ağaçları	DC

Tahmin performanslarını değerlendirmek için iki farklı ölçüt kullanacağız : Kök Ortalama Kare Hatası (RMSE) ve Ortalama Mutlak Hata (MAE)⁷⁹

⁷⁸ <https://scikit-learn.org/stable>.

⁷⁹ Weijie Wang, Yanmin Yu, Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model, **IOP Conf. Ser.: Mater. Sci. Eng.**, pp 324, 2018, s.2.

$$MAE = \frac{\sum_{n=1}^N |\hat{r}_n - r_n|}{N} \quad RMSE = \sqrt{\frac{\sum_{n=1}^N (\hat{r}_n - r_n)^2}{N}}$$

r_n : Gerçekleşen Değer, \hat{r}_n : Tahmin Değeri, N : örneklem sayısı olarak alınmıştır.

Her iki ölçüt için de tahmin başarısı gerçekleşen değerlerle tahmin edilen değerlerin arasındaki farkın minimum tutulmasıyla özdeşdir. Tablo 7’de 40 firmayı içeren 1038 dönemlik tahmin sonuçlarıyla 1038 adet gerçekleşen değer arasındaki farklardan yola çıkarak hem Kök Ortalama Kare Hatası değerleri hem de Ortalama Mutlak Hata Değerleri hesaplanmıştır.

Tablo 7: Kök Ortalama Kare Hatası (KOKH) ve Ortalama Mutlak Hata (OMH) Değerleri (Milyon)

Type	OMH	KOKH
LR	9,6	27,6
DC	13,8	39,2
KNN	10,3	28,9
RF	12,0	33,7
AB	12,7	36,1
GB	13,5	38,2
XGB	13,8	39,1
T1	13,1	37,3
T2	13,7	38,8
T3	9,9	28,0

Kök Ortalama Kare Hatası sonuçlarına göre, en başarılı tahmin performansına sahip (en az hataya sahip) üç modelin (27,6 hata değeri ile) Doğrusal Regresyon, (28,0 hata değeri ile) Toplu Öğrenme-3 ve (28,9 hata değeri ile) En Yakın Komşuluk modelleri olduğu söylenebilir. Diğer Makine Öğrenmesi temelli modellerin toplam hata değeri daha fazladır. Ortalama Mutlak Hata sonuçlarına göre ise, en başarılı tahmin performansına sahip (en az hataya sahip) üç modelin (9,6 hata değeri ile) Doğrusal Regresyon, (9,9 hata değeri ile) Toplu Öğrenme-3 ve (10,3 hata değeri ile) En Yakın Komşuluk modelleri olduğunu söyleyebiliriz. Diğer Makine Öğrenmesi temelli modellerin toplam hata değeri daha fazladır. Böylece her iki hata ölçütüne göre de bu üç model

daha iyi bir tahmin performansı ortaya koymaktadır. Bu bakımdan, Doğrusal Regresyona alternatif oluşturabilecek tek farklı yöntem En Yakın Komşuluk algoritmasıdır. Toplu Öğrenme-3 modelinde ise En Yakın Komşuluk ve Doğrusal Regresyon birlikte kullanılmıştır. Buradan şu sonucu çıkarabiliriz: En Yakın Komşuluk ve Doğrusal Regresyon gibi görece basit yöntemler daha iyi sonuçlar vermektedir.

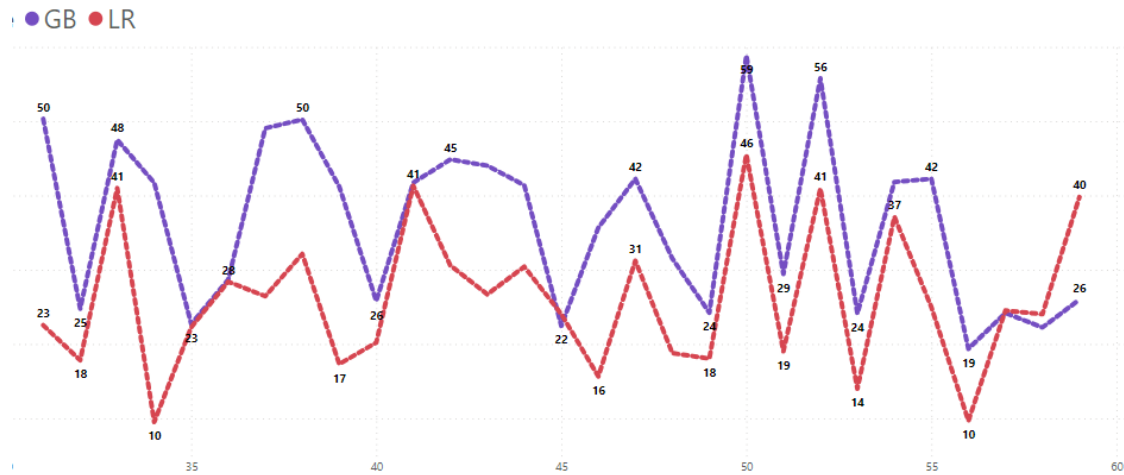
Hata değeri açısından sadece son dönem (2022 4. Çeyrek) tahminleriyle gerçekleşen değerleri Kök Ortalama Hata Karesi üzerinden karşılaştırdığımızda ise Makine Öğrenmesi temelli karmaşık algoritmaların biraz daha iyi bir tahmin performans (daha düşük hata değeri) verdiğini görmekteyiz.

Tablo 8: 2022 4.Çeyrek için Kök Ortalama Kare Hatası Değerleri (Milyon)

AB	DC	GB	KNN	LR	RF	XGB	T1	T2	T3
31,6	22,3	23,3	29,6	35,6	22,6	22,3	25,8	22,6	32,2

Bunun sebebi olarak veri sayısı arttıkça daha karmaşık algoritmaların tahmin başarılarının da artmasını gösterebiliriz. Şekil 15’te Doğrusal Regresyon ve Gradient Boosting modellerinin tüm firmalar için yaptıkları tahminlere göre KOKH skorları görülmektedir. Buna göre son döneme yaklaştıkça (veri sayısı arttıkça) Gradient Boosting’in tahmin performansındaki iyileşme görülmektedir. Doğrusal Regresyon ise geçmiş dönemlere kıyasla daha fazla hata yapma eğilimindedir.

