



**GÜNCEL AKILLI OPTİMİZASYON
ALGORİTMALARIYLA DUYGU SINIFLANDIRILMASI**

Sinem AKYOL

**Doktora Tezi
Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Doç. Dr. Bilal ALATAŞ
KASIM-2018**

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**GÜNCEL AKILLI OPTİMİZASYON ALGORİTMALARIYLA DUYGU
SINIFLANDIRILMASI**

DOKTORA TEZİ

Sinem AKYOL

131129202

**Tezin Enstitüye Verildiği Tarih : 08 Ekim 2018
Tezin Savunulduğu Tarih : 09 Kasım 2018**

Tez Danışmanı : Doç. Dr. Bilal ALATAŞ (F.Ü.)
Diğer Jüri Üyeleri : Prof. Dr. A. Bedri ÖZER (F.Ü.)

Doç. Dr. Burhan ERGEN (F.Ü.)

Dr. Öğr. Üyesi Soner KIZILOLUK (M.Ü.)

Dr. Öğr. Üyesi Kenan İNCE (İ.Ü.)

KASIM-2018

ÖNSÖZ

Bireyler arasında kişisel ya da profesyonel ilişkilerin oluşturduğu ağ olarak adlandırılabilir olan sosyal ağlar, günlük yaşam tarzının önemli bir parçası haline gelmiş ve gittikçe önem kazanmaya başlamıştır. Kullanıcılar, sosyal ağlar üzerinden yeni iletişim ortamlarının sunduğu hemen hemen tüm özelliklerden yararlanmakta, sürekli bilgi paylaşımı yapmaktadırlar. Bilgi paylaşımının bu kadar yaygın hale gelmesi her geçen gün yeni sosyal ağların doğmasına yol açmaktadır. Sosyal ağlar eğitim, akademik, sağlık ve daha birçok alanda bireyler tarafından aktif bir şekilde kullanılmaktadır. Sosyal ağ kavramının bu kadar çok kullanılması sosyal ağları oluşturan yapıların incelenip analiz edilmesi konusunu beraberinde getirmiştir. Bu tez çalışmasında sosyal ağ analizi problemleri incelenmiş olup, en bilinenlerinden biri olan duygu analizi - fikir madenciliği ilk defa bir optimizasyon problemi olarak ele alınmıştır.

Doktora çalışmalarım boyunca, değerli görüş ve katkılarıyla beni yönlendiren, her konuda desteklerini hiçbir zaman esirgemeyen, çalışma disiplini bize kazandıran ve kıymetli tecrübelerinden faydalandığım tez danışmanım Sayın Doç. Dr. Bilal ALATAŞ'a teşekkürü bir borç bilirim.

Sinem AKYOL
ELAZIĞ-2018

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	II
İÇİNDEKİLER.....	III
ÖZET	V
SUMMARY	VI
ŞEKİLLER LİSTESİ	VII
TABLolar LİSTESİ	IX
KISALTMALAR.....	X
SEMBOLLER LİSTESİ	XI
1. GİRİŞ.....	1
1.1. Genel Bilgiler	1
1.2. Tezin Amacı	2
1.3. Tezin Yapısı	3
2. SOSYAL AĞ ANALİZİ PROBLEMLERİ.....	4
3. DUYGU ANALİZİ-FİKİR MADENCİLİĞİ.....	8
4. KULLANILAN AKILLI OPTİMİZASYON ALGORİTMALARI	13
4.1. Sosyal Etki Teorisi tabanlı Optimizasyon Algoritması	16
4.2. Balina Optimizasyonu Algoritması	19
4.2.1. Avın Çevrelenmesi	21
4.2.2. Hava Kabarcığı Saldırı Yöntemi	22
4.2.3. Av Arama	23
5. DENEYSEL ÇALIŞMADA KULLANILAN VERİ SETLERİ VE ÖN İŞLEMLER.....	26
5.1. KNIME’da Yapılan Ön İşlemler	26
5.1.1. Dosya Okuma (File Reader).....	27
5.1.2. Belge Oluşturma (Document Creation).....	27
5.1.3. Ön İşleme (Preprocessing)	28
5.1.4. Belge Vektörü (Document Vector)	30
5.1.5. Kategoriden Sınıfa (Category to Class).....	31
5.1.6. Sütun Filtreleme (Column Filter)	31
5.1.7. Bölümleme (Partitioning).....	31
5.1.8. ARFF Yazıcı (ARFF Writer)	31
5.2. Duygu Analizi için Kullanılan Veri Setleri	31

5.2.1.	IMDB Yorumları Veri Seti.....	32
5.2.2.	Polarity Veri Seti	33
5.2.3.	Amazon Veri Seti	35
6.	PROBLEMİN OPTİMİZASYON ALGORİTMALARINA UYARLANMASI	38
6.1.	Kelimelerin Ağırlıklandırılması	38
6.2.	Başlangıç Popülasyonu ve Kodlama	39
6.3.	Jaccard Benzerliği	40
6.4.	Uygunluk Fonksiyonu	41
7.	DENEYSEL SONUÇLAR.....	42
7.1.	IMDB Veri Setinden Elde Edilen Sonuçlar	43
7.2.	Polarity Veri Setinden Elde Edilen Sonuçlar	48
7.3.	Amazon Veri Setinden Elde Edilen Sonuçlar	54
7.4.	SETO ve BOA'nın Yakınsama Grafikleri.....	60
8.	SONUÇLAR VE TARTIŞMA	68
	KAYNAKLAR.....	70
	ÖZGEÇMİŞ.....	79

ÖZET

Sosyal ağlar, son yıllarda günlük yaşam tarzının önemli bir parçası haline gelmiştir ve gittikçe önem kazanmaktadır. Ağ analizi, bir sosyal sistemi incelemek için en iyi yolun sistem üyeleri arasındaki bağları incelemek olduğu varsayımıyla toplumsal yapıyı inceleyen bir yaklaşımdır. Günümüzde sağlık, eğitim, akademik gibi birçok alanda kullanılmaktadır. Sosyal ağların aktif bir şekilde kullanımı sosyal ağ analizi gereksinimi ortaya çıkarmıştır. En bilinen sosyal ağ analizi problemlerinden biri olan duygu analizi – fikir madenciliği insanların, ürünler, hizmetler, kuruluşlar, bireyler, konular, etkinlikler, başlıklar ve nitelikleri gibi varlıklara yönelik fikirlerini, duygularını, değerlendirmelerini, derecelendirmelerini, tutumlarını ve hislerini analiz eden çalışma alanıdır.

Bu tez çalışmasında duygu analizi ilk defa bir optimizasyon problemi olarak ele alınmıştır. Duygu analizi problemi için, güncel akıllı optimizasyon algoritmalarından Balina Optimizasyonu Algoritması ve Sosyal Etki Teorisi tabanlı Optimizasyon Algoritması uyarlanmıştır. Daha verimli sonuçlar alabilmek için Sosyal Etki Teorisi tabanlı Optimizasyon Algoritmasına hafıza özelliği eklenilmiştir. Çalışmada IMDB, Polarity ve Amazon olmak üzere üç adet veri seti kullanılmıştır. Sonuçların performanslarını değerlendirmek için literatürde en çok bilinen değerlendirme ölçütleri olan doğruluk yüzdesi, kesinlik, hassasiyet, F-Ölçütü ve MCC kullanılmıştır. Elde edilen sonuçlar otuz üç tane denetimli öğrenme algoritmalarıyla karşılaştırılmıştır. Sonuçlar incelendiğinde, metasezgisel optimizasyon algoritmalarının duygu analizinde başarılı sonuçlar verdiği görülmüştür.

Anahtar Kelimeler: Sosyal Ağ Analizi, Duygu Analizi, Fikir Madenciliği, Balina Optimizasyonu Algoritması, Sosyal Etki Teorisi tabanlı Optimizasyon Algoritması

SUMMARY

Sentiment Classification with Current Intelligent Optimization Algorithms

In recent years, social networks have become an important part of daily life and are becoming increasingly important. Network analysis is an approach that examines social structure with the assumption that the best way to examine a social system is to examine the links between actors. Nowadays, it is used in many areas such as health, education and academic. The active use of social network has revealed requirement of social network analysis. Sentiment Analysis - Opinion Mining which is one of the most well-known social network analysis problems, is a field of study that analyzes people's opinions, sentiments, evaluations, ratings, attitudes, and feelings about assets such as products, services, organizations, individuals, subjects, activities, titles, and qualifications.

In this thesis, sentiment analysis is considered as an optimization problem for the first time. Whale Optimization Algorithm and Social Impact Theory based Optimization Algorithm which are one of the current intelligent optimization algorithms, have been adapted for the sentiment analysis problem. Memory feature has been integrated into Social Impact Theory based Optimization Algorithm in order to obtain effective results. IMDB, Polarity, and Amazon datasets were used in this study. In order to evaluate the performance of the results, accuracy percentage, precision, recall, F-Measure, and MCC which are the five most commonly used measure in the literature were used. The results were compared with thirty-three supervised learning algorithms. When the results were examined, it was seen that metaheuristic optimization algorithms gave successful results in sentiment analysis.

Key Words: Social Network Analysis, Sentiment Analysis, Opinion Mining, Whale Optimization Algorithm, Social Impact Theory based Optimization Algorithm

ŞEKİLLER LİSTESİ

	<u>Sayfa No</u>
Şekil 4.1. Metasezgisel yöntemler	15
Şekil 4.2. SETO'nun sözde kodu	18
Şekil 4.3. BOA'nın sözde kodu	25
Şekil 5.1. KNIME'da yapılan işlemlerin akış şeması	27
Şekil 5.2. Belge oluşturma düğümleri.....	28
Şekil 5.3. KNIME'da ön işleme adımları	29
Şekil 5.4. IMDB yorumları veri setinin KNIME'daki ekran görüntüsü	32
Şekil 5.5. IMDB yorumları veri setinin doküman matrisi	33
Şekil 5.6. Polarity veri setinin KNIME'daki ekran görüntüsü	34
Şekil 5.7. Polarity veri setinden elde edilen doküman matrisi.....	35
Şekil 5.8. Amazon veri setinin KNIME'daki ekran görüntüsü.....	36
Şekil 5.9. Amazon veri setinden elde edilen doküman matrisi.....	37
Şekil 6.1. Kelimelerin ağırlıklandırılması.....	39
Şekil 7.1. Doğruluk yüzdesi ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar.....	44
Şekil 7.2. Kesinlik ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar.....	45
Şekil 7.3. Hassasiyet ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar	46
Şekil 7.4. F-Ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar.....	47
Şekil 7.5. MCC ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar	48
Şekil 7.6. Doğruluk yüzdesi ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar.....	50
Şekil 7.7. Kesinlik ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar	51
Şekil 7.8. Hassasiyet ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar.....	52
Şekil 7.9. F-Ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar.....	53
Şekil 7.10. MCC ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar	54
Şekil 7.11. Doğruluk yüzdesi ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar.....	56
Şekil 7.12. Kesinlik ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar.....	57

Şekil 7.13. Hassasiyet ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar	58
Şekil 7.14. F-Ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar.....	59
Şekil 7.15. MCC ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar	60
Şekil 7.16. SETO ve BOA'nın IMDB veri seti için yakınsama grafiği.....	61
Şekil 7.17. SETO ve BOA'nın Polarity veri seti için yakınsama grafiği	62
Şekil 7.18. SETO ve BOA'nın Amazon veri seti için yakınsama grafiği.....	63
Şekil 7.19. IMDB veri setinde SETO'nun farklı κ değerleri için yakınsaması	64
Şekil 7.20. IMDB veri setinde BOA'nın farklı p değerleri için yakınsaması.....	65
Şekil 7.21. Polarity veri setinde SETO'nun farklı κ değerleri için yakınsaması	65
Şekil 7.22. Polarity veri setinde BOA'nın farklı p değerleri için yakınsaması.....	66
Şekil 7.23. Amazon veri setinde SETO'nun farklı κ değerleri için yakınsaması	66
Şekil 7.24. Amazon veri setinde BOA'nın farklı p değerleri için yakınsaması.....	67

TABLÖLAR LİSTESİ

Sayfa No

Tablo 7.1. Sonuçları karşılaştırmada kullanılan ölçütler	42
Tablo 7.2. IMDB veri seti için kullanılan katsayılar	43
Tablo 7.3. Polarity veri seti için kullanılan katsayılar	49
Tablo 7.4. Amazon veri seti için kullanılan katsayılar	55



KISALTMALAR

ACO	: Karınca Koloni Algoritması (Ant Colony Algorithm)
ACROA	: Yapay Kimyasal Reaksiyon Optimizasyon Algoritması (Artificial Chemical Reaction Optimization Algorithm)
ALO	: Karınca Aslanı Optimizasyonu (Ant Lion Optimization)
BOA	: Balina Optimizasyon Algoritması (Whale Optimization Algorithm)
BP	: Geri Yayılım Algoritması (Back-Propagation Algorithm)
CSO	: Kedi Sürüsü Optimizasyonu (Cat Swarm Optimization)
DE	: Diferansiyel Gelişim (Differential Evolution)
ES	: Evrim Stratejisi (Evolution Strategy)
GA	: Genetik Algoritma (Genetic Algorithm)
GSA	: Yerçekimsel Arama Algoritması (Gravitational Search Algorithm)
HS	: Armoni Arama (Harmony Search)
ICA	: Emperyalist Yarışmacı Algoritma (Imperialist Competitive Algorithm)
IWD	: Zeki Su Damlacıkları (Intelligent Water Drops)
LCA	: Lig Şampiyonluğu Algoritması (League Championship Algorithm)
MCC	: Matthews Korelasyon Katsayısı (Matthews Correlation Coefficient)
POA	: Parlamenter Optimizasyon Algoritması (Parliamentary Optimization Algorithm)
PSO	: Parçacık Sürü Optimizasyonu (Particle Swarm Optimization)
SETO	: Sosyal Etki Teorisi Tabanlı Algoritma (Social Impact Theory based Optimization Algorithm)
WEKA	: Bilgi Analizi için Waikato Ortamı (Waikato Environment for Knowledge Analysis)

SEMBOLLER LİSTESİ

$\hat{I}_p$	: İkna edici etki
$\hat{I}_s$	: Destekleyici etki
$\vec{X}_{rand}$	: Gelişigüzel seçilen bir balinanın pozisyon vektörü
$\vec{A}, \vec{C}$	: Katsayı Vektörleri
D_j	: Doküman matrisindeki j . yorum satırı
N_s	: Aynı tutuma sahip bireylerin sayısı
N_o	: Karşıt tutuma sahip bireylerin sayısı
$\vec{X}$	: Pozisyon vektörü
X_i	: Popülasyondaki i . birey
c_1, c_2, c_3	: Uygunluk denkleminin katsayıları
d_i	: i . birey ile mevcut birey arasındaki mesafesi
f_{max}, f_{min}	: Popülasyondaki maksimum ve minimum uygunluk değerleri
k_1, k_2	: Benzerlik denkleminin katsayıları
p_i	: i . bireyin ikna ediciliği
q_i	: i . bireyin gücü
s_i	: i . bireyin destekleyiciliği
d	: Boyut sayısı
DN	: Doğru Negatif
DP	: Doğru Pozitif
YN	: Yanlış Negatif
YP	: Yanlış Pozitif
X^*	: Mevcut en iyi çözümün pozisyon vektörü
b	: Logaritmik spiralın şeklini tanımlamak için bir sabit
f	: Bireyin uygunluk değeri
p	: $[0, 1]$ arasında gelişigüzel bir sayı
$uygunluk$	: Bir bireyin uygunluk değeri

1. GİRİŞ

1.1. Genel Bilgiler

Bireyler arasında kişisel ya da profesyonel ilişkilerin oluşturduğu ağ olarak adlandırılabilir olan sosyal ağlar, günlük yaşam tarzının önemli bir parçası haline gelmiş ve gittikçe önem kazanmaya başlamıştır. Kullanıcılar, sosyal ağlar üzerinden yeni iletişim ortamlarının sunduğu hemen hemen tüm özelliklerden yararlanmakta, sürekli bilgi paylaşımı yaparak sosyal ağı veri sağlamaktadırlar. Bilgi değişiminin bu kadar yaygın hale gelmesi her geçen gün yeni sosyal ağların doğmasına yol açmaktadır. Sosyal ağlar eğitim, akademik, sağlık ve daha birçok alanda insanlar tarafından aktif bir şekilde kullanılmaktadır. Sosyal ağ kavramının bu kadar çok kullanılması sosyal ağları oluşturan yapıların incelenip analiz edilmesi konusunu beraberinde getirmiştir (Wellman, 1983; McGloin ve Kirk, 2010).

Sosyal ağ analizi problemlerinden en bilineni olan duygu analizi – fikir madenciliği; insanların, ürünler, hizmetler, kuruluşlar, bireyler, konular, etkinlikler, başlıklar ve nitelikleri gibi varlıklara yönelik fikirlerini, duygularını, değerlendirmelerini, derecelendirmelerini, tutumlarını ve hislerini analiz eden çalışma alanıdır. Günümüzde, Twitter, Facebook, mesaj panoları, bloglar ve kullanıcı forumları gibi sosyal medyadan gelen büyük bir ‘duygu’ patlaması olmaktadır. Bu metin parçaları, saygınlıklarını izlemek ve ürün ve eylemleri hakkında zamanında geri bildirim almak isteyen şirketler ve bireyler için bir altın madeni olmaktadır. Duygu analizi, bu organizasyonlara farklı sosyal medya sitelerini gerçek zamanlı olarak izleyebilmelerini ve buna göre hareket edebilmelerini sağlamaktadır. Pazarlama yöneticileri, halkla ilişkiler firmaları, kampanya yöneticileri, politikacılar, hatta sermaye yatırımcıları ve çevrimiçi alışverişçiler, duygu analizi teknolojisinin doğrudan yararlanıcılarıdır. Son yıllarda duygu analizi uygulamaları nedeniyle sanayi faaliyetlerinde gelişme gözlemlenmiştir. Duygu analizi uygulamaları, tüketici ürünleri, sağlık hizmetleri ve finansal hizmetler ile toplumsal olaylar ve siyasi seçimler de dâhil olmak üzere neredeyse her alana yayılmıştır. Bu pratik uygulamalar ve endüstriyel çıkarlar, duygu analizi araştırmaları için güçlü bir motivasyon sağlamıştır (Liu 2012; Feldman, 2013).

Metasezgisel algoritmalar, basit bir yaklaşımla, arama ya da optimizasyon problemlerinin çözüm tekniği olarak kullanan yapılardır. Son yıllarda gittikçe popüler olmaktadır ve güçlenmektedirler. Literatürde çok başarılı metasezgisel algoritmaların bulunmasına rağmen daima daha iyiyi arama ve bilimsel alanda sürekli iyileşme için yeni teknikler tasarlanıp geliştirilmiştir (Alataş, 2007; Cura, 2008; Akyol, 2013). Bu tez çalışmasında sürü tabanlı Balina Optimizasyon Algoritması (BOA) ve sosyal tabanlı Sosyal Etki Teorisi tabanlı Optimizasyon (SETO) duygu analizi problemi için uyarlanmıştır. BOA, kambur balinaların beslenme hareketinden esinlenerek geliştirilmiş bir metasezgisel algoritmadır (Mafarja ve Mirjalili, 2017). SETO ise gerçek, ima edilen veya hayal edilen bir varlığın veya diğer bireylerin eylemlerinin bir sonucu olarak bir bireyin davranışlarında, inançlarında, hislerinde, tutum ve düşüncelerinde meydana gelen değişikliklerden yola çıkılarak geliştirilmiş bir metasezgisel algoritmadır (Macas ve Lhotska, 2008).

1.2. Tezin Amacı

Bu tez çalışmasında, sosyal ağ analizi problemlerinden olan duygu analizi - fikir madenciliği ilk defa bir optimizasyon problemi olarak ele alınarak, başarılı bir duygu analizinin yapılabilmesi amaçlanmıştır. Duygu analizi çok amaçlı bir problem olarak ele alınmıştır. Doğruluk yüzdesi, kesinlik, hassasiyet, F-Ölçütü ve MCC ölçütlerinin hepsi için de iyi sonuçlar elde etmek hedeflenmiştir. Duygu analizi için BOA ve SETO olmak üzere iki farklı metasezgisel optimizasyon algoritması uyarlanmıştır. SETO'ya hafıza özelliği eklenerek daha iyi sonuçların alınması sağlanmıştır. Yapılan çalışmada üç farklı veri seti kullanılmıştır. Alınan sonuçlar 33 tane gözetimli makine öğrenmesi algoritmalarından alınan sonuçlar ile karşılaştırılmıştır. Karşılaştırma için literatürde en çok bilinen doğruluk yüzdesi, kesinlik, hassasiyet, F-Ölçütü ve MCC ölçütleri kullanılmıştır. Sonuçlara bakıldığında BOA ve SETO'nun iyi sonuçlar verdiği gözlemlenmiştir.

Bu tezin literatüre katkıları genel olarak şu şekilde özetlenebilir:

- Sosyal Ağ Analizi ve Duygu Analizi hakkında temel bilgiler sunmaktadır.
- Duygu Analizi problemi ilk defa optimizasyon problemi olarak ele alınmaktadır.
- Duygu analizi çok amaçlı bir problem olarak ele alınmıştır.
- Güncel akıllı optimizasyon algoritmalarından olan SETO ve BOA bu problem için ilk kez uyarlanarak duygu analizi veri setlerinde performansları test edilmiştir.

- SETO'ya hafıza özelliği eklenerek daha iyi sonuçlar elde edilmiştir.
- SETO ve BOA'dan elde edilen sonuçlar 33 tane gözetimli makine öğrenmesi algoritması ile doğruluk yüzdesi, kesinlik, hassasiyet, F-ölçütü ve MCC ölçütleri ile karşılaştırılması ilk kez bu tez çalışmasında yapılmıştır.

1.3. Tezin Yapısı

Bu tez çalışması yedi temel bölümden oluşmaktadır. Birinci bölümde, metasezgisel optimizasyon algoritmaları, sosyal ağ problemleri ve duygu analizi hakkında temel bilgiler verilmiştir ve daha sonra tezin amacı belirtilmiştir.

İkinci bölümde, optimizasyon ve metasezgisel algoritmalar hakkında daha detaylı bilgi verilmiştir ve bu tez çalışmasında probleme uyarlanan BOA ve SETO algoritmaları açıklanarak literatürde nerelerde kullanıldıkları hakkında bir çalışma yapılmıştır.

Üçüncü bölümde, sosyal ağ analizi hakkında bilgi verilmiştir ve literatürdeki diğer sosyal ağ analizi problemlerinden bahsedilmiştir.

Dördüncü bölümde, duygu analizi problemi hakkında detaylı bilgi verilmiştir. Literatürde duygu analizi için yapılan çalışmalar incelenmiştir.

Beşinci bölümde, deneysel çalışmalarda kullanılan veri setleri ve KNIME programı kullanılarak bu veri setleri üzerinde yapılan ön işlemlerden bahsedilmiştir.

Altıncı bölümde, problemin optimizasyon algoritmalarına nasıl uyarlandığı anlatılmıştır. Kelimelerin ağırlıklandırılması, başlangıç popülasyonunun oluşturulması ve kodlama hakkında bilgiler verilerek uygunluk fonksiyonu ve Jaccard değerinin nasıl hesaplandığı anlatılmıştır.

Yedinci bölümde, BOA ve SETO'nun uyarlanmasıyla elde edilen programın sonuçları 33 tane gözetimli makine öğrenmesi algoritmasından alınan sonuçlarla karşılaştırılarak verilmiştir. 50, 100 ve 200 iterasyon boyunca BOA ve SETO'nun uygunluk değerinin değişim grafikleri üç veri seti için de verilmiştir. Ayrıca BOA'da p , SETO'da ise κ parametrelerinin parametre analizleri yapılmıştır.

Sekizinci bölümde ise, bu tez çalışmasında yapılanlar ve elde edilen sonuçlar değerlendirilmiş, gelecekteki çalışmalar için önerilerde bulunulmuştur.

2. SOSYAL AĞ ANALİZİ PROBLEMLERİ

Sosyal ağ analizi, aktörler arasındaki ilişkilerin kalıpları üzerine odaklanmaktadır. Kaynakların kullanılabilirliğini ve bu aktörler arasındaki kaynak alışverişini incelemektedir (Scott, 1991; Wasserman & Faust, 1994; Wellman & Berkowitz, 1988; Haythornthwaite, 1996). Ağ analizi, bir sosyal sistemi incelemek için en iyi yolun sistem üyeleri arasındaki bağları incelemek olduğu varsayımıyla toplumsal yapıyı inceleyen bir yaklaşımdır. Sosyal yapı olarak tanımlanan bu toplumsal ilişkilerin modelinin davranışa etkileri olduğu varsayılmaktadır. Buna karşın, sosyal bilimciler geleneksel olarak aktör özelliklerini ve normlarını davranışın açıklayıcı değişkenleri olarak sınırlandırmışlardır. Böyle bir yaklaşımla varsayım, benzer niteliklere (örneğin, cinsiyet, sosyo-ekonomik durum) veya benzer normlara sahip olan bireylerin benzer şekilde davrandıkları ve bireyler arasındaki davranış farklılıklarının farklı nitelik ve normlarla açıklandığı şeklinde olmaktadır (Wellman ve Berkowitz 1988). Ayrıca, suç davranışı da dahil olmak üzere davranışa ilişkin en deneysel açıklamalar, bireylerin bağımsız, özerk birim olduğunu varsayan istatistiksel yöntemlere dayanmaktadır. Bununla birlikte teorik açıdan, sosyal ağ analizi düğümler arasındaki ilişkilere ve bağımlılıklara ve model ilişkilerden türetilen kısıtlamaların ve fırsatların nihai olarak davranışı nasıl etkilediğine odaklanmaktadır (Wellman, 1983; McGloin ve Kirk, 2010).

Sosyal ağ analizi problemlerinden bazıları şunlardır: Ayrım keşfi (Segregation discovery) problemi, minimum bir eşik değerinin üzerindeki bir referans ayrım indeksi için alt grafikleri ve alt grupları aramayı içermektedir. Örnek olarak azınlık gruplarının (gençler, yaşlılar, kadınlar) şirketlerin yönetim kurullarında ayrılması verilebilir (Baroni and Ruggieri, 2015). Sosyal istenmeyen mesaj (spam) göndericiler, geleneksel istenmeyen mesajlardan farklı olarak, sosyal medyadaki spamlar bazen karşılıklı rızalar olmaksızın kolaylıkla birbirleriyle bağlantı kurabilmektedir. Çok sayıda “insan” arkadaşını hızla biriktirerek, normal kullanıcıları taklit etmek için birbirleriyle sohbet etmektedir. Sosyal istenmeyen mesaj gönderici algılamada (Social spammer detection), spam gönderenler ve normal kullanıcılar arasında duyarlılık analizi yapılmaktadır (Hu ve arkadaşları, 2014). Kullanıcı tarama desenleri, daha önce belirli bir konuya ilişkin olarak tanımlanmış içerikten, ilgilenilen belirli bir konu ile ilgili tespit edilmemiş içeriğe işaret edebilir. İçerik tespitinde (Content detection), tarama desenlerine dayanarak, bir sosyal ağ üzerindeki diğer sayfaların belirli bir ilgi alanına göre içerik barındırma olasılığı belirlenmektedir (Kunal

Ve Panesar, 2012). Fikir liderleri algılama (Opinion leaders detecting), görüşleri fikir liderleri tarafından büyük ölçüde etkilenen kullanıcıları anlamayı hedeflemektedir (Zhang ve arkadaşları, 2016).

Gerçek zamanlı eğilim algılama (Real-time trend detection), sosyal paylaşım ağ sisteminin eğilimli bir nesneyi erken belirlemesine yardımcı olarak, kullanıcıların daha ilgi çekici bir ortam sağlayan eğilim verici nesneyle etkileşime girmesine izin vermektedir (Li ve arkadaşları, 2013). Dolandırıcılık algılama (Fraud Detection) problemi, çevrimiçi sitelerdeki yorumları göz önüne alarak, her bir yorum için; metnine, yazarına, hangi ürün için yazıldığına, yayın zaman damgasına ve yıldız derecesine göre, sahte yorumları belirlemeye çalışmaktadır (Akoglu ve arkadaşları, 2013). Söylenti yayılım algılama sistemi (Rumor propagation detection system), bazı bilgilerin bir söylenti olup olmadığını veya genel bir bilgi olup olmadığını ve bu yanlış bilgileri yayan hesapları belirlemektedir (Yang ve arkadaşları, 2016).

Geleneksel yakınlık tespit yöntemlerinin çoğu, konum sunucusu için aralarındaki göreceli uzaklığı hesaplayarak komşularını bulmada kullanıcıların lokasyonlarına ihtiyaç duymaktadır. Gizlilik koruma yakınlık tespiti yöntemi (Privacy-preserving proximity detection method), geleneksel yakınlık tespitindeki sunucunun güvenilir olması durumunda oluşabilecek gizlilik sızıntısı sorununu çözmek için kablosuz erişim noktalarına dayanarak yakınlık tespiti yapmaktadır (Chen ve arkadaşları, 2016). Çevrimiçi iletişimlerde siber suç tespiti (Cybercrime detection in online communications), bir kişi veya bir grup kullanıcı tarafından diğer kullanıcıları rahatsız etmek için bilgi ve iletişim teknolojilerini kullanan siber zorbalıkları belirlemektedir (Al-garadi ve arkadaşları, 2016). Bu probleme benzer diğer bir problem olan çevrimiçi sosyal ağ zorbalık algılama (Online social network bullying detection) sosyal medyada elektronik iletişim kullanarak doğaya zarar veren veya tehdit eden mesajlar göndererek bir kişiyi taciz eden veya hakaret eden siber zorbalıkların faaliyetlerini tespit etmektedir (Nandhini ve Sheeba, 2015).

