

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YASSI ALÜMİNYUM ÜRETİMİNDE KALİTE SINIFLARININ
MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNLENMESİ

DOKTORA TEZİ

Alperen AYTATLI

Endüstri Mühendisliği Anabilim Dalı

HAZİRAN 2024

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YASSI ALÜMİNYUM ÜRETİMİNDE KALİTE SINIFLARININ
MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNLENMESİ

DOKTORA TEZİ

Alperen AYTATLI

Endüstri Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Alper KİRAZ

HAZİRAN 2024

Alperen AYTATLI tarafından hazırlanan “Yassı Alüminyum Üretiminde Kalite Sınıflarının Makine Öğrenmesi Yöntemleri ile Tahminlenmesi” adlı tez çalışması 24.06.2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	Doç. Dr. İ. Hakan SELVİ Sakarya Üniversitesi	
Jüri Üyesi :	Doç. Dr. Fuat ŞİMŞİR Yalova Üniversitesi	
Jüri Üyesi :	Doç Dr. Alper KİRAZ (Danışman) Sakarya Üniversitesi	
Jüri Üyesi :	Dr. Öğr. Üyesi M. Rıza ADALI Sakarya Üniversitesi	
Jüri Üyesi :	Dr. Öğr. Üyesi Çağatay TEKE Bayburt Üniversitesi	



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “Bir Tesiste Veri Madenciliği Ve Yapay Zekâ Teknikleri Kullanılarak Planlama Esnekliğinin ve Hızının Artırılması” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, etik kurul onay belgesi aldığımı, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(24/06/2024)

Alperen AYTATLI



TEŞEKKÜR

Öncelikle, bundan 7 yıl önce, beni bu programa kabul ederek bana bu fırsatı sunan ve emekliliğine kadar bana her konuda yol gösteren değerli hocam, Sayın Prof. Dr. Cemalettin KUBAT'a teşekkür etmeliyim.

Doktoram süresince sahip olduğu bilgi ve deneyimlerinden yararlandığım, daima bilgi ve desteğini veren, değerli danışmanım Sayın Doç. Dr. Alper KİRAZ'a, araştırmanın planlanmasından yazılmasına kadar geçen süreçteki destekleri, teşvikleri ve titiz rehberliği için içten teşekkürlerimi sunarım.

Tez izleme süreçlerinde yapıcı ve yol gösterici fikirleriyle, tezde kullanılan yöntemlerin kapsamının genişletilmesinde katkıları olan tez izleme jürisindeki değerli hocalarım Doç. Dr. İhsan Hakan SELVİ ve Dr. Öğr. Üyesi Mehmet Rıza ADALI'ya da ayrıca teşekkürü borç bilirim.

Tez savunmamda verdikleri önerilerle tezime yaptıkları katkılarından dolayı değerli hocalarım Doç. Dr. Fuat ŞİMŞİR'e ve Dr. Öğr. Üyesi Çağatay TEKE'ye de ayrıca teşekkür ederim.

Daima yoğun olmuş ve olacak hayatımda, öncelikle eşime ve aileme de bana bu süreçte verdikleri destek ve anlayış için teşekkür ederim.

Bu Dünya'da herkesin bir kişiden şansı varsa, sanırım benim de şansım annem. Bugün akademik ve özel sektörde elde ettiğim tüm başarıların arkasında onun fedakarlığı yatar. Kendisine, özellikle bu vesile ile de asla ödeyemeyeceğim sonsuz minnet duyduğumu belirtmek isterim.

Sürecinin neredeyse tamamını, kimi zaman tökezleyerek, ağırlıklı olarak hafta içi özel sektörde çalışırken, akşamları ve hafta sonlarını da doktora eğitimime ayırarak tamamladığım doktora eğitimimi ve sonucu olan bu tezimi, bu yoğun tempolu süreçte, sürece katıldığı andan itibaren yeri geldiğinde bir gülüşü ile bile beni hep ayakta tutabilen canım kızıma adıyorum. Eğer her düştüğünde ayağa kalkmayı bilir ve asla vazgeçmez ise başaramayacağı, üstesinden gelemeyeceği hiçbirşey olmadığını anlaması ve asla unutmaması için. Attığım her adım, yaptığım her şey senin için güzel kızım...

Alperen AYTATLI



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
SİMGELER	xiii
TABLO LİSTESİ	xv
ÖZET.....	xix
SUMMARY	xxi
1. GİRİŞ.....	1
2. LİTERATÜR ARAŞTIRMASI	3
2.1. Alüminyum ve Metal Dışı Üretim Süreçlerinde Yapılan Çalışmalar	3
2.2. Metal ve Metal Döküm Süreçlerinde Yapılan Çalışmalar	6
2.3. Alüminyum ve Alüminyum Döküm Süreçlerinde Yapılan Çalışmalar	9
3. MATERYAL VE YÖNTEM.....	13
3.1. Kullanılan Materyal, Veri Setleri	13
3.2. Kullanılan Araç-Gereçler	13
3.3. Kullanılan Yöntemler	14
3.3.1. Veri ön işleme	15
3.3.2. Kümeleme yöntemleri.....	16
3.3.3. Derin öğrenme ve makine öğrenmesi sınıflandırma yöntemleri.....	18
3.3.4. Değerlendirme metrikleri	21
3.3.4.1. Karmaşıklık matrisi	21
3.3.4.2. Siluet (Silhouette) skoru.....	22
3.3.4.3. AUC-ROC eğrisi.....	23
3.3.4.4. Algoritma çalışma süresi.....	24
3.3.5. Veritabanı analiz yöntemleri	25
4. ARAŞTIRMA BULGULARI.....	27
4.1. Proses Özeti.....	27
4.2. Yassı Mamul Alüminyum Ürünlerinde Yapılan Kalite Kontrol Testleri.....	28
4.3. Analizler	32
4.3.1. Veri ön işleme	32
4.3.2. Toplanan verilerin düzenlenmesi	33
4.3.2.1. Eksik verilerin analizi ve etkisiz parametrelerin ayıklanması.....	35
4.3.2.2. Yanlış tasnif edilmiş yada mantıksız değerler	39
4.3.3. Veri madenciliği.....	41
4.3.3.1. Veri seti (veri çerçevesi) üzerinde istatistiksel analizler.....	41
4.3.3.2. Kümeleme analizi ve sonuçları	43
4.3.3.3. Sınıflandırma analizleri	50
4.3.4. Çalışmanın sonuçlarının uygulanması	66
5. TARTIŞMA VE SONUÇ.....	69
KAYNAKLAR	73

EKLER.....	81
ÖZGEÇMİŞ.....	95



KISALTMALAR

AUC	: Eğri Altındaki Alan
ADABOOST	: Uyarlamalı Artırma
BIRCH	: Hiyerarşilerle Dengeli Yinelemeli Azaltma ve Kümeleme
DT	: Karar Ağacı
IF	: İzolasyon Ormanı
FFNN	: İleri Beslemeli Yapay Sinir Ağı
FPR	: Hatalı Pozitif Oranı
GBT	: Gradyan Artırma Ağaçları
GB	: Gigabyte
GBM	: Gradyan Artırma Makineleri
GPU	: Grafik İşlem Birimi
HAC	: Hiyerarşik Kümeleme
KNN	: K-En Yakın Komşu
LR	: Lojistik Regresyon
LSTM	: Uzun Kısa Süreli Bellek
NAR	: Doğrusal Olmayan Oto Regresif Model
NB	: Naive Bayes
PCB	: Baskılı Devre Kartı
OCSVM	: Tek Sınıf Destek Vektör Makinesi
RAM	: Rastgele Erişim Belleği
RF	: Rastgele Orman
RNN	: Tekrarlayan Sinir Ağı
ROC	: Alıcı Çalışma Karakteristiği
SVM	: Destek Vektör Makinesi
TPR	: Duyarlılık
XGBOOST	: Aşırı Gradyan Artırma
YSA	: Yapay Sinir Ağı
YZ	: Yapay Zekâ



SİMGELER

S	: Siluet İndeksi
χ^2	: Ki-kare
T_n	: Gerçekte Yanlış iken Yanlış Tahmin Edilen Değerler, Doğru Tahmin
T_p	: Gerçekte Doğru iken Doğru Tahmin Edilen Değerler, Doğru Tahmin
F_n	: Gerçekte Doğru iken Yanlış Tahmin Edilen Değerler, Doğru Tahmin
F_p	: Gerçekte Yanlış iken Doğru Tahmin Edilen Değerler, Doğru Tahmin
C_n	: Gerçek Negatif Değerlerin Sayısı
C_p	: Gerçek Pozitif Değerlerin Sayısı
R_n	: Tahmin Negatif Değerlerin Sayısı
R_p	: Tahmin Pozitif Değerlerin Sayısı
α	: Yanlış Pozitif Oranı
β	: Doğru Tahmin Edilen Değerlerin 1'den Farkı
$\Delta\alpha$	: Birbirini Takip Eden İki Tahmin Pozitif Değeri Farkı



TABLO LİSTESİ

Sayfa

Tablo 3.1. Karmaşıklık matrisi.....	21
Tablo 3.2. ROC eğrisi karmaşıklık matrisi.....	23
Tablo 4.1. Kalite tipi parametreleri	31
Tablo 4.2. Yarı mamullere ait üretim dönemleri.....	33
Tablo 4.3. Çalışmaya girdi teşkil eden parametreler	35
Tablo 4.4. Tespit edilen girdi değişkenleri.....	37
Tablo 4.5. Korelasyon analizi sonucu ayıklanan girdi parametreleri.....	39
Tablo 4.6. Korelasyon analizi sonucu kalan girdi değişkenleri	42
Tablo 4.7. Seçilen kümeleme algoritmaları.....	45
Tablo 4.8. Kümeleme algoritmaları siluet skor ve auc değerleri karşılaştırması	46
Tablo 4.9. Yarı mamul bazlı uygulanan kümeleme algoritmaları sonuçları	48
Tablo 4.10. Tüm veri sınıflandırma analizi sonuçları	51
Tablo 4.11. B01 Yarı mamülü sınıflandırma analiz sonuçları	53
Tablo 4.12. B02 Yarı mamülü sınıflandırma analiz sonuçları	54
Tablo 4.13. B03 Yarı mamülü sınıflandırma analiz sonuçları	54
Tablo 4.14. B04 Yarı mamülü sınıflandırma analiz sonuçları	55
Tablo 4.15. B05 Yarı mamülü sınıflandırma analiz sonuçları	55
Tablo 4.16. B06 Yarı mamülü sınıflandırma analiz sonuçları	56
Tablo 4.17. B07 Yarı mamülü sınıflandırma analiz sonuçları	56
Tablo 4.18. B08 Yarı mamülü sınıflandırma analiz sonuçları	57
Tablo 4.19. B09 Yarı mamülü sınıflandırma analiz sonuçları	57
Tablo 4.20. B10 Yarı mamülü sınıflandırma analiz sonuçları	58
Tablo 4.21. B11 Yarı mamülü sınıflandırma analiz sonuçları	58
Tablo 4.22. B12 Yarı mamülü sınıflandırma analiz sonuçları	59
Tablo 4.23. B13 Yarı mamülü sınıflandırma analiz sonuçları	59
Tablo 4.24. B14 Yarı mamülü sınıflandırma analiz sonuçları	60
Tablo 4.25. B15 Yarı mamülü sınıflandırma analiz sonuçları	60
Tablo 4.26. B16 Yarı mamülü sınıflandırma analiz sonuçları	61
Tablo 4.27. B17 Yarı mamülü sınıflandırma analiz sonuçları	61
Tablo 4.28. B18 Yarı mamülü sınıflandırma analiz sonuçları	62
Tablo 4.29. B19 Yarı mamülü sınıflandırma analiz sonuçları	62
Tablo 4.30. B20 Yarı mamülü sınıflandırma analiz sonuçları	63
Tablo 4.31. B21 Yarı mamülü sınıflandırma analiz sonuçları	63
Tablo 4.32. B22 Yarı mamülü sınıflandırma analiz sonuçları	64
Tablo 4.33. B23 Yarı mamülü sınıflandırma analiz sonuçları	64
Tablo 4.34. B24 Yarı mamülü sınıflandırma analiz sonuçları	65
Tablo 4.35. Planlama hazırlık aşamasında iyileşme.....	67
Tablo 4.36. Malzeme değişiminde azalma	67
Tablo 4.37. Nihai ürün stoğunda azalma	67
Tablo 4.38. Rework miktarında azalma	67
Tablo 4.39. Özet tablo	68



ŞEKİL LİSTESİ

Sayfa

Şekil 3.1. Çalışmanın akış şeması.....	15
Şekil 3.2. ROC eğrisi (Bradley, 1997).....	23
Şekil 3.3. Büyük veri işleme yaşanan ana problemler (Cielen ve ark., 2016).....	25
Şekil 3.4. Ram kısıtı problemine olası çözüm yolları (Cielen ve ark., 2016).....	26
Şekil 4.1. Alüminyum yassı mamul üretim süreci.....	27
Şekil 4.2. Döküm hattı.....	27
Şekil 4.3. Bir görsel kontrol sisteminin yakaladığı yüzey kalite özü.....	29
Şekil 4.4. Yassı rulo takip mekanizması.....	32
Şekil 4.5. Girdi parametrelerinin takip edildiği iba sistemi.....	34
Şekil 4.6. Df veri çerçevesi işlemleri.....	36
Şekil 4.7. Df üzerindeki boş ve tanımsız satırlar.....	36
Şekil 4.8. Diğer değişkenler tanımsız ve boş satır sayıları.....	38
Şekil 4.9. Değişkenlerin tanımsız ve boş değerlerinin ayıklanması.....	38
Şekil 4.10. B01 malzemesi tandış metal seviyesi profilreport () analizi.....	40
Şekil 4.11. Veri setinde teknik limitler dışı değerlerin temizlenmesi.....	41
Şekil 4.12. Girdi parametreleri/çıktı (kalitetipi) arasındaki korelasyon matrisi.....	42
Şekil 4.13. Elbow yöntemi ile optimal küme analizi.....	44
Şekil 4.14. Tüm veri üzerinde, ward kümeleme denemesi sonucu.....	45
Şekil 4.15. Tüm veri seti için k-means kümeleme.....	46
Şekil 4.16. Tüm veri kümeleme algoritmaları roc eğrileri.....	47
Şekil 4.17. B01-B06 arası yarı mamül kümeleme roc eğrileri.....	49
Şekil 4.18. B07-B12 arası yarı mamül kümeleme roc eğrileri.....	49
Şekil 4.19. B13-B18 arası yarı mamül kümeleme roc eğrileri.....	50
Şekil 4.20. B19-B24 arası yarı mamül kümeleme roc eğrileri.....	50
Şekil 4.21. Tüm veri sınıflandırma roc eğrileri.....	52
Şekil 4.22. Algoritmaların veri seti üzerindeki başarıları.....	66



YASSI ALÜMİNYUM ÜRETİMİNDE KALİTE SINIFLARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNLENMESİ

ÖZET

Rekabetçi ve esnek üretim ortamında, ürünü müşteriye hızlı, istenilen kalitede ve oluşabilecek en düşük maliyette üreterek sevk edebilmek, dün de önemli olmakla birlikte günümüzde üreticiler için olmazsa olmaz bir hale gelmiştir. Bazı üretim süreçlerinde, üretimin çıktısının istenilen kalitede olup olmadığını önceden tahmin edebilmek ise, bu hedefe ulaşmada en etkin yollardan biridir. Günümüzde bu konuda genel yaklaşım uzman kişilerin bilgi ve tecrübelerinden yararlanmaya çalışmak olsa da, bu tarz yaklaşımlarla elde edilen başarılar, oldukça sınırlı kalmaktadır.

Alüminyum, Dünya’da yükselen bir metaldir. Bu metalden elden edilen yassı rulo ürünlerin de yükselişi, alüminyumun cazibesinin artmasına paralel olarak devam etmektedir. Ortada olan yüksek pazar, yüksek rekabeti de beraberinde getirmektedir. Sürekli üretimin yapıldığı bu ürünlerin üretim prosesindeki en küçük iyileştirmeler bile, yüksek karlılık ve rekabet avantajı olarak firmalara geri dönmektedir.

Bu çalışmada, Alüminyum Yassı Mamul üretiminin ilk aşaması olan Döküm Sıcak Haddelme sürecine ilişkin, süreçteki girdi değişkenleri parametrelerine bağlı olarak nihai ürünün kalitesini (1. Kalite, 2. Kalite ya da 3. Kalite), nihai ürün sok kontrol aşamasına gelmeden, daha döküm aşamasında iken tespit edecek, böylelikle istenmeyen üretimlerin ve bunlara bağlı oluşan maliyetlerin önüne de geçerek üretim planlama esnekliğini artıracak, öğrenebilen akıllı bir yapı oluşturulmaya çalışılmıştır.

Araştırma esnasında, Eğitici-siz Öğrenme metodu olan Kümeleme Yöntemleri ve Eğitici-li Öğrenme, Sınıflandırma yöntemleri olan Makine Öğrenmesi ve Derin Öğrenme yöntemleri denenmiştir. Yöntemlerin uygulanmasından önce verinin sistemden alınması sonrasında veri üzerinde veri ön işleme süreçleri uygulanmıştır.

Araştırma sonuçlarına göre, Kalite Tipi’nin eğitici-siz öğrenme teknikleri ile belirli bir seviyede ama düşük bir seviyede olmakla birlikte, eğitici-li öğrenme teknikleri ile yüksek seviyede tahmin edilebileceği görülmüştür. Araştırma, mevcut durumda Eğitici-li Öğrenme tekniklerinin yarı mamul bazında uygulanmasının en iyi sonuçları ortaya koyduğunu, bu durum içerisinde de farklı yarı mamul tiplerinde farklı algoritmaların en başarılı algoritmalar olabileceğini göstermiştir. Elde edilen sonuçlar ışığında yüksek derecede kazanımlar elde edilmiştir. Sonuçlar, fiili olarak kullanıma geçmiş olmakla birlikte, kullanılmakta olan yazılımlara entegrasyon süreci devam etmektedir.



PREDICTION OF QUALITY CLASSIFICATION IN FLAT ROLLED ALUMINIUM PRODUCTION USING MACHINE LEARNING METHODS

SUMMARY

In a competitive and flexible production environment, it is essential for today's producers to ship products to customers quickly, at the desired quality, and at the lowest possible cost. One of the most effective ways to achieve this goal in some production processes is to predict if the output of the production is of the desired quality.

In the casting process, the raw flat aluminum coil produced undergoes different processes according to the variety of the final product and transforms into the end product. At this point, looking at the production process, it is generally observed that these raw coils are subjected to long-term casting, and the quality of these components can only be understood after the production process. If the resulting final product is not of the desired quality, it is observed that this semi-finished component, if not scrap, can be served as 2nd or 3rd quality.

During the process, the ability to understand the quality of the semi-finished component that will emerge is extremely important. For example, if a coil expected to meet the final product with a 1st quality expectation is not of the desired quality during production, the entire planning will need to be changed. As a matter of fact, this situation will also cause the production of other semi-finished coils waiting for other customers to be delayed. This is a situation that creates bottlenecks in terms of planning.

On the other hand, predicting the quality of the semi-finished coil that will emerge after casting will have many advantages. The most important of these is that if the coil that will emerge is scrap, urgent measures can be taken beforehand, and if a 2nd or 3rd quality semi-finished product will emerge, it can be quickly analyzed whether it can be substituted for other orders waiting in the plan, thus allowing these orders to be removed from the plan.

In reality, expert personnel also try to make these predictions and form some forecasts. However, the complexity of the problem makes it very difficult for these predictions to be accurate.

Today, using Artificial Intelligence and Data Mining techniques, high prediction success has been achieved with prediction algorithms in many different fields through data obtained from systems. Of course, for this, the available data must first be transformed into meaningful data through preprocessing steps using data mining methods, and then the most appropriate algorithms must be applied to the obtained data.

In this study, considering many input variables and the quality of the output semi-finished rolls that are thought to affect the result tracked instantly during the casting process, it was tried to predict the quality level of the output roll based on the input variables. With the correct prediction, flexibility provided in the production planning

of the semi-finished product increased efficiency in the production process, and some delays in the delivery times of the final product to the customer were prevented.

In the first sections of the thesis, a Literature Review is presented. In the literature, it is possible to find many studies on predictions using Data Mining and Machine Learning methods. Within the scope of the literature review, the studies conducted were first examined in general on classification and prediction studies, then studies conducted specifically in the metal sector, and finally, studies on data mining and prediction algorithms on aluminum casting, which is the subject of our study, were examined.

In the continuation of the thesis, the materials and methods used in the thesis are introduced. In the research, data from the casting process of a factory producing Flat Aluminum Coil was used. Within the data obtained, there are 63 different input variables. One year of data for these variables was taken from the system. The data obtained from the system were analyzed with personal computers and rented super powerful computers. The data obtained from the system were taken as Excel files and uploaded to the Jupyter Notebook library. Then, Clustering methods, which are among the Unsupervised Learning methods, were tried on the data. Clustering analysis is a group of multivariate techniques whose primary aim is to group objects (units) based on their characteristics. Clustering analysis groups objects in a very similar manner within the cluster and differently between clusters. If the clustering process is successful, when a geometric drawing is made, the objects will be very close to each other within the cluster and far from each other between clusters. At this point, Clarans, Birch, Spectral Clustering, and Hierarchical Clustering Methods were used. After testing the Unsupervised Learning algorithms, machine learning and deep learning algorithms, which are Supervised Learning methods, were tested. Here, twelve different algorithms were tested. In both cases where Unsupervised Learning algorithms were tested and where Supervised Learning algorithms were tested, the algorithms were first applied to the data set, considering the materials as a single data frame, then clustering and classification algorithms were applied to the data, and then the materials were taken one by one from the data set as individual data frames, and the clustering and classification algorithms were applied to the data belonging to the materials.

To evaluate the performance of the methods, evaluation metrics from the literature were used. The silhouette score was used to evaluate the clustering algorithms. The Silhouette Score was developed to find the appropriateness of the clustered data within their respective clusters. Additionally, to evaluate the performance of both Supervised and Unsupervised Learning algorithms, the AUC Score and ROC Curve were used. This method is a good way to visualize the performance of a classifier to select an appropriate operating point or decision threshold. Another evaluation method used was the confusion matrix. The F1 Score, Recall, and Precision values calculated using the confusion matrix were used to compare the performance of the classification algorithms. In the world of machine learning, the runtime of the algorithm to be used should always be considered when selecting the most accurate algorithm for the data among the available algorithms. Therefore, when comparing algorithm performance, the runtime of the algorithms was also considered as a comparison criterion.

The results of the study were implemented in the facility. Accordingly, the successful algorithms from the study were used during the planning phase based on input parameters through Jupyter Notebook. The success of the algorithms was tested by

making necessary planning changes based on the findings obtained. At this point, very successful results were also achieved in the real-life application of these algorithms. The gains achieved include improvements in the planning preparation phase, a reduction in material changes, a decrease in the inventory of final products held, and a reduction in the amount of rework. At the end of the section, these gains and their contributions are presented to the reader.

In conclusion, looking at the significant findings obtained from the thesis, the first noteworthy point is that this study has highly accelerated a process that was previously predicted based on individuals' experiences and took a considerable amount of time even then, making it possible to predict without human errors. In this context, the thesis is one of the few examples conducted specifically for the sector and problem. The thesis demonstrated that clustering algorithms applied to the entire data set performed poorly. However, the classification tested with machine learning and deep learning in the continuation of the thesis gave very positive results. In the initial evaluation conducted without separating semi-finished products as a single data set, the AUC value, accuracy value, and F1 score, which are the first evaluation criteria, ranged between 0.80-0.91 for the best algorithms. This is a very good success. Among the algorithms, the CATBOOST and GBM algorithms achieved these values, making them the most successful algorithms. However, in terms of values, the CATBOOST algorithm is slightly better than GBM. Therefore, it is stated that the CATBOOST algorithm is the most successful algorithm. Finally, when classification was applied with machine learning and deep learning algorithms on a semi-finished product basis, the obtained AUC, accuracy, and F1 score values ranged between 0.90-1.00 for most types of semi-finished products. At this point, it has been shown that the most successful solution for the problem in the study is the classification with machine learning and deep learning on a semi-finished product basis. The most successful algorithm at this point has been identified as the RNN algorithm.

In light of these results, a software structure has been created where, in reality, a program that makes predictions by running the most successful algorithm from our study for the semi-finished product to be included in the production plan is used. This structure is currently in use and has yielded very positive results in real life. These results have reflected in reduced costs and increased profitability.



1. GİRİŞ

Herhangi bir ürünün üretiminde kaliteli ve düşük hurda ile üretim yapmak ciddi bir rekabet avantajı sağlamaktadır. Yüksek kaliteli ürünler üretmenin yolu, onları doğru girdi parametreleri ile üretmekten geçmektedir. Bu nedenle, uygun parametrelerin tasarlanması, üretim kalitesinin erken bir üretim aşamasında tahmin edilmesine bağlıdır. Bu aynı zamanda yüksek kaliteli ürünlerin daha düşük maliyetlerle optimize edilmesine olanak sağlar. Son zamanlarda, yapay zekâ gibi bilgisayar teknolojilerinin gelişimi, üretim süreçlerini hesaba katması gerekmeyen minimum bilgi gereksinimleri nedeniyle araştırmacıların dikkatini çekmiştir (Deng, Bai ve Li 2020).

Alüminyum, ekonomik ve çevreci bir metaldir. Hayatın her alanında içecek kutusundan mutfağımıza, ofisimizden uçağımıza yüzlerce alanda kullanılmaktadır. Diğer yandan bakıldığında oldukça yeni diyebileceğimiz bir sektördür. Aynı zamanda Alüminyum, yeryüzünde en bol bulunan 3. element olma özelliği ile, üretim teknolojilerinde oldukça avantajlı bir hammadde konumunda durmaktadır (Bayseçkin, 2002).

Bugün yassı mamul olarak isimlendirdiğimiz alüminyum ürünü, alüminyumun çeşitli proseslerden geçmesi sonucu farklı kalınlıklarda elde edilen, rulo halinde alüminyumdur. Çalışmamıza konu olan süreç ise, işte bu alüminyum rulo üretiminin ilk başladığı nokta olan döküm sürecidir.

Döküm sürecinde üretilen ham yassı rulo alüminyum, sonrasında nihai ürünün çeşitliliğine göre farklı süreçlerden geçer ve son ürüne dönüşür. Bu noktada üretim sürecine bakıldığında, genel olarak bu ham rulonun uzun süreli dökümlerinin yapıldığı ve bu bileşen kalitesinin ise üretim sürecinin sonrasında anlaşılabilirdiği görülmektedir. Çıkan nihai ürünün istenilen kalitede olmaması durumunda, bu yarı mamul bileşenin hurda değil ise 2. ya da 3. Kalite olarak servis edilebildiği de görülmektedir.

Süreç esnasında, ortaya çıkacak yarı mamül bileşenin kalitesinin anlaşılabilirliği oldukça önemlidir. Örneğin, 1. Kalite beklenti ile son ürüne hitap etmesi istenen bir rulonun üretim esnasında istenilen kalitede olmaması, bütün planlamanın yeniden değiştirilmesi gerekliliğini doğuracaktır. Nitekim bu durum aynı zamanda başka

müşterilerin beklemekte olan başka yarı mamul rulo üretimlerinin de gecikmesine neden olacaktır, olmaktadır. Bu da planlama anlamında darboğazlar doğuran bir durumdur.

Döküm sonrası ortaya çıkacak yarı mamul rulonun kalitesinin ne olacağının tahmin edilebilir olmasının ise diğer yandan birçok avantajı olacaktır. Bunlardan en önemlisi ise, eğer çıkacak rulo hurda olacak ise acil olarak önden buna önlem alınabileceği gibi, eğer 2. ya da 3. Kalite bir yarı mamul çıkacak ise, bunun planda bekleyen başka siparişlere ikame edilebilirliğinin hızlıca analiz edilmesi ile bu siparişlerin planda aradan çıkarılmasını sağlayabilmek olacaktır.

Aslında bakıldığında gerçek hayatta da uzman personeller bu tahmini yapmaya çalışmakta ve bazı öngörüler oluşturmaktadır. Ancak problemin karmaşıklığı bu tahminin doğru olmasını oldukça zorlaştırmaktadır.

Günümüzde YZ ve Veri Madenciliği teknikleri kullanılarak, sistemlerden elde edilmiş veri vasıtasıyla, birçok farklı alanda tahmin algoritmaları ile yüksek tahmin başarısı elde edilebilmiştir. Tabii ki bunun için, sahip olunan verinin veri madenciliği metotları ile anlamlı bir veriye doğru ön işleme adımları uygulanarak önce dönüştürülmesi ve sonrasında da elde edilen veri üzerinde en uygun algoritmaların uygulanması gerekmektedir.

Bu çalışmada, döküm sürecinde anlık olarak takip edilen sonuca etkisi olabileceği düşünülmekte olan birçok girdi değişkeni ve çıktı yarı mamul roluların kaliteleri göz önünde bulundurularak, girdi değişkenlerine bağlı olarak çıktı rulonun kalite seviyesi tahmin edilmeye çalışılmıştır. Doğru tahmin ile yarı mamulün üretim planlamasında sağlanacak esneklik ile üretim sürecinde verimlilik artırılmıştır ve müşteriye nihai ürünün teslim sürelerinde yaşanan bazı gecikmeler önlenmiştir.

2. LİTERATÜR ARAŞTIRMASI

Literatürde Veri Madenciliği ve Makine Öğrenmesi yöntemleri kullanılarak yapılan tahminler üzerine farklı çalışmalar bulunmaktadır. Bu tahmin, gerek üretim süreçlerinde bir sürece ait bir tahmin, gerekse farklı bir konu olabilir. Örneğin literatürde kişinin spor yaptığı esnada oksijen tüketiminin ulaşacağı maksimum seviyeyi tahmin ederek, kişinin spor esnasında aşırı yüklenmesinin önüne geçmeye çalışan makalelere bile rastlanmaktadır (İşler, 2014). Bu nedenle kaynak araştırması kapsamında tezimiz içerisinde tezin de inceleme alanı olan üretim süreçlerinde yapılan tahmin çalışmalarından daha ziyade bahsedilmiştir.

Literatür araştırması kapsamında, yapılan çalışmalar, öncelikle genel olarak sınıflandırma ve tahmin üzerine yapılan çalışmalar incelendikten sonra, metal sektörü özelinde yapılan çalışmaların incelenmesi ve en son olarak da çalışmamıza konu olan alüminyum döküm üzerinde yapılan veri madenciliği ve tahmin algoritma çalışmalarının incelenmesi şeklindedir.

2.1. Alüminyum ve Metal Dışı Üretim Süreçlerinde Yapılan Çalışmalar

Üretim süreçlerinde kullanılan veri madenciliği ve tahmin çalışmaları incelenmeye başlandığında, ilk göze çarpan, farklı üretim süreçlerinde ve farklı sektörlerde birçok çalışma yapılmış olduğunun görülmesidir.

Yapılan bir çalışmada lineer model, ekipmandaki aşınmanın kalite üzerindeki etkisini modellemek üzere kullanılmıştır. İstatistik model konseptinin amacı bu istatistik öğeler ile (varyans, ortalama, sağa/sola yatıklık, maksimum/minimum vs.) hedef arasında bir ilişki kurabilmektir. Fakat üretimin artan kompleksliği ile birlikte, veri, çok boyutluluk, dinamiklik, kesin olmama muğlaklık gibi istatistik modellerin açıklamakta zorlandığı modern üretimde sıkça rastlanılan konular nedeniyle istatistik model ile bir noktaya ulaşılammıştır (Hao ve ark., 2017).

Bilgisayar sistemlerindeki gelişmeler ve akıllı üretim metotlarına artan talebin sonucu olarak, kalite tahmininde yapay zekâ tabanlı çalışmaları hızlandırdığı (Lieber ve ark.,

2013) ve örnek olarak, yarı iletken bir ürünün verimini modelleme için yaratıcı bir YSA modeli olarak kullandığı görülmüştür (Chen,2017).

Başka bir çalışmada da kaynak dikişinin tipini anlayabilmek için SVM kullanılmış (Fan, Lin ve Tsai, 2016) iken, bir başka çalışmada Wang ve arkadaşları, üretimi devam eden ürünlerde kalite tahmini yapabilmek için genel bir YSA önermişlerdir (Li ve ark., 2019). Deng ve arkadaşları ise üretim kalitesini tahmin etmek için otomatik kodlayıcı tabanlı bir derin regresyon modeli oluşturmuştur (Deng, Bai ve Li, 2020). Bahsedilenlere ek olarak ayrıca, öznitelik çıkarma ve model optimizasyonu dahil olmak üzere iki açıdan tahmin doğruluğunu geliştirmek için birçok çalışma yapılmıştır. Örneğin Kao ve arkadaşları öznitelik çıkartabilmek için Temel Bileşen Analizi'ni desen tanımayı iyileştirmek için SVM yöntemini kullanarak yapmışlardır (Kao, Lee ve Lu ,2016).

Üretim kalitesinin tahminini yapacak bir model inşa etmeyi amaçlayan bir başka çalışmada ise öncelikle hesaplama yükünü hafifletmek için faydalı olan nitelik (farklı süreçteki parametreler) önem analizi için Kaba Küme Teorisi yapılmış, sonrasında da çıkarılan bilgi LSTM algoritması kullanılarak kalitenin tahmininde kullanılmıştır. LSTM perfromansını iyileştirmek için ADABOOST kullanılmıştır ve böylece daha güçlü bir regresyon modeli elde edilmiştir (Bai ve ark., 2021). Başka bir çalışmada yine LSTM algoritması ile bir Alman otomobil fabrikasında bulunan sacı bükerek şekil veren bir makinada, makinanın çalışması esnasında sensörlerden alınan verileri kullanarak, makinanın çıktısının kalitesinin ne olacağının tahmin edilmesine çalışılmıştır (Meyes ve ark., 2019). LSTM modeli kullanılarak ekipman kullanılabilir ömrünün tespit edildiği başarılı bir çalışma da literatürde bulunmaktadır (Zhou, Zhao ve Gao, 2019). Ayrıca LSTM kullanılarak makinelerin üretim performansının gerçekleşen makine OEE değerleri girdi olarak kullanılarak yapılan bir çalışma da bulunmaktadır (Brunelli ve ark., 2019).

Literatürde Kestirimci Bakım'a yardımcı olması amacıyla yapılan çalışmalara da yer verilmektedir. Örneğin, kestirimci bakım yapılması amacıyla bu amaca yardımcı olacak bir makine öğrenmesi yöntemi üzerinde çalışılmıştır. En Küçük Kareler Destek Vektör Makinleri algoritması regresyon için kullanılmıştır. Öncesinde ise algoritmayı besleyecek veri üzerinde Çekirdek Spektral Kümeleme yapılmıştır. Analiz için aynı zamanda Doğrusal olmayan Oto Regresif Model kullanılmıştır. Doğrusal olmayan Oto Regresif Model, çıktı değişkeninin doğrusal olmayan bir şekilde kendi önceki

değerlerine bağlı olduğunu belirten zamanla değişen fenomenleri tanımlar (Langone ve ark.,2015). SVM algoritmasının kalite tahmini konusunda yarı iletken üretimi ile ilgili kullanıldığı ve başarılı olduğu çalışmalar da mevcuttur (Hu ve ark.,2022). Diğer bir kestirimci bakım çalışmasında da örnek bir takım tezgâhı kestirimci bakımı için girdi değerleri arızayı etkileyen parametreler olarak belirlenmiş ve bunların kümeleme yöntemi ile sınıflandırılması noktasında da Makine Öğrenmesi kullanılmıştır. Eğitici öğrenme yöntemlerinden: KNN, FFNN, DT, NB, RF ve SVM denenmiştir, KNN algoritması, algoritmalar içerisinde en iyi sonucu vermiştir (Schmidt ve Wang, 2018).

Yapılan araştırmada, kendinden uyarlamalı tahmin modellerine dayalı çok aşamalı bir üretim sürecinde ürün kalitesinin otonom denetimi ve optimizasyonu için bir uygulama yapıldığı saptanmıştır. Bu uygulamanın en önemli tarafı da diğer uygulamalarda kullanılan yöntemlere ek olarak içerisinde Bulanık Mantık kullanımını da barındırmasıdır. Çalışmada üretim devam ederken eldeki bilgi ile bir Zaman Serisi üzerinden tahmin yapılmakta, sonrasında ise üretim son aşamaya geldiğinde ve gelene kadar da en başta belirlediği kalite kriteri değerini güncellemektedir (Lughofer ve ark., 2019).

Başka bir çalışmada ise endüstride kullanılan robot kollardaki ortaya çıkabilecek arızalar için tahmin algoritması kullanılmıştır. GBM kullanılarak, aynı zamanda istatistik kalite kontrol ile de tahmin eşik değerleri birleştirilmiştir. Böylelikle, kolun arıza yapma potansiyel tahmini alt ve üst kontrol limitlerine istinaden, değerler bu noktalara yaklaştığında robot kol kontrolü yapılacak şekilde süreç organize edilmiştir. Gerçek şu ki, robot kol arızalarının bu şekilde önlenmesi, daha sonrasında arıza ortaya çıktığında yol açacağı büyük mekanik arızaların ve iş kazalarının önüne geçilmesi noktasında çok önemli ve başarılı olmuştur. RF ve XGBoost algoritmalarının kullanıldığı hibrit bir sınıflandırma modeli uygulanmıştır. ABB firmasının robot kollarına uygulanmış bir çalışmadır (Costa ve ark., 2019).

