

Black Sea Journal of Engineering and Science





**BLACK SEA JOURNAL OF ENGINEERING AND SCIENCE
(BSJ ENGIN SCI)**



Black Sea Journal of Engineering and Science (BSJ Eng Sci) is a double-blind peer-reviewed, open-access international journal published electronically 4 times (January, April, July and October) in a year by since January 2018. It publishes, in English and Turkish, full-length original research articles, innovative papers, conference papers, reviews, mini-reviews, rapid communications or technical note on advances in a wide range of scientific disciplines from all fields of engineering and science and from any source.

ISSN 2619 - 8991

Phone: +90 362 408 25 15

Fax: +90 362 408 25 15

Email: bsjsci@blackseapublishers.com

Web site: <http://dergipark.gov.tr/bsengineering>

Sort of publication: Periodically 4 times (January, April, July and October) in a year

Publication date and place: October 01, 2022 - Samsun, TÜRKİYE

Publishing kind: Electronically

OWNER

Assoc. Prof. Uğur ŞEN

DIRECTOR IN CHARGE

Prof. Dr. Hasan ÖNDER

EDITOR BOARDS

EDITOR IN CHIEF

Prof. Dr. Hasan ÖNDER, Ondokuz Mayıs University, TÜRKİYE

Assoc. Prof. Uğur ŞEN, Ondokuz, Mayıs University, TÜRKİYE

SECTION EDITORS*

Prof. Dr. Ahmet UYANIK, Ondokuz Mayıs University, TÜRKİYE

Prof. Dr. Berna KILIÇ, Ege University, TÜRKİYE

Prof. Dr. Çiğdem TAKMA, Ege University, TÜRKİYE

Prof. Dr. Ertan BUYRUK, Sivas Cumhuriyet University, TÜRKİYE

Prof. Dr. Gökhan CİVELEKOĞLU, Akdeniz University, TÜRKİYE

Prof. Dr. Hasan TANAK, Amasya University, TÜRKİYE

Prof. Dr. Hasan TEMİZ, Ondokuz Mayıs University, TÜRKİYE

Prof. Dr. İbrahim Özgür DENEME, Aksaray University, TÜRKİYE

Prof. Dr. İbrahim UĞUR, Süleyman Demirel University, TÜRKİYE

Prof. Dr. Theodore A. TSİLİGİRİDİS, Athens University of Agriculture, GREECE

Prof. Dr. Ümit Cafer YILDIZ, Karadeniz Technical University, TÜRKİYE

Assoc. Prof. Dr. Mehmet EBEOĞLUGİL, Dokuz Eylül University, TÜRKİYE

Assoc. Prof. Dr. Belgin KARABACAKOĞLU, Eskişehir Osmangazi University, TÜRKİYE

Assoc. Prof. Dr. Bülent BOSTANCI, Erciyes University, TÜRKİYE

Assoc. Prof. Dr. Ergün EKİCİ, Çanakkale Onsekiz Mart University, TÜRKİYE

Assoc. Prof. Dr. Kadyrbay CHEKİROV, Kyrgyz Turkish Manas University, KYRGYZSTAN

Assoc. Prof. Dr. Nilüfer YURTAY, Sakarya University, TÜRKİYE

Assoc. Prof. Dr. Özgür Hakan AYDOĞMUŞ, Social Sciences University of Ankara, TÜRKİYE

Assoc. Prof. Dr. Sinan AKISKA, Ankara University, TÜRKİYE

Asst. Prof. Dr. Mehmet GÜÇYETMEZ, Kırşehir Ahi Evran University, TÜRKİYE

Asst. Prof. Dr. Melahat CİHAN, Samsun University, TÜRKİYE

Asst. Prof. Dr. Sedat KARADAVUT, Trakya University, TÜRKİYE

* The ranking is arranged alphabetically within the academic title

EDITORIAL - ADVISORY BOARD*

Prof. Dr. Aglaia (Litsa) LIOPA-TSAKALIDI, Institute of Western Greece, GREECE

Prof. Dr. Ercan EFE, Kahramanmaraş Sutcu Imam University, TÜRKİYE

Prof. Dr. Mohammad Masood TARIQ, University of Balochistan, PAKISTAN

Prof. Dr. Mustafa Çağatay TUFAN, Ondokuz Mayıs University, TÜRKİYE

Prof. Dr. Özkan GÖRGÜLÜ, Ahi Evran University, TÜRKİYE

Assoc. Prof. Dr. Taner TUNÇ, Ondokuz Mayıs University, TÜRKİYE

Asst. Prof. Dr. Emil OMURZAK, Kyrgyz-Turkish Manas University, KYRGYZSTAN

Asst. Prof. Dr. Yılmaz KAYA, Ondokuz Mayıs University, TÜRKİYE

* The ranking is arranged alphabetically within the academic title

STATISTIC EDITOR

Prof. Dr. Mehmet TOPAL, Kastamonu University, TÜRKİYE

ENGLISH EDITOR

Asst. Prof. Dr. Betül ÖZCAN DOST, Ondokuz Mayıs University, TÜRKİYE

TURKISH EDITOR

Prof. Dr. Serkan ŞEN, Ondokuz Mayıs University, TÜRKİYE

REVIEWERS OF THE ISSUE*

Prof. Dr. Elijah OMIDIORA, Ladoké Akintola University of Tehnology, NIGERIA

Prof. Dr. Hakan ARSLAN, Ondokuz Mayıs University, TÜRKİYE

Prof. Dr. Olabiyisi STEPHEN, Ladoké Akintola University of Tehnology, NIGERIA

Prof. Dr. Yakup Emre ÇORUHLU, Karadeniz Technical University, TÜRKİYE

Assoc. Prof. Dr. Okan YILDIZ, Karadeniz Technical University, TÜRKİYE

Assist. Prof. Dr. Aycan Mutlu YAĞANOĞLU, Atatürk University, TÜRKİYE

Assist. Prof. Dr. Esra YAVUZ, Şırnak University, TÜRKİYE

Dr. Cem TIRINK, Iğdır University, TÜRKİYE






Dr. Ebru ERGÜNEŞ BERKİN, Islahiye Direct. of District Agri. and Forestry, TÜRKİYE

Dr. Hasan Alp SAHİN, University of Szeged, HUNGARY

* The ranking is arranged alphabetically within the academic title

Table of Contents

Research Article

- 1. USE OF GENERALIZED ADDITIVE MODELS IN ANIMAL SCIENCE** 
(Genelleştirilmiş toplamsal modellerin zooteknide kullanımı)
Nursen KURDAL, Hasan ÖNDER.....137-145
- 2. AN EVALUATION OF SUPERMARKETS FROM THE LENS OF MULTIPLE CRITERIA: THE INTUITIONISTIC FUZZY TOPSIS METHOD** 
(Çoklu kriter merceğinden süpermarketlerin değerlendirilmesi: sezgisel bulanık TOPSIS yöntemi)
Feride TUĞRUL.....146-150
- 3. YORUBÁ CHARACTER RECOGNITION SYSTEM USING CONVOLUTIONAL RECURRENT NEURAL NETWORK** 
(Evrşimli tekrarlayan sinir ağı kullanan Yorúbá karakter tanıma sistemi)
Jumoke Falilat AJAO, Shakirat Ronke YUSSUF, Abdulazeez Olorundare AJAO.....151-157
- 4. PERFORMANCE COMPARISON OF K-MEANS AND DBSCAN METHODS FOR AIRLINE CUSTOMER SEGMENTATION** 
(Havayolu müşteri segmentasyonu için K-Means ve DBSCAN yöntemlerinin performans karşılaştırması)
Kevser ŞAHİNBAŞ.....158-165
- 5. USING REAL OPTIONS IN WIND POWER PLANT EVALUATION** 
(Rüzgâr enerji santrali değerlemesinde reel opsiyonların kullanılması)
Duygu BIYIKLI, Faik Ahmet SESLİ, Pelin KASAP.....166-173



GENELLEŞTİRİLMİŞ TOPLAMSAL MODELLERİN ZOOTEKNİDE KULLANIMI

Nursen KURDAL¹, Hasan ÖNDER^{1*}

¹Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 55139, Samsun, Türkiye

Özet: Regresyon analizi pek çok bilim dalında olduğu gibi zootekni alanında da en sık kullanılan istatistik yöntemlerden birisidir. Ancak pek çok biyolojik çalışma için doğrusal regresyon varsayımlarının sağlanması mümkün olmamaktadır. Özellikle kesikli verilerin açıklayıcı değişken olarak kullanılması gereken durumlarda doğrusal regresyon kullanılarak elde edilen modellerin istatistiksel olarak hatalı sonuçlar üretebileceği bilinmektedir. Bu gibi durumlar için doğrusal regresyon yerine parametrik olmayan ya da yarı parametrik yöntemlerin kullanılması önerilmektedir. Bu çalışmada, kesikli açıklayıcı değişken varlığında önerilen genelleştirilmiş toplamsal modellerin zootekni alanından elde edilmiş bir veri kümesi kullanılarak tanıtılması amaçlanmıştır. Sonuç değişkeni olarak süt keçilerinden alınan laktasyon süt verimi kullanılırken açıklayıcı değişkenler olarak laktasyon süresi ve ana yaşı kullanılmıştır (n=166). Laktasyon süt verimi, laktasyon süresi ve ana yaşı değişkenlerinin normal dağılım göstermediği belirlenmiştir. Elde edilen doğrusal regresyon modelinin istatistiksel olarak anlamlı olduğu (P<0,001), Hata Kareler Ortalamasının (HKO) 5826,039 ve belirtme katsayısının (R²) 0,16 olduğu belirlenmiştir. Düşük bir belirtme katsayısına rağmen modelin anlamlı olmasının örnek büyüklüğüne bağlı olabileceği yorumlanmıştır. Elde edilen genelleştirilmiş toplamsal model için Ana yaşı değişkeninin laktasyon süresi değişkenine göre daha başarılı düzleştirme değerine sahip olduğu belirlenmiştir. Doğrusal regresyon ve genelleştirilmiş toplamsal modelleri karşılaştırabilmek adına Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE) ve ortalama mutlak sapma (MAD) değerleri hesaplanmıştır. Elde edilen bulgulara göre doğrusal regresyon analizinin tahmin başarısının daha yüksek olduğu anlaşılmaktadır. Ancak açıklayıcı değişkenler arasında çoklu bağlantı problemi olmamasına (VIF=1.000) ve sapmaların normal dağılıma sahip olmasına (P>0,05) rağmen değişkenlerin normal dağılmadığı bilinmektedir. Bu durumda genelleştirilmiş toplamsal modellerin kullanılmasının önerilebileceği söylenebilir. Sonuç olarak zootekni çalışmalarında kesikli değişkenler ile model oluşturmak gerektiğinde doğrusal modeller yerine genelleştirilmiş toplamsal modellerin kullanılmasının elde edilecek modelin güvenilirliğini artıracığı için önerilebileceği belirlenmiştir.

Anahtar kelimeler: Genelleştirilmiş toplamsal modeller, Regresyon, Kesikli veri, Zootekni

Use of Generalized Additive Models in Animal Science

Abstract: Regression analysis is one of the most frequently used statistical methods in the field of animal science, as it is in many branches of science. However, it is not possible to provide linear regression assumptions for many biological studies. It is known that especially in cases where discrete data should be used as explanatory variables, the models obtained using linear regression can produce statistically incorrect results. For such cases, it is recommended to use non-parametric or semi-parametric methods instead of linear regression. In this study, it is aimed to introduce the generalized additive models proposed in the presence of discrete explanatory variables, using a dataset obtained from the field of animal science (n=166). While lactation milk yield was used as the outcome variable, lactation period and maternal age taken from dairy goat were used as explanatory variables. It was determined that the variables of lactation milk yield, lactation period and maternal age did not show normal distribution. It was determined that the obtained linear regression model was statistically significant (P<0.001), the Mean Square Error (MSE) was 5826.039 and the coefficient of determination (R²) was 0.16. It has been interpreted that the significance of the model, despite a low coefficient of determination, may depend on the sample size. For the generalized additive model obtained, it was determined that the maternal age variable had a more successful smoothing value than the lactation period variable. Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE) and mean absolute deviation (MAD) values were calculated to compare linear regression and generalized additive models. According to the findings obtained, it was understood that the estimation success of the linear regression analysis was higher. However, it was known that although there was no multicollinearity problem between the explanatory variables (VIF=1.000) and the deviations had a normal distribution (P>0.05), the variables were not normally distributed. In this case, it can be said to recommend to use of generalized additive models. As a result, it has been determined that when it is necessary to create a model with discrete variables in animal science studies, using generalized additive models instead of linear models can be recommended as it will increase the reliability of the model to be obtained.

Keywords: Generalized additive models, Regression, Discrete data, Animal science

*Sorumlu yazar (Corresponding author): Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 55139, Samsun, Türkiye

E mail: honder@omu.edu.tr (H. ÖNDER)

Nursen KURDAL <https://orcid.org/0000-0001-7198-2329>

Hasan ÖNDER <https://orcid.org/0000-0002-8404-8700>

Gönderi: 18 Şubat 2022

Kabul: 23 Nisan 2022

Yayınlanma: 01 Ekim 2022

Received: February 18, 2022

Accepted: April 23, 2022

Published: October 01, 2022

Cite as: Kurdal N, Önder H. 2022. Use of generalized additive models in animal science. BSJ Eng Sci, 5(4): 137-145.



1. Giriş

Regresyon analizi, aralarında sebep sonuç ilişkisi bulunan değişkenler arasındaki ilişkiyi incelemek, tespit etmek, modellemek ve bu modeli kullanarak ilgili konu üzerinde tahminler (estimation) ya da kestirimler (prediction) yapabilmek amacıyla kullanılan istatistiksel yöntemlerden birisidir (Bağdatlı, 2010; Arı ve Önder, 2013). Regresyon analizi, hemen hemen tüm bilim dallarında en çok kullanılan yöntemlerden birisidir. Regresyon analizinin metodolojisi, sebep değişkenlerin sonuç değişken üzerindeki etkisini ve katkısını değerlendirerek sonuç değişkeni ve açıklayıcı değişkenler arasındaki ilişkiyi ortaya koymak ve modellemektir (Durmuş, 2018). Birçok biyolojik olayda sebep sonuç ilişkisine rastlamak mümkündür. Örneğin; yem miktarının (sebebi), süt miktarına (sonuç) etkisi bir örnek olarak verilebilir. Bu analiz tekniğinde iki (basit doğrusal regresyon) veya daha fazla değişken (çoklu doğrusal regresyon) arasındaki ilişkiyi açıklamak için matematiksel bir model kullanılır. Zootekni alanında da verilerin değerlendirilmesi işleminde değişkenler arası ilişkilerin incelenmesi amacıyla en çok kullanılan istatistiksel yöntemlerden birisi çoklu doğrusal regresyon analizidir.

Çoklu regresyon analizi, bağımlı değişkenle ilişkili olan iki ya da daha fazla açıklayıcı değişkene dayalı olarak, bağımlı değişkenin tahmin edilmesine yönelik bir analiz yöntemidir. Açıklayıcı değişkenler tarafından sonuç değişkeni de açıklanan toplam varyansın yorumlanmasına ve açıklayıcı değişkenlerle sonuç değişkeni arasındaki ilişkinin yönüne ve miktarına dair yorum yapma imkanı verir (Kanıt ve Baykan, 2004).

Doğrusal regresyon modelleri için açıklayıcı değişkenlerin birbirlerinden bağımsız olduğu ve benzer dağılıma sahip oldukları varsayımı kabullenilmektedir. Parametrik regresyon analizinin bir diğer varsayımı ise; regresyon katsayısının test edilmek istenmesi durumunda hata terimlerinin normal dağılıma sahip olmasıdır. Bu varsayımlar altında regresyon katsayısının En Küçük Kareler (EKK) tahmin edicisi $n-k$ serbestlik dereceli t dağılımına sahiptir (burada, n ve k sırasıyla örnek büyüklüğü ve açıklayıcı değişken sayısını göstermektedir). Ayrıca, merkezi limit teoreminden bilindiği gibi regresyon katsayısı optimum örnek büyüklüğüne ulaşıldığında normal dağılıma sahip olmaktadır. Ancak, hata terimlerinin normal dağılıma sahip olmadığı ve örnek büyüklüğünün yeterli düzeye ulaşmadığı durumlarda, EKK tahmin edicisi hatalı sonuçlar üretebilmektedir (Önder, 2007). Yukarıda bahsi geçen özelliklere ek olarak, çoklu doğrusal regresyon analizi için temel bir varsayım olan, açıklayıcı değişkenle sonuç değişkeni arasındaki ilişkilerin doğrusal, karesel veya küresel gibi tanımlı bir matematiksel fonksiyon olması varsayımının da sağlanması gerekir. Bu varsayımlar sağlandığında, doğrusal regresyon analizi oldukça güçlü bir parametrik yöntemdir. Ancak biyolojik çalışmalarda elde edilen örneklere matematiksel fonksiyon tanımlamak çok zor olmakta hatta bazı

durumlarda mümkün olmamaktadır. Bu çerçevede sonuç değişkeni ve açıklayıcı değişkenler arasındaki ilişkinin fonksiyonel şekli biliniyorsa parametrik regresyon modelleri, bilinmiyorsa parametrik olmayan regresyon modelleri kullanılır (Durmuş, 2018). Ancak, veri yapısına uygun olmayan regresyon yöntemlerinin kullanılması yanıltıcı sonuçların elde edilmesine neden olabilmektedir. Regresyon analizinde üzerinde çalışılan değişkenler sürekli ya da kesikli yapıda olabilmektedir. Bu durumda, veri yapısına bağlı olarak farklı regresyon modelleri kullanılabilir (Arı ve Önder, 2013). Sonuç değişkeninin sürekli olmadığı durumlarda regresyon analizi uygulayabilmek için de genelleştirilmiş modellerde bulunmaktadır. Değişkenlerin sürekli olduğu ancak normal dağılım göstermediği durumlarda veri analizi için geliştirilen Genelleştirilmiş Doğrusal Modeller (GDM) kullanılmaktadır (Savaş ve Cengiz, 2009). Şenel, ve ark. (2009)'nın bildirdiğine göre; GDM, ilk kez 1997 yılında Nelder ve Wedderburn tarafından ileri sürülmüştür. Çok geniş uygulama alanlarında kullanılmıştır. Bu alandaki ilk detaylı kitap McCullagh ve Nelder (1989) tarafından yazılmıştır.

Doğrusal modeller de her bir tahmin edicide sonuç değişkenini doğrusal kabul ederken toplamsal modeller de ise sadece düz bir şekilde bağımlı değişkenin her bir tahmin edici tarafından etkilediğini varsayılmaktadır. Genelleştirilmiş Toplamsal Modeller (GTM), tahmin ediciler ve bağımlı değişkenler arasındaki esnek toplamsal olmayan ilişkileri modelleyen, doğrusal modeller ve GDM'in genişletilmiş bir ifadesidir (Savaş ve Cengiz, 2009). Doğrusal regresyon modellerinin genelleştirilmiş doğrusal modellerin uzantısı olması gibi genelleştirilmiş toplamsal modeller de toplamsal modellerin bir uzantısıdır. Bu tür modellerin veriye uyumu için kullanılan değişik yaklaşımlardan biri olan düzleştirme, model matrisinde çeşitli düzleştirme fonksiyonlarının kullanıldığı yaklaşımlardandır (Asfha, 2017).

Ağırlıklı en küçük kareler kullanılarak oluşturulan ve kesin bir çözüme sahip olan doğrusal regresyon modellerinin aksine, bir GTM (veya bir GDM) için tahmin yöntemi, optimal tahminleri bulmak için yinelemeli (iteratif) yaklaşımlar gerektirir. GTM'de tahmin, yerel puanlama (local scoring) algoritması ile geriye uydurma (backfitting) algoritmasının bir bileşimine dayanır. Yerel puanlama algoritması, GDM'de maksimum olabilirlik tahminlerini bulmak için Fisher puanlama prosedürünün bir genellemesidir. Geri uydurma algoritması, herhangi bir toplamsal modeline uydurmak için uygundur ve GTM'de, modele birkaç düzgün fonksiyon dahil edildiğinde yerel puanlama yinelemesinde kullanılır. GTM, mevsimsellik, eğilimler ve hava değişkenlerinin doğrusal olmayan karıştırıcı etkileri için parametrik olmayan düzeltmelere izin verdiği için en yaygın olarak uygulanan yöntem olmuştur ve parametrik alternatiflere göre daha esnek bir yaklaşımdır. GTM, birçok zaman serisi analizinde yaygın olarak kullanılmaktadır (Dominici, ve ark., 2002).

Literatür incelendiğinde; Omay (2007)'ın çalışmasında Los Angeles'daki hava kirliliği ile kaydedilen ölüm oranları arasındaki ilişki incelenmiştir. Gözlenen ölüm oranlarına karşılık gelen ozon seviyeleri, karbon monoksit seviyeleri, partikül seviyeleri, ortalama sıcaklık ve zaman ölüm üzerine etkili açıklayıcı değişken olarak belirlenmiştir. Hava kirliliğinin ölüm oranını nasıl etkilediğini test etmek için farklı regresyon modelleri kullanılmış. Modeller sırasıyla doğrusal model, GDM, toplamsal model, GTM ve ince tabakalı splayn ile GTM'dir. Elde edilen sonuçlar incelendiğinde genelleştirilmiş toplamsal modelin istatistiksel olarak daha anlamlı sonuçlar verdiği ve semiparametrik toplamsal modele göre daha açıklayıcı olduğu görülmüştür.

Savaş ve Cengiz (2009) çalışmalarında zeka testleri yapılan bir merkezde 118 kişiye yapılan anket çalışmasında kişinin hayat memnuniyet derecesi üzerine etkisini incelemek için yaş, zeka testi, eğitim, aylık gelir, ağırlık gibi değişkenlere ikili lojistik regresyon uygulamışlar ve sonucunda yaşın anlamlı olmadığı görülmüştür. Bu ilişkiyi daha az sınırlı hale getirmek için GTM kullanılmış ve Ki-kare test değerine bakıldığında %5 anlamlılık seviyesinde gelirin etkisi anlamsız çıkmıştır. Yapılan analizlerden elde edilen sonuçta zeka, ağırlık ve yaş etkileri kuadratik alınırken gelirin etkisi doğrusal olarak tekrar test edilmiş ve parametreler %5 anlamlılık seviyesinde anlamlı bulunmuştur.

Ma ve Yan (2014)'ın çalışmasında, doğrusal model durumunda anlamlı bulunmayan sürücü yaşı değişkeninin trafik kazalarında çarpan ya da çarpılan taraf olma üzerine etkisinin olup olmadığı durumu, kübik splayn düzeltmesi kullanılarak toplamsal lojistik regresyon model ile incelemişlerdir. İnceleme sırasında sürücü yaşı değişkeni anlamlı bir etkiye sahip olup, sürücü yaşı küçüldükçe çarpan (hatalı) taraf olma olasılığının arttığı gözlemlenmiştir.

Kan Kılınc ve Çavuş (2017) yapmış oldukları çalışmada, çoklu bağlantının varlığında, genelleştirilmiş doğrusal model, toplamsal lojistik regresyon model ve sınıflandırma ağaçlarının bir benzetim çalışması ile küçük, orta ve büyük veri kümeleri için modellerin yeterliliğini, bahsedilen yöntemler yardımıyla ortaya koymayı amaçlamışlardır. Sonuç olarak toplamsal modellerden elde edilen sonuçların CART modellerinde düzeltme fonksiyonlarının etkisiyle başarılı sonuçlar elde etmiş olduklarını bildirmişlerdir.

Kovalchik ve Varadhan (2013) çalışmasında toplamsal lojistik modellerin genişletilmiş bir hali olan toplamsal binom modellerini kullanarak mesane kanseri üzerine bir veri seti kullanarak R istatistik programının "blm" paketinin kullanımını göstermişlerdir.

Omay (2014)'ın çalışmasında Dünya Bankası, Dünya Kalkınma Göstergeleri'nden alınan verilerle çevresel göstergeler olarak CO₂ ve PM10 emisyonları, ekonomik göstergeler olarak kişi başına düşen GSYH (Gayri safi yıllık hasıla) ve enerji kullanımı, sağlık göstergesi olarak ise kadınlar ve erkeklere ilişkin ölüm oranları alınmıştır. Bu değişkenler dikkate alınarak regresyon modelleri

oluşturulmuş ve OECD (Ekonomik Kalkınma ve İşbirliği Örgütü)'ye üye ülkeler kullanılmıştır. Yapılan çalışmada OECD ülkeleri için PM10, CO₂ çevresel göstergelerin ve enerji kullanımı, GSYH ekonomik göstergelerin sağlık üzerindeki etkisi modellenmiş ve farklı varyanslılık problemi gözlenmiş bu ortadan kaldırmak için panel genelleştirilmiş toplamsal model (PGAM) kullanılarak problemin aşıldığı görülmüştür.

