



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

**GRİ KURT OPTİMİZASYON ALGORİTMASI İLE ÇIKARIMSAL
METİN ÖZETLEME VE ÖZETLERİN DERİN ÖĞRENME İLE
SINIFLANDIRILMASI**

EBRU DUDAK

**YÜKSEK LİSANS TEZİ
ELEKTRİK-ELEKTRONİK VE BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI**

**DANIŞMAN
PROF. DR. PAKİZE ERDOĞMUŞ**

DÜZCE, 2020

T.C.
DÜZCE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GRİ KURT OPTİMİZASYON ALGORİTMASI İLE ÇIKARIMSAL
METİN ÖZETLEME VE ÖZETLERİN DERİN ÖĞRENME İLE
SINIFLANDIRILMASI

Ebru DUDAK tarafından hazırlanan tez çalışması aşağıdaki jüri tarafından Düzce Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Prof. Dr. Pakize ERDOĞMUŞ

Düzce Üniversitesi

Jüri Üyeleri

Prof. Dr. Pakize ERDOĞMUŞ

Düzce Üniversitesi

Dr. Öğr. Üyesi Abdullah Talha KABAKUŞ

Düzce Üniversitesi

Dr. Öğr. Üyesi Şafak KAYIKÇI

Bolu Abant İzzet Baysal Üniversitesi

Tez Savunma Tarihi: 07/07/2020

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

07 Temmuz 2020

Ebru DUDAK

TEŐEKKÜR

Yüksek lisans öğrenimimde ve bu tezin hazırlanmasında gösterdiği her türlü destek ve yardımdan dolayı çok değerli hocam Prof. Dr. Pakize ERDOĞMUŐ'a en içten dileklerle teşekkür ederim.

Bu çalışma boyunca yardımlarını ve desteklerini esirgemeyen sevgili aileme ve çalışma arkadaşlarıma sonsuz teşekkürlerimi sunarım.

07 Temmuz 2020

Ebru Dudak



İÇİNDEKİLER

Sayfa No

ŞEKİL LİSTESİ	vii
ÇİZELGE LİSTESİ	viii
KISALTMALAR.....	ix
SİMGELER.....	x
ÖZET	xi
ABSTRACT	xii
1. GİRİŞ.....	1
1.1. METİN ÖZETLEME	2
1.1.1. Metin Özetleme Nedir?	2
1.1.2. Metin Özetleme Çeşitleri.....	2
1.2. LİTERATÜR ÖZETİ	3
2. MATERYAL VE YÖNTEM	9
2.1. KULLANILAN ALGORİTMALAR VE ÇALIŞMA ORTAMI	9
2.1.1. Gri Kurt Optimizasyon Algoritması.....	9
2.1.1.1. <i>Avı Çevreleme (Encircling prey)</i>	11
2.1.1.2. <i>Avlama (Hunting)</i>	12
2.1.1.3. <i>Ava Saldırma (Attacking Prey)</i>	13
2.1.1.4. <i>Arama (Search for prey)</i>	13
2.1.1.5. <i>GWO Uygulamaları</i>	14
2.1.2. K-means.....	15
2.1.3. Veri Seti	16
2.2. BAŞARI DEĞERLENDİRME ÖLÇÜTÜ	16
2.2.1. Ngram birliktelik istatistiği (ROUGE)	17
3. GRİ KURT OPTİMİZASYON ALGORİTMASI İLE ÇIKARIMSAL METİN ÖZETLEME	19
3.1. SİSTEMİN TASARLANMASI	19
3.2. ÖNİŞLEME ADIMLARI	19
3.2.1. Büyük/ Küçük Harf Dönüşümü	20
3.2.2. Durma Kelimelerinin Kaldırılması.....	20
3.2.3. Belgelerin Cümlelere Bölünmesi.....	20
3.2.4. Noktalama İşaretlerinin Kaldırılması	20
3.2.5. Metni Kelimelere Bölme İşlemi.....	20
3.3. CÜMLELERİN PUANLANDIRILMASI.....	20
3.3.1. Cümle Uzunluğu	21
3.3.2. Cümle Konumu	21
3.3.3. Başğa Olan Eşlik	21
3.3.4. Terim Sıklığı	22
3.3.5. Cümlelerin Toplam Puanı	22
3.4. CÜMLELERE KOSİNÜS BENZERLİĞİNİN UYGULANMASI.....	22
3.5. GWO İLE SİSTEMİN ÇALIŞTIRILMASI.....	23

3.6. K-MEANS İLE SİSTEMİN ÇALIŞTIRILMASI	24
3.7. ÖZETİN DEĞERLENDİRİLMESİ.....	25
4. DERİN ÖĞRENME.....	26
4.1. DOĞAL DİL İŞLEME ALANINDA DERİN ÖĞRENME UYGULAMALARI.....	27
4.2. DERİN ÖĞRENME MİMARİLERİ.....	28
4.2.1. Derin Sinir Ağları.....	28
4.2.2. Derin Oto Kodlayıcılar.....	28
4.2.3. Derin İnanç Ağları.....	29
4.2.4. Derin Boltzmann Makinesi	29
4.2.5. Evrişimsel Sinir Ağları (CNN)	29
4.2.6. Yinelenen Sinir Ağları (RNN).....	29
4.2.6.1. LSTM Ağları	30
5. UZUN KISA SÜRELİ BELLEK AĞLARI İLE ÖZET METNİN SINIFLANDIRILMASI.....	32
5.1. ÖNİŞLEME ADIMLARI	32
5.2. MODELİN TASARLANMASI VE EĞİTİLMESİ.....	32
5.2.1. Metinleri Dizilere Dönüştürme	33
5.2.2. LSTM Ağının Oluşturması ve Eğitimi.....	34
6. SONUÇLAR VE ÖNERİLER.....	38
7. KAYNAKLAR	40
ÖZGEÇMİŞ.....	46

ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 2.1. GWO algoritması hiyerarşi zinciri.	9
Şekil 2.2. Gri kurtların avlanma davranışları: (a) avı kovalamak, yaklaşmak ve avı izlemek (b-d) takip etmek, rahatsız etmek ve kuşatmak	11
Şekil 2.3. Gri kurtların 2D ve 3D uzayındaki olası konumları [43].	12
Şekil 2.4. Ava saldırının aranması [43].	13
Şekil 2.5. GWO algoritmasının sözde kodu.	14
Şekil 2.6. Veri seti örneği.	16
Şekil 3.1. Sistemin akış şeması.	19
Şekil 3.2. GWO ile oluşturulan özet örneği.	24
Şekil 3.3. K-means ile oluşturulan özet örneği	25
Şekil 4.1. Derin sinir ağı mimarisinin genel görünümü.	26
Şekil 4.2. Standart bir RNN'de katmanlar.	30
Şekil 4.3. LSTM katmanları.	30
Şekil 5.1. BBC News kategorileri ve dağılımları.	33
Şekil 5.2. BBC News kelime uzunluğu histogramı.	34
Şekil 5.3. Özetsiz veri setine LSTM uygulanması.	36
Şekil 5.4. GWO özetlerine LSTM uygulanması.	36
Şekil 5.5. K-means özetlerine LSTM uygulanması.	37

ÇİZELGE LİSTESİ

	<u>Sayfa No</u>
Çizelge 3.1. Cümle final fkoru örneği.	22
Çizelge 3.2. Kosinüs benzerlik matrisi örneği.	23
Çizelge 3.3. GWO ve K-means'in özetleme başarıları.	25
Çizelge 5.1. LSTM ağında kullanılan parametreler.	35



KISALTMALAR

CNN	Evrişimsel sinir ağı
DNN	Derin sinir ağı
ELD	Elektronik yük sevkiyatı
GWO	Gri kurt optimizasyon algoritması
LSTM	Long-short term memory
NLP	Doğal dil işleme
RNN	Yinelemeli sinir ağı
ROUGE	Recall-oriented understudy for gisting evaluation
RTS	Uyumlu metin özeti



SİMGELER

α
 β
 ω
 δ

Alfa
Beta
Omega
Delta



ÖZET

GRİ KURT OPTİMİZASYON ALGORİTMASI İLE ÇIKARIMSAL METİN ÖZETLEME VE ÖZETLERİN DERİN ÖĞRENME İLE SINIFLANDIRILMASI

Ebru DUDAK

Düzce Üniversitesi

Fen Bilimleri Enstitüsü, Elektrik-Elektronik ve Bilgisayar Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Prof. Dr. Pakize ERDOĞMUŞ

Temmuz 2020, 45 sayfa

Günümüzde özellikle internet ortamında metinsel veri hızla çoğalmaktadır. Büyük bir veri seti haline gelen bu metinlerde istenen parçaya ulaşmak gitgide zorlamıştır. Otomatik metin özetleme teknikleri, bu büyük verilerden ilgili bilgilerin çıkarılmasında büyük rol oynamaktadır. Bu tez kapsamında çıkarımsal metin özetleme tekniği olarak, sezgisel Gri Kurt Optimizasyon (GWO) algoritması önerilmiştir. GWO algoritmasının kümeleme yeteneği tekli metin özetleme uygulamasıyla test edilmiştir. Özetleme sisteminde cümle sıralaması, kelime uzunluğu gibi istatistiksel anahtar kelime çıkarma yöntemleri ile GWO'nun küme çıkarımı birleştirilmiştir. Özet sistemi K-means kümeleme algoritması ile de denenmiş ve sonuçlar ROUGE değerlendirme metriği ile ölçülmüştür. 2225 haber makalesinden oluşan BBC News veri setiyle test edilen çalışmada elde edilen sonuçlar doğrultusunda GWO algoritmasının oldukça iyi performans gösterdiğini görülmüştür. Elde edilen özet metinler derin öğrenme yöntemlerinden biri olan Uzun kısa süreli bellek (LSTM) ağları ile sınıflandırılmıştır. Deneysel sonuçlar LSTM ağının sınıflandırma başarısının GWO algoritması ile oluşturulan özetlerde arttığını göstermiştir. GWO algoritması çıkarımsal özetleme yöntemleri için yeni bir yaklaşım olarak sunulmuştur.

Anahtar sözcükler: Gri kurt optimizasyon algoritması, Çıkarımsal metin özetleme, Uzun kısa süreli bellek, ROUGE metriği, Derin öğrenme.

ABSTRACT

EXTRACTİVE TEXT SUMMARİZATION BY GRAY WOLF OPTİMİZATION ALGORITHM AND CLASSİFICATION OF ABSTRACTS WITH DEEP LEARNING

Ebru DUDAK

Duzce University

Graduate School of Natural and Applied Sciences, Department of Electrical-Electronics
and Computer Engineering

Master's Thesis

Supervisor: Prof. Dr. Pakize ERDOĞMUŞ

July 2020, 45 pages

Today, especially in the internet environment, textual data is increasing rapidly. It is getting more and more difficult to reach the desired piece in these texts, which have become a large data set. Automated text summarization techniques play a major role in extracting relevant information from these big data. In this thesis, the intuitive Gray Wolf Optimization (GWO) algorithm is proposed as inferential text summarization technique. Clustering ability GWO algorithm has been tested with single text summarization application. In the summarization system, statistical keyword extraction methods such as sentence order, word length and cluster extraction of GWO are combined. The summary system was also tested with the K-means clustering algorithm and the results were measured with the ROUGE evaluation metric. In accordance with the results obtained in the study tested with BBC News data set consisting of 2225 news articles, GWO algorithm was observed to perform quite well. The summarized texts are classified with one of the deep learning methods, Long- Short term memory (LSTM) networks. Experimental results showed that the classification success of the LSTM network increased in the summaries created with the GWO algorithm. GWO algorithm is presented as a new approach for inferential summarization methods.

Keywords: Gray wolf optimization algorithm, Extractive text summarization, Long-Short term memory, ROUGE metric, Deep learning.

1. GİRİŞ

Metin, video, resim gibi farklı formatlarda bilgi boyutu günümüzde hızla artmaktadır. Özellikle mobil cihazlar, internet sayfaları, kişisel bloglar, sosyal medya uygulamaları vasıtası ile üretilen veriler yüksek boyutlara ulaşmıştır. Dünya çapında binlerce exabayt büyüklüğünde olan ve her geçen gün katlanarak artan bu verileri yönetmek ve depolamak, işlemek, bilimsel veya ticari kullanım amacıyla analiz etmek ve sonuçlarını açıklayabilmek oldukça zordur ve büyük bir zaman gerektirmektedir.

Elektronik ortamda biriken bu veriler metin, ses, görüntü gibi türlerde olabilmektedir. Bu türlerden biri olan metin verileri, günümüzde en sık kullanılan ve erişim sağlanan kısımdır. Çağımızda insanlar ulaşmak istedikleri metinsel verilere en yalın ve doğru haliyle en kısa zamanda ulaşmak istemektedir. Ulaşılmak istenen metin verisinde temel nokta gereksiz verilerden arındırılmış olmasıdır. Okunmak istenen haberin temel noktalarına hemen ulaşılması, bir mailin sadeleştirilmesi, akademik bir makalenin fazla zaman kaybetmeden incelenip, gözden geçirilmesi gibi birçok alanda büyük verinin sadeleştirilmesine ihtiyaç duyulmaktadır. Veri yığınındaki istenen bilgiye ulaşma problemi günümüzde bilgi erişimi isimli araştırma konusu olarak karşımıza çıkmaktadır.

Bilgi erişimi kullanıcının talep olarak ifade ettiği ve kullanılan sisteme, ihtiyaca yönelik bilginin saptaması ve bu bilgilere erişimi ile ilgilenmektedir. Bir bilgi erişimi sistemi, belgeleri ve içindeki bilgileri temsil etmek için asıl belgenin yerine geçen bir örnek içinde barındırır [1].

Bilgi erişimi araştırma konusunun alt bir dalı olarak metin özetleme sistemleri yukarıda anlatmış olan veri boyutunu küçültme amacıyla yaygın olarak uzun yıllardır kullanılmaktadır. Metin özetleme günümüzde de hala önemli kabul edilen bir araştırma dalıdır.

Sunulmakta olan bu çalışma genel olarak iki aşamadan oluşmaktadır. İlk aşamada sezgisel bir algoritma olan Gri Kurt Optimizasyon (GWO) algoritmasının metin özetleme başarısı ele alınmıştır. İkinci aşamada ise oluşturulan özet metinler, derin öğrenmenin bir kolu olan Uzun kısa süreli bellek (LSTM) kullanılarak sınıflandırılmıştır. Temel olarak GWO'nun sınıflandırma başarımına etkisini de test

etmek amaçlanmıştır.

1.1. METİN ÖZETLEME

1.1.1. Metin Özetleme Nedir?

Metin özetleme, metin belgesinin kısa, öz ve metnin ana fikrini temsil eden sıkıştırılmış bir versiyonunu otomatik olarak oluşturma işlemidir [2]. Metin özetlemeye duyulan ihtiyaç birçok alandaki metinsel bilgi kaynaklarındaki hızlı ve sürekli artışla birlikte artmaktadır. Otomatik metin özetleme, belge indeksleme, soru cevaplama sistemleri, yardım alma sistemleri ve belge sınıflandırma gibi çeşitli uygulamalar için yararlı olabilmektedir.

Girdi olarak sisteme bir belge verilir, sistem boyutundan daha az olacak şekilde o belgeyi temsil edebilecek en önemli ve gerekli cümleleri içeren bir belge ortaya çıkarmak metin özetlemenin temel işlevidir. Özetlemeyi ifade eden temel kelimeler "önem" veya "ihtiyaç"dır. Bir kelimenin neden gerekli veya önemli olduğuna verilecek karar özetin başarısını ortaya koyan temel unsurdur. Metin içinde yer alan bir cümle ifade edilen konu ile ne kadar ilişkili ise o kadar önemlidir ve kullanıcının ulaşmak istediği, ihtiyaç duyduğu ham metindir.

Kısa metin özetleme "metnin bütünü ifade eden en uygun parçasını oluşturmak" olarak ifade edilebilir. Genellikle bir sıkıştırma-küçültme işlemi olarak görülen metin özetleme aynı zamanda metni kavrama işlemidir. Bir metnin özetinin çıkarılabilmesi için o metnin ilk olarak kavranması gerekmektedir. Metin özetleme bu açıdan hem kavrama hem sıkıştırma işlemlerinin birlikte gerçekleştirilmesidir. Doğal Dil İşleme (NLP) ve bunun bir kolu olan metin özetlemenin amacı, insanların yaptığı gibi anlayabilen sistemler oluşturmaktır.

