

Mevcut Olmayan Bir Ürün İçin Fizibilite Analizi ve Karar Destek Sistemi Tasarımı

Feasibility Analysis and Decision Support System Design for a Non-Existent Product

Efe Baskın

Bahçeşehir Üniversitesi Bilgisayar Mühendisliği, efebaskin2005@gmail.com

ÖZET

Bu çalışmada, henüz var olmayan ürünler için kapsamlı bir fizibilite analizi ve karar destek sistemi (DSS) tasarlanmıştır. Literatürdeki 16 fizibilite raporundan çıkarılan ortak şablonlar oluşturularak LLM tabanlı sentetik veri üretimi gerçekleştirilmiş; elde edilen veri üzerinde keşifsel veri analizi, korelasyon ve dağılım incelemeleri yapılmıştır. Regresyon modelleri (Lineer, Random Forest, XGBoost, Stacking) ile ürünün beklenen yıllık geliri tahmin edilmiş, Stacking regressor ile R^2 0.98 düzeyinde yüksek doğruluk sunmuştur. K-Means kümeleme ile beş yatırım profili tanımlanmış; Monte Carlo simülasyonu (%52 pozitif NPV, %88 5 yıl altı ROI) belirsizlikleri nicelendirirken lojistik regresyon destekli Markov zinciri ile karar aşamaları arasındaki geçiş olasılıkları modellenmiştir. Sonuç olarak önerilen DSS, ürünün orta ve uzun vadeli performansına dair net içgörüler sağlayarak yatırım kararlarını desteklemektedir.

Anahtar Kelimeler: Fizibilite Analizi, Karar Destek Sistemi, Sentetik Veri, Büyük Dil Modelleri, Makine Öğrenmesi, Monte Carlo Simülasyonu, Markov Zinciri

ABSTRACT

In this study, a comprehensive feasibility analysis and decision support system (DSS) are developed for a non-existent product. Shared templates extracted from 16 literature-based feasibility reports guide the generation of synthetic data via large language models (LLMs). The dataset undergoes exploratory data analysis, including correlation and distribution assessments. Regression techniques (Linear, Random Forest, XGBoost, Stacking) forecast the expected annual revenue, with ensemble methods achieving $R^2 \approx 0.98$. Five investment profiles are identified using K-Means clustering; Monte Carlo simulation quantifies uncertainty (%52 positive NPV, %88 ROI within five years), and a logistic regression-driven Markov chain models dynamic transition probabilities between decision stages. The resulting DSS delivers clear, actionable insights into the product's mid- and long-term viability, supporting informed investment decisions.

Keywords: Feasibility Analysis, Decision Support System, Synthetic Data, Large Language Models, Machine Learning, Monte Carlo Simulation, Markov Chain

1 GİRİŞ

Yeni bir ürün ortaya konulacağında ürün hakkında çeşitli analizler yapılmaktadır. Finansal olarak yeterli olup olmadığını tespit etmek için yapılan analiz türüne finansal fizibilite analizi denir. Potansiyel gelirin beklenen maliyetten yüksek olduğu durumlar ilgili projenin finansal olarak uygulanabileceğine işaret eder [1]. Finansal fizibilite analizi gibi fizibilite analiz çeşitleri mevcuttur. Teknik analiz, market analizi ve etki değerlendirmesi fizibilite analizinin kapsadığı analizlerdendir [3]. Fizibilite çalışmalarının yapılma nedenleri planlama ile kontrol kolaylığı sağlaması, denetimi basitleştirmesi ve kayıp riskinin azaltılmasıdır [2].

Bu çalışmanın bir diğer konusu olan karar destek sistemleri, işletmelerin karar süreçlerine yardımcı olmayı amaçlar. Yapay zeka alanında yaşanan gelişmelerle birlikte karar destek sistemleri, makine öğrenmesi teknolojilerinden faydalanarak kullanıcıya göre kendini uyarlayabilme ve değişen koşullara adapte edebilme özelliklerine sahip olur [4].

Bu çalışmamızda piyasada olmayan bir ürüne çeşitli fizibilite analizleri uygulanacaktır. Ürüne benzer ürünler ve müşteri davranışları incelenerek müşteri değer tahmini yapılacaktır. Ardından farklı makine öğrenmesi teknikleri kullanılarak bir karar destek sistemi geliştirilecek ve sonuçlar değerlendirilecektir.

2 LİTERATÜR ARAŞTIRMASI

Günümüzde küçük ve orta büyüklükteki işletmeler (KOBİ), gelişen teknolojiyle beraber müşteri kitlesini ve işleyiş süreçlerini çeşitli yöntemlerle daha net analiz etmeye başlamıştır. Böylece KOBİ'ler teknolojiyi iş süreçlerine uygulayarak müşteri davranışlarını daha iyi anlamasına ve pazarlama stratejilerini özelleştirmelerine yardımcı olmaktadır [16].

Katerina Doroshkevych ve ekibi bir işletmenin almayı düşündüğü yenilikçi kararın planlanma süreçlerini araştırmıştır. Çalışma için veri, Lviv Politeknik Ulusal Üniversitesi'nden uzmanlar ve Stratejik Araştırma ve Analiz Merkezi temsilcilerinden oluşan toplam 12 uzmanın katılımıyla anket yöntemi uygulanarak toplanmıştır. Şirketin kaynak yeterliliğinin ve market talebinin fikre uygunluğuna bakılmıştır. Stratejik planlamalar etkileşimli durum tanımlayıcıları olarak sunulmuş ve Markov zincir grafiği oluşturulmuştur. Bu sayede Markov geçiş olasılık matrisi kullanılarak planlama sürecinin belirli bir zaman aralığındaki her olası durumda bulunma olasılığı tahmin edilmektedir. Önerilen modelin avantajı stratejik planlama süreçlerinin modellenmesiyle ilgili hesaplamaları basitleştirmesi olarak gözlemlenmiştir. Model sonuçlarının analizinde elde edilen geçiş olasılıklarının simülasyon sonuçlarına göre varyansı hesaplanmış ve standart sapma kullanılarak hesaplanan göreceli hata oranları üzerinden değerlendirilmiştir [5].

Markov zincirlerinin kullanıldığı başka bir çalışma da C.Cheng tarafından yapılmıştır. Yaptıkları çalışmada bir oto tamir ve bakım firması için müşterilerin yaşam boyu değer hesaplaması yapan bir sistem kurulmaya çalışılmıştır. Veriler, bir oto tamir ve bakım şirketine ait müşteri veritabanlarından elde edilmiştir. Kullanılan veri kaynakları; müşteri demografik bilgileri (müşteri veritabanı), satın alma kayıtları ve tamir/bakım kayıtlarını içeren işlem verileridir. Müşteri yaşam boyu değeri müşterinin mevcut ve gelecekteki değerlerini içerecek şekilde tanımlanmıştır. Müşteri davranışında ise bir Markov zincir modeli kullanılmıştır. Model müşterinin belirli bir satın alma davranışından başka bir davranışa geçiş olasılığını tahmin etmektedir. Kurulan Markov zincir modelinde, müşteri durumu yalnızca bir değişken -yıllık ziyaret sayısı- ile temsil edilmiştir. Tek değişken durumu ilgili şirketin işlem veritabanının eksik olmasından kaynaklanmaktadır. Genellikle bir müşterinin satın alma davranışını iyi temsil etmek için birden fazla değişken gerekmektedir. Ancak çalışmada birden fazla değişken kullanımının fazla olası durum yaratmasından dolayı Markov modelinin yönetimini zorlaştırabileceği belirtilmiştir [6].

Bu çalışmaya benzer şekilde müşteri satın alım davranışlarını inceleyen başka bir araştırmayı da Hodjat Hamidi ve ekip arkadaşları yapmıştır. Çalışma bir giyim şirketi üzerinde gerçekleştirilmiştir ve ana amacı müşteri yaşam boyu değer tahmin doğruluğunu artırmaktır. Çalışma verileri 2018'den 2023'e kadar, tanınmış bir giyim şirketinin satış bilgilerini içeren ham

veri setinden elde edilmiştir. Toplanan veriler; müşteri ID, satın alma yılı, satın alma sıklığı, yıllık işlem tutarı, evlilik tarihi, doğum tarihi, il, ilk ve son satın alma tarihi, cinsiyet, meslek ve müşterinin en çok ilgi gösterdiği ürün kategorisi gibi demografik ve işlem bilgilerini kapsamaktadır. CLV hesaplaması için, en az iki satın alma gerçekleştiren müşteriler seçilmiştir ve ek olarak veri üzerinde özellik mühendisliği uygulanarak eksik değerler tamamlanmış, Label Encoding ile kategorik kodlama yapılmış, Standart Scaler ile ölçeklendirme yapılmıştır. Temel bileşen analizi (PCA) da uygulanarak veri boyutu azaltılmıştır ve aşırı uyum önlenmiştir. 4 farklı model ile stacking ensemble learning yöntemi kullanılarak tahmin doğruluğunun yükseldiği gözlemlenmiştir. Modeller ve kullanılan özellikler sırasıyla BG-NBD modelinden elde edilen tahmin sayısı, Gamma modelinden hesaplanan ortalama değer, K-means modeliyle belirlenen müşteri kümeleme etiketleri ve Markov zincir modelinden üretilen müşteri durum indeksidir. Model performans değerlendirmesinde normalize edilmiş Kök Ortalama Kare Hatası (nRMSE), normalize edilmiş Ortalama Mutlak Hata (nMAE) ve Determinasyon Katsayısı (R^2) kullanılmıştır. nRMSE ve nMAE değerlerinin düşük olması modelin tahmin doğruluğunun yüksek olduğunu ayrıca R^2 değerinin yüksek olması ise modelin veri varyansının büyük bir kısmını açıkladığını göstermektedir. Bu çalışmada da kullanılacak olan markov zinciri ve kmeans modeli sayesinde elde edilen özelliklerin çalışmaya bağlamsal bilgi kattığı belirtilmiştir [7].