Kullanıcılar, haber hikâyelerine yorum yazabilmekte, çevrimiçi olarak sağlanan tüm bilgilere ilişkin görüş sunabilmekte ve fikirlerini paylaşabilmektedir. İlgi madenciliğinin (interest mining) temel hedefi, ilgi alanı aynı kişileri aynı gruba toplamak olan sanal topluluklar oluşturmaktır (Mittal ve arkadaşları, 2016). Olay tespitinde (Event detection), temel varsayım sosyal medyadaki bazı olaylarla ilgili sözcüklerin, bir olay meydana geldiğinde kullanımında bir artış göstereceği varsayımdır. Dolayısıyla, bir etkinlik, geleneksel olarak, görünüş sayısında patlama gösteren birkaç anahtar kelime ile temsil

edilmektedir (Wang ve arkadaşları, 2013). Rol madenciliği (Role mining), güvenlik sistemleri koleksiyonunun erişim denetimi bilgilerinden roller üretmek için veri madenciliği tekniklerinin kullanılmasıdır (Kuhlmann ve arkadaşları, 2003; Colantonio ve arkadaşları, 2010). Kullanıcı yeri tahmini (User location prediction), kullanıcıların en önemli yerlere ziyaretlerinin doğrusal olmayan zaman serileri analizine dayanmaktadır. Kullanıcı yeri tahminine yeni bir yaklaşım olan Sonraki yer (NextPlace) ise, gelecekteki ziyaretlerin süresini ve bu yerlerde beklenen bekleme süresini tahmin etmektedir (Scellato ve arkadaşları, 2011). Bağlam tahmini (Context prediction), önceki davranışına dayanarak gözlemlenen bir bağlam zaman serisinin ilerlemesi hakkında bilgi çıkarma görevidir (Sigg ve arkadaşları, 2012).

İşbirlikçi yazılım geliştirme, işin daha paralel hale getirilmesini ve daha hızlı tamamlanmasını sağlarken, geliştiricilerin tutarsız değişiklikler paylaşmasıyla ortaya çatışmalar çıkabilmektedir. Çatışma tahmini (Conflict prediction) veya çatışma algılama (Conflict detection) ile çatışmalar yönetilerek çözüm pratiği ortaya koyulmaktadır (Brun ve arkadaşları, 2012). Etki madenciliği (Influence mining), sosyal ağlardaki etkili düğümleri belirleyerek, bir ağın davranışını ve düğümler boyunca ne kadar bilginin yayıldığını ve benzer olarak ağa bilginin ulaşmasının ne kadar zaman alacağını incelemektedir (Hussain ve arkadaşları, 2013). Etkileşim madenciliği (Interaction mining) yapılandırılmamış müşteri etkileşimlerinden, yani çeşitli iletişim kanalları (örneğin, telefon görüşmeleri, e-posta, sosyal medya, kayıtlı toplantılar) aracılığıyla insanlar ve kuruluşlar arasında yapılan sohbetlerden zengin bilgi çıkarmaktadır (Pallotta ve arkadaşları, 2013). Duygu tespiti (emotion detection), sosyal medyada tüm sözcüklerin yazarları tarafından olumlu veya olumsuz duyguları hakkındaki bilgileri etkili bir şekilde okuyucuya iletmesini ve okuyuculardan karşılık gelen keyifli ya da acı verici bir tepki yaratmasını incelemektedir (Rao ve arkadaşları, 2014).

Konu madenciliği (Topic mining), kümeleme tabanlı tekniklerle haber dizilerindeki konuları bulmayı ve takip etmeyi amaçlayarak ortaya çıkarılmıştır (Wang ve arkadaşları, 2012). İtibar madenciliği (Reputation mining), sanal toplulukların ezici boyutlarından dolayı, forumlarda, yayınlanan çok sayıda spam ve önerge için üyelere kişilerin güvenilirliği hakkında bilgi çıkarmaktadır (Skopik ve arkadaşları, 2009). Müşteri hareketi tespiti (Churn detection), mevcut servis sağlayıcısından başka servis sağlayıcısına geçmek üzere olan müşterileri tespit ederek, müşteri sadakatini arttırmayı ve geçiş olayının olasılığını azaltmayı amaçlamaktadır (Birtolo ve arkadaşları, 2013). Anomali tespiti

(Anomaly detection), sosyal veya iletişim ağılarındaki anormalliklere ve grafiklerin varlıkları arasındaki bağlantılara odaklanmaktadır (Heard ve arkadaşları, 2010). Sosyal ağ değişim tespiti (Social network change detection), örgütsel yapılarında önemli değişiklikler meydana geldiğinde ve bunlara neden olan şeyleri belirlemede ağıları izleme sürecidir (Culloh ve Carley, 2008). Algı madenciliği (Perception mining), sosyal medya mesajlarından algılar çıkararak, toplu taşıma, acil durumlar ve politik eylemler ya da diğer faaliyetlere tepkiler gibi konularda değerli bilgiler sağlamaktadır (Doran ve arkadaşları, 2014).



3. DUYGU ANALİZİ – FİKİR MADENCİLİĞİ

Duygu analizi – fikir madenciliği; insanların ürünler, hizmetler, kuruluşlar, bireyler, konular, etkinlikler, başlıklar ve nitelikleri gibi varlıklara yönelik fikirlerini, duygularını, değerlendirmelerini, derecelendirmelerini, tutumlarını ve hislerini analiz eden çalışma alanıdır (Liu, 2012). İnsanların karar verme süreci, düşünce liderleri ve sıradan insanlar tarafından oluşturulan görüşlerden etkilenmektedir. Örneğin, bir birey çevrimiçi bir ürün satın almak istediğinde, başkaları tarafından çeşitli tekliflerde yazılan inceleme ve görüşleri arayarak başlamaktadır (Feldman, 2013).

Fikirler neredeyse tüm insan faaliyetlerinin merkezinde yer almaktadır; çünkü bunlar davranışların anahtar etkenidir. Bireyler karar verirken diğer bireylerin de görüşlerini önemsemektedirler. Gerçek dünyada, şirketler ve kuruluşlar her zaman ürün ve hizmetleri hakkında tüketici yorumlarını veya kamuoyunu bulmak istemektedirler. Bireysel tüketiciler, bir ürünün satın alınmadan önceki mevcut kullanıcılarının görüşlerini veya siyasi seçimlerde oy kullanmadan önce siyasi adaylarla ilgili diğer seçmenlerin görüşlerini bilmek istemektedirler. Geçmişte, bir kişi görüşe ihtiyacı olduğunda arkadaşlarına veya ailesine sormaktaydı. Benzer şekilde bir kuruluş veya bir işletme, kamu veya tüketici görüşlerine ihtiyaç duyduğunda anketler, kamuoyu yoklamaları ve odak gruplar düzenlenmekteydi. Kamuoyu ve tüketici görüşlerini edinmek; uzun zamandır pazarlama, halkla ilişkiler ve siyasi kampanya şirketleri için büyük bir ticaret olmuştur (Liu, 2012).

Günümüzde; Twitter, Facebook, mesaj panoları, bloglar ve kullanıcı forumları gibi sosyal medya ortamları insanların duygularını yansıtan çok fazla metin içermektedir. Bu metin parçaları, saygınlıklarını izlemek ve ürün ve eylemleri hakkında zamanında geri bildirim almak isteyen şirketler ve bireyler için bir altın madeni olmaktadır. Duygu analizi, bu organizasyonlara farklı sosyal medya sitelerini gerçek zamanlı olarak izleyebilmelerini ve buna göre hareket edebilmelerini sağlamaktadır. Pazarlama yöneticileri, halkla ilişkiler firmaları, kampanya yöneticileri, politikacılar, hatta sermaye yatırımcıları ve çevrimiçi alışverişçiler, duygu analizi teknolojisinin doğrudan yararlanıcılarıdır (Feldman, 2013).

Son yıllarda duygu analizi uygulamaları nedeniyle otomotiv, besin ve giyim gibi birçok endüstri faaliyetlerinde gelişme gözlemlenmiştir. Örneğin otomotiv endüstrisinde, şirketler pazarda rekabet avantajı elde etmek için tüketicilerin sosyal medyadaki görüşlerini analiz ederek pazarlama hedeflerini geliştirmeleri için bir yol sağlamaktadır.

Duygu analizi uygulamaları; tüketici ürünleri, hizmetler, sağlık hizmetleri ve finansal hizmetler ile toplumsal olaylar ve siyasi seçimler de dahil olmak üzere neredeyse her alana yayılmıştır. Bu pratik uygulamalar ve endüstriyel çıkarlar, duygu analizi araştırmaları için güçlü motivasyonlar sağlamıştır (Liu, 2012).

Genel bir duygu analizi sisteminin mimarisi şöyle verilebilmektedir; ilk önce sisteme yapılan girdi HTML, XML, PDF, Word vb. formatlarındaki belgelerin bir derlemesi olmaktadır. Bu belgeler metin haline dönüştürülerek; kökten çıkarma, simge haline getirme, varlık çıkarma ve ilişki çıkarma gibi çeşitli dil araçları kullanılarak ön işlemden geçirilmektedir. Sistemin temel bileşeni, duygu açıklamalarıyla önceden işlenmiş belgeleri açıklamak için dilsel kaynakları kullanan belge analizi modülü olmaktadır. Açıklamalar; tüm belgelere, bireysel cümlelere veya varlıkların belirli yönlerine eklenebilmektedir. Bu ek açıklamalar sistemin çıktısıdır ve çeşitli görselleştirme araçları kullanılarak kullanıcıya sunulabilmektedirler (Feldman, 2013).

Temel olarak duygu analizinin yapılabileceği üç seviye vardır: Belge (doküman) seviyesi; tüm belgeleri pozitif ve negatif olarak sınıflandırmaktadır ve yaygın olarak belge düzeyinde duygu sınıflandırması olarak bilinmektedir (Pang ve arkadaşları, 2002; Turney, 2002; Liu, 2012). En basit biçimdeki duyarlılık analizi olmakla birlikte gözetimli öğrenme ve gözetimsiz öğrenme olmak üzere iki temel yaklaşımı vardır. Cümle seviyesi, cümleleri pozitif, negatif veya nötr olarak yaygın olarak sınıflandırmaktadır. Cümle seviyesinde duygu sınıflandırma olarak bilinmektedir. Tek bir belge, aynı varlıklar hakkında bile çok sayıda görüş içeriyor olabilmektedir. Varlıklar hakkında belgede ifade edilen farklı görüşlerin daha iyi incelenebilmesi için cümle seviyesine geçilmektedir. Görünüm-özellik seviyesi; genel olarak görünüm duygu sınıflandırması olarak bilinen bu seviyede cümleler/belgeler görünümlerine dayalı olarak pozitif, negatif veya tarafsız olarak sınıflandırılmaktadır. Önceki iki yaklaşım, belgenin tamamı ya da her bir cümle tek bir varlığı belirttiğinde iyi çalışmaktadır. Bununla birlikte, pek çok durumda insanlar birçok özelliğe (özniteliklere) sahip varlıklardan bahsetmektedirler ve insanların her yönüyle ilgili farklı görüşleri olabilmektedir. Bu tarz durumlarda görünüm duygu sınıflandırması daha iyi çalışmaktadır (Feldman, 2013).

Gerçek hayatta kullanılan uygulamaların yanı sıra, birçok uygulamaya yönelik araştırma makaleleri de yayımlanmaktadır. Örneğin Pak ve Paroubek (2010) duygu analizi için günümüzdeki en popüler mikroblog platformlarından biri olan Twitter verilerini kullanmışlardır. Duyarlılık analizi ve görüş bildirme amaçlı belgelerin nasıl otomatik

olarak toplanabileceği gösterilmiştir ve toplanan belgelerin dilbilimsel analizini yapmışlardır. Belgeleri kullanarak pozitif, negatif ve nötr duyguları belirleyebilen bir duygu sınıflandırıcısı oluşturmuşlardır (Pak ve Paroubek, 2010). Esuli ve Sebastiani (2007), duygu sınıflandırması ve fikir madenciliği uygulamalarını desteklemek için halka açık bir sözcük kaynağı olan Sentiwordnet 1.0'ı geliştirmişlerdir (Esuli ve Sebastiani, 2007). Daha sonra ise yeni sürümü olan Sentiwordnet 3.0 geliştirilmiştir (Baccianella vd., 2010).

Eirinaki ve arkadaşları (2012), yalnızca bir belgenin/inceleminin genel düşüncelerini analiz etmekle kalmayıp, belirli bir duyarlılığa yol açan belgenin belirli bileşenlerinin semantik yönelimini tanımlayan bir algoritma sunmuşlardır. Algoritma, sonuçları genel tonuyla ve en önemli özelliklerin duygularının bir özetiyle birlikte bir soruna sonuç veren bir fikir arama motoruna entegre edilmiştir (Eirinaki vd., 2012). Jin ve Ho (2009), daha çok doğal dil işleme tekniklerine veya istatistik bilgilerine dayanan önceki yaklaşımlardan farklı olarak, sözcük üreten gizli Markov modelleri kullanıp yeni bir makine öğrenme çerçevesi önererek, bir ürünün müşteri yorumlarını incelemeyi ve son derece spesifik ürünlerle ilgili varlıkları çıkarmayı hedeflemişlerdir (Jin ve Ho, 2009). Grefenstette ve arkadaşları, niş tarama teknolojisini kavramanın ve analiz teknolojisini etkilemenin popüler basında kamuoyundaki kişilere verilen görüşte eğilimi ölçen bir uygulama geliştirmelerini sağladığını göstermişlerdir (Grefenstette vd., 2004). Nielsen, Twitter'dan alınan duygu için elle etiketlenmiş gönderileri kullanarak yeni bir ANEW (İngilizce Kelimeler için Duyarlı Normlar) önermiştir (Nielsen, 2011).

SemEval (Semantic Evaluation) hesaplamalı semantik analiz sistemlerinin devam eden bir dizi değerlendirmesidir. Bu değerlendirmeler dizisinin amacı semantik analiz sistemlerini değerlendirmektir. Metinlerin anlamsal analizinde konuları incelemek için bir mekanizma sağlamaktadır (Rosenthal vd., 2017). SemEval'deki ilgili görevlerden bazıları şunlardır: Pontiki ve arkadaşları ürün incelemesi ve görüşlerini (Pontiki vd, 2014, 2015, 2016), Ghosh ve arkadaşları Twitter'da figüratif dilin duygu analizini (Ghosh vd., 2015), Russo ve arkadaşları ima edilen olgu polaritesini (Russo vd., 2015), Mohammed ve arkadaşları tweetlerde tutumun saptanmasını (Mohammed, 2016), Kiritchenko ve arkadaşları kelime ve ifadelerin bağlam dışı duygu yoğunluğunu (Kiritchenko vd., 2016) araştırmışlardır.

Khan ve arkadaşları (2016) Twitter verilerine özgü bir sözdizim tabanlı yöntemi duygu analizini için kullanmışlardır. Daha sonra ise bir ikili duyarlılık sınıflandırıcısı, eğitim

verileri sözdizim tabanlı yöntemle sağlanmış olan yeni tanımlanmış tweetlere duygu polaritelerini atamak için eğitmişlerdir. Önerdikleri yöntem hassasiyet ve F-Ölçütünü önemli ölçüde geliştirmiştir. (Khan vd., 2016). Santos ve Gatti (2014), kısa metinlerin duygu analizini gerçekleştirmek için karakter-cümle düzeyinde bilgilerden yararlanan ve iki tane birbirinden bağımsız katman kullanan yeni bir derin konvolüsyonel sinir ağı önermişlerdir. Harften cümleye konvolüsyonel sinir ağı olarak adlandırdıkları önerilen ağ iki farklı alandaki iki şirket için uygulanmıştır ve yüksek doğruluklar elde edilmiştir (dosSantos ve Gatti, 2014). Gilbert (2014), genel duygu analizi için basit bir kural tabanlı model olan VADER'i sunmuştur ve bu modelin etkinliğini on bir tipik uygulama ölçütü ile karşılaştırmıştır. Karmaşık makine öğrenme teknikleri ile karşılaştırıldığında, bu modelin, doğruluktan ödün vermeden hem hız açısından hem de hesaplama açısından ekonomik olduğunu belirtmiştir (Gilbert, 2014).

Cambria ve arkadaşları (2014), grafik madenciliği ve boyutsal indirgeme teknikleri kullanmak yerine, genişletilmiş ortak ve sağduyulu bilgi sunumlarının çeşitli kısımlarını birbirine bağlamak için enerji akışını kullandıkları SenticNet 3'ü önermişlerdir. SenticNet 3, enerji tabanlı COGBASE sağduyu bilgi formalizmini temel alarak, 30,000 çok sözcüklü ifadeler için anlam ve sentezler sağlayarak doğal dil fikirlerinin daha derin ve çok yönlü analizini mümkün kılmışlardır (Cambria vd., 2014). Severyn ve Moschitti (2015), derin konvolüsyonel sinir ağı ile Twitter verilerinin duygu analizini yapmışlardır. Ağın parametrelerini üç aşamalı süreçte eğiterek daha başarılı sonuçlar elde etmişlerdir (Severyn ve Moschitti, 2015). Poria ve arkadaşları, video kliplerde insanların konuştuklarının duygu polaritesini belirlemek için söyledikleri sözcükleri, yüz ifadelerini ve ses tonunu kullanan yeni bir yöntem sunmuşlardır.

Poria ve arkadaşlarının, çalışmalarının temel yeniliği metinden özellikleri çıkarmak için derin konvolüsyonel sinir ağının kullanılması ve çok modlu heterojen karma özellikleri sınıflandırmada çoklu çekirdek öğrenmeyi kullanmalarıdır (Poria vd., 2015). Daha sonra ise, web videolarından elde edilen duyguları toplamak için çok yönlü duygu analizi yapan, ses, görsel ve metinsel yöntemler kullanan, bir model önermişlerdir. Önerilen modeli Youtube veri seti için uygulamışlardır (Poria vd., 2016). Kiritchenko ve arkadaşları (2014), hem kısa, gayri resmi metinsel mesajların duygularını tespit etmek, hem de bir kelimenin veya ifadenin içindeki duyguları tespit etmek için bir sistem önermişlerdir. Sistem çeşitli yüzey biçimleri, anlamsal ve duygusal özelliklerinden yararlanarak denetlenen bir istatistiksel metin sınıflandırma yaklaşımını dayanmaktadır. SemEval-2013 test kümesinde ve film

yorumları alıntılarında test edildiğinde karşılaştırıldığı algoritmalarından daha iyi sonuç verdiği gözlemlenmiştir (Kiritchenko vd., 2014).



4. KULLANILAN AKILLI OPTİMİZASYON ALGORİTMALARI

En iyileme anlamına gelen optimizasyon, belirli koşullarda bir problem için tüm çözümler arasından en iyisini bulma işidir. Optimizasyon problemi, belirli kısıtlamaları sağlamak şartıyla, bilinmeyen değişken değerlerinin bulunmasını hedefleyen herhangi bir problem olarak ifade edilebilmektedir (Murty, 2003). Birçok optimizasyon algoritması, amaç fonksiyonu ve sistem modeli için matematiksel modellere ihtiyaç duymaktadır ve matematiksel modellerin kurulması genellikle zor olmaktadır ya da yüksek maliyetli olduğundan dolayı kullanılamamaktadır. (Güden vd., 2005). Doğrusal olmayan ve büyük ölçekli tümleşik problemlerde ve tamsayı ya da ayrık karar parametrelerinin kullanıldığı birçok doğrusal optimizasyon modellerinde klasik optimizasyon yeterli olmamaktadır. Bu nedenle klasik optimizasyon algoritmaları, verilen bir probleme çözüm uyarlamada etkin olamamaktadır.

Çözüm yöntemi, klasik optimizasyon algoritmalarında daha çok problemi modellemede kullanılan değişkenlerin tipine (tamsayı, gerçel), amaç ve sınırlayıcıların tipine (doğrusal, doğrusal olmayan vb.) bağlı olurken, etkinliliği de benzer şekilde problem modellemede çözüm uzayı (içbükey, dışbükey vb.), karar değişken sayısı ve sınırlayıcı sayısına oldukça bağlıdır. Diğer önemli bir eksiklik ise; farklı tipte karar değişkenleri, amaç ve sınırlayıcıların olması durumunda problem formülasyonlarına uygulanabilecek genel çözüm stratejileri sunmamalarıdır. Yani çoğu algoritma belirli tipteki amaç fonksiyonu ya da sınırlayıcıların olduğu modelleri çözmektedir. Ancak çoğu yönetim bilimi, bilgisayar, mühendislik gibi bir çok farklı alandaki optimizasyon problemleri eşzamanlı olarak formülasyonlarında farklı tipteki karar değişkenleri, amaç fonksiyonu ve sınırlayıcıları gerektirir. Bu yüzden klasik ve genel amaçlı metasezgisel optimizasyon algoritmaları önerilmiştir. Bunlar son yıllarda oldukça popüler yöntemler haline gelmiştir. Çünkü, bu yöntemlerin hesaplama gücü iyidir ve dönüşümleri kolaydır. Yani tek amaç fonksiyonlu bir problem için yazılmış bir metasezgisel program, kolaylıkla çok amaçlı bir probleme ya da farklı bir probleme uyarlanabilmektedir (Akyol, 2013; Alataş, 2007).

Gerçek yaşam problemlerinin çoğunda problemin çözüm uzayı sonsuz veya tüm çözümlerin değerlendirilemeyeceği kadar büyük olur. Bunun için kabul edilebilir bir sürede çözümlerin değerlendirilerek iyi bir çözümün bulunması gerekmektedir. Böyle problemler için kabul edilebilir bir sürede çözümlerin değerlendirilmesiyle aslında tüm çözüm uzayında “bazı çözümlerin” değerlendirilmesi aynı anlama gelmektedir. Bazı

çözümlerin neye göre ve nasıl seçileceği metasezgisel tekniğe göre değişmektedir. Değerlendirmeye dahil olan çözümlerin içerisinde optimal çözümün yer alması garanti edilememektedir. Bu sebeple de metasezgisel tekniklerin bir optimizasyon problemine önerdiği çözüm, optimal değil iyi çözüm olarak algılanmalıdır (Cura, 2008).

Akıllı optimizasyon herhangi bir amacı gerçekleştirmek veya hedefe varmak için çeşitli alternatif hareketlerden etkili olanlara karar vermek amacı ile tanımlanan kriterler veya bilgisayar metotlarıdır. Bu algoritmalar, çözüm uzayında optimum çözüme yakınsaması ispat edilemeyen algoritmalar olarak da adlandırılır. Bu tür algoritmalar yakınsama özelliğine sahiptir, ama kesin çözümü garanti edemezler ve sadece kesin çözüm yakınındaki bir çözümü garanti edebilirler.

Metasezgisel algoritmalara gerek duyulmasının sebepleri şunlardır:

a) Optimizasyon problemi kesin çözümü bulma işleminin tanımlanamadığı bir yapıya sahip olabilir.

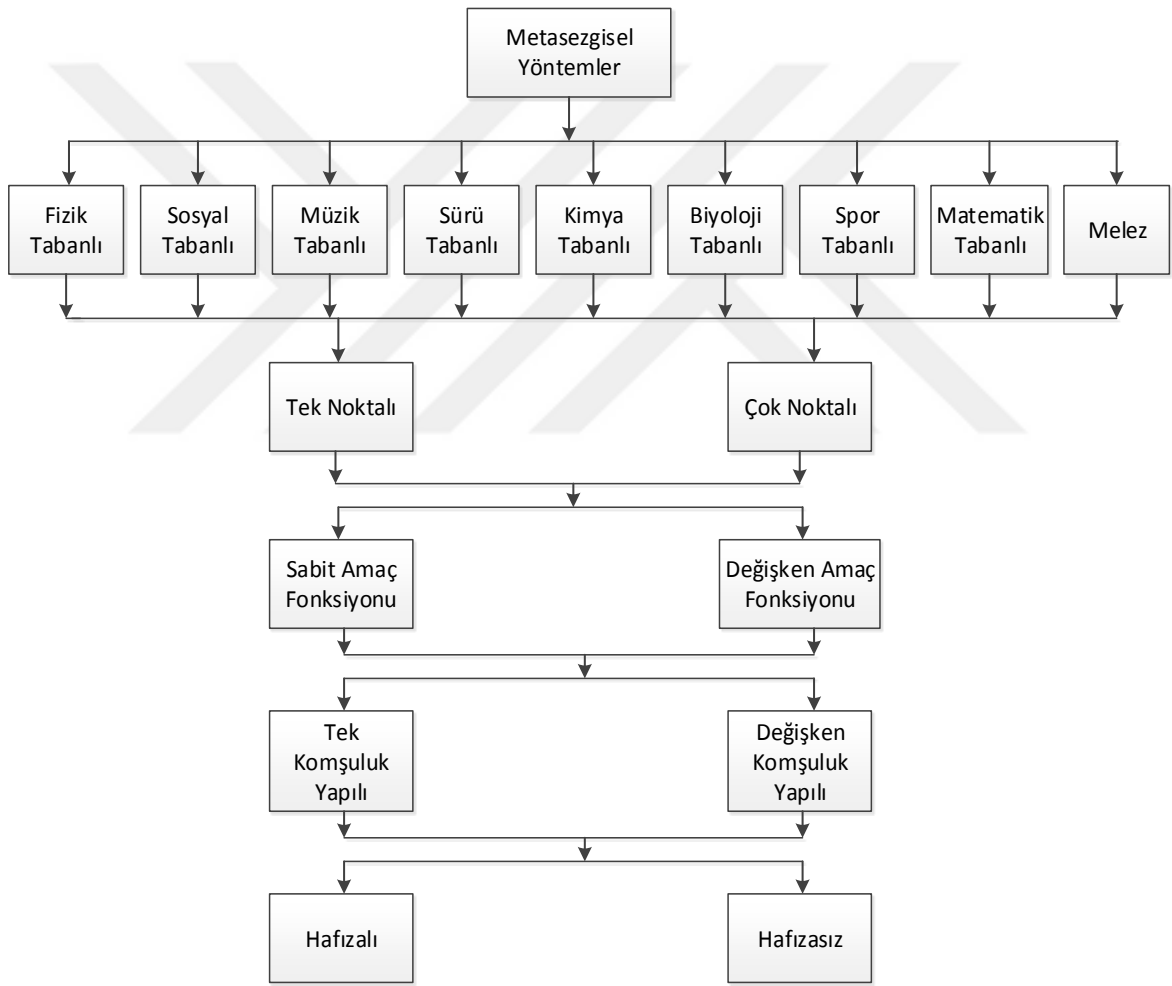
b) Anlaşılabilirlik yönünden metasezgisel algoritmalar karar vericiler için çok daha basit olabilir.

c) Metasezgisel algoritmalar, öğrenme amaçlı ve kesin çözümü bulma işleminin bir parçası olarak kullanılabilir.

d) Matematik formülleriyle yapılan tanımlamalarda genellikle gerçek dünya problemlerinin en zor tarafları, hangi amaçlar ve hangi sınırlamalar kullanılmalı, hangi alternatifler test edilmeli, problem verisi nasıl toplanmalı, ihmal edilir. Model parametrelerini belirleme aşamasında kullanılan verinin hatalı olması, metasezgisel yaklaşımın üretebileceği alt optimal çözümden daha büyük hatalara sebep olabilir (Akyol, 2013; Karaboğa, 2011).

Genel amaçlı metasezgisel yöntemler; biyoloji tabanlı, fizik tabanlı, sürü tabanlı, sosyal tabanlı, müzik tabanlı, kimya tabanlı, spor tabanlı ve matematik tabanlı olmak üzere sekiz farklı grupta değerlendirilebilmektedir. Ayrıca bunların birleşimi olan melez yöntemler de vardır. Bahsedilen bu yöntemler Şekil 4.1’de sunulmaktadır (Akyol, 2013; Alataş, 2007). Genetik Algoritma (GA) (Holland, 1975), Diferansiyel Gelişim Algoritması (DEA) (Storn ve Price, 1995) ve Karınca Koloni Algoritması (ACO) (Dorigo vd., 1991) biyolojik tabanlı; Emperyalist Yarışmacı Algoritma (ICA) (Atashpaz-Gargari ve Lucas, 2007) ve Parlamenter Optimizasyon Algoritması (POA) (Borji, 2007) sosyal tabanlı; Yapay Kimyasal Reaksiyon Algoritması (ACROA) (Alataş, 2011) kimya tabanlı; Armoni Arama

Algoritması (HS) (Geem vd., 2001) müzik tabanlı; Yerçekimsel Aarama Algoritması (GSA) (Rashedi vd., 2009) ve Zeki Su Damlacıkları Algoritması (IWD) (Shah-Hosseini, 2009) fizik tabanlı, Parçacık Sürü Optimizasyonu (PSO) (Kennedy ve Eberhart, 1995), Kedi Sürüsü Optimizasyonu (CSO) (Chu vd., 2006) sürü tabanlı, Lig Şampiyonluğu Algoritması (LCA) (Kashan, 2009) spor tabanlı ve Baz Optimizasyon Algoritması (Salem, 2012) ve Matheuristics (Maniezzo vd., 2009) Matematik Tabanlı Algoritma ve modellerdir. Kültürel Algoritma da hem biyoloji hem de sosyal tabanlı algoritma olarak sınıflandırılabilir (Akyol, 2013; Alataş, 2007).