1.1.2. Metin Özetleme Çeşitleri

Metin özetleme çeşitli kıstaslara göre türlere ayrılmaktadır. Bu kıstaslar sistemin oluşturduğu çıktı, sistemin girdisi olan metnin kaynağı, çıktının stili ve etki alanıdır.

Metin özetleme sistemleri sistemin oluşturduğu sonuç olan çıktıya türüne göre çıkarımsal (extractive) ve yoruma dayalı (abstractive) olmak üzere ikiye ayrılmaktadır. Çıkarımsal metin özetleme yönteminde özeti çıkarılacak metnin içinden o metni en iyi temsil ettiği düşünülen cümleler seçilip özet metin oluşturulur. Metin üzerinde herhangi

bir deęişim ya da yorumla yapılmaz. Bu özetleme türünde metindeki önemli cümleler, istatistiksel yöntemlerle, sezgisel çıkarımlarla ya da bu yöntemlerin birlikte kullanılmasıyla gerçekleştirilmektedir [3]. Yoruma dayalı metin özetlemede ise özetlenecek metin yorumlanarak tekrar oluşturulur. Bu yüzden metni ayrıştırmak ve özet metni oluşturmak için gramerler ve dil bilgileri de dahil olmak üzere doğal dil işleme yöntemleri kullanan sistemlere ihtiyaç duyulmaktadır.

Metin özetleri sisteme giren metin sayısına göre tekli metin özetleme (single document summarization) veya çoklu metin özetleme (multiple document summarization) olarak ayrılmaktadır. Tekli metin özetlemede özeti çıkarılacak tek bir metin varken, çoklu metin özetlemede ise birbirleri ile ilgili olan birden fazla metin kaynağı kullanılmaktadır. Haberler gibi birçok web tabanlı sistemlerde çoklu metin özetinden yararlanılmıştır. Fakat çoklu belge özetleme nihai özetin kendi içinde tutarlı ve eksiksiz olmasını amaçladığından özeti ilişkilendirme yönünden tekli metin özetlemeye göre daha zordur [4].

Bir özet çıktının stiline göre, gösterici (indicative) veya bilgi verici (informative) olabilir. Gösterici özetleme, metnin ana fikrini kullanıcıya verir. Bu özetlemenin uzunluğu, verilen metnin yaklaşık %5'i kadardır ve genel başlıklar belirlenir. Bilgi verici özette ise sistem ana metin hakkında kısa bilgi verir. Bilgilendirici özetin uzunluğu verilen metnin yaklaşık %20'sidir [5].

Özet çıkarma yöntemlerinde bir diğer ayrım genel özetler ve sorgu tabanlı özetler olarak yapılmaktadır. Sorguya dayalı özetlerde özeti çıkarılacak metinler sorguya bağlı içeriklerle oluşturulurken genel amaçlı özetler belgenin tüm içeriği hakkında genel bilgi vermektedir.

Bu tez kapsamında tek kaynaklı (tekli metin özetleme), çıkarıma dayalı (extractive), bilgi verici (informative) ve genel özetlerin çıkarıldığı bir sistem üzerinde çalışılmıştır.

1.2. LİTERATÜR ÖZETİ

Metin özetleme uzun yıllardır araştırmacılar tarafından ele alınan ve günümüzde hala daha geçerliliğini sürdüren bir araştırma dalıdır. Gelişen teknoloji ve dil bilimsel çalışmaların neticesinde popüler olmaya devam etmektedir. Bu bölümde metin özetleme alanında yapılmış olan akademik çalışmalar incelemiştir.

Metin özetlemeyle ilgili en eski çalışmalar cümle puanlama tabanlıdır ve istatistiksel öğrenme yaklaşımları da dahil olmak üzere çeşitli yaklaşımlar araştırmacılar tarafından denenmiştir. Çıkarımsal özetlemenin tarihi yoruma göre özetlemeye oranla daha eskidir. Sezgisel algoritmaların ve sınıflandırma tekniklerinin özetleme başarısı birçok çalışmada test edilmiştir. İfadelerin veya cümlelerin puanlanması çıkarımsal özetlemede kullanılan en yaygın yöntemdir.

Cümle puanlaması bugün uygulanan yöntemlerin çoğunda benimsenmiştir. Puanlama yöntemleri kelime puanlama, cümle puanlama olarak sınıflandırılır [6]. Kelime puanlama yöntemlerinde, metinde bir kelimenin sıklığını içeren cümlelerin önemi dikkate alınarak, belirleyici olarak kabul edilen özel isimler, yerler ve nesnelere gibi kelimeler daha yüksek puanlanır [7].

Kelime sıklığının cümle seçimine etkisini ilk olarak 1958 yılında Luhn incelemiştir [8]. Bu çalışmada cümleler terim frekanslarına göre puanlandırılmıştır. Luhn çalışmasında terimlerin yüksek sıklıktaki değerleri için bir kesme değeri belirlemiş ve bu değerin üzerinde olan terimlerin alınmamasını önermiştir. Çalışmada aynı zamanda bir alt kesme değeri de belirlenmiş ve bu değerin altındaki terimler de dikkate alınmamıştır.

Edmunson [9] ise Luhn'un çalışmasında bulunan kelime sıklığı değerlendirme kriterine ek olarak "sözcük öbekleri", "başlık" ve "cümle konumu" gibi üç yeni özelliğe yer vermiştir. Metinlerin karakteristik özellikleri göz önüne alındığında başlık ve alt başlıkların metnin konusu hakkında bilgi verdiği sonucu ortaya çıkmıştır. Buna dayanarak metinde geçen kelimelerin metnin başlık ya da alt başlıklarında geçen kelimelerden herhangi biri olup olmadığına bakılır. Eğer cümleler bu kelimeleri barındırıyorsa bu cümlelerin puanları daha fazla olacaktır. Bu yöntem Başlık Terimleri Yöntemi denir ve Edmundson çalışmalarında bu yöntemden yararlanmıştır. Edmundson'un bir başka görüşü de metindeki başlığın ve cümle konumunun önemidir. Bir metinde en önemli cümleler giriş ve sonuç cümleleridir.

Pollack ve Zamora [10] terim sıklığına ek olarak çalışmalarında ipucu söz öbeklerine yer vermişlerdir. Cümle seçiminde söz öbeklerinin önemine değinmişlerdir. Aynı şekilde Brandow'un [11] yaptığı çalışmada ise ana metinde yer alan ilk cümle özete dahil edilmiştir. Kupiec ve arkadaşları [12] büyük harf özelliği, paragraf özelliği gibi metin özetinde puanlama için yeni kriterler kullanmıştır. Ayrıca özetlemede Bayes sınıflandırıcısı kullanmışlardır.

Baldwin ve arkadaşları [13] sorgu tabanlı bir özetleme sistemi kurmuştur. Sorguda bulunan söz öbeklerini referans eden kısaltmalarla söz öbekleri eşleştirilmiştir. Özetleme işlemi skor değerlerine göre yapılmıştır. Kan ve McKeow [14] bilgi çıkarımı ve cümle çıkarımı tekniklerini birlikte kullanmıştır. Radev ve arkadaşları [15] birbirleri ile çok benzeyen cümleleri dikkate almamışlardır. Ayrıca ilk cümle ile olan ortak kelime sayısına algoritmalarında yer vermişlerdir.

Gupta [16] tarafından yapılan çalışmada metin puanlama yönteminde kelimelerin biçimsel özellikleri (kalın, eğik, altı çizili) dikkate alınmıştır. Kalın, eğik ya da altı çizili cümlelerin metinlerde önemli noktaları belirttiği ve özetle yer alma olasılığının daha yüksek olduğu ortaya çıkmıştır. Ayrıca metindeki “Sonunda” ve “Sonuç olarak” gibi ifadelerle başlayan cümleler işaret cümleleri olarak tanımlanmakta ve bu ifadeleri izleyen cümlelerin önemli cümleler olduğu gözlemlenmiştir. Benzer şekilde değerlendirme, özetlenecek metnin başlığıyla da ilişkilendirilmiştir.

Özet çıkarılacak olan metinde bulunan tarih, sayı, nümerik karakter, önemli kelime grupları, kelime ve cümle konumu, cümle uzunluğu, başlık bilgisi, pozitif ve negatif anlam taşıyan kelimeler, merkezlik bir cümlenin önemini tespit etmede kullanılan özelliklerdir. Bahsedilen bu yöntemler dilin yapısal özelliklerini kullanan yöntemlerdir. Çoğu İngilizce metinler üzerine olan bu çalışmalar incelendiğinde, bu yöntemlerin birleştirilerek kullanıldığı görülmektedir.

Nandhini ve arkadaşları [17], hibrit bir yaklaşımı okuma güçlüğü çeken öğrencilere yardımcı olma amacıyla kullanmıştır. Özetleme yöntemiyle okunabilirliği artırarak okuma güçlüğü çeken öğrencilere fayda sağlamayı amaçlamıştır.

Grafik tabanlı temsiller, çok etkili çözümler sundukları için metin analizi yöntemlerinde sıklıkla kullanılmaktadır. Mihalcea ve arkadaşları [18], metin içeriğinin keşişimlerini kullanarak özetleme için grafik tabanlı gösterim içeren TextRank'ı önermişlerdir. Hem TextRank hem de LexRank algoritmalarında, kelime ve cümleler arasındaki karşılıklı bilgileri kullanarak bir belgedeki merkezi cümleleri elde etmek için sunulan bir belge özetleme çerçevesi olan PageRank [19] algoritmasından esinlenilmiştir.

Literatürde metin özetleme işlemini makine öğrenmesi tekniklerini kullanılarak gerçekleştiren çalışmalar yer almaktadır. Copeck ve arkadaşları [20] çalışmalarında temel görevlerini yerine getirmek için alternatif modüller kullanan bir özetleyici

sunmuştur. Hangi yapılandırmanın en iyi performansı gösterdiğini görmek için makine öğrenme tekniklerinden yararlanmışlardır. Otomatik olarak oluşturulan bir özetin kalitesini, kaliteli olduğu bilinen bir model özetinde içerik ifadelerinin kapsamının bir ölçüsünü hesaplayarak belirlemişlerdir. Wong ve arkadaşları [21] denetimli ve yarı denetimli makine öğrenme tekniği kullanarak çıkarımsal metin özetleme üzerinde çalışmıştır. Lal ve Reuger [22], tekli metin özetleme için Bayes sınıflandırıcısını kullanmışlardır.

2002 yılında Hirao ve arkadaşları [23] destek vektör makinelerini kullanarak cümleleri özetle ilgili ya da özetle ilgisi olmayan olarak ayırmıştır. Ayrıca önemli fiilleri içermesi gibi özellikleri de dikkate almıştır. Karamuftuoğlu [24] da aynı şekilde çalışmasında destek vektör makinelerini kullanmıştır. İki cümle arasındaki ortak kelimelerin sözlüksel bağına değinmiştir.

Silla ve arkadaşları [25] özet çıkarmada makine öğrenmesi tekniği kullanan bir diğer çalışmayı ortaya koymuştur. Cümle önemini belirleyen yedi özellik kullanmış ve Naive Bayes ve C4.5 karar ağaçlarını kullanmışlardır. 2017 yılında Luis [26] Uyumlu Metin Özeti (RTS) kavramını ortaya atmıştır. RTS TextRank algoritmasını uygulamaktadır ve arama aygıtının boyutuna göre koşullandırılmış bir metin özetleme tekniği sunmuştur.

2018 yılında Mozghan ve arkadaşları [27] çalışmalarında grafik tabanlı kümeleme yaklaşımı ile biyomedikal metinlerin özetlemesi yaklaşımını ortaya atmıştır. Özetleyici sistem kaynak belgenin kavram tabanlı bir modelini oluşturmak ve belgeyi kavramlarla eşleştirmek için Birleşik Tıp Dili Sisteminden yararlanmıştır.

Hernandez ve Ledeneva [28] cümleleri K-means ile kümelemiş ve özetin çıkarılması için bu kümeler arasından en önemli cümleler seçilmiştir. Çalışmalarında terim sıklığı özelliği kullanılmıştır ve cümleler benzer gruplar altında toplanmıştır. Kümeler oluşturulduktan sonra grup içindeki en iyi cümle belirlenmiştir. Makine öğrenmesi ile yapılan çalışmalar genellikle, bayes sınıflandırıcı, destek vektör makineleri, yapay sinir ağları gibi tekniklerle birlikte kullanılmıştır.

2019 yılında Oussama ve arkadaşları [29] kümeleme ve terim sıklığı madenciliğine dayalı yeni bir biyomedikal metin özeti yöntemi önermiştir. Kümelemenin terim sıklığının keşfinde büyük öneme sahip olduğu görülmüştür. Kümeleme işlemi için K-means kümeleme algoritması kullanmış ve diğer kümeleme algoritmalarıyla karşılaştırmasını yapmıştır.

Özetleme sistemlerinde algoritmalara da sıklıkla yer verilmiştir. Filatova ve arkadaşı [30] özet çıkarmak için Greedy algoritmasını kullanmıştır. Ayrıca olay tabanlı özetleme sistemi tasarlamıştır. Yeh ve arkadaşları [31] cümle konumu, anahtar kelimeler, merkezîyet gibi özelliklerin genetik algoritmayla birleştirildiği hibrit bir model kullanmıştır. Kiani ve Akbarzadeh [32] da çalışmalarında genetik algoritmayı kullanmıştır. Genetik algoritma aracılığıyla bulanık kümelerin optimize edilmesini sağlamışlardır. McDonald [33] Greedy algoritmasını ve tam sayılı lineer programlamayı kullanarak çoklu doküman özetleme sistemini önermiştir. Çalışmalarında genel ve kullanıcı sorgularına dayalı özetler çıkarılmıştır. Berker ve Güngör [34] özet çıkarmak için 11 özellik kullanmış ve özetleri birleştirmek için genetik algorithmadan yararlanmıştır.

Kumar ve Chandrakala [35] çalışmalarında özet çıkarmak için optimizasyon algoritmalarından yararlanmıştır. Parçacık sürü optimizasyonu, yapay arı kolonisi algoritmaları, genetik algoritmalar ve karınca kolonisi optimizasyonu gibi optimizasyon algoritmalarının özet çıkarımında başarılı sonuçlar verdiğini belirtmişlerdir.

Metin özetleme üzerine yapılan çalışmalarda önışleme işlemleri de yer almaktadır. Suanmali ve arkadaşları [36] çalışmalarında özet çıkarımında önışleme aşamalarına değinmiştir. Cümle çıkarımında çeşitli özelliklerle kullanmış ve bu özellikleri bulanık mantık kurallarına göre birleştirmişlerdir. Kyoomarsi ve arkadaşları [37] ana metni önüşleme aşamasından geçirmiş ve veri setlerini eğitim ve test kümelerine ayırırken Nive Bayes ve C4.5 karar ağaçlarını kullanmıştır.

Milad ve arkadaşları [38] etki alanına özgü kelime düğümleri ve grafik sıralaması kullanılarak biyomedikal makalelerin özetlenmesi için grafik tabanlı bir özetleyici sistem kurmuşlardır.

Yakın geçmişte derin öğrenme yöntemleri metin özetleme görevine uygulanmış ve yüksek başarı oranına ulaşmıştır. Özellikle yoruma dayalı özetleme yöntemlerinde derin öğrenme ve yapay zeka sıklıkla kullanılmaktadır. Derin öğrenmeyi içeren yaklaşımlar, odağını yeterli derinliğe sahip bir sinir ağının kullanıldığı ve cümleleri özete konulması için önemli veya önemsiz olarak sınıflandırılması için kullandıkları bir yaklaşımdır.

Svore ve arkadaşları [39] cümle özelliklerini birleştirmek için yapay sinir ağları kullanmıştır. Dokümanı en iyi ifade eden üç cümle özete dahil edilmiştir.

2017 yılında Mahmood ve arkadaşları [40] gözetimsiz derin öğrenme yöntemini

kullanarak metin özetleme sistemi oluşturmuştur ve elde edilen özetler son derece anlamlı ve insan tarafından oluşturulan özetlere benzer çıkmıştır.

Sinha ve arkadaşları [41] tek belge özetlemesi için ileri beslemeli sinir ağlarını kullanarak tamamen veriye dayalı bir yaklaşım önermiştir. Önerilen model ölçeklenebilir ve orijinal belgeyi sabit boyutlu parçalara bölerek ve sonra tekrar ağı eğiterek farklı boyutlarda özet çıkarılmasını sağlamıştır. 2020 yılında Rupal ve arkadaşları [42] Generative Adversarial Networks (GAN) kullanarak metin özetleme sistemi oluşturmuşlardır.