Şekil 15: İki algoritmanın Dönemsel KOKH Skorları



Araştırma sorumuza cevap olarak şunu söyleyebiliriz: Örneklemimize göre, firmalar bazında 30. Dönemden itibaren yaptığımız tahminleri incelediğimizde, geçmiş tahminlerde doğrusal regresyondan farklı olarak Makine Öğrenmesi algoritmalarının kullanılması tahmin performansı açısından bir ilerleme getirmemiştir. Ancak veri sayısı arttıkça, son iki dönem tahmininde, Makine Öğrenmesi algoritmalarının daha iyi sonuçlar (daha düşük hata) verdiğini gözlemlemekteyiz. Ancak bu durumun geçici bir eğilim mi yoksa yapısal ve kalıcı bir eğilim mi olduğunu kesin olarak söyleyebilmek için daha sonraki dönemlerin de verisine ihtiyaç duymaktayız.

4.2.RMNA Sonuçları ve Değerlendirme

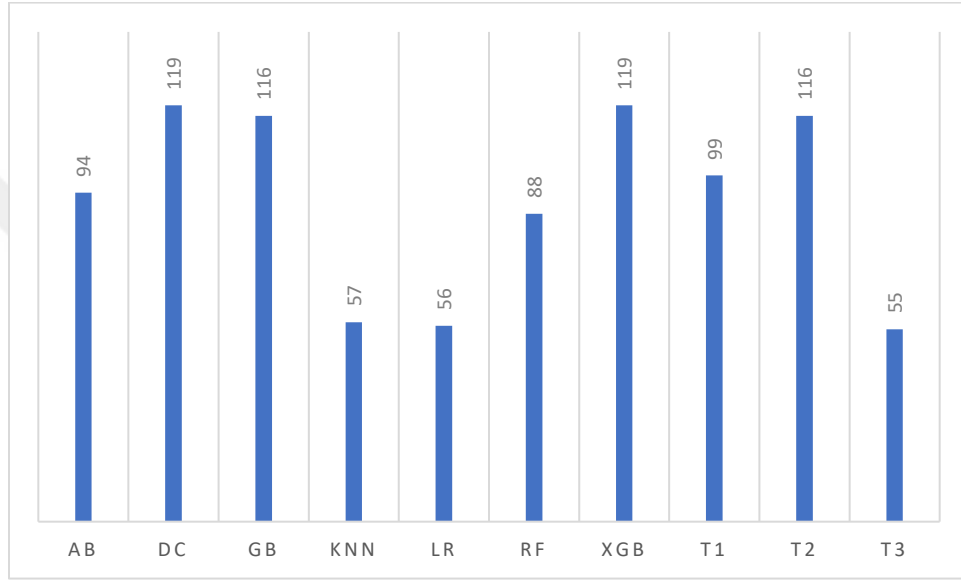
RMNA sonuçlarını değerlendirirken, %95 güven aralığındaki alt sınır değerleri kontrol edilecektir. Bu sınır değeri x ise, %95 ihtimalle x değerinden daha büyük bir nakit riski ile karşılaşmayacağımızı varsayabiliriz. Eğer gerçekleşen çeyreklik nakit akış değişimi RMNA değerinin altında kalırsa, bu durumu RMNA değerinin ihlali (başarısızlığı) olarak düşünebiliriz. Tablo 9’da modellerin ihlal sayıları işletme bazında görülmektedir.

Tablo 9: İşletme bazında Model İhlal Sayıları

	AB	DC	GB	KNN	LR	RF	XGB	T1	T2	T3
ALCAR	0	0	0	0	0	0	0	0	0	0
BFREN	1	1	1	1	1	1	1	1	1	1
BNTAS	0	0	0	0	0	0	0	0	0	0
DITAS	0	2	1	0	0	1	2	1	2	0
EGEEN	6	7	6	5	5	4	6	6	6	5
EMKEL	1	1	1	0	0	0	1	1	1	0
GEREL	1	2	2	0	0	2	2	2	2	0
IHEVA	2	3	3	2	2	2	3	2	3	2
JANTS	3	4	4	2	2	2	4	4	4	2
KARSN	4	4	4	3	3	5	4	4	4	3
KATMR	2	2	2	1	0	1	2	2	2	0
KLMSN	6	7	7	1	1	6	7	6	7	1
MAKTK	0	0	0	0	0	0	0	0	0	0
OTKAR	1	2	2	2	1	1	2	1	2	1
PRKAB	3	4	4	0	0	3	4	3	4	0
SAYAS	1	1	1	0	0	0	1	1	1	0
SILVR	0	1	1	0	0	0	1	0	1	0
TMSN	3	3	3	1	1	1	3	3	2	1
TOASO	3	4	3	2	0	2	5	3	3	1
TTRAK	3	3	3	3	2	2	3	2	3	2
VESBE	0	2	2	1	1	1	2	1	2	1
VESTL	3	4	4	1	1	3	4	3	4	1
AVOD	4	5	5	1	1	4	5	5	5	1
BANVT	2	2	2	1	1	2	2	2	2	1
DARDL	3	3	4	2	3	4	3	3	3	2
EKIZ	2	3	3	2	2	2	3	2	3	2
ERSU	4	4	4	3	4	3	4	4	4	4
FRIGO	2	2	2	1	2	2	2	2	2	1
KENT	6	5	6	4	4	5	6	6	6	4
KERVY	2	3	2	1	1	2	3	1	3	1
KNFRT	1	2	2	1	1	1	2	1	2	1
KRSTL	2	2	2	2	2	2	2	2	2	2
OYLUM	2	3	3	1	1	3	3	3	3	1
PETUN	1	2	2	1	1	1	2	1	2	1
PINSU	4	5	5	2	1	4	5	4	5	1
PNSUT	3	4	4	1	1	3	4	3	4	1
SELGD	3	4	4	3	3	4	4	4	4	3
TATGD	5	5	4	2	2	3	4	4	4	2
TBORG	4	4	4	4	5	4	4	4	4	5
TUKAS	1	4	4	0	1	2	4	2	4	1
TOPLAM	94	119	116	57	56	88	119	99	116	55

Şekil 16’da ise modellerin toplam ihlal sayıları grafik biçiminde görülmektedir. Bu verilere göre en az ihlal sayısı (55) Toplu Öğrenme-3 modelinde görülmektedir. Doğrusal Regresyon ve En Yakın Komşuluk modelleri sırasıyla 56 ve 57 ihlal sayılarına sahiptir. Karmaşık Makine Öğrenmesi temelli modeller ise aynı RMNA tahmin başarısını gösterememişlerdir.