Makine öğrenmesi algoritmaları yalnızca müşteriye bağlı analizlerde değil ayrıca şirketin finansal durumunu analiz etmekte de kullanılır. Guido Perboli ve arkadaşları KOBİ'leri hedef alarak 60 aya kadar olan orta ve uzun vadeli iflas tahminine odaklanmıştır. Kullanılan finansal veri 2012'den 2018'e kadar ki italyan firmalarına aittir. Çalışma iki aşamaya ayrılmıştır. İlk aşamada her şirket için önündeki 5 yıl içinde iflas etme olasılığı hesaplanmıştır ve aktif şirketler arasında gelecek 5 yıl içerisinde iflasa girebilecek olanlar da bulunduğu gözlemlenmiştir. İkinci aşamada da belirli eşik değerleri kullanılarak 5 yıl boyunca iflas olasılığı eşik değerin altında kalan şirketler seçilmiş ve aktif şirketler daha net ayrılmıştır. En optimal eşik değerinin 60 olduğu görülmüştür. Bu iki aşamalı yöntem veri setindeki dengesizliği azaltmak ve aktif ile iflas eden şirketler arasındaki farkları belirginleştirmek amacıyla uygulanmıştır. Örneğin ilk aşamada random forest Area Under Curve (AUC) değeri için yüzde 78 sonuç verirken ikinci aşamada yüzde 88'e çıkmıştır. Test edilen algoritmalar Random Forest, XGBoost, Lojistik regresyon ve nöral ağlar olmaktadır. Aralarından en kötü sonucu veren log loss metriğine göre 0.41 ile lojistik regresyon, en iyi sonucu veren de 0.25 ile XGBoost olmuştur. [8].

Benzer bir çalışma da Anti Merisalo tarafından yapılmıştır. Çalışmanın amacı, Business to Business (B2B) hizmet sektöründe faaliyet gösteren Finlandiya'dan bir şirketin satış sürecini çok aşamalı olarak modelleyip Monte Carlo simülasyonu kullanarak kısa vadeli satış tahminleri yapmak ve bu modeli karar destek sistemi olarak kullanmaktır. Mevcut satış tahmin yöntemleri incelenerek klasik deterministik modeller yerine belirsizlik ve rastgelelik içeren stokastik modeller tercih edilmiştir. Veri olarak Customer Relationship Management (CRM) verisi, Enterprise Resource Planning (ERP) verisi ve personel verileri kullanılmıştır. Personel verileri kullanılarak çalışanların aylık rolleri belirlenmiş ve her rol için aylık ortalama arama sayısı, standart sapma, aramadan toplantıya dönüş oranı, toplantıdan vaka açmaya dönüşüm oranı gibi özellikleri içeren veri tablosu oluşturulmuştur. Kullanılan Monte Carlo simülasyonu sayesinde tekrarlı rastgele örnekleme yöntemi kullanılarak satış sürecindeki olası sonuçların olasılık dağılımlarına ulaşılmıştır. Excel tabanlı bir arayüz ile modelin hem aylık hem de 3 aylık senaryo tahminleri görselleştirilmiştir. Simülasyonun performansı 3 aylık hareketli ortalama ve iki farklı regresyon modeliyle karşılaştırılmıştır ve simülasyon modelinin tahmin doğruluğunun Mean Squared Error (MSE), Mean Absolute Deviation (MAD) ve Mean Absolute Percentage Error (MAPE) gibi metriklerle bakılarak daha yüksek sonuç verdiği gözlemlenmiştir [9].

Karar destek sistemleri yalnızca müşteri bazlı ve finansal analizler için değil, ayrıca işletmelerin yönetim sistemlerinde kullanılmaktadır. Yönetimin daha net kararlar vermesinde de makine öğrenmesi algoritmalarından yararlanılmaktadır. Jian Cao ve arkadaşı Yuxi Zhang'ın yaptığı araştırmada farklı işletmelerden aldığı hem gerçek zamanlı hem de geçmiş verilere

ön işleme ve özellik seçimi uygulayarak yönetim kararlarının veri odaklı iyileştirilmesi hedeflenmiştir. Özellik seçiminde korelasyon analizi ve PCA gibi yöntemler kullanılarak yönetimdeki kararları etkileyen en önemli değişkenler belirlenmiştir. Makine öğrenimi modelleri olarak karar ağaçları, regresyon analizi ve nöral ağları içeren farklı yöntemler kullanılmıştır. Sistem performansı Mesafe Doğruluk Oranı ile gerçek değerler ve tahminler arasındaki mesafe hassasiyeti değerlendirilmiştir. Eğitim ve doğrulama sürecinde aşırı uyumu önlemek için çapraz doğrulama uygulanmıştır. Ardından sistem parametrelerini iyileştirmek amacıyla Bayes optimizasyonu ve grid search gibi yöntemler kullanılmıştır. Yönetim sistemlerinin performanslarını gözlemek amacıyla da konvolüsyonel sinir ağı ile takip yapılmıştır. Kurulan karar destek sisteminin performansını iyileştirmede verimlilik optimizasyonu ve artan iterasyon sayılarının etkileri gözlemlenmiştir [10].

Son olarak işletmeler için riskli projelerin değerlendirilmesiyle ilgili çalışma Qian Wang tarafından yapılmıştır. Proje için veriler piyasa verilerinden ve operasyonel maliyetler ile vergi oranları gibi finansal parametrelerden oluşmaktadır. Veriden hesaplanan pazar riski ve uzman tahminlerine dayanan özel riskin birleşimiyle oluşturulan hibrit değişkenler üzerinden sayısal değerlendirme yapmıştır. Projede gerçek opsiyon modellemesinin, geleneksel yöntemlere göre belirsizliği ve yöneticilerin esneklik gösterdiği durumları hesaba katması vurgulanmaktadır. Model olarak genişletilmiş LSM algoritması ve Monte Carlo simülasyonu kullanılmıştır. Monte Carlo ile örnek yollar üretilmiş ve her yol için opsiyon değeri hesaplanarak tüm yolların beklenen değerinin alınması sağlanmıştır. LSM algoritmasında ise belirsizlik ile rastgelelik iki seviye olarak ayrılarak bir hibrit süreç tanımlanmıştır. Bu süreç üzerinde de sayısal değerlendirmeler yapılmıştır. Genişletilmiş LSM algoritması, fuzzy gerçek opsiyonlarının sayısal olarak değerlendirilmesinde etkin bir yöntem olarak uygulanmıştır [11].

Bu çalışmanın bir diğer konusu da fizibilite analizidir. Literatür taraması devamında fizibilite analizinin KOBİ'lerde uygulanış tarzına bakılacaktır. O. Nainggolan, Gobor'da KOBİ'lerin ayakkağı ve terlik üretiminin fizibilitesini; yatırım büyüklüğü, işletme faaliyetleri ve elde edilecek faydalar açısından incelemiştir. Fizibilite analizi; teknik fizibilite, ekonomik ve sosyal fizibilite, finansal fizibilite ve SWOT analizi olmak üzere farklı tekniklerle gerçekleştirilmiştir. Teknik fizibilite analizinde hammadde kalitesi, altyapı, elektrik, fiyat ve insan kaynakları gibi unsurlar incelenerek ortalama teknik değerler elde edilmiştir. Örneğin hammadde ile insan kaynaklarından elde edilen değer 3.6 iken elektrik ve diğerlerinde 4 elde edilmiştir. Ekonomi ve sosyal fizibilitede; hükümetin zenginleşmesine, istihdam olanaklarına, sosyal yapı unsurlarına değer olarak 4 belirlenmiştir. Finansal fizibilite analizinde; Geri Ödeme Süresi (Payback Period), NPV (Net Bugünkü Değer), IRR (İç Verim Oranı) ve Ortalama getiri gibi ölçütler 5 değerini almaktadır. Öte yandan Profitability Index'in fizibiliteyi gösterecek yeterli kanıt sunmadığı gözlemlenmiştir. Tüm bu yapılan analizler neticesinde KOBİ'lerin faaliyet göstermeye elverişli olduğu fakat finansal fizibilite de yalnızca Profitability Index açısından eksiklik olduğu görülmüştür. Ayrıca yapılan SWOT analizi sonucunda üretimde ve pazarlamada çeşitlilik sağlanması, pazarlama kanallarının optimize edilmesi, ilişki pazarlaması ağlarının genişletilmesi gibi konularda stratejiler belirlenmiştir [12].

Daha yazılım sektörüne yönelik yapılan bir çalışmada Leandro ve arkadaşları, KOBİ'lerin yazılım sistemlerinin dağıtımında ortaya çıkabilecek sorunları tespit edip düzeltme veya revizyon gerektiren unsurları belirleyebilecek bir fizibilite analiz yöntemi tanımlamaktadırlar. Yöntem organizasyon, süreç, insan kaynakları, ürün ve uygulamalar-araçlar-teknikler olmak üzere beş açıdan analiz edilmiştir. Fizibilite analiz yöntemi çok boyutlu değişkenlerin kullanıcı yorumu ile tanımlanan sözsöz değerler üzerinden değerlendirilmesine dayanmakta olup her soruya belirli ağırlıklar ve eşik değerler atanmakta, sonuçlar fuzzy (bulanık) aralıklar şeklinde ifade edilmektedir. Sonuçlar tüm boyutların kısmi puanlarının birleştirilmesiyle genel bir eşik değer ile karşılaştırılarak; Red: Uygulanamaz, Yellow: Potansiyel olarak uygulanabilir değil ve Green: Potansiyel olarak uygulanabilir şeklinde görsel uyarılarla sunulmaktadır. Yöntem, KOBİ'lerin yazılım dağıtım süreçlerinde karşılaşılabilecekleri sorunları kapsamlı bir biçimde değerlendirebilmek için geliştirilmiş olup ön testler iki

Arjantinli yazılım geliştirme KOBİ'sinde uygulanarak, Perceived Ease of Use (Algılanan Kullanım Kolaylığı) açısından olumlu, People boyutu açısından ise daha düşük sonuçlar vermiştir [13].