Şenel ve ark. (2009)'ın Samsun il merkezinde hava kirliliğine neden olan SO₂ ölçümlerindeki değişimin nispi nem, ortalama rüzgar hızı, ortalama sıcaklık, düşen yağış miktarı gibi değişkenler ile ilişkisini ortaya koymak için çoklu doğrusal regresyon modeli ve GAM karşılaştırılmıştır. Çoklu doğrusal regresyon modelinde SO₂'yi açıklamada ortalama nispi nem ve aylık yağış miktarı anlamsız olduğu görülmüştür. GAM kullanılarak açıklayıcı değişkenler arasındaki ilişkinin ortalama rüzgar hızı ve ortalama sıcaklığın doğrusal, ortalama nispi nem ve aylık yağışın kuadratik olduğu ve %5 anlamlılık seviyesinde tüm açıklayıcıların istatistiksel olarak anlamlı olduğu belirlenmiştir.

Dominici ve ark. (2002) hava kirliliği ve sağlık ilişkisini inceledikleri çalışma sonucunda analiz bulgularının, epidemiyolojik veya diğer araştırmalarda GTM ve diğer parametrik olmayan regresyon tekniklerinin faydasını ortaya koymuştur. Çalışmalarında tanımlanan sorunlara karşı önlem almak için, yakınsama kriterleri, önemli ölçüde daha güvenilir hale getirilmesi gerektiğini bildirmişlerdir. Analizleri gerçekleştirmek için S-Plus yazılımında "gam" fonksiyonunu kullandıklarını belirtmişlerdir.

Bu çalışmada genelleştirilmiş toplamsal modellerin tanıtılması ve zootekni alanında bir örnek ile doğrusal regresyon analizi ile karşılaştırılması amaçlanmıştır.

2. Materyal ve Yöntem

2.1. Materyal

Bu çalışmada veri olarak 2014 yılında tamamlanan bir doktora tezi için alınan veriler (Tozlu Çelik, 2014) danışman onayı alınarak kullanılmıştır. Sonuç değişkeni olarak süt keçilerinden alınan laktasyon süt verimi (kg) alınırken açıklayıcı değişkenler olarak laktasyon süresi (gün) ve ana yaşı (yıl) kullanılmıştır. Verilerin analizinde SAS yazılımı kullanılmıştır.

2.2. Yöntem

Temel amaçlar ise en genel biçimde, sonuç değişkenindeki değişimi açıklama, herhangi bir gözleme karşılık gelen ortalama y değerini bulma, noktalar için en iyi eğriyi uydurma olarak sıralanabilir (Bağdathı, 2010). Geleneksel çoklu doğrusal model (Eşitlik 1);

$$E(Y|X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (1)$$

olarak gösterilebilir (Arı ve Önder, 2013; Kan Kılınc ve Çavuş, 2017). Bu model oldukça basit olmasına rağmen, varsayımlarının sağlanması durumunda çok güvenilir bir yöntemdir. Doğrusal regresyon modelleri için açıklayıcı değişkenlerin birbirlerinden bağımsız olduğu ve benzer

dağılıma sahip oldukları varsayımı kabullenilmektedir. Parametrik regresyon analizinin bir diğer varsayımı ise; regresyon katsayısının test edilmek istenmesi durumunda hata terimlerinin normal dağılıma sahip olmasıdır. Bu varsayımlar altında regresyon katsayısının EKK tahmin edicisi $n-k$ serbestlik dereceli t dağılımına sahiptir (burada, n ve k sırasıyla örnek büyüklüğü ve açıklayıcı değişken sayısını göstermektedir). Ayrıca, merkezi limit teoreminden bilindiği gibi regresyon katsayısı optimum örnek büyüklüğüne ulaşıldığında normal dağılıma sahip olmaktadır. Ancak, hata terimlerinin normal dağılıma sahip olmadığı ve örnek büyüklüğünün yeterli düzeye ulaşmadığı durumlarda, EKK tahmin edicisi hatalı sonuçlar üretebilmektedir (Önder, 2007). Biyolojik örneklerin genellikle doğrusal olmaması nedeniyle, daha esnek bir yaklaşım olan istatistiksel yöntemlerden parametrik olmayan yöntemler ve geliştirilmiş toplamsal modeller alternatif olarak kullanılabilir (Hastie ve Tibshirani, 1990; Kan Kılınç ve Çavuş, 2017). Parametrik olmayan çoklu regresyon modellerinin tahmininde çok boyutluluğun yarattığı güçlükler (curse of dimensionality) nedeniyle zorluklar yaşanabilmektedir. Bu sorunun çözümü için toplamsal modeller kullanılabilir (Savaş ve Cengiz, 2009).

Genelleştirilmiş toplamsal modeller (GTM) Eşitlik 2'de olduğu gibi gösterilebilir (Hastie ve Tibshirani, 1990; Savaş ve Cengiz, 2009);

$$E(Y|X_1, X_2, \dots, X_k) = f_1(X_1) + \dots + f_k(X_k) \quad (2)$$

- X_i : Bağımsız değişken
- Y : Bağımlı değişken
- f_i : Düzleştirici fonksiyonlar (parametrik olmayan) gösterir.

Burada X_i 'nin rastgele seçilen fonksiyon için terimler $f_i(X_i)$, doğrusal denklemdeki $\beta_i X_i$ terimlerine karşılık gelmektedir. Buradaki yaklaşımın farklılığı, her fonksiyona serpilme diyagramı düzleştiricisinin uydurulması ve bu şekilde her f_i fonksiyonunu tahmin edilmesini sağlamaktır (Bağdatlı, 2010). Bu modeller, bağımlı değişkenin ortalamasının, doğrusal olmayan bir bağlantı işlevi aracılığıyla toplamsal bir tahmin ediciye bağlı olduğunu varsayar. Genelleştirilmiş toplamsal modeller GTM'i doğrusal $\sum_i \beta_i X_i$ ifadesini $\sum_i f_i(X_i)$ ile ikame ederek genişletir. GTM'deki lineer regresyon adımı, verilerin uygun düzgün fonksiyonu f belirlemek için kullanıldığı, parametrik olmayan bir eklemeli regresyon adımı ile değiştirilir. Bu, yinelemeli (iteratif) düzleştirme işlemleri aracılığıyla yapılır ve açıklayıcı değişkenlerin çeşitli doğrusal olmayan etkilerine izin verir (Berg, 2007).

Geleneksel bağlantı fonksiyonları ve ilgili genelleştirilmiş toplamsal modeller şu şekildedir;

- $g(\mu) = \mu = \sum_j f_j(x_j)$, geleneksel doğrusal modelin toplamsal genişlemesi

- $g(\mu) = \log(\mu) = \sum_j f_j(x_j)$, poisson için logaritmik toplamsal model
- $\log\left[\frac{\mu}{1-\mu}\right] = \alpha + f_1(x_1) + \dots + f_p(x_p)$ toplamsal logistik regresyon modeli

Yukarıda verilen ifadeler negatif binom ve gamma dağılımını içeren üssel ailesinden örnek modellerinden oluşturulabilir. Bunlar genelleştirilmiş toplamsal modeller ile uyumlu şekilde genişletilmiş doğrusal modeller olarak bilinen sınıfı oluştururlar (McCullagh ve Nelder, 1989).

f_i fonksiyonları bir algoritma için temel yapı bloklarını meydana getiren serpilme diyagramını kullanarak daha esnek bir şekilde tahmin edilebilir. x_j 'nin etkisinde tahmin edilen $\hat{f}_j(x_j)$ fonksiyonu genellikle doğrusal olmayan durumlar ortaya koyabilir. Burada sürekli ve bazı değişkenlerin kesikli olduğu durumlar için diğer parametrik durumlarda da kolaylıkla kıyaslanabilir olduğunu göstermeyi amaçlamaktadır. Doğrusal olmayan ifadeler ana etkilerle sınıflandırılmamaktadır. İki ya da daha fazla değişken için doğrusal olmayan bileşenler mevcut olabilir ya da kesikli faktörün her seviye düzeyinde eğrilerin kategorize edilebilir özellikte olabilir. Sonuç olarak sınıflandırma amacıyla aşağıda verilenlerden biri kullanılabilir;

- $g(\mu) = X\beta + \alpha_k + f(z)$: Semi-parametrik bir model; burada X doğrusal olarak modellenirken tahmin ediciler için bir faktör olan, α_k , kesikli bir faktörün k 'nci derecesi için etki ve tahmin edici Z 'nin etkisi non-parametrik şekilde modellenir.
- $g(\mu) = f(x) + g_k(z)$: Burada, k bir faktörü için seviye sayısı ve sonuçta k ve Z 'nin bir etkisi için bir etkileşim ifadesini ortaya koyar.
- $g(\mu) = f(x) + g(z, w)$: Burada g iki değişken için non-parametrik bir fonksiyon olarak tanımlanabilir.

Genel doğrusal olmayan modeller Eşitlik 3'deki gibi gösterilir;

$$Y = (X_1, X_2, \dots, X_p) + \varepsilon \quad (3)$$

- X_1, X_2, \dots, X_p : Açıklayıcı değişken
- ε : rastgele hata
- f : $R^p \rightarrow R$ keyfi seçilen düzgün bir fonksiyon

Ancak $p > 2$ olduğu durumlarda doğrusal olmayan model için uyumsuzluk gösterir. Ayrıca açıklayıcı değişken sayısının fazla olması doğrusal olmayan regresyonu yorumlamayı zorlaştırır. Toplamsal modelde; f_j 'yi tahmin etmek için kullanılan X_j 'nin rastgele seçilen fonksiyonuna ait terimler olan $f_j(X_j)$ doğrusal denklemdeki $\beta_j X_j$ 'nin yer değiştirmesidir (Eşitlik 4).

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \varepsilon \quad (4)$$

Bağımlı değişken Y , bağımsız değişkenleri X_1, X_2, \dots, X_p 'ye kadar olan toplamsal model, yukarıdaki forma sahiptir. Hatalar burada X_j 'den bağımsız, $E(\varepsilon) = 0$ ve $\text{Var}(\varepsilon) = \sigma^2$ 'dir (Eşitlik 5).

$$E(Y|X_1, X_2, \dots, X_p) = \sum_{j=1}^p f_j(X_j) \quad (5)$$

$E(Y)$ üzerinde X_j 'nin etkisi diğer açıklayıcı değişkenlerin her değeri için benzerdir. Y bir bağımlı değişken olsun ve X_1, X_2, \dots, X_p açıklayıcı değişkenler olarak tanımlandığında standart doğrusal regresyon modeli Eşitlik 6'daki şekilde;

$$E(Y|X_1, \dots, X_p) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (6)$$

koşullu beklentisi için doğrusal bir form önerir. $\beta_1, \beta_2, \dots, \beta_p$ 'nin tahminleri EKK yöntemi ile elde edilir. Toplamsal modeller (Eşitlik 7);

$$E(Y|X_1, \dots, X_p) = s_0 + s_1(X_1) + \dots + s_p(X_p) \quad (7)$$

de verilen koşullu beklenen değer modellemesi ile doğrusal modelleri genelleştirir. Burada $i=1, \dots, p$ olmak üzere $s_i(X_i)$ düzleştirme fonksiyonlarıdır. Düzleştirme fonksiyonları s_i 'lerin $E s_i(X_i) = 0$ standartlaştırılmış koşulları sağlaması gerekir.

2.3. Düzleştirici Fonksiyonlar

Genelleştirilmiş toplamsal modellerde düzleştirici fonksiyonlar, bağımlı değişkendeki değişkenliği bağımsız değişkenin bir fonksiyonuyla açıklar. Düzleştirici fonksiyonlar splayn olarak da bilinir. Splaynlar sonuç değişkeni ile bağımsız değişkenler arasındaki ilişkiyi açıklar. Farklı düzleştiriciler vardır burada yaygın kullanılan düzleştiriciler verilecektir.

2.3.1. Kübik splaynlar

Basit şekilde farklı kübik polinomlardan oluşan bir eğri olarak tanımlanabilir (Eşitlik 8).

$$\sum_{i=1}^n (n(x_i))^2 + \lambda \int_a^b (n''(t))^2 dt \quad (8)$$

λ bir sabit ve $a \leq x_1 \leq \dots \leq x_n \leq b$ 'dir. İlk terim veriye yakınlığı ölçer, ikinci terim fonksiyondaki eğriyi düzeltir. $\lambda/(1+\lambda)$ düzleştirme parametrisidir ve λ büyüdükçe düzleştirme parametresi 1'e yaklaşır ve daha düzgün eğri oluşturur. λ 'nın küçük değeri için düzleştirme parametresi 0'a yaklaşır.

2.3.2. Yerel regresyon

Yerel regresyon Cleveland ve ark. (1988) tarafından ortaya konmuştur. Yerel regresyon, bir x açıklayıcı değişkeninde regresyon fonksiyonu $\eta(x)$ 'in bir fonksiyonunun değeriyle hesaplanacağı fikrine dayanır. Böyle bir yerel yaklaşım, x noktasının seçilen bir komşuluğu içinde veri noktalarına bir regresyon yüzeyi düzleştirme ile elde edilir. Ağırlıklı En Küçük Kareler (EKK) algoritması açıklayıcı değişkenlerin doğrusal fonksiyonlarını düzleştirme için kullanılır. Yerel regresyon süreci için düzleştirme parametresi tahmin edilen eğrinin düzgünlüğünü kontrol eder (Savaş ve Cengiz, 2009).

2.3.3. Thin-plane zincir (TPSPALYN) prosedürü

Thin-Plane Zincir Prosedürü (TPSPALYN) düzeltilmiş EKK metodunu kullanarak parametrik olmayan

regresyon modeli uydurur. Thin-Plane düzleştirme zincirini kullanır. Modelin bir parametrik formu için Genelleştirilmiş Çapraz Geçerlilik (GCV) kullanılır. Regresyon model tanımlanmış ve model parametre sayısı bilindiğinde kolaylıkla kullanıldığı gibi model hakkında bir bilgi yokken, veri sabit sayıda parametrelili bir modelle belirtilmiyorsa veriyi modellemek için de kullanılabilir.

Temel özellikleri;

- Düzeltmiş en küçük kareler tahminlerini verir.
- Çok boyutlu veri kullanımını sağlar.
- Çoklu skor ifadelerini verir.
- Hem semiparametrik modelleri hem de non-parametrik modelleri düzleştirir.
- Geniş veri setlerini yönetme de seçenekler sunar.
- Çoklu bağlantılı değişkenleri kullanabilir.
- Model serbestlik derecesini veya düzleştirme parametresinin belirlenmesi ile özel bir modelin seçilmesine imkan sağlar.

Farz edelim ki H_m m'inci dereceden kısmi türevleri $L_2(E_d)$ 'de olan fonksiyonların bir uzayı olsun. Burada E^d x 'in tanım kümesidir. Model (Eşitlik 9),

$$y_i = f(x_{1(i)}, \dots, x_{d(i)}) + \varepsilon_i, \quad i=1, \dots, n \text{ dir.} \quad (9)$$

burada $f \in H_m$ dir. Sabit bir λ için (Eşitlik 10)

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i) - z_i \beta) + \lambda J_m(f) \quad (10)$$

ile verilen düzeltilmiş en küçük kareler fonksiyonunun minimum yapılması ile f 'yi tahmin etmek mümkündür. Burada $J_m(f)$ 'yi tanımlamak için birkaç tane yol vardır. TP düzleştirme zinciri için d boyutlu x için $J_m(f)$ (Eşitlik 11),

$$J_m(f) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \sum_{a_1! \dots a_d!} \frac{m!}{a_1! \dots a_d!} \left[\frac{\partial^m f}{\partial x_1^{a_1} \dots \partial x_d^{a_d}} \right]^2 dx_1 \dots dx_d \quad (11)$$

gibi tanımlanır. Burada $\sum_i a_i = m$ 'dir. $d=2$ ve $m=2$ için $J_m(f)$ (Eşitlik 12),

$$J_2(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\left[\frac{\partial^2 f}{\partial x_1^2} \right]^2 + 2 \left[\frac{\partial^2 f}{\partial x_1 \partial x_2} \right]^2 + \left[\frac{\partial^2 f}{\partial x_2^2} \right]^2 \right) dx_1 dx_2 \quad (12)$$

deki gibidir.

Genelde m ve d , $2m-d > 0$ koşulunu sağlamalıdır. Kolaylaştırmak için $m=2$ varsayımı kabul edilir (Wahba 1990, Bates ve ark., 1987).

2.4. Düzleştirme Parametresinin Seçimi

2.4.1. Çapraz geçerlilik (CV) ve genelleştirilmiş çapraz geçerlilik (GCV)

Düzleştirme parametresinin seçiminde CV kullanılabilir. $(x_j y_j)$ miktardan her seferinde biri dışarıda tutularak geriye kalan $n-1$ noktadan x_i 'deki düzleştirme fonksiyonu için artık karelerin tahmini elde edilir ve bu artık karelerin toplamını minimum yapacak düzleştirici seçilir. CV fonksiyonu (Eşitlik 13),

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - n_{\lambda}^{-i}(x_i))^2 \quad (13)$$

olarak kullanılır. Buradaki $n_{\lambda}^{-i}(x_i)$, x_i deki uyumu gösterir. $nCV(\lambda)$ tahmin kareler toplamı olarak da adlandırılır. GTM prosedürü tarafından uydurulan bütün düzleştiriciler örnek bağımlı değişkenlerin bir doğrusal kombinasyonu olarak Eşitlik 14'deki şekilde formüle edilebilir (Allen, 1974).

$$n(x) = A(\lambda)Y \quad (14)$$

burada $A(\lambda)$, λ' ya bağlı matristir. $A(\lambda)$ ' nin diagonal elemanları a_{ii} ile gösterilirse CV fonksiyonu (Eşitlik 15),

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - n_{\lambda}(x_i)}{1 - a_{ii}} \right)^2 \quad (15)$$

deki gibi ifade edilebilir. Çoğu durumlarda a_{ii} 'yi hesaplamak oldukça zaman alıcıdır. Bu hesaplama problemini çözmek için Wahba (1980, 1990) GCV fonksiyonunu önermiştir. GVC fonksiyonu (Eşitlik 16),

$$GVC(\lambda) = \frac{n \sum_{i=1}^n (y_i - n_{\lambda}(x_i))^2}{(n - iz(A(\lambda)))^2} \quad (16)$$

deki gibi ifade edilebilir. GCV, CV' nin ağırlıklı oranıdır. Çoğu durumlarda GCV, CV' ye yakındır. Ancak hesaplaması daha kolaydır (Savaş ve Cengiz, 2009). Kesikli bağımlı değişkenlerin modellenmesi amacıyla normal dağılım uygun değildir. GTM bahsedilen sorunların aşılması için yardımcı olabilecek bir yöntemdir. Pek çok dağılım için toplamsal modeller uygulanabilir (SAS Institute Inc., 1999a, 1999b, 1999c, 2004).

GTM ve GDM aynı amaçla kullanılabilir. Fakat farklı çözümlenme amaçlarına hitap ederler. GDM model parametreleri için tahmine önem verirken GTM veriyi non-parametrik olarak açıklamayı amaçlar. GTM açıklayıcı değişkenler ile sonuç değişkeni arasındaki ilişkinin ortaya konulmasında ve verinin açıklanmasında daha uygun bir yöntemdir (Savaş ve Cengiz, 2009).

2.5. GTM Nasıl Yorumlanmalı?

GTM için yazılan modelde regresyon parametreleri β_j 'lerin bulunmaması GTM' in önemli özelliklerinden birini yansıtır. GTM kullanıldığında regresyon parametreleri kümesi elde edilemez. Fakat buna rağmen, $\hat{f}_j(X_j)$ olarak yazılan her X_j değeri için $f_j(X_j)$ tahminleri vardır. Her $f_j(X_j)$ için katsayı model oluştururken belirlenir ve dahası $\hat{f}_j(X_j)$, X_j ve sonuç değişkeni arasındaki ilişkiyi gösterir. Çoğu uygulamada standart regresyon uygulamasında olduğu gibi bazı tahmin edicilerin doğrusal etkileri olduğu varsayılır.

Grafiksel yöntemler GTM' in non-parametrik bileşenlerini yorumlamak için kullanılır. X_j ve $\hat{f}_j(X_j)$ arasındaki grafik

X_j ve sonuç değişkeni arasındaki ilişkinin tahmin edilen doğrusal olmayan yapısını gösterir. Standart hatalar ve güven aralıkları $\hat{f}_j(X_j)$ hesaplanabilir ve grafik çizilebilir. Böylece X_j 'nin yükselmesi ve azalması ya da doğrusal uyumundan elde edilen fonksiyonun uyumluluğu değerlendirilebilir. Regresyon katsayıları tablosunun incelenmesi \hat{f}_j grafiğine göre daha basit görünse de bu kullanım kolaylığı muhtemel olarak güvenilir olmayan, kısıtlayıcı ve gereksiz doğrusallık varsayımları karşılığında elde edilir. Diğer yandan eklemelilik GTM'ın tahmin edicilerinin her birinin etkisinin doğrusal regresyonda olduğu gibi diğer tahmin edicilerin etkileriyle oluşan ağırlık yorumlanmasına olanak sağlar (Beck ve Jackman, 1998).

2.6. Analiz için Kullanılan SAS Kodları

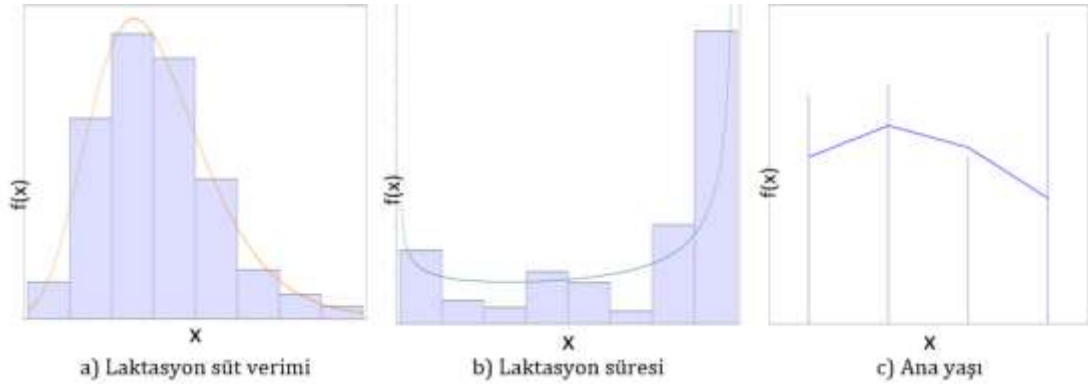
```
analiz 'tez';
data tez;
input lsv ay ls;
logCP = log(lsv);
datalines;
272.60 2 235
217.60 2 235
231.00 2 231

301.5 5 201
179.40 5 195
146.60 5 229
;
ods graphics on;
proc gam data=tez;
model logCP = spline(ay) spline(ls);
run;
```

3. Bulgular ve Tartışma

Sonuç değişkeni olarak laktasyon süt verimi (kg) alınırken açıklayıcı değişkenler olarak laktasyon süresi (gün) ve ana yaşı (yıl) değişkenleri alınmıştır. Bu değişkenlerin dağılışı EasyFit yazılımı ile belirlenmiş olup Şekil 1'de verilmiştir.

Laktasyon süt veriminin $\sigma=0,24582$, $\mu=5,7345$ ve $\gamma=-99,815$ parametreleri ile lognormal (3P) dağılışı gösterdiği belirlenmiştir. Kolmogorov-Smirnov normallik testi sonucunda LSV değişkeninin normal dağılışa uygun olmadığı ($P=0,038$) belirlenmiştir. laktasyon süresinin $\alpha_1=0,68651$, $\alpha_2=0,34817$, $a=187,0$ ve $b=235,0$ parametreleri ile Beta dağılışı gösterdiği belirlenmiştir. Kolmogorov-Smirnov normallik testi sonucunda LSV değişkeninin normal dağılışa uygun olmadığı ($P<0,0001$) belirlenmiştir. Ana yaşının $\lambda=3,5602$ parametresi ile Poisson dağılışı gösterdiği belirlenmiştir. AY değişkeninin normal dağılışa uygun olmadığı ($P<0,0001$) belirlenmiştir (Şekil 1).



Şekil 1. İncelenen değişkenlere ait dağılım grafikleri.

Doğrusal regresyon analizi ve parametre tahminleri Tablo 1’de verilmiştir. Elde edilen doğrusal regresyon modelinin istatistiksel olarak anlamlı olduğu ($P < 0,001$), Hata Kareler Ortalaması (HKO) 5826,039 ve belirtme katsayısının (R^2) 0,16 olduğu belirlenmiştir. Düşük bir belirtme katsayısına rağmen modelin anlamlı olmasının örnek büyüklüğüne bağlı olabileceği ($n=166$) düşünülmektedir (Bishara ve Hittner, 2012).

Laktasyon süt verimini(kg) tahmin etmek için laktasyon süresi(gün) ve ana yaşı değişkenleri kullanılarak elde edilen genelleştirilmiş toplamsal model için iterasyon özeti ve uyum istatistikleri; Backfitting Iterasyon sayısı=5, Final Backfitting Kriteri= $2,271223E-10$ ve sonuç tahminine ait sapma= $20,128886164$ olarak elde edilmiş olup iterasyon sonucunda uyum sağlandı (convergence) bilgisine ulaşılmıştır.

GTM modeli için düzeltirici parametre tahminleri ve bunlara karşılık gelen genelleştirilmiş çapraz geçerlilik değerleri Tablo 2’de ve GTM analiz istatistikleri Tablo 3’de verilmiştir.