Şekil 16: Modeller ve İhlal Sayıları



Her model 40 firma için toplam 1038 farklı dönem için tahmin yapmıştır. Tablo 10’da 1038 dönem için model bazında ihlal oranları görülmektedir:

Tablo 10: Model İhlal Oranları

	AB	DC	GB	KNN	LR	RF	XGB	ENS1	ENS2	ENS3
İhlal Sayısı Toplamı	94	119	116	57	56	88	119	99	116	55
Toplam Dönem Sayısı	1038	1038	1038	1038	1038	1038	1038	1038	1038	1038
Toplam İhlal Yüzdesi	9%	11%	11%	5%	5%	8%	11%	10%	11%	5%

Tablo 10’daki sonuçlara göre, en az ihlal oranları Doğrusal Regresyon, En Yakın Komşuluk ve Toplu Öğrenme-3 modellerine aittir.

4.3.Kupiec Testi

Kupiec testi esas olarak Paul Kupiec tarafından RMD tahminlerini test etmek üzere geliştirilmiştir.⁸⁰ Sıfır Hipotezi uygulanan RMD model tahmininin başarılı olduğunu varsayar.

⁸⁰ Kupiec, Paul, **Techniques for Verifying the Accuracy of Risk Measurement Models**, Journal of Derivatives, 1995, pp. 73-84.

Sıfır hipotezi reddedilirse model test sonucunu “başarısız” olarak kabul edebiliriz. 10 adet modelimizin 40 adet firma için bulmuş olduğu sonuçlar Tablo 11’de görülmektedir.

Tablo 11: Kupiec Testi Sonuçları

	AB	DC	GB	KNN	LR	RF	XGB	T1	T2	T3
ALCAR										
BFREN										
BNTAS										
DITAS										
EGEEN	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi		Reddedildi	Reddedildi	Reddedildi	Reddedildi
EMKEL										
GEREL										
IHEVA										
JANTS	Reddedildi	Reddedildi	Reddedildi				Reddedildi	Reddedildi	Reddedildi	
KARSN						Reddedildi				
KATMR										
KLMSN	Reddedildi	Reddedildi	Reddedildi			Reddedildi	Reddedildi	Reddedildi	Reddedildi	
MAKTK										
OTKAR										
PRKAB										
SAYAS										
SILVR										
TMSN	Reddedildi	Reddedildi	Reddedildi				Reddedildi	Reddedildi		
TOASO							Reddedildi			
TTRAK										
VESBE										
VESTL										
AVOD	Reddedildi	Reddedildi	Reddedildi			Reddedildi	Reddedildi	Reddedildi	Reddedildi	
BANVT										
DARDL										
EKIZ										
ERSU										
FRIGO										
KENT	Reddedildi	Reddedildi	Reddedildi			Reddedildi	Reddedildi	Reddedildi	Reddedildi	
KERTV										
KNFRT										
KRSTL										
OYLUM		Reddedildi	Reddedildi			Reddedildi	Reddedildi	Reddedildi	Reddedildi	
PETUN										
PINSU		Reddedildi	Reddedildi				Reddedildi		Reddedildi	
PNSUT										
SELGD										
TATGD	Reddedildi	Reddedildi								
TBORG					Reddedildi					Reddedildi
TUKAS										

Tablo 11’deki sonuçlara göre sıfır hipotezi (model tahmininin başarılı olduğu savı) 40 firmadan 29 tanesinde tüm modeller tarafından geçerli sayılmıştır. Sadece 11 adet firma için bazı modeller testi geçememişlerdir.

Sadece model bazında incelediğimizde ise Tablo 12’deki sonuçlara ulaşmaktayız:

Tablo 12: Model bazında Kupiec Testi Başarı Sayısı

Model	Başarılı	Başarısız
AB	33	7
DC	31	9
T1	33	7
T2	33	7
T3	38	2
GB	32	8
KNN	39	1
LR	38	2
RF	35	5
XGB	31	9

Tabloya göre 40 firma için uygulanan testlerde, Doğrusal Regresyon 38, En Yakın Komşuluk 39 ve Toplu Öğrenme-3 modeli 38 defa Kupiec testini geçmişlerdir. Diğer modellerin test başarı oranı daha düşüktür.

4.4.Süre Testi

Peter Christoffersen tarafından hali hazırda mevcut bulunan RMD risk modeli testlerine alternatif olarak önerilmiştir. Bu testte, RMD modelinin toplam ihlal sayısından ziyade iki ihlal dönemi arasında geçen süreye odaklanılmıştır.⁸¹ Sıfır Hipotezi uygulanan RMD model tahmininin başarılı olduğunu varsayar. Süre testi uyguladığımızda, 10 adet modelimizin 40 adet firma için bulmuş olduğu sonuçlar Tablo 13'te görülmektedir.

⁸¹ Christoffersen, Peter and Pelletier, Denis, Backtesting Value-at-Risk: A Duration-Based Approach, Montreal, 2003.

Tablo 13: Süre Testi Sonuçları

	AB	DC	GB	KNN	LR	RF	XGB	T1	T2	T3
ALCAR										
BFREN										
BNTAS										
DITAS										
EGEEN	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi
EMKEL										
GEREL		Reddedildi	Reddedildi				Reddedildi	Reddedildi	Reddedildi	
IHEVA										
JANTS				Reddedildi	Reddedildi	Reddedildi				Reddedildi
KARSN				Reddedildi	Reddedildi					Reddedildi
KATMR	Reddedildi	Reddedildi	Reddedildi				Reddedildi	Reddedildi	Reddedildi	
KLMSN		Reddedildi	Reddedildi			Reddedildi	Reddedildi		Reddedildi	
MAKTK										
OTKAR				Reddedildi						
PRKAB										
SAYAS										
SILVR										
TMSN			Reddedildi							
TOASO				Reddedildi		Reddedildi				
TTRAK	Reddedildi			Reddedildi		Reddedildi		Reddedildi		Reddedildi
VESBE										
VESTL										
AVOD										
BANVT										
DARDL				Reddedildi						
EKIZ	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi
ERSU										
FRIGO		Reddedildi	Reddedildi		Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	
KENT										
KERVT										
KNFRT		Reddedildi	Reddedildi				Reddedildi		Reddedildi	
KRSTL										
OYLUM	Reddedildi					Reddedildi				
PETUN		Reddedildi	Reddedildi				Reddedildi		Reddedildi	
PINSU				Reddedildi						
PNSUT	Reddedildi									
SELGD										
TATGD										
TBORG	Reddedildi	Reddedildi	Reddedildi		Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi	Reddedildi
TUKAS		Reddedildi	Reddedildi			Reddedildi	Reddedildi	Reddedildi	Reddedildi	

Bu sonuçlara göre 40 firmadan 20 tanesi için tüm modeller süre testinde başarılı olmuş, ancak diğer 20 firmada en az bir model testi geçememiştir.