Daha küçük ölçekli bir işletmeye yapılan çalışmaya bakılırsa Amir Halid ve ekibi küçük ölçekli sığır yetiştiriciliği işletmesinin finansal fizibilitesinin incelemişlerdir. Finansal analizde NPV, IRR, net faydanın maliyete oranı (Net B/C Oranı), Geri Ödeme Süresi gibi göstergeler kullanılmıştır. Göstergeler incelenirse NPV tüm ölçeklerde pozitif olup işletmenin finansal karlı olduğunu göstermiştir. IRR, belirlenen %12 bankalar arası oranı aşmıştır. Hesaplanan net B/C oranı tüm ölçeklerde 1'in üzerinde bulunmuş ve geri ödeme süresinin de en fazla 8 yıl 11 ay 26 gün olduğu gözlemlenmiştir [14].

3 UYGULANAN YÖNTEM VE DEĞERLENDİRME

Bu çalışmada mevcut olmayan bir ürün için fizibilite analizi yapılması ve karar destek sisteminin oluşturulması amaçlanmıştır. Bu doğrultuda çalışmanın veri oluşturma ve model değerlendirme süreci 5 aşamadan oluşmaktadır. Bu aşamalar veri kümesi oluşturulması, veri anlayışı, few-shot öğrenme uygulaması ve büyük dil modellerinin (LLM) test edilip model performanslarının değerlendirilmesidir.

3.1 Veri Kümesi Oluşturulması

İlk aşama olarak farklı sektörlerden seçilen toplam 16 adet fizibilite raporu analiz edilerek bir veri kümesi oluşturulmuştur. Raporlarda tarım, sanayi, enerji, gıda ve hizmet sektörleri gibi çeşitli sektörler bulunmaktadır. Bu raporlar üzerinden fizibilite çalışmalarının yapı ve içerik anlamında özellikleri çıkarılarak raporların veri türleri ve başlıkları belirlenmiştir.

3.2 Veri Anlayışı

Fizibilite raporlarının incelenmesi sonucunda aşağıdaki başlıkların ortak olarak her raporda yer aldığı gözlemlenmiştir.

1. Yatırım Künyesi: Projenin temel tanımı, yatırım yeri, yatırım tutarı, üretim/hizmet kapasitesi gibi bilgiler
2. Ekonomik Analiz: Pazar durumu, talep tahminleri, rekabet analizi, ekonomik katkı ve potansiyel getiriler
3. Teknik Analiz: Üretim/hizmet süreçleri, kullanılacak teknoloji ve ekipmanlar, tesis yapısı, üretim kapasitesi
4. Finansal Analiz: Sabit ve değişken maliyetler, gelir projeksiyonları, karlılık oranları ve yatırımın geri dönüş süresi
5. Çevresel ve Sosyal Etki Analizi: Çevre dostu uygulamalar, istihdam yaratılması, sosyal kalkınmaya katkılar

Bu başlıklar altında her yatırım için yatırım tutarı, üretim kapasitesi, yıllık satış gelirleri gibi sayısal veriler ve çevresel etki, toplumsal faydalar gibi niteliksel bilgiler yer almaktadır. Veri anlayışı aşamasında, fizibilite çalışmalarının hem ekonomik hem de sosyal özellikler içeren çok boyutlu bir yapıda olduğu görülmüştür. Ayrıca bu başlıklar modellerin eğitileceği genel şablonun oluşturulmasında kullanılmıştır.

3.3 Few-Shot Öğrenme Uygulaması

Few-shot learning (FSL), modelin az sayıda örnekten öğrenerek yeni görevleri yerine getirme yeteneği olarak tanımlanmıştır. FSL özellikle veri toplamanın maliyetli, zor veya etik açıdan kısıtlı olduğu durumlarda önem kazanmış bir öğrenme paradigmasıdır. Geleneksel makine öğrenimi yöntemlerinin büyük miktarda etiketli veriye ihtiyaç duymasının aksine few-shot learning az sayıda örnekten genelleme yapabilmeyi hedefler [15].

Bu çalışmada belirli bir sektör veya ürünle doğrudan ilişkili büyük ölçekli veri setleri olmadığından dolayı few-shot prompting yaklaşımı tercih edilmiştir. Çalışmada few-shot prompting kullanılması çeşitli nedenlere dayanmaktadır. Yeni geliştirilen ürünlere yönelik yeterli miktarda erişilebilir geçmiş fizibilite raporu verisi bulunmaması, veri kısıtlılığına neden

olmaktadır. Bu durumdan dolayı few-shot prompting ile modellerin az sayıda örnekle eğitilerek görev hakkında eğitilmesi sağlanmıştır. Çalışmada ChatGPT, DeepSeek V3, Gemini ve Claude 3.7 Sonnet gibi farklı LLM performanslarını karşılaştırmak için standart referans noktası oluşturulmuştur ve her modele aynı örnekler sunulmuştur. Ayrıca modellerin özgün ve fizibilite rapor standartlarına uygun, olmayan bir ürün için yeni fikirler üretebilmeleri için yaratıcılarının belirli bir çerçevede yönlendirilmesi amaçlanmıştır. Böylece seçilen fizibilite raporlarından oluşturulan genel şablon kullanılarak modeller yönlendirilmiş ve örneklerden genelleme yapılarak fizibilite raporu oluşturma yetenekleri test edilmiştir.

3.4 Uygulama

Uygulama sürecinde ChatGPT 4o, DeepSeek V3, Gemini ve Claude 3.7 Sonnet modelleri değerlendirilmiştir. Her modele few-shot örnekler gösterilerek belirli yönergeler verilmiştir. Modeller 2 aşamayla test edilmiştir. İlk aşama, few-shot örnekler analiz edilerek fizibilite raporlarında bulunması gereken ortak özellik ve içerikleri belirlemek ve ikinci aşama, özgün ürün ya da hizmet fikri bularak belirlenen yapıya uygun bir fizibilite raporu hazırlamaktır. Tüm modellere aynı örnekler ve aynı test soruları sorulmuştur böylece karşılaştırma için uygun ortam sağlanmıştır. Modellerin ürettiği çıktılar toplanmış ve en iyi performans gösteren LLM çıktıları veri kümesi olarak kullanılmıştır.

3.5 Değerlendirme

Karşılaştırma sonucunda tüm modeller ilk aşamadaki göreve doğru ve net cevaplar vermiştir. Ancak ikinci aşamada belirgin farklılıklar oluşmuştur. Genel olarak en başarılı model ChatGPT 4o olmuştur. ChatGPT, diğer modellere kıyasla fizibilite raporu formatını tam yakalamıştır. Yatırım künyesi, teknik analiz, ekonomik analiz, finansal analiz ve çevresel-sosyal etki bölümlerinde eksiksiz ve anlaşılır bir rapor yapısı sunmaktadır. Ayrıca finansal analiz bölümünde sunduğu sabit ve değişken maliyetler, yıllık gelir projeksiyonu ve yatırımın geri dönüş süresi gibi değerler net bir şekilde belirtilmiştir. ChatGPT'den sonra en başarılı bir diğer model de Claude 3.7 Sonnet olmuştur. Claude özellikle finansal analiz bölümünde oldukça detaylı ve sayısal verilerle çıktılar sunmuştur fakat tüm raporu uzun ve yoğun metinler halinde vermesi anlaşılabilirlikten uzaklaşmasına sebep olmuştur. Bu nedenle ChatGPT'nin gerisinde kalmıştır.

DeepSeek ve Gemini ise diğerlerine göre oldukça yüzeysel ve detaylandırma konusunda eksik kalmıştır. Sonuç olarak ChatGPT'nin az sayıda örnek üzerinden eğitime ile sade çıktı formatı, yapılandırılmış ve uygulanabilir fizibilite raporları üretebilme kapasitesinin diğer modellere göre daha yüksek olduğu görülmüştür.

3.6 Verilerin Yapılandırılması

Few-shot prompting yöntemi ile GPT modeliyle oluşturulan fizibilite raporları karar destek sistemi uygulamaları için yapılandırılmış tablosal verilere dönüştürülmüştür. Bu aşamada öncelikle raporların yapısal bölümleri incelenmiş sonrasında sayısal ve kategorik değerler içeren öz nitelikler ayrıştırılmıştır. Sayısal öz nitelikler arasında yatırım tutarı, yıllık üretim kapasitesi, satış miktarı, ROI süresi, beklenen yıllık gelir gibi birçok metrik yer alırken; kategorik öz nitelikler arasında çevresel etki düzeyi, sosyal katkı düzeyi, pazar potansiyeli ve teknolojik yenilik düzeyi gibi ifadeler bulunmuştur. Bu yapılandırma aşaması, literatürde GPT benzeri büyük dil modelleriyle oluşturulan sentetik verilerin yapılandırılması yaklaşımıyla benzerlik göstermektedir. Örneğin Hurst ve arkadaşları, GPT-4o modelinden elde ettikleri sentetik tıbbi verileri istatistiksel analize uygun olacak şekilde yapılandırmıştır [17]. Benzer biçimde Sufi, GPT destekli veri üretimi sonrasında bu verileri özellik çıkarımı yoluyla tablolaştırarak makine öğrenmesi modellerinin eğitimi için kullanmıştır [18]. Bu çalışmada uygulanan süreçte fizibilite raporları manuel etiketleme gibi yaklaşımlarla belirli sütunlara ayrılmıştır. Her ürün için bir veri satırı oluşturulmuştur. Böylece GPT tabanlı modelden elde edilen metinsel içerik çok adımlı

analizlerde kullanılabilecek biçimde kullanılmaya hazır hale getirilmiştir. Hazırlanan verilerin örnekleri Tablo 1’de gösterilmektedir.