Ana yaşı değişkeninin laktasyon süresi değişkenine göre daha başarılı düzeltirme değerine sahip olduğu söylenebilir çünkü sıfıra yakın değerler düzeltirme parametresinin bir fonksiyonu olarak GCV’nin başarılı bir şekilde en aza indirilmesini sağlamak ve modele uyumu artırmak için oldukça önemlidir (Jansen, 2015).

GTM sonuçlarının en önemli kısmı olarak ele alınabilecek olan bilgilere sapma analizi tablosundan ulaşılabilir. Modeldeki düzeltiriciler için tablo tam model ve

ilgilenilen değişken yokluğunda oluşturulan model arasındaki sapmayı karşılaştıran bir χ^2 test değeri üretmektedir. Elde edilen sonuçlara göre 0,05 önem düzeyinde ana yaşı ve laktasyon süresi değişkenlerinin model üzerindeki etkisinin istatistiksel olarak anlamlı olduğu anlaşılmıştır.

Ana yaşı ve laktasyon süresine ait düzeltirme grafikleri Şekil 2’de verilmiştir.

Şekil 2’de verilen grafikler yorumlandığında ana yaşı ve laktasyon süresi değişkenlerinin genelleştirilmiş toplamsal modellerle değerlendirilmesi gerektiği anlaşılmaktadır.

Doğrusal regresyon ve genelleştirilmiş toplamsal modelleri karşılaştırabilmek adına Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE) ve ortalama mutlak sapma (MAD) değerleri hesaplanmış ve Tablo 4’de verilmiştir.

Elde edilen bulgulara göre doğrusal regresyon analizinin tahmin başarısının daha yüksek olduğu Tablo 4’deki düşük karşılaştırma değerlerinden anlaşılmaktadır. Ancak açıklayıcı değişkenler arasında çoklu bağlantı problemi olmamasına ($VIF=1,000$) ve sapmaların normal dağılışa sahip olmasına ($P > 0,05$) rağmen değişkenlerin normal dağılmadığı bilinmektedir. Doğrusal regresyon analizinin gerekli varsayımları sağlayamadığından sonuçların güvenilir olmadığı bu nedenle Genelleştirilmiş Toplamsal Modellerin kullanılmasının önerilebileceği söylenebilir.

Tablo 1. Doğrusal regresyon analizi ve parametre tahminleri

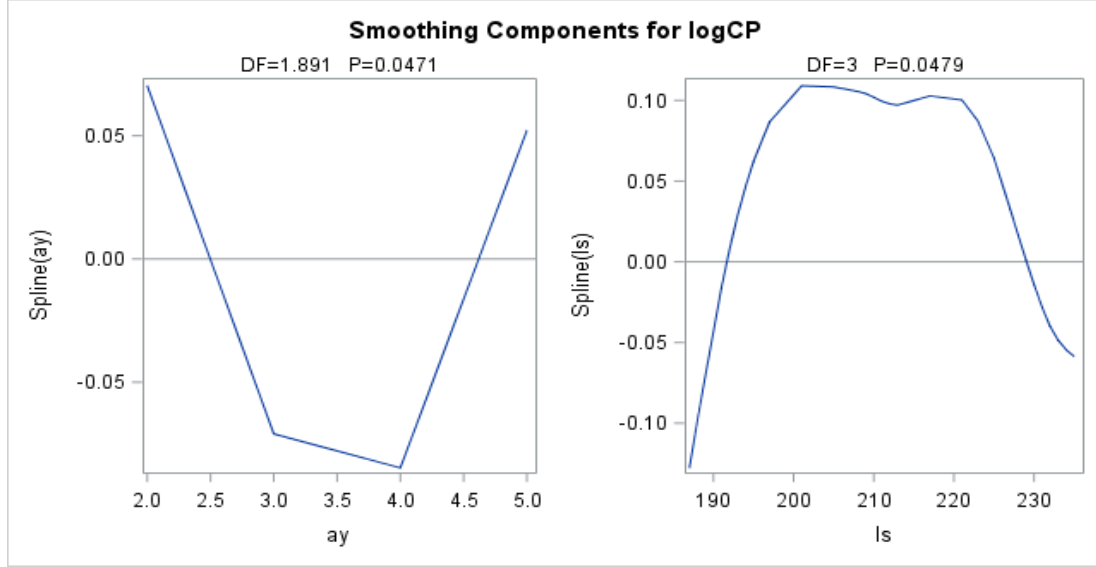
Parametreler	Parametre Tahmini	Standart Hata	t değeri	P
Sabit	-157,322	84,686	-1,858	0,065
Ana yaşı	21,589	5,068	4,26	<0,0001
Laktasyon süresi	1,364	0,374	3,647	<0,0001

Tablo 2. GTM modeli için düzeltirme bileşenleri ve uyum istatistikleri

Bileşen	Düzteltirme Parametresi	Serbestlik Derecesi	GCV	Tekil gözlem sayısı
Spline (ana yaşı)	0.009901	1.890859	0.001077	4
Spline (laktasyon süresi)	0.994764	3.000000	0.099842	30

Tablo 3. GTM analiz istatistikleri (sapma analizi)

Bileşen	Serbestlik Derecesi	Kareler Toplamı	Ki-Kare	P
Spline (ana yaşı)	1.89086	0.750597	5.8958	0.0471
Spline (laktasyon süresi)	3.00000	1.006934	7.9093	0.0479



Şekil 2. Ana yaşı ve laktasyon süresine ait düzleştirme grafikleri.

Tablo 4. Modeller için karşılaştırma istatistikleri

Model	RMSE	rRMSE	MAD
LR	75,6357	34,4214	58,1024
GTM	79,9291	36,3753	61,9424

4. Sonuç

Toplamsal modeller doğrusal regresyon modellerine göre daha kısıtsız modellerdir ancak parametrik olmayan regresyon modellerinden daha kısıtlı modellerdir. Toplamsal modeller açıklayıcı değişkenler arasındaki ilişkiyi dikkate almamaktadır. Bu durum toplamsal modeli parametrik olmayan modellerden daha fazla kısıtlı duruma getirmektedir. Toplamsal modellerin genel gösteriminde açıklayıcı değişkenler arasındaki ilişki dikkate alınmaz ancak bu durum açıklayıcı değişkenler arasında ilişki olmadığı şeklinde yorumlanmamalıdır. Açıklayıcı değişkenler arasında ilişki yok ise, her açıklayıcı değişkenin sonuç değişkenine karşı düzeltilmesiyle elde edilen fonksiyonların toplamı toplamsal modelin tahminini oluşturur. Açıklayıcı değişkenler arasında ilişki olduğu durumda ise modelin tahmininde bu ilişkinin de dikkate alınması gerekir. Bu durumda kısmi regresyonlar yardımıyla hesaplanan kısmi artıklar kullanılabilir (Çağlayan, 2002; Bağdatlı, 2010).

Genelleştirilmiş toplamsal modeller doğrusal regresyon modellerine göre daha kısıtsız modellerdir ancak parametrik olmayan regresyon modellerinden daha kısıtlı modellerdir. Açıklayıcı değişkenlerin normal dağılım göstermediği durumlarda genelleştirilmiş toplamsal modellerin doğrusal modeller yerine

kullanılmasının daha doğru ve güvenilir sonuçlar üretebileceği yapılan bu çalışma ile literatürü destekler nitelikte ortaya konulmuştur.

Katkı Oranı Beyanı

Tüm yazarlar eşit oranda katkıya sahiptir. Tüm yazarlar makaleyi inceledi ve onayladı.

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Etik Onay

Bu çalışmada hayvan materyali kullanılmadığından etik izin gerekli değildir. Çalışmada 2020 yılı öncesinde tamamlanan bir doktora tezi için alınan veriler danışman onayı alınarak kullanılmıştır (Tozlu Çelik, 2014).

Destek ve Teşekkür Beyanı

Makalede değerlendirilen verilerin kullanılmasına izin veren Prof. Dr. Mustafa OLFAZ ve Dr. Öğr. Üyesi Hilal TOZLU ÇELİK'e teşekkür ederiz. Bu çalışma birinci yazarın yüksek lisans tezinden özetlenmiştir.

Kaynaklar

- Allen DM. 1974. The relationship between variable selection and data augmentation and a method of prediction. *Technometrics*, 16: 125-127.
- Arı A, Önder H. 2013. Farklı veri yapılarında kullanılabilir regresyon yöntemleri. *Anadolu J Agr Sci*, 28(3): 168-174.
- Asfha HD. 2017. Performance of spline-based gam in the presence of outliers and multicollinearity. *Yüksek Lisans Tezi, Eskişehir Anadolu Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir, Türkiye*, pp: 73.
- Bağdath S. 2010. Semiparametrik regresyon ve bir uygulama. *Yüksek Lisans Tezi, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye*, pp: 81.
- Bates DM, Lindstrom MJ, Wahba G, Yandell BS. 1987. GCVPACK Routines for generalized cross-validation. *Commun Statist Simul*, 16: 263-97.
- Beck N, Jackman S. 1998. Beyond Linearity by default: generalized additive models. *American J Polit Sci*, 42(2): 596-627.
- Berg D. 2007. Bankruptcy prediction by generalized additive models. *Appl. Stochastic Models Bus Ind*, 23: 129-143.
- Bishara AJ, Hittner JB. 2012. Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychol Meth*, 17(3): 399-417.
- Cleveland WS, Devlin SJ, Grosse EH. 1988. Regression by local fitting methods, properties and computational algorithms. *J Economet*, 37: 87-114.
- Çağlayan E. 2002. Yarı parametrik regresyon modelleri ile yasadışı gelir hipotezinin Türkiye uygulaması. *Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, Türkiye*, pp: 171.
- Dominici F, McDermott A, Zeger SL, Samet JM. 2002. On the use of generalized additive models in time-series studies of air pollution and health. *American J Epidemiol*, 156(3): 193-203.
- Durmuş S. 2018. Değişken katsayılı regresyon modeli ile gebelerin ortalama arteriyel kan basıncına etki eden risk faktörlerinin belirlenmesi. *Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, Türkiye*, pp: 62.
- Hastie TJ, Tibshirani RJ. 1990. *Generalized additive models*. Chapman & Hall New York, US, pp: 352.
- Jansen M. 2015. Generalized Cross Validation in variable selection with and without shrinkage. URL: <http://homepages.ulb.ac.be/~majansen/publications/jansen15gcvreprint.pdf> (erişim tarihi: 14 Kasım 2021).
- Kan Kılıç B, Çavuş M. 2017. İkili yanıt değişkenine sahip modellerin yeterliliklerine ilişkin benzetim çalışması - parametrik olmayan yöntemler. *Sakarya Üniv Fen Bil Enst Derg*, 21(2): 169-177.
- Kant R, Baykan UN. 2004. Bina yaklaşık maliyetinin çoklu doğrusal regresyon ile belirlenmesi. *Politeknik Derg*, 7(4): 359-367.
- Kovalchik S, Varadhan R. 2013. Fitting additive binomial regression models with the R package blm. *J Stat Software*, 54(1): 1-18.
- Ma L, Yan X. 2014. Examining the nonparametric effect of drivers' age in rear-end accidents through an additive logistic regression model. *Accident Anal Prevent*, 67: 129-136.
- McCullagh P, Nelder JA. 1989. *Generalized linear models*. Chapman & Hall, London, UK, pp: 532.
- Omay RE. 2007. Regresyonda pürüzlülük ceza yaklaşımı. *Doktora Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir, Türkiye*, pp: 121.
- Omay RE. 2014. OECD ülkeleri için hava kirliliği-sağlık ilişkisinin incelenmesi. *Hava Kirl Araş Derg*, 3: 39-46.
- Önder H. 2007. Permütasyon testlerinin doğrusal regresyonda kullanılabilirliğinin irdelenmesi. *OMÜ Zir Fak Derg*, 22(2): 157-161.
- SAS Institute Inc. 1999a. *SAS language reference: Concepts, Version 8*, Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 1999b. *SAS language reference: Dictionary, Version 8*, Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 1999c. *SAS procedures guide, Version 8*, Cary, NC: SAS Ins.Inc.
- SAS Institute Inc. 2004. *SAS/STAT 9.1 User's guide*, Cary, NC: SAS Institute Inc.
- Savaş N, Cengiz MA. 2009. Yaşam memnuniyetini etkileyen faktörlerin belirlenmesinde geliştirilmiş toplamsal modellerin kullanımı. *EÜFBED*, 2(1): 105-119.
- Şenel T, Cengiz MA, Savaş N, Terzi N. 2009. Çoklu doğrusal regresyonda model seçiminde geliştirilmiş toplamsal modellerin kullanımı. *EÜFBED*, 2(2): 217-227.
- Tozlu Çelik H. 2014. Kıl keçi ve Saanen x Kıl keçi melezlerinin (F1, G1, G2) üretici şartlarında çeşitli verim özellikleri bakımından karşılaştırılması. *Doktora Tezi, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun, Türkiye*, pp: 103.
- Wahba G. 1980. *Spline bases, regularization and generalized cross-validation for solving approximation problems with large quantities of noisy data*. Academic Press, New York, US, pp: 14.
- Wahba G. 1990. *Spline functions for observational data*. SIAM, Philadelphia, US, pp: 161.



AN EVALUATION OF SUPERMARKETS FROM THE LENS OF MULTIPLE CRITERIA: THE INTUITIONISTIC FUZZY TOPSIS METHOD

Feride TUĞRUL^{1*}

¹Kahramanmaraş Sütçü İmam University, Faculty of Science and Arts, Department of Mathematics, 46050, Kahramanmaraş, Türkiye

Abstract: This study is an attempt to evaluate supermarkets by both the consumer and the market researcher through the intuitionistic fuzzy TOPSIS. In this regard, the decision makers' views were expressed in linguistic terms and ranked from the most preferred supermarket to the least preferred supermarket. The satisfaction of using linguistic terms is that it gives decision makers the chance to express views that they cannot express in numerical values. Furthermore, undecided situations were interpreted through the IFS. This method, which may be easily used in many application areas, offers researchers more consistent results than the others due to its advantages. Besides, this study attempts to innovative implementation of the intuitionistic fuzzy TOPSIS method on its algorithm and mathematical basis.

Keywords: Intuitionistic fuzzy set, Market selection, Intuitionistic fuzzy TOPSIS

*Corresponding author: Kahramanmaraş Sütçü İmam University, Faculty of Science and Arts, Department of Mathematics, 46050, Kahramanmaraş, Türkiye

E mail: feridetugrul@gmail.com (F. TUĞRUL)

Feride TUĞRUL  <https://orcid.org/0000-0001-7690-8080>

Received: August 04, 2022

Accepted: September 07, 2022

Published: October 01, 2022

Cite as: Tuğrul F. 2022. An evaluation of supermarkets from the lens of multiple criteria: the intuitionistic fuzzy TOPSIS method. BJS Eng Sci, 5(4): 146-150.

1. Introduction

Fuzzy logic was described by Zadeh (1965), furthermore Atanassov described intuitionistic fuzzy (IF) sets (Atanassov, 2016). Recently, there are many MCDM methods defined and have been the focus of attention for many researchers (Majumder, 2015). TOPSIS method makes a ranking built on the positive ideal and negative ideal relationship (Hwang and Yoon, 1981). Besides IF TOPSIS method is chosen since decision makers are free to express their ideas in linguistic terms. Numerous researchers have benefited from the TOPSIS method in their application areas such as supplier selection, location selection, renewable energy technologies, mobile phone selection, product concept selection, wind power plants, etc. (Boran et al., 2009; Boran, 2011; Rouyendegh, 2011; Boran et al., 2012; Rouyendegh and Saputro, 2014; Büyüközkan and Güleriyüz, 2015; Efe et al., 2015; Rouyendegh, 2015; Damgacı et al., 2017; Rouyendegh et al., 2018; Rouyendegh et al., 2020). Today, the issue of market selection that meets the basic needs of people has been the focus of the attention of researchers. There are studies conducted with different methods in the related literature such as: Market segment evaluation, e-marketplace selection, the competitiveness of supermarket chains, market selection in international expansion, the consumer market for business, opinion of supermarket executives, market segment evaluation with CODAS method, VIKOR method with research on the fresh fruit-vegetable sector, market

evaluation using by AHP and COPRAS-G method, etc. (Büyükoğuzkan, 2004; Gorecka and Szałucka, 2013; Tosun, 2015; Ghorabae et al., 2017). The intuitionistic fuzzy sets and MCDM methods were utilized by researchers in many decision making processes and applications, and very favorable results were obtained since the sensitivity makes sense (Çuvalcıoğlu, 2014; Tuğrul and Çitil, 2021; Sözeyatarlar et al., 2021; Yavuz and Şahin, 2022).

2. Materials and Methods

2.1. Material

This study utilized the 3 most preferred chain supermarkets and 1 chain local market in Turkey. These supermarkets were evaluated according to certain criteria by consumers and market researchers. From many aspects, supermarkets have a very significant role in people's lives. Chain supermarkets, on the other hand, are designed to serve the needs of people.

Alternatives and explanations about them are as follows:

- M_1 : Local Market -has 7 stores and approximately 150 employees
- M_2 : Supermarket Chain-has approximately 11000 stores and approximately 61000 employees
- M_3 : Supermarket Chain-has approximately 2500 stores and approximately 50000 employees
- M_4 : Supermarket Chain-has approximately 8500 stores and approximately 60000 employees

Criteria (Super Market) and explanations about them are



as follows:

- MC_1 : Pricing
 - Greengrocer
 - Meat and meat products
 - Milk and milk products
 - Legumes
- MC_2 : Location
 - Accessibility
 - Number of stores
- MC_3 : Expiry dates of products
- MC_4 : Wide product range
- MC_5 : Freshness of products and quality of products
- MC_6 : Campaigns
 - Campaign application
 - Advertising on social media
- MC_7 : Parking facility
- MC_8 : Working principle of the staff
 - The friendliness of the staff
 - The helpfulness of the staff
 - The sufficient number of staff
- MC_9 : Clean and comfortable atmosphere
- MC_{10} : Clutter-free, organized shelf-compartments
- MC_{11} : Image of the market
- MC_{12} : Easy access to the product in the market
 - Guide member
 - Guide signs
 - Lighting
 - Arrows
- MC_{13} : Product variety
- MC_{14} : Having a children's play area

2.2. Methods

Definition 1: (Atanassov, 1986; 2016) (Equation 1) Let $X \neq \emptyset$. An IFS A in X ;

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \} \tag{1}$$

The algorithm of the IF TOPSIS method was shown in this section (Rouyendegh, 2015).

Step 1: By specifying the contribution interest of the DMs on the practice, the DM importance value was determined in linguistic terms (Wu et al., 2016).

Table 1. Expressions for DMs

Expressions	IFNs
VI	(0.8,0.1)
I	(0.5,0.2)
M	(0.5,0.5)
B	(0.3,0.5)
VB	(0.2,0.7)

The table above represents I important, VI very I, M medium, B bad, VB very bad. The value of DMs views

$Dl = [\mu l, \nu l, \pi l]$ was calculated as follows (Equation 2):

$$\lambda l = \frac{[\mu l + \pi l (\frac{\mu l}{\mu l + \nu l})]}{\sum_{l=1}^k [\mu l + \pi l (\frac{\mu l}{\mu l + \nu l})]} \tag{2}$$

where, $\lambda l \in [0, 1]$ and $\sum_{l=1}^k \lambda l = 1$.

Step 2: DMs indicated their opinions on the criteria in linguistic terms based on Table 2.

Table 2. Expressions for criterion evaluation

Expressions	IFNs
VI	(0.9,0.1)
I	(0.75,0.2)
M	(0.5,0.45)
U	(0.35,0.6)
VU	(0.1,0.9)

Table 2, unlike Table 1, represents U is unimportant, VU is very U. Afterwards, while calculating the criterion weight, and the IFWA operator defined by Xu (2007) was used (Equation 3).

$$w_j = IFWA_{\lambda} (w_j^{(1)}, w_j^{(2)}, \dots, w_j^{(k)}) = \lambda_1 w_j^{(1)} \oplus \lambda_2 w_j^{(2)} \oplus \dots, \lambda_k w_j^{(k)} \tag{3}$$

Step 3: With the help of Table 3, the views of the DMs about the alternatives were determined.

Table 3. Expressions for alternative assessment

Expressions	IFNs
VG	(1.00,0.00)
G	(0.85,0.05)
MG	(0.70,0.20)
F	(0.50,0.50)
MP	(0.40,0.50)
P	(0.25,0.60)
VP	(0.00,0.90)

With these values, Intuitionistic Fuzzy Decision Matrix (IFDM) was acquired. Table 3 indicates that VG is very G, G is good, MG is medium G, F is fair, P is poor, MP is medium P and VP is very P. Aggregated IFDM was acquired with the help of the IFWA operator (Equation 4).

$R^{(l)} = (r_{ij}^{(l)})_{m \times n}$ is the IFDM.

$$R = (r_{ij})_{m \times n} \tag{4}$$

$$r_{ij} = IFWA_{\lambda} (r_{ij}^{(1)}, r_{ij}^{(2)}, \dots, r_{ij}^{(k)}) = \lambda_1 r_{ij}^{(1)} \oplus \lambda_2 r_{ij}^{(2)} \oplus \dots, \lambda_k r_{ij}^{(k)}$$

Step 4: S matrix was determined as seen in Equation 5.

$$S = R \otimes W \tag{5}$$

Step 5 The ideal solution means close to positive and far from negative. A^+ (Equation 6) and A^- (Equation 7) are formed in which J_1 : benefit and J_2 : cost criteria:

$$A^+ = (r_1^+, r_2^+, \dots, r_n^+), r_j^+ = (\mu_j^+, v_j^+, \pi_j^+), j = 1, 2, \dots, n \quad (6)$$

$$A^- = (r_1^-, r_2^-, \dots, r_n^-), r_j^- = (\mu_j^-, v_j^-, \pi_j^-), j = 1, 2, \dots, n \quad (7)$$

where;

$$\mu_j^+ = \{(\max\{\mu_{ij}\} | j \in J_1), (\min\{\mu_{ij}\} | j \in J_2)\} \quad (8)$$

$$v_j^+ = \{(\min\{v_{ij}\} | j \in J_1), (\max\{v_{ij}\} | j \in J_2)\} \quad (9)$$

$$\mu_j^- = \{(\min\{\mu_{ij}\} | j \in J_1), (\max\{\mu_{ij}\} | j \in J_2)\} \quad (10)$$

$$v_j^- = \{(\max\{v_{ij}\} | j \in J_1), (\min\{v_{ij}\} | j \in J_2)\} \quad (11)$$

Step 6: Researchers utilized many different methods to calculate separation measures (Szmidt and Kacprzyk, 2000; Çitil, 2019). This study varies across other research since the normalized Hamming distance was used. Separation measures, S_i^+ and S_i^- respectively, were gained as in Equations 12 and 13:

$$S_i^+ = \frac{1}{2n} \sum_{i=1}^n [|\mu_{ij} - \mu_{ij}^+| + |v_{ij} - v_{ij}^+| + |\pi_{ij} - \pi_{ij}^+|] \quad (12)$$

$$S_i^- = \frac{1}{2n} \sum_{i=1}^n [|\mu_{ij} - \mu_{ij}^-| + |v_{ij} - v_{ij}^-| + |\pi_{ij} - \pi_{ij}^-|] \quad (13)$$

Step 7: C_i^* was calculated by the Equation 14:

$$C_i^* = \frac{S_i^-}{S_i^+ + S_i^-}, \text{ and } 0 \leq C_i^* \leq 1 \quad (14)$$

A higher value of C_i^* represents the better alternative.

3. Results

For this study, the views of 3 decision makers were taken into account. DM_1 : Market researcher, DM_2 and DM_3 : Consumer. After specifying the importance of DMs with linguistic variables according to Table 1, it was converted into numerical data with the help of Equation (2) and was shown in Table 4 as follows:

Table 4. Importance of DMs

DM_1	DM_2	DM_3
VI	I	I
λ_1	λ_2	λ_3
0.383561644	0.308219178	0.308219178

Decision makers specified when determining the weights for criteria as follows:

Table 5. Weights of Criteria

Criteria	DM_1	DM_2	DM_3
MC_1	VI	VI	U
MC_2	M	I	U
MC_3	VI	VI	VI
MC_4	VI	I	I
MC_5	VI	M	VI
MC_6	I	I	M
MC_7	I	U	VI
MC_8	I	VI	I
MC_9	VI	U	VI
MC_{10}	VI	M	VI
MC_{11}	VI	U	VI
MC_{12}	VI	I	I
MC_{13}	VI	I	I
MC_{14}	I	U	VI

The views of the DMs were given in Table 6-7-8. For each criterion, decision makers evaluated all alternatives separately.