İncelemeler içerisinde anomali tespitinde makine öğrenmesi uygulamasına da rastlanılmıştır. Çalışma, eğitim süresi, tahmin süresi, uygulama kolaylığı ve performansa dayalı üç Makine Öğrenmesi yöntemini değerlendirerek pres sertleştirme sürecinde anomali tespiti için uygun bir yöntem seçerek literatüre katkıda bulunmaktadır. Anomalilerin doğru zamanda tespiti ile makinalara kestirimci bakım yapılarak büyük arızaların ve arıza bakımında yaşanabilecek potansiyel kayıpların önüne geçilmesi amaçtır. Makine Öğrenmesi anomali tespit metodu olarak çalışmada YSA,

OCSVM, ve IF metotları denenmiştir. Metotlar scikit-learn üzerinden denenmiştir (Lejon, Kyösti ve Lindström, 2018). YSA algoritmasının başarılı olduğu, alüminyum alaşımlarının birbirine yapıştırıldığı sürtünme karıştırma kaynağı sürecinde, yapıştırılan parçaların alaşım kalitesini etkileyen parametrelerinin tahmininin yapıldığı bir çalışma da literatürde mevcuttur (D'Souza, Rao ve Herbert, 2023).

İncelenen çalışmalar içerisinde Kalite Tahmini konulu olarak rastlanan bir başka çalışmada ise Makine Öğrenimi tekniklerini ve bulut bilişim ve ağ teknolojileri alanında kullanılan Uç Bulut Bilişimi teknolojisini kullanarak endüstriyel üretimde tahmine dayalı model tabanlı kalite denetiminin yeni bir entegre çözümünü araştırılmıştır. PCB'ler için geliştirilmiş tahmine dayalı model tabanlı kalite kontrol uygulamasında DT, NB, LR, SVM ve GBT algoritmaları kullanılmıştır. Tahmin doğruluğu açısından bakıldığında, SVM ve GBT'nin en iyi sonuçları verdiği görülmüştür (Schmitt ve ark., 2020).

Günümüzde üretim planlamada makine öğrenmesinin kullanımı ve bu konuda yapılan akademik çalışmalara bakıldığında ise ağaç bazlı modeller, YSA'lar ve Kümeleme Yöntemleri üretim planlamada sıklıkla kullanılan yöntemlerden biridir. Son olarak, Temel Bileşen Analizi ve İlişkilendirme Kuralının Makine Öğrenmesi destekli Üretim Planlama Kontrol'de nadiren de olsa kullanıldığını söyleyebiliriz. Birliktelik Kuralı da bilgi ve bağlamsal analiz oluşturmak için uygun olan yorumlanabilir sonuçlar sağladığından keza Üretim Planlama Kontrol içerisinde az da olsa kullanım alanı elde etmesiyle göz önüne çıkmıştır (Cadavid ve ark., 2019).

2.2. Metal ve Metal Döküm Süreçlerinde Yapılan Çalışmalar

Bu bölümde, genel olarak tahmin ve kalite tahmininde yapılan çalışmaların literatür incelemesi sonrasında, konuyu biraz daha özelleştirerek, metal üretim ve metal döküm süreçleri özelinde yapılan çalışmalar incelenmiştir.

Çalışmalar incelendiğinde metal sektöründe bir çok konuda veri madenciliği ve yapay zeka tekniklerinin farklı alanlarda tahmin için kullanılmakta olduğu görülmektedir. Üretim planlamadan üretime, hatta iş güvenliğinde bile çalışmalar bulunmaktadır. Örneğin iş kazalarının oldukça yüksek miktarda görüldüğü metal sektöründe iş kazalarının sınıflandırılması ve sonrasında da tahmini amacıyla yapılan bir çalışmada, temel bileşen analizi ve kümeleme yöntemlerinin kullanılarak kazalara neden olan değişken sayıları azaltılmış, sonrasında elde edilen değişkenler ile çeşitli makine

öğrenmesi ve derin öğrenme algoritmaları uygulanarak, kaza tahmini yapacak en iyi model bulunmuştur. Bulunan ileri beslemeli yapay sinir ağı modeli, tahminde kullanılarak iş kazalarının tahmininde yüzde doksan başarı sağlamıştır (Ayanoğlu ve Kurt, 2019).

İstatistik model, genel olarak üretimde kalite tahminini de içerir şekilde farklı alanlarda ortak bir modelleme metodu olmuştur. Örneğin, Paul Surajit Kumar, lineer olmayan regresyon modelini farklı çelik levhaların çekme özelliklerini tahmin etmek amacıyla uygulamıştır (Paul, 2016).

Çelikler yapısal malzemeler olarak yaygın olarak kullanılır ve bu da onları hayatımızı ve endüstrilerimizi desteklemek için vazgeçilmez kılar. Ancak, çelik üretiminde yer alan sürekli geliştirme ve çok sayıda işleme parametresi nedeniyle çeliğin kapsamlı özelliklerini geleneksel deneme yanılma yöntemleriyle daha da iyileştirmek zorlayıcı hale gelir. Bu zorluğun üstesinden gelmek için, üretim süreçleri ve çelik performansı arasında karmaşık ilişkiler kurmada makine öğrenimi yöntemlerinin uygulanması önemli hale gelir. Yapılan bir çalışmada, makine öğrenimi yöntemlerine genel bir bakışla başlar ve ardından çelik malzemelerdeki çeşitli performans tahminlerini tanıtır. Performans tahmininin sınıflandırılması, makine öğrenimi model destekli tasarımın mevcut uygulamasını değerlendirmek için kullanılır (Fang ve ark., 2024). Benzer bir diğer çalışma, çelik üretim ortamında süreç performansına ilişkin çeşitli göstergeleri tahmin etmek için veriye dayalı modeller geliştirmeyi ve uygulamayı amaçlar. Nihai hedef, optimum performans ve kalite seviyelerine ulaşmak için süreç parametrelerini optimize etmektir. Bu süreç parametrelerinin optimizasyonu, son ürünlerdeki yüzey ve yüzey altı kusurlarını azaltarak, mikro alaşımlı çeliklerin daha iyi kalite ve performansa sahip olmasını sağlamayı amaçlamaktadır. İncelenen çalışmada bu amaçla RF, GBM, XGBOOST, ve YSA algoritmalarını kullanılır (Boto ve ark., 2022).

Üretimin birçok aşamasında ve türünde, görsel tanıma ile hata tespiti yapan akıllı yapay zeka tabanlı sistemler kullanılmaya başlamıştır. İncelenen bir çalışmada, bir metal döküm prosesinde görsel yüzey kusurlarının tespiti için CNN algoritması denendiği görülmektedir (Ferguson ve ark., 2018).

Yapılan başka bir çalışmada, Bulanık Mantık modeli kullanılarak, ülkemizin güzide demir-çelik tesislerinden bir olan Kardemir A.Ş.'de belirli bazı on adet stok kalemi

için olması gereken en uygun stok miktarlarını tahmin eden bir çalışma yapılmıştır (Şimsir, Civak ve Uygun, 2017).

Benzer şekilde yapılan başka bir çalışmada ise paslanmaz çelik sektöründe faaliyet gösteren bir firmanın satış yaptığı sektörlerle göre satış tahminleri gerçekleştirilmiştir. Veri madenciliği yöntemleriyle (Veri Ön işleme, Destek Vektör Regresyonu ve Yapay Sinir Ağları) toplam satış ve sektörlerle göre satış değerlerinin tahminleri gerçekleştirilmiştir. Modelleme sonucunda Destek Vektör Regresyon yönteminin nispeten daha başarılı olduğu görülmüştür (Ecemiş ve Irmak, 2018).

Metal malzemeler üzerinde Kalite Tahmini konusunda da çalışmalar yapılmıştır. Araştırmada, literatürde bulunan bu konudaki bir çalışmada, Kalite sınıflandırılmasında gerek metallerde gerekse plastik ürünlerde kullanılan çekme deneyinin, 4140 çelik malzeme özelinde yapılan çekme testlerinin elde edilmiş gerçek çekme kuvveti ve uzama verisi ile modellenerek farklı çekme hızları altında gerçekleşecek olası uzama miktarlarının yapay sinir ağları kullanılarak tahmin edilmesine çalışılmıştır (Kiraz, 2014).

Metal malzemeler üzerine Kalite Tahmini konusunda karşılaşılan diğer bir çalışma ise kum döküm üzerine yapılmıştır. Bu çalışmada RF algoritması kullanılarak kum dökümde karşılaşılan hatalar tahmin edilmeye çalışılmıştır (Guan ve ark., 2024). Benzer şekilde başka bir çalışmada da bir piston fabrikasının çelik pistonlarının kalıp döküm sürecinde ortaya çıkabilecek kusurların yapay zeka sistemleri ile tahminine çalışılmıştır. Burada da tahmin modeli olarak RF, YSA ve SVM kullanılmıştır (Lee ve ark., 2018). Metal Kalıp dökümü prosesinde yaşanan kalite kusurlarının tespiti ve sınıflandırılması için de ayrıca benzer bir çalışma uygulanmış, bu çalışmada da birçok makine öğrenmesi metodu denenerek model performansları AUC skoru ile ölçülmüştür (Kim ve ark., 2018). Bu çalışmalara ek olarak, Yüksek Basıncılı Kalıp Döküm sürecinde, süreçleri optimize etmek için Endüstri 4.0 teknolojilerinden, özellikle dijital gölgelerden yararlanmaya odaklanan başka bir çalışma daha bulunmaktadır. Bu çalışmada da daha önce bu süreci optimize etmek için YSA kullanılarak yapılan bir çalışmanın (Rai ve ark., 2008) üzerinden gidilmiş ve bu çalışmanın sıcaklık, erimiş metal alaşımın döküme akış hızı şeklinde seçtiği ve gözeneklilik değerini, katılaşma süresini ve akış süresini tahmin ettiği ama net bir eğilim belirleyemediği noktadaki eksikliklerini tamamlamıştır. Bunun için veri ön işleme aşamasında boyut azaltma kullanmış ama burada popüler olan Temel Bileşen

Analizi yöntemi yerine boyut azaltma için Tekil Değer Ayrışımı yöntemi tercih edilmiştir. Döküm nesnelerdeki yüzey kusurlarını doğru bir şekilde tahmin etmek ve bulmak için Dijital Gölgeleme adı ile yeni bir görüntü tabanlı piksel sınıflandırma yaklaşımı tanıtılmıştır. Önerilen Dijital Gölge modelleme yaklaşımı, yüzey kusurlarını tahmin etmede gelişmiş doğruluk göstererek, gelişmiş üretim izleme için alan spesifik bilginin gerçek zamanlı veri analitiğiyle bütünleştirilmesinin etkinliğini kanıtlamaktadır. (Chakrabarti ve ark., 2021).

Metal sektöründe gerek çelik sac üretilsin, gerekse alüminyum sac üretilsin, üretilecek olan ürünün üretimi için ilk önce Plaka (Slab) adı verilen külçeler üretilir. Bu külçelerin kalitesi ise, takip eden proseslerde üretilecek metal ürünlerin kalitesini de etkiler. İlgili plakaların kalitesinin tahmini üzerine de literatürde LSTM algoritmasının kullanıldığı çalışma bulunmaktadır (Wu ve ark.,2020). Benzer bir mantıkla başka bir çalışmada da şerit çeliklerde dengesiz veri üzerinde, prosesin mekanik özelliklerinin tahmini ve optimizasyonu amacıyla RF algoritması denenmiştir (Li ve ark., 2023).

Literatürde rastalanan örneklerde son olarak metallerin diğer metallerle kullanıldığı proseslere ait tahminlerden de bahsedebiliriz. Buna bir örnek olarak, literatürde yine son dönemde yapılan bir çalışmada, çelik-alüminyum akışlı matkap vidasının farklı plaka kalınlıkları üzerindeki mekanik davranışlarını tahmin etmek üzere YSA modellerinin kullanıldığı bir çalışmaya rastlanmaktadır (Zhou ve ark., 2023).

2.3. Alüminyum ve Alüminyum Döküm Süreçlerinde Yapılan Çalışmalar

Literatürde alüminyum üzerine yapılan çalışmalar incelendiğinde sınırlı miktarda olsa da, alüminyum üretim süreçlerinde de denenmiş bir miktar tahmin algoritması olan çalışmaların olduğu görülmektedir.

Yapılan bir çalışmada, yeni bir sınıflandırma metodu tanıtılmıştır. Bu metotta, bir veri kümeleme işlemini ve elde edilen kümelerin daha sonra etiketlenmesini birleştirerek, sınıflandırma amaçları için kullanılabilecek bir yapı oluşturur. Bu yöntem, düzensiz veri setleri üzerinde bulanık mantık ile bir eğitici-siz öğrenme tekniğinin bir arada kullanılması ile uygulanan bir yöntemdir. Bu yöntem, Göğüs Kanseri hücreleri tespitinde başarı ile uygulandıktan sonra sürekli dökümde alüminyumda ergimiş alüminyumun tandište tıkanıklık oluşumunun tahmininde de başarılı olmuş, tıkanıklığın olacağını girdi değerlerine istinaden yüksek oranda tahmin edebilmiştir (Vannucci ve Colla, 2011).

Çekilen alüminyum rulonun şeklinin stabil kalmasında, diğer birçok üretim faktörünün etkisi olduğu gibi, rollerin hızının artmasının da etkisi vardır. merdanenin şeridinin düzlüğü esas olarak şeridin genişliği boyunca rulo aralığının konfigürasyonuna bağlıdır. Merdanelerinin şekli, merdane boşluğu konturu üzerinde önemli bir etkiye sahip olan merdane boşluğunun şeklini etkileyen ana faktörlerden biridir. Bununla birlikte, sürekli şerit haddeleme işleminde merdanelerin termal deformasyonu ve aşınması nedeniyle iş merdanelerinin şekli dinamik olarak değişir. Bu nedenle, merdanelerin termal değerlerindeki değişimin anlık takibi çok önemlidir. Normalde aşırı ısınmalarda devreye giren soğutma sistemleri bu değişimi yönetmekte oldukça iyidir. Ancak, Haddeleme işleminin hızlı olması ve iş merdanelerinin termal taçlarının çevrimiçi olarak ölçülmesinin imkansız olması nedeniyle, termal taç tahmini için kesin hesaplama verimliliğine sahip çevrimiçi bir model çok önemlidir. Bu durum literatürde geçici ısı transferi problemi olarak geçmektedir. Süreçte ısı transferinin, iş merdanelerinin iletim ve çevresinde meydana geldiği kabul edilir. Ayrıca merdanelerin dönüşünün etkisi ihmal edilebilir olarak kabul edilmekle birlikte, etkili olan parametreleri Sonlu Farklar Yöntemi ile ölçmeye çalışmışlardır. İncelediğimiz bu çalışmada, YSA modeli ile eğitici öğrenme uygulanmış, geçmiş veriler ile bir model oluşturulup bilinmeyen durumları sistemin anlaması için model eğitilmiştir. Model çevrimiçi olarak çalışarak sıcaklık değişimlerine istinaden ani sıcaklık yükselmesinin ne zaman olabileceğini öngörmekte ve soğutma sistemi buna göre devreye girmektedir (Jiang ve ark.,2014).

Alüminyum üretimini bu seviyede etkileyen merdanelerin, alüminyuma uyguladığı kuvveti etkileyen parametrelerin bu değeri nasıl etkilediği konusunda da YSA algoritmalarının kullanıldığı bir çalışmaya da literatürde rastlanılmaktadır (Bagheripoor ve Bisadi, 2013).

Üretim esnasında yassı ruloda sıcak haddelemede oluşan kamburluk, üretim verimliliğini yüksek derecede kötü etkilemektedir. Bu durum benzer şekilde ruloya da zarar vermektedir. Başka bir çalışmada bu konu üzerine girdi parametreleri ile kamburluğu kontrol etmek amacıyla YSA tabanlı bir algoritma geliştirilmiştir (Lee ve ark., 2018).

İncelenen diğer bir çalışma ise Alüminyum Kompozitlerde Yüzey Pürüzlülüğü ve Malzeme Kaldırma Oranının Taguchi Eşleştirilmiş Gri İlişkisel Analiz ve YSA Kullanılarak Tahmini üzerinedir (Thankachan ve ark., 2019).

Bir başka araştırma, alüminyum ve metal sektöründe ürün kalitesinin tahmin edilmesi üzerine yapılmıştır. Bu çalışmada çelik çubukların kalite sorunlarının denetimli ve denetimsiz makine öğrenimi teknikleri ile son kontrol öncesinde tahmini yapılmıştır. Denenen yöntemler içerisinde KNN yöntemi ile başarı elde edilebilmiştir (Lieber ve ark.,2013).

Litetarür araştırmasında bulunan son örnek ise, alüminyum döküm işlemi için kalite tahmini yapılan çalışma olmuştur. Bu çalışmada da bir otomotiv fabrikasında üretimi yapılan alüminyum vana gövdesi dökümünde ürün kalitesinin tespiti üzerinde çalışılmıştır. Döküm prosesinin daha iyi ürün kalitesi için bir tahmin algoritması geliştirmeyi amaçlanmıştır (Bak, Roy ve Son, 2021).





3. MATERYAL VE YÖNTEM

3.1. Kullanılan Materyal, Veri Setleri

Araştırmada, Alüminyum Yassı Mamul Rulo üretimi yapan bir fabrikaya ait, döküm prosesinin verisi kullanılmıştır. Alınan veri içerisinde, 63 farklı girdi değişkeni vardır. Bu değişkenlerin 2021 yılına ait 1 yıllık verisi sistemden alınmıştır. Çalışmada kullanılan veri büyüklüğü yaklaşık 2 terabayttır. Çalışmada kullanılan veri, 69 sütun ve 4.839.168 satırdan meydana gelmektedir. Kullanılan ilk ham veri, içerisinde Sayaç Bilgisi, Tarih ve benzeri kategorik değişkenler ile birlikte, ağırlıklı olarak nümerik değişkenler içermektedir. Çalışmada kullanılan veri setinin örnek bir kısmına aşağıdaki adresten ulaşılabilir:

[Flat_Roll_Data_Sample](<https://www.kaggle.com/datasets/alperaytatli/flat-roll-data-sample>).

3.2. Kullanılan Araç-Gereçler

Çalışmada kullanılan başlıca ekipmanlardan biri olarak, 1 Adet Intel(R) Core (TM) i7-7500U CPU @ 2.70GHz, 2.90 GHz işlemci, 8.00 GB RAM ve 64-bit işletim sistemine sahip bir bilgisayar kullanılmıştır. Çalışmada ayrıca Google Colab üzerinde, 83,5 gb sistem ve 40 gb GPU RAM'i olan NVIDIA A100 Tensor Core GPU ve 52 GB Sistem RAM'li PNY NVIDIA TESLA V100 GPU sahibi üst düzey gelişmiş bilgisayarlar da kullanılmıştır. Verilerin makinadan elde edilmesinde makineye ait yazılım ve PLC sistemleri kullanılmıştır.

Yapılan analizler Jupyter Notebook üzerinden Python kütüphaneleri kullanılarak gerçekleştirilmiştir. Kullanılan kütüphanelere genel olarak bakıldığında; bilimsel hesaplamalar için kullanılan Numpy, veri analizi ve manipülasyonu için kullanılan Pandas, veri görselleştirme için Seaborn, makine öğrenme algoritmaları için Scikitlearn, kod parçalarının çalışma sürelerini ölçmek için timeit, tarih ve saat işlemleri için datetime, çeşitli kümeleme algoritmaları, ayrık optimizasyon, nörolojik ağlar ve diğer algoritmalar için pyclustering, bilimsel ve teknik hesaplamalar için kullanılan scipy, Google tarafından geliştirilen ve açık kaynak kodlu olan derin

öğrenme algoritmaları için kullanılan Tensorflow, istatistiksel modelleme, veri analizi ve ekonometrik analizler yapmak için kullanılan statsmodel kütüphaneleri ile birlikte, uyarıları kontrol etmek için warning modülü kullanılmıştır. Sistemden alınan veriler excel dosyaları halinde alınarak Jupyter Notebook kütüphanesine atılmıştır.

3.3. Kullanılan Yöntemler

Veri Madenciliği, temel olarak, anlamlı bir veri üzerinden, bir konu ya da problem hakkında “Bilgi” elde etmektir. Bu bilginin elde edilebilmesi için, öncelikle verinin tekniklerin üzerinde uygulanabilir hale getirildiği bir veri ön işleme sürecinden geçmesi gerekir. Çünkü özellikle gerçek hayata dair problemlerde, kullanılan gerçek veri içeriğinde yüksek olasılıkla bir dizi problemi barındırır. Bu problemler temel olarak sensörlerden gelen verilerin aktarımı esnasında oluşabilecek problemler nedeniyle sistemin boş sürüm veri yada geçersiz sayı verisi ataması olabilir. Yada proses kaynaklı olarak, proseste anlamsız bir şekilde ilgili parametre için olmaması gereken bir veri bir parametreye yanlış aktarılmış da olabilir. Bu ve benzeri durumlar nedeni ile, veri ön işleme yapılması çok önemlidir.

Bu çalışmada da öncelikli olarak toplanan verilere ait veri ön işleme yapılmış, veri ön işlemesi yapılan veriler üzerinde veri madenciliği teknikleri uygulanmıştır. Sonrasında ise verinin sınıflandırılması için kümeleme ve sınıflandırma algoritmaları veri üzerinde uygulanmıştır. Çalışmanın devamında ise uygulanan algoritmaların performansı, literatürde belirtilen ölçüm yöntemleri ile kıyaslanmıştır. Son etapta ise olası çıktılar için en iyi tahmin algoritmaları belirlenmiştir.

Belirlenen en iyi tahmin algoritmaları, belirli bir süre sistem içerisinde de kullanılarak fiili durum test edilmiştir. Bu süre zarfında yaşanan değişimler ve elde edilen getiriler de ölçülerek çalışmanın getirisi hesaplanmıştır. Çalışmanın işleyişi Şekil 3.1’de gösterilmiştir.



Şekil 3.1. Çalışmanın akış şeması

3.3.1. Veri ön işleme

Veri, ön işleme aşamasına geçmeden önce çeşitli sorunlarla karşılaşabilir. Bu sorunlar arasında gereksiz veya kullanılmayan alanlar, eksik değerler, aykırı değerler, mantıksız değerler ve model için uygun olmayan girdi formatındaki veriler bulunabilir. (Akküçük, 2011).

Mantıksız değerler, çeşitli yollarla meydana gelebilir. Nitel ve nicel değişkenlerde, farklı biçimlerde mantıksız değerler girilmiş olabilir. Eğer bir veritabanında gelir

değişkeni negatif olarak kaydedilmişse ya da sayısal olması gereken yaş değişkeni bir harfle belirtilmişse bu durumda bu kayıta bir aksaklık olduğu düşünülebilir. Bu mantıksız değerlerin düzeltilerek eksik değere dönüştürülmesi bir çözümdür (Akküçük, 2011).

Öncelikle ham verinin, bu potansiyel problemlerden temizlenmesi gerekir. Veri ön işleme aşamasında ise, öncelikle verinin içerisindeki gürültü adını verdiğimiz verinin oluşumu esnasında oluşan hatalı yada tutarsız verilerin temizlenmesi gerekir. Aynı zamanda kullanılacak veri seti birden fazla kaynaktan temin edilen bir veri ise, bu kaynaklardaki verilerin de bir araya getirilmesi gereklidir (Han, Kamber ve Pei, 2011).

Oluşan veri setinde, bazı girdi değişkenler, başka girdi değişkenleri ile yüksek korelasyona sahip olabilir. Böyle durumlarda, özellikle boyutu yüksek veri setlerinde, birbiri ile yüksek ilişkili girdi değişkenlerinden bir tanesi diğerlerini temsilen alınabilir. Burada yüksek ilişki için değer, literature de uygun olacak şekilde korelasyon değeri 0,8 ve üzeri olan korelasyon değerleri yüksek değer olarak kabul edilmiştir (Statistics How To, Tarih Yok).

3.3.2. Kümeleme yöntemleri

Kümeleme analizi, nesnelerin (birimlerin) sahip oldukları özelliklere dayanarak gruplanmasını sağlayan çok değişkenli tekniklerin bir grubudur. Bu analizde, nesneler benzer özelliklere sahip oldukları için aynı kümeye, farklı özelliklere sahip oldukları için ise farklı kümeler arasına yerleştirilir. Başarılı bir kümeleme işlemi, nesnelerin birbirlerine yakın olduğu ve kümelerin birbirinden uzak olduğu bir geometrik yapı oluşturur (Hair ve ark., 1998).

Kümeleme analizinde, verilerin normal dağılıma sahip olduğu varsayımı yaygın bir kabuldür, ancak bu varsayım sadece prensipte geçerlidir. Uzaklık ölçümlerinin normal dağılıma sahip olması genellikle yeterli görülmektedir. Ayrıca, kümeleme analizinde kovaryans matrisiyle ilgili herhangi bir özel varsayım bulunmamaktadır (Tatlıdil, 1996).

Yapılan araştırmalarda, kümeleme analizi yaklaşımı ile alüminyum üretim süreçlerinde, alüminyum ergitmede hataların tespiti için yapılan bir çalışmanın belirli ölçüde başarılı olduğunu göstermiştir (Majid ve ark., 2012). Bu çalışmadan ilham alınarak kümelemenin çalışmamız üzerinde de ne derece etkisi olacağı görülmek istenilmiştir.

Bazı arařtırmalarda, normal dađılıma uymayan veri setlerinin faktör ve kümeleme analizleri öncesinde normal dađılıma yaklařtırılması için karekök ve logaritma dönüşümleri yapıldığı tespit edilmiştir (Sađlam, 2013).

K-Means kümeleme yöntemi, kümeleme algoritmaları içerisinde en eski olan algoritmalarındandır. Eğiticiz bir öğrenme algoritmasıdır. Büyük verilerde hızlı çalışması onu tercih edilen yaygın bir algoritma yapmıştır. Ancak küme sayısını algoritmanın başlangıcında uygulayıcıya bırakması gibi bir dezavantajı vardır (Takalođlu ve Takalođlu, 2019). Bu dezavantaj da bir nebze Elbow Yöntemi ile aşılmıştır. K küme sayısının belirlenmesinde kullanılan bu yöntem, her bir noktanın oluşturulan küme merkezlerine olan uzaklıklarının karelerinin toplamı ile hesaplanır. Bu yöntemle göre küme merkezlerine olan uzaklığının karelerinin toplamındaki deđişim miktarının azaldığı nokta dirsek noktasıdır ve bu dirsek noktası, en iyi k küme sayısını ifade eder (Aslanyürek ve Mesut, 2021). Genel Formül ifadesi ařađıdaki gibidir:

$$J(V) = \sum_{i=1}^k \sum_{j=1}^n ||x_{ij} - v_j||^2 \quad (3.1)$$

Clarans Kümeleme Algoritması, Clara algoritmasının bir türevidir. Clara algoritması, tüm veri kümesini deđil, küçük bir bölümünü temsil ederek PAM algoritmasını uygulamak için kullanılır. Bu temsilci sabittir. Clarans algoritmasında ise her seferinde farklı bir temsilci belirlenir ve daha iyi kümeleme sađlanması amaçlanır (Aslanyürek ve Mesut, 2021).

BIRCH Algoritması, özellikle çok büyük verilerde, hızlı ve hassas ölçüm yapan bir algoritma olarak sunulmuştur. Aynı zamanda bilgisayar hafıza kullanımı da oldukça düşüktür. BIRCH, her kümeleme kararının tüm veri noktalarını veya řu anda var olan tüm kümeleri taratmadan yapılması nedeniyle yereldir. Noktaların dođal yakınlığını yansıtan ölçümleri kullanır ve aynı zamanda kümeleme işlemi sırasında kademeli olarak korunabilir (Zhang, Ramakrishnan ve Livny, 1997).

Spektral Kümeleme Algoritması, düşük benzerliğe sahip farklı kümelerde bir benzerlik matrisinin bölme noktalarına özyapısına, aynı kümedeki noktaların yüksek benzerlik ve noktalara sahip olduđu ayırık kümelere dayanan bir teknik sınıfını ifade eder. Geleneksel küme algoritmalarının çođunlukla küme nesnelerinin belirli özelliklere sahip olduđunu ve bir dizi farklı bölge oluşturabileceđini ve dışbükey küresel örnek uzayına dayanabileceđini varsayması gerekir. Ancak örnek uzayı

dışbükey olmadığında, algoritma yerel optimumda sıkışacaktır. Bu sorunu çözmek için Spectral Kümeleme algoritması olarak bilinen yeni bir küme algoritması önerilmiştir. Spektral Kümeleme algoritmaları spektrum grafiği teorisine dayanmaktadır. Veri kümelemesini, veri kümelerinin biçimi hakkında herhangi bir varsayım yapmadan, bir grafik bölümlendirme problemi olarak ele alırlar, yani, veri kümelerinin kümelenebilir Laplacian matriksinin satır vektörüne eşlenmiştir ki bu da yumruk özellik vektörlerinden oluşur (Ding, Zhang ve Zhang, 2010).

Yoğunluk Tabanlı Gürültüden Ayırıştırma ve Kümeleme Algoritması, nesnelerin komşuları ile olan mesafelerini hesaplayarak belirli bir bölgede önceden belirlenmiş bir eşik değerden daha fazla nesne bulunan alanları gruplandırarak kümeleme işlemini gerçekleştirir. Sipariş Tabanlı Yoğunluk Kümeleme algoritması, Yoğunluk Tabanlı Gürültüden Ayırıştırma ve Kümeleme Algoritması'nın geliştirilmiş bir versiyonu olarak kabul edilebilir. (Bilgin ve Çamurcu, 2005).

HAC Algoritması, mantık olarak başlangıçta her bir veriyi bir küme kabul eder ve farklı özelliklerine göre bu kümeleri ikililer halinde birleştirerek ilerler. En sonunda ikili birleştirilebilecek küme kalmadığında ise algoritma çalışmayı durdurur. Verileri kümeler halinde birleştirmek için, verilerin herhangi bir özelliği seçilebilir. Ward algoritması, bunun için küme içi varyansı minimize eder. Her adımda mesafesi en az olan iki küme birleştirilir (Takaloğlu ve Takaloğlu, 2019).

3.3.3. Derin öğrenme ve makine öğrenmesi sınıflandırma yöntemleri

Bu bölümde, veri seti içerisinde yer alan yarı mamullere, ilk bölümde literatür araştırması sonucunda tespit edilen daha önce benzer sınıflandırmaya problemlerinde kullanılmış Derin Öğrenme ve Makine Öğrenmesi algoritmaları uygulanmıştır. Literatürde problem tespitinde kullanılan algoritmalar IF ve OCSVM algoritmaları veri seti içerisinde anomali tespitinde kullanılan eğitimci olmayan öğrenme algoritmalarıdır. Bizim örneğimizde herhangi bir anomali tespitine gerek yoktur. Çünkü veri ön işleme aşamasında Teknik limitler de dahil olmak üzere veri setinde aykırı değerler ayıklanmıştır.

Rastgele Orman yöntemi çoklu karar ağaçlarının birleşiminden oluşur. Breiman'ın geliştirdiği bu algoritmada tek bir ağaç oluşturmak yerine birkaç çok parametrelili kararın alınmasıyla bir ağaç oluşur. Bu yöntem, çok değişkenli karar ağaçları oluşturmak için sınıflandırma ve regresyon ağaçları algoritmasını kullanır. Nitelikler,

ağacın her seviyesi için hesaplanarak tanımlanır ve daha sonra bu nitelikler, karar ağacının niteliğini seçmek için birleştirilir, ve bu işlem ağacın her seviyesi için tekrarlanır. CART algoritması bilgi kazancını hesaplar ve ağacın dallanacağı parametrelere karar verir. Ağacı oluşturduktan sonra veriler bu ağaç modeline göre analiz edilir ve sınıflandırma veya kümeleme yapılır (Bakırarar ve ark., 2019).

İstatistikte LR, binom olarak dağıtılmış yanıt bağımlı değişkenler için bir regresyon modelidir. Diğer faktörlerin bir fonksiyonu olarak meydana gelen bir olayın olasılığını modellemek için kullanışlıdır. LR, tıp ve sosyal bilimlerde yaygın olarak kullanılmaktadır. Diğer çeşitli uygulama alanlarında kullanılan LR için diğer isimler arasında lojistik model, logit modeli ve maksimum entropi sınıflandırıcı bulunur. LR, bağımlı bir değişkeni sürekli ve temelinde tahmin etmek veya kategorik bağımsız değişkenler ve bağımsız değişkenler tarafından açıklanan bağımlı değişkendeki varyans yüzdesini belirlemek, bağımsız değişkenlerin göreceli önemini sıralamak, etkileşim etkilerini değerlendirmek ve kovaryasyon kontrol değişkenlerinin etkisini anlamak için kullanılabilir (Chandra, Ravi ve Bose, 2009).

KNN algoritması, test seti ile eğitim seti arasındaki uzaklıkları ölçüt kabul ederek, test verisine en yakın k adet örneği baz değer kabul eder. Baz değerlere en yüksek sınıflandırma doğruluğu veren parametreler elde edilir. Hızlı çalışan bir algoritmadır (Takci, 2022).

YSA'lar başlangıçta insan beyninin bilgi işleme yeteneklerini taklit etmek için tasarlanmış bilgisayar sistemlerinin bir bileşeni olarak tasarlanmıştır. Günümüzde YSA algoritmaları, sadece örüntü tanıma, optimizasyon ve tahmin gibi alanlarda değil, bu alanlarla sınırlı olmamak üzere geniş bir uygulama yelpazesinde yaygın olarak kullanılan bir algoritmik çerçeve haline gelmiştir. YSA'lar, verilerden öğrenme ve bu öğrenmeyi yeni durumlara genelleme yetenekleri nedeniyle yaygın olarak kullanılmaktadır. YSA'lar böylece çeşitli alanlardaki karmaşık problemleri çözmek için önemli bir araç haline gelmiştir. Bir YSA tipik olarak üç katmandan oluşur: bir giriş katmanı, birden fazla gizli katman ve bir çıkış katmanı. Bir YSA'nın işleyişi şu şekilde özetlenebilir: veriler giriş katmanı tarafından alınır, bu daha sonra ilk gizli katmana aktarılır; gizli katmandaki her nöron, giriş verilerinin ağırlıklı bir doğrusal kombinasyonunu içeren bir hesaplama gerçekleştirir, ardından aktivasyon fonksiyonu gelir ve elde edilen çıktı bir sonraki gizli katmana yayılır; bu işlem, çıktı katmanına ulaşılan kadar yinelemeli olarak tekrarlanır, bu da YSA'nın son tahminini veya

çıkışını üretir. Ağın yapısı ışığında, YSA hesaplamaları genel olarak sinir ağlarına ve derin öğrenme algoritmalarına ayrılabilir. FFNN olarak da bilinen geleneksel sinir ağı, yalnızca bir gizli katman içerir. Tersine, Derin Öğrenme algoritmaları, Tekrarlayan Sinir Ağı, Derin Sinir Ağı ve Evrimsel Sinir Ağı'nı içerir. FFNN, bilginin giriş düğümlerinden çıkış düğümlerine, var olabilecek herhangi bir gizli düğümden geçerek yalnızca bir yönde aktığı temel ve en basit YSA tasarımıdır. RNN, mimarisi FFNN'den türetilen bir sinir ağı biçimidir. Geçerli gizli katmanda girdi olarak kullanılmak üzere önceki gizli katmanlardan gelen verileri saklayan bellekleri, ayırt edici özellikleridir. Basit bir ifadeyle, geçerli durumun tahmini girdi ve önceki durum tarafından yapılır. Bu bellek yeniliği ile RNN'ler, sıralı girdiyi işlemek için özel bir yeteneğe sahip güçlü modeller haline gelir. Bununla birlikte, sıralı girdileri işleme yeteneklerine rağmen, RNN'ler uzun vadeli zaman dizileri için uzun vadeli bağımlılık probleminin zorluğuyla karşı karşıya kalır ve bu da gradyanın kaybolmasına neden olabilir. Bu sorunun üstesinden gelmek için, kontrollü depolama kullanan LSTM yaklaşımı önerilmiştir.

LSTM katmanları üç kapı içerir: giriş kapısı, unut kapısı ve çıkış kapısı (Kurucan ve ark.,2024). LSTM, YZ yöntemleri arasında uzun süreli bağımsızlık ve kısa vadeli ilişki için daha iyi tahmin yeteneği gösterdiğinden, giderek dikkat çekmiş ve kalan faydalı ömür tahmininde (Zhou, Zhao ve Gao, 2019), aynı zamanda anomali tespitinde (Li, 2019) kullanılmıştır.

NB algoritması bize önceki olasılık ve koşullu olasılıkları tek bir formülde birleştirmenin bir yolunu vermektedir. Böylelikle sırayla olası sınıflandırmaların her birinin olasılığını hesaplamak için bunu kullanabilmektedir. Pratikte bazı durumlar böyle olmayabilir. Bu teorik zayıflığa rağmen, NB yöntemi pratik kullanımda genellikle iyi sonuçlar vermektedir. Yöntem koşullu olasılıklar kullanmaktadır (Bramer, 2013). Güçlendirme, bir yöntemler ailesini kapsar. Bu yöntemlerin odak noktası bir dizi sınıflandırıcı üretmektir. Serinin her üyesi için kullanılan eğitim seti, serideki önceki sınıflandırıcının yada sınıflandırıcıların performansına göre seçilir.