Tablo 1: Veri Seti

Ürün Adı	Yatırım tutarı (M TL)	ROI Süresi (yıl)	Yıllık Üretim	Beklenen Yıllık Gelir (M tl)	Çevresel Risk	Sosyal katkı	Pazar Potansiyeli	Yenilik Seviyesi	Sektör
MedPulse Patch V11	6.4	3	4800	12.1	Düşük	Yüksek	Orta	Yüksek	Sağlık
CityFlux Scooter V11	8.2	2	6500	18.9	Orta	Yüksek	Yüksek	Yüksek	Ulaşım
NanoForge Printer V1	16.6	3	3300	22.5	Orta	Düşük	Orta	Yüksek	Üretim
MedPulse Patch V12	14.3	5	1200	7.8	Orta	Orta	Düşük	Orta	Sağlık

3.7 Keşifsel Veri Analizi

Bu bölümde oluşturulan sentetik veri setinden elde edilen temel görsellerle çıkarımları değerlendirilecektir.

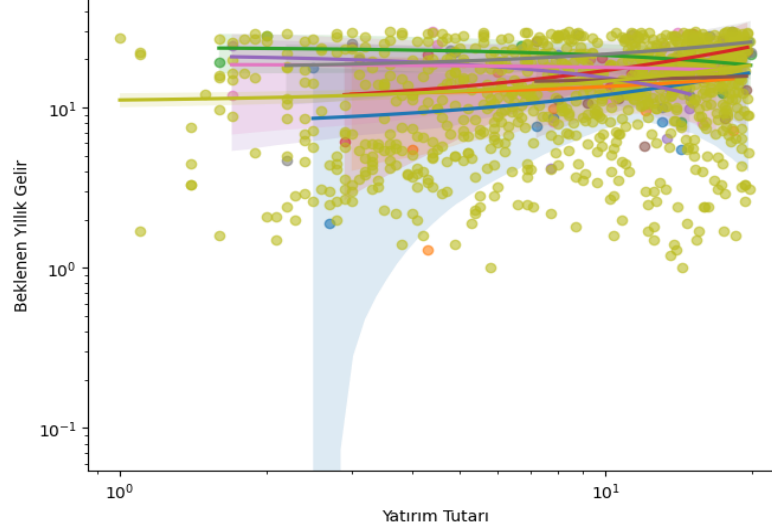
Veri kümesine ilk olarak üç temel düzenleme uygulanmıştır. Ürün adlarının ilk iki sözcüğü “*ürün ailesi*” değişkeni, sonda yer alan “*V*” etiketi ise “*versiyon*” değişkeni olarak ayrıştırılarak hem aile düzeyinde hem de sürüm düzeyinde inceleme yapılabilmiştir. Bu işlemler ile veri seti ön işleme ve veri analizi için hazır hale getirilmiştir.

Tablo 2: Yeni Değişkenli Veri Seti

Ürün Adı	Yatırım tutarı (M TL)	ROI Süresi (yıl)	Yıllık Üretim	Beklenen Yıllık Gelir (M TL)	Çevresel Risk	Sosyal katkı	Pazar Potansiyeli	Yenilik Seviyesi	Sektör	Ürün Ailesi	Versiyon
MedPulse Patch V11	6.4	3	4800	12.1	Düşük	Yüksek	Orta	Yüksek	Sağlık	MedPulse Patch	V11
CityFlux Scooter V11	8.2	2	6500	18.9	Orta	Yüksek	Yüksek	Yüksek	Ulaşım	CityFlux Scooter	V11
NanoFore Printer V1	16.6	3	3300	22.5	Orta	Düşük	Orta	Yüksek	Üretim	NanoForce Printer	V1
MedPulse Patch V12	14.3	5	1200	7.8	Orta	Orta	Düşük	Orta	Sağlık	MedPulse Patch	V12

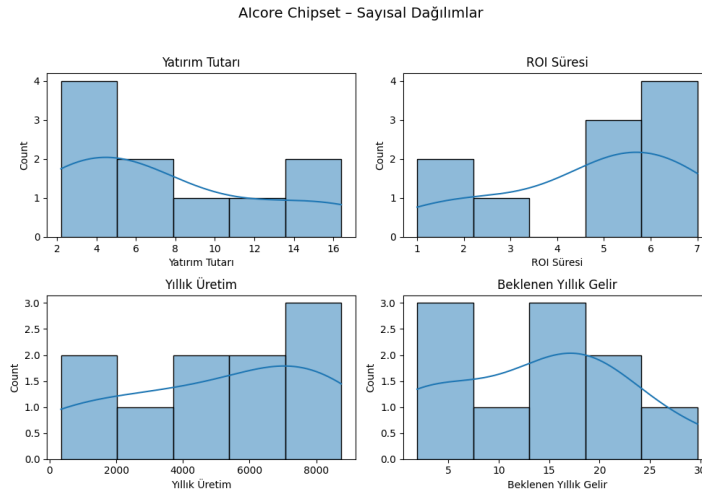
3.7.1 Grafikler

Bu bölümde elde edilen temel grafikler incelenecek ve yorumlanacaktır. İlk grafik olan yatırım tutarı-Beklenen Gelir grafiği Figür 1’de görülmektedir.



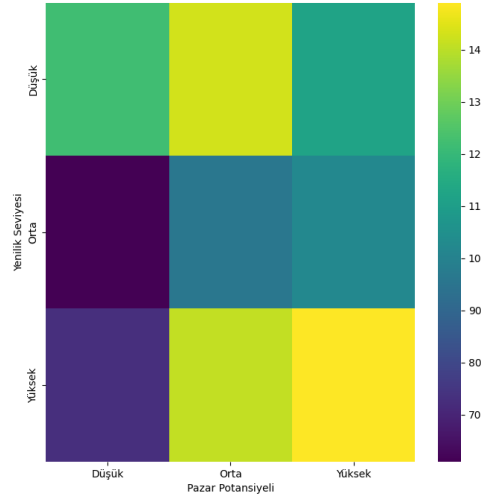
Figür 1: Yatırım Tutarı-Beklenen Gelir Grafiği

Figür 1’deki grafik incelendiğinde tüm ürün ailelerini kapsayan doğrusal bir eğilim sergilediği görülmektedir. 1-10 M TL aralığında getiri orantılı olarak artarken 10 M TL seviyesinden sonra giderek eğimin yataylaştığı görülmektedir. Bu durum beklenen gelirin yüksek bütçeli projelerde bile bir süreden sonra farklı faktörlere bağlı olduğuna işaret edebilmektedir. Bir sonraki grafikte veri setinden örnek bir ürün olan Alcore Chipset’in belirli özellikleri Figür 2’de incelenmiştir.



Figür 2: Alcore Chipset Ürün Ailesinin Dağılımı

Figür 2’de görülen grafik veri setinden örnek bir ürün olan Alcore Chipset’e ait olan dört temel sayısal değişkenin dağılımını sunmaktadır. 6.5 M TL medyana sahip bir yatırım tutarına sahipken bazı sürümlerde bu değer 16 M TL’ye çıkmaktadır. ROI süresi gözlemlerin çoğunda 5-7 yıl bandında toplanmış; 1-3 yıl gibi erken geri dönüş örnekleri ise belirli sürümlerin daha çok verimli olduğuna işaret eder. Figür 3’te yer alan ısı haritası sayesinde Yenilik Seviyesi ile Pazar Potansiyeli kategorilerinin birbirine ne kadar bağlı olduğu gözlemlenebilmektedir.



Figür 3: Yenilik Seviyesi-Pazar Potansiyeli Matrisi

Yüksek yenilik ile yüksek pazar potansiyelinin kesiştiği hücrenin en pozitif bölge olduğu görülmektedir ve bu güçlü ticari fırsatların olduğunu gösterebilir ve inovatif ürünlerin piyasada tutunacağına işaret eder. Diğer yandan en çok negatiflik gösteren hücre orta yenilik seviyesi düşük pazar potansiyeline denk gelmektedir. Bu daha temkinli yatırım yapılması yapılmasını veya farklı stratejiler denenmesi gerektiğini gösterebilir.

3.8 Veri Ön işleme

Veri ön işleme süreci makine öğrenimi modellerinin başarısını etkilemede önemli bir adımdır.

Bu çalışmada ilk olarak veri setindeki aykırı değerlerin tespiti için Interquartile Range (IQR) yöntemi tercih edilmiştir. IQR, verilerin orta %50’lik kısmının yayılımını ölçerek uç değerlerin belirlenmesinde etkili bir araçtır. Bu yöntem özellikle normal dağılıma harici de çarpık dağılımlarda da etkili olduğu için geniş bir uygulama alanına sahiptir [19].

Çalışmada aykırı değer hiç tespit edilmemiştir. Bu durumda verilerin sentetik olmasının da etkisi vardır.