Sadece model bazında incelediğimizde ise Tablo 14'teki sonuçlara ulaşıyoruz:

Tablo 14: Model bazında Süre Testi Başarı Sayısı

Model	Başarılı	Başarısız
AB	33	7
DC	30	10
T1	32	8
T2	30	10
T3	34	6
GB	29	11
KNN	31	9
LR	34	6
RF	30	10
XGB	30	10

Tabloya göre 40 firma için uygulanan testlerde, Doğrusal Regresyon ve Toplu Öğrenme-3 modelleri 34 kez, Adaboost 33 kez ve Toplu Öğrenme-1 modeli ise 32 kez süre testini geçmişlerdir. Diğer modellerin test başarı oranı daha düşüktür ancak tabloda da görüldüğü üzere model başarı sayıları birbirine yakın seyretmektedir.

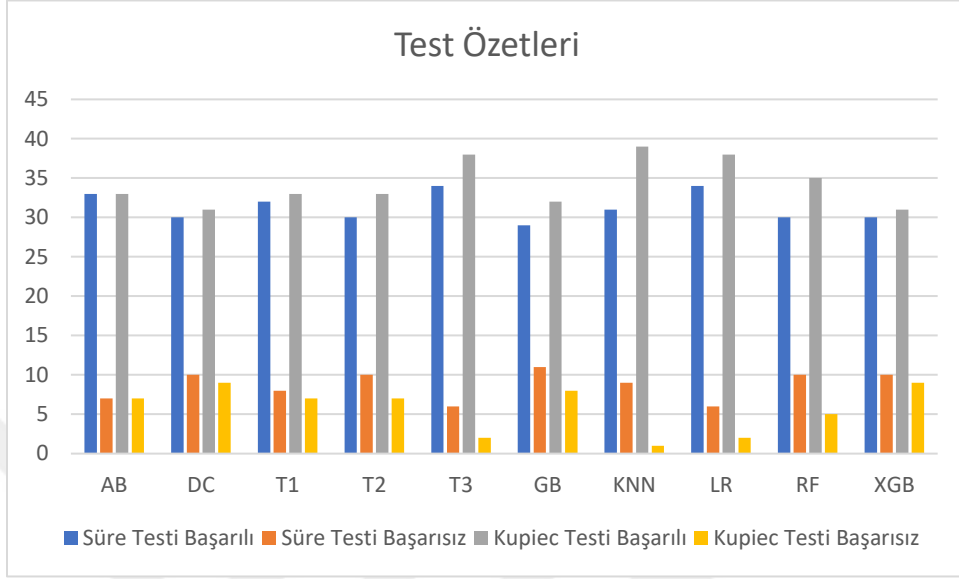
4.5. Test Sonuçları Özeti ve Değerlendirme

Her iki test sonucunun karşılaştırılması Tablo 15 ve Şekil 17’de görülmektedir:

Tablo 15: Kupiec Testi ve Süre Testi Özeti

Model	Süre Testi		Kupiec Testi	
	Başarılı	Başarısız	Başarılı	Başarısız
AB	33	7	33	7
DC	30	10	31	9
T1	32	8	33	7
T2	30	10	33	7
T3	34	6	38	2
GB	29	11	32	8
KNN	31	9	39	1
LR	34	6	38	2
RF	30	10	35	5
XGB	30	10	31	9
Toplam	313	87	343	57

Şekil 17: Kupiec Testi ve Süre Testi Özeti



Her iki test sonucunu karşılaştırdığımızda Süre Testindeki başarılı model sayısı Kupiec testine göre daha düşüktür. Kupiec testinden geçen her modelin Süre testinde başarılı olamadığını düşünürsek Süre Testinin daha hassas olduğu söylenebilir. Süre Testi sonuçlarında Doğrusal Regresyonun başarı oranı düşmüş ve Makine Öğrenmesi temelli modellere yakın seyretmiştir. Dolayısıyla Süre Testi sonuçlarına göre, Makine Öğrenmesi temelli modeller ile Doğrusal Regresyon temelli modeller birbirine yakın derecede başarı oranı üretmişlerdir. Kupiec testi daha çok model ihlal sayılarıyla ilgiliyken, Süre testi ihlal dönemleri arasında geçen süre ile ilgilidir. Bu yüzden, Süre testinde ihlal sayılarından daha bağımsız sonuçlara ulaşmaktayız. Örneğin ihlal sayıları ve Kupiec testi baz alındığında, en başarılı modellerden birisi En Yakın Komşuluk temelli modellerken, Süre testine baktığımızda bu tür modeller aynı ölçüde başarılı görünmemektedir.

4.6.Uygulama Sonuçları ve Değerlendirme

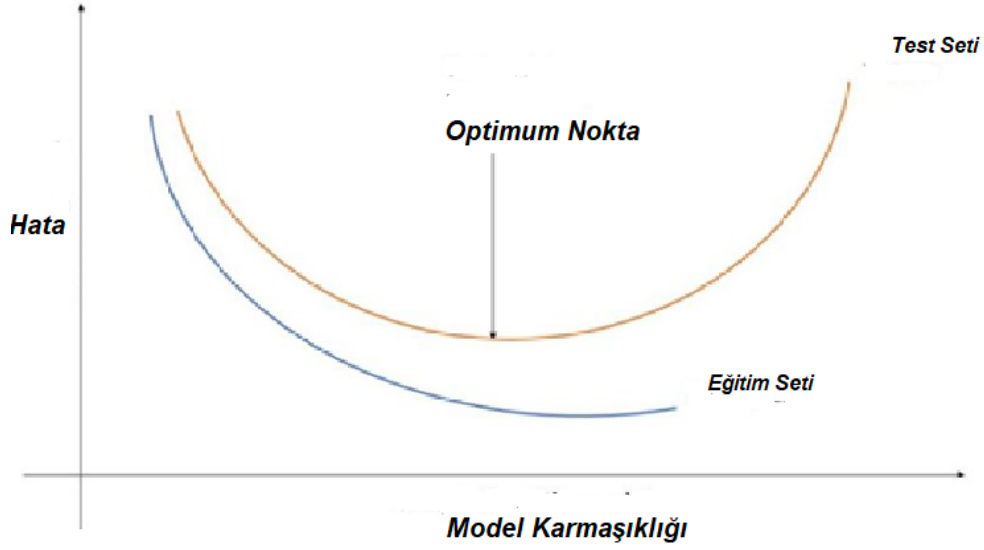
RMNA uygulamalarında RMD modellerine benzer şekilde olası değişimlerin dağılımı ve bu dağılıma göre bir güven aralığı modeli ortaya konulur. Amaç, gelecek dönemlerdeki olası değişimleri ortaya koyabilmektir. Tez çalışmamızda oluşturduğumuz tahmin modelleri, Doğrusal Regresyonun yanı sıra, Makine Öğrenmesi uygulamaları üzerinden oluşturulmuştur. Bu modeller ekseninde, nakit akış tahminleri açıklanmış ve Doğrusal Regresyondan elde edilen sonuçlarla karşılaştırılmıştır. Bu şekilde görece daha karmaşık modeller oluşturulmuş ve model karmaşıklığı ile tahmin ya da RMNA ölçüm başarısı arasındaki ilişki incelenmiştir.