Literatürdeki çalışmalar incelendiğinde günümüzde metin özetleme araştırma konusunun popülerliğini sürdürdüğü net bir şekilde görülebilmektedir. İlk yapılan çalışmalar çıkarımsal özetlemede yoğunlaşırken günümüzde yoruma dayalı özetleme daha yaygın araştırılmaktadır.

2. MATERYAL VE YÖNTEM

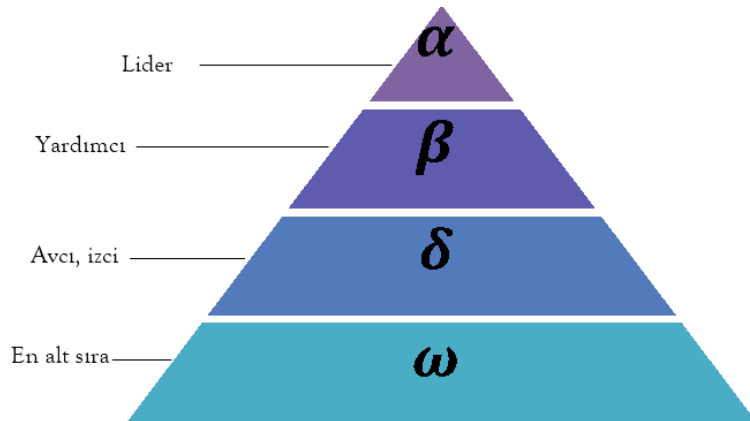
2.1. KULLANILAN ALGORİTMALAR VE ÇALIŞMA ORTAMI

Günümüzde metin özetleme teknikleri büyük verilerden ilgili bilgilerin çıkarılmasında büyük rol oynamaktadır. Sezgisel algoritmalar arařtırmacıların özetleme sistemlerinde sıklıkla çalıştığı bir alandır. Bu çalışmada, sezgisel bir algoritma olan GWO tekli metin özetleme başarısını test edilmiştir. Bu kapsamda GWO kullanılmıştır. Aynı zamanda GWO'nun başarımını karşılařtırabilmek için K-means algoritması ile de sistem test edilmiştir.

Bu tezde bahsi geçen sistem Matlab uygulaması üzerinde test edilmiştir. Windows 7 işletim sistemli, i7 işlemcili ve 16 gb Ram'e sahip bir bilgisayar bu test işlemleri sırasında kullanılmıştır.

2.1.1. Gri Kurt Optimizasyon Algoritması

Günümüzde optimizasyon problemlerinin çözümünde sezgisel algoritmalar sıklıkla kullanılmaktadır. Sezgisel algoritmalar genellikle doğadan ilham alınarak geliştirilen algoritmalarlardır. Yakın gelecekte bu sezgisel optimizasyon algoritmaları arasına Gri Kurt GWO da katılmıştır. GWO Mirjalili [43] tarafından geliştirilmiş meta sezgisel bir optimizasyon algoritmasıdır. Geliştirilmesinde gri kurtların doğadaki yaşam ve avlanma stratejilerinden esinlenilmiştir [44]. Gri kurtların sürüleri hiyerarşik bir yapıya sahiptir, bu yapı Şekil 2.1'de yer almaktadır.



Şekil 2.1. GWO algoritması hiyerarşi zinciri.

Gri kurtların toplumsal hiyerarşik yapısı alfa, delta, beta ve omega olarak ayrılmıştır. Alfa grubu lider olarak adlandırılır ve diğer tüm kurt grupları üzerinde söz sahibidir. Beta grubunun görevi, alfaya karar vermede veya diğer faaliyetlerde yardımcı olmaktır. Üçüncü seviyede ise delta denilen kurtlar yer almaktadır. Bu kategorideki kurtlar izciler, nöbetçiler, avcılar, yaşlılar ve bakıcılardan oluşur. İzciler bölgenin sınırlarını gözlemek ve herhangi bir tehlike durumunda sürüyü uyarmakla yükümlüdür. Nöbetçiler, paketin güvenliğini korur ve garanti eder. Yaşlılar, eskiden alfa veya beta olan uzman kurtlardır. Avcılar av avlarken ve paket için yiyecek sağlarken alfalara ve betalara yardım ederler ve bakıcılar paketteki zayıf, hasta ve yaralı kurtların bakımından sorumludur. En düşük ise seviye omega grubudur. Omega kurtları diğer tüm baskın kurtlara uymak zorundadır.

Gri kurtlar, av konumlarını ezberleme ve kuşatma yeteneğine sahiptir. Alfa ava liderlik etmektedir. Gri kurtların avlanma davranışını matematiksel modele benzetmek için en iyi çözüm alfa olarak kabul edilir. Beta ve delta sırasıyla ikinci ve üçüncü optimal çözümler kabul edilir. Aday çözümlerin geri kalanının omega olduğu varsayılmaktadır. Av, alfa, beta ve delta tarafından yönlendirilirken, omega kurtları bu üç en iyi çözümün pozisyonlarını dikkate alarak pozisyonlarını güncellemektedir.

Kurtların sosyal hiyerarşisine ek olarak, grupça avlama gri kurtların bir başka ilginç sosyal davranışıdır. Muro ve arkadaşları [45] gri kurt avının ana aşamaları şunlardır:

- Avı izleme, takip etme ve yaklaşma
- Av hareket etmeyi durdurana kadar takip etmek, kuşatmak
- Ava Saldırma

Bu aşamalar aşağıdaki Şekil 2.2'de gösterilmiştir.



Şekil 2.2. Gri kurtların avlanma davranışları: a) Avı kovalamak, yaklaşmak ve avı izlemek b-d) Takip etmek, rahatsız etmek ve kuşatmak e) Konumu sabitleme ve saldırı [45].

2.1.1.1. Avı Çevreleme (*Encircling prey*)

Gri kurtların avlanma mekanizmaları temel olarak pozisyonlarını güncelleme üzerine kuruludur. Gri kurtlar, avın çevresinde bulunacakları pozisyonları Denklem (2.1) ve (2.2)'yi kullanarak rasgele güncelleyebilir:

$$D = |C \cdot X_p(t) - X(t)| \quad (2.1)$$

$$X(t+1) = |X_p(t) - A \cdot D| \quad (2.2)$$

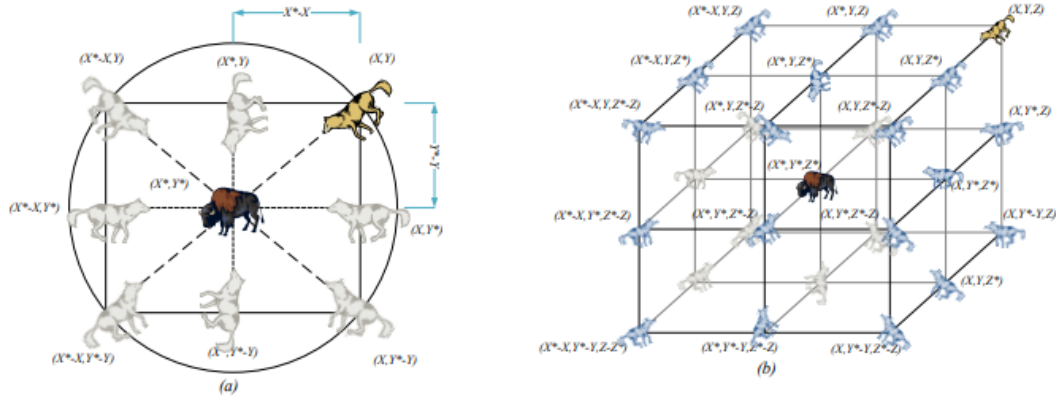
Denklemden t parametresi mevcut iterasyonu, A ve C parametresi katsayı vektörlerini, X_p avın bulunduğu konum vektörünü ve X bir gri kurdun konumunu temsil göstermektedir. A ve C değerleri Denklem (2.3) ve (2.4)'teki gibi sırasıyla hesaplanır:

$$A = |2a \cdot r1 - a| \quad (2.3)$$

$$C = |2a \cdot r2| \quad (2.4)$$

(x, y) pozisyonunda gri kurt av pozisyonuna göre konumunu günceller. En iyi ajan A ve C vektör değerlerini güncel konumuna göre değiştirerek farklı değerlere ulaşabilir. Aynı kavram n boyutlara sahip bir arama alanında uzatılabilir. Şekil 2.3'de verilen 2D

ve 3D uzayı, n boyutlu bir arama uzayına genişletilebilmektedir ve gri kurtların şimdiye kadar elde edilen en iyi çözüm çevresinde hiper-küplerde hareket etmesi sağlanabilir.



Şekil 2.3. Gri kurtların 2D ve 3D uzayındaki olası konumları [43].

2.1.1.2. Avlama (Hunting)

Gri kurtların alfa, beta ve delta türleri avlama sırasında avın mevcut konumu hakkında büyük bir bilgiye sahiptirler. Bu yüzden, ilk üç en iyi çözüm kaydedilir ve sonrasında diğer kurtların pozisyonlarını en iyi arama ajanlarının pozisyonlarına göre güncellemeleri sağlanır. Bu işlemin adımları aşağıdaki denklemlerde belirtilmiştir.

$$D\alpha = |C1. X\alpha - X| \quad (2.5)$$

$$D\beta = |C2. X\beta - X| \quad (2.6)$$

$$D\delta = |C3. X\delta - X| \quad (2.7)$$

$$X1 = |X\alpha - A1. D\alpha| \quad (2.8)$$

$$X2 = |X\beta - A2. D\beta| \quad (2.9)$$

$$D\alpha = |C1. X\alpha - X| \quad (2.10)$$

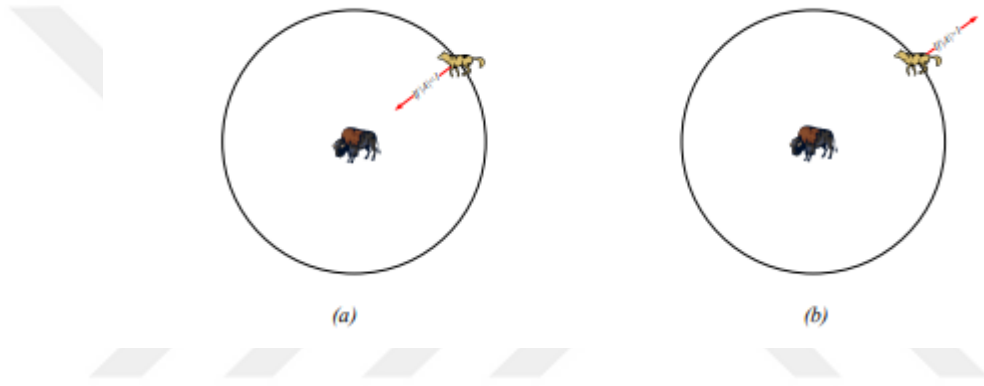
$$D\alpha = |C1. X\alpha - X| \quad (2.11)$$

Nihai konumun, arama alanındaki alfa, beta ve delta konumları ile tanımlanan bir daire içinde rastgele bir yerde olacağı gözlemlenebilir. Başka bir deyişle, alfa, beta ve delta avın konumunu tahmin eder ve diğer kurtlar avın etrafındaki konumlarını rastgele

günceller.

2.1.1.3. Ava Saldırma (*Attacking Prey*)

Gri kurt av durduğunda ava saldırarak avı öldürür. Sırayla matematiksel modelle yaklaşan avın a vektörünün değerini azaltılır. Böylelikle a vektörüne bağlı A vektörü de azalır. Başka bir deyişle A vektörü $[-2a, 2a]$ değeri arasında rastgele bir sayıdır ve (a) yineleme boyunca 2'den 0 a düşürülmüştür. A vektörü $[-1,1]$ içinde rastgele değer olduğunda, arama ajanının bir sonraki pozisyonu geçerli pozisyonu ve avın pozisyonu arasında herhangi bir pozisyonda olabilir. Şekil 2.4, $|A| < 1$ 'in kurtları avına saldırmaya zorladığını göstermektedir.



Şekil 2.4. Ava saldırının aranması [43].

2.1.1.4. Arama (*Search for prey*)

Gri kurtlar kurtların konumuna göre arama yaparlar. Gri kurtlar arama yapmak için dağılır ve avı bulduğunda saldırmak için toplanırlar[43]. Dağılımı matematiksel olarak modelleyebilmek için, rastgele değerlere sahip A parametresi kullanılır. Yapılan bu işlem, aramayı önemli kılar ve GWO algoritmasının global aramasına ortam sağlar. GWO algoritmasının sözde kodu (akış şeması) Şekil 2.5'de verilmiştir.

```
Gri Kurt Populasyonunu Başlat
a, A C parametrelerini ata
Her kurtun uygunluk değerini hesapla
Birinci, ikinci ve üçüncü en iyi çözümleri belirle
while (t< Maksimum İterasyon sayısı)
    for Her bir ajan
        Pozisyon güncelle
    end
    a, A ve C parametrelerini güncelle
    Her bir kurtun uygunluk değerini güncelle
    Birinci, ikinci ve üçüncü en iyi çözümleri güncelle
    t = t+1;
end
```

Şekil 2.5. GWO algoritmasının sözde kodu.

2.1.1.5. GWO Uygulamaları

GWO algoritması araştırmacılar tarafından yakın geçmişte matematiksel problemlerin çözülmesinde yaygın olarak kullanılmıştır ve günümüzde popüler bir sezgisel algoritmadır. GWO 2016 yılında Moumita ve arkadaşları tarafından [46] Ekonomik Yük Sevkiyatının (ELD) optimal çalışma stratejisi için kullanılmıştır.

Ramji ve arkadaşları [47] çalışmalarında Otonom Mobil Robotların (AMR) yol optimizasyonlarında GWO'dan faydalanmıştır ve engellerin aşılmasında etkili sonuçlar elde edildiği gözlemlenmiştir. Dipayan ve arkadaşları[48] büyük ölçekli güç sisteminin yük frekansı kontrolünde GWO'yu kullanmıştır.

Shubham ve arkadaşları [49] çalışmalarında uydu görüntü segmentasyonu için GWO algoritmasının bir uygulamasını önermiştir. GWO, otomatik bir kümeleme algoritması olarak çalışacak şekilde değiştirilmiştir.

GWO görüntü işlemede de kullanılmıştır. Shankar ve arkadaşları [50] beynin görüntü analizine dayanan Alzheimer tespiti için yeni bir mode önermiştir. Çalışmada Alzheimer hastalığının doğru tespiti için%96.23 hassasiyet elde edilmiştir.

Sürekli ve ayrık optimizasyon problemlerinin çözümünde kullanılan GWO ile diğer sezgisel algoritmalar ile birleştirilerek de çalışılmıştır [50,51]. Ayrıca lineer olmayan denklem sistemleri için yeni bir çözüm yaklaşımı olarak sunulmuştur [52]. GWO, sınıflandırma [53], sistem tanımlaması ve filtre tasarımı [54], özellik seçimi [55] gibi birçok mühendislik probleminin çözümü için başarıyla uygulanmıştır.

2.1.2. K-means

Kümeleme kısaca benzer özellikleri bulunan verilerin, kendi aralarında gruplara ayrılması işlemidir. Yani kümeleme işleminin olabilmesi için benzer özellik gösteren bireylerin aynı kümede yer alması sağlanmalıdır. Kümeleme problemi araştırmacıların araştırdığı yaygın bir optimizasyon problemidir. K-means ise her verinin yalnızca tek kümeye ait olmasına izin vermesi özelliğiyle keskin bir kümeleme algoritmasıdır[56].

K-means'in genel mantığı şu şekildedir. N adet veriden oluşan bir veri seti kümelenecek, k adet giriş parametresi kadar kümeye bölünmektedir. Burada amaç, kümeleme işlemi sonucunda ortaya çıkan kümelerin, kendi aralarındaki benzerliklerinin maksimum ve yine bu kümelerin birbirleriyle olan benzerliklerinin minimum olmasını sağlamaktır. Bu kümeleme işlemi en iyi çözüme ulaşıncaya kadar devam etmektedir [57].