Kategorik değişkenlerin sayısal formata dönüştürülmesi, makine öğrenimi algoritmalarının veriyi etkili bir şekilde işlemesi için kritik bir adımdır [20]. Çalışmadaki veride farklı kategorik değişkenlere farklı kodlama teknikleri uygulanmıştır. Sıralı kategorilere sahip olan “Çevresel Risk”, “Sosyal Katkı”, “Pazar Potansiyeli” ve “Yenilik Seviyesi” değişkenleri için ordinal kodlama kullanılmıştır. Düşük 1, orta 2, yüksek 3 değerlerini almıştır. Böylece doğal sıralama korunmuştur. Sırasız ve düşük sayıda kategorik değişkenlere one-hot kodlama yöntemiyle dönüştürülmüştür. Örneğin “Sektör” için bu yöntem kullanılarak her kategori için ikili değişken oluşturulmuştur.

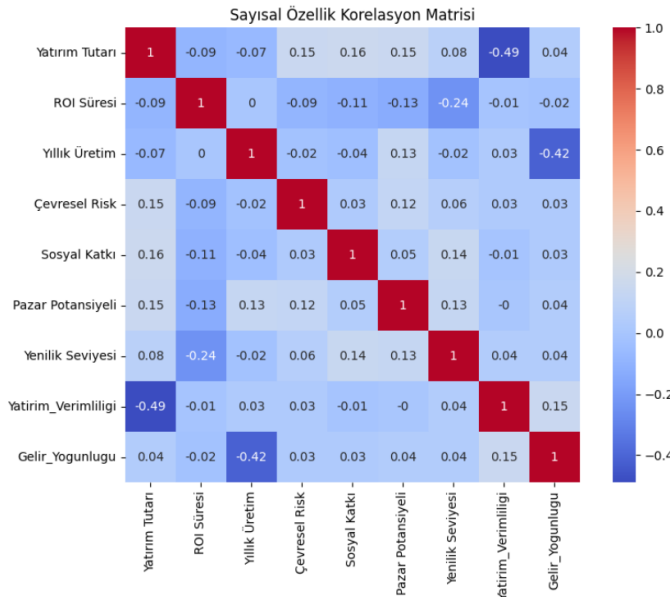
Yüksek sayıda kategorik değişkenler için de hedef değişkenin ortalamasına dayalı target encoding yöntemi

uygulanmıştır. Özellikle “ürün ailesi” gibi değişkenlerde her kategoriye karşılık gelen hedef değişkenin ortalamasıyla modelin daha anlamlı ilişkiler kurmasını sağlamaktadır.

Çoklu doğrusal bağlantılı regresyon, modellerde tahminlerin güvenilirliğini azaltan bir durumdur [21]. Bu durumu tespit etmek için Varyans Enflasyon Faktörü hesaplanmıştır. VIF değeri 10’un üzerinde olan değişkenler yüksek çoklu doğrusal bağlantı olmasına yol açar ve bu değişkenlerin modelden çıkarılması önerilir [22]. Fakat bakıldığında veride “Yatirim_Uretim_ETK” 14.7 VIF değeri, “Yatirim tutarı” 16.5 ile çok yüksek korelasyon göstermiştir. Bu nedenle “Yatirim_Uretim_ETK” veriden çıkarılmıştır ve tüm değişkenlerin VIF değeri 1-10 arasına düşmüştür.

Özellik mühendisliği kapsamında veri setine yeni değişkenler eklenmiştir. Bunlar “Yatirim_Verimliliği” (Beklenen Yıllık Gelir / Yatirim Tutarı), “Gelir_Yogunlugu” (Beklenen Yıllık Gelir / Yıllık Üretim) ve “Yatirim_Uretim_ETK” (Yatirim Tutarı*Yıllık Üretim) olmaktadır. “ROI_Kategori” değişkeni de “ROI_Süresi”nin kategorik temsildir ve ordinal şekilde kodlanmıştır. Yatırımın geri dönüş süresine göre sınıflandırma yapmaktadır. Son olarak “Risk yenilik” değişkeni ise “Çevresel Risk” ile “Yenilik Seviyesi” değişkenlerinin birleşiminden oluşarak etkileşimlerini yansıtmaktadır. Bu tür özellik mühendislikleri modelin karmaşık bağları öğrenmesinde yardımcı olurlar.

Son olarak değişkenler arasında ilişkileri görselleştirmek ve çoklu doğrusal bağlantıyı daha iyi anlamak için korelasyon ısı haritası kullanılmıştır. Bu görselleştirme değişkenler arasındaki pozitif ve negatif ilişkileri renkleriyle beraber göstererek modelleme sürecinde hangi değişkenlerin birlikte kullanılmasının uygun olacağını belirlemede faydalı olmaktadır [22]. Matris figür 4’te gösterilmektedir.



Figür 4: Korelasyon Isı Haritası

Korelasyon ısı haritası sayısal değişkenler arasındaki doğrusal ilişkilerin görselini sunar. Renk aralığı değişken çiftler arasındaki Pearson korelasyon katsayısını temsil eder; +1 koyu kırmızı ile güçlü pozitif ilişkiyi, -1 ise koyu mavi ile güçlü negatif ilişkiyi gösterir. Şekil üzerinde yer alan simetrik matris yapısı hangi değişken çiftlerinin birlikte yükseldiğini ya da ters hareket ettiğini tespit etmeyi sağlar.

Bu çalışmada, korelasyon ısı haritası sayesinde “Yatırım Tutarı” ile “Yatırım_Verimliliği” arasında negatif (–0.49) ve “Yıllık Üretim” ile “Gelir_Yogunluğu” arasında da belirgin negatif (–0.42) ilişki gözlenmiştir. Diğer değişkenler arasındaki korelasyonlar ise genellikle zayıf düzeydedir ve bu da çoklu doğrusal bağlantı riskinin sınırlı olduğunu göstermektedir. Korelasyon matrisleri, modelin girdi değişkenlerinin birbirinden ne ölçüde bağımsız olduğunu anlamak ve gerektiğinde bazı değişkenleri modelden çıkarmak için temel analiz aracıdır [22].

4 MODELLEME

Aşağıda uygulanan modelleme tekniklerine ve sonuçlarına değinilmektedir.

4.1 Modellemeye Giriş

Modelleme sürecindeki temel hedef “Beklenen Yıllık Gelir” değişkeninin tahmin edilmesidir. Bu değişken bir fizibilite raporunda proje yatırımının beklenen ekonomik geri dönüşünü ifade eder. Bu tahmin, karar destek sisteminin temel bir yapısını oluşturur. Yeni bir yatırım kararı alınırken projenin beklenen getirisine dair öngörüler karar verenlerin finansal risk ve kazancı daha iyi değerlendirmesini sağlar. Bu doğrultuda çalışmaya model performanslarını test etmek ve karşılaştırmak için %80 eğitim, %20 test şeklinde bölerek başlanmıştır. Bu ayırım sayesinde modellerin bilinmeyen verilerdeki genelleme yeteneği de gözlenmiştir. Model performansları karşılaştırırken R^2 skoru, ortalama mutlak hata (MAE), RMSE, MAPE metrikleri kullanılmıştır.

4.1.1 Lineer Regresyon

Modelleme ilk olarak lineer regresyon ile başlamıştır. Lineer regresyon bağımlı değişken ile bağımsız değişkenler arasındaki ilişkileri doğrusal bağlamda modellemeye çalışan bir yöntemdir [23]. Fakat lineer regresyon varsaydığı doğrusal ilişkiler nedeniyle çoğunlukla karmaşık ve çok boyutlu fizibilite verilerini yansıtmakta yetersiz kalabilmektedir. Burada da benzer bir durum gözlemlenmiştir. Model, verinin yalnızca %37’ sini açıklayabilmiştir ($R^2=0.372$) ve ortalama 4.91 M TL sapma ile beklentinin oldukça altında bir başarı göstermiştir. Bu da doğrusal yaklaşımın yoğun ve karmaşık proje verilerinde yetersiz kaldığını göstermektedir.

4.1.2 Random Forest

Lineer modelin yetersiz performansından sonra daha esnek modeller denenmiştir. Bu yüzden Random Forest Regressor kullanılmıştır. Bu yöntem çok sayıda karar ağacını rastgele alt örnekler üzerinden eğitir ve bunların tahminlerinin ortalamasını alarak sonucu belirler [24]. Random Forest, karar ağaçlarının yüksek varyanslı yapısını azaltarak daha istikrarlı tahminler sunarlar. Parametre olarak $n_estimators=300$ ve $random_state=42$ olarak belirlenmiştir. Sonuç olarak model R^2 skorunda %97.7’lük bir açıklama oranı yakalamış ve MAE 0.82 M TL sapma gibi düşük bir seviyeye düşürülmüştür. Bu sonuç modelin karmaşık ve doğrusal olmayan yapıları çok başarılı öğrendiğini göstermektedir.

4.1.3 XGBoost

Random Forest modelini daha ileri düzeye taşımak için boosting tabanlı XGBoost modeli uygulanmıştır. XGBoost hızlı hesaplaması, overfitting kontrolü ve ağaç düzenlemesi gibi özellikleriyle oldukça yaygın kullanılmaktadır [25]. Model ilk önce varsayılan parametrelerle test edilmiştir. Parametreler; $n_estimators=100$, $max_depth=3$, $learning_rate=0.1$, $subsample=1$ ve $random_state=42$ ’dir. Genel olarak metriklerde Random Forest’a göre düşük sonuçlanmıştır. Elde edilen sonuçlar Tablo 3’te gösterilmektedir. Modelin daha da iyileştirilmesi için hiper parametre ayarlanması BayesSearchCV ile yapılmıştır. Bayesian optimizasyonu her denemeden sonra öğrendikleriyle bir sonraki aşamayı daha iyi tahmin etmeye

çalışan olasılık temelli bir yaklaşımdır [26]. Bu çalışmada 30 deneme yapılmıştır ve model %98.6 açıklama oranına ve 0.64 M TL sapma düzeyine ulaşmıştır. Random Forest’a göre daha düşük hatayla daha yüksek genelleme sağlandığı görülmektedir. Bu, özellikle verilerin aşırı sapma gösterdiğini ayrıca bazı ürün türlerinde modelin dayanıklılığını ortaya koymaktadır.