Şekil 4.1. Metasezgisel yöntemler

4.1. Sosyal Etki Teorisi tabanlı Optimizasyon Algoritması

Sosyal etki, gerçek, ima edilen veya hayal edilen bir varlığın veya diğer bireylerin eylemlerinin bir sonucu olarak bir bireyin davranışlarında, inançlarında, hislerinde, tutum ve düşüncelerinde meydana gelen değişikliklerden herhangi biridir (Latane, 1981). Dinamik sosyal etki teorisi, inançların sosyal sistemler aracılığıyla yayılımını tanımlamaya ve tahmin etmeye çalışmaktadır. Bu teorinin, grup ve kişi arasındaki sosyal etkileşimi açıklamasında önemli olan üç temel unsuru bulunmaktadır:

- Kuvvet (güç), etkileyen bireylerin, etkilenen birey için ne kadar önemli olduğunu ifade etmektedir. Sosyal faktörün kuvveti arttıkça, sosyal etki daha fazla olacaktır.
- Yakınlık, etkileyici bireylerin uzamsal yakınlığını temsil etmektedir.
- Sayı, bir bireyin kaç birey tarafından etkilendiğini tanımlamaktadır (Macas ve Lhotska, 2008).

Andrzej Nowak, Jacek Szamrej ve Bibb Latane tarafından sosyal etki kavramına dayalı önemli bir fikir oluşumu sınıfı geliştirilmiştir (Macas vd., 2014). Bu yapı ilk olarak Andrzej Nowak ve Chris Szamrej tarafından 1990 yılında simüle edilmiştir (Nowak vd., 1990). Simülasyonlar her bir bireyi dört parametre olarak temsil etmektedir: bireyin tutumu, iki güç göstergesi (ikna edici ve destekleyici) ve sosyal yapı içindeki konumu.

Tutum (fikir), ikili parametredir ve yorumdan bağımsız olarak, iki değerden yalnızca bir tanesini alabilmektedir, örneğin evet veya hayır. Kuvvet, karşıt inançlara sahip olan bireylerin düşüncelerini değiştirmek için ikna etme ve benzer inançlara sahip olan bireyler içinse sosyal destek sağlayabilme becerileridir. Yakınlık kavramı ise, her bir hücrenin bir bireyi temsil ettiği bir kare matris olarak düşünülmüştür. İki bireyin yakınlığı, karşılık gelen hücreler arasındaki Öklid uzaklığı olmaktadır (Macas ve Lhotska, 2008).

Her bir iterasyon boyunca, Öklid uzaklığı kullanılarak aynı ve zıt tutumlu bireylerin sayısı ile iki tane sosyal etki hesaplanmaktadır. Birinci sosyal etki, farklı tutumdaki N adet bireyin tek bir birey üzerindeki toplam güç ve yakınlığı olmaktadır ve denklem (4.1)'de ifade edilmektedir.

$$\hat{I}_p = N_o^{1/2} [\sum (p_i/d_i^2)/N_o] \quad (4.1)$$

$\hat{I}_p$ ikna edici etki, N_o karşıt tutuma sahip bireylerin sayısı, p_i , i . bireyin ikna ediciliği ve d_i ise i . birey ile mevcut birey arasındaki mesafesi olmaktadır. Destekleyici etki ise denklem (4.2)'de ifade edilmektedir.

$$\hat{I}_s = N_s^{1/2} [\sum (s_i/d_i^2)/N_s] \quad (4.2)$$

$\hat{I}_s$ destekleyici etki, N_s aynı tutuma sahip bireylerin sayısı s_i , i . bireyin destekleyiciliği olmaktadır (Macas ve Lhotska, 2008).

Nowak ve arkadaşları (1990), bireylerin birbirlerinden farklı düşüncelere sahip olmalarının önemli etkilerini öne sürerek dinamik sosyal etki teorisini önermektedirler. Aradaki fark kısmen tutum ve mekânsal konumlandırma ve kısmen de güç faktöründedir. Güç terimi, bir bireyi etkili kılan tüm bireysel etkenlerin net etkisine atıfta bulunmaktadır. Orijinal simülasyonda, her birinin gücü, birbirinden bağımsız iki özellik olan ikna edici ve destekleyici tarafından temsil edilmektedir (Nowak vd., 1990).

Macas ve arkadaşları, bir sınıflandırıcı girdisinin optimum özellik alt kümesi seçimi için bu teoriyi modifiye ederek uygulamışlardır (Macas vd., 2008). Bhondekar ve arkadaşları, bu teoriyi bir empedans dilinin, iTongue, optimizasyonu için kullanmışlardır (Bhondekar vd., 2011). Kaur ve arkadaşları, bir elektronik koku algılayıcı performans artışı için SETO'yu ve hareketli pencere zaman dilimlemeyi kullanan yeni bir çok amaçlı pakete dayalı yaklaşım sunmaktadır (Kaur vd., 2012).

Macas ve Lhotska ise sosyal etkilerin hesaplanmasında sürecin basitleştirilmesi için sadece bir parametre kullanılmasını önermişlerdir. Bu iki özellik, birçok bileşenin karmaşık işlevleridir ve farklıdır ama bağımsız değildir. Bu yüzden hem destekleyici hem de ikna edici etkinin hesaplanması için yalnızca bir parametre kullanılmaktadır. Her bir i bireyi q_i güç özelliğine sahiptir ve $q_i = s_i = p_i$ (Macas ve Lhotska, 2008).

Uygunluk fonksiyonu, her bir bireyi bir ikili optimizasyon problemi açısından değerlendirilebilmeli ve güç faktörlerinin değişimi güncel uygunluk değerinden ve geçmiş davranışından etkilenmelidir. Güç faktörü, denklem (4.3)'te gösterildiği gibi hesaplanmaktadır ve $[0, 1]$ aralığında değerler almaktadır.

$$q = \frac{f_{max} - f}{f_{max} - f_{min}} \quad (4.3)$$

q mevcut bireyin gücü, f bireyin uygunluk değeri, f_{max} ve f_{min} ise sırasıyla popülasyondaki maksimum ve minimum uygunluk değerleridir (Macas ve Lhotska, 2008).

Algoritmanın ilk başında, d boyuttan oluşan N adet birey gelişigüzel bir biçimde 0 veya 1 değerini alarak popülasyon oluşturulmaktadır. Bu aynı zamanda 0 ve 1'lerden oluşan bir matrise denk gelmektedir. Her iterasyonda, ikna edici etki destekleyici etkiden daha büyükse, bir birey tutumunu değiştirmektedir. Tutum değişikliği durumunda her bireyin gücü de yeniden hesaplanacaktır (Macas ve Lhotska, 2008). Algoritmanın sözde kodu Şekil 4.2'de gösterilmektedir.

SETO'nun Sözde Kodu

```

Tutumlara 0 veya 1 ikili değerler atayarak  $s_i(0)$ 'ı oluştur
while sonlandırma koşulları sağlanıncaya kadar do
  for bütün  $i$ 'ler için do
    Uygunluğu  $f_i(t)$  değerlendir
     $f_{max} = \max(f_i(t))$  ve  $f_{min} = \min(f_i(t))$ 
    Gücü  $q_i(t)$  hesapla
  end for
  for bütün  $i, d$ 'ler için do
    İkna edicilerin sayısını  $N_o$  ve destekleyicilerin sayısını  $N_s$  hesapla
    Toplam ikna edici etki  $\hat{I}_p$ 'yi hesapla
    Toplam destekleyici etki  $\hat{I}_s$ 'yi hesapla
    if  $\hat{I}_p > \hat{I}_s$  do
      Bireyin  $d$  boyutundaki tutumunu  $\hat{s}_i^d(t+1)$  tersine çevir
    end if
    if  $\hat{I}_p > \hat{I}_s$  do
      Bireyin  $d$  boyutundaki tutumunu  $s_i^d(t+1)$ ,  $1-\kappa$  olasılıkla
      tersine çevir
    else
      Bireyin  $d$  boyutundaki tutumunu  $s_i^d(t+1)$ ,  $\kappa$  olasılıkla
      tersine çevir
    end if
     $\hat{s}_i^d(t+1)$ 'i hesapla
  end for
   $t \leftarrow t + 1$ 
end while

```

Şekil 4.2. SETO'nun sözde kodu

SETO Algoritmasının, tutumların başlatılması dışında tamamen deterministik olduğu gözlemlenebilir. Ayrıca, optimizasyon açısından, stokastik öge keşif kabiliyetini arttırabilmektedir ve çeşitliliğin kaybolmasını önleyebilmektedir. Bu nedenle, ikna edici etki, destekleyici etkiden daha düşük olsa bile bir tutum değişikliğinin oluşabilmesi sağlanmıştır. Benzer bir şekilde, ikna edici etki egemen ise, değişiklik olmayabilir. Kendiliğinden olan tutum değişikliğinin olasılığını temsil eden ek olasılık parametresi κ ile etkinleştirilmektedir. Böylece, ikna edici etki daha büyük ise, değişim olasılığı $1-\kappa$ ile gerçekleşmektedir, aksi takdirde değişim olasılık ile gerçekleşmektedir (Macas ve Lhotska, 2008).

4.2. Balina Optimizasyonu Algoritması

Dünyanın en büyük memelisi olarak kabul edilen balinalar son derece zeki hayvanlardır. Balinaların yedi türü bulunmaktadır: katil, minke, kuzey, kambur, gerçek, oluklu ve mavi. BOA'da sadece kambur balinalarda gözlemlenebilen spiral hava kabarcığı adı verilen beslenme hareketi, optimizasyon amacıyla modellenmiştir (Mirjalili ve Lewis, 2016).

Literatürde BOA kullanılarak yapılmış çalışmalar bulunmaktadır. Örneğin; Aljarah ve arkadaşları (2018), BOA'yı kullanarak sinir ağlarındaki bağlantı ağırlığını optimize etmişlerdir. BOA'yı test etmek için zorluk dereceleri farklı olan, UCI veri ambarından alınan Diabetes, Blood, Breast Cancer, Hepatitis gibi 20 veri kümesi seçilmiştir ve sonuçlar GA, PSO, ACO, DE, ES (Evrin Stratejisi) ve BP (Geri Yayılım Algoritması) ile karşılaştırılmıştır. Sonuçlara göre hem yerel optimumdan kaçınmada hem de yakınsama hızı açısından veri setlerin çoğunda BOA'nın daha iyi bir performans sergilediği görülmüştür (Aljarah vd., 2018). Oliva ve arkadaşları (2017), BOA'nın kaotik versiyonunu güneş pilleri parametrelerinin tahmininde kullanmışlardır. Elde edilen sonuçlar; Yapay Arı Koloni Algoritması, CSO, Tavlama Benzetimi, Kaotik PSO, vb., algoritmalarından elde edilen sonuçlar ile karşılaştırılmıştır. Önerdikleri yöntem doğruluk ve sağlamlık konusunda daha yüksek bir performans göstermiştir (Oliva vd., 2017). Veri madenciliği ön işlemlerinden biri olan özellik seçimi; bir veri kümesi içindeki gereksiz, alakasız değişkenleri ortadan kaldırmayı amaçlamaktadır. Mafarja ve Mirjalili (2017), özellik seçimi için BOA ve Tavlama Benzetiminin hibridi olan iki farklı model önermişlerdir. Birinci modelde yerel optimumu bulmak için Tavlama Benzetimi BOA'nın içine

gömülüken, ikinci modelde BOA'daki her bir yinelemeden sonra bulunan en iyi çözümü geliştirmek için Tavlama Benzetimi kullanmışlardır. Önerilen modelleri test etmek için UCI veri ambarından alınan 18 standart referans veri setini kullanmışlardır. Elde edilen bu iki hibrit model BOA, GA, PSO ve ALO (Karıncı Aslanı Optimizasyonu) algoritmaları ile karşılaştırılmıştır. En iyi sonucu ikinci modelin, en iyi ikinci sonucu ise birinci modelin verdiği görülmüştür (Mafarja ve Mirjalili, 2017).

Kaveh ve Ghazaan (2017), iskelet yapılarının optimizasyonunu ayarlamak için keşif aşamasını geliştirerek BOA'yı kullanmışlardır. Geliştirilen BOA iki kafes optimizasyon problemi ve iki çerçeve optimizasyon problemi olmak üzere dört tane problemde test edilmiştir. Diğer optimizasyon algoritmalarıyla karşılaştırıldığında Geliştirilen BOA'nın daha iyi sonuçlar verdiği gözlemlenmiştir (Kaveh ve Ghazaan, 2017). Prakash ve Lakshminarayana (2017), elektrik dağıtım sistemlerindeki hat kayıplarını azaltmak amacıyla tipik bir radyal sisteminde optimum boyutlandırma ve kondansatörlerin yerleştirilmesi için BOA'yı önermişlerdir. Önerilen yöntem, maliyet azaltma, kayıp minimizasyonu ve voltaj profili iyileştirilmesi amaçları göz önüne alınarak IEEE'nin standart 34-bus ve 5-bus test sistemlerinde test edilmiştir. Mevcut diğer algoritmalarla kıyaslandığında, kondansatörlerin, BOA'dan elde edilen orijinal boyut ve yerlerinin maliyeti düşürmesi ve gerilim profilini koruması göz önüne alındığında BOA'nın daha iyi bir performans sergilediğini belirtmişlerdir (Prakash ve Lakshminarayana, 2017).

El Aziz ve arkadaşları (2017), görüntü segmentasyonunda çok seviyeli optimal eşğin belirlenmesi için BOA'yı ve Güve-Alev Optimizasyonunu kullanmışlardır. Sonuçları birkaç kıyaslama görüntüsü kullanarak değerlendirmişlerdir. Karşılaştırılan diğer yedi adet sürü optimizasyonuna göre en iyi sonucu Güve-Alev Optimizasyonu, sonrasında ise BOA vermiştir (El Aziz vd., 2017). Touma (2016), BOA'yı ekonomik sevk problemini çözmek için kullanmıştır. IEEE 30-bus standart test sistemi kullanarak algoritmanın etkinliğini doğrulamıştır. Elde edilen sonuçları PSO, KKA ve GA gibi akıllı yöntemlerle karşılaştırmış ve yakın sonuçlar elde edildiğini ifade etmiştir. (Touma, 2016). Trivedi ve arkadaşları (2016), BOA'nın hem keşif hem de sömürü aşamasında rol oynayan gelişigüzellik için yeni bir teknik önermişlerdir. Bu teknikle optimum çözüme, yerel minimumdan kaçınmaya ve daha hızlı yakınsamaya ulaşmak için daha az hesaplama zamanı harcamıştır (Trivedi, 2016).

Wang ve arkadaşları (2017), rüzgar hızı tahmini için çok amaçlı BOA'yı kullanarak yeni bir hibrit sistem önermişlerdir. Bu hibrit sistem veri ön işleme modülü, optimizasyon

modülü, tahmin modülü ve değerlendirme modülü olmak üzere dört modülden oluşmuştur; BOA optimizasyon modülünde kullanılmıştır. Sonuçlar iki yeni metasezgisel algoritmayla karşılaştırılmıştır ve BOA'nın daha başarılı performans sergilediğini belirtmişlerdir (Wang vd., 2017). Ling ve arkadaşları (2017), Levy uçuş yörüngesine dayanan geliştirilmiş BOA'yı önermişlerdir. Bu yöntemle erken yakınsamaya karşı popülasyonun çeşitliliğini ve yerel optimumdan çıkma yeteneğini artırarak BOA'nın keşif ve sömürü arasında daha iyi bir denge kurmuşlardır. 23 tane test fonksiyonu kullanılarak elde edilen sonuçlar, BOA'nın diğer algoritmalara göre daha iyi performans sergilediğini göstermiştir (Ling vd., 2017). Ben Oualid Medani ve arkadaşları (2018), elektrik şebekelerindeki toplam aktif güç kayıplarının en aza indirgenmesinde BOA'yı kullanmışlardır. Elde edilen sonuçları PSO ve kendi geliştirdikleri değişken kat sayılı PSO ile karşılaştırmışlardır. Elde edilen sonuçlara göre BOA'nın bu problemi çözmede sağlamlığını ve etkinliğini kanıtladığını ifade etmişlerdir (Ben Oualid Medani, 2018).

Mostafa ve arkadaşları (2017), MR görüntülerinde karaciğer segmentasyonu için BOA'ya dayalı bir yöntem önermişlerdir. Abdominal görüntüdeki potansiyel karaciğer pozisyonunu tanımlamak için istatistiksel bir görüntü hazırlanmıştır. Daha sonra, BOA, görüntüyü önceden tanımlanmış sayıda kümeye ayırmaktadır. Önerilen yöntemden elde edilen yüksek başarılı sonuçlar radyoloji uzmanları tarafından açıklanmıştır (Mostafa vd., 2017). Ahmed ve arkadaşları (2017), kablosuz sensör ağlarının ömrünü uzatmada BOA temelli bir algoritma önermişlerdir. Bu algoritmada her bir balinanın konumu ikili kod olarak ifade edilmiştir. Uygunluk fonksiyonu hesaplanırken aktif düğümlerin sayısının en aza indirgenmesi ve bu düğümler içindeki düşük enerji tüketimi özellikleri dikkate alınmıştır (Ahmed vd., 2017).

BOA, kambur balinaların üç temel hareketi baz alınarak geliştirilmiştir. Bu hareketler avın çevrelenmesi, hava kabarcığı saldırısı ve av için arama olmaktadır.

4.2.1. Avın Çevrelenmesi

Kambur balinalar avın yerini görerek çevreleyebilmektedir. En iyi sonuç bilinmediğinden dolayı, BOA mevcut en iyi aday çözümü hedef av olarak ya da optimum değere yakın olarak varsaymaktadır. En iyi aday çözüm tanımlandıktan sonra diğer aday çözümler en iyi aday çözüme doğru pozisyonlarını güncellemeye çalışacaktır. Bu hareket, denklem (4.4) ve (4.5)'teki denklemlerle temsil edilmektedir.

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (4.4)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (4.5)$$

t mevcut iterasyonu, $\vec{A}$ ve $\vec{C}$ katsayı vektörünü belirtmektedir, $\vec{X}^*$ şimdiye kadar elde edilen en iyi çözümün pozisyon vektörü, $\vec{X}$ pozisyon vektörü, $||$ mutlak değeri ve $\cdot$ bir eleman-eleman çarpımını ifade etmektedir. $\vec{A}$ ve $\vec{C}$ vektörleri denklem (4.6) ve (4.7)'deki gibi hesaplanmaktadır (Mirjalili ve Lewis, 2016).

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (4.6)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (4.7)$$

$\vec{a}$ iterasyonların seyri boyunca 2'den 0'a doğru doğrusal olarak azaltılmaktadır. $\vec{r} [0, 1]$ aralığında gelişigüzel bir vektördür.

4.2.2. Hava Kabarcığı Saldırı Yöntemi

Bu yöntemde suyun altında soluk vererek avlarının etrafında kabarcıklar oluşturmaktadırlar. Daha sonra çemberin altından su yüzeyine doğru ağzı açık bir şekilde dalarak avlarını ele geçirmektedirler. Saldırıya benzer bir davranış sergiledikleri bu yöneme sömürü aşaması da denilmektedir. Kambur balinaların hava kabarcığı davranışını matematiksel olarak modellemek için iki yaklaşım tasarlanmıştır:

1. Küçülen Sarmal mekanizması: Bir grup balina avlarının altında daire şeklinde yüzmektedirler ve çemberin boyutunu küçülterek avlarını bir baloncuk halkasına sıkıştırmaktadırlar. Böylece tek seferde birçok av avlanabilmektedir. Bu davranış, denklem (4.6)'daki $\vec{a}$ değerinin azaltılmasıyla elde edilmiştir. $\vec{A}$ 'nın oynama oranı da $\vec{a}$ tarafından azaltılmış olmaktadır. Başka bir deyişle $\vec{A}$, iterasyonlar süresince 2'den 0'a düşürülen $[-a, a]$ aralığında gelişigüzel bir değerdir.

2. Spiral Güncelleme Pozisyonu: Bu yaklaşım ilk olarak (X, Y) konumunda bulunan balinayla, (X^*, Y^*) konumunda bulunan av arasındaki mesafeyi hesaplamaktadır. Daha sonra, bir spiral denklem (4.8), kambur balinanın helezon şeklindeki hareketini taklit etmek için balina ile avın konumu arasında oluşturulmuştur (Mirjalili ve Lewis, 2016).

$$\vec{X}(t+1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (4.8)$$

$\vec{D}' = |\vec{X}^*(t) - \vec{X}(t)|$ olmaktadır ve i 'ninci balina ile av (mevcut en iyi çözüm) arasındaki uzaklığı ifade etmektedir. b logaritmik spiralin şeklini tanımlamak için bir sabittir, l $[-1, 1]$ arasında gelişigüzel bir sayıdır ve $\cdot$ ise eleman-eleman çarpımıdır.

Kambur balinalar daralan bir daire içinde ve avın etrafında spiral bir şekilde yüzmektedir. Bu eşzamanlı hareketi modellemek için optimizasyon boyunca balinaların pozisyonunu güncellenirken küçülen çevreleme mekanizması veya spiral modelin hangisinin seçileceği %50 olasılık olarak düşünülmüştür. Matematiksel modeli denklem (4.9)'da verilmiştir (Mirjalili ve Lewis, 2016).

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (4.9)$$

p , $[0, 1]$ arasında gelişigüzel bir sayıdır.

4.2.3. Av Arama

Av arama, kambur balinaların diğer yiyecek kaynaklarını keşfetmeye yönlendirdiğinden keşif aşaması olarak da adlandırılmaktadır. $\vec{A}$ vektörü için sömürü aşamasındaki benzer yaklaşım kullanılmaktadır. Kambur balinalar birbirlerinin pozisyonlarına göre avlarını gelişigüzel aradıkları için bir referans balinadan uzaklaşmaya zorlamak için, $\vec{A}$ 1'den büyük veya -1'den küçük gelişigüzel değerlerle kullanılmaktadır. Sömürü aşamasından farklı olarak, mevcut en iyi uygunluklu balina yerine gelişigüzel seçilen bir balinanın pozisyonuna göre güncellenmektedir. Bu teknik ve $|\vec{A}| > 1$ ile keşiflere önem verilmektedir ve BOA'nın global arama yapmasına imkan vermektedir. Keşif aşamasının

matematiksel modeli denklem (4.10) ve (4.11)'de verilmektedir (Mirjalili ve Lewis, 2016).

$$\vec{D} = |\vec{C} \cdot \overrightarrow{X_{rand}} - \vec{X}| \quad (4.10)$$

$$\vec{X}(t + 1) = \overrightarrow{X_{rand}} - \vec{A} \cdot \vec{D} \quad (4.11)$$

$\overrightarrow{X_{rand}}$ mevcut popülasyondan gelişigüzel seçilen bir balinanın pozisyon vektörüdür.

BOA gelişigüzel çözümler kümesi ile başlamaktadır. Her bir iterasyonda, balinalar pozisyonlarını mevcut en iyi çözüme göre veya gelişigüzel seçilen bir balinanın pozisyon vektörüne göre güncellemektedir. a parametresi sırasıyla keşif ve sömürüyü sağlamak için 2'den 0'a doğru azaltılmaktadır. Arama bireyinin pozisyonu güncellenirken eğer $|\vec{A}| > 1$ ise gelişigüzel A balinası, eğer $|\vec{A}| < 1$ ise en iyi çözüm seçilmektedir. p değerine bağlı olarak, BOA spiral veya dairesel hareketten birini seçmektedir. Son olarak, sonlandırma koşulları sağlandığında BOA sonlanmaktadır (Mirjalili ve Lewis, 2016). BOA'nın sözde kodu Şekil 4.3'teki gibidir.

BOA'nın Sözde Kodu

```
Balina popülasyonunu oluştur  $X_i (i = 1, 2, \dots, n)$   
Her bir arama bireyinin uygunluğunu hesapla  
 $X^* =$  en iyi arama bireyi  
while ( $t < \text{maksimum iterasyon sayısı}$ )  
  for her bir arama bireyi  
     $a, A, C, l$  ve  $p$ 'yi güncelle  
    if ( $p < 0.5$ )  
      if ( $|A| < 1$ )  
        Mevcut arama bireyinin pozisyonunu (4.4)'e göre güncelle  
      else if ( $|A| \geq 1$ )  
        Rasgele bir arama bireyi ( $X_{rand}$ ) seç  
        Mevcut arama bireyinin pozisyonunu (4.11)'e göre güncelle  
      end  
    else if ( $p \geq 0.5$ )  
      Mevcut arama bireyinin pozisyonunu (4.8)'e göre güncelle  
    end  
  end  
  Herhangi bir arama bireyinin arama uzayının ötesine geçip geçmediğini  
  kontrol et ve düzelt  
  Her bir arama bireyinin uygunluğunu hesapla  
  Daha iyi bir çözüm bulunduysa  $X^*$ 'i güncelle  
     $t = t + 1$   
end  
 $X^*$ 'i döndür
```

Şekil 4.3. BOA'nın sözde kodu

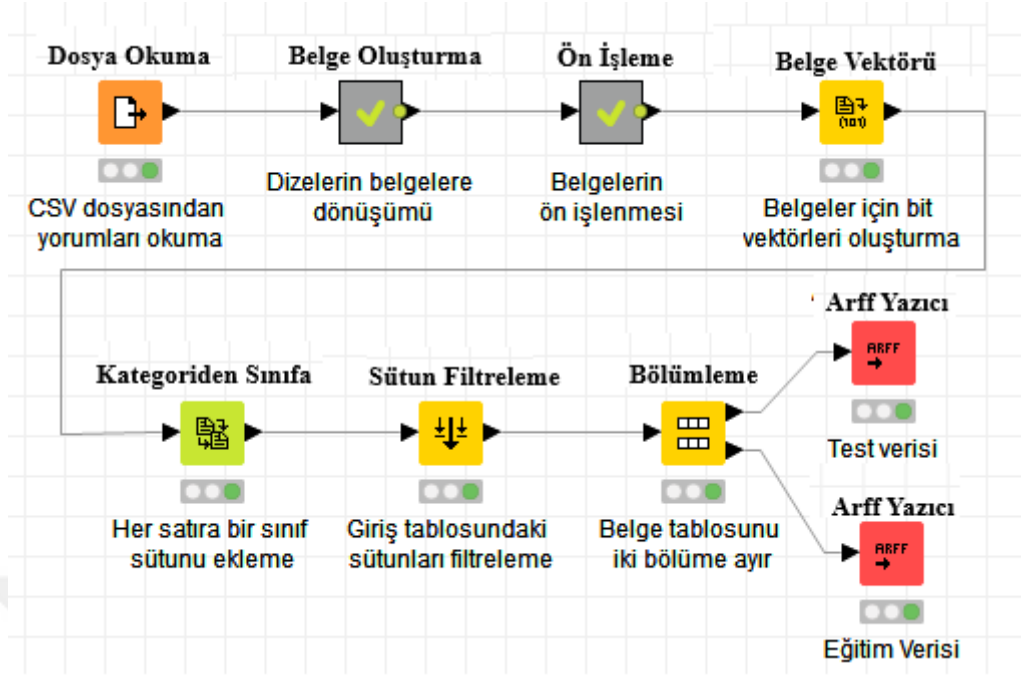
5. DENEYSEL ÇALIŞMADA KULLANILAN VERİ SETLERİ VE ÖN İŞLEMLER

Bu tez çalışmasında SETO ve BOA ile duygu analizinin yapılması MATLAB platformunda yazılan program ile gerçekleştirilmiştir. Bu bölümde öncelikle KNIME (URL-1) programı kullanılarak yapılan ön işlemler anlatılacaktır. Daha sonra ise uygulamada kullanılan veri setleri ve uygulamanın adım adım nasıl oluşturulduğundan bahsedilecektir.

5.1. KNIME’da Yapılan Ön İşlemler

Açılımı “Konstanz Information Miner” olan KNIME, Almanya’daki Konstanz Üniversitesi’nde geliştirilmiş, grafik arayüzlü ve açık kaynak kodlu bir veri madenciliği uygulama yazılımıdır (URL-1). KNIME Analytics Platformu, veriye dayalı yenilikler için, veride saklı olan potansiyelin keşfedilmesi, yeni iç görüler elde etmek veya yeni gelecekleri tahmin etmek için tasarlanmış lider bir açık çözümdür.

KNIME, kullanıcıların görsel olarak veri akışları oluşturmaya, seçici olarak bazı analiz adımlarını yürütmesine ve daha sonra sonuçları, modelleri ve etkileşimli görünümeleri incelemesine olanak tanımaktadır. Bu tez çalışmasında KNIME yazılımı kullanılarak duygu analizinde kullanılacak veriler gerekli ön işlemlerden geçirilerek doküman matrisi elde edilmiştir. Bu doküman matrisinin satırları veri setindeki her bir yoruma, sütunları ise yorumlardan çıkarılan kelimelere karşılık gelmektedir. İlgili sütundaki kelime o satırda varsa değer 1 yok ise değer 0 olmaktadır. KNIME’da yapılan işlemlerin iş akış şeması Şekil 5.1’deki gibidir.



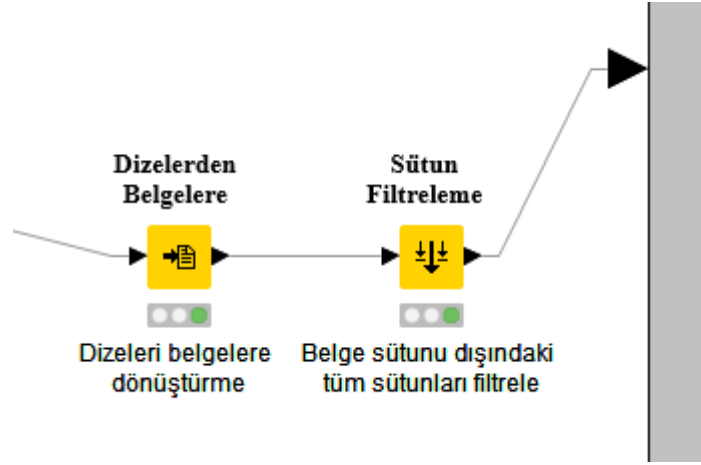
Şekil 5.1. KNIME’da yapılan işlemlerin akış şeması

5.1.1. Dosya Okuma (File Reader)

İş akışı bir File Reader düğümü ile başlamaktadır. Bu düğüm bir dosyadan veri okumak için kullanılmaktadır; ve çeşitli formatları okumak için yapılandırılabilir. Bu tez çalışmasında, .csv uzantılı dosyalar okutulmuştur. Düğümün yapılandırma iletişim kutusu açıldığında ve bir dosya adı sağlandığında dosyanın içeriği analiz edilerek okuyucunun ayarları tahmin edilmeye çalışılmaktadır. Bu ayarların sonucu ön izleme tablosundan da kontrol edilebilmektedir.