Güçlendirme'de, serideki önceki sınıflandırıcılar tarafından yanlış tahmin edilen örnekler, doğru tahmin edilen örneklerden daha sık seçilir. Böylece Güçlendirme, mevcut topluluğun performansının zayıf olduğu örnekleri daha iyi tahmin edebilen yeni sınıflandırıcılar üretmeye çalışır. Her bir katlama test edildikten sonra, en iyi sınıflandırıcı tarafından yanlış sınıflandırılmış olarak bırakılan numuneler tekrar başka

bir sınıflandırıcı ile ayrı ayrı test edilir, böylece numuneyi doğru bir şekilde sınıflandırmak için bir şans verir (Chandra, Ravi ve Bose, 2009).

Topluluk sınıflandırması fikri, sadece bir sınıflandırıcıyı değil, sınıflandırıcı topluluğu olarak adlandırılan bir dizi sınıflandırıcıyı, sonrasında bir tür oylama kullanarak görünmeyen örneklerin sınıflandırılması için tahminlerini birleştirmek için öğrenmektir. Topluluğun toplu olarak bireysel sınıflandırıcılardan herhangi birinden daha yüksek bir tahmin doğruluğu seviyesine sahip olacağı umulmaktadır, ancak bu garanti edilmez (Bramer, 2013).

İlgili algoritmalar, mevcut veri tabanlarına uygulanmadan önce, Pima Indian Veri Çerçevesi üzerinde daha önce yapılan çalışmalardan KNN algoritması (Kaggle, 2019), LR algoritması (Smith ve ark., 1988), (Kaggle, 2017), SVM algoritması (Kaggle, 2021) ve DT algoritması (Khanam ve Foo, 2021) ile test edilmiştir ve yapılan test sonucunda ilgili algoritmaların doğru çalıştığı gözlemlenmiştir.

3.3.4. Değerlendirme metrikleri

3.3.4.1. Karmaşıklık matrisi

Karmaşıklık matrisi veya hata matrisi (Stehman, 1997), bir modelin performansını ölçmede önemli bir değerlendirme yöntemidir. Tablo 3.1. karmaşıklık matrisini göstermektedir.

TP ve TN modelin doğru olarak tahmin ettiği, FP ve FN ise modelin yanlış olarak tahmin ettiği değerlerdir (Powers, 2007).

Tablo 3.1. Karmaşıklık matrisi

	Tahmin Edilen	
	TP FP	FN TN

- TP : Doğru Pozitif. Pozitif olarak tahmin edilen ve gerçekte de pozitif olan değerlerdir.
- TN: Doğru Negatif. Negatif olarak tahmin edilen ve gerçekte de negatif olan değerlerdir.
- FP (Tip1 Hata): Hatalı Pozitif. Pozitif tahmin edilen ama gerçekte negatif olan değerlerdir.
- FN (Tip2 Hata): Hatalı Negatif. Negatif tahmin edilen ama gerçekte pozitif olan değerlerdir.

Doğruluk: Doğruluk değeri modelde doğru tahmin edilen değerlerin veri kümesindeki toplam değerlere oranı ile hesaplanmaktadır.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.2)$$

Kesinlik: Pozitif olarak tahmin edilen değerlerin gerçekte de kaç tanesinin pozitif olduğunu göstermektedir.

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (3.3)$$

Duyarlılık: Pozitif olarak tahmin edilmesi gereken değerlerin ne kadarının pozitif olarak tahmin edildiğini gösteren bir metriktir.

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (3.4)$$

F1-skor: F1-skor değeri, kesinlik (Precision) ve duyarlılık (Recall) değerlerinin harmonik ortalamasını göstermektedir.

$$F_{1\text{-skor}} = 2 * \frac{\text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (3.5)$$

3.3.4.2.Siluet (Silhouette) skoru

Kümelenen verilerin bulunduğu kümedeki uygunluğunu bulmak için geliştirilen Siluet indeksi aşağıdaki eşitlik ile hesaplanır:

$$S = \frac{(b - a)}{\max(a, b)} \quad (3.6)$$

a: bir örnek ile aynı kümedeki diğer tüm noktalar arasındaki ortalama mesafe.

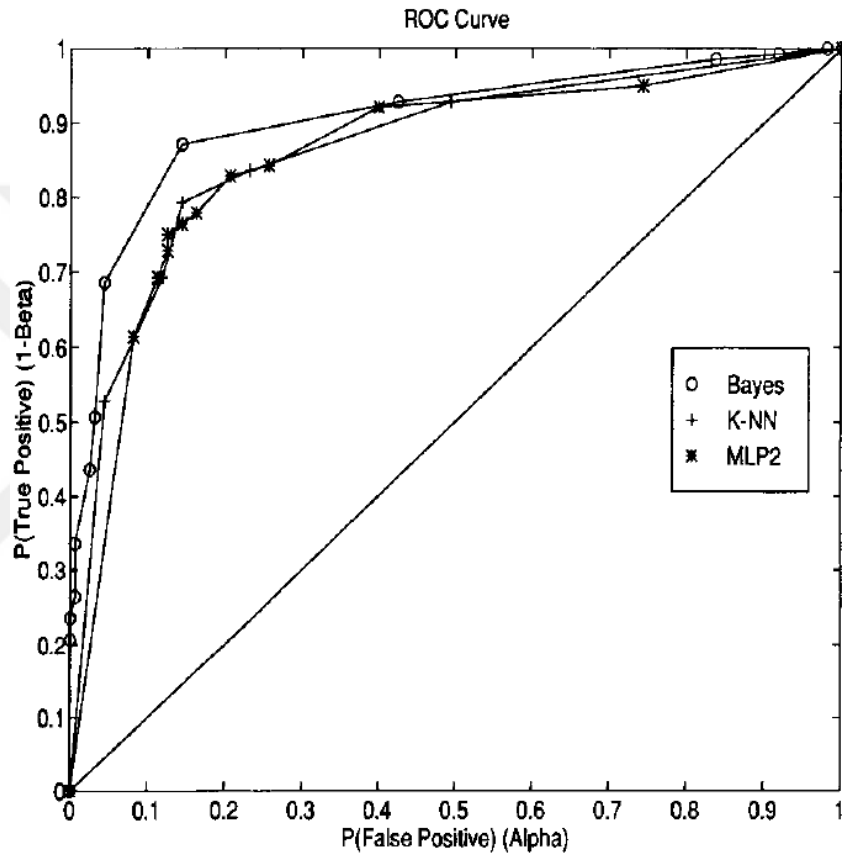
b: bir örnek ile en yakın kümedeki diğer tüm noktalar arasındaki ortalama mesafe.

S: Siluet İndeksi

Elde edilen S değeri -1 ile 1 arasında bir sayıdır. 1'e yakın değerler kümelemenin iyi olduğunu ifade eder (Aslanyürek ve Mesut, 2021).

3.3.4.3. AUC-ROC eğrisi

ROC Eğrisi sınıflandırma problemlerinde, algoritmanın performansını değerlendirmek için kullanılan bir araçtır. Uygun bir çalışma noktası veya karar eşiği seçmek için bir sınıflandırıcının performansını görselleştirmenin iyi bir yoludur. Bu araç, Bölüm 3.2’de de bahsettiğimiz TP ve FP değerlerini Şekil 3.2’de de gösterildiği gibi kullanır. AUC değeri, yani başka bir değişle eğrinin altında kalan alan ne kadar artarsa, modelin performansı o derece iyi demektir.



Şekil 3.2. ROC eğrisi (Bradley, 1997)

Bölüm 3.3.4.1’de yazdığımız Karmaşıklık Matrisi’ni daha açık bir ifade ile Tablo 3.2’de aşağıdaki gibi yazarsak;

Tablo 3.2. ROC eğrisi karmaşıklık matrisi

		Tahmin		Toplam
		-ve	+ve	
Gerçek	-ve	T_n	F_p	C_n
	+ve	F_n	T_p	C_p
Toplam		R_n	R_p	N

T_n ve T_p : Doğru tahmin edilen değerlerin sayısı

F_n ve F_p : Yanlış Tahmin edilen değerlerin sayısı

C_n ve C_p : Gerçek Negatif ve Pozitif değerlerin sayısı

R_n ve R_p : Tahmin Negatif ve Pozitif değerlerin sayısı

$$N = C_p + C_n + R_n + R_p \quad (3.7)$$

Burada amaç, yanlış sınıflandırma maliyetini (C) minimize etmektir. Bu maliyet;

$$C = F_p * C_{F_p} + F_n * C_{F_n} \quad (3.8)$$

C_{F_p} : F_p maliyeti

C_{F_n} : F_n maliyeti

Ne yazık ki, burada C_{F_p} ve C_{F_n} maliyetleri kısmen biliriz. Bu nedenle hesaplama gerekli yanlış pozitif ve yanlış negatif oranlar, $P(F_p)$ ve $P(F_n)$ cinsinden belirtilir. $P(F_n)$ ve $P(F_p)$ gereksinimlerini en yakından karşılayan çalışma noktasını (karar eşiği) seçilir. Karar eşiği değiştikçe çizilen $P(T_p)$ ve $P(F_p)$ değerlerine ROC eğrisinin kendisidir. Genel formül olarak ROC eğrisinin altında kalan AUC alanı formülü ise şu şekildedir;

$$AUC = \sum_i \left\{ (1 - \beta_i * \Delta\alpha) + \frac{1}{2} [\Delta(1 - \beta) * \Delta\alpha] \right\} \quad (3.9)$$

ki burada;

$$P(F_p) = \alpha \text{ ve } P(T_p) = 1 - \beta \quad (3.10)$$

olarak verilir (Bradley, 1997).

3.3.4.4. Algoritma çalışma süresi

Makine öğrenimi dünyasında eldeki algoritmalar içerisinde veri için en doğru algoritmanın seçilmesi noktasında kullanılacak algoritmanın çalışma süresi her zaman göz önünde tutulmalıdır. Bazı algoritmalar, çözümü bulabilmek için günlerce çalışma gerektirebilmektedir. Ancak ideal çözüm için istenilen durum, bu süreyi

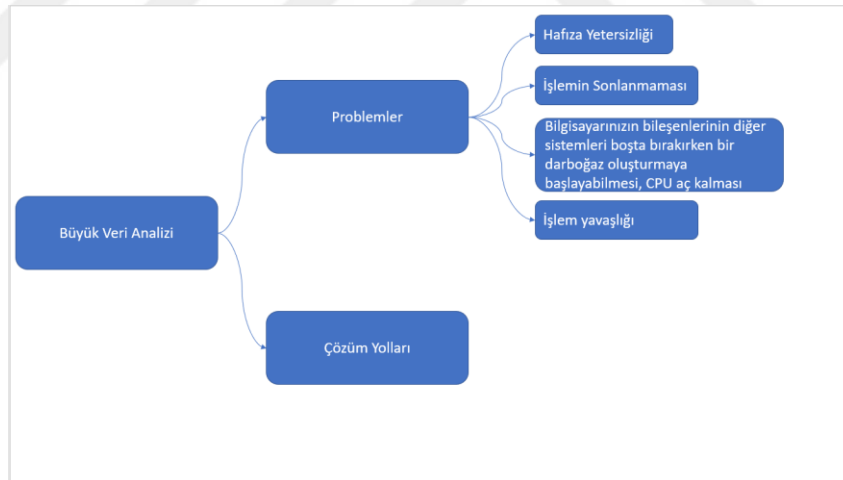
bekleyemeyecek koşullar içerebilir. Buna ek olarak, uzun süreli çalışmak ilgili problemin çözümünde zaman kısıtı nedeniyle çok maliyetli olabilir.

Literatürde de algoritmaların performanslarının kıyaslanması noktasında, bir performans ölçütü olarak bu parametrenin kullanımına örnekler mevcuttur. Kimi çalışmalar bu karşılaştırmayı algoritmaları eğitim ve test sürelerinin karşılaştırması (Carlin, Kane ve Sezer, 2019) şeklinde yaparken, kimileri algoritmaların çalışma sürelerinin direk karşılaştırılması şeklinde yapmaktadır (Cavallaro ve ark., 2024).

3.3.5. Veritabanı analiz yöntemleri

Veri miktarı çok yüksek olduğu durumlarda, RAM yetersizliğinden dolayı analiz yapılamamaktadır. Literatür bu durumu ve bu durumda yapılması gerekenleri özetlemiştir.

Büyük veri ile ilgili yaşanan problemlerin ana sebepleri, aşağıdaki Şekil 3.3’de gösterildiği gibi sıralanmıştır. Bunlardan en önemlilerinden, en sık karşılaşılanlarından biri de algoritmaların çalışması esnasında yaşanan hafıza yetersizliğidir (Cielen ve ark., 2016).

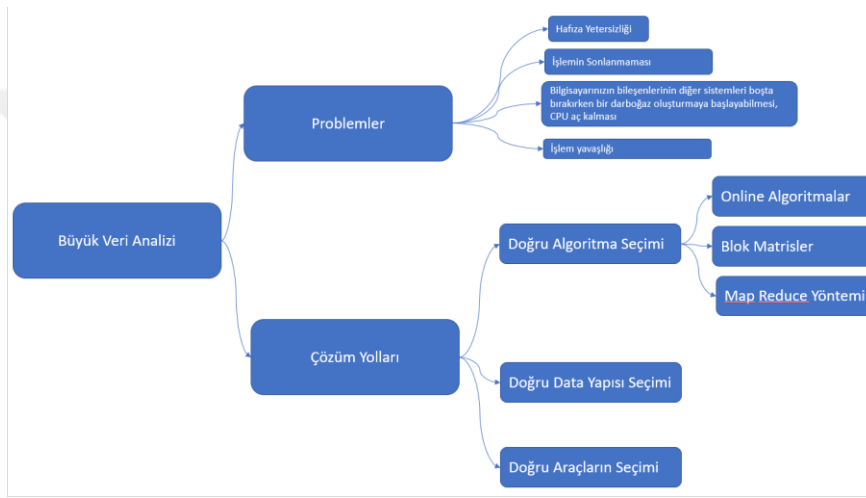


Şekil 3.3. Büyük veri işleme yaşanan ana problemler (Cielen ve ark., 2016)

Bir bilgisayarda yalnızca sınırlı miktarda RAM bulunur. Bir belleğe gerçekte sığdığından daha fazla veri sıkıştırmaya çalıştığınızda, işletim sistemi bellek bloklarını disklere aktarmaya başlamaktadır, bu da hepsini bellekte tutmaktan çok daha az verimlidir. Ancak büyük veri kümelerini işlemek için yalnızca birkaç algoritma tasarlanmıştır. Algoritmaları çoğu, tüm veri kümesini bir kerede belleğe yükler, bu da

bellek yetersiz hatasına neden olmaktadır. Diğer algoritmaların, verilerin birden fazla kopyasını bellekte tutması veya ara sonuçları saklaması gerekmektedir.

Aynı kaynak, bu problemlere verilebilecek olası çözüm yollarını da Şekil 3.4’deki gibi şematize etmiştir. Burada bahsedilen çözüm yollarından bir tanesi de veri Doğru Data Yapısı Seçimi başlığı altında şu şekilde geçmektedir. Burada örneğin, büyük bir veri tablosunu küçük matrislere bölerek doğrusal bir regresyon yapabiliriz denilmektedir. Yani veri yapısı işlem için çok büyük olduğunda, veriyi anlamlı küçük parçalara ayırmak da önerilen yöntemlerden biri olmuştur (Cielen ve ark., 2016)

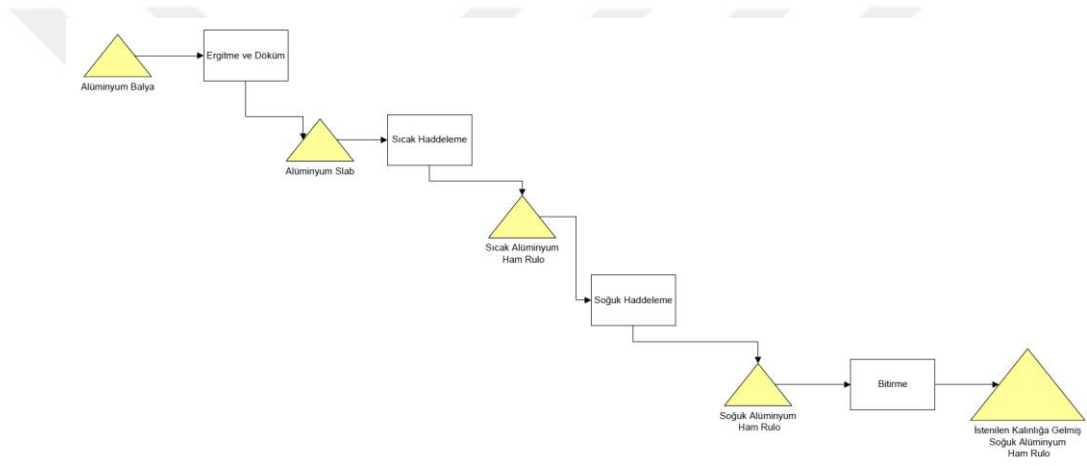


Şekil 3.4. Ram kısıtı problemine olası çözüm yolları (Cielen ve ark., 2016)

4. ARAŞTIRMA BULGULARI

4.1. Proses Özeti

Sürece temel olarak baktığımızda, süreç, alüminyum hammaddesinin, alüminyumun eritilmesi ve dökümü sonrası önce sıcak sonra da soğuk haddelemeye uğraması ve Bitirme dediğimiz bölgede temperleme, tesviye, yüzey işlem ve kaplama proseslerinden geçmesi sonrasında, son olarak nihai ürüne dönüşmesidir. Tüm proses akışı Şekil 4.1’de gösterildiği gibidir.



Şekil 4.1. Alüminyum yassı mamul üretim süreci

Sürecini incelemekte olduğumuz Döküm Hattı ise Şekil 4.2’de görülmektedir.



Şekil 4.2. Döküm hattı

4.2. Yassı Mamul Alüminyum Ürünlerinde Yapılan Kalite Kontrol Testleri

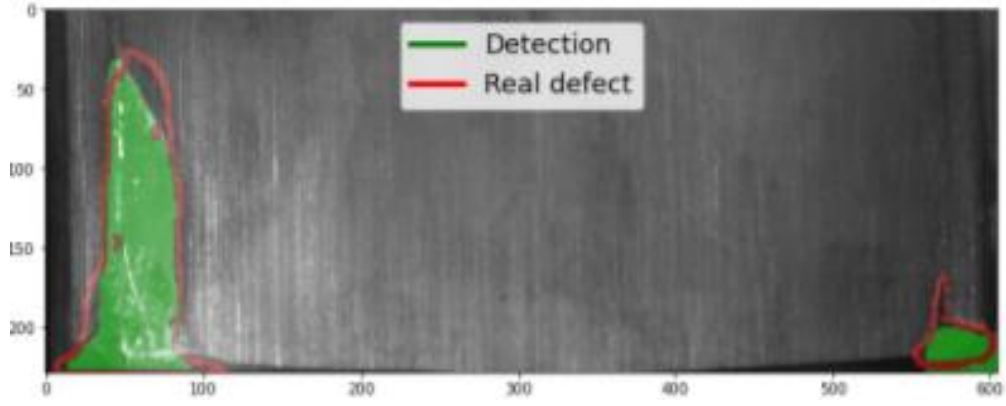
Ürün kalitesi, bir ürünün hedefine uygun olarak kullanılması için gerekli olan ürün özelliklerinin toplamı olarak tanımlanmaktadır. Yassı çubuğun mekanik özellikleri ve geometrik sıfatları, haddeleme için bu tür özelliklerdir (Mazur ve Koinov, 2016).

Sıcak haddelemede bilgisayar kontrolü 1960'ların başında geliştirilmiş ve fiilen faaliyete geçmiştir. Günümüzde hemen hemen tüm haddehaneler bilgisayarlar tarafından kontrol edilmektedir. Bilgisayarın fonksiyonlarının sınırlı olduğu ilk dönemlerden itibaren, artan gelişmelere paralel olarak çeşitli iyileştirmeler ve fonksiyonel genişlemeler yapılmıştır. Geline nokta sistem, verimlilik, yüksek kalite, işçilikten tasarruf, kaynak tasarrufu, enerji tasarrufu gibi ihtiyaçların karşılandığı yüksek verimli bir sistemdir (Kitao, Hirose ve Hamada, 1981).

Metal üretiminin ana üreticileri, yapılan üretimin belirtilen kalite derecelendirmelerini alıcıya garanti eder ve belgeler. Bu garantinin verilmesinde, teknolojiye geline nokta otomatik kontrol sistemlerinin gelişimi ve teknolojik sürecin kontrolünün geldiği noktanın etkisi büyüktür.

Genel itibariyle yassı mamul metalin kalitesi, yüzey kalitesi ile ilişkilidir. Bir diğer problem de bu kalitenin, sürecin ilerleyen aşamalarında (soğuk haddeleme, tavlama vs. sonrası) tespit edilmesinden dolayı geç aksiyon alınmak durumunda kalınmasıdır. Bu tür tutarsızlıkların tespit edildiği süreç aşaması ise daha derine inerse, reddedilmelerin neden olduğu kayıplar daha derin olmakta ve bu nedenle tüm süreç aşamalarında kontrol sistemlerinin devreye alınması gerekliliği ortaya çıkmaktadır. Genel olarak kalitenin tespitinde yüzey pürüzlülüğüne bakılmakta, burada da görsel tarama sistemleri kullanılmaktadır (Mazur ve Koinov, 2016).

Şekil 4.3, bir görsel kontrol sisteminin yakaladığı kalite problemlerini göstermektedir. Görüldüğü gibi kırmızı işaretlenen bölgeler, yüzeyde özür yakalanmış alanlardır (Han, Kamber ve Pei, 2011).



Şekil 4.3. Bir görsel kontrol sisteminin yakaladığı yüzey kalite özrü

Geleneksel olarak metal endüstrisinde boyutsal kalite kontrol ve diğer birçok kalite kontrol görevi, insanlar tarafından gerçekleştirilmiştir. Uzman teknisyenler, bir ürünün belirli bir özellik setinden sapıp saptığını görsel olarak belirlemek için eğitilir. Bununla birlikte, yetenekli teknisyenler bile zaman zaman yorgunluk nedeniyle önyargılı yanıtlar verebilir. Bu durum ürün kalitesinde olduğu gibi ilerleyen bir üretim hattında ürünün kalitesinin ne olacağının tahmin edilmesinde de bu şekilde yürür.

Alüminyum Yassı Mamullerin kalite tipinin belirlenmesi noktasında son kontrolde uygulanan birçok test vardır. İlgili testlerin genel çerçevede neler oldukları Tablo 4.1.'de gösterilmiştir.

Mamulun kalite tipinin ne olacağına bakılan öncelikle ilk parametre, yarı mamulun, yarı mamul reçetesinde belirtilen alaşımda, sahip olması gereken kimyasal kompozisyon ne ise, ona sahip olup olmadığıdır. Alüminyum alaşımlarının kimyasal kompozisyonları ürünün reçetesine göre belirlenmekle birlikte, genel olarak alaşımların olması gereken değerleri ASTM B209 standartlarına göre belirlenir (B07 Comitee, 2022).

Alüminyum ürünün kalite tipini belirlerken bazı mekanik testler uygulanır. Bu testlerden çekme testi, ASTM B557 standartını kullanarak istenilen çekme mukavemeti, akma mukavemeti ve uzama değerlerine göre ürünün kalite tipinin sınıflandırılmasında yardımcı olur (Singh, 2021). Mekanik testlerden olan Sertlik testi ise, Brinelli Sertlik indeksini kullanır (Rajesh ve ark., 2021). Ürünün istenilen sertlik derecesine yakınlığına göre kalite tipini belirler.

Alüminyum mamul kalitesinin tipini belirlemede, mamulün yüzey kalitesi de önemli bir parametredir. Günümüzde yüzey kalitesi konusunda, kalite kontrolleri yapan bazı

online sistemler bulunmaktadır. Örneğin Şekil 4.4’de de görülebildiği gibi PSI firmasına ait sistem, Alüminyum Yassı Rulo’ ya ait veriyi anlık olarak çevrimiçi takip etmektedir (PSI, 2021).

Alüminyum yüzey kalitesini belirlemede temel olarak bakılan konular; yüzeyde portakallanma, lüders çizgileri, looper çizgileri ve çatlamalardır. Prosesin sonucu olarak bazı durumlarda dış yüzeylerdeki aşırı deformasyon çatlamaya ve portakal yüzeyi gibi pürüzlü bir yüzeyin oluşmasına neden olur. Oluşan Lüders çizgileri ise, genellikle levhaların derin çekilmesinde karşılaşılan bir durumdur. Levhaların akış yönünde uzaması nedeniyle levhanın akışı yönünde yüzeyde düz çizgi şeklinde pürüzlülüklerdir. Derin çekmede karşılaşılan yüzey hatalarından bir diğeri olan looper çizgileri ise, yüzeyde oluşan halka biçimli izler olarak tanımlanmaktadır (Alper, 2003).

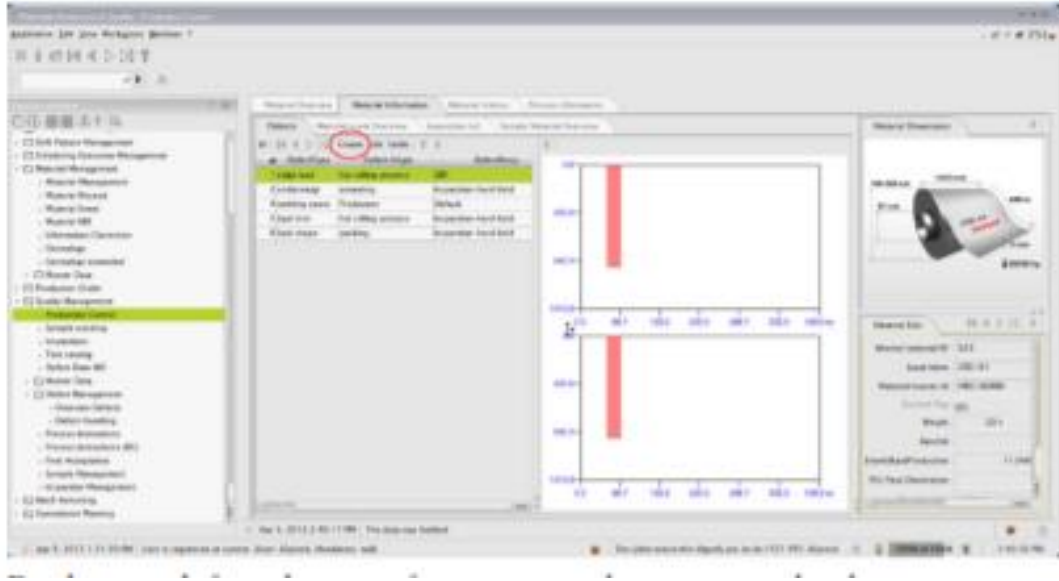
Üretilen mamulün, istenilen kalınlık değerlerinde olup olmaması da kalite tipini belirleyici diğeri bir faktördür. İstenilen değer genel olarak mamulün cinsine göre değişiklik gösterebilir. Ancak genel olarak bakıldığında ideal olarak istenilen kalınlığa göre $\pm 0,12$ mm sapma ile üretilmesi istenilen durumdur (B07 Comitee, 2022).

Bağlı bulunduğu mamulün spesifik değerlerine ve alaşımına bağlı olarak değişkenlik göstermekle birlikte, iyi bir kalite tipine haiz bir mamulde yapısal bir homojenlik ve tane boyutu beklenir. Buradaki kontroller ASTM E3 standartında belirtilen ilkeler doğrusunda gerçekleştirilir (E04 Committee, 2023).

Alüminyum alaşımı, klorür içeren ortamlarda korozyona karşı çok hassas bir metaldir. Bu nedenle elde edilen mamuller üzerinde tuz testi uygulanarak korozyon oranları ölçülür. Kısaca açıklamak gerekirse, tuz püskürtme testi, kapalı bir ortam içerisine konulmuş alüminyum metal üzerine belirli bir süre içerisinde tuz çözeltisinin bir tuz bulutu gibi uygulanmasıdır. Test için uygulanan süre, testin uygulanışına göre değişkenlik gösterebilir. Ancak iyi bir korozyon oranına sahip, yani iyi bir kalitede bir mamul için genel olarak 500 saatlik bir tuz püskürtme testi neticesinde %5 korozyon oranına sahip bir mamul denilebilir (Seetharaman, Ravisankar, ve Balasubramanian, 2015).

Tablo 4.1. Kalite tipi parametreleri

Test Ana Grup	Kalite Belirleme Parametreleri	Testin Amacı	Ürün Kalitesi		
			1. Kalite	2. Kalite	3. Kalite
Kimyasal Kompozisyon Analizi	İstenilen Kimyasal Kompozisyona Uygunluk Testi	Yarımamulun reçetesinde belirtilen değerlere ürün uygun mu?	Toleranslar Dahilinde	Sapma<±%5	Sapma>±%5
Mekanik Testler	Çekme Mukavemeti	Yarımamul türüne göre istenilen Çekme Mpa değerinden sapma	Sapma < ±10 Mpa	±10 Mpa<Sapma < ±20 Mpa	Sapma > ±20 Mpa
Mekanik Testler	Akma Mukavemeti	Yarımamul türüne göre istenilen akma mukavemeti mpa değerinden sapma	Sapma < ±10 Mpa	±10 Mpa<Sapma < ±20 Mpa	Sapma > ±20 Mpa
Mekanik Testler	Uzama Yüzdesi	Yarı mamul türüne göre istenilen Uzama Yüzdesi değerinden sapma	<%1	>%1	>%1
Mekanik Testler	Sertlik Testi	Yarımamul türüne göre istenilen sertlik değerinden sapma	<5HB	>5HB	>5HB
Yüzey Kalitesi ve Görsel İnceleme	Portakallanma	Yüzeyde Portakallanma var mı?	Hiç Yok	Aralıklı Olarak Var	Yoğun Derecede Var
Yüzey Kalitesi ve Görsel İnceleme	Lüders Çizgileri	Yüzeyde Lüders Çizgileri var mı?	Hiç Yok	Aralıklı Olarak Var	Yoğun Derecede Var
Yüzey Kalitesi ve Görsel İnceleme	Looper Çizgileri	Yüzeyde Looper Çizgileri var mı?	Hiç Yok	Aralıklı Olarak Var	Yoğun Derecede Var
Yüzey Kalitesi ve Görsel İnceleme	Çatlamalar	Yüzeyde Çatlamalar var mı?	Hiç Yok	Aralıklı Olarak Var	Yoğun Derecede Var
Boyutsal Kontrol	Kalınlık Toleransı	Yarımamul türüne göre istenilen kalınlık değerinden ± mm Sapma	±0,12 mm	<±0,20 mm	> ±0,20 mm
Mikro Yapı Analizi	Tane Boyutu	Tane boyutu ve yapısal homojenlik mevcut mu?	Mevcut	Kısmen Kötü	Kötü
Korozyon Testleri	Korozyon Oranı	Yapılan tuz püskürtme testi sonrasında korozyon oranı belirli bir seviyenin altında mı?	<%5	<%10	>%10



Şekil 4.4. Yassı rulo takip mekanizması

4.3. Analizler

4.3.1. Veri ön işleme

Çalışmanın yapılmış olduğu veri setine bakıldığında, veri setinin, toplamda 24 farklı çeşit yarı mamulden oluştuğu görülmektedir. Çalışmada bu yarı mamuller, yarı mamul kodları B01'den B24 'e kadar numaralandırılarak kullanılmıştır.

Veri setinin toplu halde incelenmesine bakıldığında, ilk bakışta bazı yarı mamullerin yılda 1 defa üretime girip çıktığı bazılarının ise yıl içerisinde farklı zamanlarda birden fazla defa üretildiği söylenebilir. Genel olarak üretim mantığı içerisinde bir ürün üretime girdiğinde değişim sürelerini en düşüğe tutabilmek için uzun soluklu üretim yapılmaya çalışılmaktadır. Ancak üretim planlamaya gelen siparişler doğrultusunda kısa süreli aralıklı üretimler de yapmak durumunda kalındığı olmaktadır. Bu durumu da minimize etmek için, ürünlerin yıl içerisinde uzun soluklu üretimi yapılarak, mümkün olan seviyede mevcut durumda ürünlerden emniyet stok tutulmaktadır.

Yarı mamullerin üretimi yapıldığı dönemler Tablo 4.2'de sunulmuştur. İlgili yarı mamullerin üretildiği üretim döküm hattı dışında, fabrika bünyesinde farklı yarı mamuller üreten, 2 ayrı üretim hattı daha bulunmaktadır.

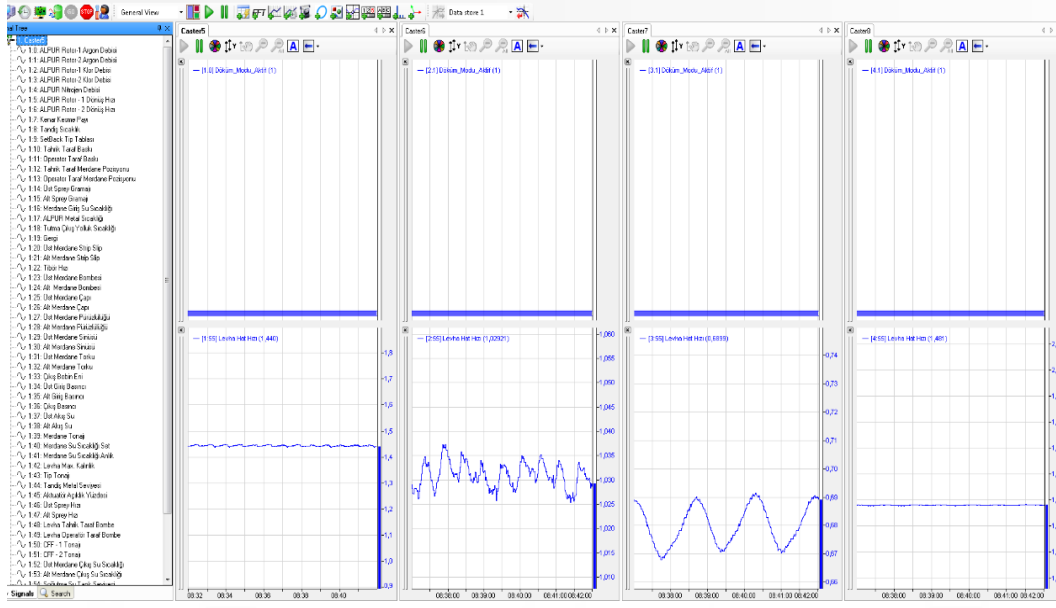
Tablo 4.2. Yarı mamullere ait üretim dönemleri

Üretildiği Ay (2021)	Yarı Mamül	Üretildiği Ay (2021)	Yarı Mamül
Ocak	B01	Nisan	B19
Ocak	B02	Mayıs	B20
Ocak	B03	Mayıs	B21
Şubat	B04	Haziran	B20
Şubat	B05	Haziran	B21
Şubat	B06	Temmuz	B22
Şubat	B07	Temmuz	B10
Şubat	B08	Temmuz	B23
Şubat	B09	Ağustos	B22
Şubat	B10	Ağustos	B10
Şubat	B24	Ağustos	B23
Şubat	B11	Eylül	B03
Şubat	B08	Eylül	B04
Şubat	B07	Eylül	B11
Şubat	B08	Eylül	B12
Şubat	B24	Ekim	B03
Şubat	B07	Ekim	B04
Şubat	B08	Ekim	B11
Şubat	B12	Ekim	B12
Mart	B13	Kasım	B01
Mart	B14	Kasım	B02
Mart	B15	Kasım	B04
Mart	B16	Aralık	B05
Mart	B17	Aralık	B06
Nisan	B18	Aralık	B09

Çalışmanın başlangıç aşamasında, verilerin toparlanması, düzenlenmesi ve bu süreci takiben eksik verilerin ve aykırı verilerin düzenlenmesi işlemleri yapılmıştır. Yapılan işlemlerin detayı, alt maddeler olarak aşağıda açıklanmıştır.

4.3.2. Toplanan verilerin düzenlenmesi

Bu sürecin ilk başlangıcı olan döküm sürecinde, makinamızdan sensörler vasıtası ile anlık veri alınmaktadır. Süreç içerisinde, çıktının kalitesine etkisi olduğu ve takip edilmesinin gerekli olduğuna karar verilmiş parametrelerin anlık takibi, Şekil 4.5’de ekran görüntüsü de görülen IBA sistemi ile yapılmaktadır.



Şekil 4.5. Girdi parametrelerinin takip edildiği iba sistemi

Öncelikli olarak IBA sisteminin 2021 yılı başından sonuna kadar, Tablo 4.3’de de görülebilecek olan girdi değerler için ürettiği değerler sistemden excel dosyalarına aktarılmıştır.

İlgili veri tek bir excel dosyasına sığamayacak kadar büyük bir veridir. Bu nedenle sistemden alınan veri, parçalar halinde excel dosyaları olarak alınmıştır. Alınan excel dosyaları öncelikle açılarak yüzeysel olarak dosyalarda bir problem var olup olmadığı incelenmiştir.

Sonrasında ise dosyalar jupyter notebook sistemine aktarılarak, python programı kullanılmak suretiyle öncelikle birleştirilmiş ve tek bir veri çerçevesi haline getirilmiştir. Oluşturulan veri çerçevesi üzerinde oluşan verinin boyutu 2 terabayt civarındadır. Oluşturulan veri çerçevesinin ilerleyen aşamalarda da kullanılabilmesi amacıyla veri çerçevesi pickle formatına dönüştürülmüştür.