4.1.4 Ensemble Stacking

Literatürde farklı modellerin Ensemble Stacking yöntemiyle sonuç vermesinden bahsedilmektedir. Bu çalışmada da farklı modellerin avantajlarını bir araya getirerek genel tahmin kalitesini artırmak için Stacking Regressor uygulanmıştır. Bu yöntem alt modellerin çıktısını meta modelin girdisi olarak kullanır [27]. Alt model olarak Lineer Regresyon, Random Forest ve optimize edilmiş XGBoost kullanılmıştır. Meta öğrenci olarak RidgeCV tercih edilmiştir. RidgeCV, ridge regresyonun çapraz doğrulamayla birleşik versiyonudur. Farklı alfa parametreleri ile en uygun olanı belirleyerek modelin performansını artırır [28]. Model, %98.7 R^2 başarısı ve yalnızca 0.62 M TL sapma seviyesiyle en iyi sonucu vermiştir. Bu model farklı modellerin farklı avantajlarını birleştirerek üst düzey bir başarı elde etmiştir.

Modeller	R^2	MAE	RMSE	MAPE (%)
Lineer Regresyon	0.372	4.91	6.45	56.14
Random Forest	0.977	0.82	1.24	7.22
XGBoost	0.962	1.19	1.58	0.11
Optimize edilmiş XGBoost	0.986	0.64	0.95	0.05
Stacking Regressor	0.987	0.62	0.91	0.05

Tablo 4 : Modeller ve Sonuçları

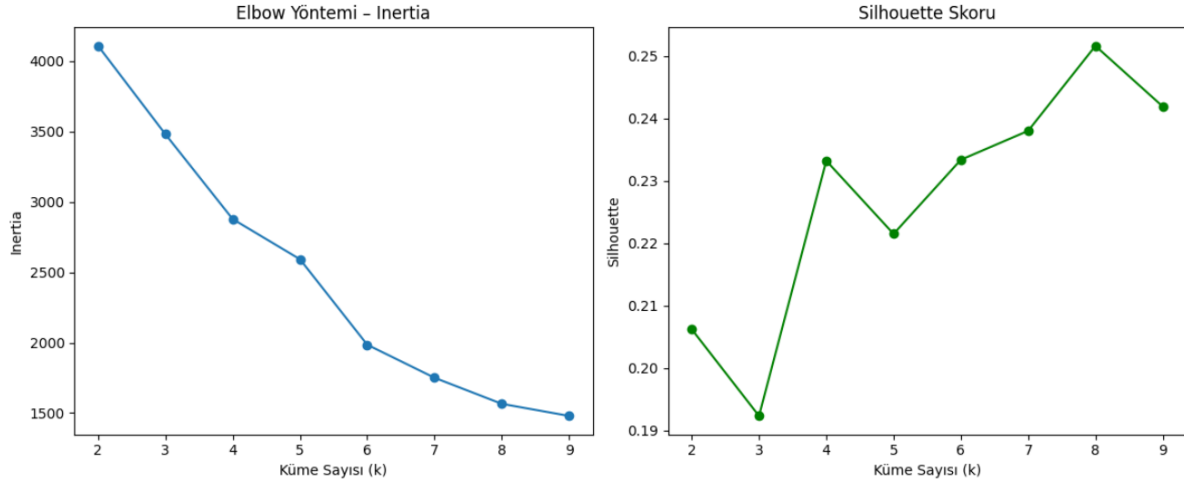
4.1.5 Sonuç Değerlendirmesi

Lineer regresyon açık ara doğrusal varsayımların yetersizliği nedeniyle en düşük sonucu veren model olmuştur. Random Forest ise ağaç tabanlı yaklaşımı sayesinde karmaşık ilişkileri yakalayarak daha başarılı olmuştur. XGBoost ilk aşamada Random Forest’in gerisinde kalsa da hiper parametre optimizasyonu sonrasında hata düzeyini yüksek oranda düşürmüştür. Son durumda Stacking Regressor ise tüm bu modellerin öğrenme yapılarını birleştirerek kararlı bir tahmin yapısı sunmuştur. Hem R^2 hem de hata metriklerine göre en düşük sapmayla en öne çıkan model olmuştur. Bir sonraki adım da, modelleme sonuçlarını segmentasyon analizleriyle desteklenmesi için kümeleme yöntemlerine geçilmiştir.

4.2 Kümeleme Analizi

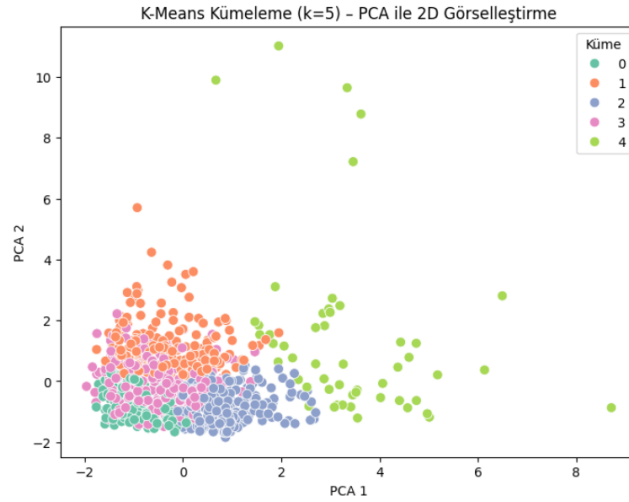
Bu çalışmada projeleri özellik bazlı segmentlere ayırmak ve benzer yapıya sahip girişimlerin birlikte analiz edilmesini sağlamak amacıyla K-Means algoritması kullanılmıştır. K-means önceden belirlenen k adet merkez etrafında verileri gruplayarak her gözlemi en yakın merkezle ilişkilendirir. Bunun için yalnızca sayısal değişkenler kullanılmıştır ve ölçekleme adımı olarak StandardScaler tercih edilmiştir.

Kümeleme için öz nitelikler “Yatırım Tutarı”, “Yıllık Üretim”, “ROI Süresi”, “Yatırım Verimliliği” ve “Gelir Yoğunluğu” olmuştur. Bu değişkenler, fizibilite projelerinin finans ve üretim yanını temsil etmektedir. En uygun k değerinin belirlenmesi için Elbow yöntemi ve Silhouette skorları kullanılmıştır. Elbow yönteminde k arttıkça modelin inertia değeri azalmaktadır fakat bir seviyeden sonra azalma yavaşlar [29]. Figür 5’te gösterilen grafiklerde bu seviye k=5 için gözlemlenmiştir. Aynı zamanda Figür 5’teki Silhouette skor grafiği de k=5 noktasının anlamlı olduğunu göstermektedir.



Figür 5: Elbow Metodu ve Silhouette Skoru

K=5 olarak belirlenen model veriyi 5 kümede toplamıştır. PCA (Principal Component Analysis) ile boyut indirgeme uygulanarak veriler iki boyutlu alanda Figür 6'da gösterildiği gibi renkli olarak görselleştirilmiştir. Burada her renk bir kümeyi temsil etmektedir.



Figür 6: PCA ile kümeleme sonuçları

Küme merkezlerini daha anlamlı değerlendirmek adına her küme için seçilen özelliklerin ortalama değerleri Tablo 5'te gösterilmiştir.

cluster label	Yatırım Tutarı	Yıllık Üretim	ROI Süresi	Yatırım Verimliliği	Gelir Yoğunluğu
0	15.35	7223.73	5.06	1.00	0.00
1	11.72	1811.47	5.03	1.39	0.01
2	5.05	6902.03	4.85	2.53	0.00
3	11.79	4898.25	1.88	1.44	0.00
4	2.97	3830.40	3.62	10.03	0.03

Tablo 5: Küme Özellikleri

Bakıldığında Küme 4, ortalama 2.97 M TL yatırımla en düşük maliyetli projeleri içerirken 10.03 değeriyle çok yüksek yatırım verimliliği göstermiştir. Öte yandan küme 0 yüksek yatırım ve üretimiyle büyük çaplı projeleri temsil etmektedir. Küme 3 ise en düşük ROI süresi ile yatırımın hızlı geri döndüğü projeleri içermektedir.

Bu kümeleme karar destek sistemi açısından önemlidir. Örneğin, yatırımcılar düşük maliyetli fakat yüksek verimlilik sağlayan projelere yönelmek isterlerse Küme 4'teki projelere odaklanabilirler Veya hızlı geri dönüş isteyenler Küme 3'e yönelebilir. Bu analizlerle birlikte projelerin özellikleri daha açıklanabilir hale getirilmiştir ve yorumlanabilirlikleri artırılmıştır. Sonraki adımda Monte Carlo simülasyonu aracılığıyla belirsizlikler değerlendirilecektir.

4.3 Monte Carlo Simülasyonu

Finansal fizibilite analizlerinde projelerin ekonomik getirilerini sadece tahminlerle değil aynı zamanda belirsizlikler altındaki olası senaryolarla değerlendirmek önemlidir. Bu sebepten ötürü modelin belirli değişkenlere hassasiyetini test etmek ve olası sonuç aralığını daha iyi tahminlemek için **Monte Carlo simülasyonu** uygulanmıştır. Monte Carlo yöntemi belirsizlik içeren sistemlerde olabilecek senaryoları rastgele örnekleme yoluyla tekrar ederek simüle eden çıktının dağılımını gösteren istatistiksel yaklaşımdır [30].