Farklı ve daha karmaşık yapıdaki yöntemlerin kullanılmasıyla birlikte, varyans duyarlılığı yükseleceği için modellerin verilerdeki dalgalanmaları daha kolay bir şekilde yakalaması beklenecektir. Başka bir deyişle, tahmin performansının ve RMNA performansının birlikte artması beklenebilir. Çalışmamızda bu durumları da göz önünde bulundurarak farklı Makine Öğrenmesi temelli modelleri test ettik. Doğrusal Regresyon, En Yakın Komşuluk, Rastgele Ormanlar, Adaboost, Gradient Boosting, XGBoost ve Karar Ağaçları algoritmalarını kullandık. Bunlara ek olarak üç farklı Toplu Öğrenme modeli kullandık. Toplu Öğrenme 1 modelinde boosting temelli üç algoritmayı (Adaboost, Gradient Boosting ve XGBoost), Toplu Öğrenme Modeli 2’de ise iki karmaşık algoritmaya (Gradient Boosting ve XGBoost) ek olarak Karar Ağaçları algoritmasını kullandık. Toplu Öğrenme 3’te ise görece daha basit iki algoritmayı (Doğrusal Regresyon ve En Yakın Komşuluk) birlikte kullandık. Tüm algoritmaları Python Programı Scikit-learn kütüphanesi verili parametreleriyle uyguladık.

Tahmin performanslarını ölçümlemek için Kök Ortalama Kare Hatası ve Ortalama Mutlak Hata ölçütlerine başvurduk. Yukarıdaki bulgularımıza göre, tahminlerin çoğunda Makine Öğrenmesi temelli modeller doğrusal regresyon temelli modellere göre daha başarılı sonuçlar üretememişlerdir. Son iki dönem için Makine Öğrenmesi temelli modeller tahmin performansı açısından bir ölçüde başarılı görünse de bu durumun kalıcı ve yapısal bir durum olduğunu gözlemleyemedik. Bu durumda tahmin performansına atıfta bulunduğumuz Hipotez 1 reddedilmiştir. İkinci olarak RMNA tahminlerinin performanslarını test edebilmek için ihlal sayılarını kontrol ettik ve RMD modellerinin test edilmesinde kullanılan Kupiec testine ve Süre testine başvurduk. Bu ölçütlere göre, Makine Öğrenmesi temelli karmaşık modeller Doğrusal Regresyon temelli modellere göre daha iyi sonuçlar üretememişlerdir. Dolayısıyla Hipotez 2 reddedilmiştir.

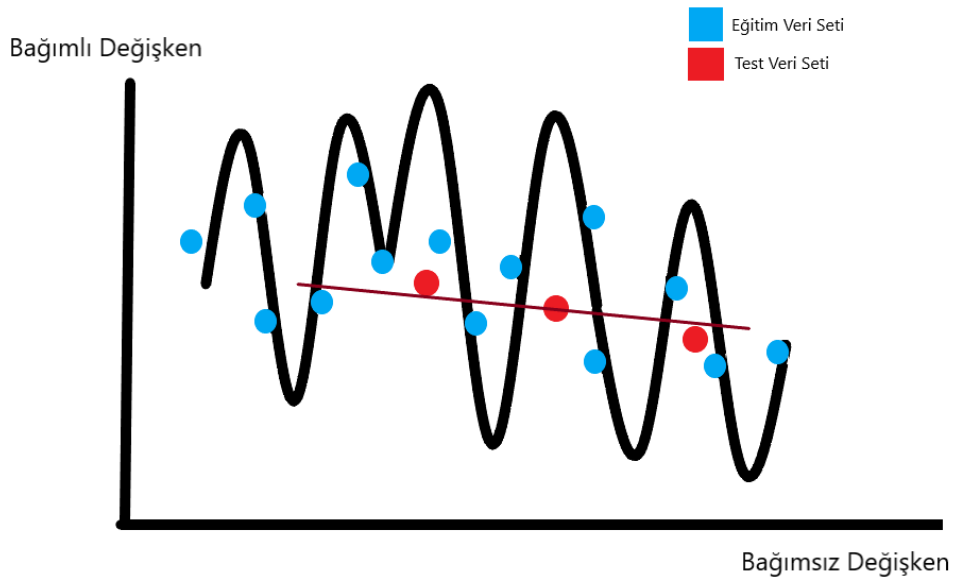
Yukarıda da bahsedildiği üzere, Makine Öğrenmesi temelli modellerde varyans ve yanlılık arasındaki dengenin iyi kurulması gerekmektedir. Karmaşıklık derecesi arttıkça (varyans derecesi arttıkça) modellerin daha iyi tahmin yapması beklenebilir. Ancak karmaşıklık derecesinin artması eğitim veri setindeki hata oranını düşürse bile belli bir noktadan sonra test veri setindeki hata oranını düşürmez; tersine arttırır. Şekil 18’de bu ilişki görülmektedir. Şekilde görüldüğü üzere, belli bir optimum noktadan sonra test veri seti tahmininin hata miktarı artmaktadır.