Tablo 4.3. Çalışmaya girdi teşkil eden parametreler

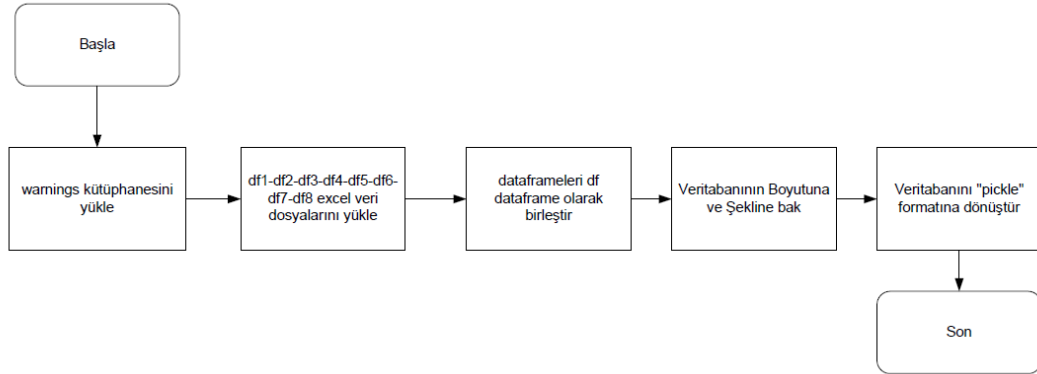
Etkili Parametreler	Etkili Parametreler
Si	Gergi
Fe	ALPURMetalSıcaklığı
Cu	TutmaÇıkışYollukSıcaklığı
Mn	ALPURRotor1ArgonDebisi
Mg	ALPURRotor2ArgonDebisi
Zn	ALPURRotor1KlorDebisi
Cr	ALPURRotor2KlorDebisi
OperatorTarafBaskı	ALPURNitrojenDebisi
TahrikTarafMerdanePozisyonu	ALPURRotor1DönüşHızı
OperatorTarafMerdanePozisyonu	ALPURRotor2DönüşHızı
MerdaneGirişSuSıcaklığı	KenarKesmePayı
ÜstMerdaneStripSlip	ÜstMerdaneBombesi
AltMerdaneStripSlip	AltMerdaneBombesi
TibörHızı	ÜstMerdaneÇapı
ÜstSpreyAtışYüzdesi	AltMerdaneÇapı
AltSpreyAtışYüzdesi	ÜstMerdanePürüzlülüğü
AktuatörAçıklıkYüzdesi	CFF1BasıncDayaimi
UstSpreyHizi	UstMerdaneÇıkışSuSıcaklığı
AltSpreyHizi	AltMerdaneÇıkışSuSıcaklığı
LevhaTahrikTarafBombe	LevhaHatHizi
CFF1TermalSokDayanimi	LevhaOperatörTarafBombe
AltMerdanePürüzlülüğü	AltMerdaneSuDebisi
ÜstMerdaneSinüsü	MerdaneTonajı
AltMerdaneSinüsü	MerdaneSuSıcaklığıSet
ÜstMerdaneTorku	LevhaMaxKalınlık
AltMerdaneTorku	TipTonajı
ÇıkışBobinEni	TandışMetalSeviyesi
UstMerdaneGirisSuBasıncı	AltSpreyGramaj
AltMerdaneGirisSuBasıncı	ÜstSpreyGramaj
MerdaneÇıkışSuBasıncı	ALPURRotor2KarışımGazAkış
UstMerdaneSuDebisi	ALPURRotor1KarışımGazAkış
	SpreyOtomatikManuelMod

4.3.2.1. Eksik verilerin analizi ve etkisiz parametrelerin ayıklanması

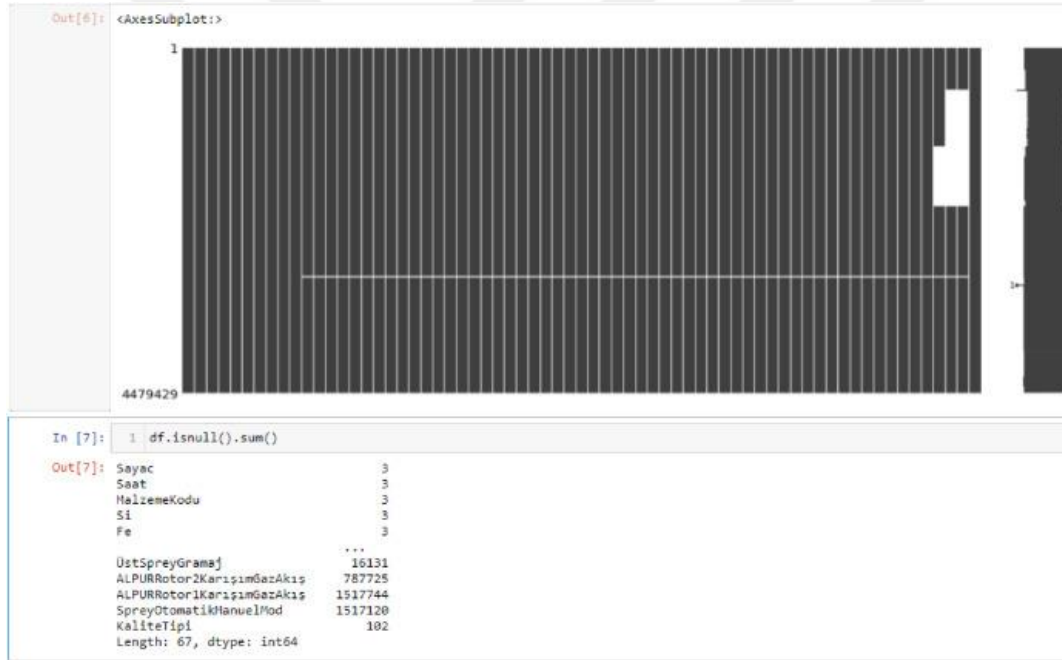
Verinin büyüklüğü nedeni ile tüm veri sisteme 8 ayrı excel dosyası şeklinde yüklenerek, başlangıçta 8 veri çerçevesi oluşturulmuştur. Birleştirme aşağıdaki Şekil 4.6’da da gözlemlenebileceği üzere, kodlama ile yapılarak sisteme yüklenen df1, df2, df3, df4, df5, ve df6 veri çerçeveleri df veri çerçevesi altında toplanılmıştır.

Çalışma kapsamında ilk etapta oluşturulan tek veri çerçevesi üzerinde istatistiksel analizler ve veri madenciliğini engelleyici olacak veri çerçevesi üzerindeki boş ve

tanımsız değerlerin analizi yapılmıştır. Yapılan analizde aşağıdaki Şekil 4.7’de de görebileceğiniz üzere sistemde boş ve tanımsız satır olduğu görülmüştür.



Şekil 4.6. Df veri çerçevesi işlemleri



Şekil 4.7. Df üzerindeki boş ve tanımsız satırlar

Bu sonuçlara dikkatlice bakıldığında, Veri Çerçevesi üzerinde, verisi oldukça eksik olan bazı değişkenlerin (örneğin SpreyOtomatikManuelMod) olduğu gözlemlenmiştir. Tespit edilen bu değişkenler, Tablo 4.4’de görülmektedir.

Eksik veri içeren bu değişkenler, ortalamaya tamamlama yada alt veya üst limite tamamlama şeklinde de tamamlanarak veri çerçevesine dahil edilebilir olmakla birlikte, bu metodun hem değişkenliği azaltarak farklı sonuçlar ortaya koymayacağı, hem de uzman görüşü sonucu çıktıya bu değişkenlerin yüksek oranda etkisinin

olmayacağı düşünülerek, ilgili değişkenler veri çerçevesinden çıkartılma yoluna gidilmiştir.

Tablo 4.4. Tespit edilen girdi değişkenleri

Etkili Parametreler	Br
UstMerdaneGirisSuBasıncı	bar
ALPURRotor2KarışımGazAkış	0
ALPURRotor1KarışımGazAkış	0
AltMerdaneBombesi	mm
ÜstMerdanePürüzlülüğü	micron
AltMerdanePürüzlülüğü	micron
ÜstMerdaneSinüsü	0
AltMerdaneSinüsü	0
UstMerdaneÇıkisSuSicakligi	°C
SpreyOtomatikManuelMod	0

İlgili değişkenler veri çerçevesinden ayıklandıktan sonra, diğer değişkenler üzerinde belirli bir miktarda kalan tanımsız ve boş satırlar aşağıdaki Şekil 4.9’da görüldüğü gibi ortaya çıkmıştır.

Şekil 4.8’da görülen bu değerlere bakıldığında, genel olarak bu değerlerin aynı satırlarda ortaya çıktığı, toplam sayılarının her sütun için aynı/yaklaşık olmasından anlaşılmaktadır. Ayrıca toplam satır sayısına bakıldığında, çoğunlukla 16622 satır veride Tanımsız/Boş veriye rastlanıldığı görülmektedir. Bu da toplam veri (4479427 satır) içerisinde binde 3,7 gibi bir veriye denk gelmektedir. Hem çoğunlukla bu değerlerin aynı satırlarda olması, hem de yüzdesel olarak düşük bir yüzdeye sahip olmaları nedeniyle bu verilerin silinmesi yoluna gidilmiştir. Şekil 4.9, verilerin silindikten sonra veri çerçevesinde de Tanımsız/Boş veri kalıp kalmadığının kontrolüdür. Görüldüğü gibi tüm Tanımsız/Boş veriler temizlenmiştir.

```
In [9]: 1 df.isnull().sum()

Out[9]: Sayac 3
        Saat 3
        MalzemeKodu 3
        Si 3
        Fe 3
        Cu 3
        Mn 3
        Mg 3
        Zn 3
        Cr 3
        OperatorTaraftBaski 16624
        TahrirTaraftMerdanePozisyonu 16624
        OperatorTaraftMerdanePozisyonu 16624
        MerdaneGirisSuSicakligi 16617
        UstMerdaneStripSliip 16624
        AltMerdaneStripSliip 16624
        TiborHizi 16624
        UstSpreyAtisYuzdesi 16624
        AltSpreyAtisYuzdesi 16624
        Gergi 16624
        ALPUMetalSicakligi 16624
        TutmaCikisYollukSicakligi 16624
        ALPURotor1ArgonDebisi 16624
        ALPURotor2ArgonDebisi 16624
        ALPURotor1KlorDebisi 16624
        ALPURotor2KlorDebisi 16624
        ALPURNitrojenDebisi 16624
        ALPURotor1DonusHizi 16624
        ALPURotor2DonusHizi 16624
        KenarKesmePayi 16624
        UstMerdaneBombesi 16624
        UstMerdaneCapı 16386
        AltMerdaneCapı 16623
        UstMerdaneTorku 16624
        AltMerdaneTorku 16624
        CikisBobinEni 16624
        AltMerdaneGirisSuBasinci 16624
        MerdaneCikisSuBasinci 16624
        UstMerdaneSuDebisi 16624
        AltMerdaneSuDebisi 16624
        MerdaneTonaJı 16624
        MerdaneSuSicakligiSet 16624
        LevhaMaxKalınlik 16624
        TipTonaJı 16621
        TandisMetalSeviyesi 21868
        AktuatörAçıklıkYuzdesi 16624
        UstSpreyHizi 16623
        AltSpreyHizi 16623
        LevhaTahrirTaraftBombe 16623
        CFFIBasincDayanimi 16622
        CFFITermalSokDayanimi 16624
        AltMerdaneCikisSuSicakligi 16624
        LevhaHatiHizi 16623
        LevhaOperatörTaraftBombe 16624
        AltSpreyGramaJı 16624
        UstSpreyGramaJı 16131
        KaliteTipi 182
        dtype: int64
```

Şekil 4.8. Diğer değişkenler tanımsız ve boş satır sayıları

```
In [10]: 1 #NaN Değerlerinin olduğu satırları sil
        2
        3 df = df.dropna()
        4 df.isnull().sum()

Out[10]: Sayac 0
        Saat 0
        MalzemeKodu 0
        Si 0
        Fe 0
        Cu 0
        Mn 0
        Mg 0
        Zn 0
        Cr 0
        OperatorTaraftBaski 0
        TahrirTaraftMerdanePozisyonu 0
        OperatorTaraftMerdanePozisyonu 0
        MerdaneGirisSuSicakligi 0
        UstMerdaneStripSliip 0
        AltMerdaneStripSliip 0
        TiborHizi 0
        UstSpreyAtisYuzdesi 0
        AltSpreyAtisYuzdesi 0
        Gergi 0
        ALPUMetalSicakligi 0
        TutmaCikisYollukSicakligi 0
        ALPURotor1ArgonDebisi 0
        ALPURotor2ArgonDebisi 0
        ALPURotor1KlorDebisi 0
        ALPURotor2KlorDebisi 0
        ALPURNitrojenDebisi 0
        ALPURotor1DonusHizi 0
        ALPURotor2DonusHizi 0
        KenarKesmePayi 0
        UstMerdaneBombesi 0
        UstMerdaneCapı 0
        AltMerdaneCapı 0
        UstMerdaneTorku 0
        AltMerdaneTorku 0
        CikisBobinEni 0
        AltMerdaneGirisSuBasinci 0
        MerdaneCikisSuBasinci 0
        UstMerdaneSuDebisi 0
        AltMerdaneSuDebisi 0
        MerdaneTonaJı 0
        MerdaneSuSicakligiSet 0
        LevhaMaxKalınlik 0
        TipTonaJı 0
        TandisMetalSeviyesi 0
        AktuatörAçıklıkYuzdesi 0
        UstSpreyHizi 0
        AltSpreyHizi 0
        LevhaTahrirTaraftBombe 0
        CFFIBasincDayanimi 0
        CFFITermalSokDayanimi 0
        AltMerdaneCikisSuSicakligi 0
        LevhaHatiHizi 0
        LevhaOperatörTaraftBombe 0
        AltSpreyGramaJı 0
        UstSpreyGramaJı 0
        KaliteTipi 0
        dtype: int64
```

Şekil 4.9. Değişkenlerin tanımsız ve boş değerlerinin ayıklanması

Kalan veriler üzerinden, değişkenlerin birbiri ile ilişkisini incelemek üzere Korelasyon Analizi yapılmıştır. Burada amaç, birbiri ile korelasyon katsayısı yüksek olacak olan değişkenler arasında ilişki yüksek olacağından, bu değişkenlerden birini göz önünde bulundurmak, aynı zamanda ilişkili diğer değişkeni de göz önünde bulundurmak olacaktır. Bu şekilde yüksek ilişkili değişkenler belirlenmiştir. Bunlar Tablo 4.5'te görülmektedir.

Tablo 4.5. Korelasyon analizi sonucu ayıklanan girdi parametreleri

Girdi Parametre 1	Girdi Parametre 2	Korelasyon
TahrikTarafMerdanePozisyonu	AlpurRotor1KlorDebisi	0,95
AlpurRotor1ArgonDebisi	ÜstMerdaneStripSlip	0,85
AlpurRotor1ArgonDebisi	AlpurRotor1KlorDebisi	0,97
ALPURRotor1DönüşHızı	ÜstMerdaneBombesi	0,96
ALPURRotor2DönüşHızı	AltMerdaneSuDebisi	0,87
ÜstMerdaneBombesi	KenarKesmePayı	-0,98
KenarKesmePayı	ALPURRotor1DönüşHızı	0,94
ÜstMerdaneBombesi	ÜstMerdaneTorku	-0,91
AltMerdaneSuDebisi	ALPURRotor2DönüşHızı	0,87
LevhaHatHizi	LevhaMaxKalınlık	-0,9
TahrikTarafMerdanePozisyonu	AlpurRotor1ArgonDebisi	0,95

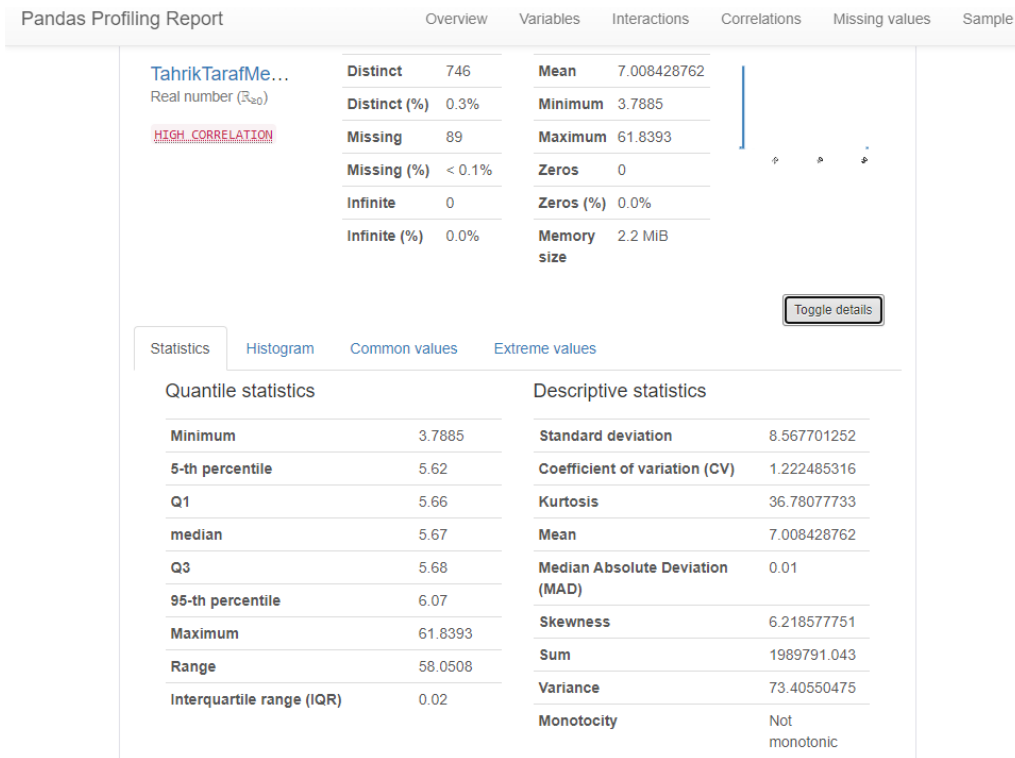
Bu değişkenlerin ilişkili olan çiftleri veri çerçevesinden ayıklanmıştır. Böylelikle veri çerçevesi daha sade bir hale getirilmiştir.

4.3.2.2. Yanlış tasnif edilmiş yada mantıksız değerler

Aslında çalışmaya başlarken ilk başta planlanan, öncelikle tüm veri tabanının birleştirilmesi ve sonrasında tüm veri ön işleme aşamalarının bu birleştirilen tek veri çerçevesi üzerinden yapılması idi. Ancak veri çerçevesini oluşturan 24 farklı yarı mamulun her birinin veri çerçevesi içerisinde ortalama ve alt/üst limitlerin dışında istatistiksel analizler sonucunda bulunabilecek aykırı değerleri olduğu gibi, aynı zamanda bu verinin içerisinde, teknik olarak olması mümkün olmayan yanlış okunmuş veriler de olabilirdi.

Bu noktada, öncelikle bu konunun tespiti için, veri çerçevesi üzerinden her bir malzemeye ait veri çerçevelerinin, geriye kalan girdi değişkenleri açısından, değişkenlerin teknik alt ve üst limitleri dışında kalan verileri olup olmadığına seaborn kütüphanesinin `profilereport()` özelliği kullanılarak bakılmıştır.

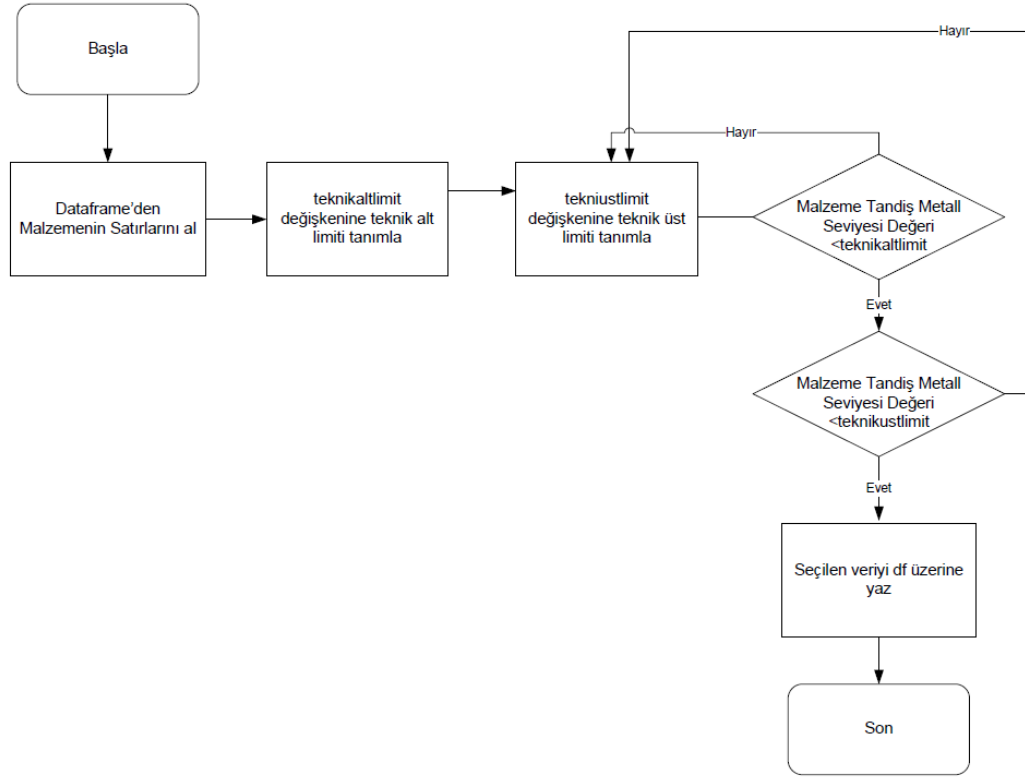
Bu yöntemle, verinin dağılımlarına malzeme bazlı detaylı baktığımızda, malzemelerde Tandiş Metal Seviyesi girdi değişkeninde ağırlıklı olmak üzere, Şekil 4.10'da da örnek olarak gösterildiği üzere, teknik limitler dışı veri olduğu tespit edilmiştir. Bu noktada, bu ve benzeri değerleri aykırı değer analizinde ortalama hesaplamasına katmamız durumunda, bu değerler ortalamayı çok yukarıya yada aşağıya çekerek bizi manipüle edebileceği düşünülmüştür. Çünkü bu durum, benzer şekilde alt ve üst limitleri de yanlış noktalara getirebilecektir. Bu nedenle, eksik verilerin ortadan kaldırılması ile birlikte, veri setinden bu değerlerin de aykırı değer analizi öncesinde ortadan kaldırılması gerektiği düşünülerek, bu şekilde hareket edilmiştir.



Şekil 4.10. B01 malzemesi tandiş metal seviyesi profilreport () analizi

Bu bağlamda veri sistemimizde analiz başında bu aşamada veri setinden çıkartılmıştır. Veri setinden bu temizleme işlemi Şekil 4.11'de görülen kod ile yapılmıştır. Yapılan temizleme sonucunda toplamda ana veri çerçevesi üzerinde 246.668 satır veri ayıklanmıştır. İlgili bu verilerin, işleme konulması için bunları ayıklamak yerine bu verileri ortlamaya eşitlemek yada baskılama metotlarını uygulamak da düşünülmüştür. Ancak bu tarz metotların değişkenliği azaltması ve verinin gerçek değerlerini olumsuz etkileyeceği düşünüldüğünden verilerin silinmesi metodu uygulanmıştır.

Son olarak, yanlış tasnif edilen veriler çıkartıldıktan sonra, parçalara (malzeme kodu bazlı veri çerçevelerine) böldüğümüz veri, tekrar bir veri çerçevesi olarak toplanılmıştır. Yapılan bu analizler sonucu oluşan veri çerçeveleri, ilerleyen bölümlerde kümeleme ve sınıflandırma algoritmalarında kullanılacaktır.

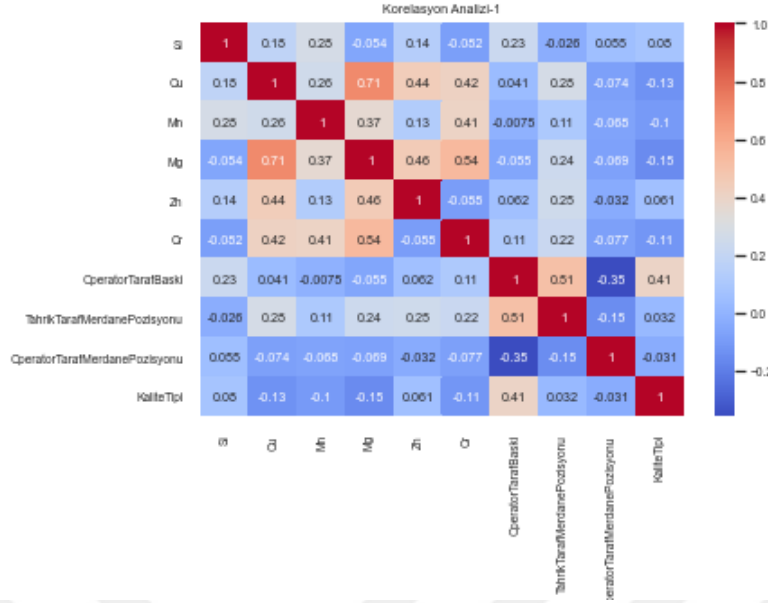


Şekil 4.11. Veri setinde teknik limitler dışı değerlerin temizlenmesi

4.3.3. Veri madenciliği

4.3.3.1. Veri seti (veri çerçevesi) üzerinde istatistiksel analizler

Daha önceki bölümlerde yapılan analizler sonucunda, başlangıçta 63 adet olan girdi parametreleri, 42 adete düşürülmüştü. Bu bölümde, verideki girdi değişkenlerin kalite üzerindeki etkileri incelenerek, değişkenlerin çıktı değişkeni üzerinde etkisi yoksa bunların elimine edilmesi amaçlanmıştır. Bunun için ise toplanan data üzerinden, elde edilen girdi parametreleri ile çıktı (KaliteTipi) arasındaki korelasyona ısı haritası matrisi ile bakılmıştır. Elde edilen matris Şekil 4.12’de gösterilmiştir.



Şekil 4.12. Girdi parametreleri/çıktı (kalitetipi) arasındaki korelasyon matrisi

Buradan da görülebileceği üzere, KaliteTipi ile bazı girdi parametreleri arasında çok zayıf bir ilişki vardır. Burada zayıf ilişki, literatürde 0,2'nin altı pearson katsayısına göre zayıf ilişki sayılmasına rağmen (Statistics How To, Tarih Yok) güvenli tarafta kalınması adına bu değerin de altında 0,18 değeri altı değerler olarak kabul edilmiştir. Bu duruma istinaden, Tablo 4.6' da da gösterilen koyu renk ile işaretli olan 10 parametre dışında kalan girdi değişkenleri, kümeleme, sınıflandırma ve makine öğrenmesi analizleri öncesinde düşük korelasyona sahip olmaları nedeniyle veri çerçevesinden çıkartılmıştır. Tüm bu analizler neticesinde, başlangıçta 63 adet olan girdi değişken sayısı analizler öncesinde 10 adede indirgenmiştir.

Tablo 4.6. Korelasyon analizi sonucu kalan girdi değişkenleri

Girdi Parametresi	Pearson Korelasyon Katsayısı
Si	0,08
Cu	-0,13
Mn	-0,1
Mg	-0,15
Zn	0,06
Cr	-0,11
OperatorTarafBaski	0,41
TahrikTarafMerdanePozisyonu	0,03
OperatorTarafMerdanePozisyonu	-0,03
MerdaneGirisSuSicakligi	0,31
AltMerdaneStripSlip	0,16

Tablo 4.5. (Devamı) Korelasyon analizi sonucu kalan girdi değişkenleri

Girdi Parametresi	Pearson Korelasyon Katsayısı
TiborHizi	-0,07
UstSpreyAtisYuzdesi	-0,18
AltSpreyAtisYuzdesi	-0,17
Gergi	-0,03
ALPURMetalSicakligi	0,06
TutmaCikisYollukSicakligi	-0,01
ALPURRotor2ArgonDebisi	0,04
ALPURRotor2KlorDebisi	-0,27
ALPURNitrojenDebisi	-0,16
UstMerdaneCapi	0,53
AltMerdaneCapi	0,51
AltMerdaneTorku	0,08
CikisBobinEni	-0,05
AltMerdaneGirisSuBasinci	0,06
MerdaneCikisSuBasinci	-0,05
UstMerdaneSuDebisi	0,09
MerdaneTonaji	0,16
MerdaneSuSicakligiSet	-0,14
TipTonaji	0,08
TandisMetalSeviyesi	0,14
AktuatorAciklikYuzdesi	0,02
UstSpreyHizi	-0,27
AltSpreyHizi	0,03
LevhaTahrikTarafBombe	0,04
LevhaOperatorTarafBombe	0,52
CFF1BasincDayanimi	-0,01
CFF1TermalSokDayanimi	-0,13
AltMerdaneCikisSuSicakligi	0,33
LevhaHatHizi	0,23
AltSpreyGramaj	-0,01
UstSpreyGramaj	-0,33

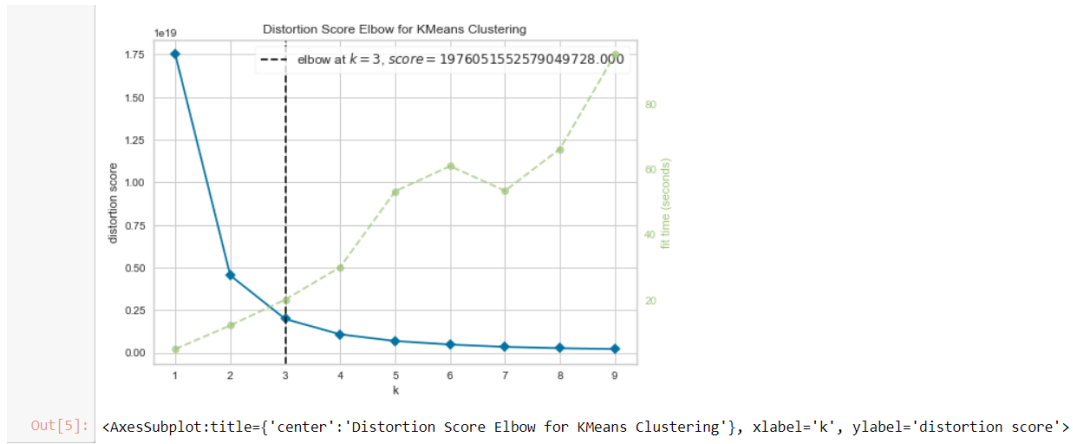
4.3.3.2. Kümeleme analizi ve sonuçları

Gelinen bu noktada, daha önce de bahsettiğimiz gibi, önceki bölümlerde anlatıldığı şekilde temizlenen veri üzerinde, tahmin algoritmalarının kullanımına geçilmiştir. Öncelikle, veri, sahip olduğu yarımamullerden bağımsız, bütün bir veri çervesi halinde düşünülerek, sanki KaliteTipi de bilinmiyormuş gibi, eğiticişiz öğrenme mantığına uygun olarak kümelenmekte ve tahmin performansına bakılmaktadır. Sonrasında ise aynı prosedür, bu sefer de veriyi yarı mamul bazlı olarak ayrı veri çerçeveleri olarak

düşünülmekte ve her bir yarıymamule ait bu veri çerçevelerine yine aynı kümeleme algoritmaları uygulanarak, kümeleme performanslarına bakılmaktadır.

- **Tüm Veri Seti Kümeleme Analizi ve Sonuçlar**

Öncelikle veri setine Elbow Yöntemi uygulanmıştır. KaliteTipi'nin bu yöntemle optimal kaç kümeden oluşması gerektiği analiz edilmiştir. Sonuçlar, Şekil 4.13'de de gösterilmiştir. Analiz, gerçekte de olduğu üzere veriyi 3 ana kümeye ayırmıştır. Gerçekten de verinin KaliteTipi çıktısı, 1,2, ve 3 tip olacak şekilde 3 gruptur. Bu da veri setinin zaten optimal olarak 3 kümeden oluşması gerekliliğini yada tespitini ortaya koymaktadır.



Şekil 4.13. Elbow yöntemi ile optimal küme analizi

Bu noktadan sonra kümeleme analizlerine Google Colab Pro ve Google Sanal Makine çözümleri ile oluşturulan yüksek güçlü bilgisayarlar ile devam edilmiştir. Çünkü veri, ilgili analiz yöntemlerinde kişisel bir bilgisayarda analiz edilemeyecek kadar büyüktür.

Çalışma kapsamında seçilen kümeleme algoritmaları aşağıdaki gibi görülmektedir. İlgili kümeleme algoritmalarının neden seçildiği de yine Tablo 4.7'de açıklanmıştır.

Yakınlık Dağılımı ve Aglomeratif Kümeleme gibi bazı algoritmaların uygulaması esnasında, aşağıdaki Şekil 4.14'de görülebileceği üzere verinin büyüklüğünden dolayı RAM yetersizliği ortaya çıkmıştır. Bunun nedeni ise örneğin Yakınlık Dağılımı algoritması tam mesafe matrisini depolamak için ikinci dereceden belleğe ihtiyaç duymaktadır. Dolayısıyla, 10000 örneğiniz ve çift kesinliğiniz varsa, yaklaşık 800.000.000 bayta ihtiyacınız vardır. Bir noktada bu matrisin bir kopyasının yapılması gerekiyorsa, kolayca 1,6 GB RAM'e ihtiya. olmaktadır (giriş verileriniz ve herhangi

bir ek yük hariç). "Yüzbinlerce" verisi olan bir durumda ise buna en az 100 faktör daha, yani 80 ila 160 GB RAM ek ihtiyaç çıkması demektir (Stackoverflow, 2021).

```
In [4]: #Gerekli Kütüphaneleri Yükledik.
from sklearn.decomposition import PCA
import scipy.cluster.hierarchy as sch
from sklearn.cluster import AgglomerativeClustering
dendrogram = sch.dendrogram(sch.linkage(df, method='ward'))

-----
MemoryError                                Traceback (most recent call last)
<ipython-input-4-e9badd9c023b> in <module>
      3 import scipy.cluster.hierarchy as sch
      4 from sklearn.cluster import AgglomerativeClustering
----> 5 dendrogram = sch.dendrogram(sch.linkage(df, method='ward'))

~\anaconda3\lib\site-packages\scipy\cluster\hierarchy.py in linkage(y, method, metric, optimal_ordering)
    1058     'matrix looks suspiciously like an uncondensed '
    1059     'distance matrix')
-> 1060     y = distance.pdist(y, metric)
    1061 else:
    1062     raise ValueError("'y' must be 1 or 2 dimensional.")

~\anaconda3\lib\site-packages\scipy\spatial\distance.py in pdist(X, metric, *args, **kwargs)
    2021     out = kwargs.pop("out", None)
    2022     if out is None:
-> 2023         dm = np.empty((m * (m - 1)) // 2, dtype=np.double)
    2024     else:
    2025         if out.shape != (m * (m - 1)) // 2,:

MemoryError: Unable to allocate 65.0 TiB for an array with shape (8939254102551,) and data type float64
```

Şekil 4.14. Tüm veri üzerinde, ward kümeleme denemesi sonucu

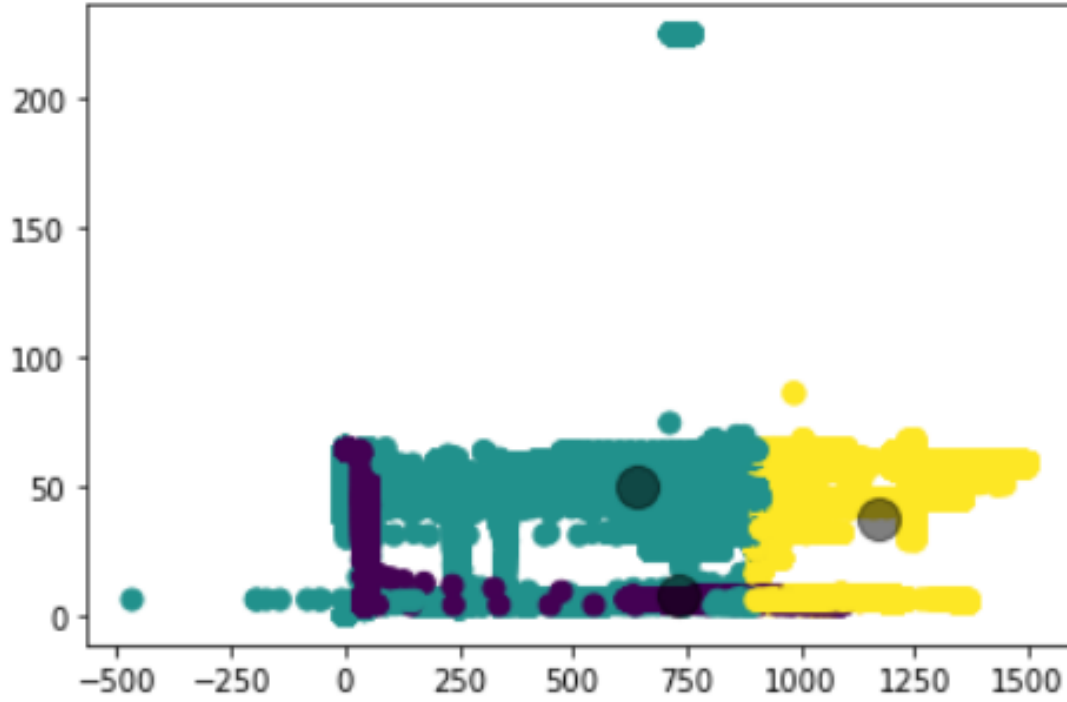
Keza durum Agglomeratif Kümelemede de benzerdir. Agglomeratif Kümeleme'nin bellek tüketimi $O(n^2)$, veri boyutuna kıyasla katlanarak büyüdüğü anlamına gelir (Stackoverflow, 2021).