Table 6. Values of alternatives for criteria according to decision maker 1

DM_1	M_1	M_2	M_3	M_4
MC_1	G	MG	MP	G
MC_2	MG	VG	MP	G
MC_3	VG	VG	VG	G
MC_4	G	MG	MG	G
MC_5	VG	MP	G	P
MC_6	P	VG	G	VG
MC_7	MP	G	MP	G
MC_8	G	MP	MG	MP
MC_9	MG	G	VG	G
MC_{10}	P	MG	G	MG
MC_{11}	G	MG	VG	MG
MC_{12}	P	MG	G	G
MC_{13}	VG	MG	VG	G
MC_{14}	VG	P	VG	P

Table 7. Values of alternatives for criteria according to decision maker 2

DM_2	M_1	M_2	M_3	M_4
MC_1	MG	MG	MP	VG
MC_2	MG	VG	P	VG
MC_3	VG	VG	VG	G
MC_4	MG	VG	MG	VG
MC_5	VG	MG	MG	MG
MC_6	P	G	MG	VG
MC_7	MP	G	MP	G
MC_8	MG	MP	MG	MP
MC_9	MP	G	MG	G
MC_{10}	P	P	F	P
MC_{11}	MG	VG	MG	VG
MC_{12}	P	MG	G	G
MC_{13}	VG	MG	F	VG
MC_{14}	F	F	F	F

Table 8. Values of alternatives for criteria according to decision maker 3

DM_3	M_1	M_2	M_3	M_4
MC_1	VG	VG	G	VG
MC_2	MG	VG	MG	VG
MC_3	VG	VG	VG	G
MC_4	G	MG	VG	MG
MC_5	VG	MP	VG	MP
MC_6	P	G	VG	G
MC_7	MP	G	MP	G
MC_8	MG	MP	G	MP
MC_9	MP	G	VG	G
MC_{10}	MG	MG	G	MG
MC_{11}	G	MG	VG	MG
MC_{12}	P	G	VG	G
MC_{13}	VG	F	VG	F
MC_{14}	VG	P	VG	P

R matrix was created with the help of Equation (4). Afterwards, the S matrix was obtained using Equation (5). A^+ and A^- solutions were calculated using Equations (6) and (7) and shown in Table 9.

Table 9. The if positive and negative ideal solution

Criteria	A^+	A^-
MC_1	(0.8219,0.1737)	(0.5003,0.3769)
MC_2	(0.5622,0.3829)	(0.2704,0.6290)
MC_3	(0.9000,0.1000)	(0.7650,0.1450)
MC_4	(0.8241,0.1533)	(0.6710,0.2182)
MC_5	(0.8358,0.1589)	(0.3946,0.4989)
MC_6	(0.6905,0.2568)	(0.1726,0.7027)
MC_7	(0.6349,0.2653)	(0.2988,0.6133)
MC_8	(0.6249,0.2601)	(0.3246,0.5808)
MC_9	(0.8219,0.1737)	(0.4439,0.4644)
MC_{10}	(0.6541,0.2445)	(0.3632,0.5186)
MC_{11}	(0.8219,0.1737)	(0.6693,0.2371)
MC_{12}	(0.8241,0.1533)	(0.2060,0.6613)
MC_{13}	(0.8241,0.1533)	(0.5347,0.3779)
MC_{14}	(0.7469,0.2266)	(0.2526,0.6653)

S^+ and S^- for alternatives determined using Equation (12) and (13) using the normalized Hamming measure are shown in Table 10, respectively. The closeness coefficient values were calculated by using the Equation (14) and are shown in Table 10. In addition, the graphs of the C_i^* are shown in Figure 1.

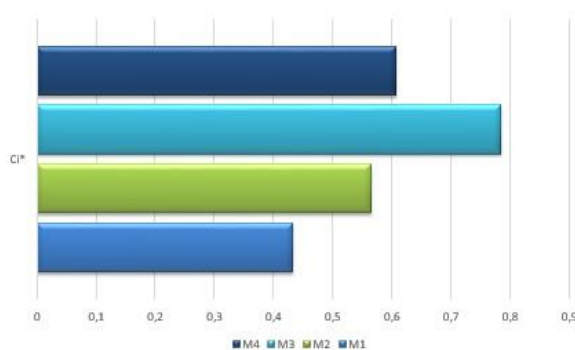


Figure 1. Closeness coefficient values

Table 10. Values of separation measures and closeness coefficient

	S^+	S^-	C_i^*
M_1	0.18685	0.14246	0.43261
M_2	0.14167	0.18463	0.56584
M_3	0.06945	0.25203	0.78395
M_4	0.12779	0.19758	0.60725

High C_i^* value means that that alternative is preferred more. According to Figure 1 and closeness coefficient values, the order of the supermarkets from the most preferred to the least preferred in terms of consumers is as follows: M_3, M_4, M_2 and M_1 .

4. Discussion and Conclusion

This study investigated the markets preferred by the consumers with the help of the IF-based TOPSIS method. While evaluating the markets, it has been handled from the perspective of both market researchers and consumers. Expressing their ideas in linguistic terms provided a freer decision making action for decision makers. Thanks to intuitionistic fuzzy sets, the study aimed to get the most efficient result by making sense of the sensitivities. The reason for using the TOPSIS is that the DMs may easily state their ideas with the help of linguistic terms, not numerical values. The study will offer various approaches the literature in many markets and marketing fields. At the same time, it will guide researchers in the field of application. The number of consumers in the study may be increased by changing, different decision makers may be added, and the number of supermarkets may be increased. This application may be developed in different supermarkets and by other methods over different criteria. By establishing a wide-ranging decision-making mechanism thanks to this method, it may be ensured that the sector across the country and the world may meet customer demands quickly and facilitate their workload.

Author Contributions

All task was done by the single author: F.T. (100%). The author reviewed and approved the manuscript.

Conflict of Interest

The author declared that there is no conflict of interest.

Acknowledgements

The author declared that has not received any financial support for the research, authorship or publication of this study.

References

- Atanassov KT. 1986. Intuitionistic fuzzy set. Fuzzy Sets Syst, 20(1): 87-96.
- Atanassov KT. 2016. Intuitionistic fuzzy sets. Int J Bioautomation, 20(S1): 1-6.
- Boran FE, Boran K, Menlik T. 2012. The evaluation of renewable

- energy technologies for electricity generation in turkey using intuitionistic fuzzy TOPSIS. *Energy Sour Part B: Econ Plan Pol*, 7(1): 81-90.
- Boran FE, Genç S, Kurt M, Akay D. 2009. A Multi-Criteria Intuitionistic fuzzy group decision making for supplier selection with TOPSIS Method. *Expert Syst App*, 36(8): 11363-11368.
- Boran FE. 2011. An integrated intuitionistic fuzzy multi criteria decision making method for facility location selection. *Math Comp App*, 16(2): 487-496.
- Büyükközkcan G, Güleriyüz S. 2015. An application of intuitionistic fuzzy topsis on mobile phone selection. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, August 2-5, 2015, Rockville, US, pp: 1-8.
- Büyükközkcan G. 2004. Multi-Criteria decision making for e-marketplace selection. *Internet Res*, 14(2): 139-154.
- Çitil M. 2019. Application of the Intuitionistic fuzzy logic in education. *Commun Math App*, 10(1): 131-143.
- Çuvalcıoğlu G. 2014. Some properties of controlled set theory. *Notes on Intuitionistic Fuzzy Set*, 20(2): 37-42.
- Damgacı E, Boran K, Boran FE. 2017. Evaluation of Turkey's renewable energy using intuitionistic fuzzy TOPSIS method. *J Polytech*, 20(3): 629-637.
- Efe B, Boran FE, Kurt M. 2015. Ergonomic product concept selection using intuitionistic fuzzy TOPSIS. *Suleyman Demirel Univ J Eng Sci Design*, 3(3): 433-440.
- Ghorabae MK, Amiri M, Zavadskas EK, Hooshmand R, Antucheviciene J. 2017. Fuzzy extension of the CODAS method for multi-criteria market segment evaluation. *J Busin Econ Manag*, 18(1): 1-19.
- Gorecka D, Szałucka M. 2013. Country market selection in international expansion using multicriteria decision aiding methods. *Mult CritDec Making*, 8: 32-55.
- Hwang CL, Yoon K. 1981. Methods for multiple attribute decision making. In: *Multiple Attribute Decision Making*. Springer, Heidelberg, Berlin, Germany, pp: 58-191.
- Majumder M. 2017. Multi criteria decision making. Chapter 2, Springer, Heidelberg, Berlin, Germany, pp: 35-47.
- Rouyendegh BD, Saputro TE. 2014. Supplier selection using integrated fuzzy TOPSIS and MCGP: a case study. *Procedia-Soc Behav Sci*, 116: 3957-3970.
- Rouyendegh BD, Yıldızbaşı A, Arıkan ÜZ. 2018. Using intuitionistic fuzzy TOPSIS in site selection of wind power plants in Turkey. *Adv Fuzzy Sys*, 2018: 6703798. DOI: 10.1155/2018/6703798.
- Rouyendegh BD, Yıldızbaşı A, Üstünyer P. 2020. Intuitionistic fuzzy TOPSIS method for green supplier selection problem. *Soft Comp*, 24(3): 2215-2228.
- Rouyendegh BD. 2011. The DEA and intuitionistic fuzzy TOPSIS approach to departments' performances: a pilot study. *J App Math*, 2011: 712194. DOI: 10.1155/2011/712194.
- Rouyendegh BD. 2015. Developing an integrated ANP and intuitionistic fuzzy TOPSIS model for supplier selection. *J Test Eval*, 43(3): 664-672.
- Sözeyatarlar M, Şahin M, Yavuz E. 2021. Statistical relations measures. *J Universal Math*, 4(2): 283-295.
- Szmidt E, Kacprzyk J. 2000. Distances between intuitionistic fuzzy sets. *Fuzzy Sets Syst*, 114: 505-518.
- Tosun N. 2017. Target market selection in fresh fruit-vegetable sector using fuzzy VIKOR method. *J Manag Mark Logis*, 4(4): 465-471.
- Tuğrul F, Çitil M. 2021. A new perspective on evaluation system in education with intuitionistic fuzzy logic and PROMETHEE algorithm. *J Univ Math*, 4(1): 13-24.
- Wu Y, Zhang J, Yuan J, Geng S, Zhang H. 2016. Study of decision framework of offshore wind power station site selection based on ELECTRE-III under intuitionistic fuzzy environment: a case of China. *Energy Convers Manag*, 113: 66-81.
- Xu ZS. 2007. Intuitionistic fuzzy aggregation operators. *IEEE Transact Fuzzy Sys*, 15: 1179-1187.
- Yavuz E, Şahin M. 2022. Semiparametric regression models and applicability in agriculture. *BSJ Agri*, 5(2): 160-166.
- Zadeh LA. 1965. Fuzzy Sets. *Info Cont*, 8: 1361-1375.



YORÙBÁ CHARACTER RECOGNITION SYSTEM USING CONVOLUTIONAL RECURRENT NEURAL NETWORK

Jumoke Falilat AJAO^{1*}, Shakirat Ronke YUSSUF¹, Abdulazeez Olorundare AJAO²

¹Kwara State University, Faculty of Information and Communication Technology, Department of Computer Science, +234, Ilorin, Nigeria

²Federal Polytechnic, School of Engineering, Department of Computer Engineering, +234, Offa, Nigeria

Abstract: Handwritten recognition systems enable automatic recognition of human handwritings, thereby increasing human-computer interaction. Despite enormous efforts in handwritten recognition, little progress has been made due to the variability of human handwriting, which presents numerous difficulties for machines to recognize. It was discovered that while tremendous progress has been made in handwritten recognition of English and Arabic languages, very little work has been done on Yorùbá handwritten characters. Those few works, in turn, made use of Hidden Markov Model (HMM), Support Vector Machine (SVM), Bayes theorem, and decision tree algorithms. To integrate and save one of Nigeria's indigenous languages from extinction, as well as to make Yorùbá documents accessible and available in the digital world, this research work was undertaken. The research presents a convolutional recurrent neural network (CRNN) for the recognition of Yorùbá handwritten characters. Data were collected from students of Kwara State University who were literate Yorùbá writers. The collected data were subjected to some level of preprocessing such as grayscale, binarization, and normalization in order to remove perturbations introduced during the digitization process. The convolutional recurrent neural network model was trained using the preprocessed images. The evaluation was conducted using the acquired Yorùbá characters, 87.5% of the acquired images were used for the training while 12.5% were used to evaluate the developed system. As there is currently no publicly available database of Yorùbá characters for validating Yorùbá recognition systems. The resulting recognition accuracy was 87.2% while the characters with under dot and diacritic signs has low recognition accuracy.

Keywords: Handwriting recognition, Yorùbá, Convolutional recurrent neural network, Document analysis, Deep learning

*Corresponding author: Kwara State University, Faculty of Information and Communication Technology, Department of Computer Science, +234, Ilorin, Nigeria

E mail: jumoke.ajao@kwasu.edu.ng (J. F. AJAO)

Jumoke Falilat AJAO



<https://orcid.org/0000-0002-4854-1901>

Shakirat Ronke YUSSUF



<https://orcid.org/0000-0001-6083-1541>

Abdulazeez Olorundare AJAO



<https://orcid.org/0000-0003-3136-6657>

Received: June 08, 2022

Accepted: September 08, 2022

Published: October 01, 2022

Cite as: Ajao JF, Yusuff SR, Ajao AO. 2022. Yorùbá character recognition system using convolutional recurrent neural network. BSJ Eng Sci, 5(4): 151-157.

1. Introduction

Optical character recognition is a subset of pattern recognition that employs a computer to recognize human handwriting. Handwriting recognition is one of the most fascinating and difficult research areas in image processing and pattern recognition, stemming from the need for humans to automate handwritten text recognition and enable computers to receive and interpret it (Mithe et al., 2013). Different writers have different writing styles; many people do not write distinct letters with clear spacing between them, making the characters more difficult to distinguish. Handwriting recognition is still considered an open research problem, owing to the wide range of individual differences in appearance. This issue has been extensively researched, and numerous viable solutions have been discovered (Chaudhari and Thakkar, 2019). Humans can read and write, as well as recognize artistic writings and graphics. Unconstrained handwriting recognition is still a difficult and critical problem for machines to solve (Graves et al., 2008). Aside from the usual problems in handwriting recognition, such as inter-writer abnormality and variation within-writer, character and word

segmentation, the presence of diacritic signs presents a significant challenge for Machine recognition (Darwish et al., 2021). Furthermore, the processing of scanned images of historical documents is hampered by variables such as poor image quality, difficult text-ink versus background separation, unknown styles for which no labelled training data sets exist, highly connected patterns that are difficult to segment horizontally and vertically, and ancient or peculiar language material for which no statistical dictionaries exist (Gatos et al., 2004; Likforman-Sulem et al., 2007; Nina et al., 2011). Numerous algorithms with varying degrees of precision have been developed to tackle these tasks (Chacko et al., 2012; Das and Behera, 2017). High precision performance on Latin, Arabic, Farsi, Chinese, and other languages has been achieved as a result of these efforts (Srihari et al., 2007; Srihari and Ball, 2012; Garoot et al., 2017). However, only a minor portion of Yorùbá language recognition accuracy has been achieved. The integration of information technology into the Yorùbá language is critical to preventing the language's extinction. Yorùbá handwriting technology can also be made available to the entire population. Yorùbá speakers



predominate in Nigeria's western region. Yorùbá mythology holds that all Yorùbá people are descended from a hero named Oodua or Oduduwa. The Yorùbá alphabet has 25 characters, 18 consonants, and 7 vowels. The vowel characters have diacritic signs such as acute, grave, and under dot, which are unique to Yorùbá alphabet. All twenty-five letters in the Yorùbá character set can be upper or lower case (Bamgbose, 2000; Ojo, 2007; Peel, 2009). The goal is to create an offline Yorùbá handwriting character recognition algorithm for recognizing handwritten characters in scanned documents and images using a Convolutional recurrent neural network. This paper is divided into six sections. The introduction is followed by related work. The third section depicts the proposed architecture, and the fourth section describes the system's methodology. The results of the experiments are shown in the following section of the paper, followed by a conclusion.

1.1. Related Works

Different solutions available in the literature to solve this problem have been reviewed. Asahiah (2014) developed, implemented, and evaluated a computational system for restoring missing diacritics in a digital text of the Standard Yoruba. Oladele et al. (2020) investigated offline Yorùbá handwritten word recognition system (OYHWR), which recognizes uppercase Yorùbá alphabets.

Various handwritten characters written by different writers were obtained using the paint application and M708 graphics tablets. Oni and Asahiah (2020) developed a Yorùbá character image recognition model using a dataset of scanned images of Standard Yorùbá printed text. The developed model was implemented and its performance was evaluated in order to create an Optical Character Recognition (OCR) model for Yorùbá printed text images. Dewa et al. (2018) developed a software that employs digital image processing methods and a convolution neural network module for offline handwritten Javanese character recognition in a computer-aided design environment.

On the basis of invariances such as translation, rotation, and scale, Oyedotun and Dimililer (2016) investigated and reviewed the performance of convolutional networks and their variant, convolutional auto encoder networks, when tasked with recognition problems involving translation, rotation, and scale. In the research, handwritten Yoruba vowel characters were used to represent the vowels.

Altwaijry and Al-Turaiki (2021) presented Hijja, a new dataset of Arabic letters written solely by children aged 7–12. The dataset includes 47,434 characters written by 591 people. Furthermore, the author proposed a model for automatic handwriting recognition based on convolutional neural networks (CNN). Hijja and the Arabic Handwritten Character Dataset (AHCD) were used to train the model. Wang et al. (2021) proposed a novel method for Chinese text recognition using cellphone-shot ID card images. It has two main contributions: A

conditional adversarial generative network designed to generate million-level synthetic ID card text line images, which preserve the diversity of synthetic data and also the inherent template pattern of ID card images. A more effective convolutional recurrent neural network (CRNN) is presented, with DenseNet using VGGNet as a replacement to extract more complex spatial features. Calvo-Zaragoza et al. (2019) proposed the use of convolutional recurrent neural networks, which have been shown to be effective in other similar domains such as handwritten text recognition. Ly et al. (2017) present a Deep Convolutional Recurrent Network (DCRN) model for recognizing offline handwritten Japanese text lines without the need for explicit character segmentation.

Bluche et al. (2014) conducted a comparison of Bidirectional LSTM-RNNs and Deep MLPs. The authors conducted experiments on two public databases of multi-line handwritten documents: Rimes and IAM, both of which are available online. They demonstrated that the proposed hybrid systems achieve performance comparable to that of the state-of-the-art regardless of the type of features used (hand-crafted or pixel values) or the optical model used (neural network) (Deep MLP or RNN)

In this study, we proposed Yoruba handwritten character Recognition system based on convolutional recurrent neural networks (CRNN).

2. Materials and Methods

2.1. Methodology

The method used for the research work are as follows:

- Data acquisition
- Image preprocessing
- CRNN module for classification

2.2. Dataset Acquisition

Data were collected from fifty literate indigenous writers who writes under a guided format. The peculiarities of Yorùbá alphabets are that, besides the twenty-five characters, there are additional variations of the vowel characters which are 11 total 36 such that, a vowel characters can carry underdot, grave accent, acute accent, and it can also combine both underdot with any of the diacritic signs which makes the total alphabet Thirty-six. The 36 characters that make up Yoruba characters were written by 50 people which amount to 1800 characters. This acquired images were digitized and replicated, since CRNN required large training dataset. Figure 1 shows samples of the digitized image.



Figure 1. Sample of digitized image.

2.3. Image Pre-processing

The data was manually labeled, and each character is made up of approximately 200 images, because deep learning requires large datasets for training. The collected sample data was replicated to form 400 images used for training. This is shown in Figure 2.

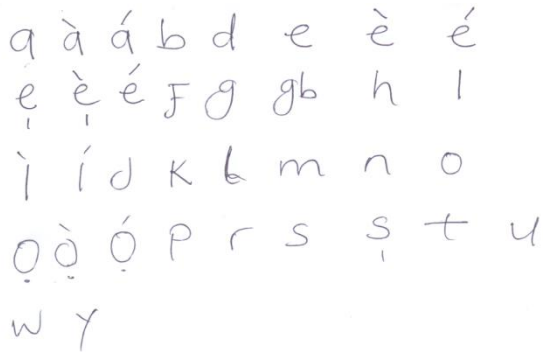


Figure 2. Sample of resized image.

2.4. RGB Image Channel Values

The acquired RGB images were preprocessed to Grayscale images each value of the grayscale value refers to a certain shade of gray with the absolute value referring to white and minimal representing black. The grayscale image was converted to binary image using Otsu method. Samples of the binarized image is shown in Figure 3.

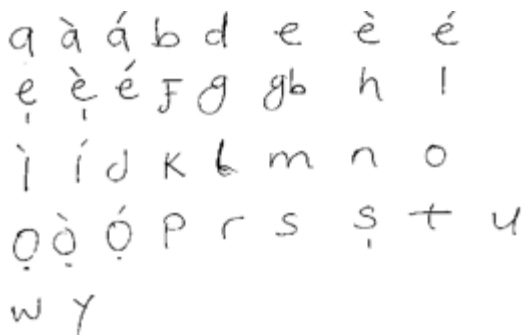


Figure 3. Binarized image

2.5. Image Normalization

In this research work, we used scale between 0 and 1 and this was achieved, by dividing our image by 255 which is the maximum possible value for any pixel.

2.6. Image Resizing

Most neural networks architectures used a fixed array shape, in our model the images were designed to use a 32x128 size image as input, so all the images were in that dimension. The implementation of the network was done using Tensor flow and keras. Figure 4 shows the model structure of CRNN. The CRNN model uses a convolutional neural network (CNN) to extract visual features, which are reshaped and fed to a long short term memory network (LSTM). The CNNLSTM were integrated using a sigmoid function through the maxpooling layer, this was passed with bias. The output of the LSTM is then mapped to character labels space with a Dense layer of softmax.

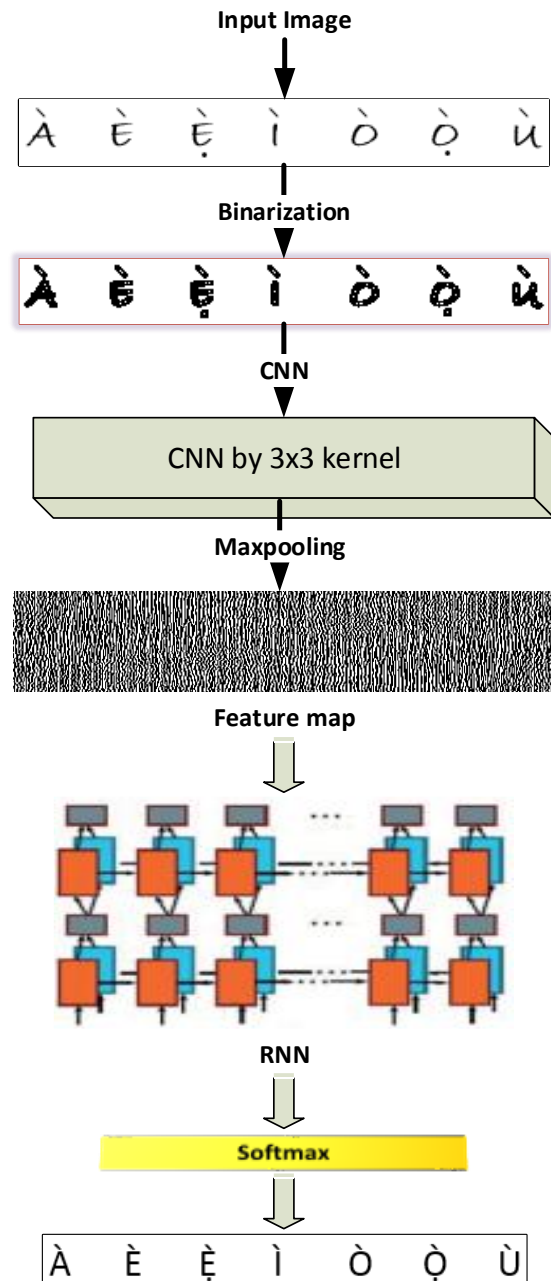


Figure 4. The CRNN architecture for Yorùbá character recognition.

Three input characters of the same character with different diacritic signs were supplied to the CNN with a filter of 3x3 for each of the characters to produce a feature sequence of size 16x16 that was also subjected to maxpooling to reduce the dimension of the data and to also reduce the training time and the testing time. After series of convolution and maxpooling of the feature vector, the final output of the maxpooling resulted into a reduced feature vectors which serve as input to the fully connected network of Recurrent Neural Network. The Convolution process is shown in Equation (1) to Equation (9)

Convolution Layer 1

$$C1_{i,j}^k = \left(\sum_{m=0}^4 \sum_{n=0}^4 w_{m,n}^k * I_{i+m,j+n} + b^k \right) \quad (1)$$

Where $C1^k$ is the feature map of the convolution layer, k is the filter number, m,n are the indices of the k^{th} filter. The training of the Convolution Neural Network was carried out as explained in equation 2.

Convolve 2: is a max-pooling layer. The output of the convolution layer $C1$ is fed to max-pooling layer. The max-pooling layer takes six features' maps from $C1$ and performs max-pooling operations on each of them. The max-pooling operations on $C1^k$ is given as $Q2 = \text{MaxPool}(C1^k)$. For each features map in $C1^k$, max-pooling performs the following operations:

$$Q2_{i,j}^k = \max \left(\begin{matrix} C1_{(2i,2j)}^k, C1_{(2i+1,2j)}^k \\ C1_{(2i,2j+1)}^k, C1_{(2i+1,2j+1)}^k \end{matrix} \right) \quad (2)$$

where (i,j) are the indices of k^{th} features map of output, and k is the features map index.