5.1.2. Belge Oluşturma (Document Creation)

Bu düğüm alt düğümlerden oluşmaktadır. Bu düğümde genel olarak dizeler belgelere dönüştürülmektedir. Bu düğümün içerdiği alt düğümler Şekil 5.2’de gösterilmektedir.



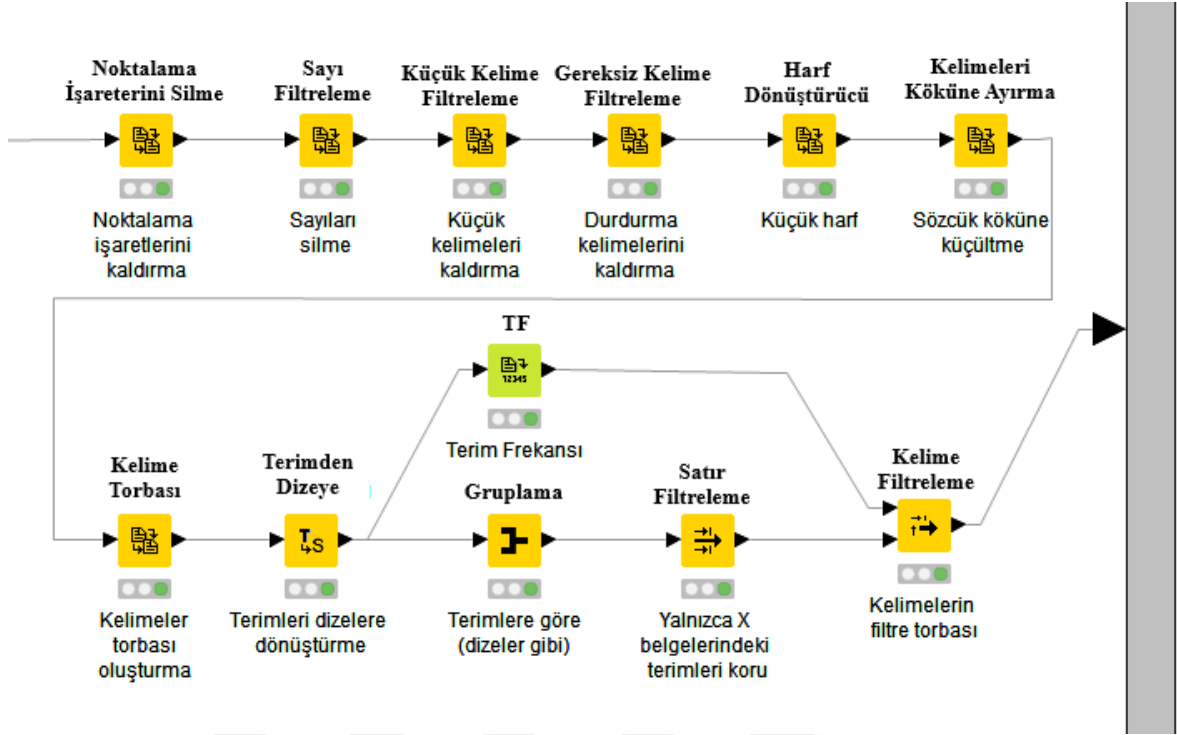
Şekil 5.2. Belge oluşturma düğümleri

Dizeleri Belgelere Dönüştürme (Strings to Document): Belirtilen dizeleri belgelere dönüştürmektedir. Her satır için bir belge oluşturulacak ve bu satıra eklenmektedir. Belirtilen sütunların dizeleri başlık, yazar ve tam metin olarak kullanılacaktır.

Sütun Filtreleme (Column Filter): Bu düğüm, sütunların giriş tablosundan filtrelenemesine izin vermektedir ve yalnızca kalan sütunlar çıkış tablosuna iletilmektedir. Burada da belge hücrelerini içeren sütun dışındaki tüm sütunlar daha sonra filtrelenmektedir.

5.1.3. Ön İşleme (Preprocessing)

Metin verileri KNIME metin işleme uzantısı tarafından sağlanan çeşitli düğümler tarafından işlenmektedir. Bu düğüm içinde yer alan tüm ön işleme adımları Şekil 5.3'te gösterilmektedir



Şekil 5.3. KNIME’da ön işleme adımları

Noktalama İşaretlerini Silme (Punctuation Erasure): Giriş belgelerinde bulunan tüm noktalama işaretlerini kaldırmaktadır.

Sayı Filtreleme (Number Filter): “.” veya “,” ondalık ayırıcıları ve “+” veya “-” işaretlerini de dahil olmak üzere rakamlardan oluşan giriş belgelerinde bulunan tüm terimleri filtrelemektedir. Ayrıca bir veya daha fazla rakam içeren tüm terimleri filtreleme seçeneği de bulunmaktadır.

Küçük Kelimeleri Kaldırma (N Chars Filter): Girdi belgelerinde yer alan tüm terimleri, belirtilen N sayıda karakterden az olanları filtrelemektedir. Bu tez çalışmasında $N=3$ olarak alınmıştır.

Gereksiz Kelimeleri Filtreleme (Stop Word Filter): Giriş tablosundaki gereksiz kelime listesinde yer alan terimleri filtrelemektedir. Bu düğüm çeşitli diller için gereksiz kelime listeleri sağlamaktadır.

Harf Dönüştürücü (Case Converter): Giriş belgelerinde yer alan tüm terimleri küçük veya büyük harfe dönüştürmektedir.

Kelimeleri Köküne Ayırma (Snowball Stemmer): Bu düğümde, terimlerin kökleri Snowball terim kütüphanesine göre belirlenmektedir.

Kelime Torbası Oluşturma (Bag of Words Creator): Bu düğüm bir grup dokümanın kelime torbasını oluşturmaktadır. İlgili belgede ortaya çıkan terimleri içeren en az bir sütundan oluşmaktadır.

Terimden Dizeye Dönüştürme (Term to String): Bu düğüm, tanımlı bir sütunun terimlerini dizelere dönüştürülmesine ve bu dizeleri içeren yeni bir sütun eklenmesine olanak tanımaktadır.

Terim Frekansı (Term Frequency-TF): Her bir terimin göreceli terim sıklığını her bir belgeye göre hesaplamaktadır ve tf değerini içeren bir sütun eklemektedir. Değer, bir belgeye göre bir terimin mutlak sıklığının, o belgenin tüm koşullarının sayısına bölünmesiyle hesaplanmaktadır.

Gruplama (GroupBy): Bir tablonun satırlarını, seçilen grup sütunlarındaki benzersiz değerlere göre gruplandırmaktadır. Seçilen grup sütunun her benzersiz değeri için bir satır oluşturmaktadır. Kalan sütunlar, belirtilen toplama ayarlarına göre toplanmaktadır. Çıkış tablosu, seçilen grup sütunlarının her bir benzersiz değer kombinasyonu için bir satır içermektedir.

Satır Filtreleme (Row Filter): Bu düğüm, belirli kriterlere göre satır filtrelemeye izin vermektedir. Bu çalışmada, en az 20 belgede geçen kelimeler alınmıştır. 20'den az belgede geçen kelimeler filtrelenmiştir.

Kelimelerin Filtre Torbası (Reference Row Filter): Bu düğüm, referans olarak ikinci tabloyu kullanarak satırların ilk tablodan filtrelenmesini sağlamaktadır. Diyalog ayarına bağlı olarak, referans tablosundaki satırlar çıktı tablosuna dahil edilmektedir veya dışlanmaktadır.

5.1.4. Belge Vektörü (Document Vector)

Bu düğüm, her bir belge için boşluk terimlerinde temsil eden bir belge vektörü oluşturmaktadır. Özellik vektörlerinin değerleri, boole değerleri olarak veya belirtilen bir sütunun değerleri olarak belirtilebilmektedir. Vektörlerin boyutu, Bag of Words Creator'daki farklı terimlerin sayısı olacaktır.

5.1.5. Kategoriden Sınıfa (Category to Class)

Bu düğüm, bir belge hücresi içeren her satıra bir sınıf sütunu eklenmesini sağlamaktadır. Sınıfın değeri belgenin dizge olarak kategorisidir.

5.1.6. Sütun Filtreleme (Column Filter)

Bu düğüm, sütunların giriş tablosundan filtrelenmesini sağlamaktadır. Yalnızca kalan sütunlar çıkış tablosundan geçirilmektedir. Bu çalışmada tablonun daha sade olması açısından giriş belgesinin olduğu sütun çıkarılmıştır.

5.1.7. Bölümleme (Partitioning)

Ayrıca bir test verisi olmayan veri setlerinde bu düğüm kullanılmıştır. Giriş tablosunu iki bölüme ayırmayı sağlamaktadır. Tablodaki veriler %70'i eğitim verisi ve %30'u test verisi olacak şekilde ikiye ayrılmıştır.

5.1.8. ARFF Yazıcı (ARFF Writer)

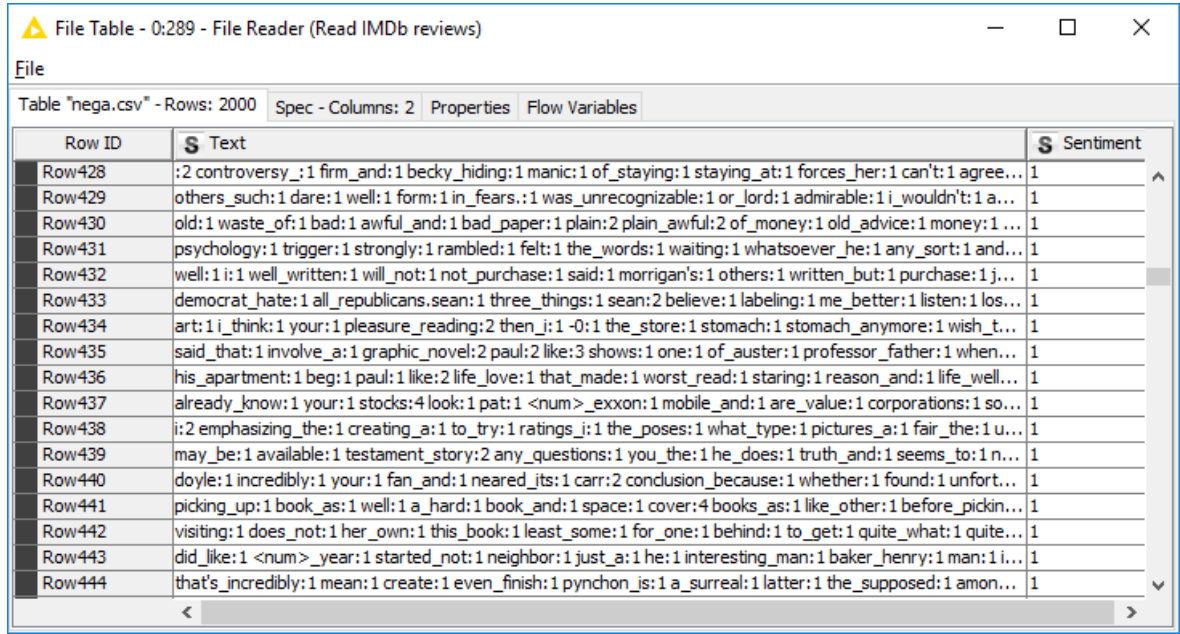
Bir dosyaya veya ARFF formatında bir URL ile belirtilen uzak bir yere veri kaydetmektedir. Bu düğümü kullanma amacı, elde edilen sonuçları WEKA programındaki gözetimli makine öğrenmesi algoritmalarından elde edilen sonuçlar ile karşılaştırabilmek için gerekli .arff dosyasının oluşturulmasıdır.

5.2. Duygu Analizi için Kullanılan Veri Setleri

Bu tez çalışmasında, IMDB (URL-2), Polarity (URL-2) ve Amazon (URL-3) yorumlarını içeren 3 adet veri seti kullanılmıştır. Bu 3 veri seti de olumlu ve olumsuz olmak kaydıyla iki duygu sınıfına sahiptir. Kullanılan veri setleri ve özellikleri şu şekildedir:

5.2.1. IMDB Yorumları Veri Seti

Bu veri seti IMDB yorumlarından oluşmaktadır. İlk olarak Bo Pang ve Lillian Lee tarafından kullanılmıştır (Pong ve Lee, 2004). 1000 pozitif ve 1000 negatif olarak toplam 2000 adet yorum bulunmaktadır. Bu veri setinin KNIME'daki ekran görüntüsü Şekil 5.4'teki gibidir.



Row ID	S Text	S Sentiment
Row428	:2 controversy_:1 firm_and:1 becky_hiding:1 manic:1 of_staying:1 staying_at:1 forces_her:1 can't:1 agree...	1
Row429	others_such:1 dare:1 well:1 form:1 in_fears.:1 was_unrecognizable:1 or_lord:1 admirable:1 i_wouldn't:1 a...	1
Row430	old:1 waste_of:1 bad:1 awful_and:1 bad_paper:1 plain:2 plain_awesome:2 of_money:1 old_advice:1 money:1 ...	1
Row431	psychology:1 trigger:1 strongly:1 rambled:1 felt:1 the_words:1 waiting:1 whatsoever_he:1 any_sort:1 and...	1
Row432	well:1 i:1 well_written:1 will_not:1 not_purchase:1 said:1 morrigan's:1 others:1 written_but:1 purchase:1 j...	1
Row433	democrat_hate:1 all_republicans.sean:1 three_things:1 sean:2 believe:1 labeling:1 me_better:1 listen:1 los...	1
Row434	art:1 i_think:1 your:1 pleasure_reading:2 then_j:1 -0:1 the_store:1 stomach:1 stomach_anymore:1 wish_t...	1
Row435	said_that:1 involve_a:1 graphic_novel:2 paul:2 like:3 shows:1 one:1 of_auster:1 professor_father:1 when...	1
Row436	his_apartment:1 beg:1 paul:1 like:2 life_love:1 that_made:1 worst_read:1 staring:1 reason_and:1 life_well...	1
Row437	already_know:1 your:1 stocks:4 look:1 pat:1 <num>_exxon:1 mobile_and:1 are_value:1 corporations:1 so...	1
Row438	i:2 emphasizing_the:1 creating_a:1 to_try:1 ratings_j:1 the_poses:1 what_type:1 pictures_a:1 fair_the:1 u...	1
Row439	may_be:1 available:1 testament_story:2 any_questions:1 you_the:1 he_does:1 truth_and:1 seems_to:1 n...	1
Row440	doyle:1 incredibly:1 your:1 fan_and:1 neared_its:1 carr:2 conclusion_because:1 whether:1 found:1 unfort...	1
Row441	picking_up:1 book_as:1 well:1 a_hard:1 book_and:1 space:1 cover:4 books_as:1 like_other:1 before_pickin...	1
Row442	visiting:1 does_not:1 her_own:1 this_book:1 least_some:1 for_one:1 behind:1 to_get:1 quite_what:1 quite...	1
Row443	did_like:1 <num>_year:1 started_not:1 neighbor:1 just_a:1 he:1 interesting_man:1 baker_henry:1 man:1 i...	1
Row444	that's_incredibly:1 mean:1 create:1 even_finish:1 pynchon_js:1 a_surreal:1 latter:1 the_supposed:1 amon...	1

Şekil 5.4. IMDB yorumları veri setinin KNIME'daki ekran görüntüsü

Bu veri setinin KNIME'da ön işlemden geçirildikten sonra elde edilen ekran görüntüsü Şekil 5.5'teki gibi olmaktadır. Buna göre ön işlemden geçirilerek elde edilen kelime sayısı 1055 olmaktadır. Yani elde edilen doküman matrisi 2000×1055 boyutundadır. Bu doküman matrisi %70'i eğitim (1400×1055 boyutlu) ve geri kalan %30'u da test (600×1055) verisi olmak üzere iki ayrı matrise ayrılmıştır.

Filtered table - 0:302:302 - Column Filter

File

Table "default" - Rows: 2000 Spec - Columns: 1055 Properties Flow Variables

Row ID	D speak	D light	D ye	D toget	D actual	D stop	D num
2001	1	1	1	1	1	1	1
2002	0	0	0	0	0	0	0
2003	0	0	0	0	0	0	1
2004	0	0	0	0	0	0	0
2005	0	0	0	0	0	0	1
2006	0	0	0	0	0	0	1
2007	0	0	0	0	0	0	0
2008	0	0	0	0	0	0	0
2009	0	0	0	0	1	0	1
2010	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	1
2012	0	0	0	0	0	0	0
2013	1	0	0	1	0	0	1
2014	0	0	0	0	0	0	0
2015	0	0	0	0	0	0	1
2016	0	0	0	0	0	0	0
2017	0	0	0	0	0	0	1
2018	0	0	0	0	0	0	1
2019	0	0	0	0	0	0	0
2020	0	0	0	0	0	0	0

Şekil 5.5. IMDB yorumları veri setinin doküman matrisi

5.2.2. Polarity Veri Seti

Bu veri seti bir önceki veri seti gibi IMDB yorumlarından oluşmaktadır. Bu veri seti ilk olarak Pang ve Lee tarafından kullanılmıştır (Pang ve Lee, 2005). Bu veri setinde 5331'i olumlu ve 5331'i olumsuz olmak üzere toplamda 10662 adet yorum bulunmaktadır. Bu veri setinin KNIME'daki ekran görüntüsü Şekil 5.6'daki gibi olmaktadır.

Row ID	Text
Row0	the rock is destined to be the 21st century's new conan and that he's going to make a splash
Row1	the gorgeously elaborate continuation of the lord of the rings trilogy is so huge that a color
Row2	effective but too-tepid biopic
Row3	if you sometimes like to go to the movies to have fun , wasabi is a good place to start .
Row4	emerges as something rare , an issue movie that's so honest and keenly observed that it do
Row5	the film provides some great insight into the neurotic mindset of all comics -- even those who
Row6	offers that rare combination of entertainment and education .
Row7	perhaps no picture ever made has more literally showed that the road to hell is paved with c
Row8	steers turns in a snappy screenplay that curls at the edges ; it's so clever you want to hate
Row9	take care of my cat offers a refreshingly different slice of asian cinema .
Row10	this is a film well worth seeing , talking and singing heads and all .
Row11	what really surprises about wisegirls is its low-key quality and genuine tenderness .
Row12	(wendigo is) why we go to the cinema : to be fed through the eye , the heart , the mind .
Row13	one of the greatest family-oriented , fantasy-adventure movies ever .
Row14	ultimately , it ponders the reasons we need stories so much .
Row15	an utterly compelling 'who wrote it' in which the reputation of the most famous author who e

Şekil 5.6. Polarity veri setinin KNIME'daki ekran görüntüsü

Bu veri setinin KNIME'da ön işlem sonucunda elde edilen doküman matrisi Şekil 5.7'deki gibidir. Elde edilen kelime sayısı 450'dir ve buna bağlı olarak elde edilen doküman matrisinin boyutu da 10662×450 olmaktadır. Bu iki veri setinin de ayrı bir test verisi olmadığı için %70'i eğitim (7076×450) ve %30'u test (3033×450) verisi olacak şekilde iki ayrı doküman matrisi olarak ayrılmıştır.

Second partition (remaining rows) - 3:307 - Partitioning (Split document table) ↔

File Hilite Navigation View

Table "default" - Rows: 7076 Spec - Columns: 450 Properties Flow Variables

Row ID	D shot	D cultur	D head	D ve	D seen	D movi	D origin
Row17498	0	1	1	1	1	1	0
Row17499	0	0	0	0	0	0	1
Row17501	0	0	0	0	0	1	0
Row17502	0	0	0	0	0	0	0
Row17503	0	0	1	0	0	0	0
Row17504	0	0	0	0	0	0	0
Row17507	0	0	0	0	0	0	0
Row17508	0	0	0	0	0	0	0
Row17512	0	0	0	0	0	0	0
Row17513	0	0	0	0	0	0	0
Row17514	0	0	0	0	0	0	0
Row17515	0	0	0	0	0	1	0
Row17516	0	0	0	0	0	0	0
Row17518	0	0	0	0	0	0	0
Row17520	0	0	0	0	0	1	0
Row17521	0	0	0	0	0	0	0
Row17523	0	0	0	0	0	0	0
Row17524	0	0	0	0	0	0	0
Row17525	0	0	0	0	0	0	0
Row17526	0	0	0	0	0	1	0
Row17527	0	0	0	0	0	0	0
Row17528	0	0	0	0	0	0	0
Row17530	0	0	0	0	0	0	0

Şekil 5.7. Polarity veri setinden elde edilen doküman matrisi

5.2.3. Amazon Veri Seti

Kullanılan son veri seti de McAuley ve arkadaşlarının tanıtmış olduğu Amazon ürün yorumları veri seti olmaktadır (McAuley vd., 2015). Bu veri seti, eğitim verisi olarak 2912'si olumlu ve 3039'u olumsuz olmak üzere toplamda 5951 tane yorum içermektedir. Test verisi olarak da 783'ü olumlu ve 791 olumsuz yorum içermektedir. KNIME'dan alınan ekran görüntüsü Şekil 5.8'deki gibidir.

File Table - 0:305:289 - File Reader

File

Table "train_binary_sent.csv" - Rows: 6920 Spec - Columns: 2 Properties Flow Variables

Row ID	S Sentiment	S Text
Row0	1	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s
Row1	0	Apparently reassembled from the cutting-room floor of any given daytime soap.
Row2	0	They presume their audience won't sit still for a sociology lesson, however entertainingly
Row3	1	This is a visually stunning rumination on love, memory, history and the war between art a
Row4	1	Jonathan Parker's Bartleby should have been the be-all-end-all of the modern-office anon
Row5	1	Campanella gets the tone just right -- funny in the middle of sad in the middle of hopeful.
Row6	0	A fan film that for the uninitiated plays better on video with the sound turned down.
Row7	1	BÄart and Berling are both superb, while Huppert ... is magnificent.
Row8	0	A little less extreme than in the past, with longer exposition sequences between them, an
Row9	0	The film is strictly routine.
Row10	1	A lyrical metaphor for cultural and personal self-discovery and a picaresque view of a little
Row11	0	The most repugnant adaptation of a classic text since Roland JoffÄ and Demi Moore's T
Row12	0	For something as splendid-looking as this particular film, the viewer expects something sp
Row13	1	This is a stunning film, a one-of-a-kind tour de force.
Row14	1	May be more genial than ingenious, but it gets the job done.
Row15	1	There is a freedom to watching stunts that are this crude, this fast-paced and this insane
Row16	0	If The Tuxedo actually were a suit, it would fit Chan like a \$99 bargain-basement special.
Row17	1	As quiet, patient and tenacious as Mr. Lopez himself, who approaches his difficult, endless
Row18	n	Final verdict: You've seen it all before

Şekil 5.8. Amazon veri setinin KNIME'daki ekran görüntüsü

Amazon veri setinin KNIME'daki ön işlemler sonucu elde edilen doküman matrisi Şekil 5.9'daki gibidir. Elde edilen doküman matrisinin boyutu eğitim verisi için 5951×164 , test verisi için de 1574×164 olmaktadır.

Filtered table - 0:305:302 - Column Filter

File

Table "default" - Rows: 5951 Spec - Columns: 164 Properties Flow Variables

Row ID	D act	D action	D actor	D actual	D american	D anim	D attempt	D au
6141	0	0	0	0	0	0	0	0
6142	0	0	0	0	0	0	0	0
6143	0	0	0	0	0	0	1	0
6144	0	0	0	0	0	0	0	0
6145	0	0	0	0	0	0	0	0
6146	0	0	0	0	0	0	0	0
6147	0	0	0	0	0	0	0	0
6148	0	0	0	0	0	0	0	0
6149	0	0	0	0	0	0	0	0
6150	0	0	0	0	0	0	0	1
6151	0	0	0	0	0	1	0	0
6152	0	0	0	0	0	0	0	0
6153	0	0	0	0	0	0	0	0
6154	0	0	0	0	0	0	0	0
6155	0	0	0	0	0	0	0	0
6156	0	0	0	0	0	0	0	0
6157	0	0	0	0	0	0	0	0
6158	0	0	0	0	0	0	0	0
6159	0	0	0	0	0	0	0	0
6160	0	0	0	0	0	0	0	0
6161	0	0	0	0	0	0	0	0
6162	0	0	0	0	0	0	0	0

Şekil 5.9. Amazon veri setinden elde edilen doküman matrisi

6. PROBLEMİN OPTİMİZASYON ALGORİTMALARINA UYARLANMASI

Bu çalışmada duygu analizi - fikir madenciliği arama problemi olarak ele alınmıştır. Bu problemin çözüm yöntemi olarak iki güncel akıllı optimizasyon algoritması olan SETO ve BOA uyarlanmıştır. Arama uzayı ise IMDB, Polarity ve Amazon veri setleri olmaktadır. Problem optimizasyon algoritmalarına uyarlanırken ilk önce her bir veri setindeki kelimeler geçme sıklığına göre ağırlıklandırılmıştır. Ön işlemlerden sonra oluşturulan doküman matrisindeki kelime sayılarına göre başlangıç popülasyonu oluşturulmuş. Uygunluk değerini hesaplamak için Jaccard benzerliği kullanılmıştır.

6.1. Kelimelerin Ağırlıklandırılması

Veri setlerinden elde edilen doküman matrisinde her bir kelime $[0, 1]$ aralığında olacak şekilde ağırlıklandırılarak, veri setindeki kelime sayısı uzunluğunda bir ağırlık vektörü oluşturulmuştur. İlk olarak her bir kelimenin kaç belgede geçtiği hesaplanmıştır, buradan bir kelimenin maksimum kaç belgede geçtiği $max(kelimesayısı)$ bulunmuştur. Daha sonra bir kelimenin kaç belgede geçtiği sayıyı $max(kelimesayısı)$ 'na bölünerek her bir kelimenin ağırlığı elde edilmiş olur. En yüksek ağırlık ise maksimum belgede geçen kelimenin ağırlığı olan 1 olacaktır. Ağırlıklandırma Şekil 6.1'deki gibi ifade edilebilmektedir. Bu şekilde polarity veri setinden kelimelerin ağırlıklandırılmasına dair bir kesit gösterilmektedir. Polarity veri setinde 1223 geçme sayısı ile belgelerde en çok geçen kelime “film” olmaktadır ve ağırlığı 1'dir. Diğer kelimelerin ağırlıkları da belgelerde en çok geçen kelimenin ağırlığına göre hesaplanmaktadır.

POLARITY VERİ SETİNDEN BİR KESİT									
	feel	lack	script	humor	film	imagin	origin	...	consider
Belge ₁	0	0	0	0	0	0	0	...	0
Belge ₂	0	0	0	0	0	0	1	...	0
Belge ₃	1	0	0	0	0	0	0	...	0
Belge ₄	0	1	1	1	0	0	0	...	0
Belge ₅	0	0	0	0	1	0	0	...	0
Belge ₆	0	0	0	0	0	1	0	...	0
Belge ₇	0	0	0	0	0	0	0	...	0
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Geçme Sayısı	237	106	119	103	1223	80	93	...	22

$$Ağırlık_i = \frac{GeçmeSayısı_i}{MaksimumGeçmeSayısı}$$

AĞIRLIK VEKTÖRÜ									
	feel	lack	script	humor	film	imagin	origin	...	consider
Ağırlık	0.194	0.0867	0.0973	0.0842	1	0.0654	0.076	...	0.0180

Şekil 6.1. Kelimelerin ağırlıklandırılması

6.2. Başlangıç Popülasyonu ve Kodlama

Kullanılan doküman matrisinin boyutuna göre popülasyonun bireyleri oluşturulmaktadır. Her bir birey bir vektörden oluşmaktadır ve bu vektörlere SETO için {0, 1} ve BOA için ise [0, 1] aralığında gelişigüzel değerler verilmektedir. Bir popülasyon şu şekilde ifade edilmektedir:

$$X_{N \times M} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1M} \\ X_{21} & X_{22} & \dots & X_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \dots & X_{NM} \end{bmatrix}$$

Burada N popülasyon sayısı olmaktadır ve bu çalışmada 100 olarak alınmıştır. M ön işleminden sonra çıkarılan doküman matrisinde olan kelimelerin sayısı olmaktadır. Bitim şartı ise 100 iterasyon olarak belirlenmiştir.

6.3. Jaccard Benzerliği

Bu çalışmada doküman benzerliklerinin hesaplanması için Genişletilmiş Jaccard benzerliği kullanılmıştır. Genişletilmiş Jaccard benzerliği, iki nesnenin paylaşılan parçalarının, nesnelerin tüm parçalarına oranıdır. Başka bir deyişle, denklem (6.1)'de görüldüğü gibi iki vektörün kesişim kümelerinin birleşim kümelerine oranı olarak tanımlanabilmektedir. $X_i = X_{i1}, X_{i2}, \dots, X_{iM}$ vektörü, popülasyondaki i . birey olsun. $D_j = D_{j1}, D_{j2}, \dots, D_{jM}$ vektörü ise doküman matrisindeki j . yorum satırına karşılık gelsin. Buna göre Genişletilmiş Jaccard benzerliğinin denklemi (6.2)'deki gibidir.

$$jaccard_değeri_{ij} = \frac{|X_i \cap D_j|}{|X_i \cup D_j|} \quad (6.1)$$

$$jaccard_değeri(X_i, D_j) = \frac{\sum_{k=1}^M X_{ik} D_{jk}}{\sum_{k=1}^M (X_{ik})^2 + \sum_{k=1}^M (D_{jk})^2 - \sum_{k=1}^M X_{ik} D_{jk}} \quad (6.2)$$

Ayrıca Jaccard benzerliği denklemine, üzerinde işlem yapılan yorumda geçen kelimelerin sayısının, doküman matrisindeki tüm kelime sayısına oranı da eklenmiştir. Bu oran denklem (6.3)'te verilmiştir.

$$oran_j = \frac{j.yorumdaki\ kelime\ sayısı}{dokümandaki\ kelime\ sayısı} \quad (6.3)$$

Elde edilen yeni Jaccard değeri denklem (6.4)'te verilmektedir.

$$benzerlik_{ij} = k_1 * jaccard_değeri_{ij} + k_2 * oran_j \quad (6.4)$$

Burada k_1 ve k_2 toplamı 1 olan ağırlıklardır.