Bu nedenlerle sadece K-Means gibi küme sayısının önceden girilebildiği algoritmaların denenmesine karar verilmiştir. Bu algoritmalar liste halinde Tablo Tablo 4.7'de sunulmuştur.

Tablo 4.7. Seçilen kümeleme algoritmaları

Algoritma Adı	Neden Seçildiği
K-Means	Genel amaçlı ve çok küme boyutu yok,
Spektral Kümeleme	Az sayıda küme ve bir neden sonuç ilişkisine bağlı ilişkili veri tipi
BIRCH	Az sayıda küme, bir neden sonuç ilişkisine bağlı ilişkili veri tipi ve çok büyük veride başarılı
HAC (Ward)	Birçok küme için uygulanmakla birlikte ki bu bize uygun değildir, büyük örneklerde de uygulaması olduğundan denendi.

İlgili algoritmalar GoogleCOLAB üzerinde çalıştırılmıştır. Çalıştırılan her bir algoritma, Şekil 4.15'de K-Means algoritması için örneği gösterildiği üzere veri seti için kümeleri hesaplar.

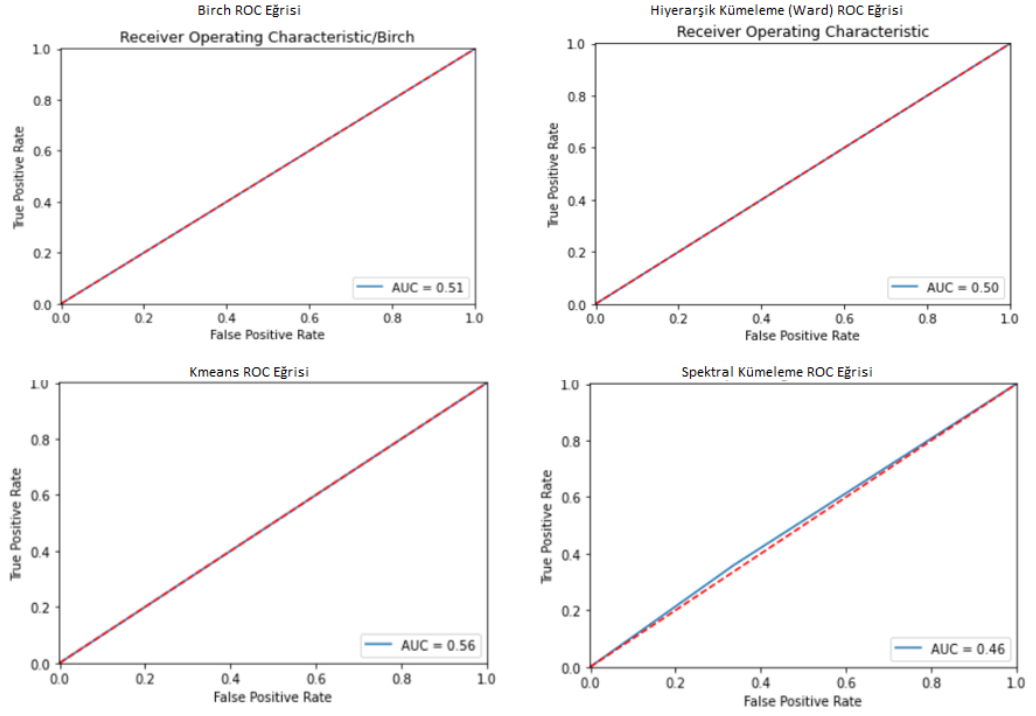


Şekil 4.15. Tüm veri seti için k-means kümeleme

İlgili algoritmalara ait Siluet Skorları ve AUC değerleri Tablo 4.8’de sunulmuştur. Yapılan analizler sonucunda, Tabloda da görüldüğü üzere siluet skor değerleri 0,85 değerinden çok uzak çıkmıştır. Benzer olarak AUC değerleri de 1’den çok uzaktır ve iyi çıkmamıştır. Bu nedenle doğruluk analizi yapma gerekliliği de duyulmamıştır. Algoritmalar için ROC Eğrileri Şekil 4.16’da gösterilmiştir.

Tablo 4.8. Kümeleme algoritmaları siluet skor ve auc değerleri karşılaştırması

Algoritma	Siluet Skor Değeri	AUC Değeri
K-Means	0,6	0,56
BIRCH	0,58	0,51
Spektral Kümeleme	0,03	0,46
HAC (Ward)	0,54	0,50



Şekil 4.16. Tüm veri kümeleme algoritmaları roc eğrileri

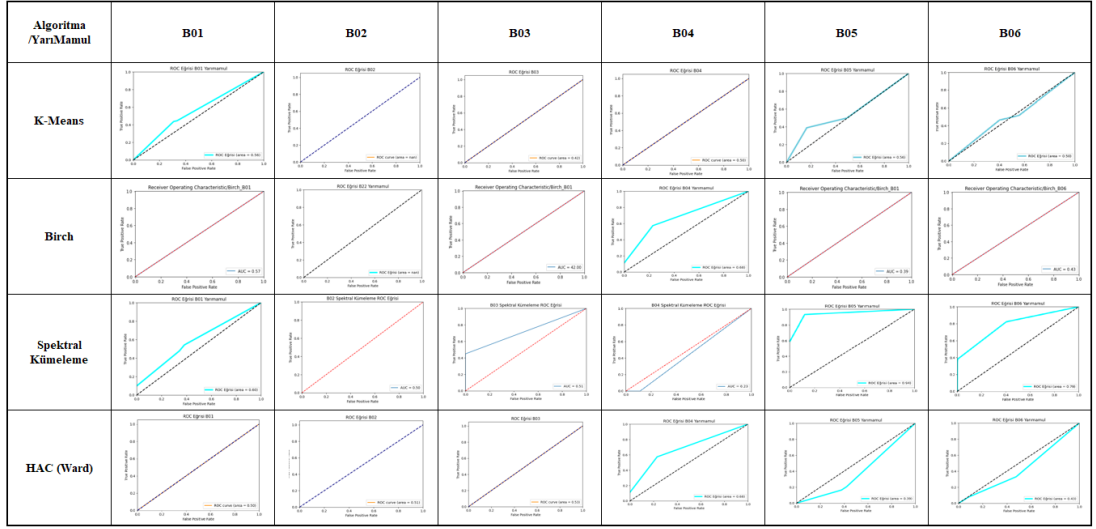
Ürün Bazlı Kümeleme Analizi Sonuçlar

Bu bölümde veri seti içerisinde malzemeler tek tek birer veri çerçevesi olarak alındıktan sonra, malzemelere ait verilere bir önceki bölümde uygulanan kümeleme algoritmaları uygulanmıştır. Uygulanan kümeleme algoritmalarının başarı performanslarının değerlendirilmesi için literatüre de uygun olarak bir önceki bölümde de tüm veri üzerinde uyguladığımız üzere algoritmaların siluet skorlarına ve algoritma AUC değerlerine, ROC eğrilerine bakılmıştır. Her bir yarı mamulde uygulanan algoritmaların siluet skorları ve AUC değerleri, her biri için ayrı ayrı Tablo 4.9’da aşağıda gösterilmiştir. Tablo üzerinde ilgili analiz edildiği her bir yarımamul için, ilgili siluet skoru ve AUC değeri, eğer başarılı olarak sayılan istenilen değer olan 0,8 değeri ve üzeri gelen değerde ise, bu değerler koyu punto ile gösterilmiştir. Ayrıca herbir yarı mamul için, yapılan algoritma analizlerine ait ortaya çıkan AUC değerleri için ROC eğrileri de çizilmiştir. Bu çizilen eğrilerin B01-B06 arası yarı mamullere ait olanları Şekil 4.17’de gösterilmiştir. Burada B05 yarı mamülüne ait yüksek AUC değeri, ROC eğrisinde de görülebilir.

Tablo 4.9. Yarı mamul bazlı uygulanan kümeleme algoritmaları sonuçları

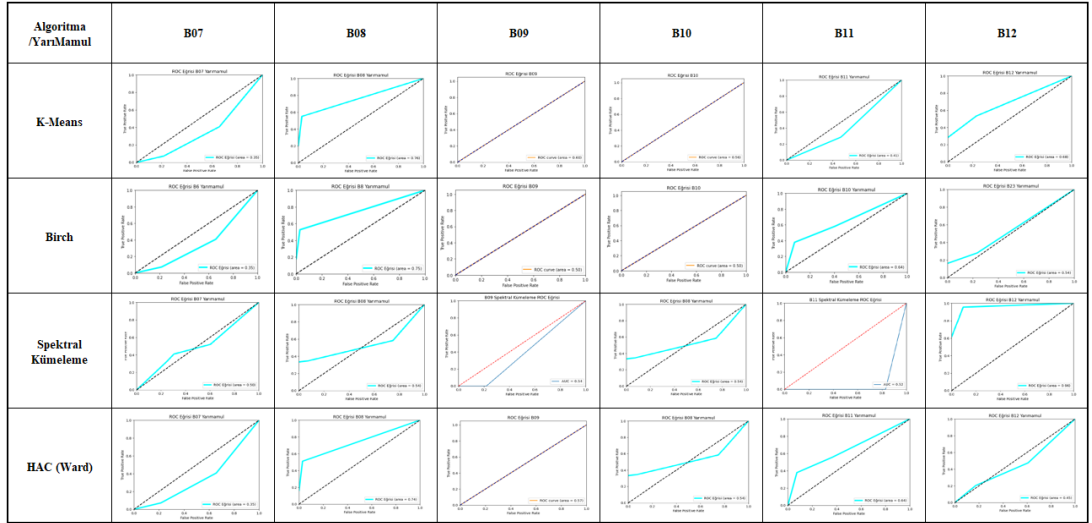
/YarıMamul	B01		B02		B03		B04		B05	
Değerlendirme	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri
K-Means	0,59	0,56	0,59	nan*	0,59	0,42	0,73	0,50	0,77	0,56
BIRCH	0,56	0,57	0,52	nan*	0,53	0,42	0,74	0,68	0,74	0,39
Spektral K.	0,01	0,60	0,20	0,49	0,07	0,51	0,24	0,23	0,69	0,94
HAC (Ward)	0,54	0,50	0,56	0,51	0,55	0,53	0,74	0,68	0,74	0,39
/YarıMamul	B06		B07		B08		B09		B10	
Değerlendirme	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri
K-Means	0,77	0,50	0,85	0,35	0,64	0,76	0,63	0,60	0,58	0,56
BIRCH	0,77	0,43	0,85	0,35	0,64	0,75	0,60	0,50	0,52	0,50
Spektral K.	0,59	0,79	0,75	0,50	0,16	0,54	0,06	0,54	0,53	0,54
HAC (Ward)	0,78	0,43	0,85	0,35	0,61	0,74	0,60	0,57	0,55	0,54
/YarıMamul	B11		B12		B13		B14		B15	
Değerlendirme	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri
K-Means	0,65	0,41	0,80	0,68	0,59	0,51	0,59	0,58	0,59	0,47
BIRCH	0,63	0,64	0,80	0,54	0,56	0,69	0,53	0,40	0,57	0,33
Spektral K.	0,07	0,52	0,43	0,96	0,52	0,38	0,40	0,43	0,46	0,27
HAC (Ward)	0,61	0,64	0,79	0,45	0,57	0,55	0,55	0,57	0,56	0,49
/YarıMamul	B16		B17		B18		B19		B20	
Değerlendirme	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri
K-Means	0,58	0,58	0,59	0,60	0,58	0,41	0,59	0,52	0,58	0,37
BIRCH	0,56	0,43	0,54	0,43	0,57	0,48	0,58	0,48	0,55	0,38
Spektral K.	0,13	0,29	0,41	nan*	0,16	0,83	0,16	0,63	0,45	0,34
HAC (Ward)	0,56	0,59	0,54	nan*	0,57	0,49	0,59	0,33	0,56	0,39
/YarıMamul	B21		B22		B23		B24			
Değerlendirme	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri	Siluet Skoru	AUC Değeri		
K-Means	0,59	0,41	0,59	0,52	0,59	0,50	0,64	0,50		
BIRCH	0,57	0,43	0,56	0,49	0,55	0,50	0,60	nan*		
Spektral K.	0,44	0,63	0,28	0,34	0,32	0,64	0,66	nan*		
HAC (Ward)	0,53	0,68	0,55	0,49	0,55	0,50	0,61	0,51		

* AUC Analizi için tahmin edilen test datası tek boyutlu yani tek sınıflı olduğundan AUC analizi yapılamamıştır.



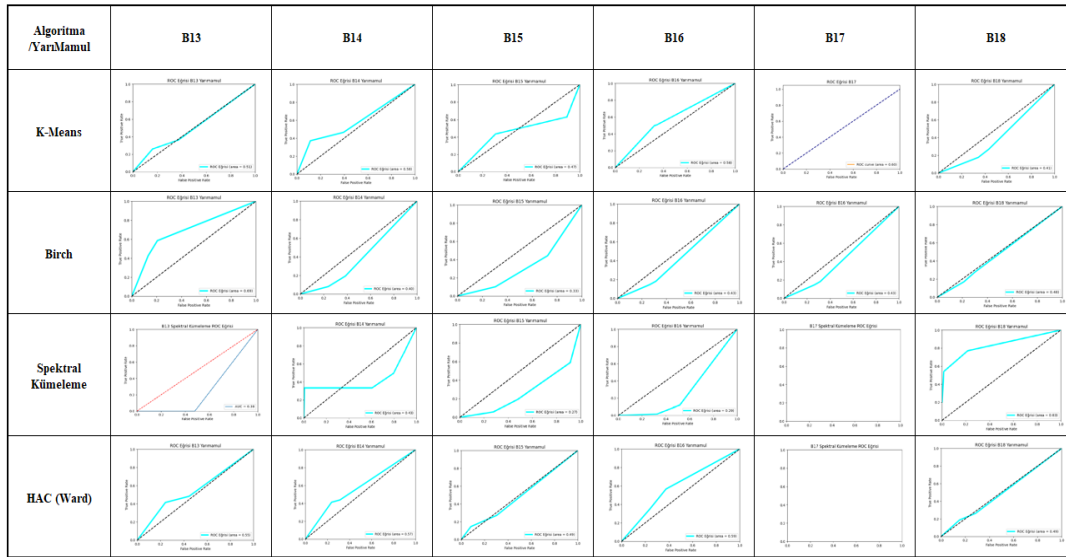
Şekil 4.17. B01-B06 arası yarı mamül kümeleme roc eğrileri

B07-B12 yarı mamülleri ROC eğrileri Şekil 4.18’de gösterilmiştir. Burada da yine B12 yarı mamülüne ait ROC eğrisi spektral kümeleme algoritması için 45 derece eğrisinin (orta kesikli çizgi) oldukça yukarıdadır. Bu durum bize burada başarılı bir sınıflandırma yapıldığını gösterir.



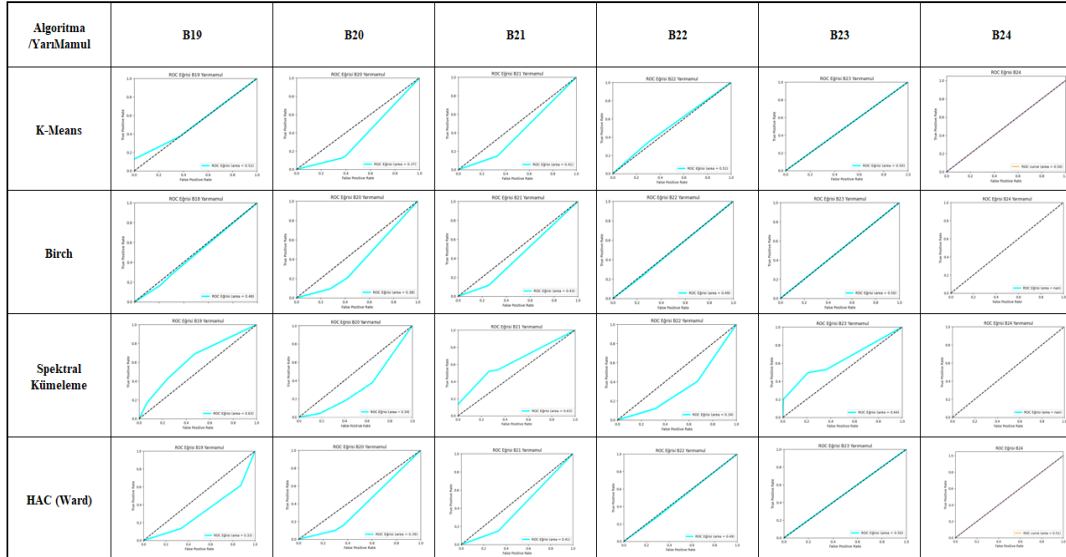
Şekil 4.18. B07-B12 arası yarı mamül kümeleme roc eğrileri

B13-B18 yarı mamülleri ROC eğrileri Şekil 4.19’de gösterilmiştir. Burada da yine B18 yarı mamülüne ait ROC eğrisi spektral kümeleme algoritması için 45 derece eğrisinin (orta kesikli çizgi) oldukça yukarıdadır. Bu durum bize burada başarılı bir sınıflandırma yapıldığını gösterir.



Şekil 4.19. B13-B18 arası yarı mamül kümeleme roc eğrileri

B19-B24 yarı mamülleri ROC eğrileri Şekil 4.20’de gösterilmiştir. Burada ROC eğrileri genel olarak 45 derece eğrisinin (orta kesikli çizgi) etrafında çıkmıştır. Bu nedenle bu grupta yapılan analizlerde ROC eğrileri açısından başarılı bir sınıflandırma yapıldığı söylenememektedir.



Şekil 4.20. B19-B24 arası yarı mamül kümeleme roc eğrileri

4.3.3.3. Sınıflandırma analizleri

Bu bölümde, veri seti üzerinde makine öğrenmesi ve yapay zeka algoritmaları ile Eğitici Öğrenme metodu denenerek veri seti sınıflandırılıp, algoritmaların kalite tahmin performanslarına bakılmıştır. Kümeleme algoritmalarında olduğu gibi, yine

önce algoritmalar tüm veri üzerinde, sonrasında da yarı mamul bazlı olarak denenmiştir.

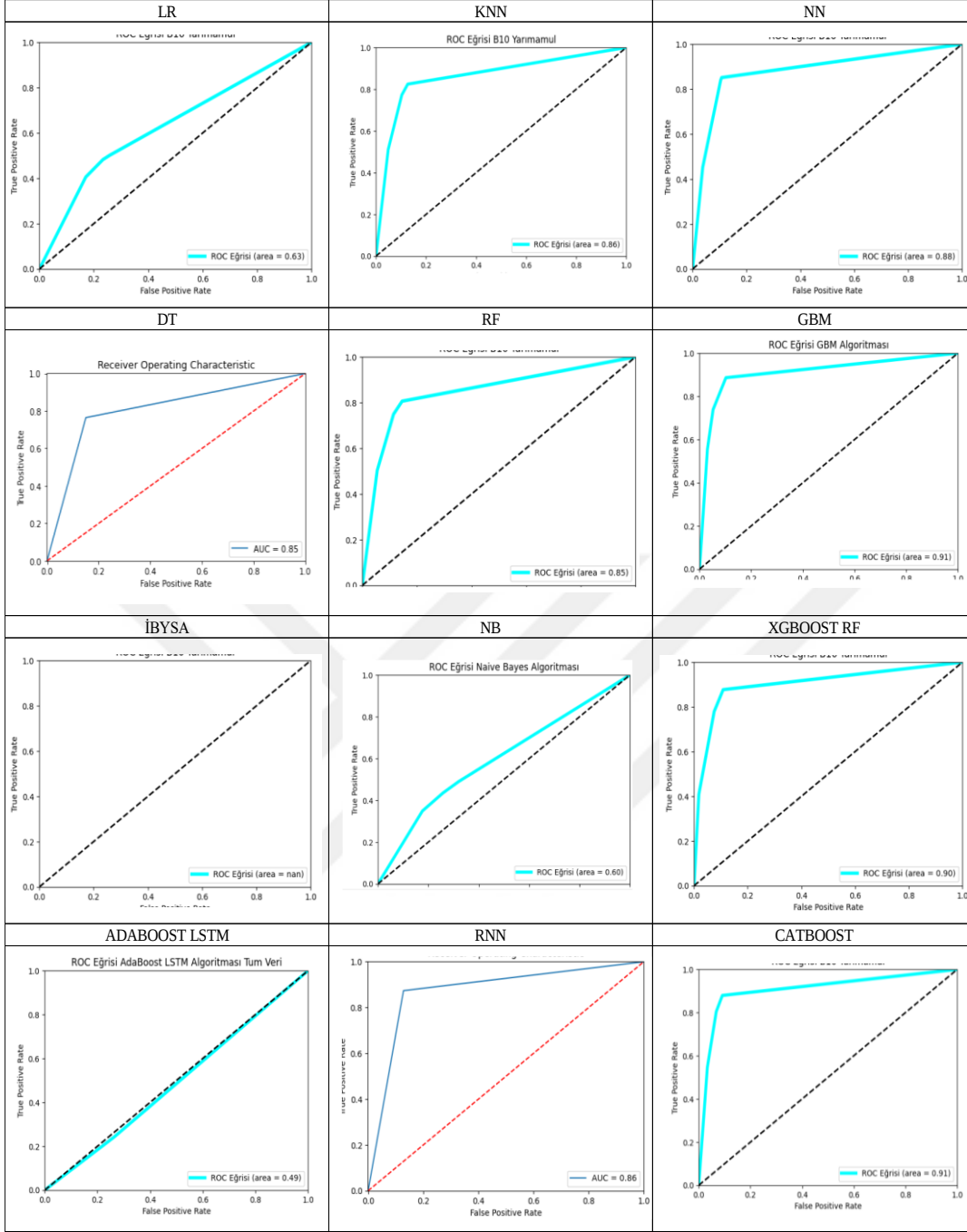
Tüm Veri Seti Sınıflandırma Analizi ve Sonuçları

Bu bölümde, seçilen derin öğrenme ve makine öğrenmesi algoritmaları, yarı mamul farkı gözetmeksizin tüm veri üzerinde denenerek, ilk öncelik olarak AUC değeri, sonrasında öncelik sırasıyla olacak şekilde doğruluk, F1 Skoru, Kesinlik ve Duyarlılık değerleri hesaplanmıştır. Ayrıca, performans kıyaslaması açısından gerekli olabileceği düşüncesi ile, algoritmaların çalışma süreleri de son değerlendirme kriteri olacak şekilde ölçülmüştür. Elde edilen sonuçlar, algoritma doğruluk değerleri ile birlikte aşağıdaki Tablo 4.10'da özetlenmiştir. Burada en başarılı algoritmalar olan GBM ve CATBOOST algoritmaları değerleri koyu punto ile gösterilmiştir.

Tablo 4.10. Tüm veri sınıflandırma analizi sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Dakika)
LR	0,55	0,35	0,63	0,36	0,38	12,0
KNN	0,82	0,82	0,86	0,81	0,83	4,0
NN	0,83	0,76	0,88	0,78	0,76	21,0
DT	0,79	0,80	0,85	0,80	0,79	1,1
RF	0,79	0,81	0,85	0,81	0,81	31,0
GBM	0,88	0,87	0,90	0,88	0,86	41,0
FFNN	0,26	0,11	0,00	0,07	0,26	32,0
NB	0,49	0,31	0,60	0,34	0,58	0,6
XGBOOST RF	0,86	0,86	0,90	0,86	0,86	48,0
ADABOOST LSTM	0,43	0,44	0,48	0,38	0,44	0,5
CATBOOST	0,87	0,87	0,91	0,85	0,87	47,0
RNN	0,81	0,61	0,86	0,83	0,81	29,0

Algoritmalara ait ROC eğrileri Şekil 4.21'de sunulmuştur. Buna göre tabloda AUC değerlerinde de görüldüğü gibi, 45 derece eğrisinin oldukça yukarısında kalan GBM ve CATBOOST algoritmaları sınıflamada başarılıdır.



Şekil 4.21. Tüm veri sınıflandırma roc eğrileri

• Yarı Mamul Bazlı Sınıflandırma Analizi ve Sonuçları

Bu bölümde, seçilen derin öğrenme ve makine öğrenmesi algoritmaları, yarı mamül farkı gözetilerek ön işleme yapılmış tüm verinin içerisinde yarı mamüllerin kendilerine ait verileri seçilip, bunların üzerinde denenerek, ilk öncelik seçim kriteri AUC değeri olacak şekilde, değerlendirme kriteri olarak önem sırasına göre Doğruluk, F1 Skoru, Kesinlik ve Duyarlılık değerleri hesaplanmıştır. Ayrıca, performans

kıyaslaması açısından gerekli olabileceği düşüncesi ile, algoritmaların Çalışma Süreleri, algoritma başarısı için son değerlendirme kriteri olacak şekilde ölçülmüştür. Elde edilen sonuçlar, algoritma doğruluk değerleri ile birlikte aşağıdaki tablolarda her yarı mamül için ayrı tablolar halinde hazırlanarak sunulmuştur.

Tablolarda gösterilen AUC değerlerine ilişkin algoritmalar için ayrıca ROC eğrileri de çizilmiştir. İlgili eğriler tez ekinde sunulmuştur. İdeal olarak, modellerde, yüksek TPR'ye sahip bir model aranır, ancak FPR'yi mümkün olduğunca düşük tutmaya çalışılır. Bu açıdan eğrilerin eğimlerinin artışı da bir yorum ortaya koyar. Ancak elde edilen AUC değerlerinin de bu yorumun genel bir sonucu olduğunu belirtmemiz gerekir.

B01 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.11'de gösterilmiştir. RNN algoritması en yüksek değerleri ile en iyi algoritma olmuştur.

Tablo 4.11. B01 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,88	0,85	0,84	0,86	0,88	5
KNN	0,94	0,94	0,81	0,94	0,94	12
NN	0,96	0,96	0,88	0,96	0,96	22
DT	0,92	0,93	0,80	0,93	0,93	35
RF	0,92	0,93	0,79	0,93	0,93	41
GBM	0,95	0,95	0,87	0,96	0,95	80
FFNN	0,00	0,00	0,32	0,01	0,00	19
NB	0,84	0,82	0,71	0,49	0,60	10
XGBOOST RF	0,95	0,96	0,38	0,96	0,96	47
ADABOOST LSTM	0,81	0,90	0,50	1,00	0,81	30
CATBOOST	0,94	0,95	0,85	0,95	0,95	28
RNN	0,95	0,94	0,91	0,92	0,96	20

B02 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.12'de gösterilmiştir. NN algoritması, en yüksek değerleri alarak en iyi sınıflandırma algoritması olmuştur.

Tablo 4.12. B02 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,94	0,91	0,95	0,94	0,94	20
KNN	0,91	0,91	0,77	0,91	0,91	19
NN	0,94	0,95	0,98	0,97	0,94	36
DT	0,90	0,90	0,75	0,91	0,90	20
RF	0,73	0,90	0,74	0,90	0,90	33
GBM	0,94	0,94	0,84	0,94	0,94	69
FFNN	0,06	0,01	0,25	0,01	0,01	99
NB	0,84	0,91	0,73	0,88	0,94	15
XGBOOST RF	0,04	0,07	0,39	0,54	0,04	65
ADABOOST LSTM	0,71	0,71	0,50	1,00	0,72	25
CATBOOST	0,91	0,92	0,77	0,92	0,91	101
RNN	0,94	0,92	0,97	0,94	0,94	17

B03 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.13’de gösterilmiştir. RNN algoritması, burada da en yüksek değerleri ile en iyi sınıflandırma algoritması olmuştur. Genel olarak bakıldığında, algoritmalarındaki F1 skorları ile doğruluk değerleri de diğer algoritmalarla göre RNN için yüksektir. Ayrıca Doğruluk değerleri ile F1 skor değerleri de analizlerde kendi içerisinde uyum göstermektedir.

Tablo 4.13. B03 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,83	0,83	0,85	0,84	0,83	29
KNN	0,84	0,83	0,85	0,83	0,83	18
NN	0,89	0,88	0,84	0,89	0,90	165
DT	0,81	0,81	0,83	0,81	0,81	10
RF	0,81	0,81	0,83	0,83	0,83	36
GBM	0,89	0,90	0,90	0,89	0,89	197
FFNN	0,04	0,10	0,25	0,10	0,10	110
NB	0,82	0,83	0,85	0,84	0,82	78
XGBOOST RF	0,92	0,92	0,93	0,92	0,92	55
ADABOOST LSTM	0,03	0,06	0,50	0,39	0,03	60
CATBOOST	0,86	0,86	0,88	0,86	0,86	121
RNN	0,89	0,88	0,97	0,89	0,90	46

B04 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.14’de gösterilmiştir. ADABOOST LSTM algoritması, burada da en yüksek değerleri ile en iyi sınıflandırma algoritması olmuştur. NB ve LR algoritmaları da AUC skoru olarak

aynı değeri elde etmiş olmakla birlikte, Doğruluk değeri en iyi olan algoritma ADABOOST LSTM algoritması olmuştur.

Tablo 4.14. B04 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,84	0,82	0,94	0,87	0,85	15
KNN	0,83	0,83	0,86	0,83	0,84	27
NN	0,85	0,83	0,79	0,88	0,85	250
DT	0,82	0,81	0,83	0,82	0,81	12
RF	0,84	0,83	0,82	0,83	0,83	21
GBM	0,85	0,84	0,80	0,87	0,85	48
FFNN	0,18	0,06	0,39	0,03	0,18	25
NB	0,84	0,83	0,94	0,87	0,84	15
XGBOOST RF	0,11	0,04	0,39	0,03	0,11	108
ADABOOST LSTM	0,88	0,83	0,94	0,87	0,84	41
CATBOOST	0,85	0,83	0,80	0,87	0,85	71
RNN	0,64	0,51	0,45	0,42	0,65	158

B05 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.15’de gösterilmiştir. Bu yarı mamül için tam AUC değeri, doğruluk ve F1 skoru ile DT algoritması, algoritmalar arasında en başarılı algoritma olmuştur.

Tablo 4.15. B05 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,68	0,63	0,62	0,59	0,68	11
KNN	0,74	0,74	0,72	0,56	0,56	24
NN	0,84	0,81	0,81	0,82	0,81	45
DT	0,83	0,82	0,96	0,83	0,82	2028
RF	0,71	0,71	0,69	0,71	0,71	68
GBM	0,68	0,69	0,67	0,69	0,68	10
FFNN	0,17	0,16	0,37	0,03	0,16	69
NB	0,60	0,61	0,70	0,75	0,60	9
XGBOOST RF	0,94	0,93	0,85	0,94	0,93	340
ADABOOST LSTM	0,00	0,00	0,00	0,00	0,00	40
CATBOOST	0,79	0,79	0,78	0,79	0,79	123
RNN	0,84	0,82	0,95	0,91	0,84	20

B06 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.16’da gösterilmiştir. Algoritmalar içerisinde RNN algoritması, gerek yüksek AUC değeri, gerekse Doğruluk değeri ile öne çıkarak en başarılı algoritma olmuştur.

Tablo 4.16. B06 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,87	0,88	0,91	0,88	0,88	163
KNN	0,91	0,92	0,94	0,92	0,92	265
NN	0,93	0,93	0,95	0,93	0,93	446
DT	0,91	0,91	0,94	0,92	0,91	176
RF	0,93	0,94	0,94	0,94	0,94	209
GBM	0,93	0,94	0,95	0,94	0,94	116
FFNN	0,33	0,17	0,50	0,11	0,34	209
NB	0,74	0,71	0,83	0,76	0,76	157
XGBOOST RF	0,93	0,94	0,95	0,94	0,94	178
ADABOOST LSTM	0,56	0,57	0,52	0,57	0,56	185
CATBOOST	0,92	0,92	0,94	0,92	0,92	188
RNN	0,97	0,97	1,00	0,98	0,97	166

B07 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.17’de gösterilmiştir. RNN algoritması tüm parametrelerde en yüksek değeri almıştır.

Tablo 4.17. B07 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,70	0,71	0,81	0,73	0,70	26
KNN	0,53	0,53	0,65	0,53	0,53	27
NN	0,70	0,70	0,78	0,70	0,70	49
DT	0,47	0,46	0,60	0,47	0,46	21
RF	0,47	0,47	0,60	0,47	0,47	33
GBM	0,65	0,65	0,75	0,66	0,66	26
FFNN	0,17	0,05	0,38	0,03	0,17	38
NB	0,67	0,67	0,82	0,75	0,67	32
XGBOOST RF	0,67	0,66	0,74	0,67	0,67	20
ADABOOST LSTM	0,01	0,00	0,50	0,00	0,01	38
CATBOOST	0,60	0,60	0,70	0,60	0,60	48
RNN	0,88	0,90	1,00	0,92	0,85	70

B08 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi, Tablo 4.18’de gösterilmiştir. RNN algoritması tüm parametrelerde en yüksek değeri almıştır.

Tablo 4.18. B08 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,80	0,72	0,79	0,66	0,80	40
KNN	0,71	0,71	0,73	0,71	0,71	11
NN	0,81	0,82	0,86	0,84	0,81	50
DT	0,67	0,66	0,68	0,65	0,67	6
RF	0,67	0,67	0,69	0,67	0,67	30
GBM	0,80	0,80	0,83	0,80	0,79	15
FFNN	0,17	0,05	0,38	0,03	0,17	40
NB	0,54	0,53	0,77	0,85	0,54	6
XGBOOST RF	0,80	0,79	0,80	0,79	0,80	30
ADABOOST LSTM	0,38	0,22	0,49	0,14	0,38	30
CATBOOST	0,74	0,74	0,76	0,74	0,74	50
RNN	0,87	0,88	0,97	0,90	0,87	18

B09 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi sonuçları, Tablo 4.19’da gösterilmiştir. Algoritmalar arasında tüm performans değerleri açısından KNN algoritması daha iyi olduğundan en başarılı algoritma olarak KNN algoritması seçilmiştir.

Tablo 4.19. B09 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,88	0,88	0,92	0,88	0,89	396
KNN	0,97	0,97	0,97	0,98	0,97	355
NN	0,96	0,97	0,94	0,97	0,97	13
DT	0,96	0,96	0,96	0,96	0,96	42
RF	0,96	0,96	0,95	0,96	0,96	152
GBM	0,96	0,97	0,95	0,98	0,97	407
FFNN	0,23	0,08	0,42	0,05	0,23	82
NB	0,96	0,96	0,95	0,96	0,96	124
XGBOOST RF	0,97	0,97	0,33	0,97	0,97	79
ADABOOST LSTM	0,38	0,56	0,50	1,00	0,39	76
CATBOOST	0,97	0,97	0,97	0,97	0,97	67
RNN	0,81	0,78	0,89	0,78	0,81	228

B10 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi sonuçları, Tablo 4.20’de gösterilmiştir. Sonuçlar, AUC değeri açısından RNN algoritmasını bir nebze XGBOOST RF modelinin önüne çıkartsa da, XGBOOST RF modeli gerek Doğruluk gerekse F1 Skoru değeri açısından RNN’e göre daha iyi performans göstermiştir. Hatta

Duyarlılık değeri bile daha iyidir. Bu bilgiler ışığında yarı mamül için en başarılı algoritma olarak XGBoost RF modeli seçilmiştir.

Tablo 4.20. B10 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,83	0,83	0,85	0,84	0,83	29
KNN	0,84	0,83	0,85	0,83	0,83	18
NN	0,89	0,88	0,84	0,89	0,90	165
DT	0,81	0,81	0,83	0,81	0,81	10
RF	0,81	0,81	0,83	0,83	0,83	26
GBM	0,89	0,90	0,90	0,89	0,89	107
FFNN	0,04	0,10	0,50	0,10	0,10	100
NB	0,82	0,83	0,85	0,84	0,82	48
XGBOOST RF	0,92	0,92	0,93	0,92	0,92	55
ADABOOST LSTM	0,03	0,06	0,50	0,39	0,03	60
CATBOOST	0,86	0,86	0,88	0,86	0,86	121
RNN	0,89	0,88	0,97	0,89	0,90	41

B11 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi sonuçları, Tablo 4.21’de gösterilmiştir. Gerek Doğruluk ve F1 Skoru değerleri açısından, gerekse AUC değeri açısından, ilgili yarı mamül için, RNN algoritması, en başarılı algoritmadır.

Tablo 4.21. B11 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,57	0,56	0,68	0,56	0,57	59
KNN	0,55	0,55	0,66	0,55	0,55	64
NN	0,57	0,57	0,68	0,58	0,57	81
DT	0,52	0,53	0,65	0,53	0,53	54
RF	0,56	0,56	0,67	0,56	0,56	73
GBM	0,60	0,59	0,67	0,60	0,60	61
FFNN	33,00	0,17	0,50	0,11	0,33	78
NB	0,54	0,50	0,73	0,67	0,54	54
XGBOOST RF	0,59	0,59	0,69	0,60	0,59	76
ADABOOST LSTM	0,00	0,01	0,46	0,14	0,00	84
CATBOOST	0,61	0,61	0,71	0,62	0,61	88
RNN	0,87	0,88	0,89	0,89	0,87	73

B12 kodlu yarı mamüle ilişkin yapılan sınıflandırma analizi sonuçları, Tablo 4.22’de gösterilmiştir. Gerek Doğruluk ve F1 Skoru değerleri açısından, gerekse AUC değeri

açısından, ilgili yarı mamül için, RNN algoritması, diğer algoritmalara göre üstünlük sağlamıştır.