Convolve 3 is the second convolution layer which produces 16 features maps as given in Equation 3.

$$C3_{i,j}^k = \sigma \left(\sum_{d=0}^5 \sum_{m=0}^4 \sum_{n=0}^4 W_{m,n}^{k,d} * P2_{i+m,j+n}^d + b^k \right) \quad (3)$$

where $C3^k$ represent the 16 output features maps of convolution layer $C3$, k is the index of the output features map, (m,n) are the indices of filter weights, (i,j) are the indices of output, and d is the index of the number of channel in the input.

Convolve 4 is a max-pooling layer which produces 16 features maps $P4^k$. The max-pooling operations is given as $P4^k = \text{MaxPool}(C3^k)$

$$P4_{i,j}^k = \max \left(\begin{matrix} C3_{(2i,2j)}^k, C3_{(2i+1,2j)}^k \\ C3_{(2i,2j+1)}^k, C3_{(2i+1,2j+1)}^k \end{matrix} \right) \quad (4)$$

where (i,j) are the indices of k^{th} features map of output, and k is the features map index.

Convolve 5 is the third convolution layer that produces 120 output feature maps and is given by;

$$C5_{i,j}^k = \sigma \left(\sum_{d=0}^{15} \sum_{m=0}^4 \sum_{n=0}^4 W_{m,n}^{k,d} * P4_{i+m,j+n}^d + b^k \right) \quad (5)$$

where $C5^k$ represents the 120 output features maps of convolution layer $C5$ of size 1×1 , k is the index of the output features, (m,n) are the indices of filter weights. d is the index of the number of channels in input and (i,j) are the indices of output, since output is only 1×1 , the index (i,j) remains $(0,0)$ for each filter. This formular can be simplified as the filter size is equal to the size of input,so no convolution stride happens.

$$C6_{i,j}^k = \sigma \left(\sum_{d=0}^{15} \sum_{m=0}^4 \sum_{n=0}^4 W_{m,n}^{k,d} * P4_{m,j}^d + b^k \right) \quad (6)$$

Convolve 6 is the fully connected layer. It consists of neurons for different classes, which is defined in equation 6.

$$F6^k = \sum_{i=1}^{120} w_i^k * C5^i \quad (7)$$

The activation functions used is softmax function given as $Z^k = \text{softmax}(F6^k)$. The softmax function is defined as;

$$Z^k = \text{softmax}(F6^k) = \frac{e^{F6^k}}{\sum_{i=1}^{10} e^{F6^i}} \quad (8)$$

The softmax activation function produces the final output of the neurons in the range $[0,1]$ and all outputs add up to 1. Each of the output represents the probability of the input belonging to a particular class. Where Z^k is the vector of size containing final output of the network.

2.7. Loss Layer

The obtained output is compared against the actual output to compute the loss function. The calculated error is then used to update the weight of the network. The process is repeated until the error is minimized. The Mean Squared Error (MSE) of CNN model can be given as. $\text{Loss} = E(Z, \text{target})$. The loss function E is defined as;

$$E(Z, \text{target}) = \frac{1}{10} \sum_{k=1}^{10} (Z^k - \text{target}^k)^2 \quad (9)$$

where Z^k is the k^{th} output (in this case the networks will produce 10 outputs representing class probabilities) generated by the CNN and the *target* is the ground truth of the input. $E(Z, \text{target})$ is the error/loss which represent how far the prediction of the network is from actual target. The number of convolution layers, the nodes in the convolution layer, the layer layout, the neurons in the fully connected layers, the activation functions, and the loss function and its parameters are all optimized in our model for a fully connected network,

The data was into divided into two categories. The model was trained using 87 %of the data. The model was created using Google's open-source deep learning library tensor flow, and it contained approximately (400×36) character parameters, as well as convolutional, Maxpooling, LSTM, and Dense layers.

Softmax is constant along diagonals: this is the dimension that is removed, and it corresponds to the softmax output being independent of the translated input scores (a choice of 0 score). To normalize the scores, the number of input scores can be used. The primary advantage of recurrent neural networks is that they have a directional circle that can recollect previous data and apply it to the current output. The samples of convolution performed on

the preprocessed Yoruba image, the maxpooling to reduce the selected features, and the number of training and non-trainable parameters are shown in Table 1. On the output shape, several convolutional layers were applied, and the total number of trainable parameters was determined at each level. The reduced feature map was passed to the bidirectional recurrent neural network with LSTM at the end of the series of convolution and maxpooling to recall and produce the output characters. 1886 parameters were trained, while 114 could not be trained.

Table 1. The numbers of training parameters and non-trainable parameters with different maxpooling layers

Layer Type	Output shape	Parameters
Input Layer	32,128, 1	0
Con2d-1	32,128,64	640
Maxpooling_1	16, 64, 64	0
Con2d-1	16,128,128	1118
Maxpooling_2	8, 64, 64	0
Con2d_3	8,32,256	1566
Maxpooling_1	8, 32, 256	0
Trainable parameters	1886	
Non-trainable	114	

The Connectionist Temporal Categorical (CTC) loss function is the objective function used to minimize the loss. As the input image varies in nature, CTC loss is specially designed to optimize both the length of the predicted sequence and the classes of the predicted sequence. During the back-propagation procedure, the convolution filters and LSTM weights are jointly learned. The Adam optimizer was used for training, with initial learning rates ranging from 0.001 to 0.002 and 0.003 to 0.005. The module was passed to the CRNN module, which runs the ROI image through a forward pass with the pre-trained model, resulting in encoded values, each of which represents a character in the Yorùbá alphabet. The encoded values from the CRNN module are received by the decoder module and decoded into Yorùbá characters.

The steps involved for the recognition process;

- Acquired the digitized image
- Preprocessed the digitized image by converting it to grey scale and binary image
- Resize and normalize the preprocessed image into 32 × 128
- Convolve the normalized image using 3x3 kernel
- Perform maximum pooling to reduce the feature map
- The feature map was passed to RNN
- The output of the RNN was passed to softmax to produce the text equivalent of the recognized characters.

3. Results and Discussion

The performance of our proposed recognition system for Yorùbá alphabet, which consist of 36 lower case characters with variant diacritic signs and under dot was evaluated. This dataset contains more than 1800 images that are classified into training and testing classes. The proposed system was implemented by taking 87.5% of character images from the class of dataset as training data and 12.5% images for the class as testing dataset. The recognition accuracy was evaluated using CRRN as shown in Table 2. The ROC curve in Figure 5 depicting the trade-off between sensitivity and specificity indicates an excellent performance of the recognition system. The resulting recognition accuracy was 87.2% while the characters with under dot and diacritic signs has low recognition accuracy.

Table 2. Recognition accuracy of Yoruba character

Yorùbá characters	Training dataset	Testing dataset	Recognized	NR
a	350	50	50	0
à	350	50	50	0
á	350	50	50	0
b	350	50	48	2
d	350	50	45	5
e	350	50	50	0
è	350	50	46	4
é	350	50	47	3
ẹ	350	50	30	20
è	350	50	26	24
é	350	50	15	25
f	350	50	49	1
g	350	50	50	0
gb	350	50	35	15
h	350	50	48	2
i	350	50	50	0
î	350	50	47	3
í	350	50	43	7
j	350	50	48	2
k	350	50	50	0
l	350	50	46	4
m	350	50	49	1
n	350	50	49	1
o	350	50	50	0
ô	350	50	48	2
õ	350	50	46	4
ọ	350	50	25	25
ó	350	50	13	37
ò	350	50	20	30
p	350	50	49	1
r	350	50	50	0
s	350	50	50	0
ş	350	50	38	12
t	350	50	49	1
u	350	50	50	0
ù	350	50	48	2
ű	350	50	46	4
w	350	50	48	2
y	350	50	50	0

NR= not recognized

An effectiveness of the normalization along with the classification technique was further supported by the area under the ROC curve (AUC) of the classification performance of the recognition systems in Table3. The normalization yielded AUC of 95% confidence interval. The performance of our model was compared with other Yoruba character recognition system as our model shows a better improvement over machine learning approach.

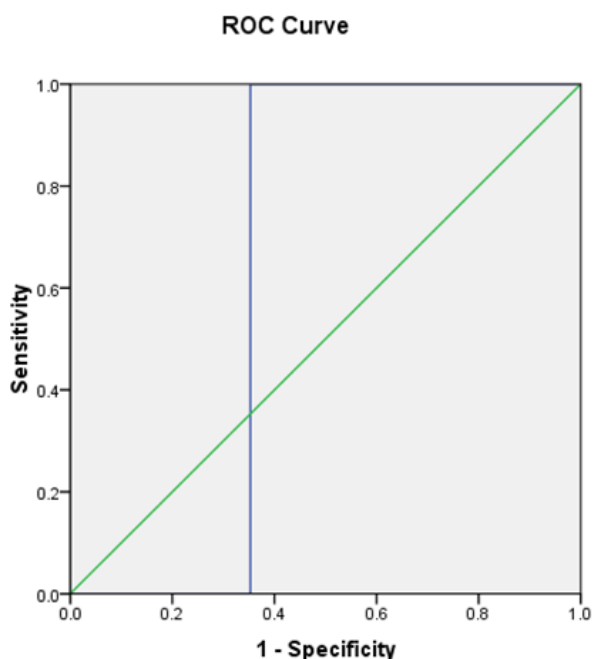


Figure 5. ROC of recognition accuracy of Yoruba character.

Table 3. Area under curve of the recognition accuracy

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.647	.082	.294	.486	.808

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Table 4. Performance evaluation of Yoruba character recognition using CRNN and other character recognition

Author	Classifiers	Database	RA
Oladele et al. (2017)	Support Vector Machine (SVM)	Yoruba Character database	76.7%
Mathias et al. (2020)	Geometric features and Support Vector Machine(SVM)	Yoruba character database	85.7%
Oyeniran and Oyebode (2021)	Transfer Learning Model(Alextnet)	Yoruba Character database	91.4%
Our Model	Convolutional Recurrent Neural Network(CRNN)	Yoruba character databse	87.2%

RA= recognition accuracy

4. Conclusion

Offline Yorùbá character recognition was implemented using CRNN, from the result, it was observed that characters that have peculiar features of Latin characters have 100% recognition accuracy. Those characters that carries unique features that are peculiar to Yorùbá characters have less recognition accuracy, those few of the characters that were recognized have the underdot and the diacritic written far apart from the characters, the characters that were written correctly with underdot directly under the characters and the diacritic signs above the characters were not recognized correctly. The recognition accuracy achieved is high compared to other classification algorithm that carries out separate feature extraction techniques separately.

Author Contributions

Concept: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%), Design: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%), Supervision: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%), Data collection and/or processing: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%), Data analysis and/or interpretation: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%), Literature search: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%), Writing: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%), Critical review: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%). Submission and revision: J.F.A. (35%), S.R.Y. (30%) and A.O.A. (30%). All authors reviewed and approved final version of the manuscript.

Conflict of Interest

The authors declared that there is no conflict of interest.

References

Ajao JF, Olawuyi DO, Odejebi OO. 2018. Yoruba handwritten character recognition using freeman chain code and k-nearest neighbor classifier. *J Teknol dan Sist Komp*, 6(4): 129-134.

Altwaijry N, Al-Turaiki I. 2021. Arabic handwriting recognition system using convolutional neural network. *Neural Comp App*, 33(7): 2249-2261.

Asahiah FO. 2014. Development of a Standard Yorùbá digital text automatic diacritic restoration system. *Signature*, 22: 02.

Bamgbose A. 2000. *A grammar of Yorùbá (Vol. 5)*. Cambridge University Press, Cambridge, UK, pp: 70.

Bluche T, Ney H, Kermorvant C. 2014. A comparison of sequence-trained deep neural networks and recurrent neural networks optical modeling for handwriting recognition. In *International conference on statistical language and speech processing*. Springer, Cham, Germany, pp: 199-210.

Calvo-Zaragoza J, Toselli AH, Vidal E. 2019. Handwritten music recognition for mensural notation with convolutional recurrent neural networks. *Pattern Recog Lett*, 128: 115-121.

Chacko BP, Krishnan VV, Raju G, Anto PB. 2012. Handwritten character recognition using wavelet energy and extreme learning machine. *Int J Machine Learn Cybernet*, 3(2): 149-161.

Chaudhari K, Thakkar A. 2019. Survey on handwriting-based personality trait identification. *Expert Sys App*, 124: 282-308.

Darwish K, Habash N, Abbas M, Al-Khalifa H, Al-Natsheh HT,

- Bouamor H, Mubarak H. 2021. A panoramic survey of natural language processing in the Arab world. *Commun ACM*, 64(4): 72-81.
- Das K, Behera RN. 2017. A survey on machine learning: concept, algorithms and applications. *Int J Innov Res Comp Commun Eng*, 5(2): 1301-1309.
- Dewa CK, Fadhilah AL, Afiahayati A. 2018. Convolutional neural networks for handwritten Javanese character recognition. *Indonesian J Comp Cybernetics Sys*, 12(1): 83-94.
- Garoot AH, Safar M, Suen CY. 2017. A comprehensive survey on handwriting and computerized graphology. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), November 9-15, 2017, Kyoto, Japan, Vol. 1, pp: 621-626.
- Gatos B, Pratikakis I, Perantonis SJ. 2004. An adaptive binarization technique for low quality historical documents. In *International Workshop on Document Analysis Systems*. Springer, Berlin, Germany, pp: 102-113.
- Graves A, Liwicki M, Fernández S, Bertolami R, Bunke H, Schmidhuber J. 2008. A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans Pattern Anal Machine Intell*, 31(5): 855-868.
- Likforman-Sulem L, Zahour A, Taconet B. 2007. Text line segmentation of historical documents: a survey. *International J Doc Anal Recog*, 9(2-4): 123-138.
- Ly NT, Nguyen CT, Nguyen KC, Nakagawa M. 2017. Deep convolutional recurrent network for segmentation-free offline handwritten Japanese text recognition. In 2017 14th IAPR International Conference on Document Analysis and Recognition, November 9-15, 2017, Kyoto, Japan, Vol. 7, pp: 5-9.
- Mithe R, Indalkar S, Divekar N. 2013. Optical character recognition. *Int J Recent Technol Eng*, 2(1): 72-75.
- Nina O, Morse B, Barrett W. 2011. A recursive Otsu thresholding method for scanned document binarization. In 2011 IEEE Workshop on Applications of Computer Vision (WACV), January 5-7, 2011, Kona, HI, US, pp: 307-314.
- Ojo O. 2007. The Yorùbá in transition: history, values, and modernity. *Africa Today*, 54(2): 151-152.
- Oladele MO, Adepoju TM, Olatoke O, Ojo OA. 2020. Offline Yorùbá handwritten word recognition using geometric feature extraction and support vector machine classifier. *Malaysian J Comp*, 5(2): 504-514.
- Óní OJ, Asahiah FQ. 2020. Computational modelling of an optical character recognition system for Yorùbá printed text images. *Sci African*, 9: e00415.
- Oyedotun OK, Dimililer K. 2016. Pattern recognition: invariance learning in convolutional auto encoder network. *Int J Image Graph Signal Proces*, 8(3): 19-27.
- Peel JDY. 2009. A Heterogeneous volume of Yorùbá history and culture-the Yorùbá in transition: history, values, and modernity. Edited by Toyin Falola and Ann Genova. Carolina Academic Press, Durham, NC, US, pp: 498.
- Srihari SN, Ball G. 2012. An assessment of Arabic handwriting recognition technology. In *Guide to OCR for Arabic Scripts*, Springer, London, UK, pp: 3-34.
- Srihari SN, Yang X, Ball GR. 2007. Offline Chinese handwriting recognition: an assessment of current technology. *Front Comp Sci China*, 1(2): 137-155.
- Wang J, Wu R, Zhang S. 2021. Robust recognition of Chinese text from cellphone-acquired low-quality identity card images using convolutional recurrent neural network. *Sensors Mater*, 33(4): 1187-1198.



PERFORMANCE COMPARISON OF K-MEANS AND DBSCAN METHODS FOR AIRLINE CUSTOMER SEGMENTATION

Kevser ŞAHİNBAŞ^{1*}


¹Istanbul Medipol University, Faculty of Business Administration and Management Sciences, Department of Management Information Systems, 34810, İstanbul, Türkiye

Abstract: Organizations are now fully embracing ideas such as customer success, customer loyalty, customer experience management and customer satisfaction. The application of these concepts must be based on three pillars of technology, process and people, to ensure that the organization ultimately has satisfied, loyal and successful customers. In today's competitive environment, as in all sectors, gaining great services in the aviation industry can provide a competitive advantage. With this study, it is aimed to help aviation companies to know how their services should meet the needs of customers and to obtain passenger satisfaction. Customer segmentation is widely used, which groups objects according to the similarity difference on each object and provides a high level of homogeneity in the same cluster or a high level of heterogeneity between each group. The aim of this study is to examine airline passenger satisfaction by using data mining methods including K-Means and Density-based spatial clustering of applications with noise (DBSCAN) clustering algorithms to reveal the service quality importance for customer satisfaction. K-Means algorithm achieved slightly better results than DBSCAN algorithm with a Silhouette value of 0.1450671.

Keywords: Clustering, Customer segmentation, K-means, DBSCAN, Data mining, Data management

*Corresponding author: Istanbul Medipol University, Faculty of Business Administration and Management Sciences, Department of Management Information Systems, 34810, İstanbul, Türkiye

E mail: ksahinbas@medipol.edu.tr (K. ŞAHİNBAŞ)

Kevser ŞAHİNBAŞ  <https://orcid.org/0000-0002-8076-3678>

Received: September 05, 2022

Accepted: September 19, 2022

Published: October 01, 2022

Cite as: Şahinbaş K. 2022. Performance comparison of K-Means and DBSCAN methods for airline customer segmentation. BSJ Eng Sci, 5(4): 158-165.

1. Introduction

The aviation, which is one of the most used modes of transportation due to its benefits such as safety, speed and comfort, has become one of the most important sectors in recent years with different factors such as developing technology, increasing aircraft companies and rising demand and decreasing flight prices. Airlines' efforts to attract and retain consumers are increasing competition in the aviation industry. Airlines must realize that price changes alone cannot win the competitive position of price competition for their competitors in the long run (Chang and Yeh, 2002). According to studies, when selecting an airline service, the passenger takes into account both the price and the quality of the service (Jou et al., 2008). Many previous studies have shown that airline service quality is an important factor in passenger satisfaction. (Jiang and Zhang, 2016; Ariffin et al., 2020). The fact that transportation services are shaped according to the wishes and needs of passengers today has caused the companies that provide this service to shift their focus from making profit to providing customer satisfaction. The fact that repeat purchasers are in the most valuable customer category for most businesses highlighted the importance of customer retention. This study examines the full-service airline business model for passenger

satisfaction. Because in today's competitive market, airline firms' ability to provide excellent customer service is a key competitive advantage. When a customer is dissatisfied with the level of service they receive, they are likely to change their mind and choose a different airline for their subsequent travels (Archana and Subha, 2012). Because the frequency of negative events will influence a customer's opinion of the business, the quality of the service provided is considered a significant factor in providing customer satisfaction (Munusamy et al., 2011).

Clustering has tremendous application. DBSCAN is especially useful for large databases and datasets including noisy objects (Cassisi et al., 2013). Besides DBSCAN efficiently discovers random sets of sizes, shapes and numbers in a large dataset (Jahirabadkar and Kulkarni, 2014). Large datasets can be handled by K-means with ease because it is simple to apply and has linear time complexity. K-Means and DBSCAN are used in many fields such as fast clustering of big data (Hanafi and Saadatfar, 2022) and data-intensive applications (Ajin and Kumar, 2016; Saeed et al., 2020), healthcare (Santhanam and Padmavathi, 2015), social networks (Hao et al., 2020), anomaly detection (Chen and Li, 2011), bioinformatics (Masood and Khan, 2015; Bustamam et al., 2017), customer segments of charging stations (Straka and Buzna, 2019), detecting grain inventory



modes (Cui et al., 2021) and so on.

Various conceptual and empirical studies have been conducted to find service quality issues in the aviation industry. Yelmen et al. (2020) provide a customer segmentation analysis using air ticket sales data with the self-organizing map method. In their study, they obtained 15 clusters by grouping customers who had similar sales behaviors. Leon and Martin (2020) used fuzzy logic and fuzzy segmentation to assess airline customer happiness and service quality in the U.S. market. The questionnaire was collected online with 624 respondents using Amazon Mechanical Turk. The findings illustrate that technical quality is not as important to passengers' satisfaction as functional quality, however both factors contribute to overall airline satisfaction. According to a survey, the quality of the food and beverage services is one of the least significant elements affecting how satisfied passengers are with their flights (Deveci and Demirel, 2018). Noviantoro and Huang (2022) proposed a model to investigate airline passenger satisfaction by applying data mining techniques. Their study showed that online/mobile boarding was the most important variables for passenger satisfaction, followed by in-flight wi-fi service second, baggage handling third, and in-flight entertainment fourth. The same constructs are relevant, and they are ranked differently from most to least effective, according to Farooq and Radovic-Markovic (2016). A study is conducted data of more than 5800 airline passengers to provide the segmentation of consumers into business and leisure did not accurately reflect the diversity of customer preferences, which resulted in an incorrect understanding of those choices (Teichert et al., 2008). A preprocessing phase is added to the Fahim (2021) approach before using the k-means algorithm to determine the number of clusters and beginning centers. The DBSCAN algorithm will be used as a preprocessing step in the suggested procedure. His article focuses on DBSCAN and k-means as a result. The final product will be of higher quality because the suggested method will eventually converge to the global minimum. The suggested approach is identical to DBSCAN in that it takes two input parameters and has a time complexity of $O(n \log n)$. In Majhi and Biswal's (2018) study a hybrid clustering approach built on K-means and Ant Lion Optimization was taken into consideration for the best cluster analysis. Based on several performance metrics, the proposed algorithm's performance is contrasted with that of the K-Means, KMeans- Particle Swarm Optimization (PSO), DBSCAN, and Revised DBSCAN clustering algorithms. Eight datasets are used in the experiment, and statistical analysis is done for each of them. The findings demonstrate that in terms of sum of intra-cluster distances and F-measurement, the hybrid of K-Means and Ant Lion Optimization method preferentially outperforms the other four algorithms. In the study of Du (2020) the global parameter selection step of the DBSCAN requires human interaction, and the regional

query procedure is difficult and prone to item loss. Based on maintaining the intrinsic nonlinearity of Internet of Thing (IoT) data, an advanced parameter adaptive and regional query density clustering method is suggested that can efficiently eliminate redundant data in the high-level complex data domain.

In this study, airline passenger satisfaction was investigated by applying data mining methods K-Means and DBSCAN algorithms. This study provides a guide to airlines, and by this way, airlines can use this survey as a reference in developing measures to aid passengers better know how they feel about airline services. This research will help improve airline management and increase the caliber of some services over competitors to give airlines a competitive advantage.

2. Materials and Methods

2.1. Dataset

In this paper, the dataset was collected from the publicly accessible Kaggle website (<https://www.kaggle.com/code/frixinglife/airline-passenger-satisfaction>). The dataset depicts customer satisfaction based on data acquired from 25.976 passenger samples who flew on full-service airlines during a survey conducted at the airport following their arrival in 2015 (Table 1).

Table 1. Variables description

Numeric Variables	
Age	
Flight distance	
Departure delay in minutes	
Arrival delay in minutes	
Categorical variables with satisfaction level (0: not rated; 1-5)	
Inflight wifi service	
Departure/Arrival time convenient	
Ease of Online booking	
Gate location	
Food and drink	
Online boarding	
Seat comfort	
Inflight entertainment	
On-board service	
Leg room service	
Baggage handling	
Checkin service	
Inflight service	
Cleanliness	
Categorical Variables	Definition
Gender	male or female
Customer type	regular or non-regular airline customer
Type of travel	personal or business travel
Class	business, economy, economy plus
Air passenger satisfaction	1- Satisfied, 0- Neutral or dissatisfied

2.2. Clustering Algorithms

Clustering, which is one of the purposes of Data Mining, is frequently implemented in many areas such as pattern recognition and statistical data analysis (Goharnejad et al., 2019). Data mining relies heavily on clustering algorithms, which group objects with related attributes together and organize data in databases into groups or clusters. Data clustering is making strong progress. Clustering analysis has recently great attention in the research of data mining.

2.2.1. K-means

One of the most widely utilized algorithms is the well-known K-means algorithm. It is a clustering method used for classifying data. The main aim of use is to divide the data to be classified into k classes determined by the researcher or clusters in terms of their properties. In the K-Means algorithm (Table 2), k, which represents the number of clusters sought, is a previously known constant and its value does not change until the clustering process is finished (Kaufman and Rosseeauw, 1990). This approach needs a predefined number of clusters, which can be determined using techniques like the elbow method or expert opinion (Han et al., 2011).