6.4. Uygunluk Fonksiyonu

Popülasyondaki her bir X_i bireyi için veri setindeki bütün yorumlarla olan benzerliğine teker teker bakılmaktadır. Eğer benzerlik değeri belirlenen değerden daha yüksek çıkmış ise o yorum pozitif olarak, düşük çıkmışsa da o yorum negatif olarak kabul edilmektedir. Bu benzerlik işlemi ile bireyin uygunluk değeri denklem (6.5) kullanılarak hesaplanmaktadır. En iyi uygunluk değerine sahip birey en iyi çözüm olmaktadır.

$$uygunluk = c_1 * \frac{DP}{DP+YN} * \frac{DN}{DN+YP} + c_2 * \frac{DP}{DP+YP} + c_3 * \frac{DN}{DN+YN} \quad (6.5)$$

DP; doğru pozitiflerdir ve kuralla aynı sınıf etiketine sahip olan, kural tarafından kapsanan örneklerin sayısıdır. YP; yanlış pozitiflerdir ve kuraldan farklı sınıf etiketine sahip olan, kural tarafından kapsanan örneklerin sayısıdır. YN; yanlış negatiflerdir ve kural tarafından kapsanmayan fakat kuralla aynı sınıf etiketine sahip örneklerin sayısıdır. DN ise; doğru negatiflerdir ve kural tarafından kapsanmayan ve kuralla da aynı etikete sahip olmayan örneklerin sayısıdır. c_1 , c_2 , ve c_3 değerleri ise ağırlıklardır ve bu üç değer toplamı 1'e eşit olmaktadır. Kullanıcı tanımlıdır ve en iyi sonucu verecek şekilde belirlenebilmektedir.

7. DENEYSEL SONUÇLAR

Bu bölümde BOA ve SETO algoritmaları ile duygu analizi yapılarak elde edilen sonuçlar gösterilmektedir. Programda eğitim verilerinden model oluşturularak test verilerinde uygulanmıştır. Bu sonuçlar ile WEKA (URL-4) programındaki gözetimli makine öğrenmesi algoritmalarından elde edilen sonuçlar karşılaştırılmıştır. Performans karşılaştırılması yapılan gözetimli makine öğrenmesi algoritmaları için standart parametre değerleri kullanılmıştır.

Tablo 7.1. Sonuçları karşılaştırmada kullanılan ölçütler

Ölçüt	MATEMATİKSEL DENKLEM
Doğru etiketlenen veri sayısı/ toplam veri sayısı	$(DP + DN)/(DP + YP + DN + YN)$
Doğruluk yüzdesi	$\frac{DP + DN}{DP + YP + DN + YN} \times 100$
Kesinlik	$\frac{DP}{DP + YP}$
Hassasiyet	$\frac{DP}{DP + YN}$
F-ölçütü	$\frac{2 \times DP}{2 \times DP + YP + YN}$
MCC	$\frac{(DP \times DN) - (YP \times YN)}{\sqrt{(DP + YP) \times (DP + YN) \times (DN + YP) \times (DN + YN)}}$

Sonuçlar karşılaştırılırken literatürde en çok kullanılan ölçütler kullanılmıştır. Bu ölçütler ve hesaplamaları Tablo 7.1’de verilmektedir. Doğruluk yüzdesi; doğru tahmin edilen veri sayısının veri setindeki toplam verilere yüzdelik oranını vermektedir. Kesinlik (Precision); doğru tahmin edilen pozitif gözlemlerin toplam tahmini pozitif gözlemlere oranı olmaktadır. Diğer bir deyişle, pozitif olarak tanımlananlardan aslında kaç tanesinin gerçekte pozitif olduğunun oranını vermektedir. YP değeri maliyetli olduğu durumlarda, kesinlik iyi bir ölçümdür.

Hassasiyet (Recall); modelin gerçek pozitiflerin kaç tanesinin pozitif (DP) olarak etiketlendiğini hesaplamaktadır. YN değeri maliyetli olduğu durumlarda, en iyi modeli seçmek için bu ölçüm kullanılmaktadır. F-Ölçütü; kesinlik ve hassasiyet arasında bir denge arandığı zaman kullanılmaktadır. Kesinlik ve hassasiyetin harmonik ortalamasıdır. MCC (Matthews Correlation Coefficient - Matthews Korelasyon Katsayısı); doğru ve yanlış dikkate almaktadır ve genellikle sınıflar çok farklı boyutta olsalar bile kullanılabilen dengeli bir ölçü olarak kabul edilmektedir.

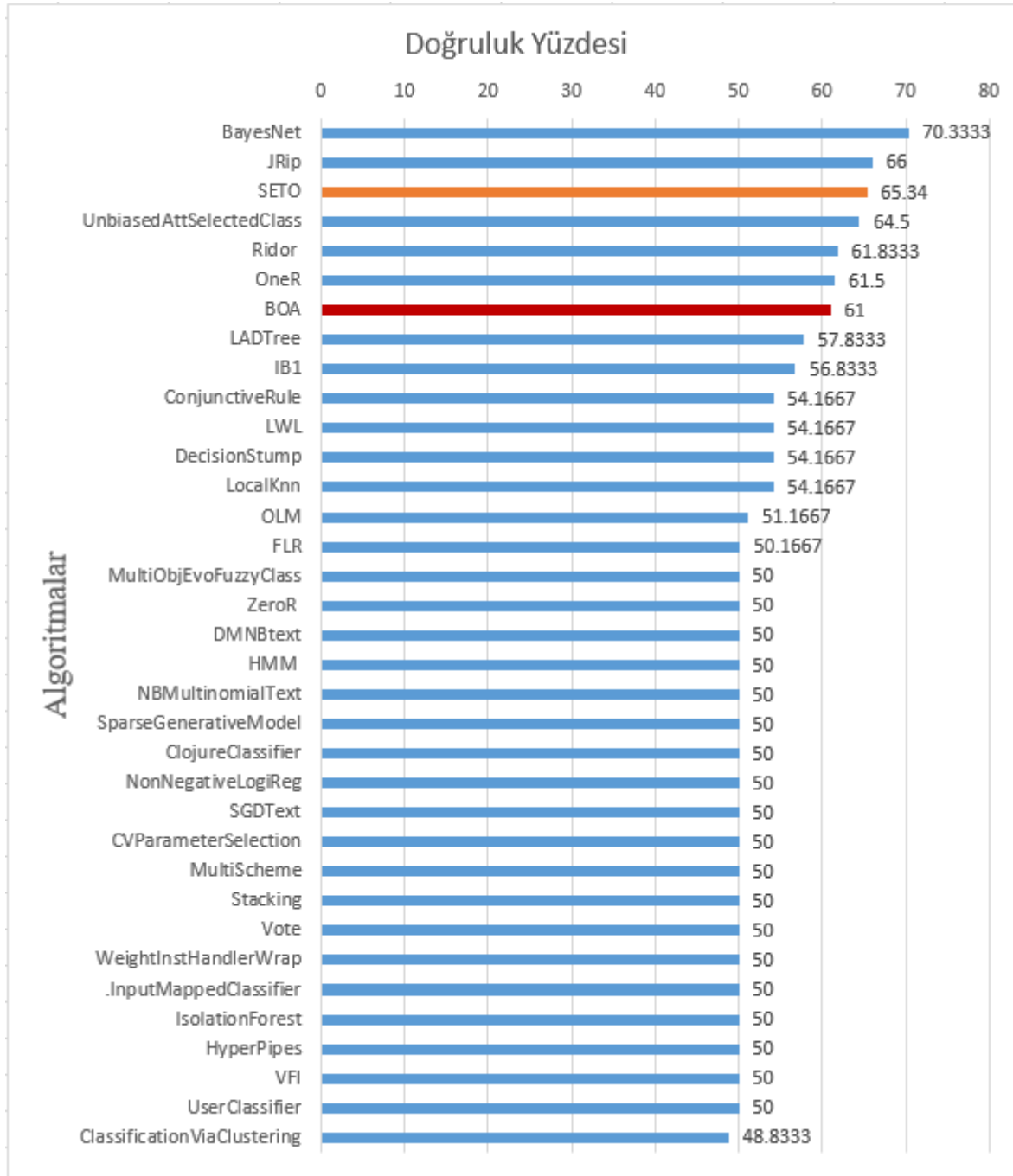
7.1. IMDB Veri Setinden Elde Edilen Sonuçlar

IMDB veri seti için geliştirilen algorithmada kullanılan uygunluk fonksiyonu ve benzerlik katsayıları Tablo 7.2’de gösterilmektedir.

Tablo 7.2. IMDB veri seti için kullanılan katsayılar

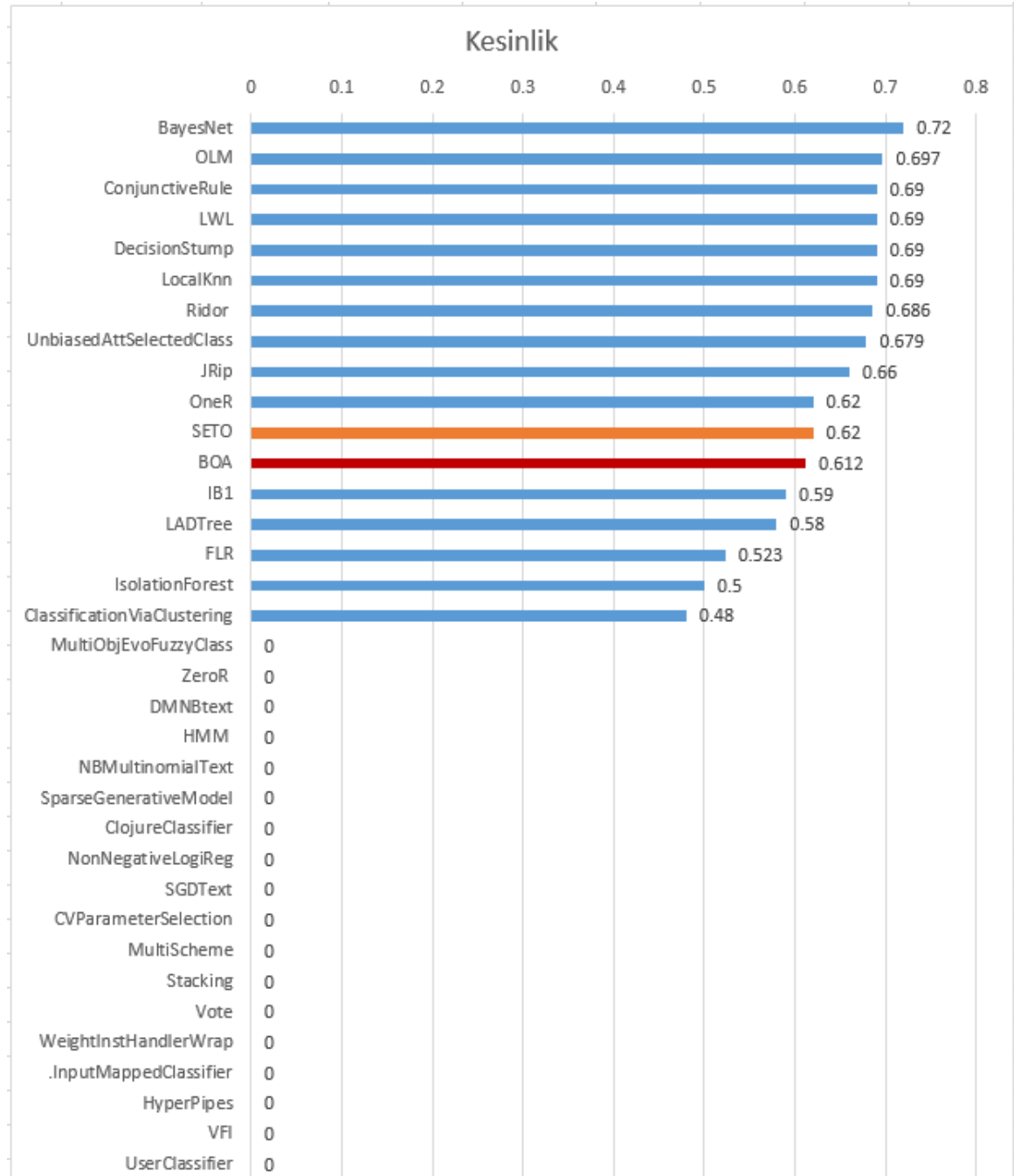
Algoritma	Benzerlik			Uygunluk Fonksiyonu		
	<i>benzerlik</i>	k_1	k_2	c_1	c_2	c_3
SETO	≥ 0.1	0.8	0.2	0.6	0.2	0.2
BOA	≥ 0.15	0.8	0.2	0.6	0.2	0.2

IMDB veri seti için SETO ve BOA’dan alınan sonuçlar ile toplam 33 tane algorithmadan elde edilen sonuçlar beş ölçüt için karşılaştırılmıştır. Sonuçlar kullanılan ölçüte göre büyükten küçüğe doğru sıralanmıştır. Doğruluk yüzdesi ölçütü göz önüne alındığında, Şekil 7.1’de görüldüğü gibi SETO bu 35 algoritma içinde en iyi üçüncü algoritma olurken, BOA ise yedinci en iyi algoritma olmuştur.



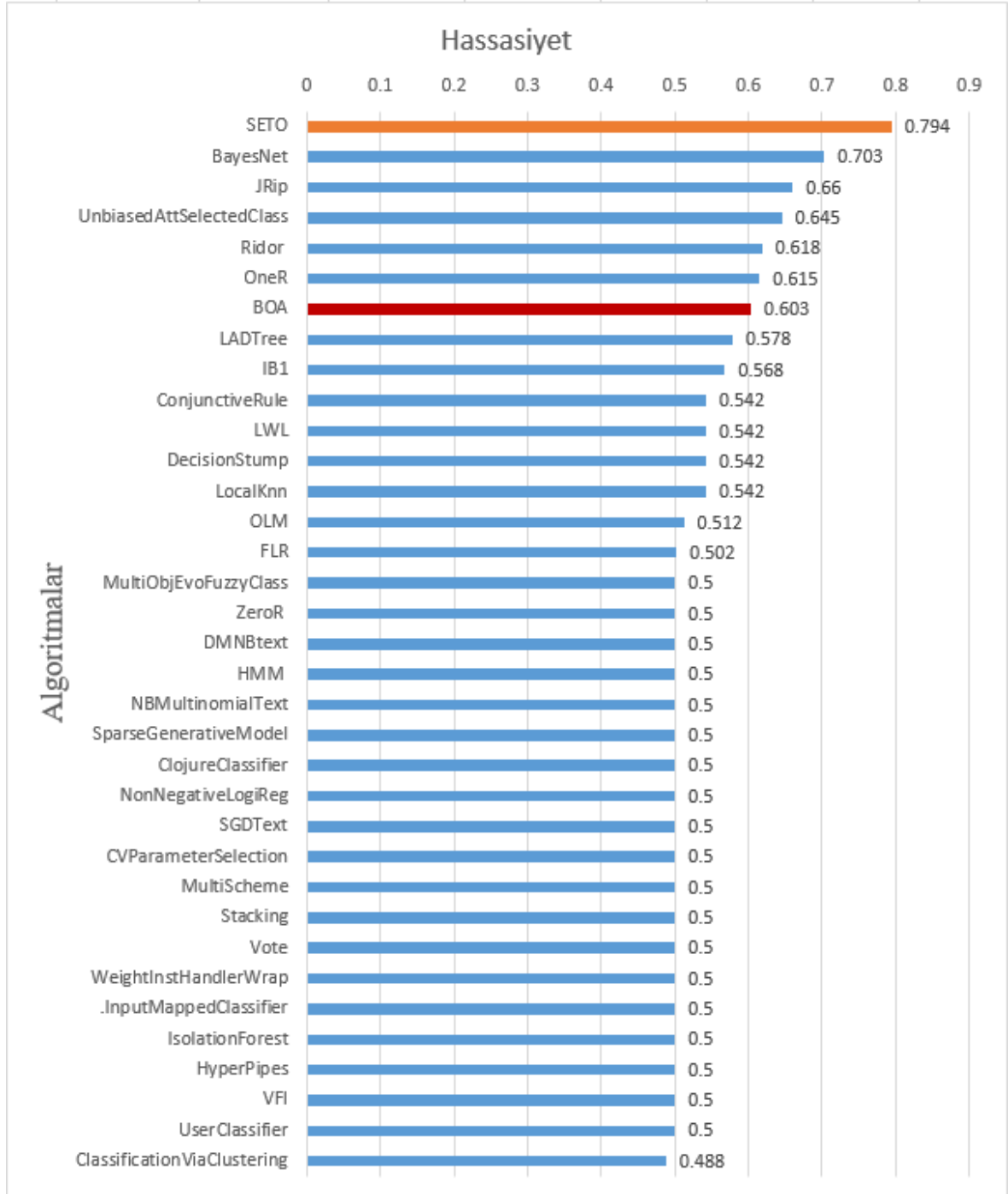
Şekil 7.1. Doğruluk yüzdesi ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar

Kesinlik ölçütüne göre, SETO ve BOA Şekil 7.2’de görüldüğü gibi ortalamanın üstünde bir değer vermiştir.



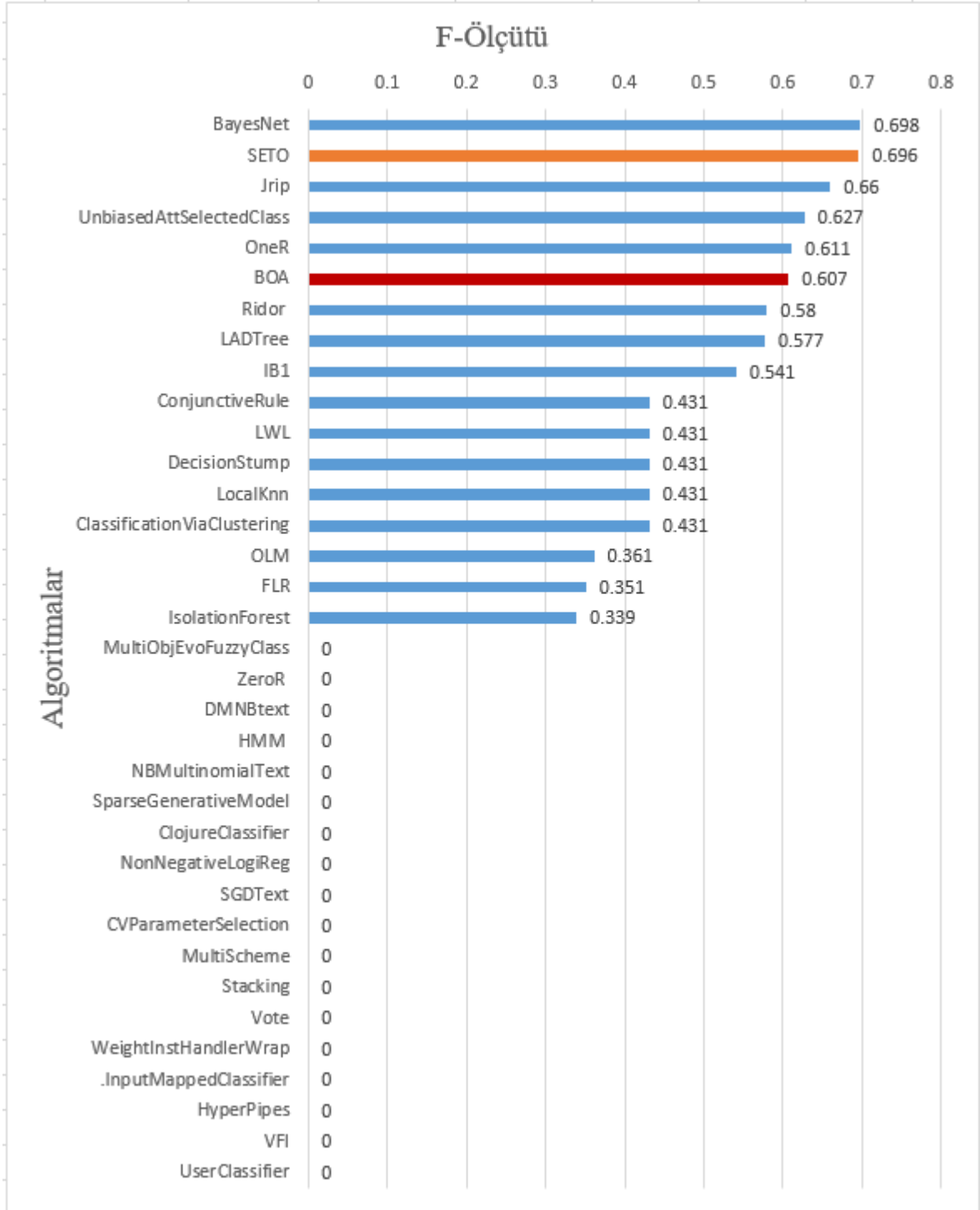
Şekil 7.2. Kesinlik ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar

Hassasiyet ölçütünde Şekil 7.3'te görüldüğü gibi SETO en iyi sonucu verirken BOA ise yedinci olmuştur.



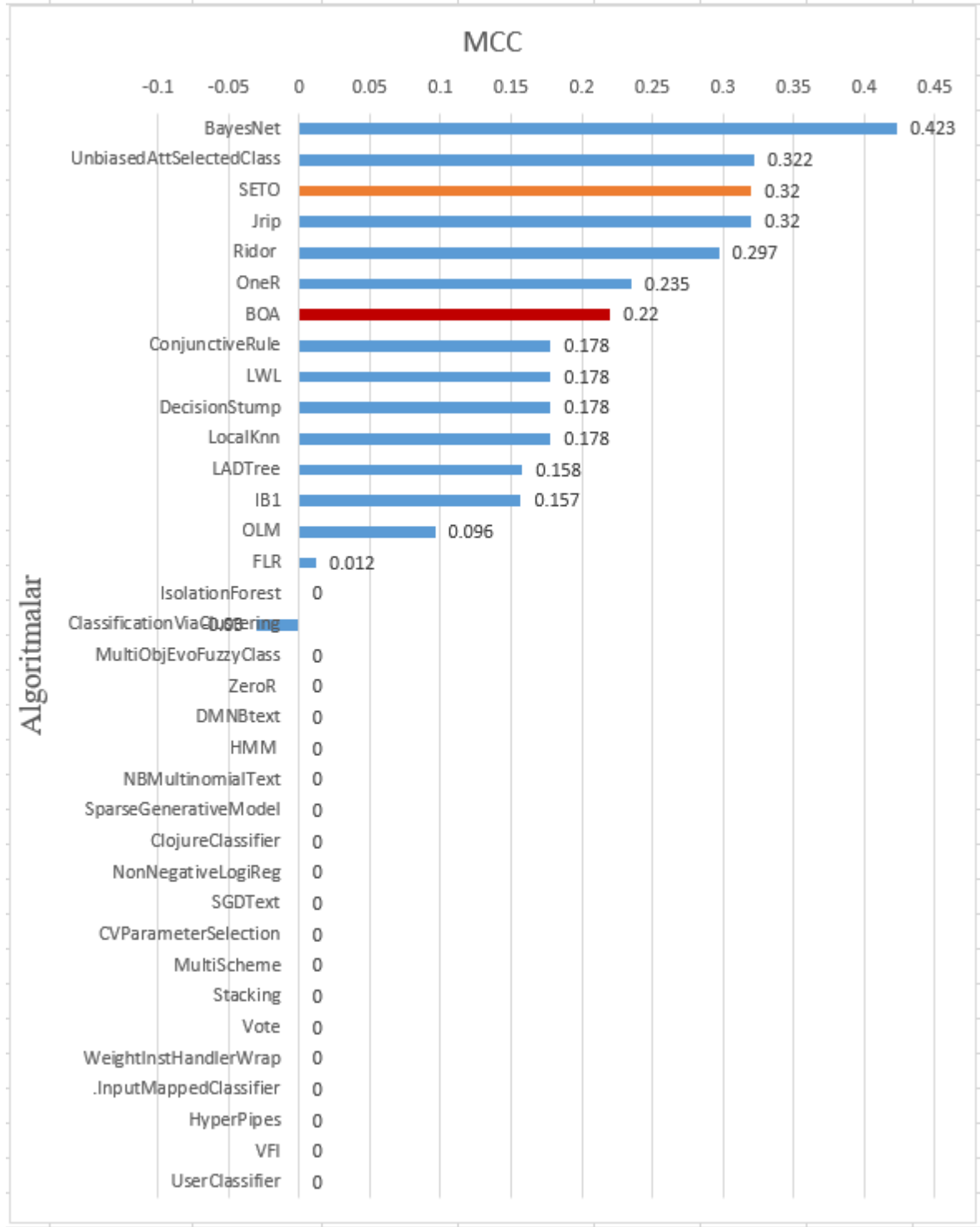
Şekil 7.3. Hassasiyet ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar

Şekil 7.4'teki F-Ölçütü değerlerine bakılacak olursa, SETO ikinci en iyi sonucu verirken BOA ise altıncı olmuştur.



Şekil 7.4. F-Ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar

MCC ölçütü için Şekil 7.5’te görüldüğü gibi SETO üçüncü en iyi sonucu verirken BOA yedinci en iyi sonucu vermiştir.



Şekil 7.5. MCC ölçütü kullanılarak IMDB veri setinden elde edilen sonuçlar

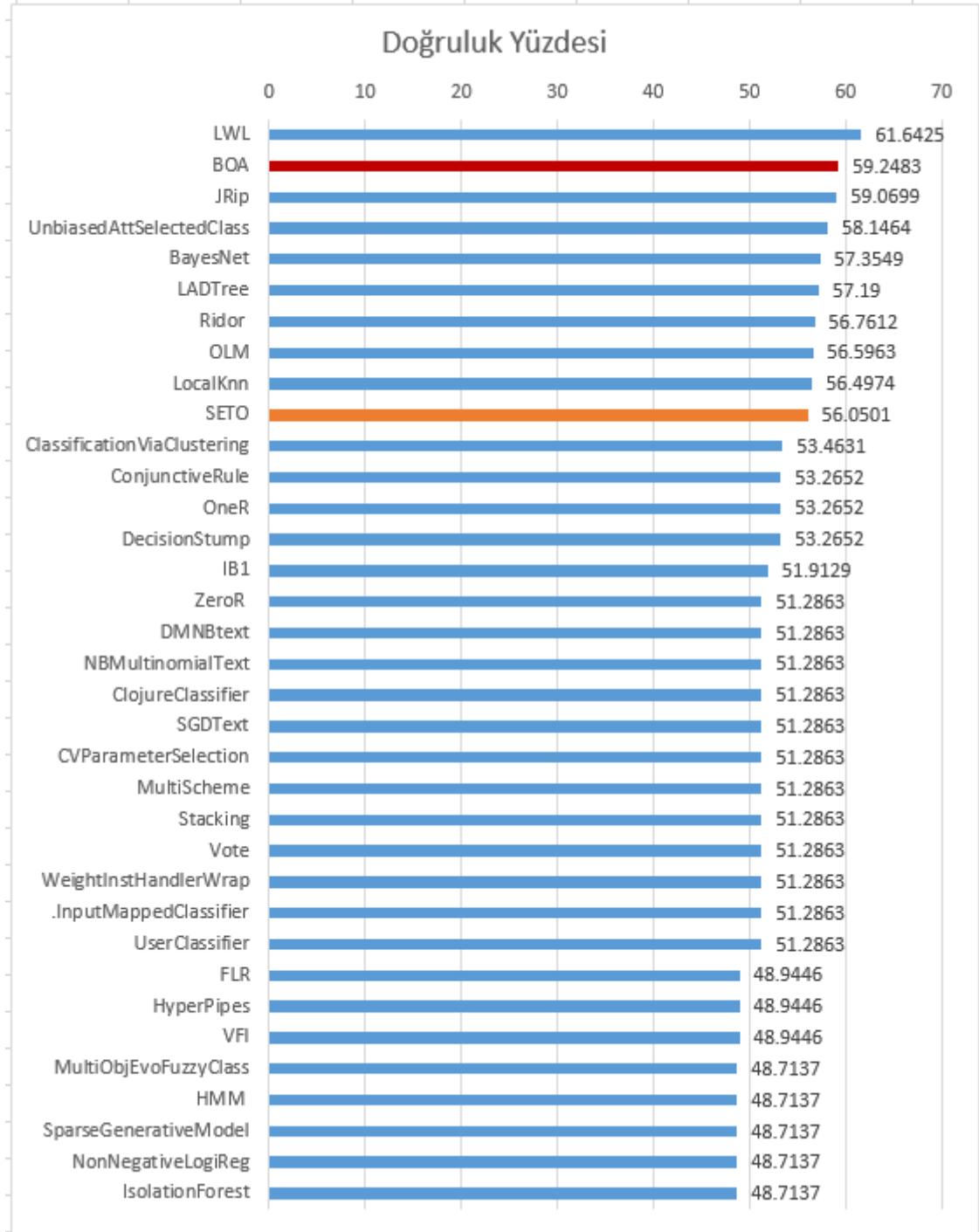
7.2. Polarity Veri Setinden Elde Edilen Sonuçlar

Polarity veri seti için kullanılan uygunluk fonksiyonu ve benzerlik katsayıları Tablo 7.3'te gösterilmektedir.