Bu çalışmada tahmin edilen yıllık gelir üzerindeki belirsizliği incelemek için Yatırım Tutarı, ROI Süresi, Gelir Yoğunluğu ve Yatırım Verimliliği gibi değişkenlerin farklı senaryolardaki durumları analiz edilmiştir. Her değişken orijinal veri setinden elde edilen istatistiklere dayanarak uygun dağılımlarla modellenmiştir ve yıllık net gelir 10.000 senaryo üzerinden hesaplanmıştır. Bu senaryolar aracılığıyla beklenen yıllık gelir değerlerinin dağılımı simüle edilerek ortalama değer, güven aralıkları ve uç değerlerin olasılıkları hesaplanmıştır.

Bu simülasyonun en büyük faydası yalnızca performansa değil aynı zamanda risk olasılığına da bakmasıdır. Özellikle belirsizlik altında karar veren yöneticilere bu yaklaşım yalnızca en doğru tahminin dışında en kötü veya en iyi olan senaryoyu da göstermektedir. Simülasyonda kullanılan tüm değişkenler olasılıksal dağılımlarla tanımlanmıştır. Aşağıda kullanılan parametrelerin detayları sunulmuştur.

4.3.1 Yıllık Üretim Miktarı

Log-normal dağılıma göre örneklenmiştir. Parametreler orijinal veri setindeki yıllık üretim değerlerinden çıkarılan ortanca ve varyans yapısına göre belirlenmiştir.

4.3.2 Yatırım Verimliliği

“Beklenen Yıllık Gelir / Yatırım Tutarı” oranı kullanılarak, normal dağılıma göre örneklenmiştir.

4.3.3 Faaliyet Gider Oranı

Ortalama %40 ve standart sapması %5 olacak şekilde normal dağılıma göre örneklenmiş ve 0-1 arasında sınırlandırılmıştır.

4.3.4 Gelir Yoğunluğu:

“Beklenen Yıllık Gelir / Yıllık Üretim” oranı üzerinden belirlenmiştir.

4.3.5 ROI Süresi

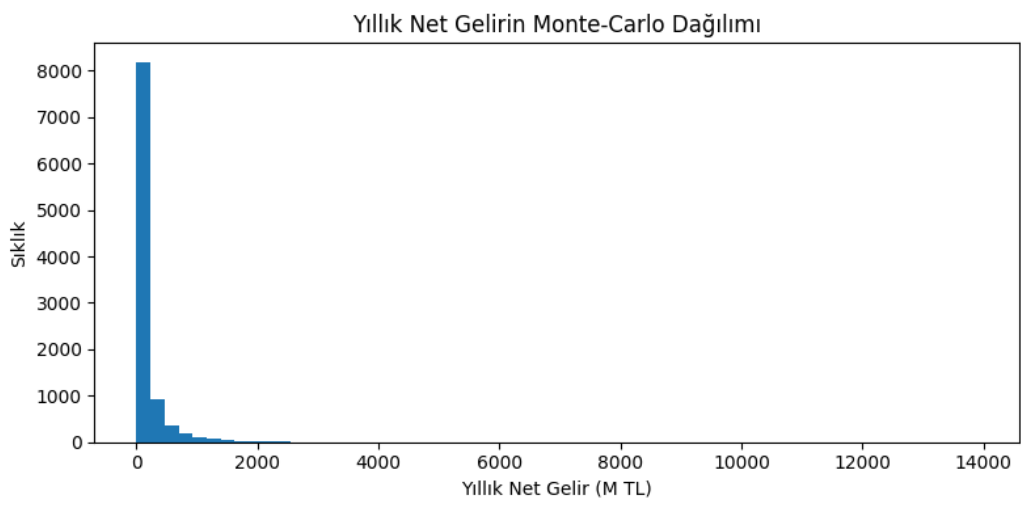
Üçgen dağılıma göre örneklenmiştir. Alt sınır 3 yıl, en olası değer 5 yıl ve üst sınır 8 yıl olarak belirlenmiştir.

4.3.6 Net Bugünkü Değer (NPV)

Her senaryo için yatırım tutarı 10 M TL kabul edilerek, %12 iskonto oranıyla yıllık net gelirin ROI süresi boyunca bugünkü değer hesaplaması yapılmıştır.

4.4 Simülasyon Sonuçları

Bu çalışmada gerçekleştirilen Monte Carlo simülasyonu sonucunda 10.000 senaryoya dayalı olarak önerilen ürünün yıllık net gelir potansiyeli değerlendirilmiştir. Simülasyonda her bir senaryo için birim katkı marjı, ROI gibi değişkenler rastgele biçimde üretilmiş ve bunlarla yıllık net gelir hesaplanmıştır. Figür 7’de gösterilen histogramda yıllık net gelirin sağa çarpık bir dağılımda olduğu görülmektedir.



Figür 7: Yıllık Net Gelir Monte Carlo Dağılımı

Bu durum gelirlerin büyük çoğunluğunun düşük ve orta seviyelerde sıklaştığını ancak düşük sayıda senaryodaysa çok fazla gelir elde edilebildiğini göstermektedir. Genele bakılırsa net gelirlerin büyük bir çoğunluğu 0-100 M TL arasına dağılmaktadır. Bu değerlere ek olarak pozitif NPV elde etme olasılığı ve 5 yıl veya daha kısa sürede yatırımın geri dönüşü hesaplanmıştır. İlk olarak 10.000 senaryo üstünden hesaplanan NPV üretme olasılığı %51.82 olarak hesaplanmıştır. Bu oran projeye yapılan yatırımın yarısından fazlasının kârlı sonuçlanacağına işaret eder. Yatırımcılar bu durumda daha rahat bir şekilde yatırımlarını yapabilmektedirler. Diğer önemli gösterge de yatırımın geri dönüş süresidir. Simülasyona göre yatırımlar 5 yıl veya daha kısa sürede %87.94 oranında geri dönüş sağlamaktadır. Dolayısıyla önerilen ürün projesi hem uzun vadede değer yaratabilme hem de orta vadede kendini amorti etme de oldukça güçlü bir potansiyele sahiptir.

Monte Carlo simülasyonu ile finansal belirsizlikler yıllık gelirler üzerinden incelenmiştir. Şimdiyse Markov zincirleri kullanılarak “Bir ürün projesi fizibilite sürecinde zaman içinde hangi karar aşamasına geçer?” sorusu cevaplanacaktır.

4.4 Lojistik Regresyon Destekli Markov Modeli

Markov zincirleri sistemin zamanla muhtemel durum geçişlerini analiz etmesini sağlayan güçlü bir stokastik modeldir. Ancak geleneksel Markov modellerinde geçiş olasılıklarının sabit olduğu varsayılır. Bu varsayım her ürün için geçiş dinamiklerinin aynı olduğu anlamına gelir fakat bu gerçek dünyada farklı özelliklere sahip projeler için geçerli değildir. Bu çalışmada, her ürünün kendine özgü yatırım profiline göre dinamik ve uyarlanabilir bir geçiş yapısı modellemek amacıyla geçiş olasılıklarının lojistik regresyon modelleri ile tahmin edilmesine dayanan bir Markov yaklaşımı benimsenmiştir. Çalışmamızda ürünlere ait zaman serisi geçmişi bulunmamaktadır. Bu nedenle geçiş olasılıklarını doğrudan sıklık analizleriyle hesaplamak mümkün olmamıştır. Bu nedenle ürünlerin sahip olduğu ROI Süresi, yatırım verimliliği, çevresel risk ve yenilik seviyesi üzerinden geçiş olasılıklarının tahmin edilmesi hedeflenmiştir. Bu amaçla literatürde önerilen regresyon destekli olasılık tahmini yaklaşımları incelenmiştir. Her hedef durum için ayrı ikili lojistik regresyon modeli kurulmuştur [31][32]. Bu modeller sayesinde her ürün için A, R, E ve I'ya geçiş olasılıkları elde edilmiştir ve sonrasında bu tahminler softmax normalizasyonu ile olasılık vektörüne dönüştürülmüştür. Durumların açıklamaları Tablo 4'te belirtilmektedir. Bu sayede sabit bir yapıdansa her ürünün türüne göre özelleşmiş bir geçiş matrisi elde edilmiştir. Bu süreçte R ve I gibi bazı karar durumlarında veriler seyrek bulunduğu için bu sınıflar model bu sınıfları öğrenmekte zorlanmıştır. Örneğin ürünlerin %60'ı onaylanırken sadece %10'u reddedilmektedir. Bu durumun önüne geçmek için class_weight paramtresi "balanced" olarak ayarlanmıştır. Bu parametre nadir görülen sınıflara daha fazla ağırlık vererek modelin dengeli kararlar almasına yardımcı olmaktadır [33].

4.4.1 Uygulama ve Simülasyon

İlk olarak ROI süresi, yatırım verimliliği, yenilik seviyesi ve çevresel risk değerleri modele girdi olarak verilmiştir. 4 ayrı lojistik regresyon modelinden elde edilen olasılık değerleri softmax ile normalize edilmiştir. Sonuç olarak elde edilen geçiş vektörleri her ürünün başlangıçta inceleme (I) durumunda olduğu varsayılarak 5 adımlık Markov zincirinde kullanılmıştır. Son olarak her adımda ürünlerin durumu özgün geçiş olasılıkları kullanılarak güncellenmiş ve kaydedilmiştir.