K-means algoritması, karesel-hata fonksiyonu minimize olacak şekilde algoritmaya girdi olarak kullanıcı tarafından gönderilen k parametresi ile toplamda n adet veriden oluşan veri setini k adet kümeye bölmektedir [58]. K-means algoritmasının kümeleme sırasında gerçekleştirdiği işlem basamakları dört adımdan oluşmaktadır. Bu adımla şu şekildedir:

- 1.Adım: İlk adım olarak küme merkezi saptanır. Oluşturulacak olan küme sayısı yani k kadar rastgele nokta belirlenir.
- 2.Adım: Veri setinde yer alan tüm nesnelerin seçilen noktalara olan uzaklığı hesaplanır ve tüm nesneler bu noktalardan kendilerine en yakın olan kümeye ait olarak atanır.
- 3.Adım: Elde edilen kümelerin belirlenecek olan yeni merkez noktaları mevcut kümede yer alan tüm nesnelerin ortalamaları ile değiştirilir.
- 4.Adım: Bu adımda ise ikinci ve üçüncü adımlar merkez noktalar sabitleninceye kadar tekrarlanır.

K-means kümeleme algoritması günümüzde kullanılmaya devam etmekte ve bir çok çalışmada başarılı sonuçlar çıkarmaktadır. Bu çalışmada K-means kümeleme algoritması metin özetlemede kullanılmış ve GWO'nun özet çıkarma başarısının karşılaştırılmasında yer almıştır.

2.1.3. Veri Seti

Bu çalışmada BBC News [59] veri seti kullanılmıştır. Veri seti BBC News internet haber sitesinden toplanan 2225 haber metnini içermektedir. Bu metinler beş kategoride sınıflandırılmıştır. Bu beş kategori iş, spor, eğlence, politika ve teknolojiden oluşmaktadır. Veri setinde yer alan toplam kelime bilgisi ise 8865'tir. Şekil 2.6'da veri setinden bir kesit gösterilmektedir. Veri seti örneğinde de görüldüğü gibi her metnin kendisi ve dahil olduğu kategori bilgisi yer almaktadır.

```
Category,Description
tech,tv future in the hands of viewers with home theat
business,worldcom boss left books alone former world
sport,tigers wary of farrell gamble leicester say the
sport,yeading face newcastle in fa cup premiership sid
entertainment,ocean s twelve raids box office ocean s
politics,howard hits back at mongrel jibe michael howa
politics,blair prepares to name poll date tony blair is
sport,henman hopes ended in dubai third seed tim henma
sport,wilkinson fit to face edinburgh england captain
entertainment,last star wars not for children the si
entertainment,berlin cheers for anti-nazi film a germa
business,virgin blue shares plummet 20% shares in aust
business,crude oil prices back above $50 cold weather
politics,hague given up his pm ambition former conse
sport,moya emotional after davis cup win carlos moya d
```

Şekil 2.6. Veri seti örneği.

Çalışmada bu veri setinin kullanılma nedeni oluşturulan sistemin hem özetleme başarısını test etmek hem de bunun sınıflandırmaya olan etkisini gözlemleyebilmektir. Kullanılan veri setinde kategori bilgisinin yer alması sınıflandırma başarısını da test edebilmemize olanak sağlamıştır.

2.2. BAŞARI DEĞERLENDİRME ÖLÇÜTÜ

Otomatik metin özetlemede kullanılan algoritmaların başarıların test edilmesini için çeşitli ölçüm yöntemleri bulunmaktadır. Bu yöntemler iki türlü olabilmektedir. Birincisi görev tabanlı olmayan yöntemlerdir ve uzmanlar tarafından oluşturulmuş yani insan gözüyle çıkarılmış özetler değerlendirme ölçütü olarak alınmaktadır. Diğeri ise görev tabanlı yöntemlerdir. Bu yöntemlerde uzman özetleri soru cevaplama, sınıflandırma gibi belirli bir alan için kullanılır.

Tez kapsamında görevden bağımsız yöntemlerden faydalanılmıştır. Otomatik sistem tarafından çıkarılan özet ile ideal yani insan gözüyle üretilen özet karşılaştırılır. Değerlendirme işlemi karşılaştırılan iki özette bulunan çakışan yani ortak cümle sayıları dikkate alınarak yapılmıştır. Literatürde keskinlik, f-ölçüm değeri, kosinüs benzerliği ve Ngram birliktelik istatistiği gibi görev tanımsız yöntemler yer almaktadır. Ayrıca literatürde Blue ve Meteor gibi metrikler de doğruluğun ölçülmesinde yaygın kullanılan yöntemlerdir. Bu çalışmada değerlendirme ölçütü olarak Ngram birliktelik istatistiği yani ROUGE kullanılmıştır.

2.2.1. Ngram birliktelik istatistiği (ROUGE)

ROUGE (Recall- Oriented Understudy for Gisting Evaluation) birliktelik istatistiği 2004 yılında Lin tarafından Perl programlama dili kullanılarak üretilmiştir[60]. ROUGE karşılaştırılacak olan iki dokümanın ortak kelime sayısına dayanan bir ölçüm yöntemidir.

ROUGE'nin beş farklı ölçüm şekli bulunmaktadır: ROUGE-N, ROUGE-L, ROUGE-S, ROUGE-W, ROUGE-SU. ROUGE metriğinde “Recall” ve “Precision” isimli iki ölçüm gerçekleştirilmektedir. Recall yani geri bildirim ölçümü ile oluşturulan sistemde yer alan özetle referans olarak alınan özetteki çakışa kelime sayıları hesaplanmaktadır. Bu hesaplama işlemi Denklem (2.12)’de görülebilmektedir. Precision yani hassaslık ölçümü ise Denklem (2.13)’deki gibi hesaplanmaktadır [60].

$$Rouge Recall = \frac{\text{Çakışan Kelime Sayısı}}{\text{Referans Özetteki Toplam Kelime Sayısı}} \quad (2.12)$$

$$Rouge Precision = \frac{\text{Çakışam Kelime Sayısı}}{\text{Sistem Özetiindeki Toplam Kelime Sayısı}} \quad (2.13)$$

ROUGE metriği hesaplamalarda n-gram kullanılmaktadır. Denklem (2.14)’te belirtilen “N” iki özet arasındaki çakışmaları değerlendirecek kelime zincirinin uzunluğunu ifade etmektedir. N değeri 1,2,3 ve 4 olabilmektedir.

$$ROUGE - N = \frac{\sum_{s \in \{\text{insan özetleri}\}} \sum_{gram_N \in S} Hesapla_{cakışan}(gram_N)}{\sum_{s \in \{\text{insan özetleri}\}} \sum_{gram_N \in S} Hesapla(gram_N)} \quad (2.14)$$

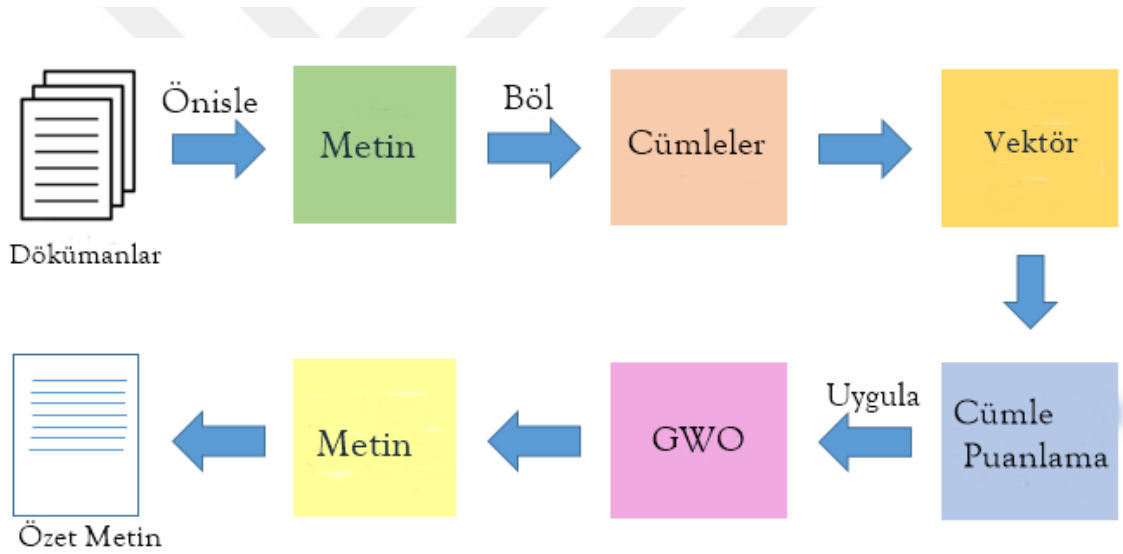
Burada $Hesapla_{cakisan}(gram_N)$ ideal özet ve sistem tarafından oluşturulan özetin ortaklaşa sahip olduğu maksimum Ngram sayısıdır (N uzunluktaki sıralı kelime grubu sayısı.) $Hesapla(gram_N)$ ise ideal özetteki toplam Ngram sayısıdır. Tez kapsamında ROUGE-1 ve ROUGE-2 ölçümleri değerlendirmek için kullanılmıştır.



3. GRİ KURT OPTİMİZASYON ALGORİTMASI İLE ÇIKARIMSAL METİN ÖZETLEME

3.1. SİSTEMİN TASARLANMASI

Bu tez çalışmasında GWO algoritması tabanlı çıkarımsal metin özetleme yöntemi kullanılmıştır. Veri setinde yer alan 2225 haber metni girdi olarak kullanılmış ve çeşitli önışleme aşamalarından geçirilmiştir. Özet metinde yer alan cümleler çeşitli parametrelere göre puanlandırılmıştır. Ardından GWO yardımıyla özetde yer alması muhtemel olan cümleler seçilerek özet metin oluşturulmuştur. Şekil 3.1’de sistemin akış şeması gösterilmektedir.



Şekil 3.1. Sistemin akış şeması.

3.2. ÖNİŞLEME ADIMLARI

Metin özetleme işleminde ilk olarak özetlenecek metnin çeşitli aşamalardan geçmesi gerekmektedir. Bu aşamalar belgelerin cümlelere bölünmesi, durma kelimelerinin kaldırılması, büyük/küçük harf dönüşümü ve noktalama işaretlerinin kaldırılmasından oluşmaktadır. Aşağıda bu aşamalar ayrıntılı anlatılmıştır ve çalışmada yer alan tüm dokümanlar bu işlemde geçirilmiştir.

3.2.1. Büyük/ Küçük Harf Dönüşümü

Oluşturulan özet sisteminde büyük/küçük harf hassasiyeti daha net sonuçlar almak amacıyla kaldırılmıştır. Bu yüzden tüm dokümanlarda her bir kelime otomatik olarak küçük olacak şekilde ayarlanmıştır.

3.2.2. Durma Kelimelerinin Kaldırılması

Durma kelimeleri metinlerde sıkça kullanılan fakat metne çok büyük anlam katmayan kelimelerdir. Bu kelimelere ve, ve ya gibi kelimeler örnek verilebilir. Bu kelimelerin metinlerden kaldırılma nedeni bir sonraki adımda anlatılacak olan cümle puanlama işleminde daha doğru ve faydalı sonuçlar elde etmektir.

3.2.3. Belgelerin Cümlelere Bölünmesi

Özet çıkarılacak olan her metin bu aşamada cümlelere bölünmüştür. Cümle sınırları virgül, noktalı virgül, soru işareti, ünlem işareti veya nokta içeren noktalama işareti kümesiyle tanımlanmıştır.

3.2.4. Noktalama İşaretlerinin Kaldırılması

Noktalama işaretleri özet çıkarılacak metinden kaldırılması gereken diğer bir özelliktir. Özet çalışmasında kelimelere bölme işlemi yapılacağından tüm noktalama işaretleri kaldırılmıştır.

3.2.5. Metni Kelimelere Bölme İşlemi

Özet çıkarılacak metinleri her biri ilk olarak cümlelere ayrıştırılmıştır. Bu her cümle içerdikleri boşluklara göre parçalara ayrılmıştır. Bu işlem kelimelere bölme (tokenizasyon) olarak adlandırılmaktadır ve token olarak adlandırılan elde edilen her parça birleştirilerek bir token listesi oluşturulmuştur.

3.3. CÜMLELERİN PUANLANDIRILMASI

Bu aşama, belgenin her cümlesinin önemini belirlemek için bir puan vermeyi amaçlamaktadır. İlgili metin cümlelerinin puanlarını hesaplamak için dört özellik kullanılmıştır. Bunlar, cümle konumu, cümle uzunluğu, terim sıklığı, başlığa olan benzerliktir. Her bir kriter alt başlıklarda ayrıntılı incelenmiştir.

3.3.1. Cümle Uzunluğu

Bu özellik, bir metindeki çok uzun ya da çok kısa cümlelerin özete dahil edilme olasılığının düşük olmasını içermektedir. Her cümlenin uzunluğuna, yani cümledeki (ön işlemeden sonra) mevcut olan kelime sayısına göre bir puan verilir. Bu özellik, daha uzun bir cümlenin önemli bilgileri içermeye olasılığının yüksek olduğu iddiasına dayanmaktadır. Bu durumda, kısa cümleler genellikle daha az bilgi aktardıkları için özete dahil edilmez. Ayrıca, çok uzun cümleler de bir özeti temsil etmek için uygun değildir. Bu puanlama Denklem (3.1) kullanılarak yapılmıştır.

$$C_u \text{ Uzunluk}(i) = \frac{\text{Cümledeki Kelime Sayısı}(i)}{\text{En uzun cümledeki kelime sayısı}} \quad (3.1)$$

C_u seçilen cümlenin(i) uzunluk puanını ifade etmektedir. Bu puan 0 ile 1 puan arasında olmaktadır.

3.3.2. Cümle Konumu

Bir cümlenin metin içerisindeki konumu, özet çıkarılacak metnin konusuyla en alakalı cümleleri bulmada önemli bir faktör olabilir. Metinlerin başındaki cümlelerin, onları takip eden cümlelerden daha fazla bilgi içermesi muhtemeldir. Metnin ilk cümlesi çok önemlidir ve özete dahil edilmelidir; bu nedenle belgenin başlangıcında yer alan cümlelere diğer cümlelere göre daha fazla ağırlık verilmiştir. Bu puanlama Denklem (3.2) kullanılarak yapılmıştır.

$$C_k \text{ Konum}(i) = 1 - \frac{\text{pos}(i)}{N} \quad (3.2)$$

Denklemden $\text{pos}(i)$ seçilen i cümlesinin başlığa olan konumunu ifade etmektedir. N ise metindeki toplam cümle sayısıdır.

3.3.3. Başlığa Olan Eşlik

Puanlama etki eden diğer bir özellik başlığa olan benzerliktir. Bir cümle başlıkla ne kadar çok benzerse özette yer alma olasılığı o kadar yüksektir. Bu özellik cümlenin başlıkla olan benzerliğine (eş kelimelere) göre hesaplanmıştır. Hesaplama Denklem (3.3) kullanılarak yapılmaktadır.

$$C_b \text{ Benzerlik}(i) = \frac{\text{Başlıkla benzer kelime sayısı}(i)}{\text{Başıktaki kelime sayısı}} \quad (3.3)$$

3.3.4. Terim Sıklığı

Terim sıklığı, bir belgedeki bir kelimenin sıklığına göre önemini değerlendirmek için kullanılır. Her cümleye, içinde görünen her kelimenin metindeki sıklığına göre bir puan verilir. Bu puan ne kadar yüksek olursa, o cümle özet için o kadar önemlidir. Terim sıklığı Denklem (3.4) kullanılarak hesaplanmıştır.

$$C_s \text{ Sıklık}(i) = \frac{\sum_{j=1}^{Tsi} Fij}{Tsi} \quad (3.4)$$

Denklemden Tsi, i cümlesindeki kelime kümesi, Fij ise i cümlesindeki j kelimesinin terim sıklığıdır.

3.3.5. Cümlenin Toplam Puanı

Cümle puanı, cümlenin gücünü temsil eder; puanın değeri ne kadar fazlaysa, özet için iyi bir aday olma olasılığı o kadar yüksektir. Her cümle için nihai puan, daha önce denklem kullanılarak çıkarılan tüm özelliklerin toplanmasıyla hesaplanır. Hesaplama Denklem (3.5)'de görülebilmektedir. Ayrıca bir cümlenin nihai puanının hazırlandığı bir örnek Çizelge 3.1'de verilmiştir.

$$C_i = C_u + C_k + C_b + C_s \quad (3.5)$$

Çizelge 3.1. Cümle final fkoru örneği.