Tablo 7.3. Polarity veri seti için kullanılan katsayılar

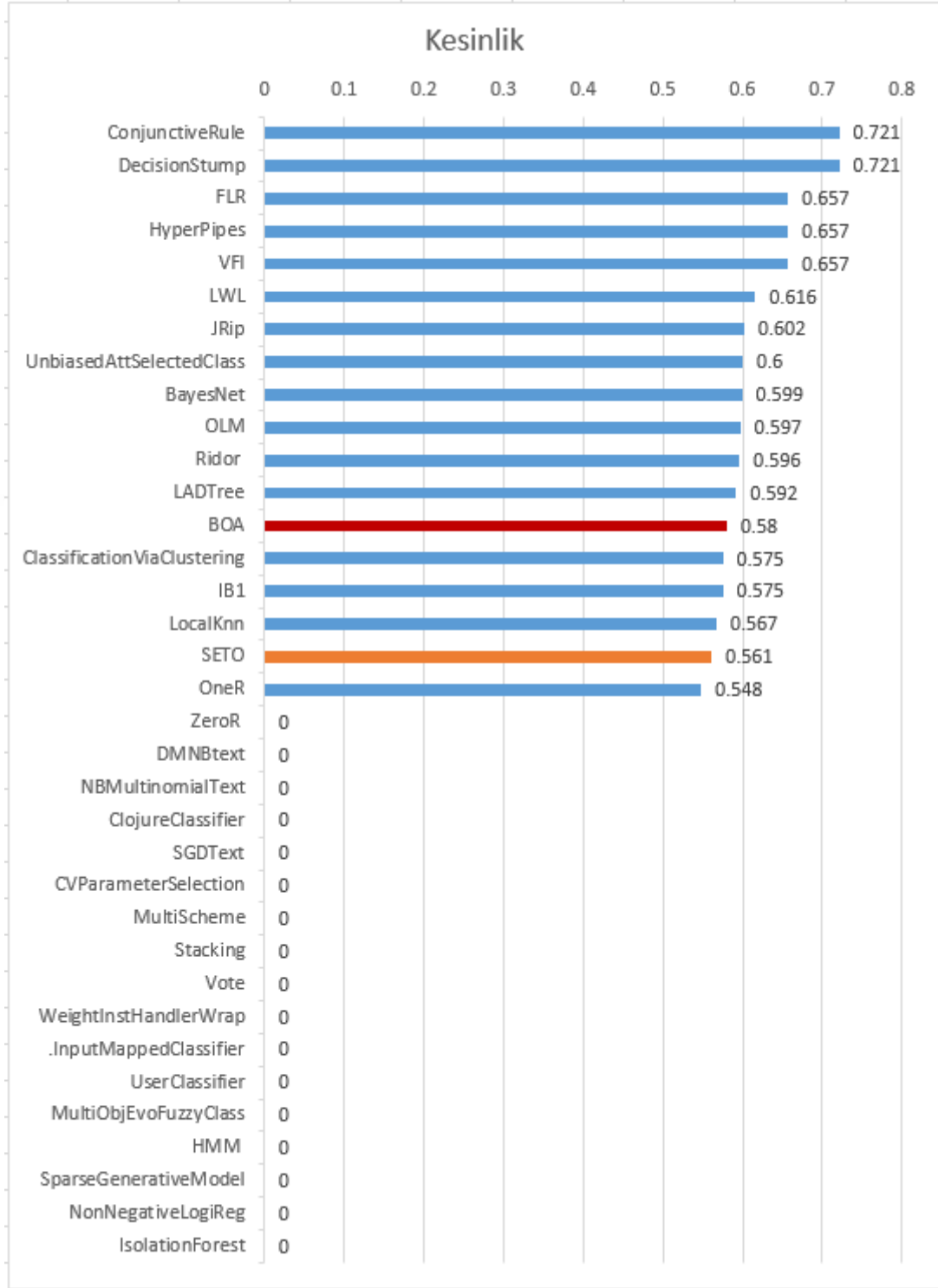
Algoritma	Benzerlik			Uygunluk Fonksiyonu		
	<i>benzerlik</i>	k_1	k_2	c_1	c_2	c_3
SETO	≥ 0.3	0.8	0.2	0.45	0.45	0.1
BOA	≥ 0.6	0.8	0.2	0.6	0.3	0.1

Polarity veri seti için algoritmalar çalıştırıldığında sonuçlar her bir ölçüte göre azalan sırada listelenmiştir. Buna göre; Şekil 7.6’da görüldüğü gibi doğruluk yüzdesi ölçütü göz önüne alındığında, 35 algoritmanın içinden BOA en iyi ikinci sonucu verirken, SETO ise onuncu sırada yer almaktadır.



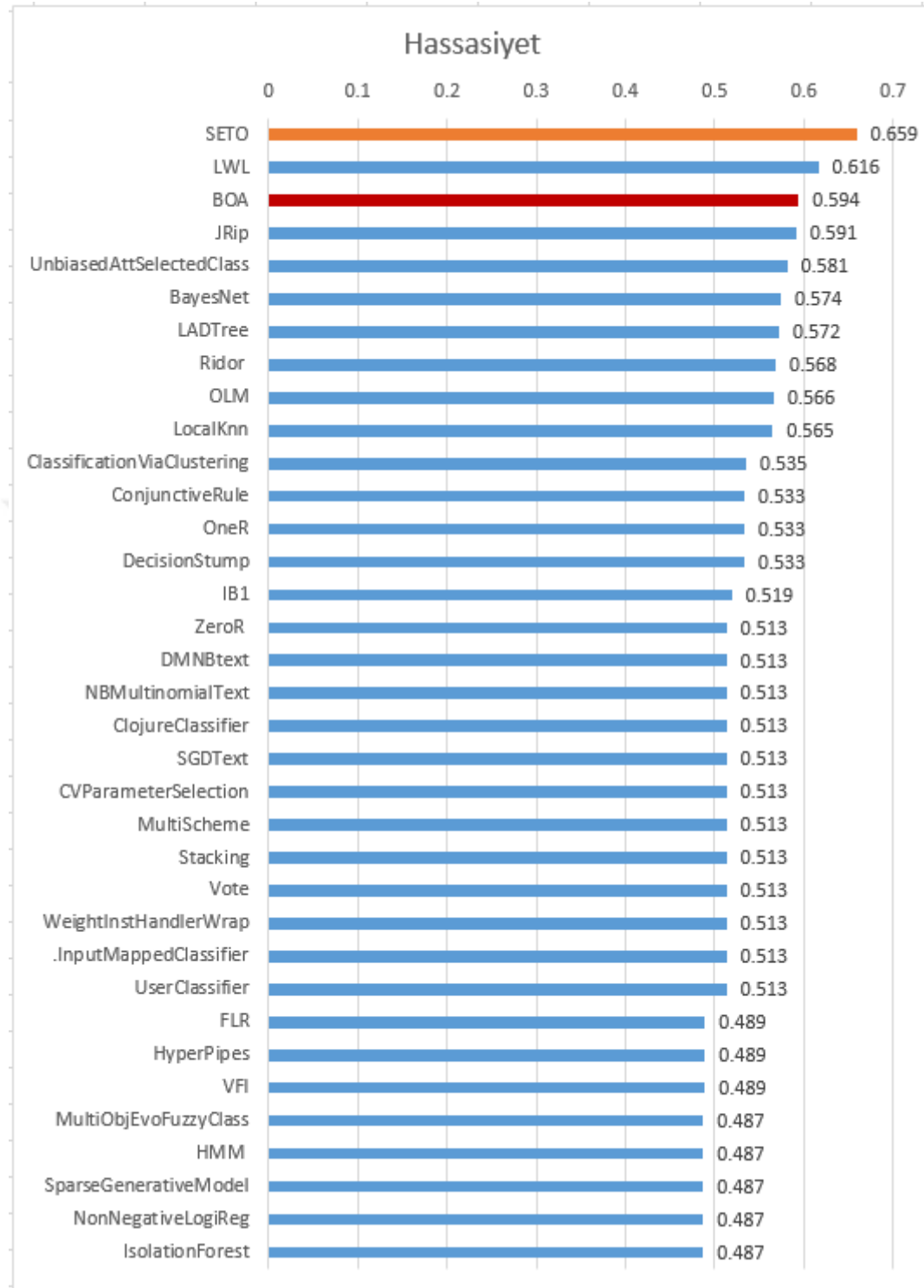
Şekil 7.6. Doğruluk yüzdesi ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar

Şekil 7.7’deki kesinlik ölçütüne bakıldığında SETO ve BOA’nın ortalamanın üstünde değer bulduğu gözlemlenmektedir.



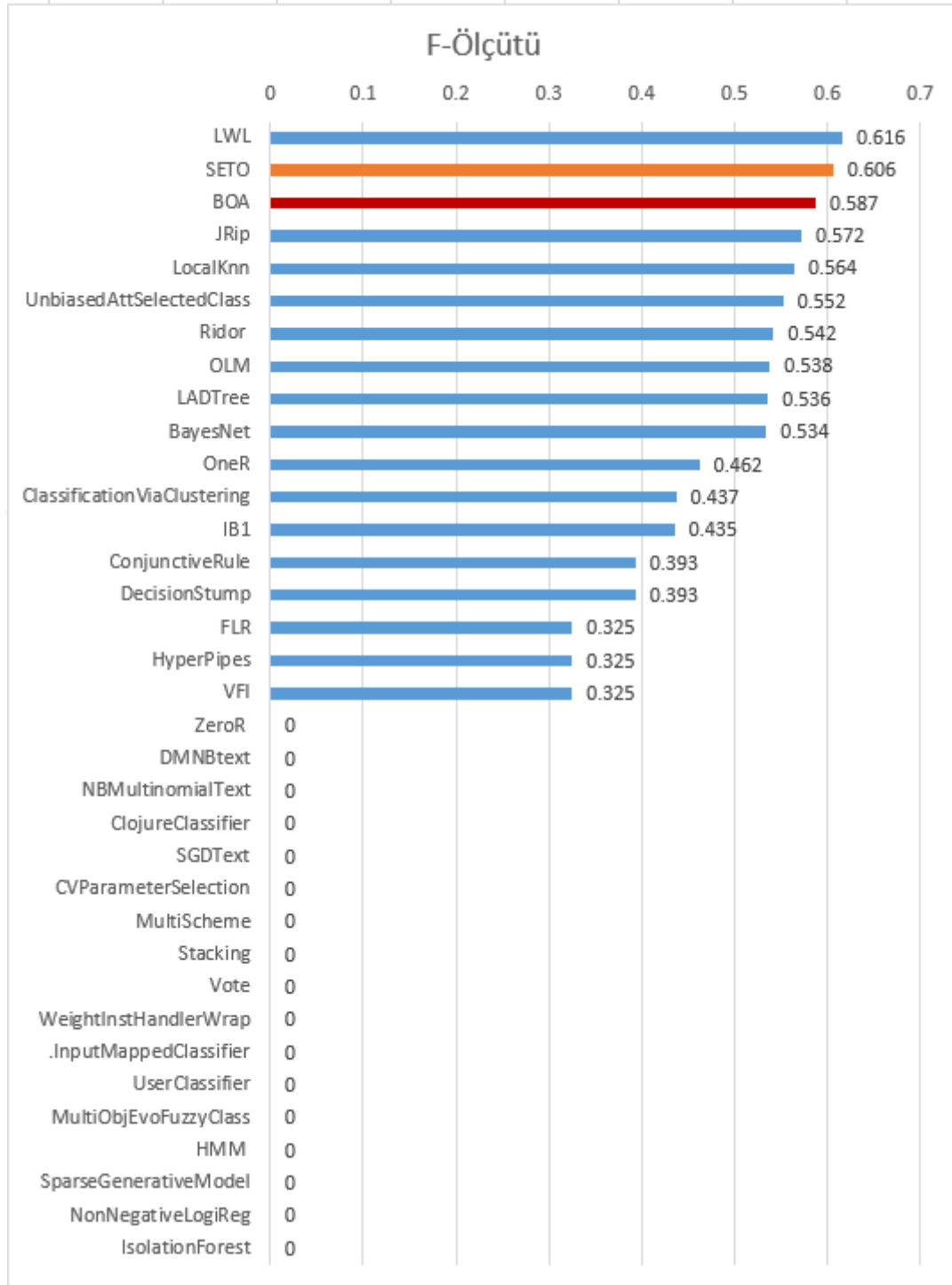
Şekil 7.7. Kesinlik ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar

Şekil 7.8'e bakılacak olunursa, hassasiyet değerinde ise en iyi sonucu BOA bulmuştur. SETO ise en iyi üçüncü sonucu bulmuştur.



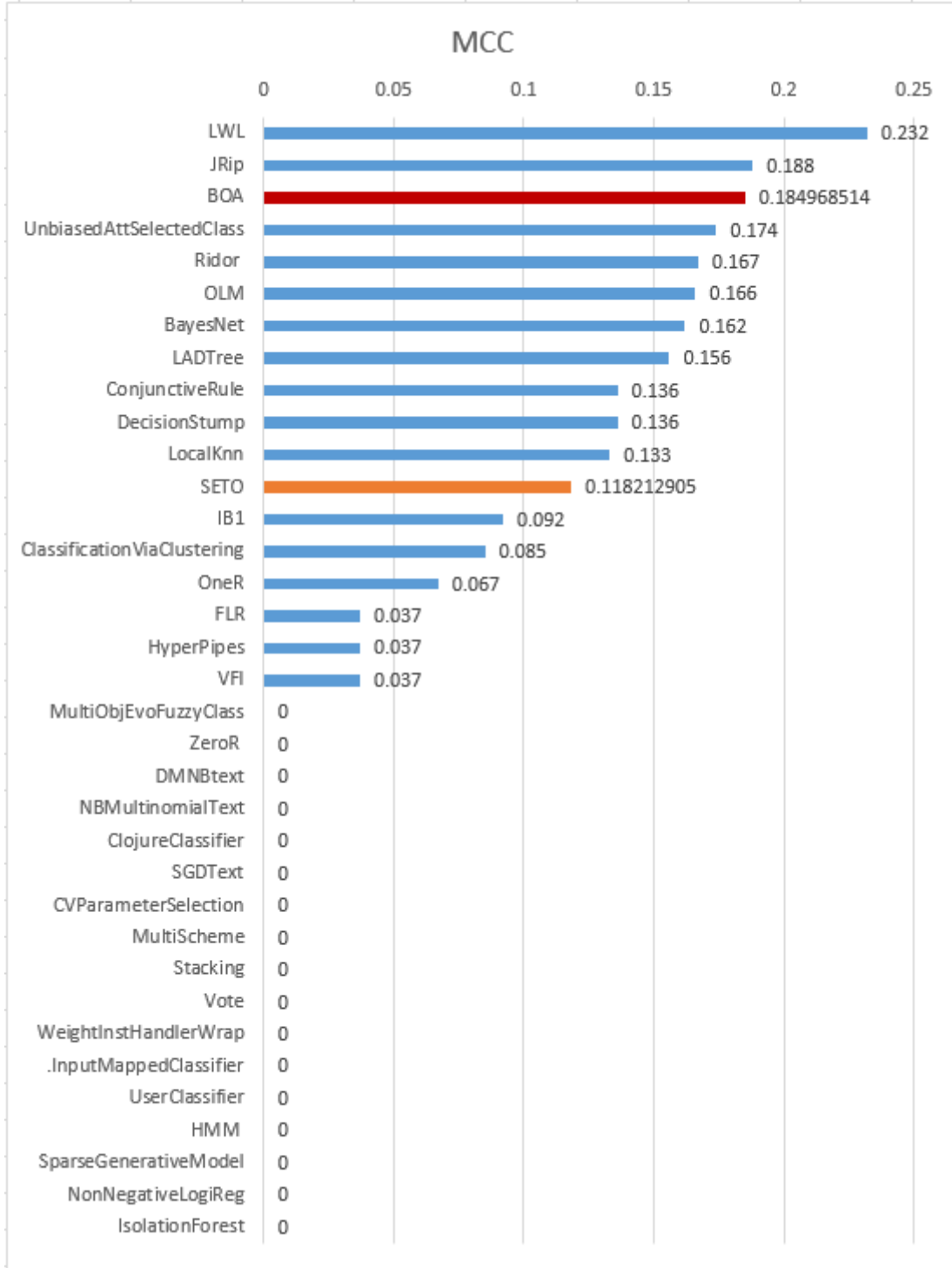
Şekil 7.8. Hassasiyet ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar

Şekil 7.9'a göre F-Ölçütü değerinde SETO en iyi ikinci sonucu bulurken, BOA ise üçüncü olmuştur.



Şekil 7.9. F-Ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar

Son olarak MCC ölçütünde ise Şekil 7.10'a göre, BOA en iyi sonuç veren üçüncü algoritma olurken, SETO ise on ikinci sırada olmaktadır.



Şekil 7.10. MCC ölçütü kullanılarak Polarity veri setinden elde edilen sonuçlar

7.3. Amazon Veri Setinden Elde Edilen Sonuçlar

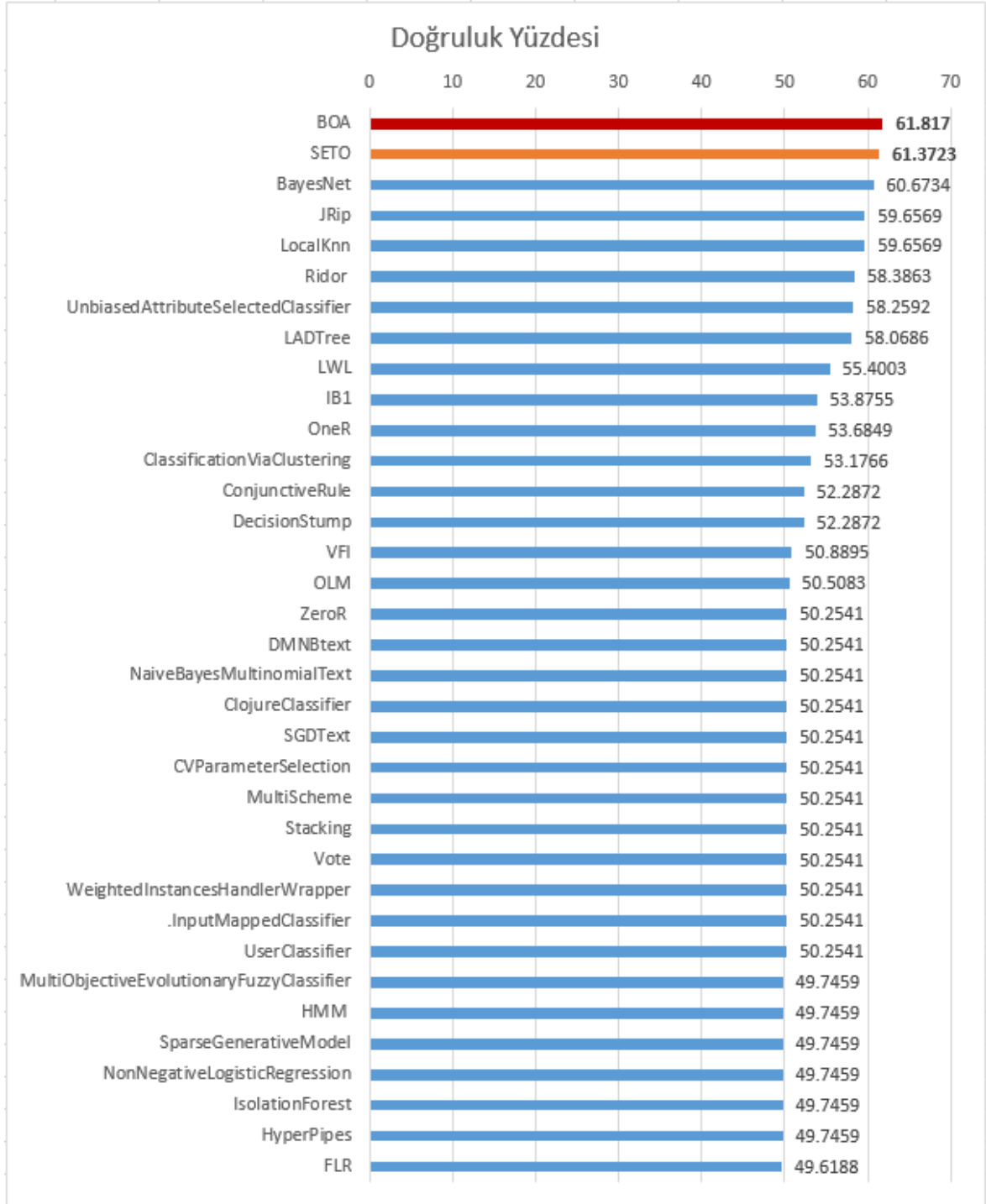
Amazon veri seti için uygulamalarda kullanılan uygunluk fonksiyonu ve benzerlik katsayıları Tablo 7.4'te gösterilmektedir.

Tablo 7.4. Amazon veri seti için kullanılan katsayılar

Algoritma	Benzerlik			Uygunluk Fonksiyonu		
	<i>benzerlik</i>	k_1	k_2	c_1	c_2	c_3
SETO	≥ 0.3	0.8	0.2	0.5	0.3	0.2
BOA	≥ 0.2	0.8	0.2	0.45	0.45	0.1

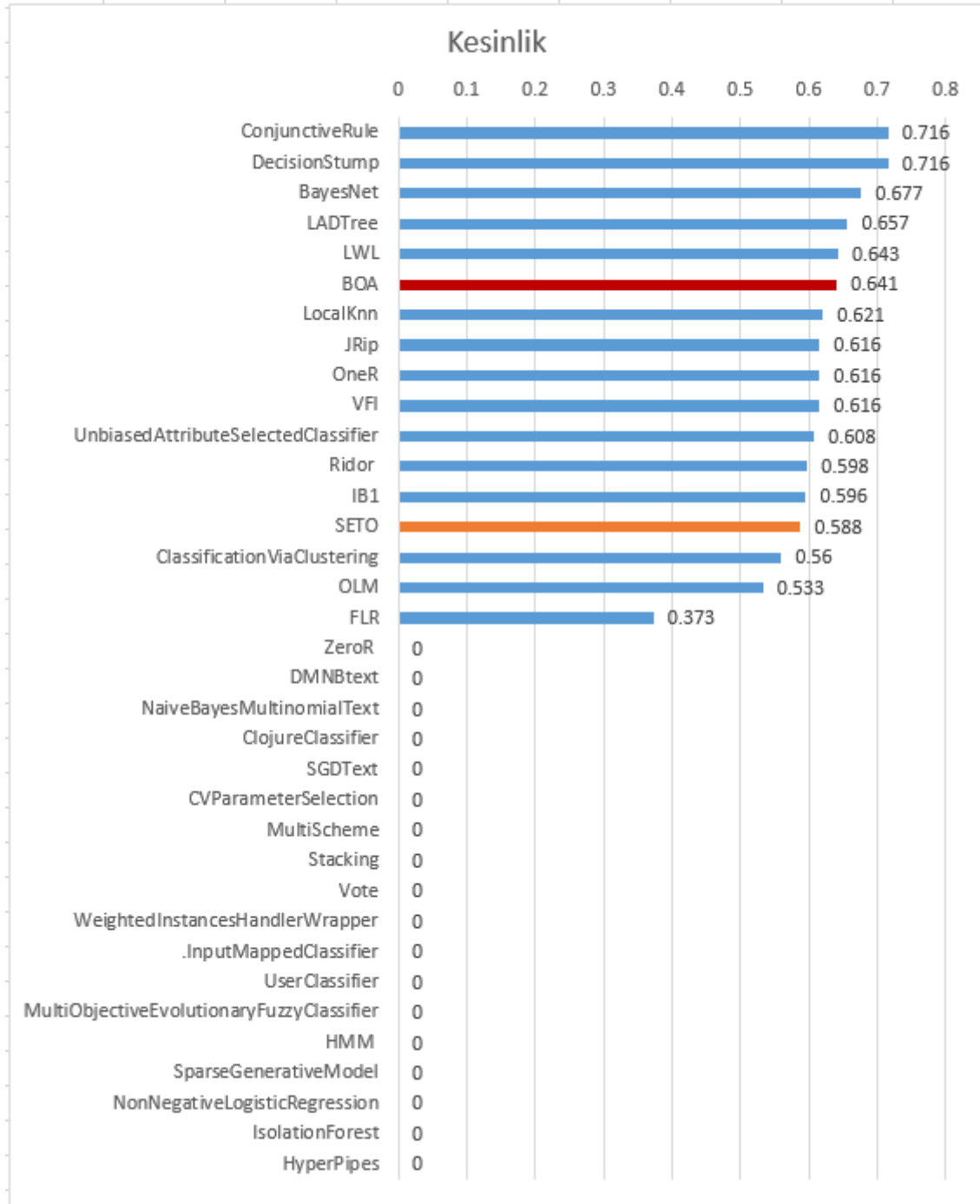
Amazon veri seti için program çalıştırıldığında elde edilen sonuçlar değerlendirme ölçütlerine göre sıralanmaktadır. Doğruluk yüzdesine göre, Şekil 7.11’de görüldüğü gibi BOA’nın en iyi sonucu verdiği SETO’nun ise ikinci en iyi sonucu verdiği görülmektedir.





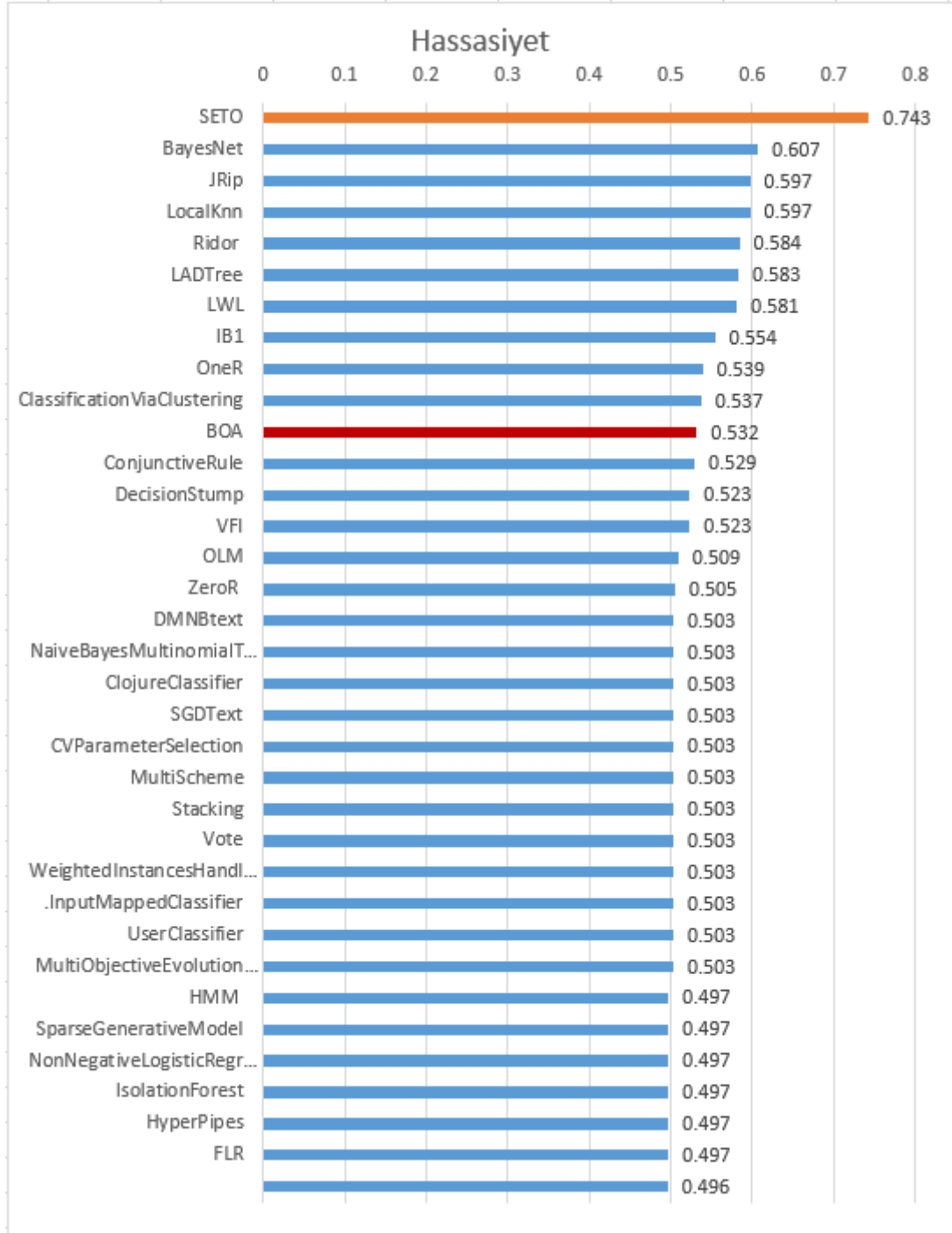
Şekil 7.11. Doğruluk yüzdesi ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar

Şekil 7.12’deki kesinlik ölçütüne bakıldığında, BOA’nın altıncı, SETO’nun ise on dördüncü olduğu görülmektedir.