4.4.2 Simülasyon Sonuçları

5 Adım sonunda ürünlerin son durum dağılımı Tablo 6'da gösterilmektedir.

Durum	Açıklama	Oran (%)
A	Onaylandı	52.27
R	Reddedildi	44.02
I	İncelemede kaldı	1.88
E	Revizyonda kaldı	1.82

Bu dağılım ile Markov simülasyonunun fazla reddedilme tespiti yaptığına ve modelin veriye bağlı kalarak gerçekçi analizler yaptığını görmek mümkündür. Böylelikle kullanılan yaklaşım sadece geçerli durumları modellemekle kalmayıp öte taraftan ürün bazında özelleştirilmiş karar simülasyonları gerçekleştirmektedir.

5 SONUÇ

Bu çalışmada var olmayan ürünler için fizibilite analizi ve karar destek sistemi tasarımı gerçekleştirilmiştir. Öncelikle 16 farklı fizibilite raporu incelenerek ortak başlıklar ve veri şablonları çıkarılmıştır ardından LLM destekli sentetik veri üretimi yapılmıştır. Bu veri kümesi üzerinde yapılan keşifsel veri analizi ile yatırım tutarı, beklenen gelir, ROI süresi gibi

temel ilişkiler görselleştirilmiştir ve korelasyon, dağılım analizleri gibi analizlerle çalışma zenginleştirilmiştir. Ayrıca hem makine öğrenmesi algoritmalarıyla hem de kümeleme kullanılarak performans karşılaştırmaları ayrıntılı olarak sunulmuştur. Markov ve Monte Carlo gibi stokastik modellerle karar verme aşamasına dair risk ve belirsizlik durumları analiz edilmiştir. Özetle makine öğrenmesi modellerinde en yüksek doğruluğu $R^2=0.987$ ile Stacking Regressor sağlamıştır. Kümeleme de 5 farklı yatırım profili belirlenerek her grubun özellikleri incelenmiştir. Ardından yapılan Monte Carlo simülasyonu ile %52 civarında pozitif NPV ve %87 ROI hesaplanmıştır. Son adımda uygulanan Markov zinciri ile karar aşamalarındaki geçiş olasılıkları modellenmiş sonucunda red riskinin ve onay riskinin fazla olduğu tespit edilmiştir. Sonuç olarak LLM destekli sentetik veri ve entegre edilmiş analitik yöntemlerle geliştirilen karar destek sistemi, ürünün fizibilitesine dair önemli içgörüler sunmaktadır.

6 REFERANSLAR

- [1] H. İ. Şahin and B. Er, "KOBİ'ler İçin Finansal Fizibilite Otomasyonu Hazırlanmasına Yönelik İhtiyaç Araştırması," *Karadeniz Teknik Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, vol. 12, no. 24, pp. 43–62, Mar. 2021
- [2] F. Hendra, S. Supriyono, R. Efendi, R. Rosalinda, and R. Indriyati, "A Business Feasibility Analysis of Small and Medium Enterprises for Product Strategy Determination," *Scientific Journal of Reflection: Economic, Accounting, Management and Business*, vol. 4, no. 3, pp. 421–431, Jul. 2021
- [3] P. K. Dey, "Integrated approach to project feasibility analysis: a case study," *Impact Assessment and Project Appraisal*, vol. 19, no. 3, pp. 235–245, Sep. 2001
- [4] J. Merkert, M. Mueller, and M. Hubl, "A Survey of the Application of Machine Learning in Decision Support," in *European Conference on Information Systems 2015*, Münster, Germany, Jun. 2015
- [5] K. Doroshkevych, O. Maslak, V. Voronovska, and I. Salata, "Use of Markov chains for modeling processes of strategic planning of innovative activity of the enterprise," *WSEAS Transactions on Environment and Development*, vol. 16, pp. 440–447, Jun. 2020
- [6] C.-J. Cheng, S. W. Chiu, C.-B. Cheng, and J.-Y. Wu, "Customer lifetime value prediction by a Markov chain based data mining model: Application to an auto repair and maintenance company in Taiwan," *Scientia Iranica*, vol. 19, no. 3, pp. 849–855, 2012
- [7] A. M. Haddadi and H. Hamidi, "A hybrid model for improving customer lifetime value prediction using stacking ensemble learning algorithm," *Computers in Human Behavior Reports*, vol. 18, p. 100616, Feb. 2025
- [8] G. Perboli and E. Arabnezhad, "A machine learning-based DSS for mid and long-term company crisis prediction," *Expert Systems with Applications*, vol. 174, p. 114758, Feb. 2021
- [9] A. Merisalo, "Applying Monte Carlo simulation to model a sales process for forecasting future sales: A case study for a Finnish recruitment consulting company," M.Sc. thesis, Aalto Univ. Sch. of Business, Espoo, Finland, 2018
- [10] J. Cao and Y. Zhang, "Optimization of management decision based on machine learning: Application of intelligent algorithms in enterprise management systems," in *Proc. 2024 Int. Conf. on Computers, Information Processing and Advanced Education (CIPAE)*, Macau, 2024, pp. 719–726
- [11] Q. Wang, D. M. Kilgour, and K. W. Hipel, "Fuzzy real options for risky project evaluation using least squares Monte-Carlo simulation," *IEEE Systems Journal*, vol. 5, no. 3, pp. 385–395, Sep. 2011
- [12] O. V. B. Nainggolan, "Feasibility analysis of small and medium enterprises of shoes and slippers in Bogor," *Advances in Economics, Business and Management Research*, vol. 117, pp. 166–173, 2020
- [13] L. Moreno, M. Panizzi, and R. Bertone, "Approach to a feasibility analysis method for the deployment of software systems for SMEs," unpublished.
- [14] A. Halid, M. Muhtar, and S. Y. Mokodompit, "Financial feasibility analysis, small business farm beef cattle livestock in Gorontalo District," *Jurnal Perspektif Pembiayaan dan Pembangunan Daerah*, vol. 5, no. 2, pp. 105–114, Oct.-Dec. 2017.
- [15] Y. Song, T. Wang, S. K. Mondal, and J. P. Sahoo, "A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities," *arXiv preprint*, arXiv:2205.06743, May 2022

- [16] R. Bayram, B. Epli, B. Tahtacı, and S. Karaman, "KOBİ Yönetim Süreçlerinde Yapay Zekâ Entegrasyonu ve SAAS Tabanlı Uygulamaların Potansiyeli," *İşletme Araştırmaları Dergisi (Journal of Business Research-Turk)*, vol. 17, no. 1, pp. 662–675, 2025
- [17] A. A. Barr, J. Quan, E. Guo and E. Sezgin, "Large language models generating synthetic clinical datasets: a feasibility and comparative analysis with real-world perioperative data," *Frontiers in Artificial Intelligence*, vol. 8, Art. no. 1533508, Feb. 2025
- [18] F. Sufi, "Addressing data scarcity in the medical domain: A GPT-based approach for synthetic data generation and feature extraction," *Information*, vol. 15, no. 5, Art. no. 264, pp. 1–31, May 2024
- [19] V. Verardi and C. Vermandele, "Outlier identification for skewed and/or heavy-tailed unimodal multivariate distributions," *Journal de la Société Française de Statistique*, vol. 157, no. 2, pp. 90–114, 2016
- [20] F. Pargent, F. Pfisterer, J. Thomas and B. Bischl, "Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features," *Computational Statistics*, vol. 37, pp. 2671–2692, 2022
- [21] D. J. Mundfrom, M. D. Smith, and L. W. Kay, "The effect of multicollinearity on prediction in regression models," *General Linear Model Journal*, vol. 44, no. 1, pp. 24–28, 2018
- [22] P. Kulan ve diğerleri, "Graphical representation of data prediction potential: correlation graphs and correlation chains," *The Visual Computer*, cilt 39, no. 5, ss. 1773–1793, Mayıs 2023.
- [23] C. K. Puteri and L. N. Safitri, "Analysis of linear regression on used car sales in Indonesia," *J. Phys.: Conf. Ser.*, vol. 1469, no. 1, Art. no. 012143, pp. 1–9, 2020
- [24] E. Scornet, G. Biau and J.-P. Vert, "Consistency of random forests," *The Annals of Statistics*, vol. 43, no. 4, pp. 1716–1741, 2015
- [25] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016
- [26] J. Snoek, H. Larochelle and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 2951–2959, 2012
- [27] R. Lin, T. Naselaris, K. Kay, and L. Wehbe, "Stacked regressions and structured variance partitioning for interpretable brain maps," *NeuroImage*, vol. 298, Art. no. 120772, 2024
- [28] P. Patil, Y. Wei, A. Rinaldo, and R. J. Tibshirani, "Uniform consistency of cross-validation estimators for high-dimensional ridge regression," in *Proc. 24th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, San Diego, CA, USA, vol. 130, pp. 1–31, 2021
- [29] Humaira, H., & Rasyidah, R. (2020, February). Determining the appropriate cluster number using elbow method for k-means algorithm. In *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018*, 24–25 January 2018
- [30] F. Diker and E. Aslan, "Monte Carlo Simülasyonu ve Zaman Serisi Yöntemlerine Göre Bir Üretim Firmasında Tahminleme," *Journal of Academic Value Studies*, vol. 8, no. 1, pp. 26–36, 2022
- [31] E. Kozłowski, A. Borucka, and A. Świdorski, "Application of the logistic regression for determining transition probability matrix of operating states in the transport systems," *Eksplotacja i Niezawodność – Maintenance and Reliability*, vol. 22, no. 2, pp. 192–200, 2020
- [32] J. R. Eastman and J. He, "A regression-based procedure for Markov transition probability estimation in land change modeling," *Land*, vol. 9, no. 11, Art. no. 407, pp. 1–12, Oct. 2020
- [33] B. Bakırarar and A. H. Elhan, "Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research," *Türkiye Klinikleri Journal of Biostatistics*, vol. 15, no. 1, pp. 19–29, 2023