Cümle Uzunluğu	Terim Sıklığı	Başlığa Benzerlik	Cümle Konumu	TOPLAM
0.65	0.24	0.67	0.56	2.12

3.4. CÜMLELERE KOSİNÜS BENZERLİĞİNİN UYGULANMASI

İyi bir özet oluşturmak için sadece yüksek bilgilendiriciliğe sahip cümleler çıkarmak yeterli değildir. Bunun yanında özel içerisindeki cümlelerin birbirleri ile alakalı olmaları gerekmektedir. Bu yüzden tutarlı bir özet çıkarmak için cümlelerin birbiriyle

benzerlikleri ölçülmelidir. Dolayısıyla her bir cümle için benzerliğini hesaplamamız gerekir. Bu benzerlik kosinüs benzerliği teoremi kullanılarak hesaplanmıştır.

İki vektör arasındaki açı ve bu vektörlerin iç çarpımının çarpılmasıyla hesaplanan uzaklık; kosinüs benzerliği olarak formüle edilir. Denklem (3.6) cümlelerin birbirleriyle olan benzerliklerini hesaplamak için kullanılmıştır.

$$\cos(X, Y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i (x_i)^2} \sqrt{\sum_i (y_i)^2}} \quad (3.6)$$

Denklemde yer alan $\cos(X, Y)$ metinde yer alan X ve Y cümlelerinin birbirleriyle olan benzerliklerini vermektedir. Çizelge 3.2’de bir metnin kosinüs benzerlik matrisi örneği verilmiştir.

Çizelge 3.2. Kosinüs benzerlik matrisi örneği.

Cümle	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	1	0.52	0.67	0.56	0.12	0.16	0.41	0.01	0.01	0
C2	0.52	1	0.20	0.04	0	0	0	0	0	0
C3	0.67	0.20	1	0.16	0	0.32	0.26	0.09	0.02	0
C4	0.56	0.04	0.16	1	0	0.09	0.24	0.03	0.14	0
C5	0.12	0	0	0	1	0.29	0.33	0.12	0.18	0
C6	0.16	0	0.32	0.09	0.29	1	0.21	0.13	0.17	0
C7	0.41	0	0.26	0.24	0.33	0.21	1	0.29	0	0
C8	0.01	0	0.09	0.03	0.13	0.15	0.28	1	0	0
C9	0.01	0	0.02	0.14	0.17	0.17	0.01	0	1	0
C10	0	0	0	0	0	0	0	0	0	1

Yukarıdaki çizelge C’ler her bir cümleyi ifade etmektedir. Çizelgeden bir cümle için kendi ile olan benzerliğinin 1 sonucunu verdiği görülebilmektedir.

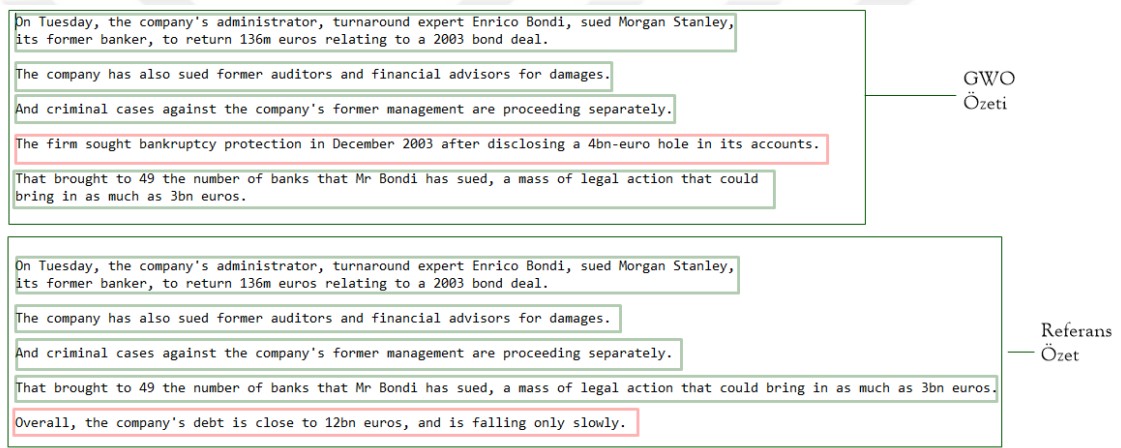
3.5. GWO İLE SİSTEMİN ÇALIŞTIRILMASI

Bu aşamada önerilen yaklaşım, cümle ve bilgilendirici puan arasındaki bütünlük kapsamında özet için optimum cümle kombinasyonunu sağlamak için GWO

algoritmasının kullanılması üzerine kurulmuştur.

Ağırlıkları hesaplanan ve kosinüs benzerlikleri belirlenen cümleler GWO yardımıyla girilen parametrelerle kümelendirilmiştir ve özet oluşturulmuştur. Çalışmada metinlerin küçültme oranı %60 olarak belirlenmiştir. Bu oran veri setinde bulunan insan gözüyle oluşturulan özete en yakın olan küçültme oranıdır. Bu oran aynı zamanda oluşturulacak olan küme sayısına etki etmektedir.

Çalışmada kullanılan diğer parametrelerin değerleri ise şu şekildedir. SearchAgents_no:5, Max_iterasyon=300, alpha=0.1, n_grid = 10, beta=4 ,gamma=2 (Algoritma için önerilen değerler kullanılmıştır.) GWO ile oluşturulmuş bir özet örneği ve referans özetle karşılaştırması Şekil 3.2’de gösterilmiştir.

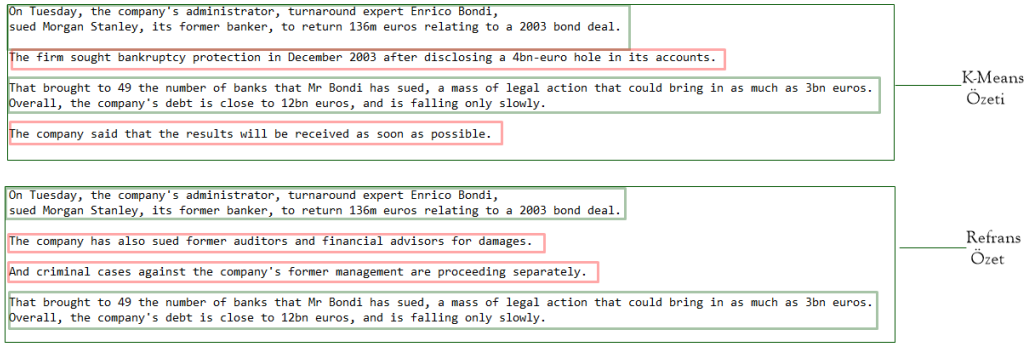


Şekil 3.2. GWO ile oluşturulan özet örneği.

3.6. K-MEANS İLE SİSTEMİN ÇALIŞTIRILMASI

GWO'nun kümeleme yöntemiyle özet çıkarım başarısını karşılaştırabilmek için aynı veri seti GWO'da uygulanan önışleme adımları uygulanarak literatürde yaygın olarak kullanılan ve kayda değer sonuçlar veren K-means kümeleme algoritması ile de test edilmiştir.

Küçültme oranı %60 olarak belirlenmiştir. Bu oran K-means algoritmasının oluşturacağı küme sayısı ile doğrudan orantılıdır. K-means ile oluşturulmuş bir özet örneği ve referans özetle karşılaştırması Şekil 3.3'de gösterilmiştir.



Şekil 3.3. K-means ile oluşturulan özet örneği

3.7. ÖZETİN DEĞERLENDİRİLMESİ

GWO ve K-means ile yapılan çalışmalar sonucu iki kümeleme algoritması ile de başarılı sonuçlar elde edildiği gözlemlenmiştir. Bu iki algoritmasının karşılaştırılması ise ROUGE metriği ile sağlanmıştır. ROUGE-1 ve ROUGE-2 değerlerine göre iki algoritmanın verdiği sonuçlar aşağıdaki tablodaki gibidir. Elde edilen sonuçların ayrıntıları ve değerlendirmesi sonuç kısmında yapılmıştır.

Çizelge 3.3. GWO ve K-means'in özetleme başarıları.

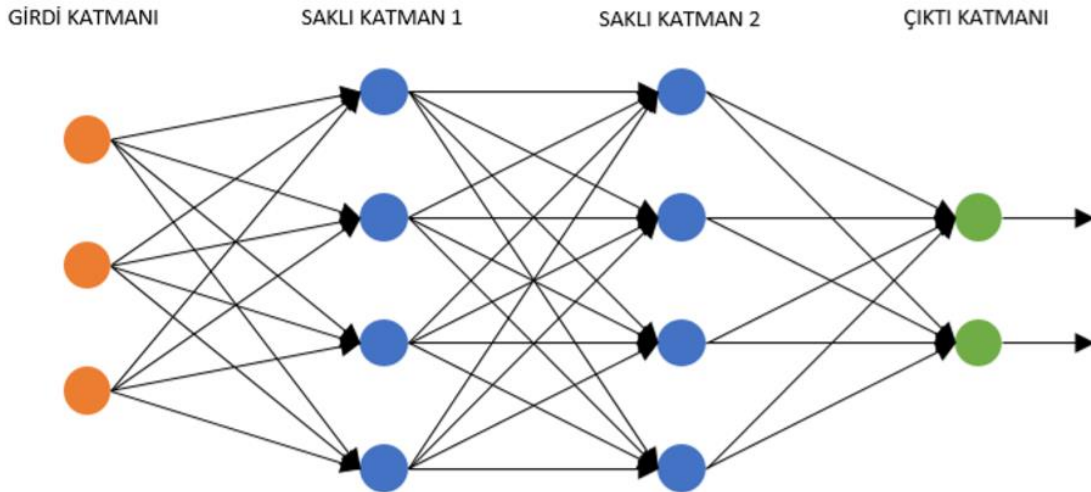
	GWO	K-means
ROUGE-1	61.67	56.18
ROUGE-2	48.14	41.76

4. DERİN ÖĞRENME

Günümüzde yapay zeka, makine öğrenmesi gibi araştırma alanları oldukça fazla kullanılmaktadır. Derin öğrenme ise popülerliğini koruyan makine öğrenmesi alanlarının önemli bir parçasıdır. İnsanın beynini ilham kaynağı olarak alan derin öğrenme, beyin neokorteksinin veriyi karmaşık bir hiyerarşiden geçirdiğini ortaya koyan sinir bilimi bulgularını kullanmaktadır [61]. Derin öğrenme yaklaşımları, temelini sinir ağlarından alır; isminde yer alan 'derin' kelimesi birbirinin üzerine yığılan birkaç katmanı ifade etmektedir.

Derin öğrenme, sistemi direkt olarak verilen veriler üzerinden öğrenmektedir. Sinir ağları, diğer algoritmaların kolayca keşfedemeyeceği özellikleri ve aralarındaki ilişkileri kolayca öğrenebilirler. Yapay sinir ağları, matematiksel olarak veriyi dönüştüren nöron katmanlarını kullanarak karmaşık bir model öğrenirler.

Sıradan bir sinir ağı modeli, girdi katmanı, çıktı katmanı ve ikisinin arasındaki bilgi akışını sağlayan gizli katmandan oluşmaktadır. Derin öğrenme terimi, birçok saklı katmana sahip modeller için kullanılmaktadır. Şekil 4.1'de bir derin öğrenme sinir ağı mimarisinin genel görünümü verilmiştir.



Şekil 4.1. Derin sinir ağı mimarisinin genel görünümü.

Derin öğrenme ağ yapısını basit bir yapay sinir ağından ayıran en temel özellik

görülebileceği gibi birden fazla gizli katmana sahip olması ve daha karmaşık bir ağ yapısına sahip olmasıdır.

4.1. DOĞAL DİL İŞLEMEDE DERİN ÖĞRENME UYGULAMALARI

Derin öğrenme uygulamaları Doğal dil işleme alanında yaygın olarak kullanılmakta ve kayda değer sonuçlar vermektedir. Derin öğrenme uzun yıllardır araştırmacıların temel ilgilendiği yapılardan biridir. Derin öğrenme yöntemleri aynı zamanda; video, ses, metin gibi farklı veri türlerinin işlenmesinde de kullanılmıştır [62]. Özellikle DAE, LSTM, CNN mimarileri etkileyici sonuçlar vermiştir. Derin öğrenmeye dayalı sinir ağı modelleri, cümle ve belge gösterimi [63], sınıflandırma [64], istatistiksel makine çevirisi metin sınıflandırması [65] gibi birçok NLP çalışmasında kullanılmıştır.

Sınıflandırma çalışmalarının içerisinde, Evrişimli Sinir Ağı (CNN) ve tekrarlayan sinir ağı(RNN) iki popüler ağıdır. Liu ve arkadaşları [66] RNN'nin çoklu görev öğrenmesine dayanan üç metin sınıflandırma yöntemi getirmiştir. Çalışmalarına her kelimenin önemini doğrudan ortaya koyan RNN modeline de eklenmiştir. Araştırmacıların çoğu CNN ve RNN'nin avantajlarını, küresel uzun vadeli bağımlılıkları ve yerel semantik özellikleri ayıklamak için kullanarak birleştirmiştir.

Tai ve arkadaşları [67] zengin anlamsal gösterimleri öğrenmek için bir LSTM'nin ağaç yapılı ağ tipolojisi olan bir Tree-LSTM modeli geliştirmiştir. Yazarlar, Tree-LSTM'nin NLP görevleri için zincir-yapılı LSTM'den daha iyi bir model olduğunu savunmaktadır, çünkü doğal dil, kelimeleri ifadelerle birleştirecek sözdizimsel özellikler sergilemektedir.

Metin sınıflandırma için de derin öğrenme yöntemleri yaygın olarak kullanılmıştır. Yapılan bir çalışmada farklı veri kümeleri üzerinde cümle sınıflandırma işlemi için basit seviyeli bir sinir ağı kullanılmıştır ve başarılı sonuçlar elde edilmiştir. Bu çalışmada CNN ile önceden eğitimi gerçekleştirilmiş olan kelime vektörleri cümlenin sınıflandırma işleminde kullanılmış ve yedi farklı kategoride yüksek başarılar göstermiştir [68].

Yinelemeli sinir ağları metin sınıflandırma işlemlerinde en çok kullanılan derin öğrenme tekniğidir. Örnek bir çalışmada cümlelerin birbirleriyle olan bağlantısı için yinelemeli sinir ağları kullanılmıştır [69]. Başka bir çalışmada ise evrişimsel sinir ağlarının sınıflandırma başarıları test edilmiştir. Elde edilen sonuçlar evrişimsel sinir

ağlarının metni sınıflandırmada başarılı bir yöntem olduğunu ortaya koymuştur [70].

Kısa metinlerin sınıflandırılması için, yinelemeli yapay sinir ağı olan LSTM ağlarının kullanılması önerilmiştir. LSTM ağları ile oluşturulan sistem farklı veri setleriyle test edilmiş ve sonuçlar LSTM'nin başarılı bir sınıflandırıcı olduğunu ortaya koymuştur [71]. Metinlerdeki anahtar kelimelerin tespit edilmesinin önemini vurgulayan bir diğer çalışmada LSTM ağları kullanılarak tasarlanan bir sistem farklı ağlarla karşılaştırılmış ve LSTM'nin daha başarılı olduğu sonucuna varılmıştır [72].

Başka bir çalışmada kullanıcının okumuş olduğu metinleri baz alarak hedef kullanıcının ilgi duyacağı düşünülen metinleri tahmin etmeyi amaçlayan CNN içeren bir sistem kurulmuştur. Bu model, her bir metin kaynağını hedeflerle eşleştirmiştir ve kaynak-hedef ilişkilerini içeren vektörlerle tanımlanmıştır. Bu sayede metinlerin aynı zamanda CNN ağlarının konu modelleme başarısını da ortaya koymuştur [73].

4.2. DERİN ÖĞRENME MİMARİLERİ

Derin öğrenme mimarileri farklı yapılarda olabilmektedir. Genel olarak derin öğrenme mimarileri altı farklı kategoriye ayrılmaktadır [74]. Alt bölümde bu mimariler incelenmiştir.

4.2.1. Derin Sinir Ağları

Bu sinir ağında, sinir ağı toplamda iki katmandan oluşmaktadır. Bu katmanlar girdi ve çıktı katmanlarıdır. Girdi katmanı direk olarak çıktı katmanına bağlıdır ve bu ağlar genellikle lineer olarak ayrılabilen problemlerde kullanılmaktadır. Daha karmaşık olan problemlerin çözümünde ise katmanların ağırlıklarının farklılaştırılması yöntemi uygulanmaktadır.