Tablo 4.22. B12 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,73	0,70	0,80	0,73	0,73	78
KNN	0,75	0,75	0,80	0,76	0,75	59
NN	0,81	0,82	0,86	0,83	0,81	195
DT	0,73	0,73	0,79	0,74	0,73	66
RF	0,71	0,71	0,77	0,71	0,71	86
GBM	0,84	0,82	0,85	0,82	0,81	176
FFNN	0,37	0,20	0,53	0,14	0,37	105
NB	0,63	0,55	0,68	0,57	0,63	76
XGBOOST RF	0,81	0,82	0,86	0,83	0,81	132
ADABOOST LSTM	0,69	0,67	0,73	0,79	0,69	115
CATBOOST	0,79	0,79	0,83	0,80	0,79	125
RNN	0,87	0,84	0,97	0,86	0,87	85

B13 numaraları yarı mamüle ait analiz grafiği Tablo 4.23’de sunulmuştur. Bu yarı mamüle ait algoritma sonuçları incelendiğinde, RNN algoritmasının diğer algoritmalara nazaran hem doğruluk ve F1 Skoru, hem de AUC değerleri açısından başarılı olduğu görülür.

Tablo 4.23. B13 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,66	0,62	0,79	0,74	0,66	42
KNN	0,87	0,87	0,91	0,88	0,87	40
NN	0,85	0,85	0,89	0,85	0,85	109
DT	0,88	0,88	0,91	0,88	0,88	40
RF	0,88	0,88	0,91	0,88	0,88	45
GBM	0,88	0,88	0,91	0,88	0,88	48
FFNN	0,36	0,19	0,52	0,13	0,36	53
NB	0,68	0,64	0,82	0,78	0,68	41
XGBOOST RF	0,88	0,88	0,91	0,88	0,88	57
ADABOOST LSTM	0,00	0,00	0,50	0,29	0,00	135
CATBOOST	0,88	0,88	0,91	0,88	0,88	79
RNN	0,88	0,88	0,98	0,89	0,88	46

B14 yarı mamül sonuçları Tablo 4.24’te sunulmuştur. Bu analize göre birden fazla algoritmanın yüksek tahmin başarısı gösterdiğini söylemek mümkündür. Ancak

çalışma süresi açısından bakıldığında DT algoritması diğer algoritmalara göre daha başarılıdır.

Tablo 4.24. B14 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,61	0,61	0,72	0,61	0,61	89
KNN	0,98	1,00	1,00	1,00	1,00	91
NN	0,99	1,00	0,99	1,00	1,00	147
DT	0,99	1,00	1,00	1,00	1,00	87
RF	0,99	1,00	0,99	1,00	1,00	205
GBM	0,99	1,00	0,99	1,00	1,00	128
FFNN	0,39	0,15	0,54	0,15	0,39	102
NB	0,73	0,72	0,87	0,84	0,73	90
XGBOOST RF	0,99	1,00	1,00	1,00	1,00	138
ADABOOST LSTM	0,15	0,15	0,41	0,15	0,15	143
CATBOOST	0,99	1,00	0,99	1,00	1,00	178
RNN	1,00	1,00	0,91	1,00	1,00	151

B15 algoritmasına ait sınıflandırma analizi sonuçları Tablo 4.25'te gösterilmiştir. Tablo'ya göre, doğruluk ve F1 Skoru açısından algoritmalar arasında çok yüksek bir fark olmamakla birlikte, AUC değeri açısından RNN algoritması, diğer algoritmalar karşısında daha yüksek bir başarı elde etmiştir.

Tablo 4.25. B15 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,71	0,73	0,78	0,77	0,71	17
KNN	0,74	0,74	0,79	0,74	0,74	25
NN	0,73	0,73	0,79	0,76	0,73	37
DT	0,74	0,74	0,79	0,74	0,74	45
RF	0,74	0,74	0,79	0,74	0,74	30
GBM	0,74	0,74	0,79	0,75	0,74	35
FFNN	0,31	0,15	0,48	0,10	0,31	31
NB	0,74	0,76	0,84	0,87	0,74	28
XGBOOST RF	0,74	0,73	0,42	0,74	0,74	41
ADABOOST LSTM	0,00	0,00	0,50	0,09	0,00	35
CATBOOST	0,73	0,73	0,79	0,73	0,73	42
RNN	0,74	0,76	0,92	0,87	0,74	22

B16 yarı mamül sonuçları Tablo 4.26'da sunulmuştur. Bu analize göre birden fazla algoritmanın yüksek tahmin başarısı gösterdiğini söylemek mümkündür. Ancak

çalışma süresi açısından bakıldığında KNN algoritması diğer algoritmalara göre daha başarılıdır.

Tablo 4.26. B16 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,99	0,72	0,50	0,72	0,72	35
KNN	0,99	1,00	0,99	1,00	1,00	29
NN	0,99	1,00	0,99	1,00	1,00	47
DT	0,99	1,00	1,00	1,00	1,00	36
RF	0,99	1,00	1,00	1,00	1,00	40
GBM	0,99	1,00	0,99	1,00	1,00	55
FFNN	0,10	0,10	0,25	0,10	0,10	48
NB	0,99	1,00	0,99	1,00	1,00	46
XGBOOST RF	0,99	1,00	0,99	1,00	1,00	45
ADABOOST LSTM	0,02	0,00	0,48	0,00	0,02	36
CATBOOST	0,99	0,98	1,00	0,99	0,99	55
RNN	0,99	1,00	0,99	1,00	1,00	37

B17 yarı mamül sonuçları Tablo 4.27’de sunulmuştur. Bu analize göre birden fazla algoritmanın yüksek tahmin başarısı gösterdiğini söylemek mümkündür. Ancak değerlerindeki bir nebze de olsa yükseklik açısından bakıldığında ADABOOST LSTM algoritması diğer algoritmalara göre daha başarılıdır.

Tablo 4.27. B17 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,97	0,97	0,99	0,97	0,97	420
KNN	0,97	0,98	0,97	0,98	0,98	373
NN	0,98	0,98	0,94	0,98	0,98	410
DT	0,97	0,98	0,96	0,98	0,98	154
RF	0,98	0,98	0,96	0,98	0,98	276
GBM	0,98	0,98	0,94	0,98	0,98	1229
FFNN	0,07	0,01	0,30	0,01	0,07	28
NB	0,98	0,98	0,94	0,98	0,98	1333
XGBOOST RF	0,97	0,98	0,98	0,98	0,98	209
ADABOOST LSTM	0,98	0,99	0,99	1,00	0,99	310
CATBOOST	0,98	0,98	0,94	0,98	0,98	191
RNN	0,85	0,78	1,00	0,72	0,85	373

B18 numaraları yarı mamüle ait analiz grafiği Tablo 4.28’de sunulmuştur. Bu yarı mamüle ait algoritma sonuçları incelendiğinde, RNN algoritmasının diğer

algoritmalarla nazaran hem doęruluk ve F1 Skoru, hem de AUC deęerleri aısından bařarılı olduęu grlr.

Tablo 4.28. B18 Yarı maml sınıflandırma analiz sonuları

Sınıflandırma Algoritması	Doęruluk	F1	AUC	Kesinlik	Duyarlılık	alıřma Sresi (Sn)
LR	0,34	0,28	0,50	0,23	0,34	102
KNN	0,23	0,24	0,43	0,27	0,24	40
NN	0,37	0,30	0,53	0,47	0,38	89
DT	0,26	0,28	0,47	0,31	0,26	46
RF	0,26	0,27	0,46	0,29	0,26	74
GBM	0,40	0,39	0,56	0,42	0,40	65
FFNN	0,33	0,16	0,50	0,33	0,11	60
NB	0,35	0,29	0,51	0,35	0,35	45
XGBOOST RF	0,40	0,39	0,54	0,43	0,40	66
ADABOOST LSTM	0,01	0,01	0,50	0,50	0,01	196
CATBOOST	0,37	0,36	0,52	0,36	0,37	122
RNN	0,97	0,97	0,91	0,97	0,98	114

B19 numaraları yarı mamle ait analiz grafięi Tablo 4.29’da sunulmuřtur. Bu yarı mamle ait algoritma sonuları incelendięinde, RNN algoritmasının dięer algoritmalarla nazaran hem doęruluk ve F1 skoru, hem de AUC deęerleri aısından bařarılı olduęu grlr.

Tablo 4.29. B19 Yarı maml sınıflandırma analiz sonuları

Sınıflandırma Algoritması	Doęruluk	F1	AUC	Kesinlik	Duyarlılık	alıřma Sresi (Sn)
LR	0,34	0,27	0,51	0,34	0,34	27,00
KNN	0,12	0,10	0,31	0,09	0,12	30,00
NN	0,35	0,27	0,50	0,33	0,35	29,00
DT	0,03	0,03	0,25	0,03	0,03	33,00
RF	0,03	0,03	0,27	0,03	0,03	39,00
GBM	0,30	0,29	0,49	0,29	0,30	54,00
FFNN	0,33	0,17	0,50	0,11	0,33	34,00
NB	0,34	0,27	0,51	0,34	0,34	18,00
XGBOOST RF	0,28	0,27	0,46	0,27	0,28	47,00
ADABOOST LSTM	0,00	0,00	0,00	0,00	0,00	25,00
CATBOOST	0,20	0,20	0,40	0,19	0,19	74,00
RNN	0,98	0,98	0,91	0,98	0,98	34,00

B20 yarı maml sonuları Tablo 4.30’da sunulmuřtur. Bu analize gre DT algoritması dięer algoritmalarla gre daha bařarılıdır.

Tablo 4.30. B20 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,74	0,74	0,80	0,74	0,74	372
KNN	0,91	0,91	0,93	0,91	0,91	245
NN	0,96	0,96	0,94	0,96	0,96	1648
DT	0,96	0,96	0,97	0,96	0,96	261
RF	0,92	0,92	0,94	0,92	0,92	556
GBM	0,95	0,95	0,97	0,95	0,95	2853
FFNN	0,27	0,00	0,46	0,00	0,00	122
NB	0,95	0,95	0,97	0,95	0,95	2760
XGBOOST RF	0,95	0,95	0,34	0,95	0,95	272
ADABOOST LSTM	0,64	0,50	0,67	0,64	0,41	30
CATBOOST	0,92	0,92	0,94	0,92	0,92	200
RNN	0,96	0,96	0,91	0,96	0,96	304

B21 yarı mamül sonuçları Tablo 4.31’de sunulmuştur. Bu analize göre RNN algoritması diğer algoritmalarla göre daha başarılıdır.

Tablo 4.31. B21 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,62	0,59	0,72	0,61	0,63	143
KNN	0,44	0,44	0,59	0,44	0,45	187
NN	0,64	0,63	0,74	0,64	0,64	261
DT	0,37	0,36	0,54	0,35	0,38	156
RF	0,38	0,37	0,54	0,36	0,38	154
GBM	0,63	0,63	0,73	0,63	0,64	156
FFNN	0,35	0,18	0,51	0,12	0,35	210
NB	0,48	0,41	0,61	0,41	0,49	149
XGBOOST RF	0,61	0,61	0,72	0,60	0,62	176
ADABOOST LSTM	0,00	0,00	0,00	0,00	0,00	159
CATBOOST	0,55	0,55	0,67	0,54	0,55	298
RNN	0,98	0,99	0,91	0,99	0,98	191

B22 yarı mamül sonuçları Tablo 4.32’de sunulmuştur. Bu analize göre en başarılı algoritmalar RF ve DT algoritmalarıdır. Ancak, çok küçük bir doğruluk değeri farkı ile RF algoritmasının daha başarılı olduğunu söylemek mümkündür.

Tablo 4.32. B22 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,41	0,35	0,57	0,47	0,41	108
KNN	0,62	0,62	0,71	0,63	0,62	205
NN	0,50	0,49	0,62	0,55	0,50	288
DT	0,902	0,90	0,93	0,90	0,90	141
RF	0,903	0,90	0,93	0,90	0,90	325
GBM	0,52	0,51	0,63	0,51	0,52	581
FFNN	0,26	0,11	0,45	0,07	0,26	144
NB	0,41	0,34	0,50	0,35	0,41	103
XGBOOST RF	0,51	0,47	0,60	0,51	0,51	435
ADABOOST LSTM	0,25	0,15	0,51	0,20	0,25	206
CATBOOST	0,59	0,59	0,69	0,59	0,59	213
RNN	0,50	0,49	0,65	0,82	0,50	221

B23 numaraları yarı mamüle ait analiz grafiği Tablo 4.33’de sunulmuştur. Bu yarı mamüle ait algoritma sonuçları incelendiğinde, ADABOOST LSTM algoritmasının diğer algoritmalara nazaran tüm değerlendirme parametreleri açısından başarılı olduğu görülmektedir.

Tablo 4.33. B23 Yarı mamülü sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,53	0,39	0,50	0,36	0,53	30
KNN	0,73	0,74	0,78	0,75	0,73	30
NN	0,75	0,75	0,78	0,75	0,75	37
DT	0,75	0,75	0,79	0,75	0,75	30
RF	0,75	0,75	0,79	0,75	0,75	40
GBM	0,75	0,74	0,79	0,74	0,75	70
FFNN	0,23	0,09	0,42	0,05	0,23	38
NB	0,52	0,47	0,57	0,47	0,52	32
XGBOOST RF	0,75	0,74	0,77	0,75	0,75	44
ADABOOST LSTM	0,97	0,97	0,97	0,97	43,00	45
CATBOOST	0,75	0,75	0,79	0,75	51,00	35
RNN	0,75	0,76	0,84	0,81	0,75	11

B24 yarı mamül sonuçları Tablo 4.34’te sunulmuştur. Analize göre ilk etapta AUC Skorları açısından KNN, DT, RF ve GBM algoritmaları en başarılı algoritmalarlardır.

Ancak diğer kriterler olan Doğruluk, F1 ve Kesinlik değerleri ile birlikte sonuçlar değerlendirildiğinde, RF algoritmasının daha başarılı olduğunu görülmektedir.

Tablo 4.34. B24 Yarı mamül sınıflandırma analiz sonuçları

Sınıflandırma Algoritması	Doğruluk	F1	AUC	Kesinlik	Duyarlılık	Çalışma Süresi (Sn)
LR	0,63	0,60	0,79	0,72	0,63	3
KNN	0,92	0,91	0,94	0,91	0,91	5
NN	0,92	0,93	0,91	0,93	0,92	4
DT	0,91	0,91	0,93	0,91	0,91	4
RF	0,92	0,92	0,93	0,92	0,92	2
GBM	0,920	0,92	0,93	0,92	0,92	13
FFNN	0,44	0,27	0,58	0,19	0,44	11
NB	0,92	0,92	0,91	0,92	0,92	13
XGBOOST RF	0,92	0,92	0,92	0,92	0,92	3
ADABOOST LSTM	0,04	0,05	0,59	0,07	0,04	59
CATBOOST	0,93	0,93	0,92	0,94	0,93	3
RNN	0,88	0,88	0,88	0,90	0,88	15

Algoritma Başarısı, Formül 4.1’de de gösterildiği gibi, Algoritmanın en başarılı olduğu yarı mamul sayısının, tüm yarı mamul sayısına oranı şeklinde hesaplanılmıştır.

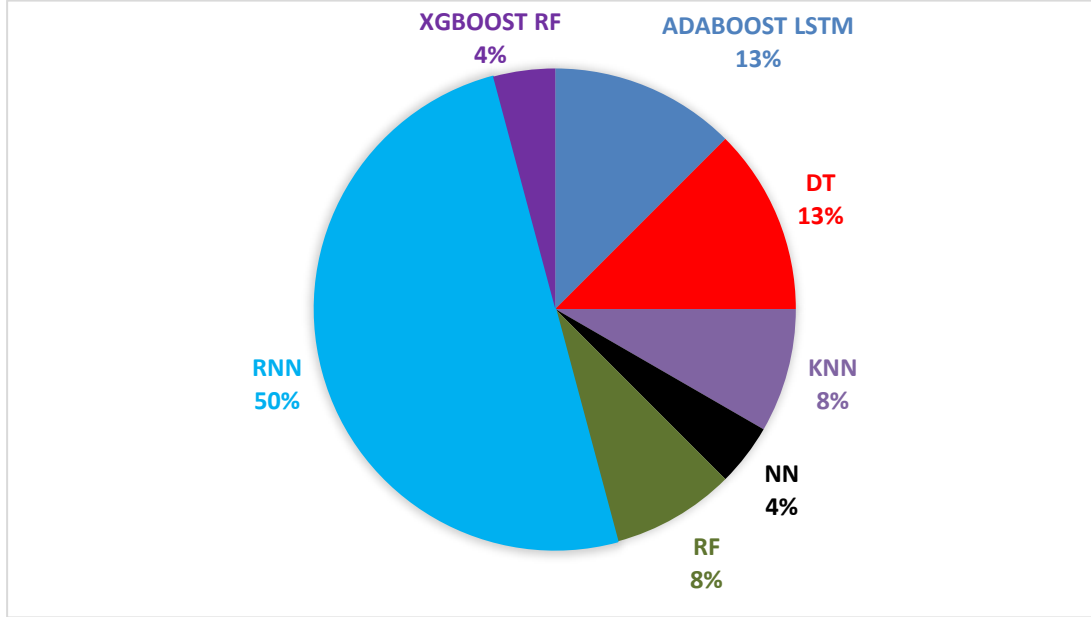
$$\text{Algoritma Başarısı (\%)} = \left(\frac{\text{Algoritmanın Başarılı Olduğu Yarı Mamul Sayısı}}{\text{Tüm Yarı Mamul Sayısı}} \right) * 100 \quad (4.1)$$

Burada örneğin 24 veri seti üzerinde RNN algoritması 12 yarı mamulde tahmin oranı en yüksek olan algoritma olduğundan, toplamda 12/24=%50 en başarılı algoritma olarak adlandırılmıştır.

Buna göre algoritmaların başarıları pasta grafiği şeklinde aşağıda Şekil 4.22’de gösterilmiştir. Burada, RNN, ADABOOST LSTM ve DT algoritmalarının, diğer algoritmalara nazaran örnek problem üzerinde yüksek çözüm etkisi olduğunu söylemek mümkündür. ADABOOST LSTM algoritması da aslında bir tür birleşik algoritmadır. Bu algoritmanın bir parçası olan LSTM algoritması, RNN algoritmasının bir türevidir. Bu bakış açısı ile bakarsak, tüm veri içerisinde %50 üzerinde en başarılı algoritmanın da RNN algoritması olduğunu söylememiz mümkündür.

Bu algoritmaları elde ettikleri kısmi başarılar ile RF, XGBOOST RF, NN ve KNN algoritmaları takip etmektedir. Burada da ADABOOST LSTM örneğinde verdiğimiz örnekte olduğu gibi, XGBOOST RF algoritması da, RF algoritmasının güçlendirilmiş

bir türevidir. Bu nedenle ikincil etken algoritmalarında da RF'in yüksek bir ağırlığı olduğunu söylemek mümkündür.



Şekil 4.22. Algoritmaların veri seti üzerindeki başarıları

4.3.4. Çalışmanın sonuçlarının uygulanması

Yapılan çalışmanın sonuçları tesiste uygulanmıştır. Buna göre, çalışmada başarılı olan algoritmalar, planlama aşamasında jupyter notebook üzerinden girdi parametreleri baz alarak kullanılmıştır. Elde edilen bulgulara istinaden de gerekli planlama değişiklikleri yapılmak sureti ile algoritmaların başarısı test edilmiştir. Bu noktada bu algoritmaların gerçek hayattaki uygulamasında da oldukça başarılı sonuçlar alınmıştır.

Elde edilen kazanımlar; planlama hazırlık aşamasında iyileşme, malzeme değişiminde azalma, tutulan nihai ürün stoğunda azalma ve rework miktarında azalma şeklinde gerçekleşmiştir.

Sistemin kullanıma başlanması ile üretim planlama hazırlama süresi düşürülmüştür. Normalde daha önce de bahsettiğimiz aksamalardan yada daha fazla insan gücüne dayalı tahminler neticesinde planlama yapıldığından, bu süreçte süreler daha uzun olmakta idi. Ancak elde edilen kazanım ile daha hızlı planlama yapılmıştır. Bu noktada elde edilen karlılık hesaplaması Tablo 4.35’de gösterilmiştir.

Tablo 4.35. Planlama hazırlık aşamasında iyileşme

Mevcut Süre (Saat/Yıl)	Çalışan Personel Sayısı (Kişi)	Kazanım (%)	Kazanım (Saat/Yıl)	Çalışan Saatlik Ücreti (€/saat)	Kazanım (€/Yıl)
285	2	50	285	10	2850

Algoritmaların devreye alınması neticesinde, makineye daha doğru plan verildiği için, üretim daha az malzeme değişimi yapılmaya başlanılmıştır. Bu da daha az hazırlık sayısını ve hazırlık süresini sağlamıştır. Böylelikle makinada daha az duruş gerçekleşmiştir. Kazanılan sürede daha fazla üretim yapılarak, kapasite artışı sağlanmıştır. Bu noktada elde edilen karlılık hesaplaması Tablo 4.36’da gösterilmiştir.

Tablo 4.36. Malzeme değişiminde azalma

Makinanın saatlik duruş maliyeti (€/saat)	Yıllık Duruş Sayısı (Adet/Yıl)	Ortalama 1 Duruş Süresi (Saat)	Kazanım (%)	Malzeme Değişimi Kazanım (€/Yıl)
10	200	12	50	12000

Algoritmaların kullanılmaya başlanılmasının ardından yapılan planın doğruluğuna güven arttığından, nihai ürün tarafında tutulan stok miktarı düşürüldü. Bu stok, aslında, öncesinde yanlış üretim yapılması durumunda müşteriye eldeki stoktan dönüş yapabilmek amacıyla tutulmakta olan bir stok idi. İstenilen ürünün istenilen zamanda üretilme yüzdesinin artışı ile, bu stoğun da çok yüksek oranda tutulma gerekliliği ortadan kaldırılmıştır. Bu noktada elde edilen karlılık hesaplaması Tablo 4.37’de gösterilmiştir.

Tablo 4.37. Nihai ürün stoğunda azalma

Mevcut Tutulan Stok Miktarı (ton)	Emniyet Stoğunda İyileşme (%)	İyileşme Sonrası Azalan Stok Miktarı (ton)	Stok Değeri (€/ton)	Stok Kazanım (€/Yıl)
1.000	20%	200	1.000	200.000

Algoritma kullanımı ile yapılmakta olan hatalı üretimlerin azalmasının bir diğer etkisi de bu nedenle yapılan reworkteki azalma olmuştur. Azaltılan rework satıldığı için bu durum karlılığı artırmıştır. Bu hesaplamalar da Tablo 4.38’de görülmektedir.

Tablo 4.38. Rework miktarında azalma

Yıllık Azaltılan Rework Miktarı (ton/Yıl)	2021 Yılı Ton Başı Satış Değeri (€/ton)	Rework Kazanımı (€/Yıl)
52	4.000	208.000

Tüm karlılık kalemleri özet tablo olarak Tablo 4.38’de gösterilmiştir. Yapılan çalışma ile yıllık kazanım yaklaşık 422.850 €/yıl olarak hesaplanmıştır.

Tablo 4.39. Özet tablo

Kalem No	Maliyet Kalemi	Kazanım (€/Yıl)
1	Planlama Hazırlık	2850
2	Malzeme Değişimi	12000
3	Nihai Ürün Stoğu	200000
4	Rework Miktarı	208000
Toplam Kazanç (€/Yıl)		422850

5. TARTIŞMA VE SONUÇ

Bu çalışmada bir alüminyum yassı mamul döküm işletmesinde, üretim esnasında ortaya çıkan yarı mamullerin Kalite Tipi'nin, üretim esnasında elde edilen girdi parametreleri ile ne seviyede tahmin edilebileceği incelenmiştir. İnceleme sonucunda bunun mümkün olduğu belirlenmiştir. Bunun yanında, sistem, eğiticişiz öğrenme kümeleme ve eğitimci öğrenme sınıflandırma yöntemleriyle analiz edilmiştir.

Sonuç olarak çalışmamızdan elde edilen önemli bulgular şu şekilde özetlenebilir:

Bu çalışma ile daha önce kişilerin tecrübelerine istinaden tahmin edilmeye çalışılan ve bu şekilde bile oldukça zaman alan bir süreç, yüksek oranda hızlandırılarak ve insani hatalardan arındırılarak tahmin edilmesi sağlanmıştır. Bu bağlamda sektör ve problem özelinde sayılı örneklerdendir.

Araştırma, tüm veri üzerinde uygulanan kümeleme algoritmalarında düşük performans alındığını göstermiştir. Bu bağlamda bu problem için tüm veri setinde kümeleme algoritması kullanmak pek seçilebilir bir yöntem değildir.

Araştırma, yarı mamül bazlı olarak kümeleme algoritmaları uygulandığında, sınırlı bir başarı elde edilebildiğini ortaya koymuştur. Burada sınırlı başarı, 2 yarı mamül cinsinde yüksek siluet (ki onlarda bile düşük AUC değeri ile) ve 3 yarı mamül cinsinde (ki onlarda bile düşük siluet) yüksek AUC değeri olarak ifade edilir. Bu durum 2 şekilde yorumlanabilir. İlk olarak, tüm veri kümelemeye nazaran, bazıısı yarı maamüllerde yüksek değerler elde edildiğinden, belki bu yarı mamül cinsleri için, kümeleme algoritmalarının bir uygulanabilirliğinden söz edilebilir. Ancak, ikinci olarak, başarı genele vurulduğunda neredeyse yok denecek kadar azdır. Ayrıca, her defasında tüm veriyi kümelemek için oldukça fazla zaman ve maliyet gerektirir. Çünkü bu sistemi çalıştırabilmek için sınıflandırma algoritmalarına göre daha güçlü bilgisayar sistemlerine (ram ve işlemci gücü) ihtiyaç duyulur.

Yarı mamül kümeleme analizlerinde 24 yarı mamülün 2 tanesinde yüksek AUC değeri ile başarılı olan kümeleme algoritması Spektral Kümeleme algoritması olmuştur.

Araştırmada, tüm veri ve yarı mamül bazlı olarak kümeleme algoritmaları ile sınıflandırma denendikten sonra, makine öğrenmesi ve derin öğrenme algoritmaları ile tüm veri üzerinde ve yarı mamül bazlı olarak sınıflandırma yapılmış ve performans ölçülmüştür. Sonuç olarak kümeleme yöntemlerine nazaran makine öğrenmesi ve derin öğrenme ile daha yüksek başarı elde edilmiştir. Bu noktada çalışmadan çıkartılabilecek önemli sonuçlardan biri de eğitici kümeleme algoritmalarına nazaran, eğitici sınıflandırma algoritmalarının güçlü bir şekilde veri üzerinde başarılı olmasıdır.

Araştırmada, tüm veri üzerinde makine öğrenmesi ve derin öğrenme algoritmaları ile sınıflandırma yapılmıştır. Sonuç olarak iyi bir başarı elde edildiği görülmüştür. İlk değerlendirme kriterleri olan AUC değeri, doğruluk değeri ve F1 Skoru açısından elde edilen değerler en iyi algorimalarda 0,80-0,91 arasında gerçekleşmiştir. Bu, oldukça iyi bir başarıdır. Algoritmalar arasından bakıldığında CATBOOST ve GBM algoritmaları bu değerlere ulaşarak en başarılı algoritmalar oluşturmaktadır. Ancak değerler açısından CATBOOST algoritması, GBM'e göre az bir farkla daha iyidir. Bu nedenle en başarılı algoritma olarak CATBOOST algoritmasının olduğu belirtilir. Elde edilen sonuçlarda algoritmalar için hesaplanan diğer kesinlik ve duyarlılık değerleri de diğer sonuçlarla paralel mantıklı değerler ortaya koyar.

Araştırmada, son olarak, yarı mamül bazlı olarak makine öğrenmesi ve derin öğrenme algoritmaları ile sınıflandırma uygulanmıştır. Burada elde edilen sonuçlarda birçok yarı mamül cinsi için tüm veri üzerinde makine öğrenmesi ve derin öğrenme ile elde edilen sınıflandırma başarısından bile çok daha yüksek derecede başarı elde edilmiştir. Elde edilen AUC, doğruluk ve F1 skor değerleri çoğu yarı mamül cinsi için 0,90-1,00 arasında gerçekleşmiştir. Bu noktada, çalışmada problem için en başarılı çözümün, yarı mamül bazında makine öğrenmesi ve derin öğrenme ile sınıflandırma yapılması olduğu ortaya konulmuştur.

Araştırma ile bize malzeme bazlı yarı mamullerde bazı algoritmaların eşit seviyelerde yüksek doğruluk ortaya konulduğu gösterilmiştir. Ancak bu algoritmalarda algoritmanın çalışma süreleri farklılık göstermiştir. En iyiyi seçerken bu durumlarda en hızlı algoritmalar tercih edilmiştir. Problemin yapısına da bakıldığında Kalite Tipi'ni mümkün olan en hızlı biçimde tahmin etmek gerekmektedir.

Makine öğrenmesi ve derin öğrenme ile yarı mamül bazlı başarılı olan algoritmalara bakıldığında, yarı mamüllerin yarısında en başarılı algoritmanın RNN algoritması olduğu saptanılmıştır. Bu sonuç, bize yarı mamüllerde sınıflandırma tahmininde bizim problemimiz için RNN algoritmasının gücünü ortaya koymaktadır. Hatta, ikinci başarılı algoritmalarından birinin de ADABOOST LSTM olması ve bu algoritmanın da bir RNN türü algoritma olması da RNN'in gücünü destekler niteliktedir.

Ortaya çıkan en önemli sonuçlardan biri de tüm veri sınıflandırmaya nazaran yarı mamul malzeme bazlı sınıflandırmanın başarısıdır. Bu durum, girdi parametrelerinin malzemeden malzemeye farklı etki gösterdiği şeklinde açıklanabilir. Kimi malzemede çıktıyı pozitif ve çok güçlü etkileyen girdi, kimi malzemede negatif ya da düşük etki verebilmektedir.

Girdiler ile Çıktılar arasındaki fonksiyonel bağlantıyı her algoritmanın analiz yöntemi farklı olduğundan, ilgili yarı mamül için bağlantıyı en iyi yakalayan algoritmalar diğerlerine göre daha başarılıdır.

Bu sonuçlar ışığında realitede üretimde de hangi yarı mamül üretim planına alınacaksa, o yarı mamülde çalışmamızda en başarılı olan algoritmanın çalıştırılarak tahmin yaptığı bir program, yazılım yapısı oluşturulmuştur. Yapı halihazırda da kullanılmaktadır ve gerçek hayatta da oldukça müspet sonuçlar alınmaktadır. Bu sonuçlar maliyetlerin düşmesine ve karlılığın da artmasına yansımıştır.



KAYNAKLAR

- Akküçük, U. (2011) *Veri madenciliği: kümeleme ve sınıflama algoritmaları*.
- Alper, M.G. (2003) 'Alüminyum Sürekli Döküm Yöntemi İle Üretilmiş 5052 – 5182 Alüminyum Alaşımlarının Şekillendirilebilirlik Kabiliyetlerinin Belirlenmesi', YL Tezi
- Aslanyürek, M. ve Mesut, A. (2021) 'Kümeleme Performansını Ölçmek için Yeni Bir Yöntem ve Metin Kümeleme için Değerlendirmesi', *European Journal of Science and Technology* [Preprint]. Available at: <https://doi.org/10.31590/ejosat.932938>.
- Ayanoğlu, C.C., Kurt, M., 2019. Metal Sektöründe Veri Madenciliği Yöntemleri İle Bir İş Kazası Tahmin Modeli Önerisi. *Ergonomi* 2, 78–87. <https://doi.org/10.33439/ergonomi.481861>
- B07 Committee, n.d. (2022) 'Specification for Aluminum and Aluminum-Alloy Sheet and Plate'. https://doi.org/10.1520/B0209_B0209M-21
- Bai, Y., Xie, J., Wang, D., Zhang, W. ve Li, C. (2021) 'A manufacturing quality prediction model based on AdaBoost-LSTM with rough knowledge', *Computers & Industrial Engineering*, 155, p. 107227. Available at: <https://doi.org/10.1016/j.cie.2021.107227>.
- Bagheripoor, M. ve Bisadi, H. (2013) 'Application of artificial neural networks for the prediction of roll force and roll torque in hot strip rolling process', *Applied Mathematical Modelling*, 37(7), pp. 4593–4607. Available at: <https://doi.org/10.1016/j.apm.2012.09.070>.
- Bak, C., Roy, A.G. ve Son, H. (2021) 'Quality prediction for aluminum diecasting process based on shallow neural network and data feature selection technique', *CIRP Journal of Manufacturing Science and Technology*, 33, pp. 327–338. Available at: <https://doi.org/10.1016/j.cirpj.2021.04.001>.
- Bakırarar, B., Kar, İ., Gökmen, D., Elhan, A.H. ve Genç, V. (2019) 'The Prediction of Breast Biopsy Outcomes Using Two Data Mining Algorithms Based on Parameter Variations', *Türkiye Klinikleri Journal of Biostatistics*, 11(2), pp. 133–142. Available at: <https://doi.org/10.5336/biostatic.2019-64754>.
- Bayseçkin, E. (2002) 'Sürekli Döküm Yöntemiyle Üretilen Alüminyum Levhalarda Döküm ve Homojenleştirme Sürecinde Mikro Yapısal Değişimin İncelenmesi', YL Tezi
- Bilgin, T. ve Çamurcu Y. (2005) 'DBSCAN, OPTICS ve K-Means Kümeleme Algoritmalarının Uygulamalı Karşılaştırılması', *Politeknik Dergisi*, Cilt 8, Sayı 2.
- Boto, F., Murua, M., Gutierrez, T., Casado, S., Carrillo, A., Arteaga, A. (2022) 'Data Driven Performance Prediction in Steel Making', *Metals* 12, 172. <https://doi.org/10.3390/met12020172>

- Bramer, M. (2013) *Principles of Data Mining*. London: Springer London (Undergraduate Topics in Computer Science). Available at: <https://doi.org/10.1007/978-1-4471-4884-5>.
- Bradley, A.P. (1997) 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognition*, 30(7), pp. 1145–1159. Available at: [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2).
- Brunelli, L., Masiero, C., Tosato, D., Beghi, A. ve Susto, G.A. (2019) 'Deep Learning-based Production Forecasting in Manufacturing: a Packaging Equipment Case Study', *Procedia Manufacturing*, 38, pp. 248–255. Available at: <https://doi.org/10.1016/j.promfg.2020.01.033>.
- Carlin, D., O’Kane, P. ve Sezer, S. (2019) 'A cost analysis of machine learning using dynamic runtime opcodes for malware detection', *Computers & Security*, 85, pp. 138–155. Available at: <https://doi.org/10.1016/j.cose.2019.04.018>.
- Cadavid, J.P.U., Lamouri, S., Grabot, B. ve Fortin, A. (2019) 'Machine Learning in Production Planning and Control: A Review of Empirical Literature', *IFAC-PapersOnLine*, 52(13), pp. 385–390. Available at: <https://doi.org/10.1016/j.ifacol.2019.11.155>.
- Cavallaro, C., Cutello V., Pavone, M. ve Zito, F. (2024) 'Machine Learning and Genetic Algorithms: A case study on image reconstruction', *Knowledge-Based Systems*, 284, p. 111194. Available at: <https://doi.org/10.1016/j.knosys.2023.111194>.
- Chandra, D.K., Ravi, V. ve Bose, I. (2009) 'Failure prediction of dotcom companies using hybrid intelligent techniques', *Expert Systems with Applications*, 36(3), pp. 4830–4837. Available at: <https://doi.org/10.1016/j.eswa.2008.05.047>.
- Chakrabarti, A., Sukumar, R.P., Jarke, M., Rudack, M., Buske, P., Holly, C. (2021) 'Efficient Modeling of Digital Shadows for Production Processes: A Case Study for Quality Prediction in High Pressure Die Casting Processes', in: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA). Presented at the 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), IEEE, Porto, Portugal, pp. 1–9. <https://doi.org/10.1109/DSAA53316.2021.9564113>
- Chen, T. (2017) 'An ANN approach for modeling the multisource yield learning process with semiconductor manufacturing as an example', *Computers & Industrial Engineering*, 103, pp. 98–104. Available at: <https://doi.org/10.1016/j.cie.2016.11.021>.
- Cielen, D., Meysman, A. ve Ali, M. (2016) *Introducing data science: big data, machine learning, and more, using Python tools*. Shelter Island, NY: Manning Publications.
- Costa, M.A., Wullt, B., Norrlöf, M. ve Gunnarsson, S. (2019) 'Failure detection in robotic arms using statistical modeling, machine learning and hybrid gradient boosting', *Measurement*, 146, pp. 425–436. Available at: <https://doi.org/10.1016/j.measurement.2019.06.039>.