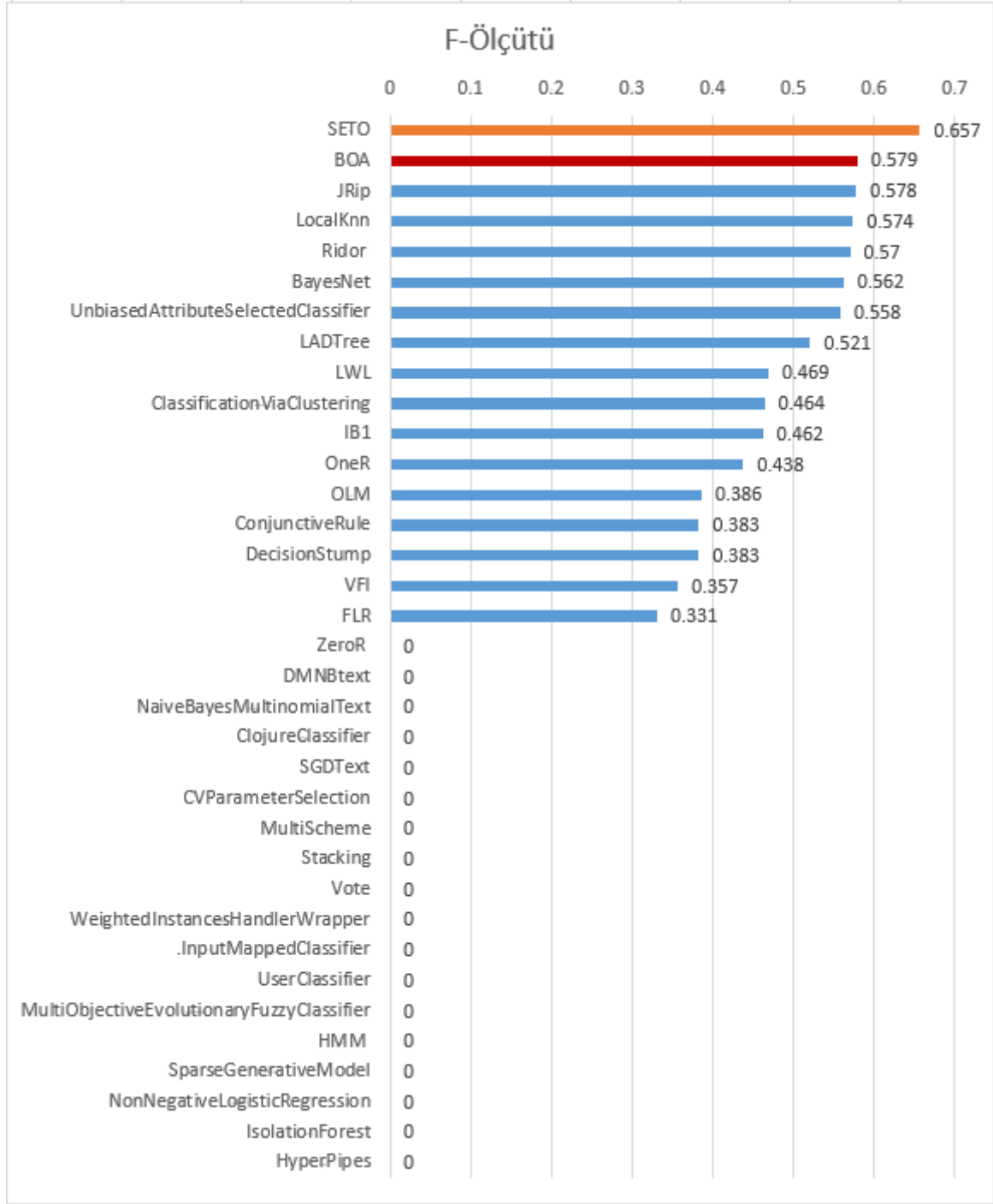
Şekil 7.12. Kesinlik ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar

Şekil 7.13'teki hassasiyet ölçütünde ise SETO'nun yine en iyi sonucu verdiği görülmektedir. BOA ise on birinci sırada yer almaktadır.



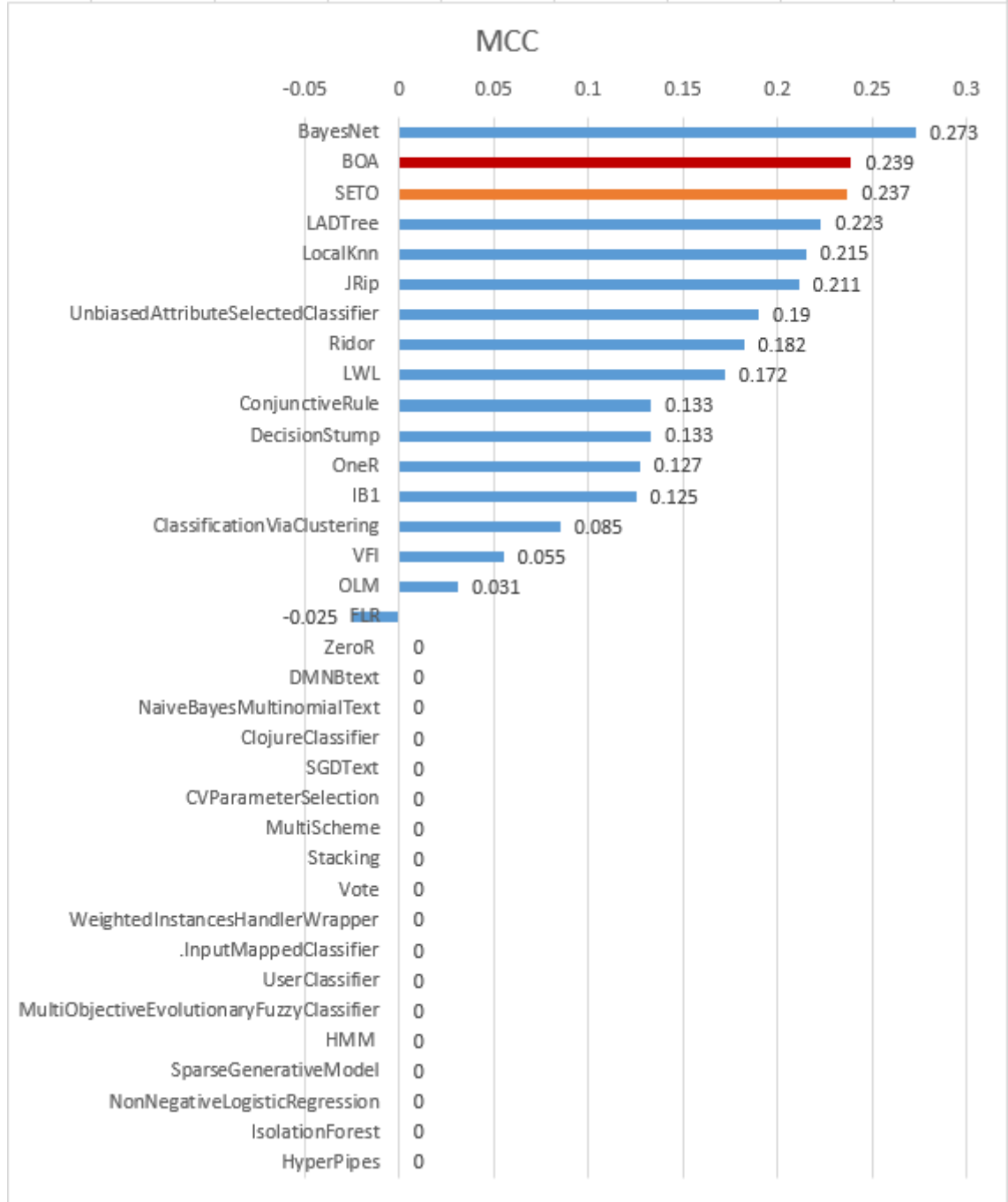
Şekil 7.13. Hassasiyet ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar

Şekil 7.14'teki F-Ölçütü'ne bakılacak olunursa; yine en iyi sonucu SETO'nun, ikinci en iyi sonucu ise BOA'nın verdiği görülmektedir.



Şekil 7.14. F-Ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar

Şekil 7.15'e bakılacak olunursa, MCC ölçütünde ise en iyi ikinci sonuç BOA'ya ait olurken üçüncü sonuç ise SETO'ya ait olmaktadır. Bütün ölçütler göz önüne alındığında SETO ve BOA'nın ikisinin de en iyi sonuçları verdiği görülmektedir.

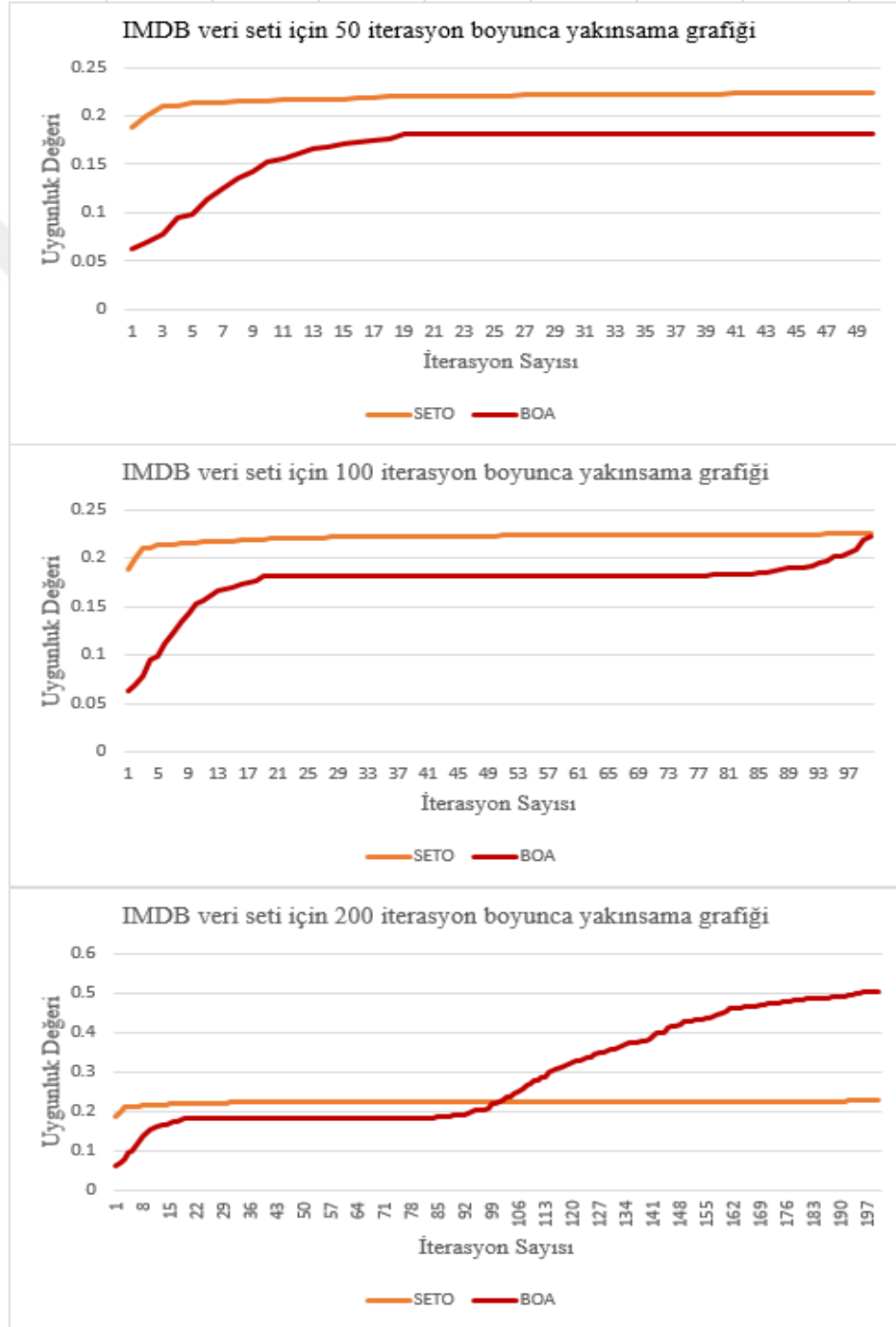


Şekil 7.15. MCC ölçütü kullanılarak Amazon veri setinden elde edilen sonuçlar

7.4. SETO ve BOA'nın Yakınsama Grafikleri

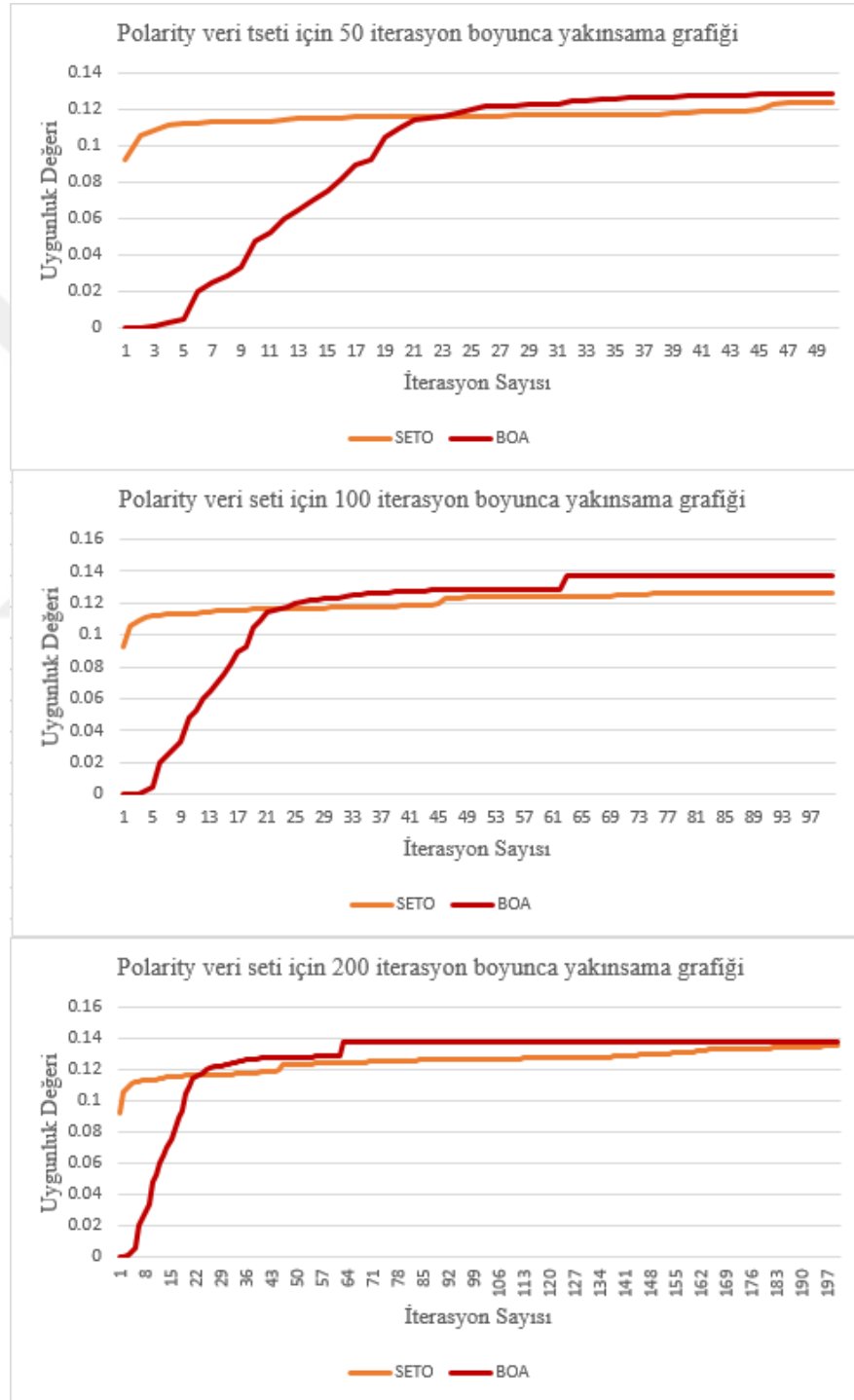
Şekil 7.16, Şekil 7.17 ve Şekil 7.18'deki grafiklerde SETO ve BOA'nın sırasıyla IMDB, Polarity ve Amazon veri setlerindeki yakınsama eğrileri görülmektedir. Algoritmalar iterasyon sayısı 50, 100 ve 200 olacak şekilde beşer defa çalıştırılmıştır ve

elde edilen uygunluk değerlerinin ortalaması alınarak grafikler oluşturulmuştur. Şekil 7.16'ya bakılacak olunursa IMDB veri seti için, SETO daha erken yakınsayıp çok değişiklik göstermemektedir. BOA, SETO'ya göre daha geç yakınsamakta, 20-85 iterasyon arasında sabit kalmaktadır ve 85. iterasyondan sonra uygunluk değerinde tekrar bir artış gözlemlenmektedir.



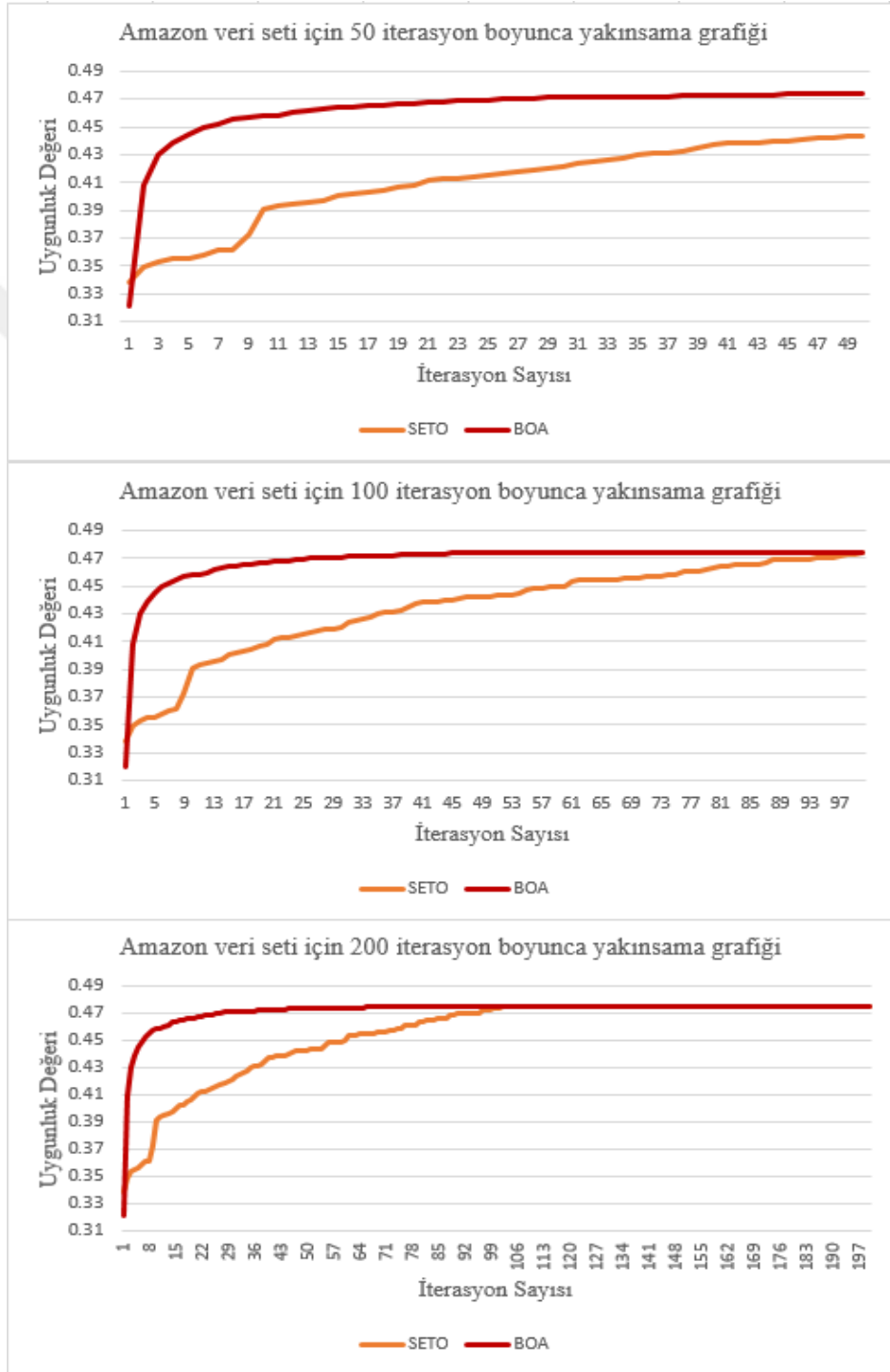
Şekil 7.16. SETO ve BOA'nın IMDB veri seti için yakınsama grafiği

Şekil 7.17’de Polarity veri seti için 50, 100 ve 200 iterasyon boyunca yakınsama grafikleri görülmektedir. Buna göre SETO’da uygunluk değeri daha yüksek bir noktadan başlayıp hızlı bir şekilde yakınsamıştır ve durağan bir eğri göstermektedir. BOA’da ise ilk 20 iterasyonda hızlı bir yakınsama görülürken, sonrasında sadece ufak değişiklikler gözlemlenmektedir.



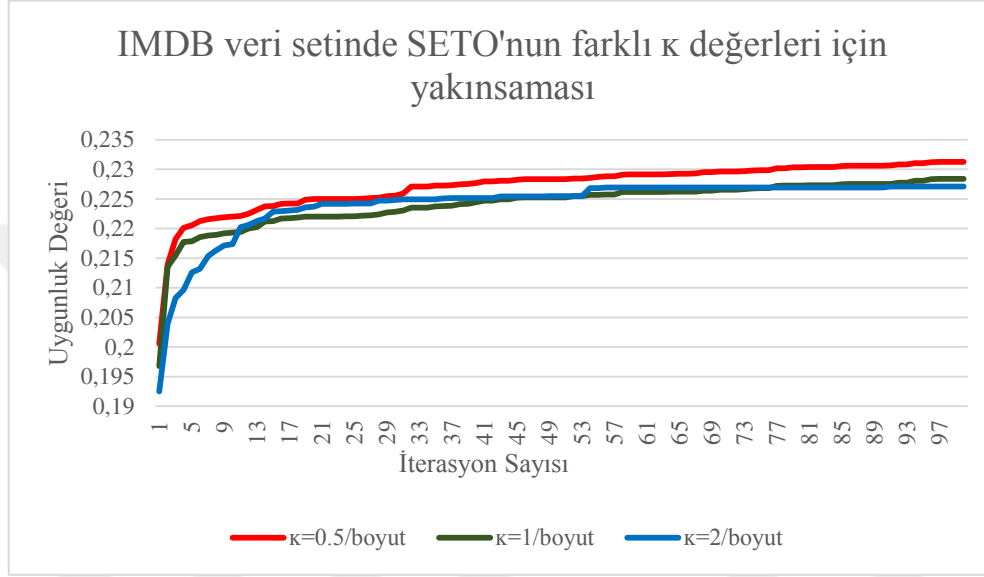
Şekil 7.17. SETO ve BOA’nın Polarity veri seti için yakınsama grafiği

Amazon veri seti için BOA ve SETO'nun 50, 100 ve 200 iterasyon boyunca elde edilen uygunluk değerlerinin yakınsama grafikleri Şekil 7.18'de görülmektedir. Bu grafiklere göre BOA, SETO'ya göre daha hızlı bir şekilde yakınsama göstermiştir. 100 iterasyondan sonra ise yakın değerler göstererek, uygunluk değerinde fazla bir değişiklik olmamıştır.



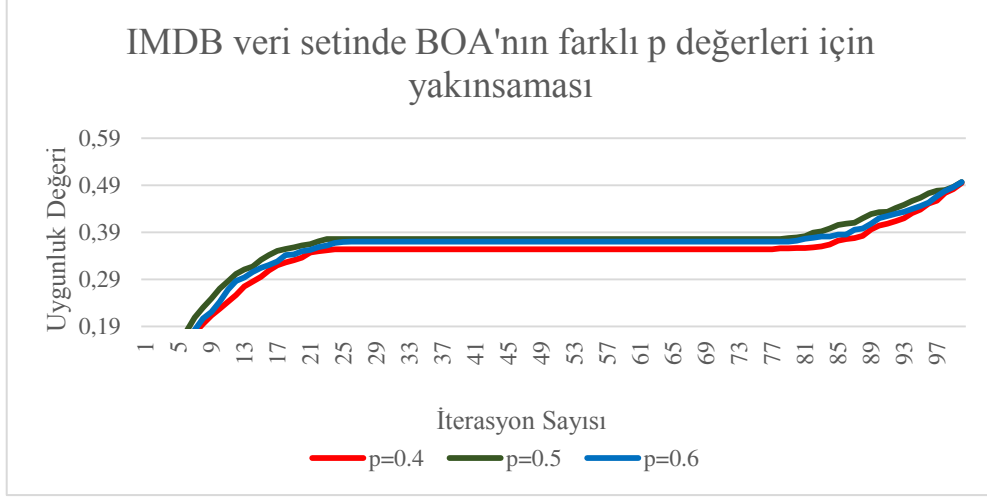
Şekil 7.18. SETO ve BOA'nın Amazon veri seti için yakınsama grafiği

Şekil 7.19’da IMDB veri setinde SETO’daki κ değişkenine 0.5/boyut, 1/boyut ve 2/boyut değerleri verilerek 100 iterasyon boyunca uygunluk değerinin yakınsaması gösterilmektedir. *boyut* veri setinden elde edilen kelime sayısına denk gelmektedir. Buna göre bu değişkene bağlı olarak, her üç değerde de birbirine yakın uygunluk değerleri elde edilmektedir.



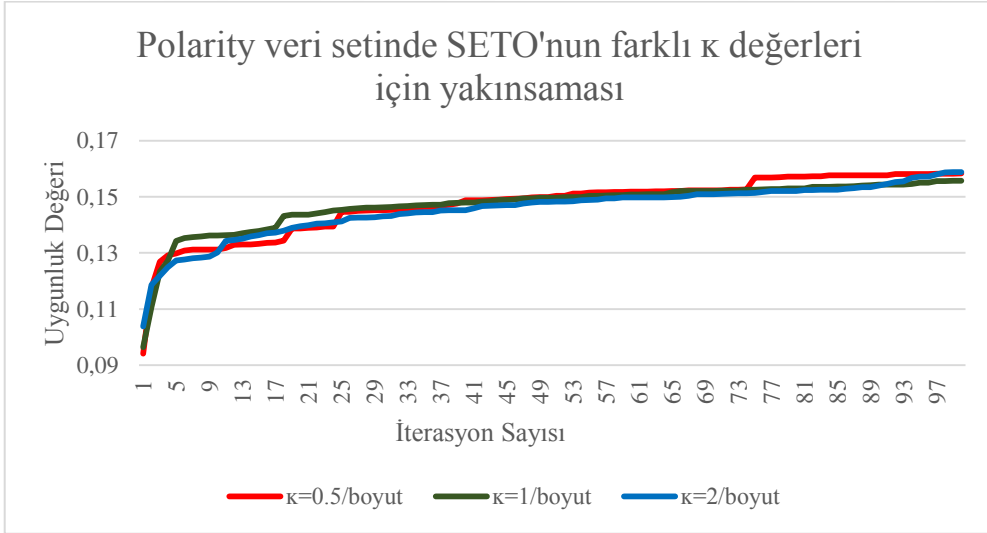
Şekil 7.19. IMDB veri setinde SETO’nun farklı κ değerleri için yakınsaması

Şekil 7.20’de IMDB veri setinde BOA’nın balinanın hangi hareketi uygulayacağını belirleyen p değişkenine 0.4, 0.5 ve 0.6 değerleri verildiğinde, 100 iterasyon boyunca uygunluk değerinin yakınsama grafiği gösterilmektedir. Buna göre her üç değerde de uygunluk değerlerinin birbirine yakın sonuçlar verdiği görülmektedir.



Şekil 7.20. IMDB veri setinde BOA'nın farklı p değerleri için yakınsaması

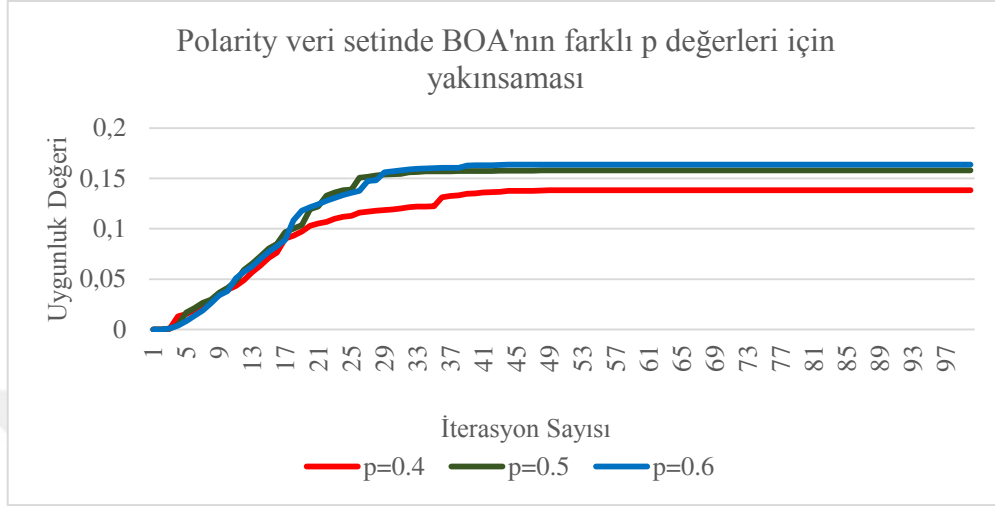
Şekil 7.21'de Polarity veri setinde 100 iterasyon boyunca SETO'nun farklı κ değişkeninin 0.5/boyut, 1/boyut ve 2/boyut değerlerini aldığındaki uygunluk değerinin yakınsama grafiği sunulmaktadır. Grafiğe göre $\kappa=1/boyut$ iken daha hızlı bir yakınsama gösterirken bir süre sonra, üç değerin birbirine yakın sonuçlar gösterdiği görülmektedir.