4.2.2. Derin Oto Kodlayıcılar

Derin Oto Kodlayıcılar, herhangi bir problemin çözüm aşamasında veri setinin öznelik kümesinin otomatik olarak çıkarılmasını sağlamaktadır. Oto Kodlayıcılar verilere sınıf etiketi atamak yerine bu verilerin girdi vektörüne göre yeniden oluşturulması temeline dayanmaktadır. Derin Oto Kodlayıcıların farklı problemlerin çözümü için geliştirilmiş birçok türü bulunmaktadır.

4.2.3. Derin İnanç Ağları

Derin İnanç Ağları (Deep Belief Networks), çok katmanlı olan ve aynı zamanda grafiksel model içeren bir derin sinir ağıdır. Bu derin sinir ağında her katmanın bağlantılı olduğu birden çok gizli katman bulunmaktadır. Her katman arasında bağlantı bulunmaktadır fakat düğümler arasında bu bağ yer almamaktadır.

4.2.4. Derin Boltzmann Makinesi

Derin Boltzmann Makineleri bir diğer derin öğrenme mimarisidir. Derin Boltzmann Makineleri temelde iki katmandan oluşmaktadır ve aynı zamanda rastlantısal bir yapay sinir ağı yapısına sahiptir. Bu ağ yapısı sınıflandırma problemlerinde kullanılabilir. Boltzmann makinelerinde zaman karmaşıklığının çok fazla olmasından dolayı öğrenme süreci yavaş işlemektedir bu da mimarinin dezavantajı olarak görülmektedir. Öğrenme sürecinin yavaşlığı büyük veriler için kullanışsız bir mimari olarak görülmesine neden olmaktadır.

4.2.5. Evrimsel Sinir Ağları (CNN)

Memelilerin görme sistemini örnek alan konvolüsyonel sinir ağı, çok sayıda konvolüsyon işlemi ve örnekleme katmanına sahip ileri beslemeli bir ağ yapısıdır. Konvolüsyonel sinir ağları görüntü sınıflandırma, nesne tanımlama gibi görüntü tabanlı çalışmalarda çok başarılıdır. Evrimsel sinir ağları özellikle sınıflandırma problemlerinde kullanılan bir ağıdır ve eğitilebilen birçok katmandan oluşmaktadır. Birçok katmandan oluşan CNN'ler her katmanda incelenen probleme dair bir öznetelik öğrenilir ve öğrenilen bir öznetelikler sistemde sonraki gelen katmana çıktı olarak verilir.

4.2.6. Yinelenen Sinir Ağları (RNN)

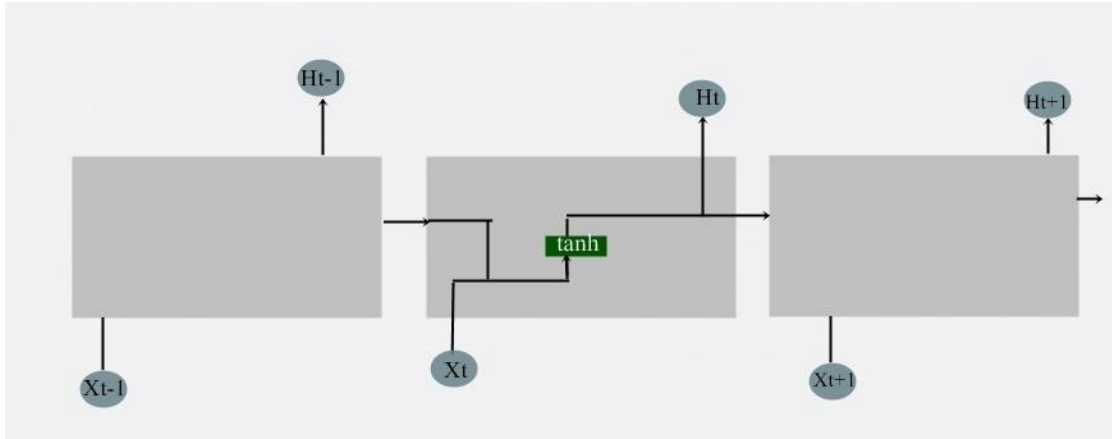
Yinelenen sinir ağları bu çalışma kapsamında da kullanılmış olan bir derin öğrenme mimarisidir. Bu ağlar verileri analiz etme kabiliyetine sahiptir. Hesaplamalar arasındaki bağ sistem boyunca saklandığından özellikle doğal dil işleme gibi çıktının önceki girdiye bağlı olduğu problemlerin çözümünde oldukça başarılı sonuçlar vermektedir. Yinelenen sinir ağları farklı problemlere uygulanmaktadır ve bu problemlerin çözülebilmesi için farklı varyasyonları geliştirilmiştir. Temel olarak RNN'ler tek yönlü ve çift yönlü RNN olarak ayrılmaktadır. Çift yönlü RNN'ler iki adet tek yönlü RNN'in birleştirilmesi olarak düşünülebilir.

Bu çalışma kapsamında metin sınıflandırmak için yinelenen sinir ağlarından biri olan LSTM kullanılmıştır.

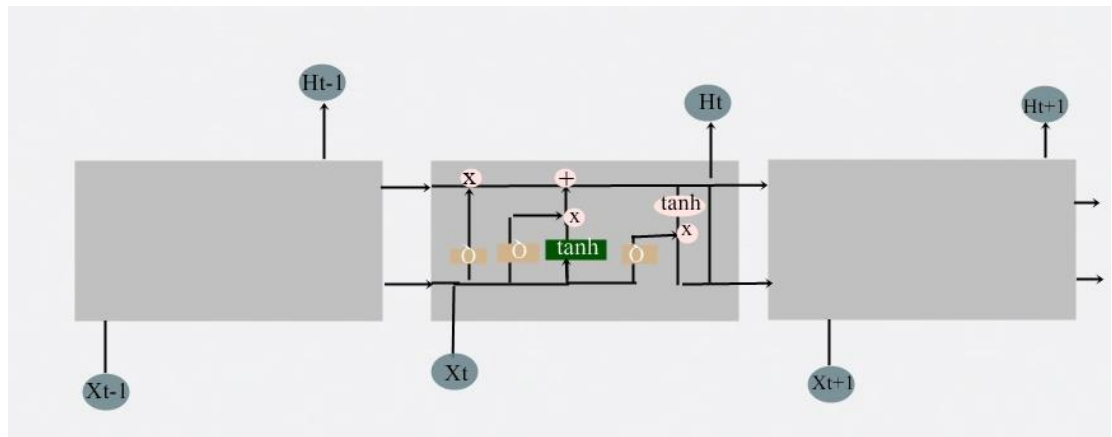
4.2.6.1. LSTM Ağları

LSTM ağları, 1997'de Hochreiter ve Schmidhuber tarafından RNN mimarisinin dezavantajlarını yok etmek için ortaya atılan sinir ağıdır [75]. LSTM ağları uzun süreli bağımlılıkları öğrenebilen özel bir RNN türüdür. Aynı zamanda RNN sinir ağlarının eğitilmesinde yaşanan sorunlar LSTM'de bütünüyle giderilmiştir. LSTM Gradyan İnişi (Gradient Descent) probleminin çözümünde de oldukça başarılı görülmektedir.

Tüm tekrarlayan sinir ağları, sinir ağının tekrarlayan modülleri zinciri şeklindedir. Standart bir RNN'de yinelenen modül tek bir katman içerir. LSTM'lerde ise tekrarlayan modül daha farklı bir yapıyla oluşturulmaktadır. LSTM ağları Tek bir sinir ağı katmanına sahip olmak yerine, birbiriyle etkileşimde olan dört katmana sahiptir [76]. Şekil 4.2 ve 4.3'de katmanlar ve aralarındaki fark net olarak görülebilmektedir.



Şekil 4.2. Standart bir RNN'de katmanlar.



Şekil 4.3. LSTM katmanları.

Yukarıdaki şekilde, her çizgi bir düğümün çıkışından diğerlerinin girişlerine kadar bir vektörü taşır.

LSTM ağları, temel olarak uzun veya kısa zaman periyotlarını hatırlamaktadır. Bu yüzden, sistem çalıştığı süre boyunca değerler her yinelemede değiştirilmez ve her bir iterasyonda edinilen bilgiler saklanmış olur. Bu sayede geri yayımlı bir eğitim gerçekleştiğinde elde edilen verilerde kayıp yaşanmamaktadır. LSTM blokları incelendiğinde, bilgi akışını kontrol eden dört adet “kapı” içerdiği görülmektedir. Bu kapıların, değerlerinin hesaplanma işlemleri Denklem (4.1)-(4.4) kullanılarak yapılmaktadır. Bir bilginin belleğe girip girmeyeceğine ya da çıkış yapmasına izin verilmesi elde edilen bu değerlere göre sağlanmaktadır. Giriş kapısı bir verinin akışının başladığı yerdir. Unut kapısı ise bu verinin bellekte hangi derecede kalacağına karar verildiği kapıdır. Çıktı kapısı ise verinin bellekte ne kadar kullanıldığının kontrol edildiği kapıdır.

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci}c_{t-1} + b_i) \quad (4.1)$$

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf}c_{t-1} + b_f) \quad (4.2)$$

$$c_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (4.3)$$

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_{t-1} + 0) \quad (4.4)$$

Burada i, o, f ve c terimleri sırasıyla giriş kapısı, çıkış kapısı, unut kapısı ve hafıza hücrelerini ifade etmektedir.

Yukarıda da anlatıldığı gibi LSTM ağları uzun vadeli bağlılık sorunlarını çözmek için oldukça uygundur. Bu çalışmada haber metinleri sınıflandırmaya çalışılmıştır ve çok-tek ilişki kurmayı gerektirdiğinden, LSTM çalışma için uygun görülmüştür.

5. UZUN KISA SÜRELİ BELLEK AĞLARI İLE ÖZET METNİN SINIFLANDIRILMASI

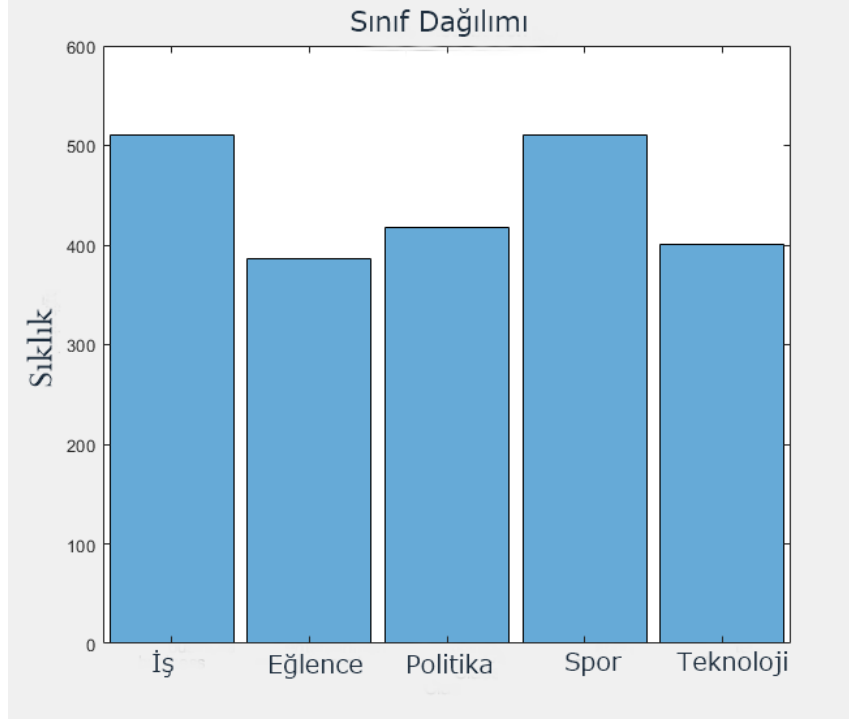
5.1. ÖNİŞLEME ADIMLARI

LSTM ağına metni girdi olarak verebilmek için ilk olarak çeşitli aşamalardan geçirilmiştir. Metinde büyük/küçük harf dönüşümü sağlanmıştır, metin verilerinin sayısal dizilere yani vektörlere dönüştürülmesi gerçekleştirilmiştir. Ayrıca metin özetlemede de uygulanmış olan durma kelimelerinin kaldırılması işlemi bu aşamada da yapılmıştır. Metinleri sayısal indeks dizileriyle eşleyen bir kelime kodlaması kullanılmıştır.

5.2. MODELİN TASARLANMASI VE EĞİTİLMESİ

LSTM ağları kullanılarak gerçekleştirilen sınıflandırma çalışma temelde üç farklı veri setiyle test edilmiştir. İlk veri seti çalışmanın ilk kısmında da kullandığımız BBC News haber metinleridir. İkinci ve üçüncü veri setleri ise sırasıyla BBC News veri setinin GWO ile özetlenmesi sonucu elde edilen metinler ve K-means ile özetlenmesi sonucu elde edilen metinlerdir. Bu üç veri seti ayrı ayrı LSTM ağları ile sınıflandırılmış ve sonuçlar karşılaştırılmıştır. Çalışmada gerçekleştirilen uygulamalar aşağıda sırasıyla incelenmiştir.

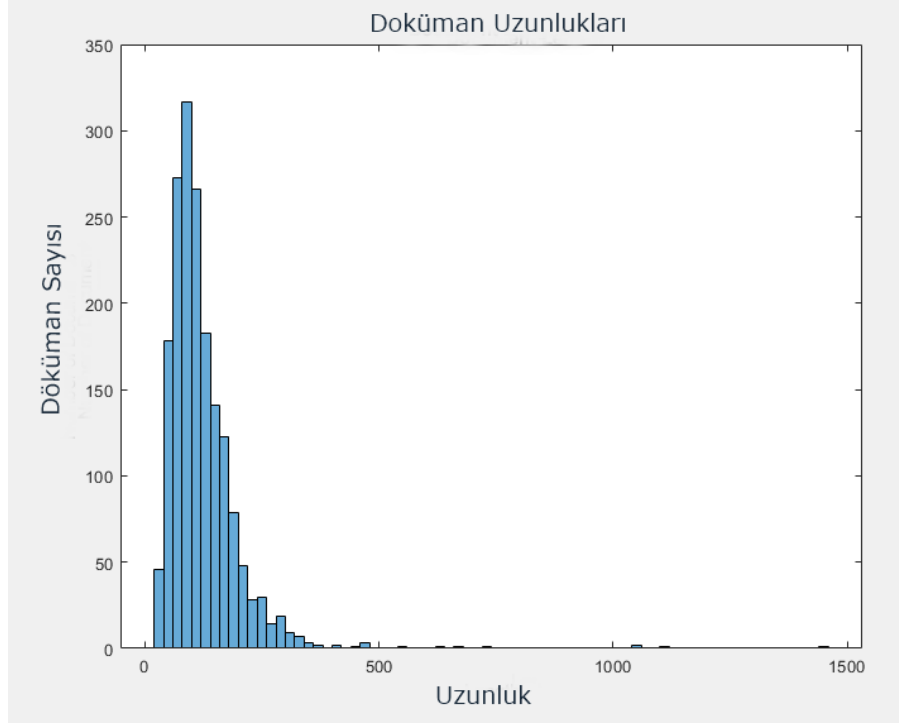
Bu üç veri setinde de her metnin ait olduğu sınıf bilgisi bulunmaktadır. Veri setlerinde yer alan kategoriler ve sayıları Şekil 5.1’de görülebilmektedir. Veri setinde 5 adet haber kategorisi yer almaktadır. Bu kategoriler: iş, eğlence, politika, spor ve eğitimidir.



Şekil 5.1. BBC News kategorileri ve dağılımları.

5.2.1. Metinleri Dizilere Dönüştürme

Metinlerin LSTM ağına girebilmesi için, sayısal indeks dizilerine dönüştürülmesi gerekmiştir. Ayrıca metinler aynı uzunlukta olacak şekilde ayarlanmalıdır. Metin uzunluklarını ayarlamak için önce bir hedef uzunluk seçilmiştir ve daha sonra ondan daha uzun olan belgeler kısaltılmıştır. Uygun bir hedef uzunluk bulmak için eğitim setindeki metinlerin uzunluklarının histogramına bakılmıştır. Şekil 5.2’de BBC News veri setinin özetsiz halinin histogram grafiği yer almaktadır.



Şekil 5.2. BBC News kelime uzunluğu histogramı.