- Deng, J., Bai, Y. ve Li, C. (2020) ‘A Deep Regression Model with Low-Dimensional Feature Extraction for Multi-Parameter Manufacturing Quality Prediction’, *Applied Sciences*, 10(7), p. 2522. Available at: <https://doi.org/10.3390/app10072522>.
- Ding, S., Zhang, L. ve Zhang, Y. (2010) ‘Research on Spectral Clustering algorithms and prospects’, in 2010 2nd International Conference on Computer Engineering and Technology. 2010 2nd International Conference on Computer Engineering and Technology, Chengdu, China: IEEE, pp. V6-149-V6-153. Available at: <https://doi.org/10.1109/ICCET.2010.5486345>.
- D’Souza, A.D., Rao, S.S. ve Herbert, M.A. (2023) ‘A study of microstructure and mechanical properties of friction stir welded joint of Al-Ce-Si-Mg aluminium alloy plates and optimization cum prediction techniques using Taguchi and ANN’, *Materials Today: Proceedings*, 92, pp. 162–176. Available at: <https://doi.org/10.1016/j.matpr.2023.04.125>.
- E04 Committee, (2023) ‘Guide for Preparation of Metallographic Specimens’, <https://doi.org/10.1520/E0003-11R17>
- Ecemiş, O., Irmak, S., 2018 ‘Paslanmaz Çelik Sektörü Satış Tahmininde Veri Madenciliği Yöntemlerinin Karşılaştırılması’, *Sosyal Bilimler Dergisi*, 8 (1), 148-169
- Fan, S., K.S., Lin, S.C. ve Tsai, P.F. (2016) ‘Wafer fault detection and key step identification for semiconductor manufacturing using principal component analysis, AdaBoost and decision tree’, *Journal of Industrial and Production Engineering*, 33(3), pp. 151–168. Available at: <https://doi.org/10.1080/21681015.2015.1126654>.
- Fang, W., Huang, J., Peng, T., Long, Y., Yin, F. (2024). Machine learning-based performance predictions for steels considering manufacturing process parameters: a review. *J. Iron Steel Res. Int.* <https://doi.org/10.1007/s42243-024-01179-5>
- Ferguson, M., Ak, R., Lee, Y.-T.T., & Law, K. H. (2018). Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning. *Smart and Sustainable Manufacturing Systems*, 2(1), 137–164. <https://doi.org/10.1520/SSMS20180033>
- Guan, B., Wang, D., Shu, D., Zhu, S., Ji, X., Sun, B., (2024) ‘Data-driven casting defect prediction model for sand casting based on random forest classification algorithm’, *China Foundry* 21, 137–146. <https://doi.org/10.1007/s41230-024-3090-1>
- Han, J., Kamber, M. ve Pei, J. (2011) ‘Data Mining: Concepts and Techniques’, 3rd. Edition, Morgan Kaufmann, USA, 5-7.
- Hao, L., Bian, L., Gebraeel, N., ve Shi, J. (2017). Residual life prediction of multistage manufacturing processes with interaction between tool wear and product quality degradation. *IEEE Transactions on Automation Science and Engineering*, 14(2), 1211–1224.
- Hair, Jr. F.J., Black, W.C., Babin, B.J. ve Anderson, R.E. (1998) ‘Multivariate Data Analysis With Readings’, 7th Ed., Prentice-Hall, USA.

- Hu, J., Qian, X., Pei, Y., Tan, C., Pardalos, P.M. ve Liu, X. (2022) ‘A novel quality prediction method based on feature selection considering high dimensional product quality data’, *Journal of Industrial and Management Optimization*, 18(4), p. 2977. Available at: <https://doi.org/10.3934/jimo.2021099>.
- İşler, Y. (2014), *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*
- Jiang, M., Li, X., Wu, J. ve Wang, G. (2014) ‘A precision on-line model for the prediction of thermal crown in hot rolling processes’, *International Journal of Heat and Mass Transfer*, 78, pp. 967–973. Available at: <https://doi.org/10.1016/j.ijheatmasstransfer.2014.07.061>.
- Kao, L.J., Lee, T.S. ve Lu, C.J. (2016) ‘A multi-stage control chart pattern recognition scheme based on independent component analysis and support vector machine’, *Journal of Intelligent Manufacturing*, 27(3), pp. 653–664. Available at: <https://doi.org/10.1007/s10845-014-0903-x>.
- Kaggle. (2017). Pima Indians Diabetes Database, Predict the onset of diabetes based on diagnostic measures. Alındı <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database> [Erişim Tarihi: 30.02.2024]
- Kaggle. (2019). Pima Indians Diabetes- Logistic Regression. Alındı <https://www.kaggle.com/logeshk/pima-indians-diabetes-logistic-regression> [Erişim Tarihi: 06.01.2022]
- Kaggle (2021). SVM Tutorial with Diabetes Dataset. Alındı <https://www.kaggle.com/code/rishidamarla/svm-tutorial-with-diabetes-dataset/notebook> [Erişim tarihi: 01.03.2024]
- Khanam, J.J. ve Foo, S.Y. (2021) ‘A comparison of machine learning algorithms for diabetes prediction’, *ICT Express*, 7(4), pp. 432–439. Available at: <https://doi.org/10.1016/j.icte.2021.02.004>.
- Kim, A., Oh, K., Jung, J.-Y., Kim, B., (2018) ‘Imbalanced classification of manufacturing quality conditions using cost-sensitive decision tree ensembles’, *International Journal of Computer Integrated Manufacturing* 31, 701–717. <https://doi.org/10.1080/0951192X.2017.1407447>
- Kiraz, A. (2014) ‘Yapay Zeka Destekli Sanal Laboratuvar Tasarımı: Çekme Deneyi Uygulaması’, *Doktora Tezi*
- Kitao, N., Hirose, Y. ve Hamada, K. (1981) ‘Flatness Control System for Hot Strip Mills’, *IFAC Proceedings Volumes*, 14(2), pp. 2483–2488. Available at: [https://doi.org/10.1016/S1474-6670\(17\)63841-2](https://doi.org/10.1016/S1474-6670(17)63841-2).
- Kurucan, M., Özbaltan, M., Yetkin, Z. ve Alkaya, A. (2024) ‘Applications of artificial neural network based battery management systems: A literature review’, *Renewable and Sustainable Energy Reviews*, 192, p. 114262. Available at: <https://doi.org/10.1016/j.rser.2023.114262>.
- Langone, R., Alzate, C., De Ketelaere, B., Vlasselaer, J., Meert, W. ve Suykens J.A.K., (2015) ‘LS-SVM based spectral clustering and regression for predicting maintenance of industrial machines’, *Engineering Applications of Artificial Intelligence*, 37, pp. 268–278. Available at: <https://doi.org/10.1016/j.engappai.2014.09.008>.

- Lee, M., Park, I.S., Kim, M.S., Lee, W.I. ve Park, P. (2018) ‘Development of Camber Control Algorithm in Hot Strip Rolling Process’, in 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Nara: IEEE, pp. 759–764. Available at: <https://doi.org/10.23919/SICE.2018.8492579>.
- Lee, J., Noh, S.D., Kim, H.-J., Kang, Y.-S., 2018. Implementation of Cyber-Physical Production Systems for Quality Prediction and Operation Control in Metal Casting. *Sensors* 18, 1428. <https://doi.org/10.3390/s18051428>
- Lejon, E., Kyösti, P. ve Lindström, J. (2018) ‘Machine learning for detection of anomalies in press-hardening: Selection of efficient methods’, *Procedia CIRP*, 72, pp. 1079–1083. Available at: <https://doi.org/10.1016/j.procir.2018.03.221>.
- Lieber, D., Stolpe, M., Konrad, B., Deuse, J., Morik, K. (2013) ‘Quality Prediction in Interlinked Manufacturing Processes based on Supervised & Unsupervised Machine Learning’, *Procedia CIRP*, 7, pp. 193–198. Available at: <https://doi.org/10.1016/j.procir.2013.05.033>.
- Li, F., He, A., Song, Y., Shen, C., Wang, F., Yuan, T., Zhang, S., Xu, X., Qiang, Y., Liu, C., Liu, P., Zhao, Q. (2023). ‘Improving imbalanced industrial datasets to enhance the accuracy of mechanical property prediction and process optimization for strip steel’. *J Intell Manuf.* <https://doi.org/10.1007/s10845-023-02275-1>
- Li, Z., Li, J., Wang, Y. ve Wang, L. (2019) ‘A deep learning approach for anomaly detection based on SAE and LSTM in mechanical equipment’, *The International Journal of Advanced Manufacturing Technology*, 103(1–4), pp. 499–510. Available at: <https://doi.org/10.1007/s00170-019-03557-w>.
- Lughofer, E., Zavoianu, A.C., Pollak, R., Pratama, M., Meyer-Heye, P., Zörrer, H., Eitzinger, C. ve Radauerd, T. (2019) ‘Autonomous supervision and optimization of product quality in a multi-stage manufacturing process based on self-adaptive prediction models’, *Journal of Process Control*, 76, pp. 27–45. Available at: <https://doi.org/10.1016/j.jprocont.2019.02.005>.
- Majid, N.A.A., Young, B.R., Taylor, M.P. ve Chen J.J. (2012) ‘K-means clustering pre-analysis for fault diagnosis in an aluminium smelting process’, in 2012 4th Conference on Data Mining and Optimization (DMO). 2012 4th Conference on Data Mining and Optimization (DMO), Langkawi: IEEE, pp. 43–46. Available at: <https://doi.org/10.1109/DMO.2012.6329796>.
- Mazur, I. ve Koinov, T. (2016) ‘Quality Control system for a hot-rolled metal surface’, *Frattura ed Integrità Strutturale*, 10(37), pp. 287–296. Available at: <https://doi.org/10.3221/IGF-ESIS.37.38>.
- Meyes, R., Donauer, J., Schmeing, A. ve Meisen T. (2019) ‘A Recurrent Neural Network Architecture for Failure Prediction in Deep Drawing Sensory Time Series Data’, *Procedia Manufacturing*, 34, pp. 789–797. Available at: <https://doi.org/10.1016/j.promfg.2019.06.205>.
- Paul, S.K. (2016) ‘Prediction of complete forming limit diagram from tensile properties of various steel sheets by a nonlinear regression based approach’, *Journal of Manufacturing Processes*, 23, pp. 192–200. Available at: <https://doi.org/10.1016/j.jmapro.2016.06.005>.

- Powers, D. (2007) 'Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation'.
- PSI. (2021). Predictive Quality. Alındı https://www.psimetals.de/fileadmin/files/downloads/PSI_BT/2_Product_Flyer/PSImetals_AI_Predictive_Quality.english.pdf [Erişim Tarihi: 23.08.2021]
- Rai, J.K., Lajimi, A.M., Xirouchakis, P. (2008). 'An intelligent system for predicting HPDC process variables in interactive environment', *Journal of Materials Processing Technology* 203, 72–79. <https://doi.org/10.1016/j.jmatprotec.2007.10.011>
- Rajesh, S., Suresh Kumar, R., Madhankumar, S., Sheshan, M., Vignesh, M., Sanjay Kumar, R., (2021) 'Study of the mechanical properties of Al7075 alloy, silicon carbide and fly ash composites manufactured by stir casting technique'. *Materials Today: Proceedings* 45, 6438–6443. <https://doi.org/10.1016/j.matpr.2020.11.278>
- Sağlam, M. (2013) 'Çok Değişkenli İstatistiksel Yöntemler ile Toprak Özelliklerinin Gruplandırılması', *Water Journal*.
- Schmidt, B. ve Wang, L. (2018) 'Predictive Maintenance of Machine Tool Linear Axes: A Case from Manufacturing Industry', *Procedia Manufacturing*, 17, pp. 118–125. Available at: <https://doi.org/10.1016/j.promfg.2018.10.022>.
- Schmitt, J., Böning, J., Borggräfe, T., Beitingner, G. ve Deuse, J. (2020) 'Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing', *Advanced Engineering Informatics*, 45, p. 101101. Available at: <https://doi.org/10.1016/j.aei.2020.101101>.
- Seetharaman, R., Ravisankar, V., Balasubramanian, V., (2015) 'Corrosion performance of friction stir welded AA2024 aluminium alloy under salt fog conditions', *Transactions of Nonferrous Metals Society of China* 25, 1427–1438. [https://doi.org/10.1016/S1003-6326\(15\)63742-6](https://doi.org/10.1016/S1003-6326(15)63742-6)
- Simsir, F., Civak, H. ve Uygun, Ö. (2017) 'Demir Çelik Sektöründe Bulanık Mantık Yaklaşımıyla Optimal Stok Seviyelerinin Belirlenmesi: Kardemir A.Ş. Örneği', *5th International Symposium on Innovative Technologies in Engineering and Science*, pp. 1797–1808.
- Singh, P.K., (2021). 'Mechanical characterization of graphene-aluminum nanocomposites.' *Materials Today: Proceedings* 44, 2304–2308. <https://doi.org/10.1016/j.matpr.2020.12.395>
- Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C. ve Johannes, R.S. (1988) 'Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus'.
- Stackoverflow. (2021). Sklearn AffinityPropagation MemoryError. Alındı <https://stackoverflow.com/questions/35901718/sklearn-affinitypropagation-memoryerror> [Erişim Tarihi: 22.02.2024]
- Stackoverflow. (2021). MemoryError: Unable to allocate 617. GiB for an array with shape (82754714206,) and data type float64 On Windows and using Python. Alındı <https://stackoverflow.com/questions/67762297/memoryerror-unable-to-allocate-617-gib-for-an-array-with-shape-82754714206> [Erişim Tarihi: 20.02.2024]

- Statistics How To. (Tarih Yok). Alındı <https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/> [Erişim Tarihi: 15.02.2024]
- Stehman, S.V. (1997) 'Selecting and interpreting measures of thematic classification accuracy', *Remote Sensing of Environment*, 62(1), pp. 77–89. Available at: [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7).
- Takaloğlu, M., Takaloğlu, F. (2019) 'K-Means ve Hiyerarşik Kümeleme Algoritmasının Weka ve Matlab Platformlarında Karşılaştırılması', *İstanbul Aydın Üniversitesi Dergisi- İAÜD- ISSN: 1309-1352*, Temmuz 2019 Cilt 11 Sayı 3 (303-317)
- Takci, H. (2022) 'Optimum parametreler yardımıyla performansı artırılmış KNN algoritması tabanlı kalp hastalığı tahmini', *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 38(1), pp. 451–460. Available at: <https://doi.org/10.17341/gazimmfd.977127>.
- Tatlıdil, H. (1996) 'Uygulamalı Çok Değişkenli İstatistiksel Analiz', Cem Web Ofset, Ankara.
- Thankachan, T., Prakash, K.S., Malini, R., Ramu, S., Sundararaj, P., Rajandran, S., Rammasamy, D. ve Joth, S. (2019) 'Prediction of surface roughness and material removal rate in wire electrical discharge machining on aluminum based alloys/composites using Taguchi coupled Grey Relational Analysis and Artificial Neural Networks', *Applied Surface Science*, 472, pp. 22–35. Available at: <https://doi.org/10.1016/j.apsusc.2018.06.117>.
- Vannucci, M. ve Colla, V. (2011) 'Novel classification method for sensitive problems and uneven datasets based on neural networks and fuzzy logic', *Applied Soft Computing*, 11(2), pp. 2383–2390. Available at: <https://doi.org/10.1016/j.asoc.2010.09.001>.
- Wu, X., Jin, H., Ye, X., Wang, Jianjia, Lei, Z., Liu, Y., Wang, Jie, Guo, Y., 2020. Multiscale Convolutional and Recurrent Neural Network for Quality Prediction of Continuous Casting Slabs. *Processes* 9, 33. <https://doi.org/10.3390/pr9010033>
- Zhou, J.T., Zhao, X. ve Gao, J. (2019) 'Tool remaining useful life prediction method based on LSTM under variable working conditions', *The International Journal of Advanced Manufacturing Technology*, 104(9–12), pp. 4715–4726. Available at: <https://doi.org/10.1007/s00170-019-04349-y>.
- Zhou, J.T., Zhao, X. ve Gao, J. (2019) 'Tool remaining useful life prediction method based on LSTM under variable working conditions', *The International Journal of Advanced Manufacturing Technology*, 104(9–12), pp. 4715–4726. Available at: <https://doi.org/10.1007/s00170-019-04349-y>.
- Zhou, Q., Yang, Z., Hu, C., Xing, J., Xu, C., Xue, Z., Ma, C., Li, L., (2023) 'Prediction of the mechanical behavior of steel-aluminum flow drill screw joints using artificial neural network', *Int J Adv Manuf Technol* 129, 4553–4567. <https://doi.org/10.1007/s00170-023-12563-y>
- Zhang, T., Ramakrishnan, R. ve Livny, M. (1997) 'BIRCH: A New Data Clustering Algorithm and Its Applications'.

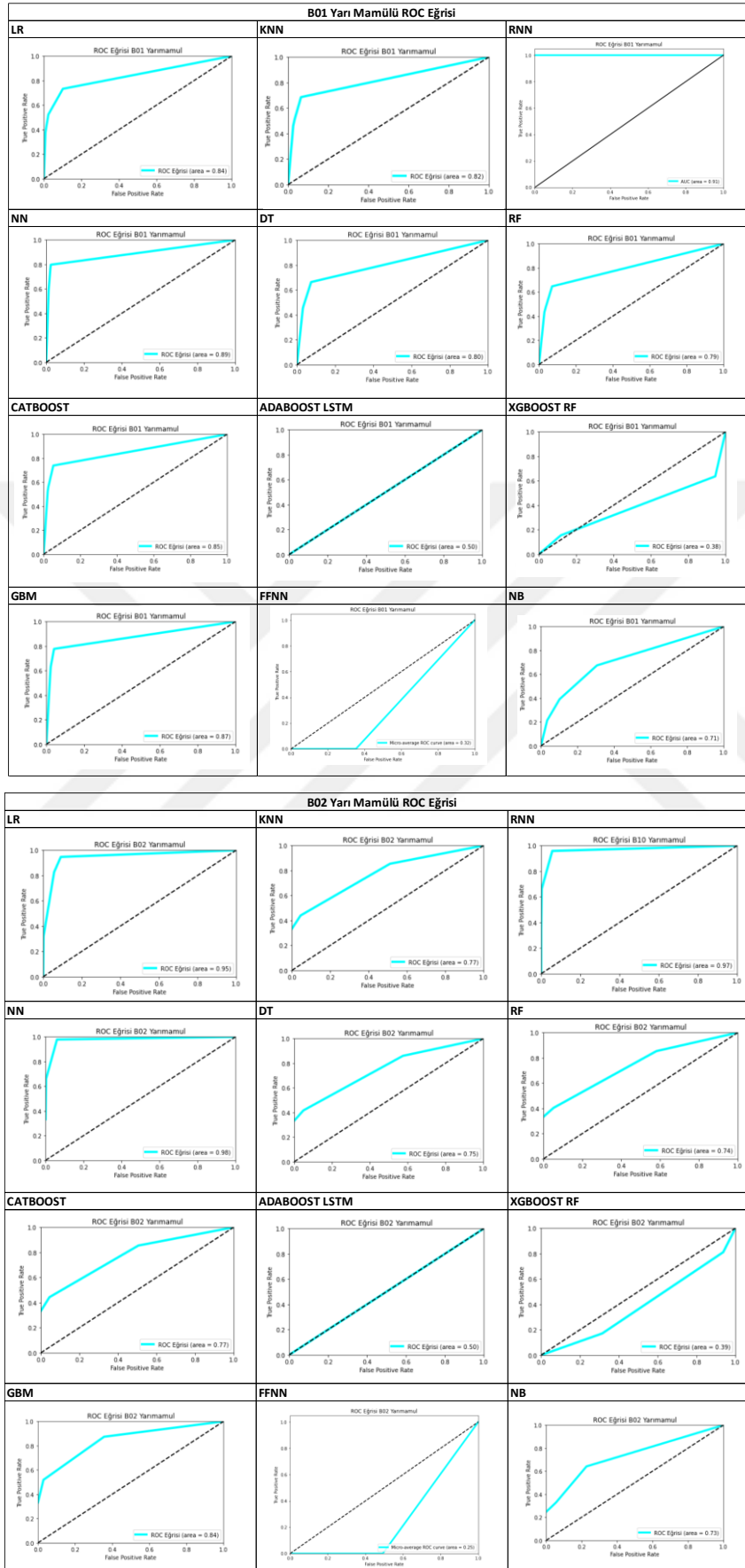


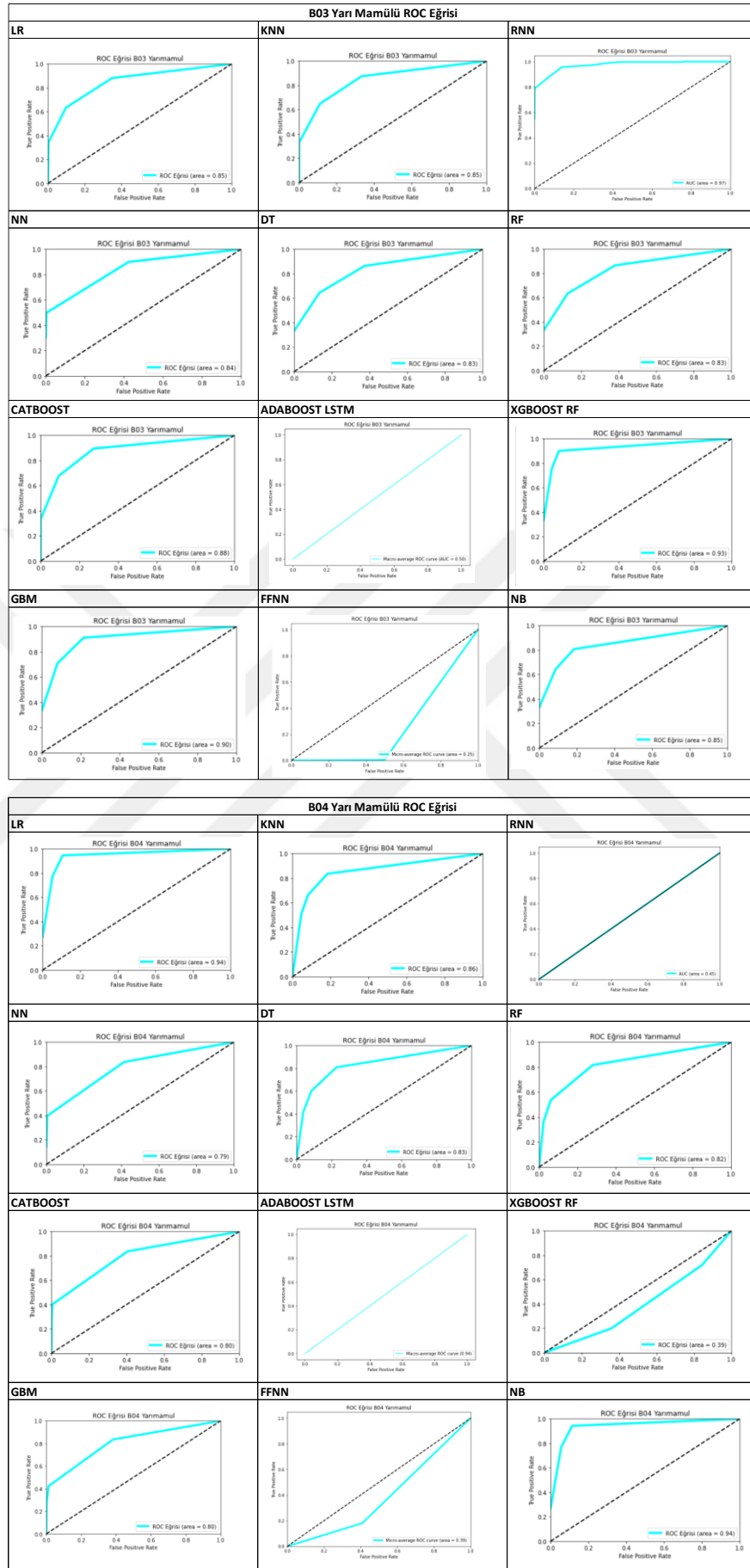
EKLER

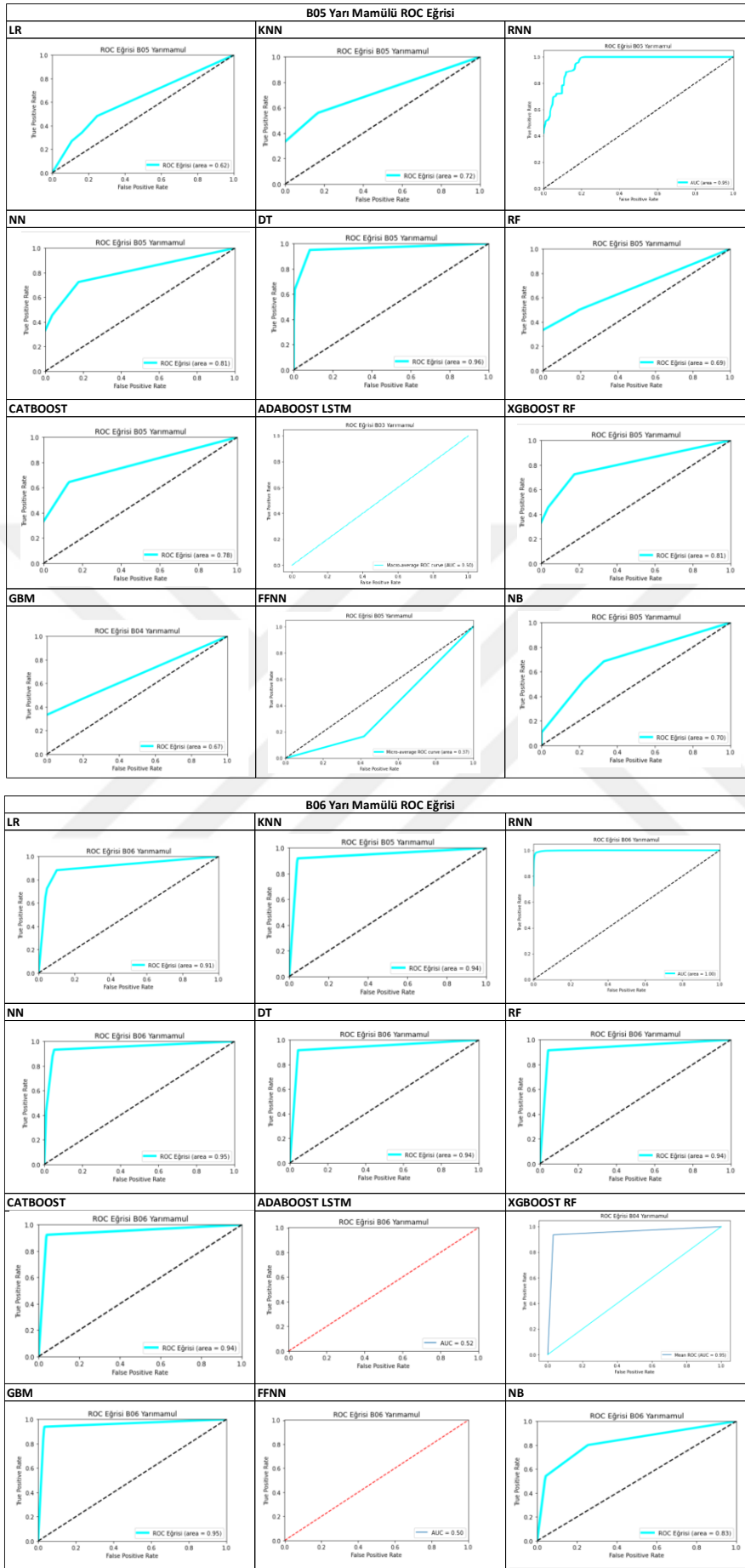
EK A1. Yarı mamül bazlı sınıflandırma analizi roc eğrileri

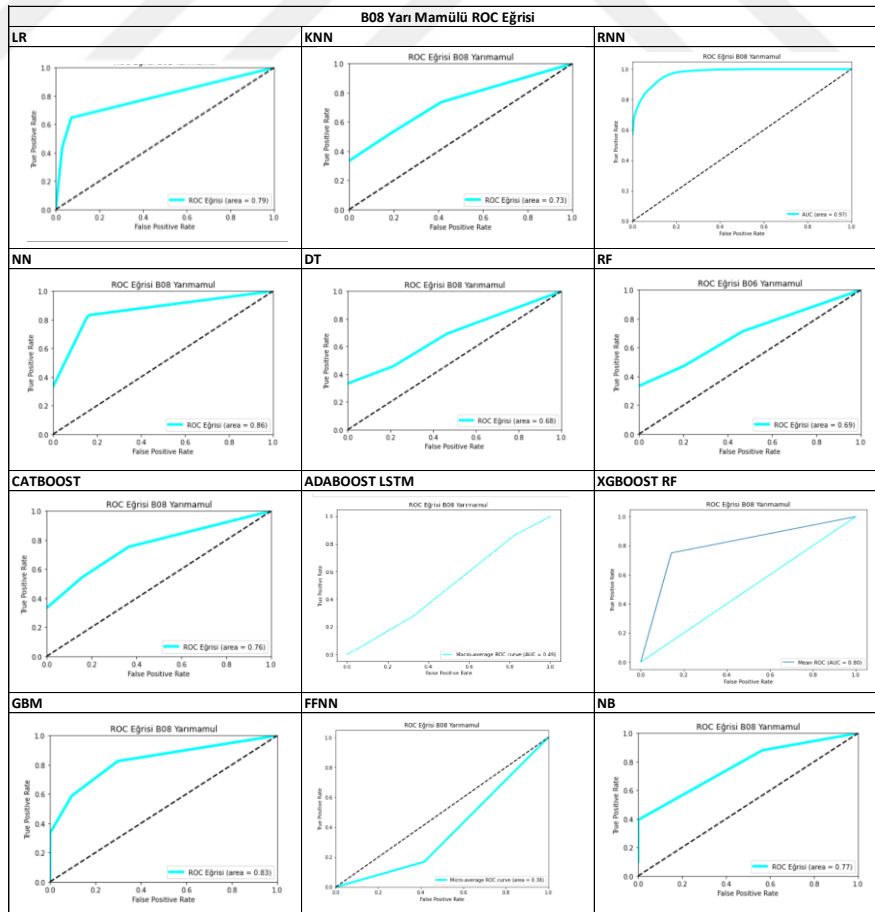
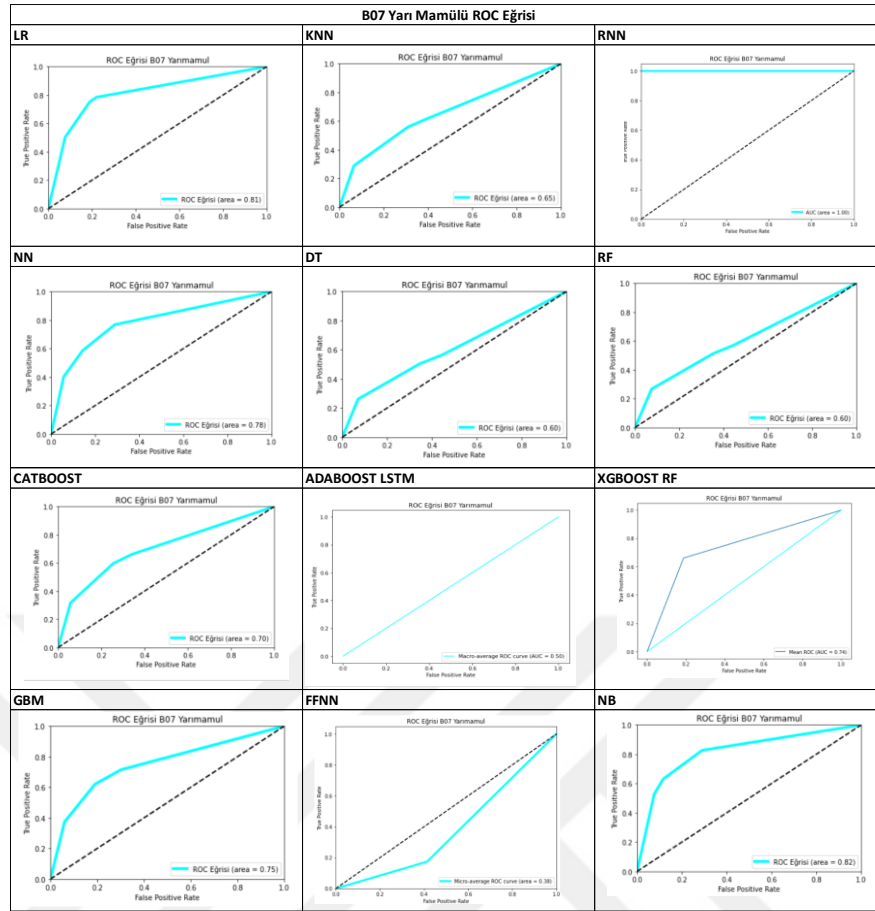


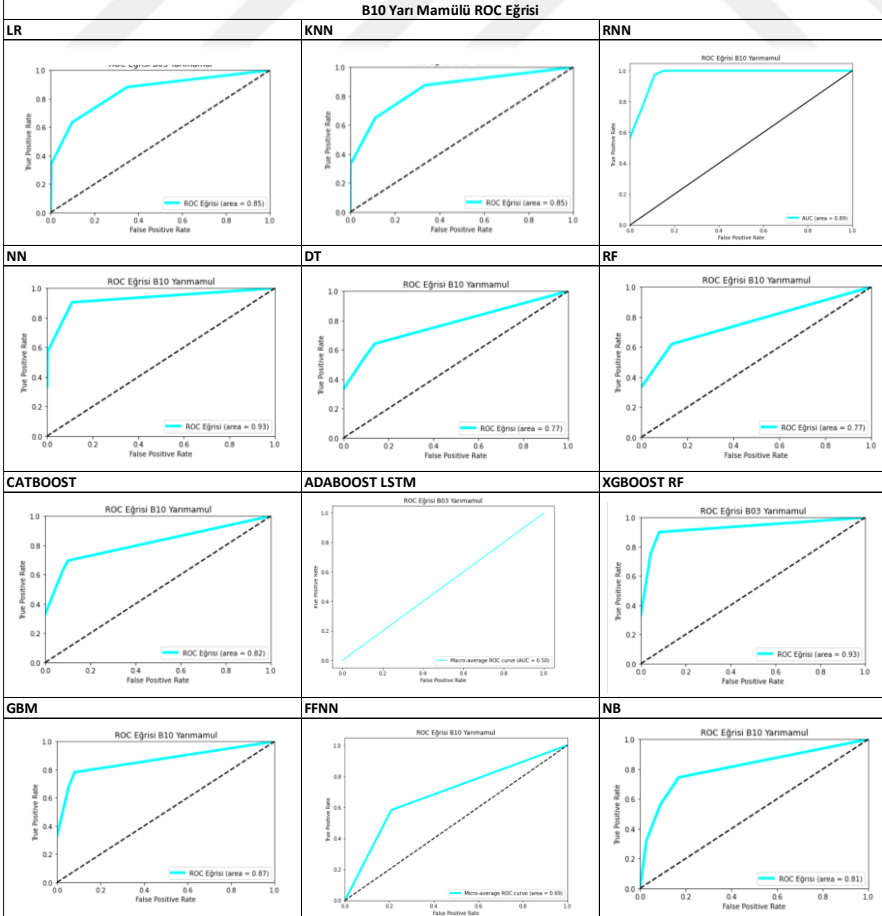
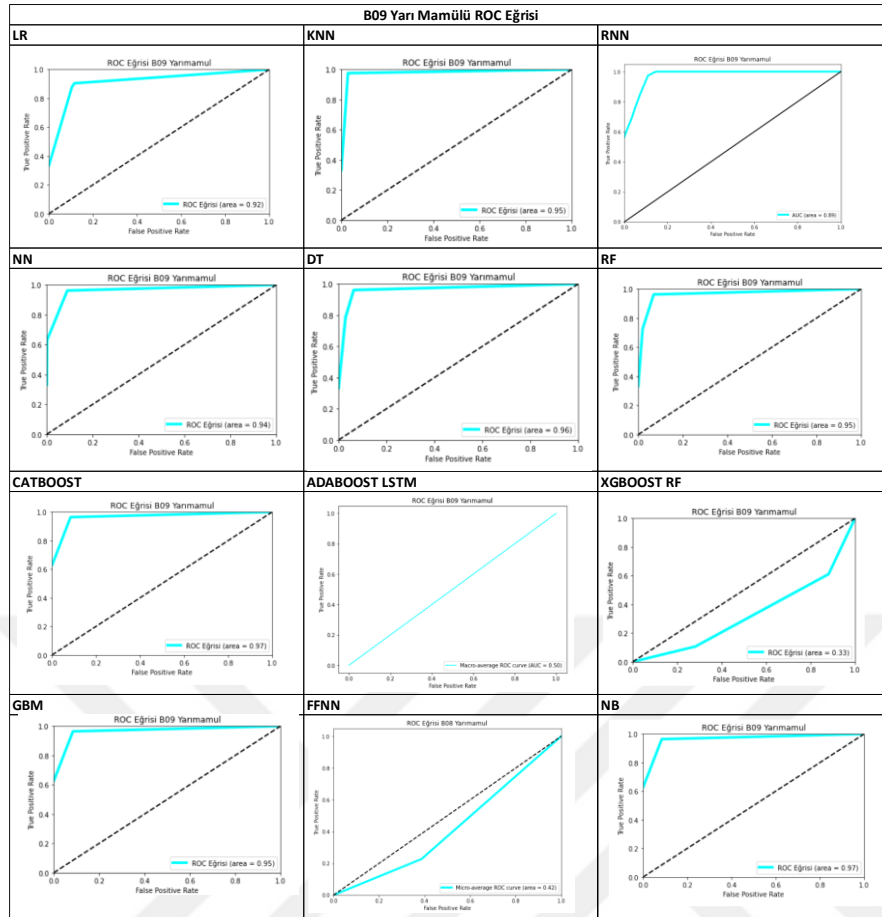
EK A1. Yarı mamül bazlı sınıflandırma analizi roc eğrileri