Şekil 18: Model Karmaşıklığı ve Hata Oranı İlişkisi



Dolayısıyla bir modelde varyans dengesi iyi kurulmalıdır. Gerekliğinde karmaşık modellerin daha iyi sonuçlar üretemeyebileceğini de göz önünde bulundurmamız gerekir. Şekil 19’da bu durumu daha açık bir şekilde görebiliriz:

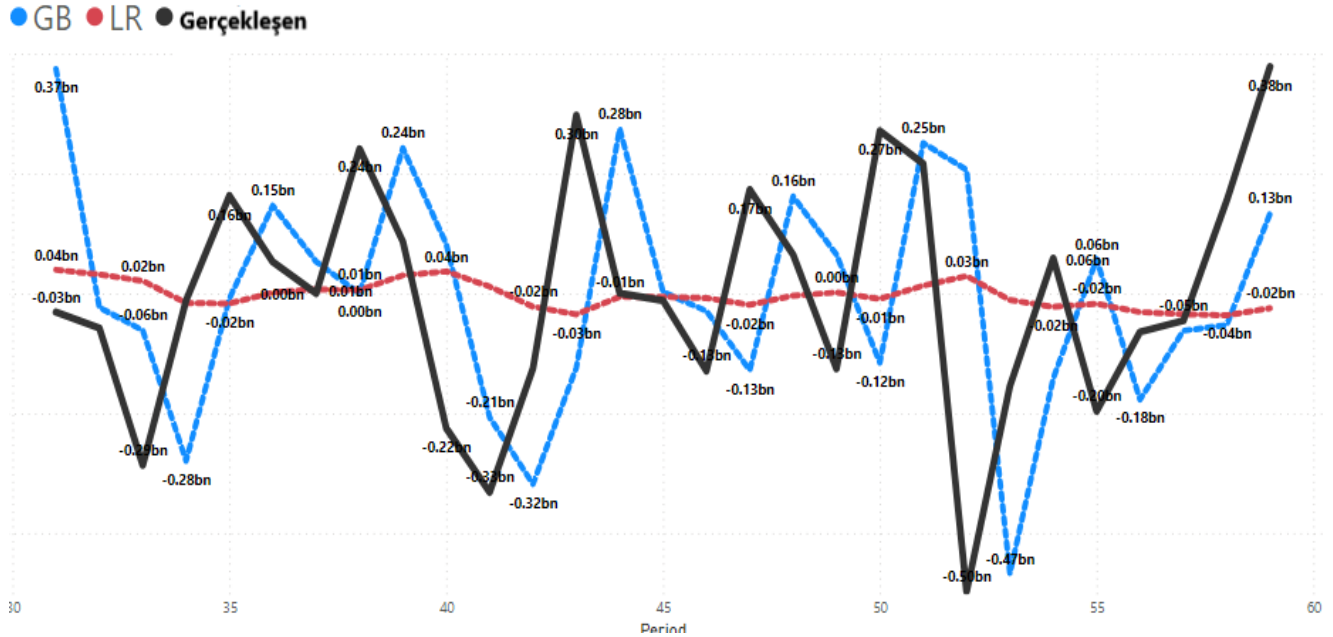
Şekil 19: Varyansı Yüksek Regresyon Regresyon ve Yanlılık Değeri Yüksek Regresyon



Şekil 19’da bir adet basit doğrusal regresyon ve bir adet polinom regresyon görülmektedir. Polinom regresyonunun varyans değeri yüksektir ve eğitim veri setini çok iyi öğrenmiştir. Yanlılık değeri yüksek olan Doğrusal Regresyon ise eğitim veri setini yansıtmakta diğeri kadar başarılı değilse de test veri setini daha iyi yansıtabilmektedir. Bunun sebebi, ilk modelin karmaşıklık düzeyinin optimum noktanın üzerinde olması ve bu durumun aşırı öğrenme problemine yol açmasıdır. Sonuç olarak, eğitim veri setini doğruluk oranı yüksek biçimde yansıtan bir modelin test setini de aynı doğrulukta yansıtabileceğinin garantisi yoktur.

Çalışmamızdaki herhangi bir Makine Öğrenmesi modeli ile Doğrusal Regresyon modelini Şekil 20’deki gibi karşılaştırdığımızda ise benzer bir durumu görebiliriz. Örnek olarak Şekil 20’de Gradient Boosting ile Doğrusal Regresyon tahminleri gerçekleşen değerlerle birlikte verilmiştir.

Şekil 20: Gradient Boosting ve Doğrusal Regresyon (Tahmin ve Gerçekleşen)



Şekil 20’deki örnekte, 40 firmanın toplam nakit değişimi ve iki modelin yaptığı ilgili tahminler yer almaktadır. Makine Öğrenmesi temelli model (Gradient Boosting) çok yüksek varyansa sahip iken Doğrusal Regresyon temelli model daha sabit bir eğilim çizmektedir. Bu sabit eğilim de tahmin performansı açısından avantaj sağlamaktadır. Örnekte de görüldüğü üzere, Gradient

Boosting modelinin tahminleri gerekleŒen deęerlere gre geriden gelmekte ve gerekleŒen verilerle orantılı olarak dalgalanmaktadır. Doğrusal Regresyon ise gerekleŒen deęerlere tam olarak cevap vermemekte olmasına raęmen sabit yapısı sebebiyle gerekleŒen deęerleri biraz daha doğru bir Őekilde yakalayabilmektedir.

Son olarak Őu da vurgulanmalıdır ki; Makine Öğrenmesi gibi karmaŒık modellerin anlamlı bir fark yaratabilmeleri için mevcut verinin yeterince artırılması gerekmektedir. Aksi takdirde bu modellerin tahmin potansiyellerinden tam olarak faydalanılamayabilir. Makine Öğrenmesi temelli modeller tam olarak belirgin olmasa da son dönem tahminlerinde daha iyi sonuçlar vermiŒtir. Bunun sebebi olarak elimizdeki eğitim veri setinin son dönemlerde daha fazla olmasını gösterebiliriz. Ayrıca, bu modellere yeni deęiŒkenler eklenmesi de Makine Öğrenmesi temelli modellerin doğruluk oranını arttıracaktır. Kısacası veri setinin yapısı da bu algoritmaların potansiyelini aŒaęı çekmiŒ olabileceęi düşünölmektedir.

SONUÇ

RMNA, RMD göstergesinden farklı olarak finansal olmayan şirketlerin nakit akışı riskine karşı önlem anlam için geliştirilmiş bir göstergedir. Amaç bir sonraki dönemde oluşabilecek nakit akışı riskini modellemek ve bir öngöründe bulunmaktır. RMNA modeli şirket yöneticisine bir sonraki dönemde gerçekleşebilecek nakit değişimini bir en kötü durum senaryosu olarak verir. Örneğin, bir RMNA modeli “%95 ihtimalle bir sonraki döneme kadar olan nakit değişimi belli bir miktardan fazla olmayacaktır” diyebilir. Böylece şirket yöneticisi nakit değişiminin belli seviyeden fazla düşmeyeceğini bilerek kurumsal risk yönetimini daha etkin bir biçimde planlayabilir. Bu tür bir gösterge sadece bir tahmin ve öngörü getirmez aynı zamanda nakit riskine karşı kurum/işletme içerisinde açık, şeffaf ve öngörülere dayalı bir tartışma ortamının da oluşmasını sağlar.