Yukarıdaki şekilde de görüldüğü gibi veri setindeki metinlerin uzunluğu 250 ile 300 kelime arasında yoğunlaşmıştır. Bu yüzden kelime uzunluğu 300 olarak seçilmiştir.

5.2.2. LSTM Ağının Oluşturması ve Eğitimi

Veri setlerinde 2.225 haber makalesi bulunmaktadır. Bu verilerden eğitim için % 80, doğrulama için ise %20'lik kısmı kullanılmıştır. LSTM ağının içerdiği katmanlar ve boyutları aşağıdaki gibidir:

- Kelime Gömme Katmanı (Word Embedding Layer): 50 boyutludur ve 423 benzersiz kelimeli içerir. Kelime gömme katmanı, sözcük başına bir vektör saklar. Kelime indeks sekanslarını vektör sekanslarına dönüştürür. Eğitimden sonra, benzer anlamları olan kelimeler genellikle benzer vektörlere sahiptir.
- LSTM: 80 gizli üniteye sahip LSTM katmanıdır. Uzun vadeli bağımlılıkları öğrenmek için kullanılır.
- Tam Bağlı Katman(Fully Connected Layer): Sınıf sayısı, softmax katmanı ve sınıflandırma katmanı ile aynı boyutta tam bağlantılı bir katmandır.
- Softmax: Softmax fonksiyonu çıkışta her bir girdinin bir sınıfa ait olma durumunu $[0,1]$ arasında çıktı üreterek göstermektedir.

LSTM ađının eđitilmesinde kullanılan parametreler izelge 5.1’de verilmiřtir.

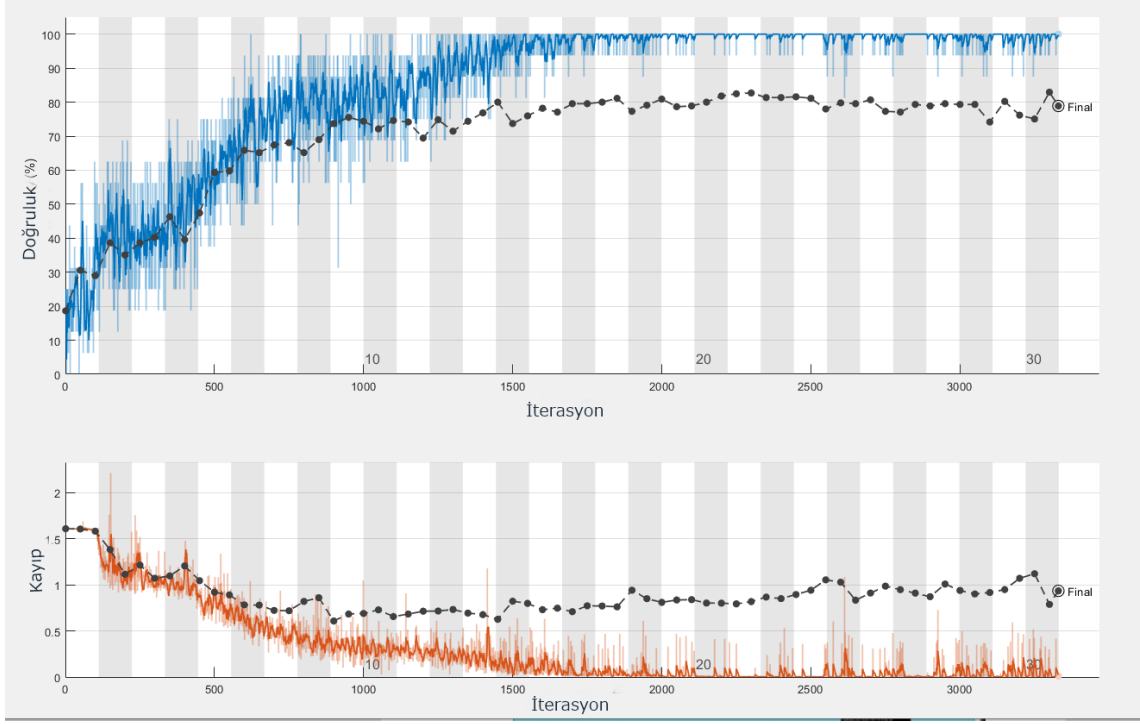
izelge 5.1. LSTM ađında kullanılan parametreler.

	Deđeri
Eđitim Verisi	%80
Test Verisi	%20
Mini-batch	16
İterasyon Sayısı	3330
Eđitim Sayısı	30
Giriř	1
ıkıř	1
Katman	4

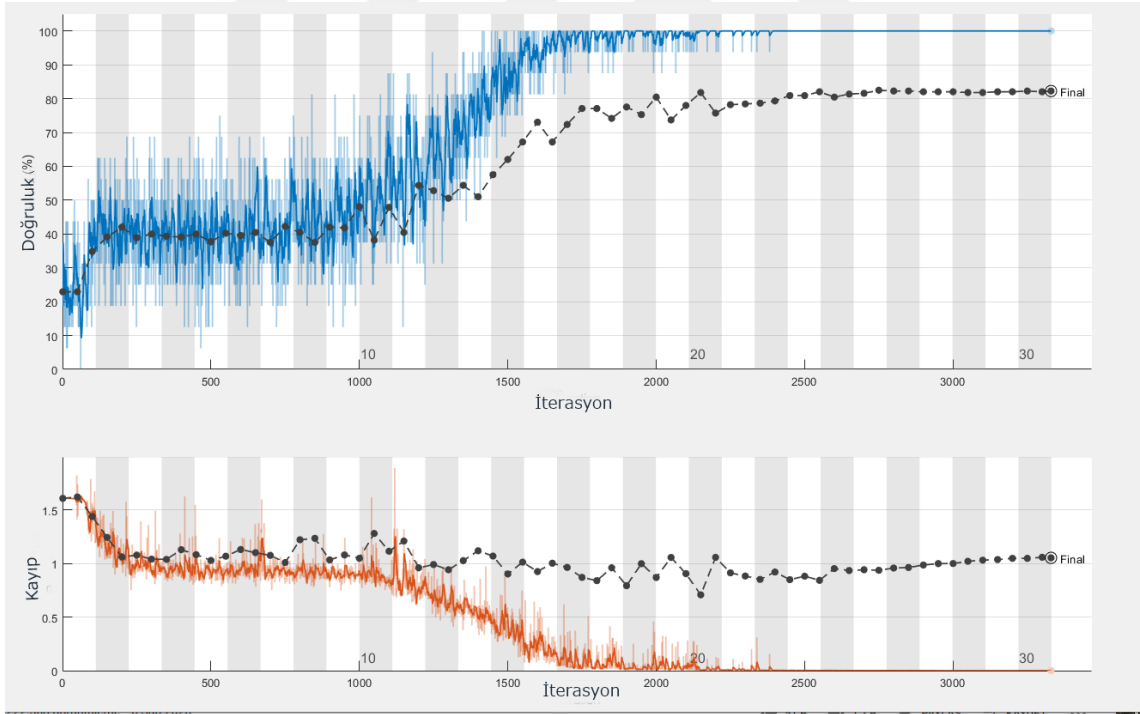
Tabloda yer alan mini-batch boyutu 16 olarak belirlenmiřtir. Bir ađ modelinin aynı anda iřleyebileceđini veri sayısı bu parametreyle belirlenmektedir. Bu veri setinin paralar halinde iřlenmesi anlamına gelmektedir. Bir derin renme sisteminde verilerin hepsinin aynı anda eđitime katılması mmkn deđildir. Bu yzden veri seti eđitime paralar halinde katılmaktadır. Her bir para ayrı ayrı eđitilmektedir ve sistemin bařarısı bu eđitim sonularının geri yayılımlarla gncellenmesi sonucu ortaya ıkmaktadır. Diđer bir parametre olan eđitim sayısı (epoch) 30 olarak kullanılmıřtır. Sistemdeki her bir eđitim adımı “epoch” olarak adlandırılmaktadır.

Oluřturulan sistemde renme hızı 0.001 olarak belirlenmiřtir. renme oranı bařlangıta 0.01 iken, eđitim sayısı 10 olduktan sonra 0.001 olarak gncellenmiřtir.

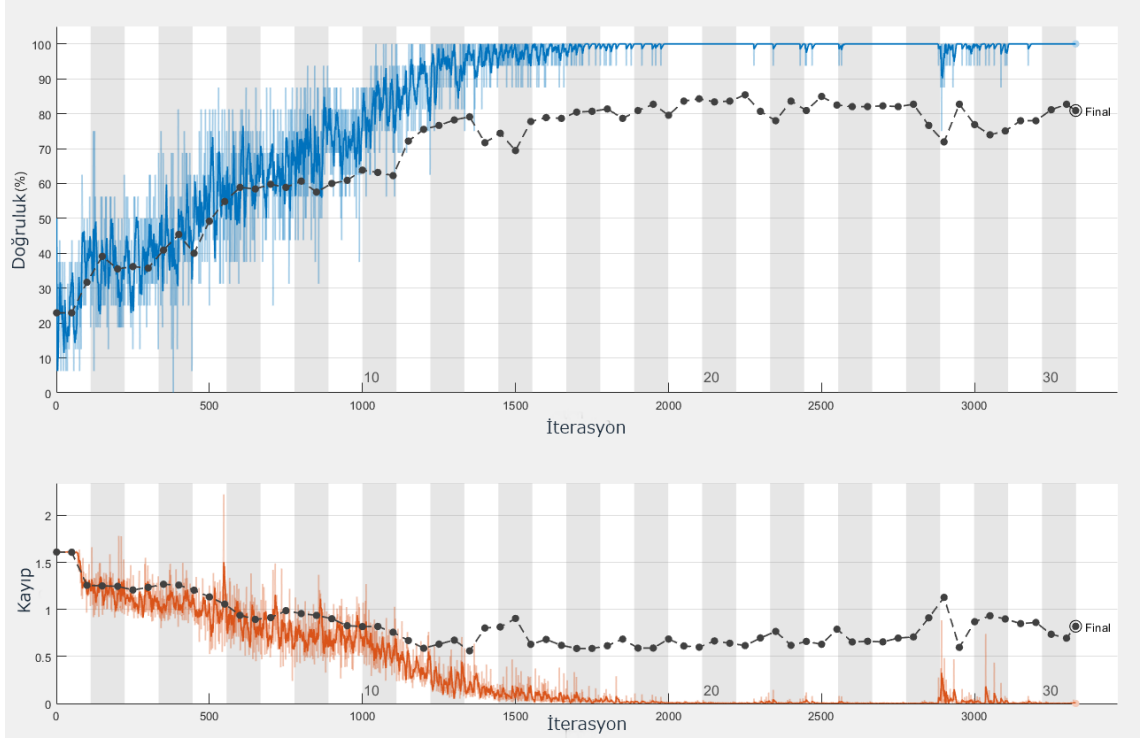
LSTM ađının  veri seti ile olan renme sırasında ortaya ıkan grafik Őekil 5.3,5.4 ve 5.5’de verilmiřtir. İterasyon arttıca dođruluk oranının arttıđı ve kayıp oranının azaldıđı grlmektedir.



Şekil 5.3. Özetsiz veri setine LSTM uygulanması.



Şekil 5.4. GWO özetlerine LSTM uygulanması.



Şekil 5.5. K-means özetlerine LSTM uygulanması.

Yukarıdaki üç şekil incelendiğinde GWO ile çıkarılan özet dokümanların LSTM ağları ile sınıflandırma başarısının K-means ile çıkarılan özetlerden daha iyi olduğu görülmektedir. Çizelge 5.2’de gerçekleştirilen bu üç testin başarı oranları gösterilmiştir.

Çizelge 5.2. LSTM ile sınıflandırma başarı yüzdesi.

	Başarı Yüzdesi
BBC News Özetsiz Veri Seti	%74.17
GWO Özet Veri Seti	%82.14
K-means Özet Veri Seti	%76.22

6. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında Doğal Dil İşleme uygulamalarında oldukça yaygın olarak kullanılan ve günümüzde veri boyutunun artmasıyla birlikte giderek önemli bir ihtiyaç haline gelen metin özetleme işlemine çözüm sağlamak için yeni bir çıkarımsal metin özetleme yöntemi geliştirilmiştir.

Oluşturduğumuz sistemde özetleme işlemi için sezgisel bir algoritma olan Gri Kurt Optimizasyon algoritması kullanılmıştır. GWO'nun kümeleme başarısı özete dahil edilecek olan cümlelerin kümeleneceği aşamasında test edilmiştir. Çıkarımsal metin özetleme işlemi Gri Kurt Optimizasyon Algoritması ve K-means kümeleme algoritması için ayrı ayrı uygulanmıştır. Bu iki özete benzerlik oranını ve çalışmanın başarısını değerlendirmek için Ngram birliktelik istatistiği (ROUGE) kullanılmıştır ve ROUGE-1 ve ROUGE-2 ölçümleri ayrı ayrı uygulanmıştır.

Metin özetleme sisteminin sonuçları incelendiğinde GWO'nun yaygın kullanılan bir kümeleme algoritması olan K-meansten daha iyi sonuçlar verdiği gözlemlenmiştir. ROUGE-2 değerleri ROUGE-1 değerlerine göre daha kötü çıkmıştır fakat bunun nedeni ROUGE-2'nin bigramları, ROUGE-1'in ise unigramları karşılaştırması dolayısıyla kelime sıralamasını değerlendirmeye katmamasıdır. GWO ROUGE-2 değerlerine göre K-means'ten %6.38 oranında daha iyi sonuçlar vermiştir. Geçtiğimiz yıllarda öne çıkan sezgisel algoritma GWO'nun bu çalışma ile kümelemede iyi bir başarıyı gösterdiğini ve çıkarıma dayalı metin özetleme için iyi bir araç olduğu gösterilmiştir.

Çalışmanın devamında hem metin özetleme işleminin metinlerin sınıflandırma başarısına etkisini gözlemleyebilmek hem de GWO algoritması ile özete çıkarılmış metinlerin sınıflandırma işlemlerinde nasıl sonuçlar verdiğini tespit etmek amaçlanmıştır.

Bu kapsamda 2225 makale ve 5 makale kategorisinden oluşan BBC News veri seti, BBC News veri setinin GWO ile elde edilmiş çıkarımsal metin özetleri ve aynı veri setinin K-means ile oluşturulmuş çıkarımsal metin özetleri kurmuş olduğumuz LSTM ağ sistemine girdi olarak verilmiştir.

Bu üç veri setinin de LSTM ağları kullanılarak oluşturulmuş sistemde iyi bir

sınıflandırma başarısı verdiği gözlemlenmiştir. Deneysel sonuçlarında daha ayrıntılı incelemeye gidildiğinde GWO'nun en iyi sınıflandırma başarısı verdiği görülmüştür. İkinci en başarılı sistem, K-means veri setidir ve ardından özet çıkarılmamış ham veri seti ile çalışılan sistem gelmektedir. GWO özetleri LSTM ağları tarafından K-means'e göre %5.92 oranında daha başarılı sonuçlar vermiştir.

Bu tez kapsamında yapılan tüm çalışmalar gözden geçirildiğinde GWO'nun çıkarımsal metin özetleme için başarılı bir yaklaşım olduğu sonucuna varılmıştır. GWO tarafından özetlenen metinlerin sınıflandırma işleminde başarı artışına neden olması GWO tarafından oluşturulan özetlerin anlamlı ve bütünü temsil eden verimli özetler olduğunu göstermektedir. LSTM ağlarıyla eğitilen özetlerin özet çıkarılmamış ham metne göre daha iyi sınıflandırma başarısı vermesi ise metin özetleme işlemlerinin sınıflandırma işlemlerinde de etkin bir role sahip olduğu kanıtlanmıştır.

Bu çalışmada elde edilen özetleme başarılarının, cümle puanlama sistemindeki ağırlıkların iyileştirilmesi ve / veya artırılması, ideal özet olarak adlandırılan insan gözüyle oluşturulmuş özetlerin daha objektif bir ortamda oluşturulması ve kelimelerin alan bilgileri dikkate katılarak alan bazlı kelime değerlendirmelerinin yapılmasıyla artırılabilirliği düşünülmektedir. Çalışmada yer alan cümle konumu parametresinin cümlenin sonuna olan yakınlığı açısından da değerlendirilmesi bundan sonraki çalışmalarda test edilecektir. İleride yapılacak olan çalışmalarda GWO'nun çoklu metin özetleme başarısının test edilmesi amaçlanmaktadır. Çalışmanın başarısının Hiper Parametre Seçim Yöntemleri ile artırılabilirliği düşünülmektedir. Ayrıca derin öğrenme aşamasında ezberlemeden önce durdurulmanın sağlanması için Erken Durdurma (Early Stopping) bundan sonraki çalışmalarda uygulanabilir.