Table 2. Algorithm steps of K-means (Hartigan and Wong, 1979)

Algorithm: K-Means

- 1: The value of k, which represents the number of clusters, is read by the algorithm. This value read is given to the algorithm as ready from the outside.
- 2: The cluster center is determined randomly. There are k cluster centers. The first k point can be the center.
- 3: The closeness of the points to the determined centers is calculated.
- 4: According to the calculated values, the points are clustered based on the centers they are close to.
- 5: In order to determine new cluster centers, the averages of the clusters are calculated.
- 6: If there are other points to be clustered, the process is repeated. But if there is no point to be clustered, the process is completed

2.2.2. DBSCAN

DBSCAN algorithm is a density-based spatial clustering algorithm and enables to reveal the neighborhood of data points in two or multidimensional space. This approach can find clusters of different shapes and does not need prior knowledge of the number of clusters (Mahesh, 2020; Hanafi and Saadatfar, 2022). Due to its focus on spatial perspectives, the database is mostly employed in the analysis of spatial data (Ester et al., 1996). Eps, MinPts, core object, direct density accessible point, density accessible point, and density connected point are the fundamental concepts for DBSCAN method (Figure 1; Table 3). Eps and MinPts are taken as input parameters. Starting with any object in the database, it examines every object. If the controlled item has previously been a

part of a cluster, it moves on without being processed to the other object. It uses a Region Query to find the Eps neighbors of the object if it has never been clustered before. This item and its neighbors are referred to as a new cluster if the number of neighbors exceeds MinPts. The process then searches the new region for each neighbor who was previously unclustered. The region query is a part of the cluster if it has more neighbors than MinPts. In the DBSCAN method, the density function of a point neighborhood eps is thus defined as $N_{eps}(p)$, which is depicted in Equation 1.

$$N_{eps}(p) = \text{card}(a \in D \mid \text{dist}(a, b) \leq eps) \quad (1)$$

where D stands for the dataset, b is a random point within D, dist(a,b) is the function measuring the distance between points a and b, and card(*) is the function calculating the number of items in a set.

After that Equation 1 and the specified threshold mpt are used to determine three different types of points.

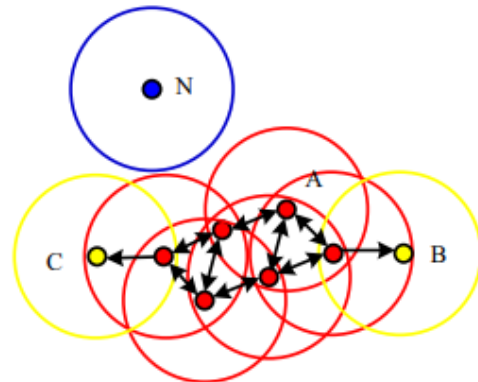


Figure 1. DBSCAN clustering method.

Table 3. Algorithm steps of DBSCAN

Algorithm: DBSCAN (D, eps, mpt)

Inputs: D, dataset of points;
eps, the neighborhood size;
mpt, the density threshold;

- 1: Start
- 2: Arbitrary select a point a 2D;
- 3: Compute the density $N_{eps}(a)$ and decide the types of point a according to mpt;
- 4: If a is a seed point, a cluster is created;
- 5: If a is a boundary point and density is accessible from other points, expand the set and visit the next point;
- 6: Continue processing until all data points have been visited;
- 7: Stop

DBSCAN is superior to other traditional clustering methods in many ways. This technique allows the management of noise patterns in the data and allows the identification of clusters of various shapes. DBSCAN typically gives reliable results. It is also used to minimize the number of computational operations.

2.3. Determination of k Cluster Number by Elbow Method

The elbow method is another simple and meaningful decision solution looking for the optimal K number. The Elbow method, which is calculated by the sum of the squares of the distance of each point from the cluster centers (WCSS: Within Clusters Sum of Square), is another simple and meaningful decision solution that seeks the optimal K number. According to this method, the point where the amount of change in WCCS decreases is the bend point and this bend point represents the best number of k clusters (Ketchen and Shook, 1996).

2.4. System Overview

The research methodology of the study is presented in Figure 2.

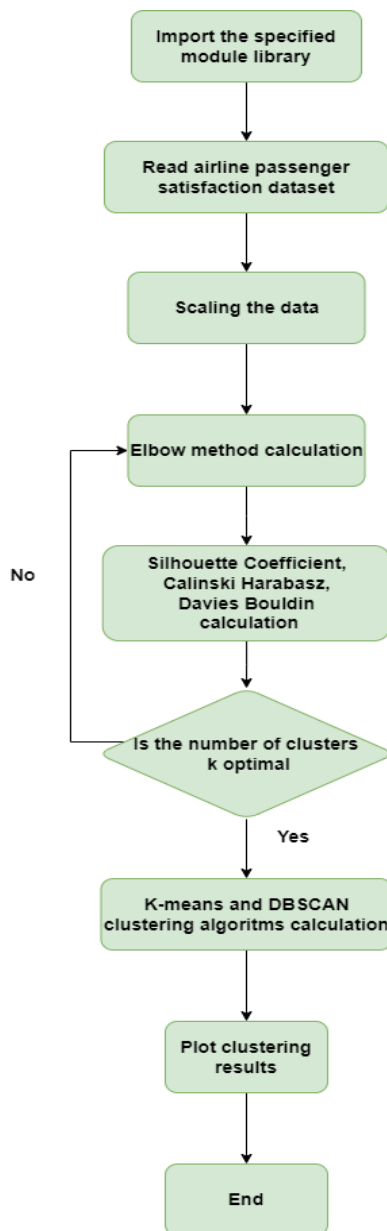


Figure 2. Workflow of cluster analysis.

Dataset was obtained from Kaggle web site. Data is preprocessed, then is scaled. Elbow method is executed to find the optimal k value for K-Means. DBSCAN and K-

Means algorithms are applied for airline passenger data. Performance metrics were found to select the algorithm. Lastly, clustering results are plotted.

2.5. Performance Metrics

In this study, Silhouette Coefficient, Calinski Harabasz, Davies Bouldin and elbow method were used to quantitatively evaluate the cluster result.

2.5.1. Silhouette Coefficient

The silhouette coefficient is a parameter that can be used to evaluate cluster performance (Rousseeuw, 1987). The clustering effect increases as the value increases; the range of this number is between -1 and 1. The solution of the resulting silhouette coefficient is shown in Equation 2.

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & a(i) < b(i) \\ 0, & a(i) = b(i) \\ \frac{a(i)}{b(i)} - 1, & a(i) > b(i) \end{cases} \quad (2)$$

Here $a(i)$ specifies the distance between element i and other samples in the same cluster. $b(i)$ is the average distance of the sample i and the color samples in other clusters. $S(i)$ indicates the average value of the entire silhouette coefficient.

2.5.2. Calinski Harabasz

The Calinski-Harabasz (CH) criterion shows a ratio between the within-cluster distribution and the between-cluster distribution. The CH criterion is used to evaluate the compactness and segregation of clusters (Equation 3).

$$ch = \frac{tr(B_k)x(n_E - k)}{tr(W_k)x(k - 1)} \quad (3)$$

$tr(B_k)$: sum of squares within-clusters

$tr(W_k)$: between-cluster sum of squares

The highest ch value represents the best cluster (Caliński and Harabasz, 1974).

2.5.3. Davies Bouldin

The following equation is used to measure the clustering validity with this method, which aims to make the distance between the cluster's minimum and the distance between clusters maximum (Equation 4):

$$db = \frac{1}{k} \sum_{i=1}^k R_i \quad (4)$$

$i = 1, 2, \dots, k$ and $j = 1, 2, \dots, k$. and the following equation determines the maximum comparison ratio between other clusters (Equation 5):

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (4)$$

d_{ij} : distance between centers in clusters

S_i and S_j : the average distance to the centers of the

cluster where the cluster observations are located. Small *db* values indicate good clustering (Davies and Bouldin, 1979).

3. Results and Discussion

In this section, dataset variables are visualized,

correlation matrix results are shown and findings from algorithms are detailed. Figure 3 indicates the visualization of data. In Figure 4, number of satisfied and dissatisfied passengers are shown. 11403 passengers are satisfied, and 14573 passengers are dissatisfied.

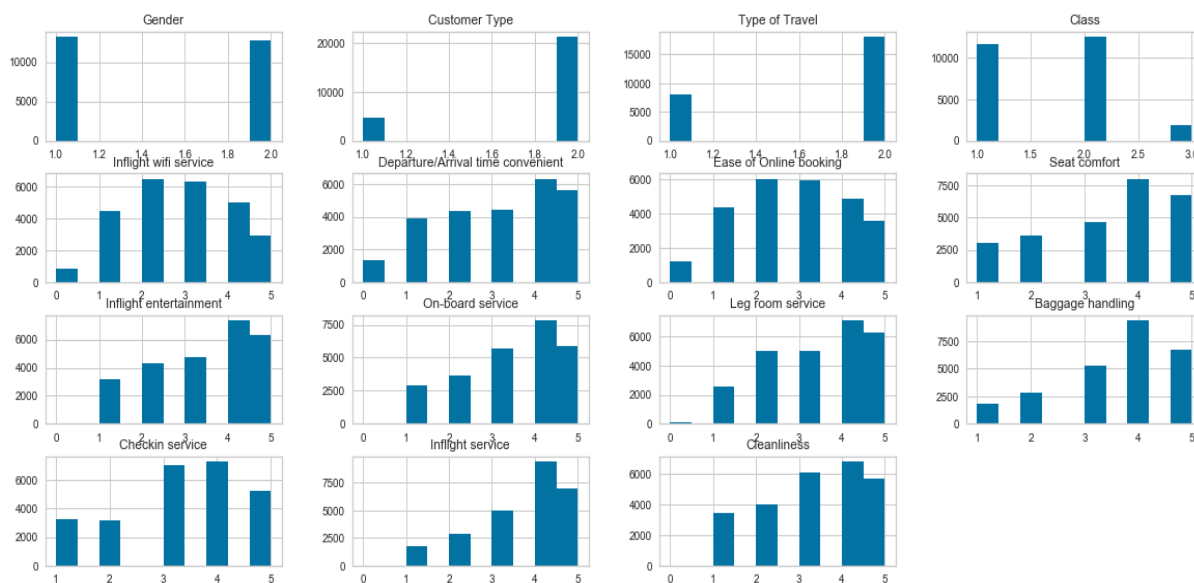


Figure 3. Data visualization.

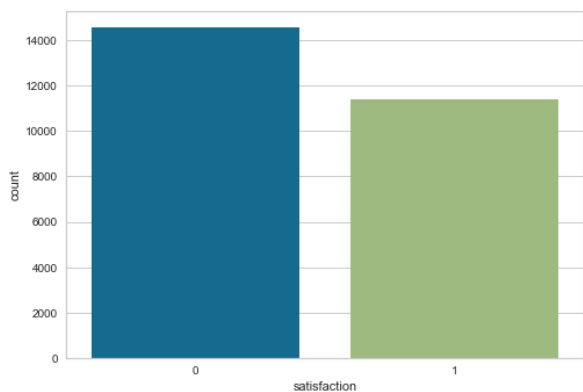


Figure 4. Number of satisfied and dissatisfied passenger.

Figure 5 illustrates correlation between variables. In correlation analysis, the aim is to see in which direction the other variable will change when the value of one variable changes (Taylor, 1990). The findings from Figure 5 indicate that the correlation coefficient between the 'age' attribute and the 'flight distance' variable is 0.099 with 0.000048 p value, indicating a weak and positive relationship between them. It means that while age increased, the flight distance increased slightly as well. Weak and positive relationship is seen between 'flight distance' and 'departure delay in minute' with 0.0034 correlation coefficient and 0.00065 P value. It shows that while flight distance increased, the departure delay in minute increased slightly. Lastly, there is low and negative relationship between 'age' and 'departure delay in minutes' with the value of -0.0043 and 0.00096 P

value. It means that while age increased, the departure delay in minute decreased slightly. P values are less than 0.05 that means our correlation coefficient is statistically significant. The relationship that emerged as a result of the correlation analysis should not be interpreted as a causality relationship.

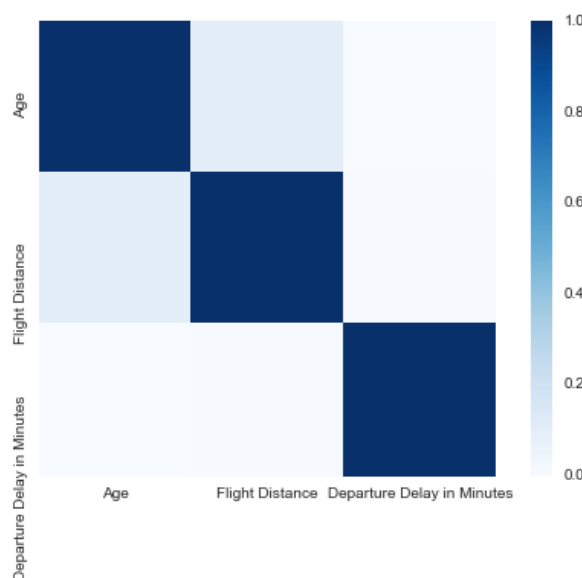


Figure 5. Correlation matrix.

As shown in Figure 6, the optimal k number is 8. Table 4 indicates the metrics values of each clustering algorithm in terms of Silhouette Coefficient, Calinski Harabasz and Davies Bouldin.

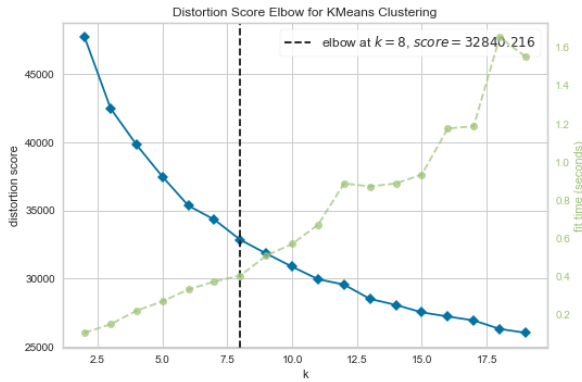


Figure 6. Elbow method.

Table 4. Clustering performance scores

Algorithm	Sihoutte Coefficient	Calinski Harabasz	Davies Bouldin
K-Means	0.145	2854.193	2.078
DBSCAN	0.133	1672.774	2.287

The findings in Table 4 demonstrate that the K-Means algorithm achieve the best performance in terms of Sihoutte Coefficient, Calinski Harabasz and Davies Bouldin with a small margin of 0.145, 2854.193 and 2.078, respectively with a small difference.

Figures 7 and 8 demonstrate the Silhouette analysis of the K-Means and DBSCAN algorithms. The Silhouette range is between -1 and 1 as described in Section 2.5.1. The silhouette coefficient graph indicates that the K-Means clustering algorithm performs better.

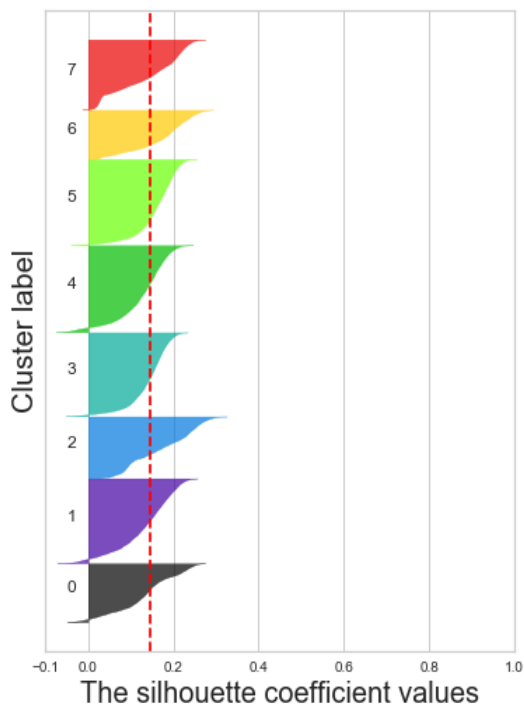


Figure 7. K-means Silhouette coefficient values.

In Figure 9, PCA is applied to the dataset. The number of clusters for K-Means is 8.

The results obtained with the application of K-means and DBSCAN algorithms are compared and the following evaluations can be clarified. These two algorithms generate similar results. The findings from analysis present that K-Means outperformed with a small difference. In fact, both algorithms provide high performance. It can be stated that K-Means and DBSCAN approaches perform quite well to cluster airline customer. In order to achieve good results in the DBSCAN algorithm, it is necessary to change the algorithm with different parameters. For example, eps value can be changed to obtain good results.

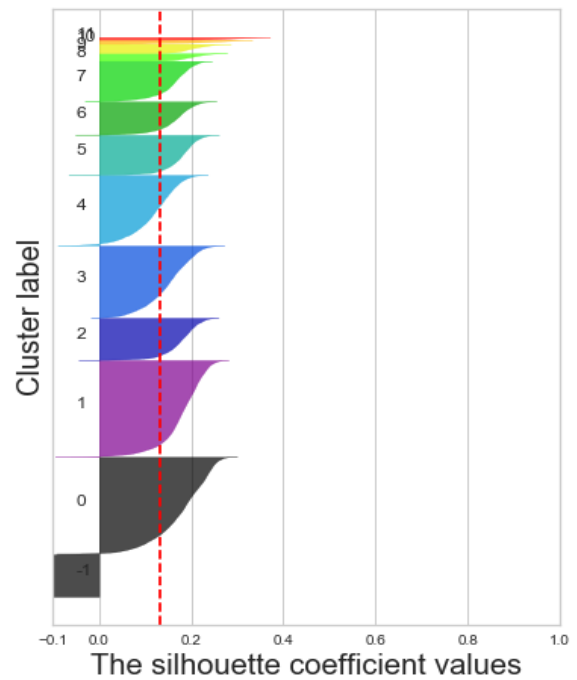


Figure 8. DBSCAN Silhouette coefficient values.

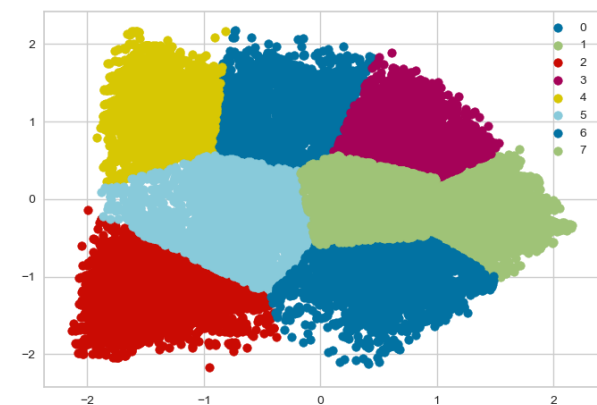


Figure 9. PCA of K-means algorithm.

Table 5. Clustering analysis results

Cluster	Member	Satisfied (25.976)		High score	Low score
1	3899	Yes	2426	Baggage Handling	Gate Location
		No	1473	Inflight wifi service	Checkin service
2	3012	Yes	2940	Inflight service	Gate Location
		No	72	Inflight service	Inflight wifi service
3	3439	Yes	3105	Seat comfort	Gate Location
		No	334	Inflight wifi service	Cleanliness
4	2456	Yes	341	Baggage Handling	Cleanliness
		No	2115	Inflight service	Inflight wifi service
5	3374	Yes	1360	Inflight entertainment	Gate Location
		No	2014	Inflight entertainment	Inflight wifi service
6	2191	Yes	509	Online Boarding	Inflight wifi service
		No	1682	Departure, Arrival time convenient	Baggage handling
7	4635	Yes	1181	Online boarding	Gate Location
		No	3454	Departure, Arrival time convenient	Gate Location
8	2970	Yes	494	Inflight service	Departure/Arrival time convenient
		No	2676	Departure, Arrival time convenient	Ease of Online booking

The definition of columns is as follow: Cluster= cluster number, Member= class member count, Satisfied= satisfied member count and dissatisfied member count, High score= the category name with the highest score from the headings used as the data source, Low score= the category name with the lowest score from the headings used as the data source.

The findings from Table 5 present the members of each cluster, satisfied and dissatisfied number for each cluster. The findings of the study can provide information that can help airline management, professionals and those related to the aviation industry see the most important and unimportant criteria and help them to develop strategies and decision-making in this direction.

4. Conclusion

Ensuring the satisfaction of passenger in air travel, which is one of the most popular transportation modes because it is a faster means of transportation, plays a significant role in increasing the number of competitors day by day.

In this paper, a publicly dataset for airline passenger is used by applying K-Means and DBSCAN clustering algorithms to cluster the airline customer dataset. The reason behind the selection of these two algorithms is that DBSCAN is the first and best-known density-based clustering algorithm and K-means handles huge datasets well since it is straightforward to implement and has linear time complexity.

This study presents a proof of concept on how data analytics can be used in customer segmentation for airline passengers. When the findings of K-Means algorithm and DBSCAN are compared, it is observed that K-Means achieved slightly better results in Sihoutte Coefficient, Calinski Harabasz and Davies Bouldin metrics in all three performances, 0.145, 2854.193 and 2.078, respectively. The study's findings can offer information that airline management, professionals, and other parties with an interest in the aviation industry can use and make decisions about. More different clustering algorithms and the cost implication of each cluster can be considered for further studies.

Author Contributions

All task was done by the single author: K.Ş. (100%). The author reviewed and approved the manuscript.

Conflict of Interest

The author declared that there is no conflict of interest.

References

- Ajin VW, Kumar LD. 2016. Big data and clustering algorithms. International conference on research advances in integrated navigation systems (RAINS) IEEE, 6-7 May 2016, Bangalore, India, pp: 1-5.
- Ariffin Mohd IA, Yajid SA, Johar MGM. 2020. Consumer preferences of airline choice: A comparison of Air Asia and Malaysia Airlines System. *Syst Rev Pharm*, 11(1): 817-826.
- Archana R, Subha MV. 2012. A study on service quality and passenger satisfaction on Indian airlines, *Int J Multidis Res*, 2(2): 50-63.
- Bustamam A, Tasman H, Yuniarti N, Mursidah I. 2017. Application of K-means clustering algorithm in grouping the DNA sequences of hepatitis B virus (HBV). *AIP Conf Proc*, 1862(1): 030134.
- Caliński T, Harabasz J. 1974. A dendrite method for cluster analysis. *Commun Stat Theo Meth*, 3(1): 1-27.
- Cassisi C, Ferro A, Giugno R, Pigola G, Pulvirenti, A. 2013. Enhancing density-based clustering: parameter reduction and outlier detection. *Inf Syst*, 38(3): 317-330.
- Chang YH, Yeh CH. 2002. A survey analysis of service quality for domestic airlines. *European J Oper Res*, 139(1): 166-177. DOI: 10.1016/S0377-2217(01)00148-5.
- Chen Z, Li YF. 2011. Anomaly detection based on enhanced DBScan algorithm. *Procedia Eng*, 15: 178-182.
- Cui H, Wu W, Zhang Z, Han F, Liu Z. 2021. Clustering and application of grain temperature statistical parameters based on the DBSCAN algorithm. *J Stored Prod Res*, 93: 101819.
- Deveci M, Demirel NÇ. 2018. A survey of the literature on airline crew scheduling. *Eng App Artif Intel*, 74: 54-69.
- Ester M, Kriegel HP, Sander J, Xu X. 1996. A density based algorithm for discovering clusters in large spatial databases.

- Int. Conference of Knowledge Discovery and Data Mining (KDD'96), Portland, USA, pp: 226-231.
- Davies DL, Bouldin DW. 1979. A cluster separation measure. *IEEE Transact Pattern Analysis Machine Intel*, 2: 224-227.
- Du Z. 2020. Energy analysis of Internet of things data mining algorithm for smart green communication networks. *Comp Commun*, 152: 223-231.
- Fahim A. 2021. K and starting means for k-means algorithm. *J Comput Sci*, 55: 101445.
- Farooq MS, Radovic-Markovic M. 2016. Modeling entrepreneurial education and entrepreneurial skills as antecedents of intention towards entrepreneurial behaviour in single mothers: a PLS-SEM approach. *ETCTFP*, 2016: 198-216.
- Goharnejad H, Shamsai A, Zakeri Niri M. 2019. Prediction of sea level rise in the south of Iran coastline: evaluation of climate change impacts. *Water Res Eng*, 12(42): 1-17.
- Jiang H, Zhang Y. 2016. An investigation of service quality, customer satisfaction and loyalty in China's airline market. *J Air Trans Manag*, 57: 80-88.
- Han J, Pei J, Kamber M. 2011. *Data mining: concepts and techniques*. Elsevier, New York, US, pp: 703.
- Hao F, Zhang J, Duan Z, Zhao L, Guo L, Park DS. 2020. Urban area function zoning based on user relationships in location-based social networks. *IEEE Access*, 8: 23487-23495.
- Hanafi N, Saadatfar H. 2022. A fast DBSCAN algorithm for big data based on efficient density calculation. *Expert Sys App*, 203: 117501.
- Hartigan JA, Wong MA. 1979. Algorithm AS 136: A k-means clustering algorithm. *J Royal Stat Soc Series c*, 28(1): 100-108.
- Jou RC, Lam SH, Hensher DA, Chen CC, Kuo CW. 2008. The effect of service quality and price on international airline competition. *Transport Res Part E*, 44(4): 580-592.
- Jahirabadkar S, Kulkarni P. 2014. Algorithm to determine ϵ -distance parameter in density based clustering. *Expert Sys App*, 41(6): 2939-2946.
- Kaufman L, Rosseeuw PJ. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley and Sons Inc., New York, US, pp: 335.
- Leon S, Martín JC. 2020. A fuzzy segmentation analysis of airline passengers in the US based on service satisfaction. *Res Transport Busin Manag*, 37: 100550.
- Ketchen DJ, Shook CL. 1996. The application of cluster analysis in strategic management research: an analysis and critique. *Strat Manag J*, 17(6): 441-458.
- Masood MA, Khan MNA. 2015. Clustering techniques in bioinformatics. *IJ Modern Educ Comp Sci*, 1: 38-46.
- Munusamy J, Chelliah S, Pandian S. 2011. Customer satisfaction delivery in airline industry in Malaysia: a case of low cost carrier. *Australian J Basic App Sci*, 5(11): 718-723.
- Noviantoro T, Huang JP. 2022. Investigating airline passenger satisfaction: Data mining method. *Res Transport Busin Manag*, 43: 100726.
- Majhi SK, Biswal S. 2018. Optimal cluster analysis using hybrid K-Means and Ant Lion Optimizer. *Karbala Int J Modern Sci*, 4(4): 347-360.
- Mahesh B. 2020. Machine learning algorithms-a review. *Int J Sci Res*, 9: 381-386.
- Straka M, Buzna LU. 2019. Clustering algorithms applied to usage related segments of electric vehicle charging stations. *Transport Res Proc*, 40: 1576-1582.
- Teichert T, Shehu E, von Wartburg I. 2008. Customer segmentation revisited: The case of the airline industry. *Transport Res Part A*, 42(1): 227-242.
- Rousseeuw PJ. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J Comput App Math*, 20: 53-65.
- Saeed MM, Al Aghbari Z, Alsharidah M. 2020. Big data clustering techniques based on spark: a literature review. *Peer J Comp Sci*, 6: e321.
- Santhanam T, Padmavathi MS. 2015. Application of K-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. *Procedia Comp Sci*, 47: 76-83.
- Taylor R. 1990. Interpretation of the correlation coefficient: a basic review. *J Diag Medic Sonograp*, 6(1): 35-39.
- Yelmen İ, Üstebay S, Zontul M. 2020. Customer segmentation based on self-organizing maps: a case study on airline passengers. *J Aeronautics Space Technol*, 13(2): 227-233.