Çalışmamızda BİST-100’de faaliyet gösteren 40 işletmeden oluşan bir örneklem üzerinden tahmin modelleri geliştirilmiştir. 2008-2022 yılları arasındaki bilanço verileri kullanılarak 40 firma için 1038 farklı dönem için tahminler yapılmıştır. Bu modeller üzerinden hem bir sonraki dönemde gerçekleşebilecek nakit akışı tahmin edilmiş hem de bir güven aralığı ortaya konularak RMNA tahminleri ortaya konulmuştur. Tahmin modellerinin oluşturulmasında Doğrusal Regresyonun yanı sıra önceki çalışmalardan farklı olarak Makine Öğrenmesi algoritmaları kullanılmıştır. Çalışmanın amaçları ve hipotezleri, Makine Öğrenmesi temelli modellerin Doğrusal Regresyon temelli modellere göre hem bir sonraki dönem nakit akış tahminlerinde hem de bir sonraki dönemle ilgili RMNA sınır değerlerinde bir iyileştirme getirip getirmediğini tespit edebilmektir. Daha önceki çalışmalarda sıklıkla rastlanmadığını düşündüğümüz için hem nakit akış tahminlerinin hem de RMNA tahminlerinin başarılı olup olmadığını görebilmek için testler uyguladık. Nakit akış tahminlerinin doğruluğunu tespit edebilmek için Kök Ortalama Kare Hatası ve Ortalama Mutlak Hata ölçütlerini kullandık. RMNA bulgularını test edebilmek için RMD uygulamalarında kullanılan ancak RMNA modelleri için bugüne kadarki çalışmalarda yaygın olarak kullanılmamış olan Kupiec testi ve Süre testi uyguladık.

Testlerin sonuçlarına göre, Doğrusal Regresyon temelli modeller hem nakit akış tahminlerinde hem de RMNA tahminlerinde daha iyi performans göstermişlerdir. Makine Öğrenmesi temelli modellerin beklenen başarıyı gösterememesinin en önemli sebebi eğitim veri setindeki varyansı yansıtmaya çalışmaları ve bu yüksek varyans duyarlılığı sebebiyle nakit akış tahminlerinde kimi

dönemlerde yüksek hata oranı vermesidir Aynı şekilde bu durum gerçekleşen nakit akış değerlerinin RMNA sınır tahminlerinin altında kalmasına yol açmaktadır. Böylece, RMNA güven aralığı bazı dönemlerde gerçekleşen nakit akışı değerini kapsayamamaktadır. Yine de sadece veri etimizin son dönemi için nakit akış tahminlerine baktığımızda, Makine Öğrenmesi temelli modeller daha başarılı olmuşlardır. Örneğin, 2022 4. Çeyrek için Doğrusal Regresyon temelli modeller 40 firma bazında en fazla tahmin hatasını veren modeller olmuşlardır. Veri sayısının artırılmasıyla Makine Öğrenmesi tipi modellerin katkısını da arttırılabildiği buradan yol çıkılarak düşünülebilir. Ayrıca Makine Öğrenmesi modelleri RMNA ihlal sayıları ve Kupiec testinde Doğrusal Regresyon kadar başarılı olmasalar da hassasiyeti daha yüksek bir test olan Süre testinde Doğrusal Regresyona yakın bir performans sergilemişlerdir.

Yapmış olduğumuz çalışma geliştirilmeye açıktır. Öncelikle, çalışmamızda 40 firma üzerinden 10 farklı tahmin modeli kullanıldı ancak hem işletme sayısı hem de model sayısı artırılması bulduğumuz sonuçları iyileştirebilir. Ayrıca, yukarıda kullandığımız ölçütlerin ve testlerin yanı sıra farklı test ve ölçütlerin kontrol edilmesi de modellerin iyileştirilmesine ve mevcut başarılarının anlaşılmasına katkı sağlayacaktır. Çalışmamızda Python programı Scikit-learn kütüphanesi verili parametreleriyle çalıştık. Ancak her firmaya özgü farklı farklı parametrelerin denenmesi sonucu değiştirecektir. Modellerde farklı farklı parametreler denenerek algoritmaların tahmin performansları ya da RMNA performansları arttırılabilir. Özellikle tek ya da az sayıda firma üzerinde çalışılırsa algoritmaların parametre optimizasyonunun yapılması tahmin başarılarını arttıracaktır. Ayrıca firmaların kendi yapılarına göre nakit akış değerlerinde uç değerde bir dalgalanma varsa bunların tespit edilmesi, o firma için kullanılacak en yakın değişkenin belirlenmesi veya firmaların bazı makroekonomik dışsal etkenlere olan duyarlılıklarının (enflasyon artışı, kur değişimi vb.) göz önüne alınması da faydalı olacaktır. Firmaların duyarlılıklarına göre yeni değişkenler eklenerek modellerin tekrar denenmesi ile Makine Öğrenmesi algoritmalarından daha fazla verim elde edilecek ve nakit akış tahmini ve RMNA tahminlerinin doğruluk oranlarının artması beklenecektir.

KAYNAKÇA

- Alam, Naquib Confidence Interval vs Prediction Interval, (Çevrimiçi), 2021, <https://medium.com/analytics-vidhya/confidence-interval-vs-prediction-interval-2f9e36f752e3,27>, 27 Nisan 2023.
- Andrén, Niclas, Exposure-Based Cash-Flow-at-Risk: An Alternative to VaR for Industrial Companies, **Lund University**, Volume:17, 2005, pp. 76-86
- Jankensgård, Håkan, Oxelheim, Lars,
- Andrén, Niclas, Exposure-Based Cash-Flow-at-Risk for Value-Creating Risk Management under Macroeconomic Uncertainty, **Research Institute of Industrial Economics**,2010.
- Jankensgård, Håkan, Oxelheim, Lars,
- Artley,Brendan Time Series Forecasting: Prediction Intervals, (Çevrimiçi), 2022, <https://towardsdatascience.com/time-series-forecasting-prediction-intervals-360b1bf4b085>, 24 Temmuz 2023.
- Avşar, Türkan, “Nakit Akış Tablosu ve Analizi Aracılığıyla İşletmelerde Nakit Yönetimi ve bir Uygulama”, Marmara Üniversitesi Yüksek Lisans Tezi, İstanbul, 2019.
- Bozkuş, Sezer “Risk Ölçümünde Alternatif Yaklaşımlar: Riske Maruz Değer (VaR) ve Beklenen Kayıp (ES) Uygulamaları”, **Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi**, Cilt: 20 Sayı: 2, 2005, s. 27-46.
- Brooks, Chris **Introductory Econometrics for Finance**, Cambridge University Press, 2019.
- Bruce, Andrew **Practical Statistics for Data Scientists**, O’Reilly, 2017