7. KAYNAKLAR

- [1] T. M. T. Sembok, "Character strings to natural language processing in information retrieval," *Lecture Notes in Computer Science*, Berlin, Almanya, 2003, ss. 26–33.
- [2] E. Hovy ve C. Y. Lin, "Automated text summarization and the SUMMARIST system," *Tipster Text Program Phase III: Proceedings of a Workshop*, Baltimore, USA, 1998, ss. 197–214.
- [3] V. Gupta ve G.S. Lehal, "A survey of text summarization extractive techniques," *Journal of Emerging Technologies in Web Intelligence*, c. 2, sayı 3, 2010.
- [4] P. Bhatia, "A survey to automatic summarization techniques," *International Journal of Engineering Research and General Science*, c. 3, sayı 5, 2015.
- [5] S. A. Babar ve P. D. Patil, "Improving performance of text summarization," *International Conference on Information and Communication Technologies*, 2014.
- [6] R. Ferreira ve diğ. "Assessing sentence scoring techniques for extractive text summarization," *Expert Systems with Applications*, c. 40, sayı 14, ss. 5755–5764, 2013.
- [7] R. Shardan ve U. Kulkarni, "Implementation and evaluation of evolutionary connectionist approaches to automated text summarization," *Journal of Computer Science*, c. 6, sayı 11, ss. 1366–1376, 2010.
- [8] H. R. Luhn, "The automatic creation of literature abstracts," *IBM Journal of Research Development*, c. 2, sayı 2, ss.159–165, 1958.
- [9] H. P. Edmundson, "New methods in automatic extracting," *Journal of the Association for Computing Machinery*, c. 16, sayı 2, ss. 264–285, 1969.
- [10] J. J. Pollock ve A. Zamora, "Automatic abstracting research at chemical abstracts," *Journal of Chemical Information and Computer Sciences*, c. 15, sayı 4, ss. 226–232, 1975.
- [11] R. Brandow, K. Mitze ve L. F. Rau, "Automatic condensation of electronic publications by sentence selection," *Information Processing Management*, c. 31, sayı 5, ss. 675–685, 1995.
- [12] J. Kupiec, O. P. Jan ve C. Francine, "A trainable document summarizer," *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, Washington, USA, 1995, ss. 68–73.
- [13] B. Baldwin ve T. S. Morton, "Dynamic coreference-based summarization," *Proceedings of the 3rd Conference on Empirical Methods in Natural Language Processing*, Granada, Spain, 1998, ss. 1–6.
- [14] M. Y. Kan ve K. McKeown, "Information extraction and summarization: Domain independence through focus types," Columbia University, New York, USA, Rep.

CUCS-030-99, 1998.

- [15] R. Radev, S. Blair-Goldensohn, ve Z. Zhang, “Experiments in single and multi-document summarization using MEAD,” *First Document Understanding Conference*, New Orleans, USA, 2001.
- [16] V. Gupta, “Hybrid algorithm for multilingual summarization of Hindi and Punjabi documents,” *Mining Intelligence and Knowledge Exploration*, 2013, ss. 717–727.
- [17] K. Nandhini ve S. R. Balasundaram, “Improving readability through extractive summarization for learners with reading difficulties,” *Egyptian Informatics Journal*, c. 14, sayı 3, ss. 195–204, 2013.
- [18] R. Mihalcea ve P. A. Tarau, “Language independent algorithm for single and multiple document summarization,” *Int'l Joint Conf. on Natural Language Processing*, 2005, ss. 19–24.
- [19] D. Parveen, H. M. Ramsel ve M. Strube, “Topical coherence for graph-based extractive summarization,” *Proceedings of the 2015 conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 2015, ss. 1949–1954.
- [20] T. Copeck, S. Szpakowicz ve N. Japkowicz, “Learning how best to summarize,” *Proceedings of the Workshop on Multi-Document Summarization Evaluation of the 2nd Document Understanding Conference at the 40th Meeting of the Association for Computational Linguistics*, Philadelphia, USA, 2002.
- [21] K. F. Wong, M. Wu ve W. Li, “Extractive summarization using supervised and semi-supervised learning,” *Proceedings of the 22nd International Conference on Computational Linguistics*, Manchester, UK, 2008, ss. 985–992.
- [22] P. Lal ve S. Rueger, “Extract-based summarization with simplification,” *Proceedings to Association for Computational Linguistics*, 2002.
- [23] T. Hirao, Y. Sasaki, and H. Isozaki, “NTT's text summarization system for DUC-2002,” *Document Understanding Conference*, 2002.
- [24] M. Karamuftuoglu, “An approach to summarizaion based on lexical bonds,” *40th Annual Meeting of the Association for Computational Linguistics*, 2002, ss. 86–89.
- [25] C. N. Silla, G. L. Pappa, A. A. Freitas ve C. A. A. Kaestner, “Automatic text summarization with genetic algorithm-based attribute selection,” *9th IberoAmerican Conference on Artificial Intelligence*, 2004, ss. 305–314.
- [26] L. A. Leiva, “Responsive text summarization,” *Information Processing Letters*, c. 130, ss. 52–57, 2018.
- [27] M. N. Azadani, N. Ghadiri ve E. Davoodijam, “Biomedical text summarization: an itemset mining and sentence clustering approach,” *Journal of Biomedical Informatics*, c. 84, ss. 42–58, 2018.
- [28] R. A. G Hernandez ve Y. Ledeneva, “Word sequence models for single text summarization,” *Proceeding of Second International Conference on Advances in Computer-Human Interactions*, Mexico, 2009, ss. 44–48.
- [29] O. Rouane, H. Belhadeif ve M. Bouakkaz, “Combine clustering and frequent itemsets mining to enhance biomedical text summarization,” *Expert Systems with*

- Applications*, c. 135, ss. 362–373, 2019.
- [30] E. Filatova ve V. Hatzivassiloglou, “Event-based extractive summarization,” *Proceedings of ACL 2004 Workshop on Summarization*, Barcelona, Spain, 2004.
- [31] J. Y. Yeh, H. R. Ke, W. P. Yang ve I. H. Meng, “Text summarization using a trainable summarizer and latent semantic analysis,” *Journal of Information Processing and Management*, c. 41, sayı 1, ss. 75–95, 2005.
- [32] A. Kiani ve M. R. Akbarzadeh, “Automatic text summarization using: hybrid fuzzy GA–GP,” *IEEE International Conference on Fuzzy Systems*, Vancouver, Canada, 2006, ss. 977–983.
- [33] R. McDonald, “A study of global inference algorithms in multidocument summarization,” *29th European Conference on IR Research*, Rome, Italy, 2007, ss. 557–564.
- [34] M. Berker ve T. Güngör, “Using genetic algorithms with lexical chains for automatic text summarization,” *4th International Conference on Agents and Artificial Intelligence*, Vilamoura, Portugal, 2012, ss. 595–600.
- [35] V. S. Raj Kumar ve D. Chandrakala, “A survey on text summarization using optimization algorithm,” *Elk Asia Pasific Journal Of Computer Science and Information Systems*, c. 2, sayı 1, ss. 2454–3047.
- [36] L. Suanmali, N. Salim ve M. S. Binwahlan, “Fuzzy logic based method for improving text summarization,” *International Journal of Computer Science and Information Security*, c. 2, sayı 1, ss. 65–70, 2009.
- [37] F. Kyoomarsi, H. Khosravi, E. Eslami ve P. K. Dehkordy, “Optimizing machine learning approach based on fuzzy logic in text summarization,” *International Journal of Hybrid Information Technology*, c. 2, sayı 2, ss. 105–116, 2009.
- [38] M. Moradi, M. Dashti ve M. Samwald, “Summarization of biomedical articles using domain-specific word embeddings and graph ranking,” *Journal of Biomedical Informatics*, c. 107, ss. 103452, 2020.
- [39] K. Svore, L. Vanderwende ve C. Burges, “Enhancing single document summarization by combining RankNet and third-party sources,” *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Prague, Czech Republic, 2007, ss. 448–457.
- [40] M. Yousefi-Azar ve L. Hamey, “Text summarization using unsupervised deep learning,” *Expert Systems with Applications*, c. 68, ss. 93–105, 2017.
- [41] A. Sinha, A. Yadav ve A. Gahlot, (2020, Şubat, 20), *Extractive text summarization using neural networks*, (Online). Erişim: <https://deepai.org/publication/extractive-text-summarization-using-neural-networks>
- [42] R. Bhargava, G. Sharma ve Y. Sharma, “Deep text summarization using generative adversarial networks in Indian languages,” *Procedia Computer Science*, c. 167, ss. 147–153, 2020.
- [43] S. Mirjalili, M.S. Mirjalili ve A. Lewis, “Grey wolf optimizer,” *Advances in Engineering Software*, c. 69, ss. 46–61, 2014.
- [44] S. Mirjalili, “How effective is the Grey Wolf optimizer in training multi-layer

- perceptrons,” *Applied Intelligence*, c. 43, sayı 1, ss. 150–161, 2015.
- [45] C. Muro, R. Escobedo, L. Spector ve R. Coppinger, “Wolf-pack (*Canis lupus*) hunting strategies emerge from simple rules in computational simulations,” *Behavioural processes*, c. 88, sayı 3, ss. 192–197, 2011.
- [46] M. Pradhan, P. K. Roy ve T. Pal, “Grey wolf optimization applied to economic load dispatch problems,” *International Journal of Electrical Power & Energy Systems*, c. 83, ss. 325–334, 2016.
- [47] A. M. Rao, K. Ramji ve T. N. Kumar, “Intelligent navigation of mobile robot using Grey Wolf Colony optimization,” *Materials Today: Proceedings*, c. 5, sayı 9, ss. 19116–19125, 2018.
- [48] D. Guha, P. K. Roy ve S. Banerjee, “Load frequency control of large scale power system using quasi-oppositional grey wolf optimization algorithm,” *Engineering Science and Technology, an International Journal*, c. 19, sayı 4, ss. 1693–1713, 2016.
- [49] S. Kapoor, I. Z. C. Singhal ve S. J. Nanda, “A Grey Wolf optimizer based automatic clustering algorithm for satellite image segmentation,” *Procedia Computer Science*, c. 115, ss. 415–422, 2017.
- [50] K. Shankar, S. K. Lakshmanaprabu, A. Khanna, S. Tanwar, J.J. Rodrigues ve N. R. Roy, “Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier,” *Computers & Electrical Engineering*, c. 77, ss. 230–243, 2019.
- [51] C. Lu, S. Xiao, X. Li ve L. Gao, “An effective multi-objective discrete grey wolf optimizer for a real-world scheduling problem in welding production,” *In Advances in Engineering Software*, c. 99, ss. 161–176, 2016.
- [52] N. Singh ve S. B. Singh, “Hybrid algorithm of Particle Swarm optimization and Grey Wolf optimizer for improving convergence performance,” *Journal of Applied Mathematics*, c. 2017, 2017.
- [53] P. Erdogmus, “A new solution approach for non-linear equation systems with Grey Wolf optimizer,” *Sakarya University Journal of Computer and Information Sciences*, c. 1. ss. 1-11, 2019.
- [54] Y. Zhang, P. Phillips, S. Wang, J. Yang ve J. Wu, “Fruit classification by biogeography-based optimization and feedforward neural network,” *Expert Systems*, c. 33, sayı 3, ss. 239–253, 2016.
- [55] A. Gotmare, S. S. Bhattacharjee, R. Patidar ve N. V. George, “Swarm and evolutionary computing algorithms for system identification and filter design: a comprehensive review,” *Swarm and Evolutionary Computation*, c. 32, ss. 68–84, 2017.
- [56] G. Sarıman, “Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: K-means ve K-medoids kümeleme algoritmalarının karşılaştırılması,” *Journal of Natural & Applied Sciences*, c. 15, sayı 3, ss.192–202, 2011.
- [57] P. Erdoğan, B. Çolak ve Z. Durdağ, “K-means algoritması ile otomatik kümeleme,” *El-Cezeri Journal of Science and Engineering*, c. 3, sayı 2, 2011.
- [58] M. Işık ve A. Y. Çamurcu, “K-means, K-medoids ve bulanık C-means algoritmalarının uygulamalı olarak performanslarının tespiti,” *İstanbul Ticaret*

Üniversitesi Fen Bilimleri Dergisi, c. 11, sayı 1, ss. 31–45, 2007.

- [59] D. Greene ve P. Cunningham. “Practical solutions to the problem of diagonal dominance in kernel document clustering,” *Proceedings of the 23rd International Conference on Machine Learning*, New York, USA, 2006, ss. 377–384.
- [60] C. Y. Lin, “ROUGE: a package for automatic evaluation of summaries,” *Proceedings of the Workshop on Text Summarization Branches Out*, Barcelona, Spain, 2004, ss. 74–81.
- [61] T. S. LEE, ve D. MUMFORD, “Hierarchical bayesian inference in the visual cortex,” *Journal of the Optical Society of America A*, c. 20, sayı 7, ss. 1434-1448, 2003.
- [62] D. H. Lee, “Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks,” *Workshop on Challenges in Representation Learning*, 2013.
- [63] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado ve J. Dean, “Distributed representations of words and phrases and their compositionality,” *Advances in Neural Information Processing Systems*, ss. 3111–3119, 2013.
- [64] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng ve C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” *Proceedings of Empirical Methods on Natural Language Processing*, c. 1631, ss. 1631–1642, 2013.
- [65] J. Devlin, R. Zbib, Z. Huang, T. Lamar, R. Schwartz ve J. Makhoul, “Fast and robust neural network joint models for statistical machine translation,” *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014, ss. 1370–1380.
- [66] Z. Li, M. Shang ve W. Yan, “News text classification model based on topic model,” *IEEE/ACIS International Conference on Computer & Information Science*, 2016, ss. 1–5.
- [67] K. S. Tai, R. Socher ve C. D. Manning, “Improved semantic representations from tree-structured long short-term memory networks”, *Association for Computational Linguistics*, Beijing, China, 2015, ss. 1556–1566.
- [68] Y. Kim, “Convolutional neural networks for sentence classification,” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014, ss. 1746–1751.
- [69] S. Lai, L. Xu, K. Liu ve J. Zhao, “Recurrent convolutional neural networks for text classification,” *Association for the Advancement of Artificial Intelligence Conference*, 2015, ss. 2267–2273.
- [70] X. Zhang, J. Zhao ve Y. Lecun, “Character-level convolutional networks for text classification,” *Advances in Neural Information Processing Systems*, 2015, ss. 649–657.
- [71] J. Y. Lee ve F. Deroncourt, “Sequential short-text classification with recurrent and convolutional neural networks,” *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, USA, 2016, ss. 515–520.
- [72] S. Fernández, A. Graves ve J. Schmidhuber, “An application of recurrent neural

networks to discriminative keyword spotting,” *International Conference on Artificial Neural Networks*, 2007, ss. 220–229.

- [73] J. Gao, L. Deng, M. Gamon ve X. He, “Modeling interestingness with deep neural networks,” USA, Patent 9 846 836, 19 Aralık, 2017.
- [74] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo ve G.Z. Yang, “Deep learning for health informatics,” *IEEE journal of Biomedical and Health Informatics* c. 21, sayı 1, ss. 4–21, 2017.
- [75] S. Hochreiter, ve J. Schmidhuber, “Long Short-term memory,” *Neural Computation*, c. 9, sayı 8, ss. 1735–1780.
- [76] F. Gers, F. Cummins, S. Fernandez, J. Bayer, D. Wierstra, J. Togelius, F. Gomez, M. Gagliolo ve A. Graves, (2020, Şubat 10), *Understanding LSTM Networks*, [Online]. Erişim: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.



ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı :Ebru DUDAK
Doğum Tarihi ve Yeri :30 Mayıs 1994 Düzce
Yabancı Dili :İngilizce
E-posta :ebrududak@hotmail.com

ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Y. Lisans	Elektrik-Bilgisayar Ortak Bölüm	Düzce Üniversitesi	2020
Lisans	Bilgisayar Müh.	Düzce Üniversitesi	2018
Lise		Düzce Anadolu Öğretmen Lisesi	2011

YAYINLAR