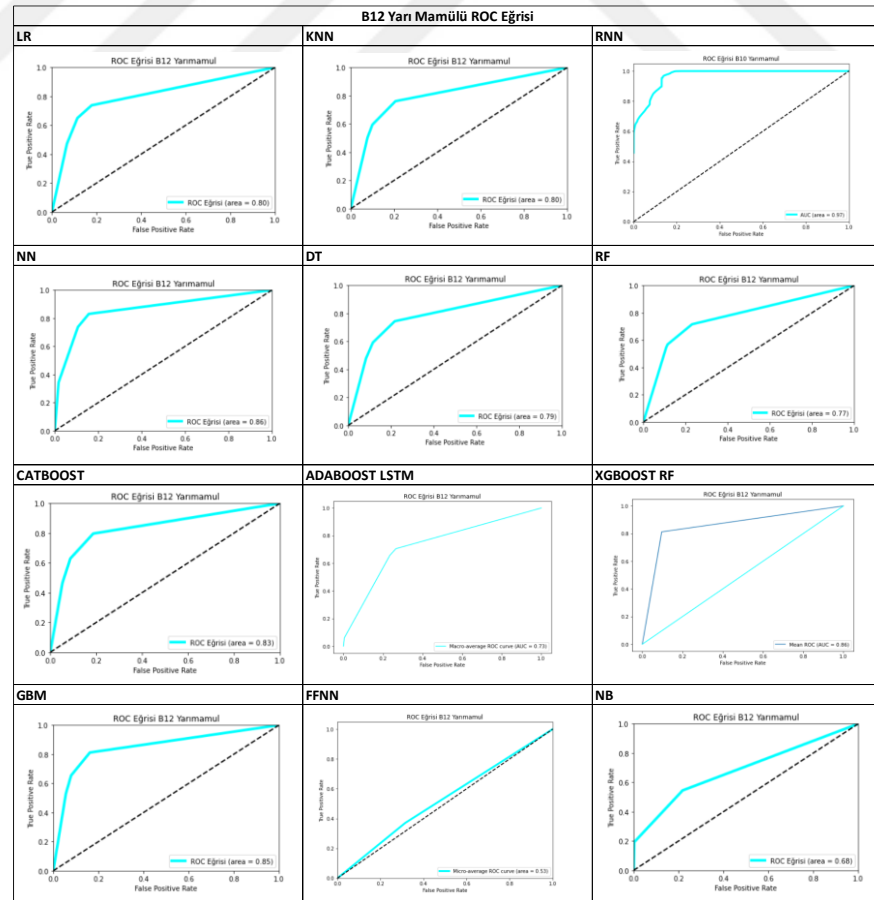
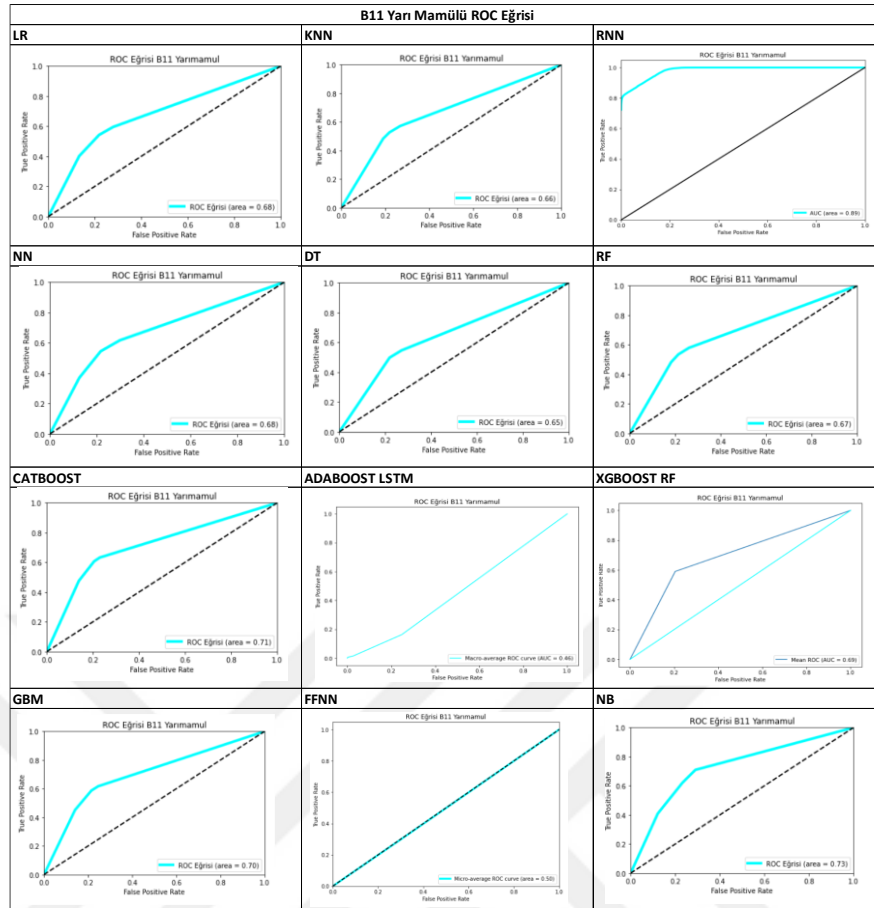


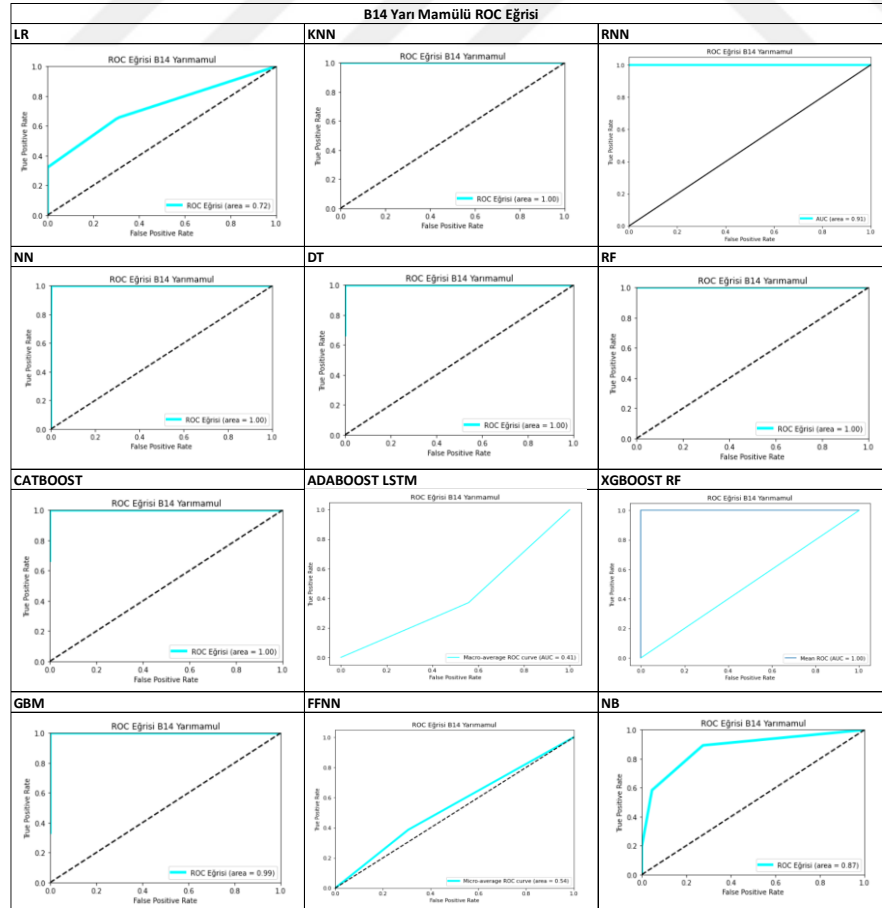
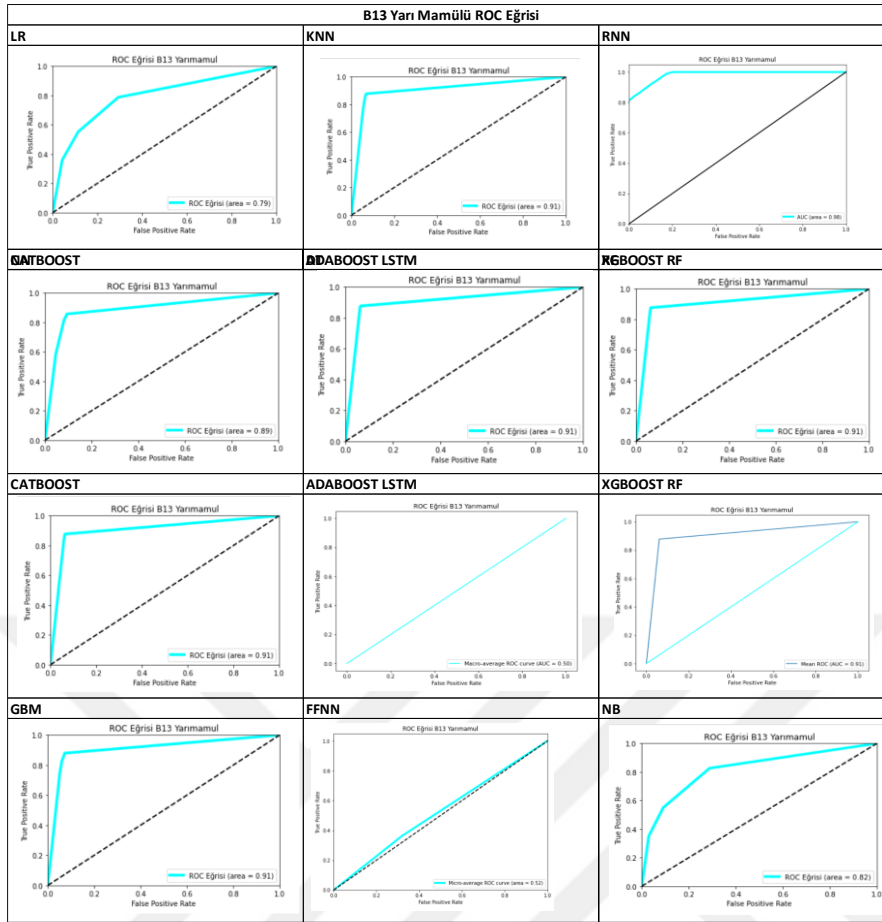


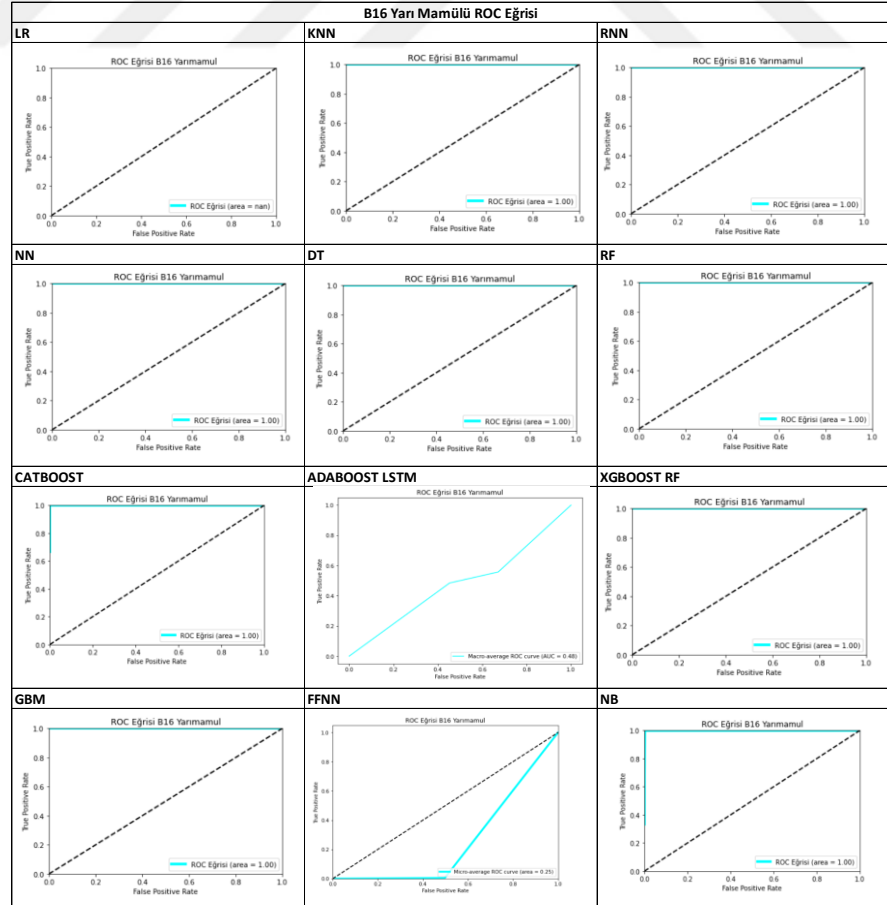
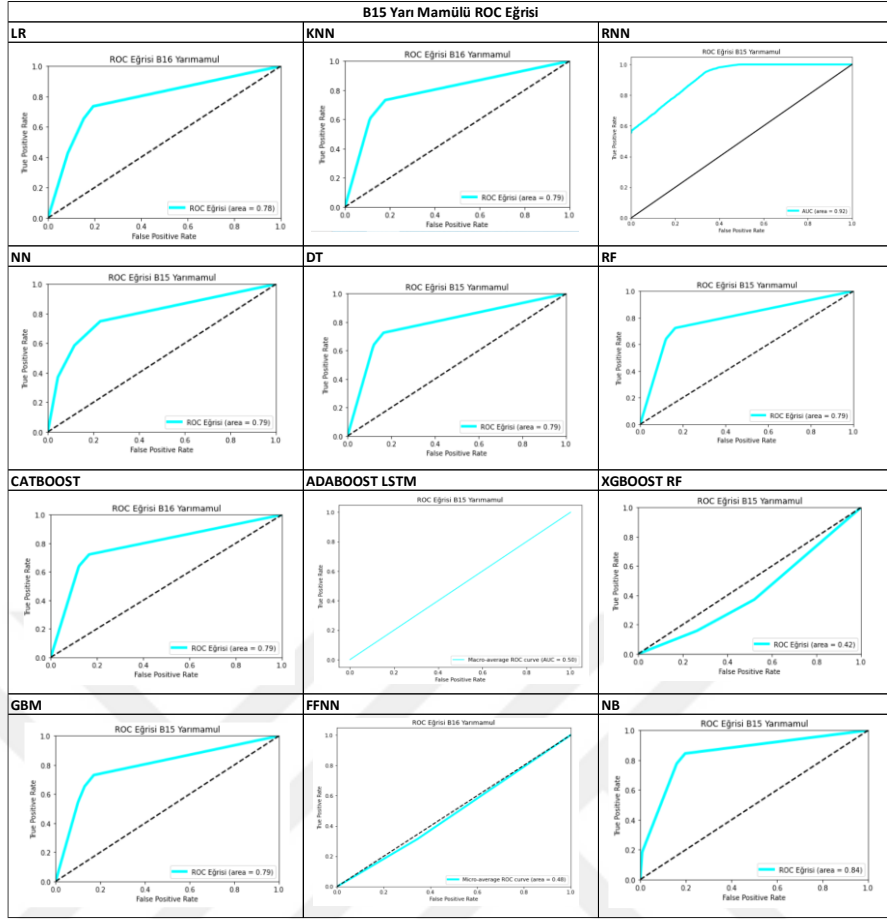


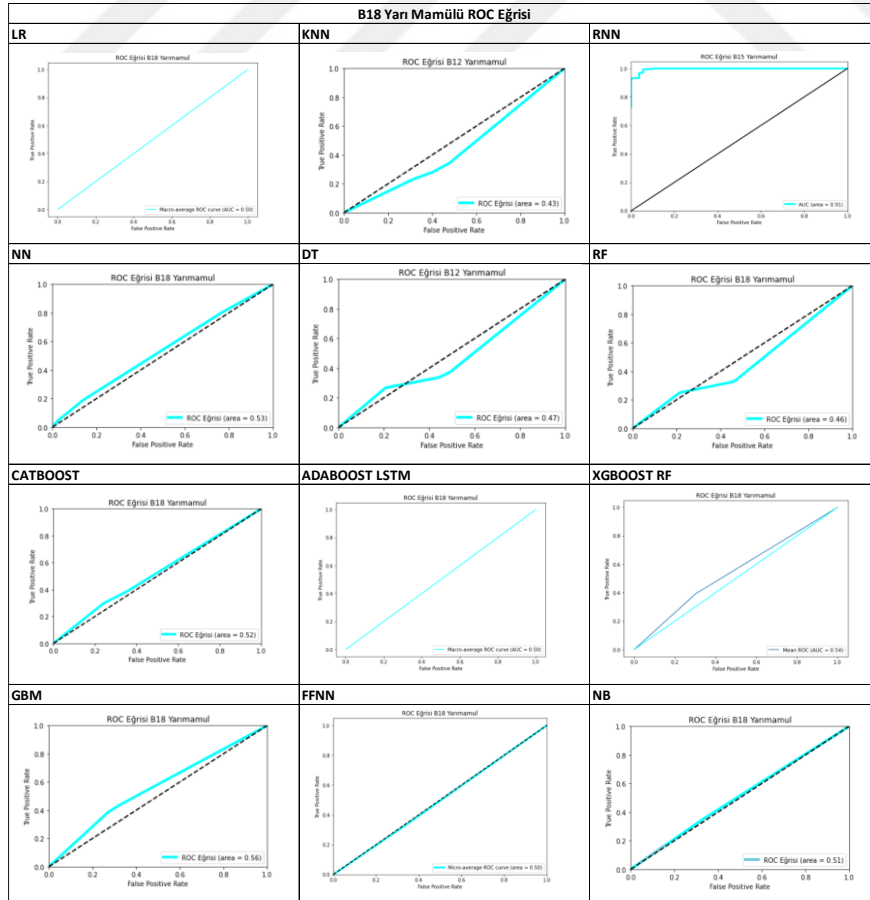
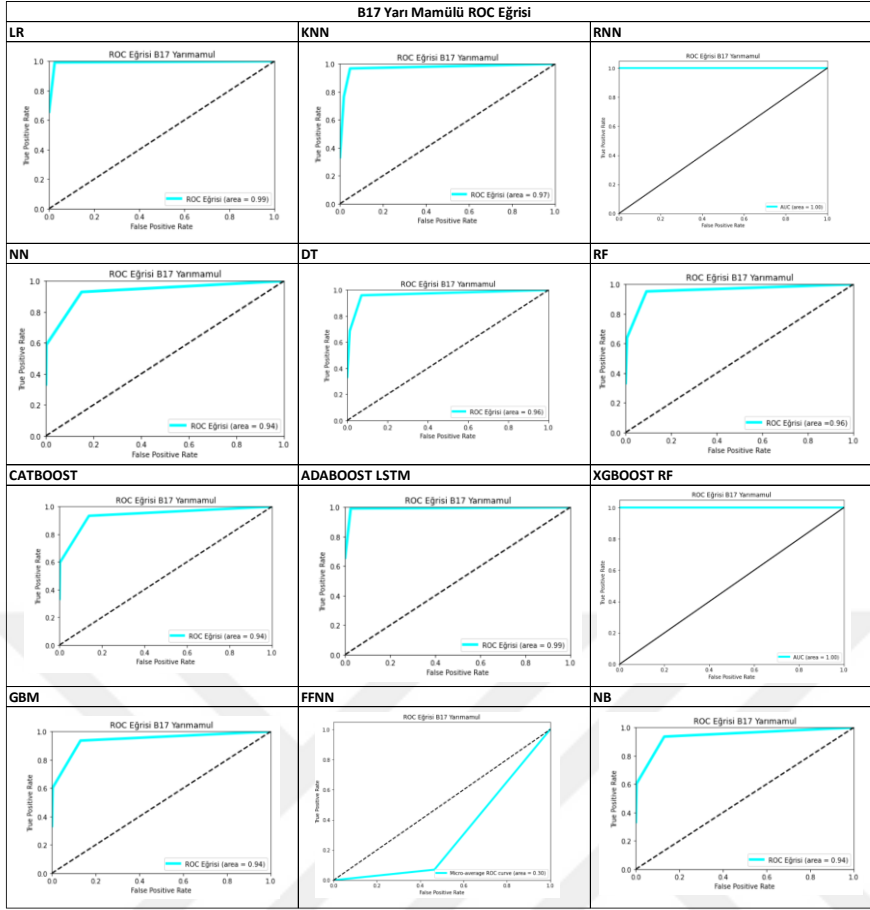


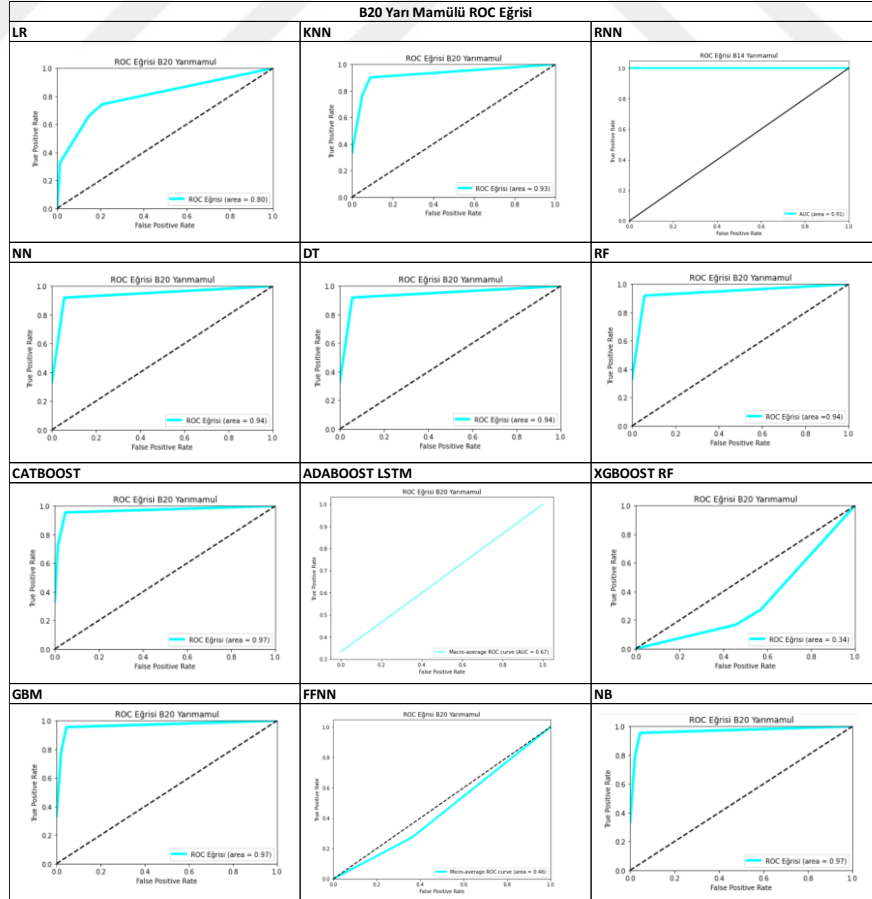
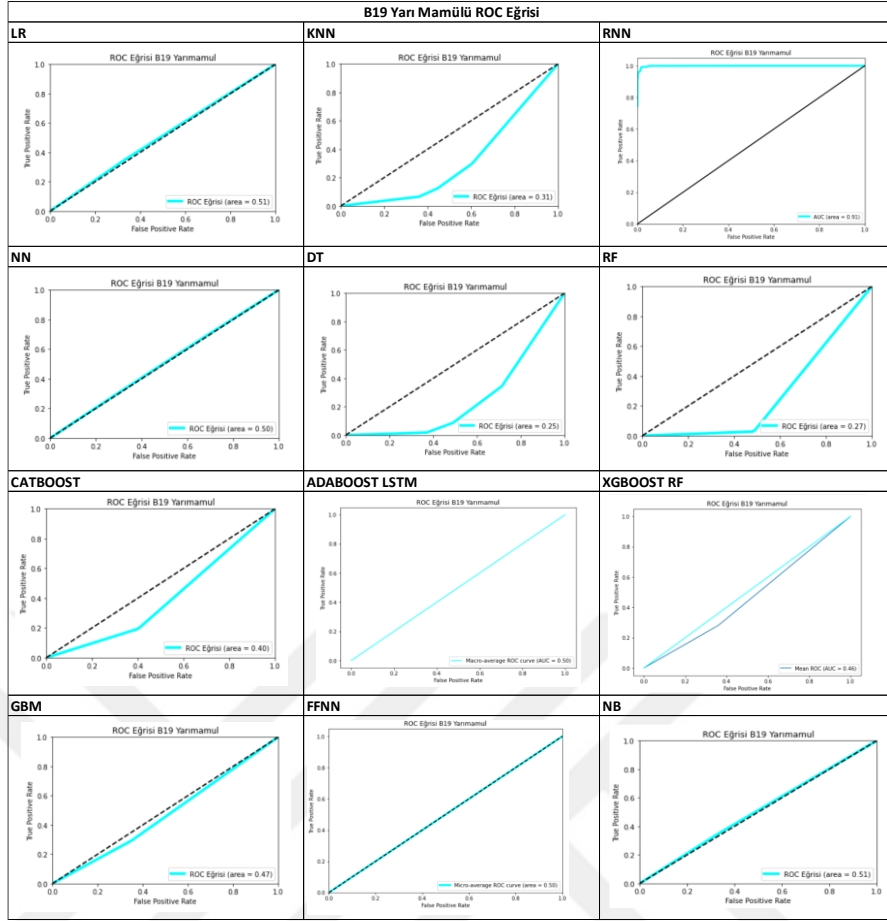


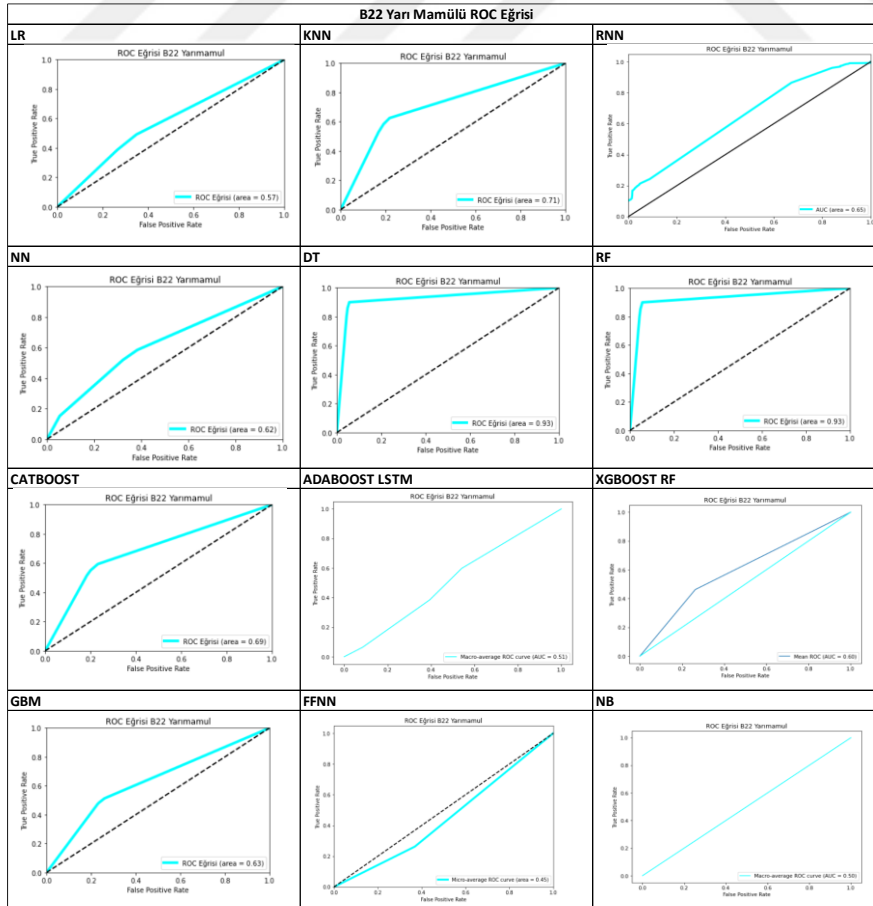
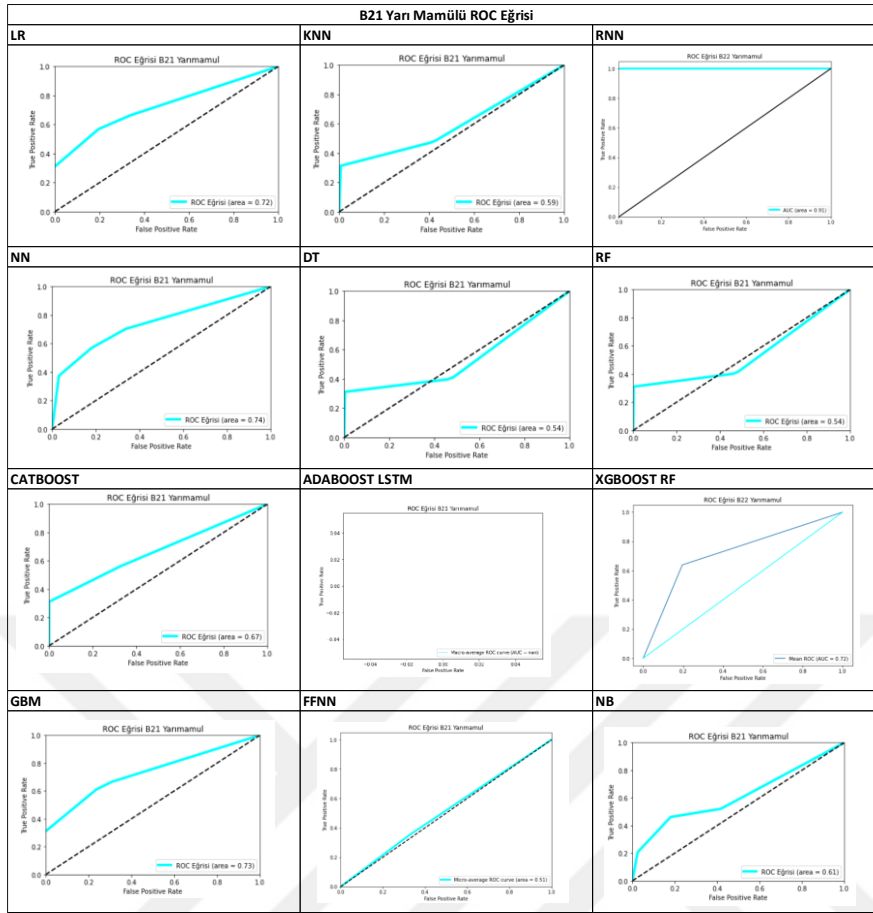


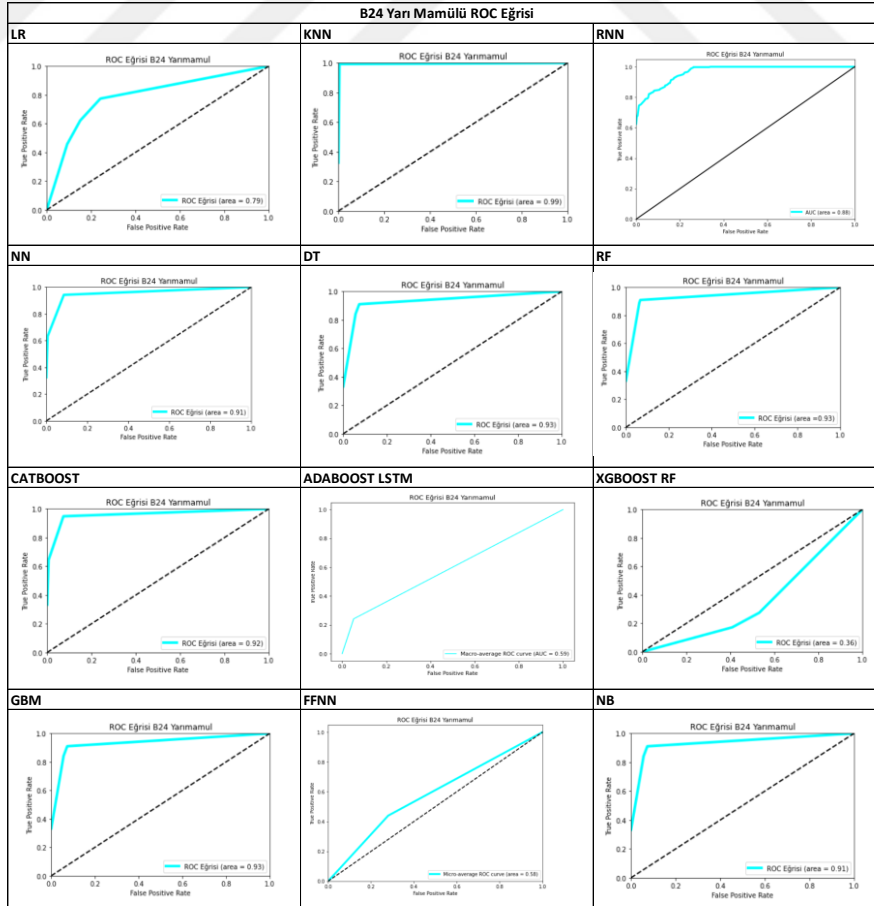
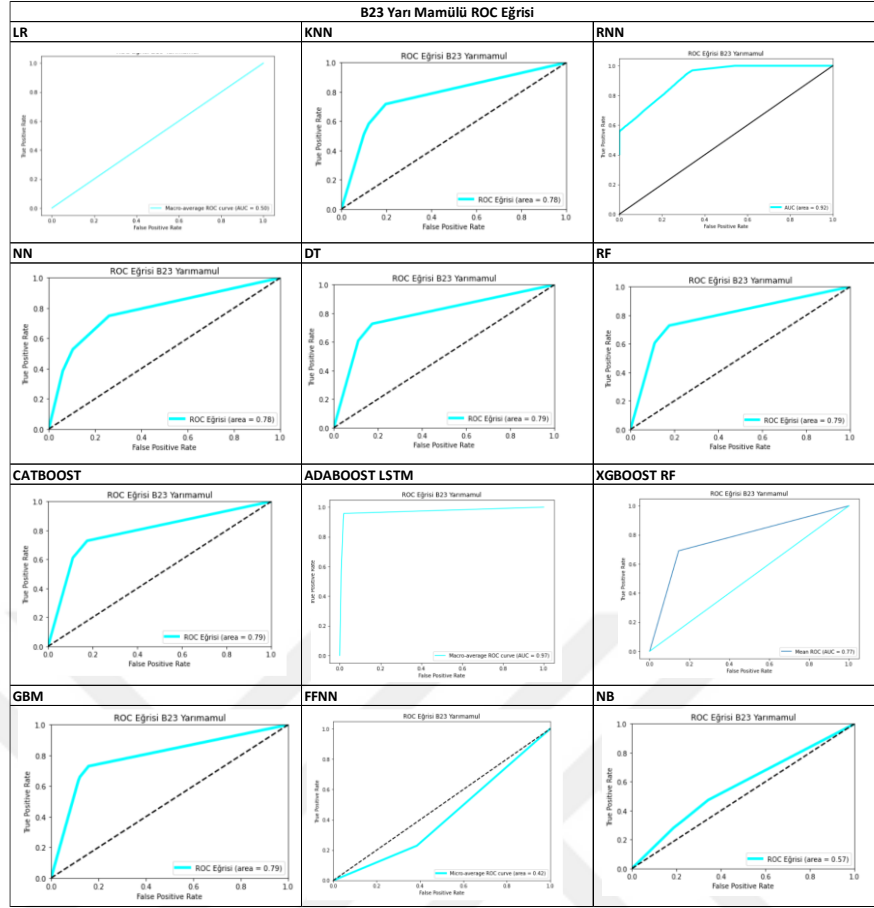














ÖZGEÇMİŞ

Ad-Soyad : Alperen AYTATLI

ÖĞRENİM DURUMU:

- **Lisans** : 2007, İstanbul Teknik Üniversitesi, Mühendislik Fakültesi
- **Yükseklisans** : 2008, İstanbul Bilgi Üniversitesi, İşletme Fakültesi, MBA
- **Yükseklisans** : 2011, İstanbul Teknik Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü

MESLEKİ DENEYİM:

- 2007-2008 yılları arası Finansbank Genel Müdürlüğü'nde Krediler Yetkili Yardımcısı olarak görev aldı.
- 2008-2016 yılları arasında Anlas Anadolu Lastik San. ve Tic. A.Ş.'de sırasıyla Üretim Planlama Uzmanı, Tedarik Zinciri Uzmanı, Kalite Uzmanı, ve Endüstri Mühendisliği Departman Yöneticisi olarak görev aldı.
- 2016-2017 yılları arası Teknorot Otomotiv Ürünleri San.ve Tic.A.Ş.'de Tedarik Zinciri Yöneticisi olarak görev aldı.
- 2017-2020 yılları arası Asaş Alüminyum San.ve Tic.A.Ş.'de Proses Kalite Müdürü olarak görev aldı.
- 2020-2021 yılları arası Orka Banyo San.ve Tic.A.Ş.'de Tedarik Zinciri Müdürü olarak görev aldı.
- 2022 yılı itibari ile Simona Plastech San.ve Tic.A.Ş.'de Fabrika Müdürü olarak görev almaktadır.
- 2024 yılı itibari ile Bahçeşehir Üniversitesi'nde Misafir Öğretim Üyesi olarak dersler vermektedir.

TEZDEN TÜRETİLEN ESERLER:

- Aytatlı, A., 2023. Machine learning-based product quality classification of the enterprise producing aluminum flat coil. Journal of Radiation Research and Applied Sciences 16, 100715. <https://doi.org/10.1016/j.jrras.2023.100715>