Şekil 7.21. Polarity veri setinde SETO'nun farklı κ değerleri için yakınsaması

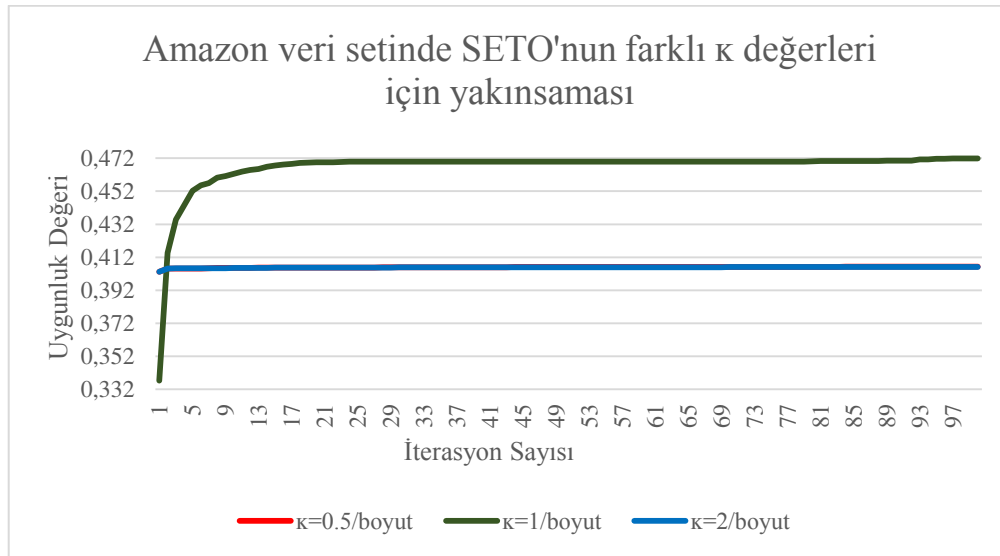
Şekil 7.22'de Polarity veri setinde BOA için p değişkeninin 0.4, 0.5 ve 0.6 değerlerini aldığı zamanki uygunluk değerinin değişim grafiği gösterilmektedir. Grafiğe göre BOA

için $p=0.4$ seçilmişken, diğerlerine göre daha düşük bir uygunluk değeri elde edilirken; 0.5 ve 0.6 değerlerini kullanan BOA’da ise birbirine yakın sonuçlar elde edilmiştir.



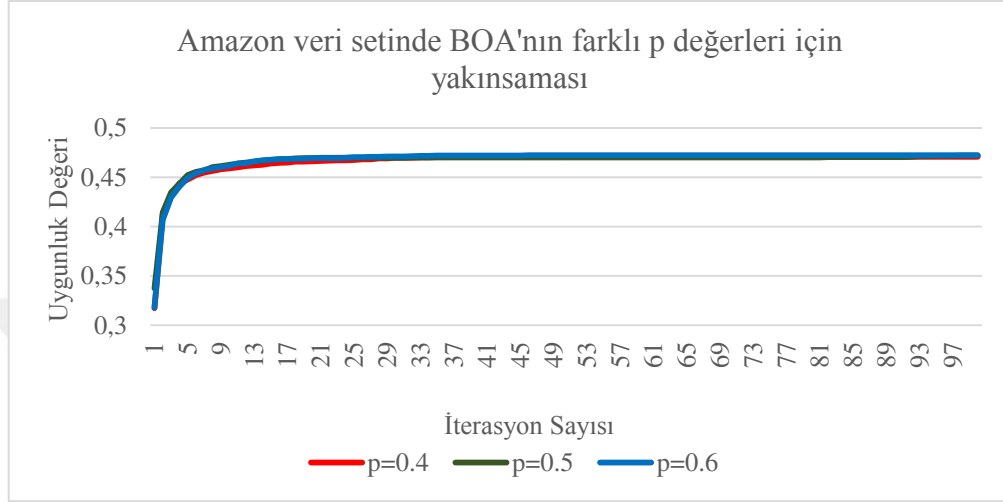
Şekil 7.22. Polarity veri setinde BOA’nın farklı p değerleri için yakınsaması

Şekil 7.23’te Amazon veri setinde SETO’nun uygunluk değerinin farklı κ değerleri için 100 iterasyon boyunca yakınsaması gösterilmektedir. Şekilde görüldüğü gibi $\kappa=1/boyut$ olduğunda uygunluk değeri diğer ikisine göre oldukça iyi sonuçlar vermektedir. $\kappa=0.5/boyut$ ve $\kappa=2/boyut$ olduğunda ise uygunluk değerleri birbirine yakın sonuçlar göstermektedir.



Şekil 7.23. Amazon veri setinde SETO’nun farklı κ değerleri için yakınsaması

Şekil 7.24'te Amazon veri setinde BOA'nın 100 iterasyon boyunca farklı p değerlerine göre uygunluk değerinin yakınsama grafiği verilmiştir. Buna göre 3 farklı p değerinde de birbirine yakın sonuçlar elde edilmiştir.



Şekil 7.24. Amazon veri setinde BOA'nın farklı p değerleri için yakınsaması

8. SONUÇLAR VE TARTIŞMA

Ağ analizi, bir sosyal sistemi incelemek için en iyi yolun sistem üyeleri arasındaki bağları incelemek olduğu varsayımıyla toplumsal yapıyı inceleyen bir yaklaşımdır. Günlük yaşam tarzının önemli bir parçası haline gelen sosyal ağlar son yıllarda gittikçe önem kazanmaktadır. Günümüzde sağlık, eğitim, akademik gibi birçok alanda kullanılmaktadır. Sosyal ağların aktif bir şekilde kullanımı sosyal ağ analizi problemlerini ortaya çıkarmıştır. En bilinen sosyal ağ analizi problemlerinden biri olan duygu analizi – fikir madenciliği insanların, ürünler, hizmetler, kuruluşlar, bireyler, konular, etkinlikler, başlıklar ve nitelikleri gibi varlıklara yönelik fikirlerini, duygularını, değerlendirmelerini, derecelendirmelerini, tutumlarını ve hislerini analiz eden çalışma alanıdır.

Bu çalışmada, duygu analizi ilk defa bir optimizasyon problemi olarak incelenmiştir. BOA ve SETO gibi iki güncel akıllı optimizasyon algoritması uyarlanarak, bu yöntemlerin duygu analizindeki performansları test edilmiştir. Daha iyi sonuçlar elde edebilmek için SETO'ya hafıza özelliği eklenmiştir. Bu problem için uyarlanan akıllı optimizasyon yöntemlerini test etmek amacıyla IMDB, Polarity ve Amazon veri setlerindeki yorumlar kullanılmıştır. İlk olarak veri setlerindeki yorumlar ön işlemlerden geçirilmiştir ve doküman matrisi oluşturulmuştur. Daha sonra BOA ve SETO algoritmalarının bu probleme uyarlanabilmesi amacıyla gerekli temsil biçimleri ve uygunluk fonksiyonları belirlenerek yorumların duygu tahminleri yapılmıştır. İki tane güncel akıllı optimizasyon algoritması ile 33 tane gözetimli makine öğrenmesi algoritmasının duygu analizindeki performans karşılaştırılması ilk kez bu tez çalışmasında gerçekleştirilmiştir.

Duygu analizi problemi çok amaçlı bir problem olarak ele alınmıştır. Doğruluk yüzdesi, hassasiyet, kesinlik, F-Ölçütü ve MCC ölçütlerinin hepsi için iyi sonuçların alınması hedeflenmiştir. Ayrıca uygunluk fonksiyonlarının beş ölçüt haricinde yeni ölçütler eklenebilecek şekilde esnek olmasından dolayı, uyarlanan bu akıllı optimizasyon algoritmaları diğer yöntemlere göre avantajlı görülmektedir. Sonuçlar incelendiğinde, doğruluk yüzdesine bakılarak BOA ve SETO algoritmalarının 33 tane gözetimli öğrenme algoritmalarına oranla doğruluk yüzdesi ölçütüne göre iyi sonuçları verdiği gözlemlenmiştir. Ayrıca diğer ölçütler de göz önüne alındığında, Hassasiyet ve F-Ölçütlerinde en iyi sonuçlar BOA ve SETO'dan elde edilmiştir. Kesinlik ve MCC ölçütlerinde de aynı şekilde başarılı sonuçlar elde edildiği görülebilmektedir.

Son olarak her bir veri seti için 50, 100 ve 200 iterasyon boyunca uygunluk değerlerinin yakınsama grafikleri verilmiştir. Ayrıca BOA'da p , SETO'da ise κ parametrelerinin parametre analizleri üç veri seti için yapılmıştır. Gelecekte, duygu analizi için diğer metasezgisel yöntemler uyarlanarak duygu tahmini yapılabilir. Algoritmaların hibrit versiyonları uygulanarak daha iyi sonuçlar elde edilebilir. Aynı şekilde diğer sosyal ağ problemleri için de metasezgisel yöntemler uyarlanabilir. Çok büyük boyutlu problemler için de MapReduce programlama uygulanarak etkili sonuçlar elde edilebilir. Kullanılan ölçütlere ek olarak yeni ölçütler eklenebilir veya mevcut ölçütler istenirse çıkarılabilir.



KAYNAKLAR

- Ahmed, M. M., Houssein, E. H., Hassanien, A. E., Taha, A. and Hassanien, E.,** 2017. Maximizing Lifetime of Wireless Sensor Networks based on Whale Optimization Algorithm, *International Conference on Advanced Intelligent Systems and Informatics*, Springer, Cham, 724-733.
- Akoglu, L., Chandy, R. and Faloutsos, C.,** 2013. Opinion Fraud Detection in Online Reviews by Network Effects. *ICWSM*, **13**, 2-11.
- Akyol, S.,** 2013. Güncel Sürü Zekâsı Algoritmalarıyla Sınıflandırma Kurallarının Keşfi, Yüksek Lisans Tezi, Tunceli Üniversitesi, Fen Bilimleri Enstitüsü, Tunceli.
- Alataş, B.,** 2007. Kaotik Haritalı Parçacık Sürü Optimizasyon Algoritmaları Geliştirme, Doktora Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ.
- Al-garadi, M. A., Varathan, K. D. and Ravana, S. D.,** 2016. Cybercrime Detection in Online Communications: The Experimental Case of Cyberbullying Detection in the Twitter Network, *Computers in Human Behavior*, **63**, 433-443.
- Aljarah, I., Faris, H. and Mirjalili, S.,** 2018. Optimizing Connection Weights in Neural Networks using The Whale Optimization Algorithm, *Soft Computing*, **22**(1), 1-15.
- Atashpaz-Gargari, E. and Lucas, C.,** 2007. Imperialist Competitive Algorithm: An Algorithm for Optimization Inspired by Imperialistic Competition, *IEEE Congress on Evolutionary Computation*, CEC 2007, 4661-4667.
- Baccianella, S., Esuli, A. and Sebastiani, F.,** 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *LREC*, **10**, 2200-2204.
- Baroni, A. and Ruggieri, S.,** 2015. Segregation Discovery in a Social Network of Companies, International Symposium on Intelligent Data Analysis, *Springer International Publishing*, 37-48.
- Ben Oualid Medani, K., Sayah, S. and Bekrar, A.,** 2018. Whale Optimization Algorithm based Optimal Reactive Power Dispatch: A Case Study of the Algerian Power System, *Electric Power Systems Research*, **163**, 696-705.
- Bhondekar, A.P., Kaur, R., Kumar, R. , Vig, R. and Kapur, P.,** 2011. A Novel Approach using Dynamic Social Impact Theory for optimization of Impedance-Tongue (iTongue), *Chemometrics and Intelligent Laboratory Systems*, **109**, 65–76.

- Birtolo, C., Diessa, V., De Chiara, D. and Ritrovato, P.,** 2013. Customer Churn Detection System: Identifying Customers Who Wish to Leave a Merchant, *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Springer Berlin Heidelberg, 411-420.
- Borji, A.,** 2007. A New Global Optimization Algorithm Inspired by Parliament Political Competitions, *Lecture Notes in Computer Science*, 4827/2007, 61-71.
- Brun, Y., Muslu, K., Holmes, R., Ernst, M. D. and Notkin, D.,** 2012. Predicting Development Trajectories to Prevent Collaboration Conflicts.
- Cambria, E., Olsher, D. and Rajagopal, D.,** 2014. SenticNet 3: A Common and Common-Sense Knowledge base for Cognition-Driven Sentiment Analysis, *Twenty-Eighth AAAI Conference on Artificial Intelligence*.
- Chen, Q., Ye, A. and Xu, L.,** 2016. A Privacy-Preserving Proximity Detection Method in Social Network. *Proceedings of the International Conference on Internet of things and Cloud Computing*, ACM, 68.
- Chu, S. C., Tsai, P. W. and Pan, J. S.,** 2006. Cat Swarm Optimization, *9th Pacific Rim International Conference on Artificial Intelligence*, LNAI, **4099**, 854-858.
- Colantonio, A., Di Pietro, R., Ocello, A. and Verde, N. V.,** 2010. Taming Role Mining Complexity in RBAC, *Computers and Security*, **29**(5), 548-564.
- Cura, T.,** 2008. Modern Sezgisel Teknikler ve Uygulamaları, Papatya Yayıncılık Eğitim.
- Doran, D., Gokhale, S. S. and Dagnino, A.,** 2014. Discovering Perceptions in Online Social Media: A Probabilistic Approach, *International Journal of Software Engineering and Knowledge Engineering*, **24**(09), 1273-1299.
- Dorigo M., Maniezzo, V. and Colorni, A.,** 1991. The Ant System: An Autocatalytic Optimizing Process. *Tech. Rep.* No. 91- 016, Dipartimento di Elettronica, Politecnico di Milano, Italy.
- dos Santos, C. and Gatti, M.,** 2014. Deep Convolutional Neural Networks for sentiment Analysis of Short Texts, *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 69-78.
- Eirinaki, M., Pisal, S. and Singh, J.,** 2012. Feature-based Opinion Mining and Ranking, *Journal of Computer and System Sciences*, **78**(4), 1175-1184.
- El Aziz, M. A., Ewees, A. A. and Hassanien, A. E.,** 2017. Whale Optimization Algorithm and Moth-Flame Optimization for Multilevel Thresholding Image Segmentation, *Expert Systems with Applications*, **83**, 242-256.

- Esuli, A. and Sebastiani, F.,** 2007. SENTIWORDNET: A High-Coverage Lexical Resource for Opinion Mining. *Evaluation*, 1-26.
- Feldman, R.,** 2013. Techniques and Applications for Sentiment Analysis, *Communications of the ACM*, **56**(4), 82-89.
- Geem, Z. W., Kim, J. H. and Loganathan, G. V.,** 2001. A New Heuristic Optimization Algorithm: Harmony Search, *Simulation*, **76**, 60-68.
- Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J. and Reyes, A.,** 2015. Semeval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter, *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 470-478.
- Gilbert, C. H. E.,** 2014. Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*.
- Grefenstette, G., Qu, Y., Shanahan, J. G. and Evans, D. A.,** 2004. Coupling Niche Browsers and Affect Analysis for an Opinion Mining Application, Coupling Approaches, Coupling Media and Coupling Languages for Information Retrieval, *Le Centre De Hautes Etudes Internationales D'informatique Documentaire*, 186-194.
- Güden, H., Vakvak, B., Özkan, B. E., Altıparmak, F. ve Dengiz, B.,** 2005. Genel Amaçlı Arama Algoritmaları ile Benzetim Eniyilemesi: En İyi Kanban Sayısının Bulunması, *Endüstri Mühendisliği Dergisi*, 16, 2-15.
- Haythornthwaite, C.,** 1996. Social Network Analysis: An Approach and technique for The Study of Information Exchange, *Library and information science research*, **18**(4), 323-342.
- Heard, N. A., Weston, D. J., Platanioti, K. and Hand, D. J.,** 2010. Bayesian Anomaly Detection Methods for Social Networks, *The Annals of Applied Statistics*, **4**(2), 645-662.
- Holland, J. H.,** 1975. Adaption in Natural and Artificial Systems, University of Michigan Pres, Ann Arbor, MI.
- Hu, X., Tang, J., Gao, H. and Liu, H.,** 2014. Social Spammer Detection with Sentiment Information, *2014 IEEE International Conference on Data Mining*, IEEE, 180-189.
- Hussain, O. A., Anwar, Z., Saleem S. and Zaidi, F.,** 2013. Empirical Analysis of Seed Selection Criterion in Influence Mining for Different Classes of Networks, *2013 International Conference on Cloud and Green Computing*, Karlsruhe, 2013., 348-353.

- Jin, W., Ho, H. H. and Srihari, R. K.,** 2009. A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining, *Proceedings of the 26th Annual International Conference on Machine Learning*, 465-472.
- Karaboğa, D.,** 2011. Yapay Zeka Optimizasyon Algoritmaları, Nobel Yayın Dağıtım.
- Kashan, A. H.,** 2009. League Championship Algorithm: A New Algorithm For Numerical Function Optimization, *IEEE International Conference of Soft Computing and Pattern Recognition (SoCPAR)*, 43–48.
- Kaur, R., Kumar, R., Gulati, A., Ghanshyam, C., Kapur, P. and Bhondekar, A. P.,** 2012. Enhancing Electronic Nose Performance: A Novel Feature Selection Approach using Dynamic Social Impact Theory and Moving Window Time Slicing for classification of Kangra Orthodox Black Tea (*Camellia sinensis* (L.) O. Kuntze). *Sensors and Actuators B: Chemical*, **166**, 309-319.
- Kaveh, A. and Ghazaan, M. I.,** 2017. Enhanced Whale Optimization Algorithm for Sizing Optimization of Skeletal Structures, *Mechanics Based Design of Structures and Machines*, **45**(3), 345-362.
- Kennedy, J. and Eberhart, R. C.,** 1995. Particle Swarm Optimization, *IEEE International Conference on Neural Networks*, Piscataway, NJ, 1942-1948.
- Khan, A. Z., Atique, M. and Thakare, V. M.,** 2015. Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis, *International Journal of Electronics, Communication and Soft Computing Science and Engineering (IJECSCE)*, **89**.
- Kiritchenko, S., Mohammad, S. and Salameh, M.,** 2016. Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 42-51.
- Kiritchenko, S., Zhu, X. and Mohammad, S. M.** (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, **50**, 723-762.
- Kuhlmann, M., Shohat, D. and Schimpf, G.,** 2003. Role Mining – Revealing Business Roles for Security Administration using Data Mining Technology, *Proceedings of the 8th ACM Symposium on Access Control Models and Technologies, SACMAT '03*, 179–186.
- Kunal, R., Panesar, K. S., Thakur, M. N. and Asgekar, A. S.,** 2012. *U.S. No: 8,311,950. Washington, DC: U.S. Patent and Trademark Office.*
- Latane, B.,** 1981. The Psychology of Social Impact, *American Psychologist*, **36**(4), 343-356.
- Li, J., Ju, L., Green, B. and Backstrom, L.,** 2013. *U.S. Patent Application No:13/961,789.*

- Ling, Y., Zhou, Y. and Luo, Q.,** 2017. Lévy Flight Trajectory-Based Whale Optimization Algorithm for Global Optimization, *IEEE Access*, **5**(99), 6168-6186.
- Liu, B.,** 2012. Sentiment Analysis and Opinion Mining, *Synthesis Lectures on Human Language Technologies*, **5**(1), 1-167.
- Macas, M., and Lhotska, L.,** 2008. Social Impact and Optimization, *International Journal of Computational Intelligence Research*. **4**(2), 129-136.
- Macaš, M., Bhondekar, A. P., Kumar, R., Kaur, R., Kuzilek, J., Gerla, V., and Kapur, P. ,** 2014. Binary Social Impact Theory based Optimization and Its Applications in Pattern Recognition, *Neurocomputing*, **132**, 85-96.
- Macaš, M., Lhotská, L. and Křemen, V.,** 2007. Social Impact based Approach to Feature Subset Selection, in: N. Krasnogor, G. Nicosia, M. Pavone, D. Pelta (Eds.), *Nature Inspired Cooperative Strategies for Optimization (NICSO 2007)*, Springer Berlin, Heidelberg, 2008, 239–248.
- Mafarja, M. M. and Mirjalili, S.,** 2017. Hybrid Whale Optimization Algorithm with Simulated Annealing for Feature Selection, *Neurocomputing*, **260**, 302-312.
- Maniezzo, V., Stützle, T. and VOB, S.,** 2009. Matheuristics: Hybridizing Metaheuristics and Mathematical Programming, **10**, Springer, New York.
- McAuley, J., Pandey, R. and Leskovec, J.** (2015, August). Inferring Networks of Substitutable and Complementary Products, *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, August, 785-794.
- McCulloh, I., A. and Carley, K., M.,** 2008. Social Network Change Detection, Carnegie Mellon University, School of Computer Science, Technical Report, CMU-CS-08-116.
- McGloin, J. M. and Kirk, D. S.,** 2010. Social Network Analysis, *Handbook of Quantitative Criminology*, Springer New York, 209-224.
- Mirjalili, S. and Lewis, A.,** 2016. The Whale Optimization Algorithm, *Advances in Engineering Software*, **95**, 51-67.
- Mittal, P., Garg, S. and Yadav, S.,** 2016. Social Network Analysis using Interest Mining: A Critical Review, *Indian Journal of Science and Technology*, **9**(16).
- Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X. and Cherry, C.,** 2016. Semeval-2016 Task 6: Detecting Stance in Tweets, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 31-41.

- Mostafa, A., Hassanien, A. E., Houseni, M. and Hefny, H.,** 2017. Liver Segmentation in MRI Images based on Whale Optimization Algorithm, *Multimedia Tools and Applications*, **76**(23), 24931-24954.
- Murty, K. G.,** 2003. Optimization Models for Decision Making, Internet Edition, Chapter 1: Models for Decision Making, 1, 1-18.
- Nielsen, F. Å.,** 2011. A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs, *Proc. ESWC-11*.
- Nowak, A., Szamrej, J., and Latané, B.,** 1990. From Private Attitude to Public Opinion: A Dynamic Theory of Social Impact. *Psychological Review*, **97**(3), 362.
- Oliva, D., El Aziz, M. A. and Hassanien, A. E.,** 2017. Parameter Estimation of Photovoltaic Cells using an Improved Chaotic Whale Optimization Algorithm, *Applied Energy*, **200**, 141-154.
- Pak, A. and Paroubek, P.,** 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining, *LREc*, 2010 May, **10**.
- Pallotta, V., Vrieling, L. and Delmonte, R.,** 2013. Interaction Mining: Making Business Sense of Customers Conversations through Semantic and Pragmatic Analysis, In *Supply Chain Management: Concepts, Methodologies, Tools, and Applications*, IGI Global, 688-709.
- Pang, B. and Lee, L.,** 2004. A Sentimental Education: Sentiment Analysis using Subjektivite Summarization based on Minimum Cuts, *Proceedings of the 42nd annual meeting on Association for Computational*, Temmuz.
- Pang, B. and Lee, L.,** 2005. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales, *Proceedings of the 43rd annual meeting on association for computational linguistics*, Temmuz, 115-124.
- Pang, B., Lee, L. and Vaithyanathan, S.,** 2002. Thumbs up?: Sentiment Classification using Machine Learning Techniques, *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)*, **10**, 79-86.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., ... and Hoste, V.,** 2016. SemEval-2016 Task 5: Aspect based Sentiment Analysis, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 19-30.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A. S., ... and Hoste, V.,** 2016. SemEval-2016 Task 5: Aspect based Sentiment Analysis, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 19-30.

- Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S. and Androutsopoulos, I.,** 2015. Semeval-2015 Task 12: Aspect based Sentiment Analysis, *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 486-495.
- Poria, S., Cambria, E. and Gelbukh, A.,** 2015. Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-Level Multimodal Sentiment Analysis, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2539-2544.
- Poria, S., Cambria, E., Howard, N., Huang, G. B. and Hussain, A.,** 2016. Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content, *Neurocomputing*, **174**, 50-59.
- Prakash, D. B. and Lakshminarayana, C.,** 2017. Optimal Siting of Capacitors in Radial Distribution Network using Whale Optimization Algorithm, *Alexandria Engineering Journal*, **56**(4), 499-509.
- Rao, Y., Li, Q., Wenyin, L., Wu, Q. and Quan, X.,** 2014. Affective Topic Model for Social Emotion Detection, *Neural Networks*, **58**, 29-37.
- Rashedi, E., Nezamabadi-pour, H. and Saryazdi, S.,** 2009. GSA: A Gravitational Search Algorithm, *IEEE Congress on Information Sciences*, **179**, 2232-2248.
- Rosenthal, S., Farra, N. and Nakov, P.,** 2017. SemEval-2017 Task 4: Sentiment Analysis in Twitter, *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 502-518.
- Salem, S. A.,** 2012. BOA: A Novel Optimization Algorithm, *IEEE 2012 International Conference on Engineering and Technology (ICET)*, **1**(5).
- Scellato, S., Musolesi, M., Mascolo, C., Latora, V. and Campbell, A. T.,** 2011. NextPlace: A Spatio-Temporal Prediction Framework for Pervasive Systems, *International Conference on Pervasive Computing*, Springer Berlin Heidelberg, 152-169.
- Scott, J.,** 1991. Social Network Analysis: A Handbook. London: Sage.
- Severyn, A. and Moschitti, A.,** 2015. Twitter Sentiment Analysis with Deep Convolutional Neural Networks. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, August, 959-962, ACM.
- Shah-Hosseini, H.,** 2009. The Intelligent Water Drops Algorithm: A Nature-Inspired Swarm-Based Optimization Algorithm, *International Journal of Bio-Inspired Computation*, **1**, 71-79.
- Sigg, S., Gordon, D., von Zengen, G., Beigl, M., Haseloff, S. and David, K.,** 2012. Investigation of Context Prediction Accuracy for Different Context

Abstraction Levels, *IEEE Transactions on Mobile Computing*, **11**(6), 1047-1059.

- Skopik, F., Truong, H. L. and Dustdar, S.,** 2009. Trust and Reputation Mining in Professional Virtual Communities, *International Conference on Web Engineering*, Springer Berlin Heidelberg, 76-90. doi: 10.1007/978-3-642-02818-2_6.
- Nandhini, B. S. and Sheeba, J.I.,** 2015. Online Social Network Bullying Detection Using Intelligence Techniques, *Procedia Computer Science*, **45**, 485-492, ISSN 1877-0509, <http://dx.doi.org/10.1016/j.procs.2015.03.085>.
- Storn, R. and Price, K.,** 1995. Differential Evolution: A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces, *Technical Report TR-95-012*, International Computer Science Institute, Berkeley.
- Strapparava, C. and Mihalcea, R.,** (2007, June). Semeval-2007 task 14: Affective Text, *Proceedings of the 4th International Workshop on Semantic Evaluations*, June, 70-74, Association for Computational Linguistics.
- Touma, H. J.,** 2016. Study of The Economic Dispatch Problem on IEEE 30-Bus System using Whale Optimization Algorithm, *International Journal of Engineering Technology and Sciences (IJETS)*, **5**(1), 11-18.
- Trivedi, I. N., Pradeep, J., Narottam, J., Arvind, K. and Dilip, L.,** 2016. Novel Adaptive Whale Optimization Algorithm for Global Optimization, *Indian Journal of Science and Technology*, **9**(38).
- Turney, Peter D.,** 2002 .Thumbs up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews, *Proceedings of Annual Meeting of the Association for Computational Linguistics (AC L-2002)*.
- Wang, J., Du, P., Niu, T. and Yang, W.,** 2017. A Novel Hybrid System based on A New Proposed Algorithm—Multi-Objective Whale Optimization Algorithm for Wind Speed Forecasting, *Applied Energy*, **208**, 344-360.
- Wang, X., Jin, X., Chen, M. E., Zhang K. and Shen, D.,** 2012. Topic Mining over Asynchronous Text Sequences, *IEEE Transactions on Knowledge and Data Engineering*, **24**(1), Jan. 2012, 156-169. doi: 10.1109/TKDE.2010.229
- Wang, X., Zhu, F., Jiang, J. and Li, S.,** 2013. Real Time Event Detection in Twitter, *International Conference on Web-Age Information Management*, Springer Berlin Heidelberg, 502-513.
- Wasserman, S. and Faust, K.,** 1994. Social Network Analysis, Cambridge, MA: Cambridge University Press.
- Wellman, B. and Berkowitz, S., D.,** 1988. Social Structures: A Network Approach. Cambridge: Cambridge University Press.

Wellman, B., 1983. Network Analysis: Some Basic Principles, *Sociol Theory*, **1**, 155–200.

Yang, H., Zhong, J., Ha, D. and Oh, H., 2016. Rumor Propagation Detection System in Social Network Services. *International Conference on Computational Social Networks*, Springer International Publishing, . 86-98.

Zhang, W., Sun, G. and Bin, S., 2016. An Opinion Leaders Detecting Algorithm in Multi-relationship Online Social Networks. *International Journal of Hybrid Information Technology*, **9**(5), 391-398.

URL-1, <https://www.knime.org/> Open for Innovation KNIME. 9 Kasım 2015.

URL-2, <http://www.cs.cornell.edu/people/pabo/movie-review-data> Movie Review Data, 15 Ocak 2016.

URL-3, <https://github.com/openai/generating-reviews-discovering-sentiment>, 21 Şubat 2017.

URL-4, <http://www.cs.waikato.ac.nz/ml/weka/> WEKA Veri Madenciliği Yazılımı, 10 Haziran 2017.

ÖZGEÇMİŞ

Sinem AKYOL

Munzur Üniversitesi
Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü

E-Posta: sakyol@munzur.edu.tr

- 1987** Tunceli’de doğdu
- 2001-2004** Tunceli Anadolu Lisesini tamamladı
- 2005-2009** Ege Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümünden mezun oldu
- 2010-2013** Munzur Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği A.B.D’de Yüksek Lisans eğitimini tamamladı
- 2013-** Fırat Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Yazılım A.B.D’de doktora eğitimine başladı ve halen devam etmektedir.
- 2010-** Munzur Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümünde Araştırma Görevlisi olarak çalışmaktadır.