RÜZGÂR ENERJİ SANTRALİ DEĞERLEMESİNDE REEL OPSİYONLARIN KULLANILMASI

Duygu BIYIKLI^{1*}, Faik Ahmet SESLİ², Pelin KASAP³

¹Kastamonu Üniversitesi, İhsangazi Meslek Yüksekokulu, Mimarlık ve Şehir Planlama Bölümü, 37250, Kastamonu, Türkiye

²Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Anabilim Dalı, 55139, Samsun, Türkiye

³Ondokuz Mayıs Üniversitesi, Fen Fakültesi, İstatistik Anabilim Dalı, 55139, Samsun, Türkiye

Özet: Rüzgâr Enerji Santrali (RES) gibi büyük bütçeli yatırım projelerinin değerlendirilmesinde, değerlendirme yönteminin seçimi en önemli kararların başında gelmektedir. Geleneksel değerlendirme yöntemlerinin, yatırımın ekonomik ömür süresince risk ve belirsizlikleri dikkate alınmaması, herhangi bir dönemde olası durumlara karşı genişleme, terk etme ya da erteleme gibi esnek çözümler sunmaması, uygulamalarda olumsuz durumlar yaratmaktadır. Reel opsiyonlar, yatırımdaki risk ve belirsizliğin, olasılık dağılımı yoluyla çeşitli analitik yöntemler ya da simülasyonlar kullanarak ele alınmasını sağlamaktadır. Yatırıma 2021 yılı Haziran ayında başladığı düşünüldüğünde, bu konuda daha önce hazırlanmış ön fizibilite raporlarından yararlanılmıştır ve piyasa değerleri de göz önünde bulundurularak proje parametreleri belirlenmiştir. Sinop ilinde ki 5 MW elektrik üretimi kapasiteli, %35 kapasite faktörüne ve 15 yıllık ekonomik ömre sahip bu RES için, ilk olarak geleneksel değerlendirme yöntemlerinden literatürde sıklıkla kullanılan, İndirgenmiş Nakit Akışlar (İNA) yöntemi ile Net Bugünkü Değer (NBD) elde edilmiştir. İkinci olarak, reel opsiyonlarla değerlendirme sürecinde ise analitik yöntemlerden Black-Scholes değerlendirme yöntemi ile opsiyon değeri ve projenin uygulanabilir olup olmadığının değerlendirilmesi için Genişletilmiş Net Bugünkü Değer (GNBD) elde edilmiştir. Yatırım projesinin uygulanabilir olması için kullanılan değerlendirme yöntemi ile elde edilen NBD'nin pozitif sonuç vermesi ve yatırımcının da ekonomik ömür süresi sonunda kâr elde etmiş olması gerekmektedir. Negatif sonuç veren yatırım için proje uygulanabilir olmamaktadır. RES için yapılan değerlendirme çalışmasında ki amaç, NBD ile negatif sonuç elde edilmesi ile birlikte proje uygulanabilir onayı alamazken, risk ve belirsizliği projeye dahil ederek reel opsiyonlarla pozitif sonuç elde edilmesi ile birlikte projenin uygulanabilir onayı alabildiği sonucu elde edilmiştir. Bu durumda reel opsiyonların geleneksel yöntemlere göre üstünlüğü gösterilmiş olmakla birlikte geleneksel yöntemlerin uygulamalarda ki yerini reel opsiyonlara bırakması gerektiği sonucuna da varılmıştır.

Anahtar kelimeler: NBD (Net bugünkü değer), Reel opsiyon, Black-scholes, RES (Rüzgâr enerji santrali)


Using Real Options in Wind Power Plant Evaluation


Abstract: The selection of the valuation method is one of the most important decisions in the valuation of large-budget investment projects such as the Wind Power Plant (WPP). The fact that traditional valuation methods do not take risks and uncertainties into account during the economic life of the investment and do not offer flexible solutions such as expansion, abandonment or postponement against possible situations in any period creates negative situations in applications. Real options enable the risk and uncertainty in the investment to be handled by using various analytical methods or simulations through probability distribution. Considering that the investment started in June 2021, pre-feasibility reports prepared beforehand were used in this regard, and the project parameters were determined considering the market values. Firstly, Net Present Value (NPV) was obtained by using the Discounted Cash Flows (DNA) method, which is one of the traditional valuation methods, for this WPP, which has a 5 MW electricity generation capacity, a capacity factor of 35% and an economic life of 15 years in the province of Sinop. are the most commonly used methods in the literature. Secondly, in the real options valuation process, the Black-Scholes valuation method, one of the analytical methods, used the option value and the Extended Net Present Value (GNPV) to evaluate whether the project is viable or not. In order for the investment project to be viable, the NPV obtained by the valuation method used must yield positive results and the investor must have earned a profit at the end of its economic life. The project is not viable for an investment with a negative result. The aim of the valuation study for RES was that while the project could not get the applicable approval with the negative result with NPV, the project was able to get the applicable approval with the real options by incorporating the risk and uncertainty into the project. In this case, although the superiority of real options over traditional methods has been demonstrated, it has also been concluded that traditional methods should be replaced by real options in practice.


Keywords: NPV (Net present value), Real option, Black-scholes, WPP (Wind power plant)

*Sorumlu yazar (Corresponding author): Kastamonu Üniversitesi, İhsangazi Meslek Yüksekokulu, Mimarlık ve Şehir Planlama Bölümü, 37250, Kastamonu, Türkiye

E mail: dbiyikli@kastamonu.edu.tr (D. BIYIKLI)

Duygu BIYIKLI  <https://orcid.org/0000-0002-0220-5101>

Faik Ahmet SESLİ  <https://orcid.org/0000-0001-8352-734X>

Pelin KASAP  <https://orcid.org/0000-0002-1106-710X>

Gönderi: 23 Ağustos 2022

Kabul: 25 Eylül 2022

Yayınlanma: 01 Ekim 2022

Received: August 23, 2022

Accepted: September 25, 2022

Published: October 01, 2022

Cite as: Biyikli D, Sesli FA, Kasap P. 2022. Using real options in wind power plant evaluation. BSJ Eng Sci, 5(4): xx-xx.



1. Giriş

Yatırımcılar proje değerini etkileyen parametrelerin, piyasa koşullarına göre nasıl etkileneceklerini ve ekonomik ömür sonunda elde edecekleri kârın ne yönde olacağını bilmek isterken, değerlendirme uzmanları tarafından bu durumla ilgili çeşitli yöntemler sunulmaktadır. Geleneksel değerlendirme yöntemleri, belirlilik koşulları altında risk, belirsizlik ve esneklik faktörlerini dahil etmeden yatırım değerlemesi yapmaktadır. Bu durum, yatırımların piyasa değerine göre daha az değerlendirilmesine neden olmakla birlikte projenin de uygulanabilir olması durumu olumsuz etkilenmektedir.

Geleneksel NBD, nakit akışlarını reel indirgeme oranı yani "riske göre düzeltilmiş iskonto oranı" ile indirgeyip, bu değerlerin toplamının yatırım maliyetinden çıkartılmasıyla elde edilir. Elde edilen NBD, yatırım projesinin kabul edilebilirliğinin bir ölçütüdür. Yatırım projesi için hesaplanan NBD'nin pozitif, negatif ve sıfır olması farklı şekillerde karar alınmasına da neden olmaktadır. NBD'si sıfır olan bir projenin uygulanması yatırımcı için ne kâr ne de zarar edebileceği bir durumdur ve ekonomik ömür sonunda bu yatırımdan etkili bir sonuç elde edilememektedir. NBD'si pozitif olan projenin uygulanması yatırımın değerini arttırarak uygulanabilir bir sonuç verirken, NBD'si negatif olan projenin uygulanması halinde de yatırımın değerinde ki azalış, uygulanabilirliği olumsuz etkilemektedir (Bilir, 2012). Değerleme sonucunda istenilen ve beklenen durum ise pozitif sonuçlu bir NBD elde etmektir.

Geleneksel değerlendirme yöntemleri, özellikle büyük ölçekli yatırım projeleri için yatırımın kârlılığını doğru bir şekilde değerlendirmekte yetersiz kalmaktadır (Dzyuma, 2012). Reel opsiyonlarla değerlendirme ise iskonto oranını kullanılması yerine, risk ve belirsizlik için risksiz faiz oranı ve değişkenlik için ise volatilité parametresinin kullanılması ve bir de yönetimsel esnekliklere izin vermesi nedeniyle öne çıkmaktadır (Colwell ve ark., 2002).

Reel opsiyonlar, proje kârlılığını hesaplamının yanı sıra, ekonomik ömür boyunca yıl bazında elde edilen verimliliğe göre bir sonra ki adım için ne yapılacağını düşünme ve uygulamaya geçmeyi sağlayan da bir yöntemdir. Örneğin 15 yıllık bir RES için 3. Yılda sonra yapılan fizibilite çalışmaları sonrası küçülme, genişleme ya da projeyi devretme ya da terk etme gibi opsiyonların kullanılabilirliğini yatırımcıya sağlamaktadır. Yanlış kararlar ile projeye devam edilmesi yerine, piyasa koşullarında nakit akış kaybının tolere edilebildiği durumlar tercih edilmelidir. Bazı durumlarda projenin ertelenebilme durumu da yatırımın kârlılığı için önemli bir faktördür. Piyasa koşullarının değiştiği, öngörülemeyen durumlarda ya da nakit akış kayıplarının kazanımlardan fazla olduğu durumlarda bu esnekliği kullanmakta yine yatırım için büyük rol oynamaktadır. Geleneksel değerlendirme yöntemleri ile riskin ve belirsizliğin yarattığı bu gibi durumlara bir çözüm getirilememekte ve yatırım bu durumdan olumsuz

etkilenmektedir (Mun, 2002). Belirsizlikler opsiyonlar için avantajlı bir durum olabilmektedir. Belirsizliğin artması, esneklikle birlikte opsiyon değerini de arttırmaktadır (Uzunlar ve Aktan, 2000).

Reel opsiyon teorisinin temeli, finansal varlıklar için hazırlanan finansal opsiyon teorisine dayanmaktadır. Myron C. Scholes ve Fisher Black tarafından 1970'lerin başında geliştirilen, Avrupa tipi satın alım opsiyonlarının fiyatlandırılmasında kullanılan Black-Scholes yöntemi, finansal opsiyonların hem literatürde hem de piyasa da geniş bir kullanım alanına sahip olmasıyla birlikte reel opsiyonların da gelişimini sağlamıştır (Schulmerich, 2010).

Reel opsiyon değerlendirme yöntemlerinden Black-Scholes ile birlikte Binomial yöntem ve Monte Carlo Simülasyonu da akademik çalışmalarda ve günümüzde sıklıkla karşımıza çıkmaktadır. Yenilenebilir enerji tesislerinin değerlemesinde daha dinamik bir yöntem olması dolayısıyla daha sık tercih edilen Black-Scholes yöntemi kadar diğer iki yöntemle de birbirlerine çok benzer sonuçlar vermeleri ile yatırım değerlendirme de kullanımlarının uygun olduğu da literatürde desteklenmektedir.

Abadie ve Chamorro (2014) çalışmalarında proje bilgilerini İngiltere'den aldıkları bir RES'in yatırımını reel opsiyon değerlendirme yöntemlerinden Monte Carlo simülasyonu yöntemi ve Binomial yöntem ile ele almış ve ekonomik ömür boyunca yaşanacak belirsizliklerden ötürü yatırımın bir süre ertelenmesi gibi bir opsiyon dahilinde değerlendirilmesini yapmıştır. Erteleme opsiyonu sonucunda vadeli işlemler piyasası üzerinde belirlenen beklenen elektrik fiyatı £ 48,91 / MWh'dan £ 85,91 / MWh'ya ulaştığını görmüş ve bunun sonucundan projenin kârlı hale geldiğini ifade etmiştir.

Brandao ve ark. (2005) problem çözmede Black-Scholes ve Binomial değerlendirme yönteminin kullanılması ile ilgili çalışmada birden çok risk ve belirsizlik durumunda daha dinamik olduğunu, yatırım projelerinin değerlemesinin geleneksel değerlendirme yöntemlerine göre daha üstün olduğunu anlatmıştır.

Chambers (2005) çalışmada yatırım projelerinin değerlemesinde reel opsiyonlar hakkında bilgi vermiş ve İNA yöntemine göre ret edilecek bir yatırım projesini genişleme opsiyonunun kullanılması ile yeniden değerlendirmiş ve sonuçta kabul edilebilir bir proje değerine sahip olduğunu ortaya koymuştur.

Fernandes ve ark. (2011) çalışmalarında enerji alanındaki yatırımlarda özellikle yenilenebilir enerji alanındaki yatırımlarda meydana gelen artış ve bu alandaki yatırımların değerlendirilmesinde reel opsiyonlar yönteminin önemini açıklamıştır. Özellikle RES yatırımları ile ilgili belirsizlikler ve risklerin, proje değerlendirilmesine en etkili reel opsiyonlar yöntemi ile yansıtıldığını ve bu nedenlerle bu alandaki yatırım projelerinin değerlendirilmesinde reel opsiyonlar yönetiminin önemi üzerinde durmuşlardır.

Kapucugil İkiz ve Deveci Kocakoç (2009) çalışmada, bilişim alanındaki yatırım projelerinin değerlemesinde

reel opsiyonların kullanılmasını ele almış ve çalışmanın sonucunda büyüme opsiyonuna sahip bir bilişim projesinin reel opsiyonlar yöntemi ve geleneksel proje değerlendirme yöntemleri ile değerlemesi yapılmış ve reel opsiyonlar yönteminin proje değerlemede daha gerçekçi sonuçlar verdiğini ve geleneksel değerlendirme yöntemleri ile ret edilebilen bir projenin reel opsiyonlar yöntemi ele kabul edilebilir şekilde hesaplandığını ortaya koymuştur. Mun (2006) çalışmasında, geleneksel değerlendirme yöntemlerini reel opsiyonlar değerlendirme yöntemleri ile karşılaştırmış ve reel opsiyonlar ile ilgili literatür bilgisi vererek avantajlarını göstermiştir. Özellikle risk ve belirsizlik durumlarında reel opsiyonlar yönteminin geleneksel yöntemlere göre üstün taraflarına dikkat çekmiş ve reel opsiyonların bu üstünlüklerini çalışmasında ele aldığı uygulamalar ile de ispatlamaya çalışmıştır.

Özcan (2013), "Muğla Bölgesinde Örnek Bir Rüzgâr Elektrik Santrali (RES) Yatırım ve Res Yatırımında Reel Opsiyonların Kullanımı Üzerine Bir İnceleme" isimli çalışmasında, bir rüzgâr enerji santrali projesini indirgenmiş nakit akımları yöntemine göre değerlemiş ve negatif çıkan yatırım sonucunu yani kabul edilemez olan projeyi reel opsiyonlar yöntemine yeniden ele almış ve opsiyon sözleşmesindeki satın alma opsiyonunun kullanılması durumunda projenin pozitif nakit akım verdiğini yani kabul edilebilir olduğunu ifade etmiştir.

Safarov (2009), çalışmasında enerji sektöründeki yatırımların reel opsiyonlar yöntemi ile değerlendirilmesini ele almış ve çalışmasında reel opsiyonlar ile geleneksel proje değerlendirme yöntemlerini karşılaştırmıştır. Çalışmanın uygulama bölümünde Barmek enerji firmasına ait yatırım projesini, Socar A.Ş' ye ait terk etme opsiyonu ve Manisa enerji firmasına ait büyüme opsiyonlarının kullanılması sonucunda yatırım projesinin analiz yapmış ve çalışma sonucunda geleneksel değerlendirme yöntemi ile ret edilebilecek bir yatırım projesinin aslında yatırım yapılabilir kârlı bir proje olduğunu ispatlamıştır.

Bu araştırma makalesinde ki çalışmanın amacı, ekonomik ömrü 15 yıl olan Sinop İlinde %35 kapasite faktörüne sahip bir RES'in değerlemesinde, geleneksel değerlendirme yöntemiyle elde edilen NBD'nin negatif sonuç vermesi, bunun yanı sıra reel opsiyonlar değerlendirme yöntemleri Black-Scholes ile opsiyon ve proje değerinin pozitif sonuç vermesi ile birlikte, reel opsiyonlar değerlendirme yönteminin geleneksel yöntemlere göre üstünlüğü ve uygulanabilirliğinin gösterilmek istenmesidir.

2. Materyal ve Yöntem

Çalışmanın bu bölümünde RES için hem geleneksel yöntemlerden NBD ile proje değeri hem de reel opsiyonlar değerlendirme yöntemlerinden Black-Scholes ile opsiyon ve opsiyon ile NBD'nin toplanarak elde edilmesiyle birlikte GNBD hesaplanacaktır. Projenin uygulanabilir olup olmadığı geleneksel yöntemde proje değeri ile belirlenirken, reel opsiyonlarda opsiyon değeri pozitif olsa dahi GNBD'nin pozitif sonuç vermesi

gerekmektedir. Bu şartlar doğrultusunda aşağıdaki bölümlerde değerlendirme yöntemleri ve kullanılan parametre ve eşitlikler detaylı bir şekilde açıklanmaktadır.

2.1. Net Bugünkü Değer (NBD) Yöntemi ile Değerleme
NBD'nin hesaplanmasının ilk aşaması projeye ait parametrelerin belirlenmesidir. Sonra ki adım İNA'nın hesaplanarak 15 yıllık net nakit akımlarının hesaplanması ve devamında iskonto oranınca indirgenerek bugünkü değerlerin oluşturulmasıdır. Bugünkü değerlerin toplamının öngörülen proje maliyetinden çıkartılması ile de proje değerini veren NBD hesaplanmaktadır. Proje parametreleri Tablo 1'de gösterilmektedir.

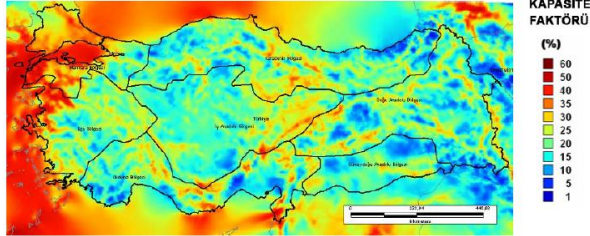
Tablo 1. Proje parametreleri (NBD).

Proje Parametreleri	
Proje Maliyeti	11.000.000 USD=70.400.000 TL
Projenin Yıllık Gelir Miktarı	7.064.640 TL
Ekonomik Ömrü	15 yıl
Kurulu Gücü	5 MW
AOSM	%10
Türbin Sayısı	2
Operasyonel Gider	306.000 USD=1.958.400 TL
Elektrik Satış Fiyatı (TL/ kWh)	(0,073*6,40TL) =0,4672 kuruş
Amortisman Değeri	4.693.333,33 TL
Kurumlar Vergisi (KV)	%25=0,25
TÜFE (Enflasyon Oranı)	%9,90=0,099
Birim Elektrik Fiyatı	0,073USD*6,40TL/USD=0,4672TL

Tablo 1'de gösterilen proje parametrelerinin hesaplanması projenin ekonomik analizi kısmında belirlenmektedir. Bir rüzgâr enerji santralinin proje gelirinin hesaplanabilmesi için öncelikle tesisin ürettiği enerji kapasitesinin bilinmesi gerekmektedir. Üretim miktarının hesaplanabilmesi için de RES'in ürettiği enerji miktarının ve santralin kurulduğu bölgeye ait rüzgâr kapasite faktörünün bilinmesi şarttır. Kapasite faktörü bölgede yapılan 1 senelik rüzgâr ölçümlerinin haricinde Rüzgâr Enerji Potansiyeli Atlası (REPA) da yatırımcılara bu bilgiyi vermektedir. Şekil 1'de Türkiye geneli 50m yükseklikteki ortalama kapasite faktörü (%) dağılımı gösterilmiştir (Çalışkan, 2020). REPA'ya göre Sino pili kapasite faktörünün %35 olduğu görülmektedir. Üretim miktarı için; 5 MW*0,35*360 gün*24 saat=15.120.000 kWh hesaplanmıştır. Burada 1 yılın 360 gün olarak alınmasının nedeni yılda 5 günün bakım onarım ile geçebileceği ve bu zaman diliminde üretimin duracağına öngörülmesindedir.

Bir diğer bilinmesi gereken parametre ise elektrik

fiyatıdır. Yenilenebilir Enerji Kaynakları Destekleme Mekanizması (YEKDEM) tarafından düzenlenen cetvellerdeki teşvik fiyatları, enerji üretim tesislerinin elektrik fiyatlarını tanımlamaktadır. RES' in 2021 yılı haziran ayı içerisinde uygulamaya girdiği düşünüldüğünde, Ek-1 sayılı cetvelinde yer alan, üretilecek elektriğin 1kWh'nin 7,3 UScent/kWh teşvik fiyatı üzerinden değerlendirilmeye alınacağı belirlenmiştir.



Şekil 1. REPA kapasite faktörü dağılımı (%) (Çalışkan, 2020).

Burada rüzgâr enerji santralleri için bilmemiz gereken bir diğer parametre değeri ise dolar kurudur. Bu çalışmada son yıllarda yüksek değerlerde dalgalanan dolar kuru üzerinden ortalama bir değer kullanmak yerine sabit bir dolar kuru kullanmanın hesaplamalarımız için daha doğru olacağına karar verilmiştir. Bunun için de T.C. Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı'nın 2020-2022 Dönemi Yatırım Programı Hazırlama Rehberi'nin 12. Maddesinde belirtilen ve yatırım kararı alınan 2021 yılı sabit kur fiyatı, 1ABD=6.40 TL'dir (SBB, 2021). Buradan proje geliri $15.120.000\text{kWh} \times 0.073\text{USD} \times 6,40\text{TL/USD} = 7.064.064\text{ TL}$ olarak hesaplanmıştır. İndirgenmiş nakit akışları hesaplanırken birim elektrik fiyatı, $0.073\text{USD} \times 6,40\text{TL/USD} = 0,4672\text{ TL}$ olarak hesaplanmıştır. RES projesinin gider unsurlarını, ilk yatırım maliyeti harcamaları ve sonrasında yatırımın sürdürülebilirliğinin sağlanabilmesi için yıllık operasyonel giderler olarak ikiye ayırmak mümkündür. Bu proje de ilk yatırım maliyeti 11.000.000 \$ yani 70.400.000 TL, operasyonel giderler için de 306.000 \$ yani 1.958.400 TL olarak belirlenmiştir. Vergi Usul Kanunu'nun 333 sıra numaralı tebliğinin 45. Bölümünde bahsi geçen türbinlerle ilgili bölüme istinaden rüzgâr enerji santrallerinin faydalı ekonomik ömürlerinin 15 yıl olduğu belirtilmektedir. Bu bilgi doğrultusunda bu projede ekonomik ömür (T) 15 yıl olarak alınmıştır. Ekonomik ömür süresinin 15 yıl olması ve proje için normal amortisman oranı kullanılacak olmasından dolayı (1/n), bu projede ki amortisman oranı $(1/15) \cong 0,067 \cong \%6,7$ olarak hesaplanmıştır. Kurumlar Vergisi (KV) oranı olarak ise, Kurumlar Vergisi Kanunu'na eklenen geçici 13. Madde uyarınca "%20 olan kurumlar vergisi oranı 2021 yılı için %25 oranınca uygulanacaktır" ifadesi üzerine, %25 olarak belirlenmiştir. Projede enflasyon oranı olarak, ticari yatırımlar için tercih edilen Tüketici Fiyat Endeksinin (TÜFE) kullanılması uygun görülmüştür. Bunun için

2007-2021 yılları arasındaki 15 yıllık ve her yılın her ayı için "on iki aylık ortalamalara göre değişim oranı" kullanılarak, en son ortalama değer bulunmuştur. Projede hem gelirlerin hem de operasyonel giderlerin yıllık artış oranı olarak kullanıldığı TÜFE oranı %9,90 olarak hesaplanmıştır. Reel indirgeme oranı olarak kullanılan iskonto oranı için herhangi bir şirket esas alınmadığından ülke taban primi ile AOSM hesaplamak ya da Türkiye Cumhuriyeti Merkez Bankası'nın (TCMB) yayınladığı altı aylık iskonto faiz oranlarını dalgalanmalardan dolayı kullanmak yerine, %10'luk genel bir iskonto oranı tercih edilmiştir. NBD için kullanılan formül "Eşitlik 1'de gösterilmektedir (Bilir, 2012).

$$A = \sum_{i=1}^n \left(\frac{A_i}{(1+r)^i} \right) - C \quad (1)$$

A, nakit girişlerinin şimdiki değeri toplamını; A_i , yatırımdan sağlanacak net nakit akımları; C, başlangıçtaki yatırımı; r, iskonto oranını; i, yılı ifade etmektedir.