- Chen, Tianqi and Guestrin, Carlos and XGBoost: A Scalable Tree Boosting System, KDD '16, August 13-17, San Francisco, 2016.
- Damodaran, Aswath **Value at Risk,** (Çevrimiçi) <https://pages.stern.nyu.edu/~adamodar/pdfiles/papers/VAR>, 24 Temmuz 2023.
- Dani'elsson, Jon, **The Theory and Practice of Forecasting Market Risk, with Implementation in R and Matlab,** Wiley, 2011
- Deisenroth, Marc **Mathematics for Machine Learning,** Cambridge University Press, Peter, Faisal, Aldo, and Soon Ong, Cheng 2021.
- Demirkapı, Hande and Betül "A Cash flow at Risk (CFaR) Model for Managing Payment Delays in Construction Projects", İzmir Yüksek Teknoloji Enstitüsü, Yüksek Lisans Tezi, 2019.
- Earley, Christine "Data analytics in auditing: Opportunities and challenges", Business Horizons, Kelly School of Business, Volume 58, Issue 5, 2015, pp. 493-500.
- Emekli, Berkay "Risk Yönetiminde Riske Maruz Değer Kavramı : Teori ve İMKB Üzerine Uygulamalar", Yıldız Teknik Üniversitesi, Yüksek Lisans Tezi, 2008.
- Fradkov, Alexander "Early History of Machine Learning, Institute for Problems in Mechanical Engineering", Preprints of the 21st IFAC World Congress (Virtual) Berlin, July 12-17, 2020.
- Scikit-Learn Guide GridSearchCV,(Çevrimiçi)https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

- Gentili, Luca, “A Dynamic Model for Cash Flow at Risk”, DO -
Giacomello, Bruno, 10.2139/ssrn.3307255 JO - SSRN Electronic Journal ER -SSRN,
Girardi, Dario, 2018
Grasselli, Martino
- Jain, Aarshay Mastering XGBoost Parameter Tuning: A Complete Guide with
Python, (Çevrimiçi),
Codes”[https://www.analyticsvidhya.com/blog/2016/03/complete-
guide-parameter-tuning-xgboost-with-codes-python/](https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/)
- Kısakürek, Mustafa **İşletme Faaliyet Nakit Akımlarının Tahmin Edilmesi – İMKB’de
bir Uygulama**, Hacettepe Üniversitesi, Doktora Tezi, Ankara, 2005.
- Krause, Timothy, **Risk management and firm value: recent theory and evidence**,
Yiuman Tse International Journal of Accounting and Information Management
Vol. 24 No. 1, 2016 pp. 56-81, 2015.
- Kupiec, Paul Techniques for Verifying the Accuracy of Risk Measurement Models,
Journal of Derivatives, 1995, pp. 73-84
- Kuti, Monica **Cash Flow at Risk, Financial Flexibility and Financing Constraint**,
University of Pécs, 2011.
- Kuzu, Serdar “Endüstri İşletmelerinde Risk Ölçüm ve Yönetiminde Riske Maruz
Nakit Akış Yönteminin Kullanımı ve Bir Uygulama Örneği”,
Marmara Üniversitesi, Doktora Tezi, İstanbul, 2013.
- Nilsson, Nils **Introduction to Machine Learning**, Stanford University, 1998.

- Özvural, Özhan “Cash Flow at Risk in Publicly Traded Non-Financial Firms in Turkey: An Application in Defense Companies”, İhsan Doğramacı Bilkent Üniversitesi, Yüksek Lisans Tezi, Ankara, 2004.
- Peixerio, Marco The Complete Guide to Time Series Analysis and Forecasting, (Çevrimiçi), 2019,<https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775> 24 Temmuz 2023
- Polat, Güven Gül “Sensitivity Analysis of Expected Shrotfall by means of a second-order Approximation”, İstanbul Teknik Üniversitesi, Yüksek Lisans Tezi, 2012.
- RiskMetrics **Technical Document**, J.P. Morgan, December 17, 1996
- RiskMetrics **The RiskMetrics 2006 Methodology**, J.P. Morgan, 2007
- Ross, Sheldon **Inroductory Statistics**, Academic Press, 2017
- Stephanie, Fook **Introduction to Bootstrap**, Department of Clinical Research, Singapore General Hospital, Singapore, 2011.
- Chong, Robin
- Stein, Jeremy, Usher, **A Comparables Approach to Measuring Cashflow-at-Risk for Non-Financial Firms**, Journal of Applied Corporate Finance Vol. 13, No. 4,2001, pp. 100-109
- Stephan, LaGatutta, Daniel, Youngen, Jeff
- Tüfekçi, Merve “Firmaların Nakit Akış Profilleri ile Finansal Performanslarının Karşılaştırılması: BİST’te bir Uygulama”, Kahramanmaraş Sütçü İmam Üniversitesi, Yüksek Lisans Tezi, 2020.
- Vural, Merve “Operasyonel Nakit Akışlarının Sistem Dinamiği Yaklaşımıyla Tahmin Edilmesi : Gıda Sektöründe bir Uygulama”, İstanbul Ticaret Üniversitesi, yüksek lisans Tezi, 2021.

Wang, Weijie and Yu, Yanmin Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model, **IOP Conf. Ser.: Mater. Sci. Eng.**, 2018, pp 324.

XGBoost Technical Document XGBoost, (Çevrimiçi),
https://xgboost.readthedocs.io/en/stable/python/python_api.html,23
Temmuz 2023

Yan, Meilan, Hall, Maximilian, Turner, J. B., Hall, Paul “Estimating Liquidity Risk Using The Exposure-Based CashFlow-at-Risk Approach: An Application To the UK Banking Sector”, Wiley Online Library Volume19, Issue3, 2014, pp. 225-238.