2.2. Black-Scholes Yöntemi ile Değerleme

Reel opsiyonlarla değerlendirme yöntemlerinden Black-Scholes, analitik bir yöntem olmakla birlikte hesaplama ve çözümleme kolaylığı açısından kullanımı en rahat yöntemdir. Black-Scholes'un 1973 yılında opsiyonların fiyatlanması için önerdiği ve o tarihten sonra gerçek yatırım projeleri için de kullanılabilen model aşağıdaki "Eşitlik 2, 3, 4 ve 5"te gösterildiği şekildedir.

$$C_0 = S_0 N(d_1) - X e^{-r_f T} N(d_2) \quad (2)$$

$$P_0 = X e^{-r_f T} N(-d_2) - S_0 N(-d_1) \quad (3)$$

$$d_1 = \frac{\ln(S_0 / X) + (r_f + \sigma^2 / 2)T}{\sigma \sqrt{T}} \quad (4)$$

$$d_2 = d_1 - \sigma \sqrt{T} \quad (5)$$

Modelde kullanılan terimler C_0 , alım opsiyonu değerini; P_0 , satım opsiyonu değerini; S_0 , dayanak varlığın değerini; X, kullanım fiyatını; T, opsiyonun vadesine kalan süreyi (yıl bazında); σ , dayanak varlığın standart sapmasını (volatilesi); r_f , risksiz faiz oranını; $N(d)$, kümülatif standart normal dağılımı ifade etmektedir. Black-Scholes değerlendirme yöntemi için kullanılan parametreler Tablo 2'de gösterilmektedir.

Tablo 2. Proje parametreleri (Reel opsiyonlar)

Girdi Parametreleri			
1	Yıllık Risksiz Faiz Oranı	r_f	%12,28=0.1228
2	Varlığın Şimdiki Değeri	S	60.810.500,15
3	Varlığın Kullanım Fiyatı	X	70.400.000,00
4	Opsiyonun Ekonomik Ömrü	T	15 yıl
5	Yıllık Standart Sapma	σ	0,4077
7	Yıllık Dönemler	Δt	1

Varlığın şimdiki değerini, NBD hesabında iskonto oranınca indirgenen bugünkü değerlerin toplamı oluşturmaktadır. Varlığın kullanım fiyatını, RES için öngörülen yatırım maliyeti oluştururken, risksiz faiz oranı içinde 2021 yılı için 10 yıllık tahvil oranı kullanılmıştır. Reel opsiyonlarla değerlendirme yöntemlerinde bilinmesi en zor parametre, değişkenliği ifade eden volatilité parametresidir. Bu çalışmada, RES yatırım projesini en çok etkileyen faktörlerinden biri olan birim elektrik fiyatı üzerinden volatilité değeri belirlenmiştir. Her yıl enflasyon oranınca artan birim elektrik fiyatının standart sapmasının bulunmasıyla elde edilmiştir. Tablo 3'te yıllara ait elektrik birim fiyatı ve

hesaplanan volatilité değeri gösterilmektedir.

Tablo 3'te hesaplanan volatilité değeri için kullanılan standart sapma formülü "Eşitlik 6" da, formüle göre volatilité değerinin hesaplanması için çözüm de "Eşitlik 7"de, aşağıda gösterildiği şekildedir (Brigham, 1995).

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (6)$$

$$\sigma = \sqrt{\frac{2.3265}{14}} = \sqrt{0.1662} = 0.4077 \quad (7)$$

Tablo 3. Volatilité hesabı

	Birim Elektrik Fiyatı (Y_i)	Ortalama (\bar{Y})	$(Y_i - \bar{Y})$	$(Y_i - \bar{Y})^2$
1.Yıl	0,4672	0.9818	(-0.5146)	0.2648
2.Yıl	0,5135	0.9818	(-0.4683)	0.2193
3.Yıl	0,5643	0.9818	(-0.4175)	0.1743
4.Yıl	0,6201	0.9818	(-0.3617)	0.1308
5.Yıl	0,6815	0.9818	(-0.3003)	0.0902
6.Yıl	0,7490	0.9818	(-0.2328)	0.0542
7.Yıl	0,8232	0.9818	(-0.1586)	0.0252
8.Yıl	0,9047	0.9818	(-0.0771)	0.0059
9.Yıl	0,9942	0.9818	(0.0124)	0.0002
10.Yıl	1,0927	0.9818	(0.1109)	0.0123
11.Yıl	1,2008	0.9818	(0.2190)	0.0480
12.Yıl	1,3197	0.9818	(0.3379)	0.1142
13.Yıl	1,4504	0.9818	(0.4686)	0.2196
14.Yıl	1,5939	0.9818	(0.6121)	0.3747
15.Yıl	1,7517	0.9818	(0.7699)	0.5928
Toplam $(Y_i - \bar{Y})^2$		2,3265		
σ		0,4077		

3. Bulgular ve Tartışma

Geleneksel değerlendirme yöntemi ile reel opsiyonlarla değerlendirme yöntemi arasındaki farkları ve üstünlükleri göstermek için yapılan bu çalışma da hesaplanan İNA ve NBD Tablo 4 ve Tablo 5'te gösterilmektedir.

Tablo 5'te görüldüğü üzere RES yatırımına ait NBD - 9.589.499,85 TL olarak negatif değer olarak, projenin bu piyasa şartları altında uygulanabilir olmadığı sonucunu vermiştir. Yatırımın bundan sonra ki aşamasında ekonomik ömür süresince piyasada ki risk ve belirsizliklerin de projeye dahil edildiği reel opsiyonlarla değerlendirme yöntemlerinden Black-Scholes yöntemi ile proje değeri hesaplanmıştır. Tablo 2'de verilen parametre değerlerine göre, 15 yıl vadeli bir RES projesine bugün başlanması durumunda ve Black-Scholes yöntemine göre yapılan hesaplamalar "Eşitlik 8"de gösterildiği gibidir.

$$d_1 = \frac{\ln(60.810.500,05 / 70.400.000) + (0.1228 + (0.4077)^2 / 2)15}{0.4077\sqrt{15}} \quad (8)$$

$$d_1 = 1.8633$$

$$d_2 = 1.8633 - 0.4077\sqrt{15}$$

$$d_2 = 0.2843$$

Hesaplamalar sonucu elde edilen d_1 ve d_2 parametrelerinin, Ek 1'de gösterilen kümülatif normal standart dağılım tablosundan elde edilen değerleri $N(d_1)=0,9688$ ve $N(d_2)=0,6119$ olarak bulunmuştur. Bu durumda opsiyon değeri "Eşitlik 9" da ki gibidir;

$$C_0 = (60.810.500,15 * 0.9688) - (70.400.000 * e^{(-0.1228 * 15)} * 0.6119) \quad (9)$$

$$C_0 = 52.085.382,91$$

Görüldüğü üzere Black-Scholes değerlendirme yöntemiyle hesaplanan hem opsiyon değeri pozitif hem de geleneksel yöntemle bulduğumuz negatif sonuçlu-9.589.499,85 TL NBD'den büyük çıkmıştır. Geleneksel NBD yöntemiyle elde edilen negatif değer sonucunda, projenin uygulanabilirliği onaylanmamış olup, opsiyon değeriyle birlikte yeni proje değerinin yani GNBD'nin aşağıda gösterildiği gibi pozitif çıkması sağlanarak, RES yatırımının uygulanabilirliğine karar verilmiştir. "Eşitlik 10" da genel GNBD'nin formülü, "Eşitlik 11"de de bu çalışmada ki RES'e ait GNBD gösterilmektedir.

$$\text{Statik NBD+Opsiyon değeri} \quad (10)$$

Tablo 4. İndirgenmiş nakit akışları (İNA) analizi

	Üretim Miktarı(kWh)	Birim Fiyat/TL	Gelir/TL	Giderler/TL	Fvök/TL	Amortisman (-)/TL	Fvök/TL	Vergi %25/TL	Net Faaliyet Kârı/TL
1.Yıl	15.120.000	0,4672	7.064.064	1.958.400	5.105.664	4.693.333,33	412.330,67	103.082,67	309.248,00
2.Yıl	15.120.000	0,5135	7.763.406,34	2.152.281,60	5.611.124,74	4.693.333,33	917.791,41	229.447,85	688.343,56
3.yıl	15.120.000	0,5643	8.531.983,56	2.365.357,48	6.166.626,08	4.693.333,33	1.473.292,75	368.323,19	1.104.969,56
4.yıl	15.120.000	0,6201	9.376.649,94	2.599.527,87	6.777.122,07	4.693.333,33	2.083.788,74	520.947,18	1.562.841,56
5.yıl	15.120.000	0,6815	10.304.938,28	2.856.881,13	7.448.057,15	4.693.333,33	2.754.723,82	688.680,95	2.066.042,87
6.yıl	15.120.000	0,7490	11.325.127,17	3.139.712,36	8.185.414,82	4.693.333,33	3.492.081,48	873.020,37	2.619.061,11
7.yıl	15.120.000	0,8232	12.446.314,76	3.450.543,88	8.995.770,88	4.693.333,33	4.302.437,55	1.075.609,39	3.226.828,16
8.yıl	15.120.000	0,9047	13.678.499,92	3.792.147,73	9.886.352,19	4.693.333,33	5.193.018,86	1.298.254,71	3.894.764,15
9.yıl	15.120.000	0,9942	15.032.671,41	4.167.570,35	10.865.101,06	4.693.333,33	6.171.767,73	1.542.941,93	4.628.825,80
10.yıl	15.120.000	10,927	16.520.905,88	4.580.159,82	11.940.746,06	4.693.333,33	7.247.412,73	1.811.853,18	5.435.559,55
11.yıl	15.120.000	12,008	18.156.475,56	5.033.595,64	13.122.879,92	4.693.333,33	8.429.546,59	2.107.386,65	6.322.159,94
12.yıl	15.120.000	13,197	19.953.966,64	5.531.921,61	14.422.045,03	4.693.333,33	9.728.711,70	2.432.177,92	7.296.533,78
13.yıl	15.120.000	14,504	21.929.409,34	6.079.581,85	15.849.827,49	4.693.333,33	11.156.494,16	2.789.123,54	8.367.370,62
14.yıl	15.120.000	15,939	24.100.420,86	6.681.460,45	17.418.960,41	4.693.333,33	12.725.627,08	3.181.406,77	9.544.220,31
15.yıl	15.120.000	17,517	26.486.362,53	7.342.925,03	19.143.437,50	4.693.333,33	14.450.104,17	3.612.526,04	10.837.578,13

Tablo 5. Net bugünkü değer (NBD) tablosu

	Net Kâr + Amortisman	Yatırım Tutarı	Net Nakit Akım	İndirgeme Oranı	Bugünkü Değer	Net Bugünkü Değer
Başlangıç		-70.400.000	-70.400.000		-70.400.000	
1.Yıl	5.002.581,33		5.002.581,33		4.547.801,21	
2.Yıl	5.381.676,89		5.381.676,89		4.447.666,85	
3.yıl	5.798.302,89		5.798.302,89		4.356.350,78	
4.yıl	6.256.174,89		6.256.174,89		4.273.051,63	
5.yıl	6.759.376,20		6.759.376,20		4.197.040,81	
6.yıl	7.312.394,44		7.312.394,44		4.127.656,03	
7.yıl	7.920.161,49		7.920.161,49		4.064.295,17	(60.810.500,15-70.400.000)=-9.589.499,85
8.yıl	8.588.097,48		8.588.097,48	%10,00	4.006.410,86	TL
9.yıl	9.322.159,13		9.322.159,13		3.953.505,49	
10.yıl	10.128.892,88		10.128.892,88		3.905.126,68	
11.yıl	11.015.493,27		11.015.493,27		3.860.863,19	
12.yıl	11.652.367,11		11.989.867,11		3.820.341,16	
13.yıl	13.060.703,95		13.060.703,95		3.783.220,71	
14.yıl	14.237.553,64		14.237.553,64		3.749.192,86	
15.yıl	15.530.911,46		15.530.911,46		3.717.976,72	

$$GNBD = -9.589.499,85 + 52.085.382,91$$

$$GNBD = 42.495.883,06$$

(11)

Her iki yöntemle de elde edilen sonuçlara bakıldığında reel opsiyonlarla yapılan değerlemenin geleneksel yöntemle göre üstünlüğü negatif çıkan proje değerinin pozitif çıkmasıyla görülmektedir. Bu durum sadece RES'ler için değil, diğer yenilenebilir enerji santralleri ve farklı yatırım projeleri için de değerlendirilmesi gereken bir durumdur. Hem geleneksel yöntemlerle hem de reel opsiyonlar yöntemleriyle değerlendirme de analitik ve simülasyonlar gibi farklı yöntemlerin kullanımına da literatürde sıkça rastlanmaktadır. Tablo 6'da reel opsiyonlar değerlendirme yöntemleri ile farklı yenilenebilir enerji santralleri için yapılan proje değerlendirme çalışmaları gösterilmektedir. Tablo'da değerlendirme üzerine geleneksel değerlendirme yöntemlerinin tercih edilmesi yerine farklı reel opsiyon değerlendirme yöntemlerinin kullanımının literatürde geniş bir kullanım alanına sahip olduğu da görülmektedir. Bu RES projesinde olduğu gibi geleneksel yöntemlerin yerine yatırımcıya daha esnek çözümler sunabilen, risk ve belirsizlik durumlarını göz ardı etmeden değerlemeye dahil eden reel opsiyonlar, yatırım için daha onaylanabilir ve uygulanabilir sonuçlar vererek, geleneksel yöntemlere oranla daha çok tercih edilebilir.

4. Sonuç

Yatırım değerlemesinde sıklıkla kullanılan geleneksel yöntemlerle negatif sonuç elde edilmesi ile birlikte, opsiyon değerinin yatırım hesaplamasında kullanılması projeyi onaylanabilir duruma getirmiştir. Birim elektrik fiyatının, enflasyon oranının bunlara bağlı olarak gelir miktarının ve maliyetlerin yıllara göre değişimi, 15 yıl içerisinde çeşitli risk ve belirsizlere maruz kalarak yatırım sonunda ki kâr miktarını değiştirebilen faktörlerdir. Bu durumun öngörülebilmesi için sabit bir iskonto oranı yerine risk ve belirsizler için risksiz faiz oranı, değişken durumlar içinde volatilité değerinin projeye dahil edilmesinin yatırıma daha sayısal olarak ta olumlu bir bakış açısı kazandırdığı görülmektedir. Bu proje de NBD'nin negatif çıktığı durum da, opsiyon değeriyle pozitif proje değerinin edildiği gösterilmiştir. NBD'si pozitif çıkan bir yatırım projesi içinde, opsiyon değerinin eklenmesiyle proje değeri yine daha kârlı olabilecek bir proje değeri verecektir. Buradan sadece negatif sonuçlu NBD elde edilen projeler için değil pozitif sonuçlu olabilecek yatırım projeleri için de reel opsiyonların kullanımının önemli olduğu belirtilmektedir. Bir başka sonuç olarak ta, geleneksel yöntemlerin büyük yatırımlı projelerin değerlendirilmesinde kullanımının yetersiz kaldığı ve belirsizliklere çözüm

sağlayacak reel opsiyonlar değerlendirme yöntemlerinin yatırımcı için daha çözüm odaklı ve olumlu sonuçlar verdiği sonuca varılarak, kullanımlarının akademik

çalışmalarda olduğu kadar yatırım değerlemede de kullanılması gerektiği sonucuna varılmıştır.

Tablo 6. Reel opsiyonlar değerlendirme yöntemlerinin yenilenebilir enerji tesislerinde ki kullanımı

Yazar	Yıl	Ülke	Proje Tipi	Değerleme Yöntemi
Anderson R. C., Weersink A.	2014	Amerika	Bioenerji	Black-Scholes
Bruno S., Ahmed S., Shapiro A., Street A.	2016	Brezilya	Hidroelektrik Rüzgâr Enerjisi	Black-Scholes
De Mare G., Manganelli B., Nesticò A.	2013	İtalya	Rüzgâr Enerjisi	Monte Carlo Simülasyonu Black-Scholes
Eissa M. A., Tian B.	2017	Mısır	Güneş Enerjisi	Monte Carlo Simülasyonu
Fleten S. E., Linnerud K., Molnár P., Nygård M. T.	2016	Norveç	Hidroelektrik Enerji	Binomial
Fleten S. E., Molnár P., Nygård M.T., Linnerud, K.	2016	Norveç	Hidroelektrik Enerji	Black-Scholes
Jeon C., Lee J., Shin J.	2015	Kore	Güneş Enerjisi	Monte Carlo Simülasyonu Black-Scholes
Kim K. T., Lee D. J., Park, S. J.	2014	Kore	Rüzgâr Enerjisi	Binomial
Kitzing L., Juul N., Drud M.	2017	Baltık Denizi	Rüzgâr Enerjisi	Black-Scholes
Li Y., Tseng C. L., Hu G.	2015	Amerika	Bioenerji	Black-Scholes
Monjas-Barroso, M., Balibrea-Iniesta, J.	2013	Danimarka Finlandiya	Rüzgâr Enerjisi	Binomial Black-Scholes
Sisodia G. S., Soares I., Ferreira P.	2016	İspanya Portekiz	Rüzgâr Enerjisi	Monte Carlo Simülasyonu Black-Scholes
Sisodia G. S., Soares I., Ferreira P., Banerji S., Prasad R.	2015	İspanya	Rüzgâr Enerjisi	Monte Carlo Simülasyonu Black-Scholes
Torani K., Rausser G., Zilberman D.	2016	Amerika	Güneş Enerjisi	Black-Scholes

Katkı Oranı Beyanı

Konsept: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Tasarım: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Denetim: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Veri toplama ve/veya işleme: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Veri analizi ve/veya yorumlama: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Kaynak taraması: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Yazım: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Eleştirel inceleme: D.B. (%40), F.A.S. (%30) ve P.K. (%30), Gönderim ve revizyon: D.B. (%40), F.A.S. (%30) ve P.K. (%30). Tüm yazarlar makalenin son halini incelemiş ve onaylamıştır.

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Destek ve Teşekkür Beyanı

Bu çalışma sorumlu yazarın "Reel Opsiyonlarla Yenilenebilir Enerji Tesislerinin Değerlemesi" başlıklı doktora tezinden üretilmiştir.

Kaynaklar

- Abadie LM, Chamorro JM. 2014. Valuation of wind energy projects: a real options approach. *Energies*, 7(5): 3218-3255.
- Anderson RC, Weersink A. 2014. A real options approach for the investment decisions of a farm-based anaerobic digester. *Canadian J Agri Econ*, 62: 69-87.
- Bilir H. 2012. Enerji yatırım projelerinin değerlendirilmesinde reel opsiyon yaklaşımı. Ankara Üniversitesi, Sosyal Bilimler Enstitüsü Doktora Tezi, Ankara, Türkiye, pp.30.
- Black F, Scholes M. 1973. The pricing of options and corporate liabilities. *J Polit Econ*, 81(3): 637-654.
- Brandao LE, Dyer JS, Hahn WJ. 2005. Using binomial decision trees to solve real-option valuation problems. *Decis Anal*, 2(2): 69-88.
- Brigham E. 1995. *Fundamentals of Financial Management*, 7th Ed., The Dryden Press, New York, US.
- Bruno S, Ahmed S, Shapiro A, Street A. 2016. Risk neutral and risk averse approaches to multistage renewable investment planning under uncertainty. *European J Oper Res*, 250(3): 979-989.
- Chambers N. 2005. Gerçek opsiyonların fiyatlandırılması. *Muhas Finans Derg*, 26: 70-80.
- Colwell D, Henker T, Ho J, Fong K. 2002. Real options valuation of Australian gold mines and mining companies. *J Alter Invest*, 6(1): 23-38.

- Çalışkan M. 2020. Türkiye rüzgâr enerji potansiyeli. URL: https://www.mgm.gov.tr/FILES/haberler/2010/retsseminer/2_Mustafa_CALISKAN_RITM.pdf (erişim tarihi: 20 Eylül 2021).
- De Mare G, Manganeli B, Nestico A. 2013. The economic evaluation of investments in the energy sector: a model for the optimization of the scenario analyses. *Comput Sci Its Appl*, 2013: 359-374.
- Dzyuma U. 2012. Real options compared to traditional company valuation methods: possibilities and constraints their use. *Financ Internet Quart*, 8(2): 51-68.
- Eissa MA, Tian B. 2017. Lobatto-Milstein numerical method in application of uncertainty investment of solar power projects. *Energies*, 10(1): 43.
- Fernandes B, Cunha J, Ferreira P. 2011. The use of real options approach in energy sector investments. *Renew Sust Energy Rev*, 15(9): 4491-4497.
- Fleten SE, Linnerud K, Molnar P, Nygard MT. 2016. Green electricity investment timing in practice: real options or net present value? *Energy*, 116(1): 498-506.
- Fleten SE, Molnar P, Nygard MT, Linnerud K. 2016. Green certificates and investments in small hydro power plants. 13th International Conference on the European Energy Market, 6-9 June 2016, Porto, Portugal, pp: 1-6.
- Jeon C, Lee J, Shin J. 2015. Optimal subsidy estimation method using system dynamics and the real option mode: photovoltaic technology case. *App Energy*, 142: 33-43.
- Kapucugil İ. A, Devci Kocakoç İ. 2009. Bilişim teknolojisi projelerinde reel opsiyonlar. *Dokuz Eylül Üniv Sos Bil Enst Derg*, 11(4): 17-51.
- Kim KT, Lee DJ, Park SJ. 2014. Evaluation of R&D investments in wind power in Korea using real option. *Renew Sustain Energy Rev*, 40: 335-347.
- Kitzing L, Juul N, Drud M, Boomsina TK. 2017. A real options approach to analyse wind energy investments under different support schemes. *App Energy*, 188: 83-96.
- Li Y, Tseng CL, Hu G. 2015. Is now a good time for Iowa to invest in cellulosic biofuels? A real options approach considering construction lead times. *Int J Prod Econ*, 167: 97-107.
- Monjas-Barroso M, Balibrea- Iniesta J. 2013. Valuation of projects for power generation with renewable energy: A comparative study based on real regulatory options. *Energy Pol*, 55: 335-352.
- Mun J. 2006. Real options analysis tools and techniques for valuing strategic investments and decisions. Second Edition, John Wiley & Sons, Inc., New Jersey, US, pp: 704.
- Özcan D. 2013. Muğla bölgesinde örnek bir rüzgâr elektrik santrali(res) yatırımı ve res yatırımında reel opsiyonların kullanımı üzerine bir inceleme. Yüksek Lisans Tezi, Muğla Sıtkı Koçman Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı Muğla, Türkiye, pp: 82.
- Safarov S. 2009. Yatırım projelerinin değerlendirilmesi reel opsiyon yöntemi ve enerji sektöründe bir uygulama. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı Finansman Programı, İzmir, Türkiye, pp: 98.
- SBB. 2021. 2020-2022 Dönemi Yatırım Programı Hazırlama Rehberi. URL: https://www.sbb.gov.tr/wp-content/uploads/2019/10/2020/2022_Donemi_Yatirim_Programi_Hazirlama_R ehberi.pdf (erişim tarihi: 01 Ekim 2021).
- Schulmerich M. 2010. Real options valuation the importance of interest. Second Ed., Springer, Munich, Germany, pp: 387.
- Sisodia GS, Soares I, Ferreira P, Banerji S, Prosad R. 2015. Projected business risk of regulatory change on wind power project: case of Spain. *Energy Procedia*, 79: 1054-1060.
- Sisodia GS, Soares I, Ferreira P. 2016. Modeling business risk: The effect of regulatory revision on renewable energy investments- The Iberian case. *Renew Energy*, 95: 303-313.
- Torani K, Rausser G, Zilberman D. 2016. Innovation subsidies versus consumer subsidies: A real options analysis of solar energy. *Energy Pol*, 92: 255-269.
- Uzunlar E, Aktan M. 2000. Finansal opsiyonlar, gerçek opsiyonlar ve uygulamaları. Gazi Kitabevi, Ankara, Türkiye, pp: 85.