KOCAELİ ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

ARAŞTIRMA PROBLEMLERİ

SORU DOKÜMANLARININ ANLAMSAL BENZERLİKLERİNE DAYALI DERİN ÖĞRENME TABANLI KÜMELEME ANALİZİ

ERAY YELMEN

KOCAELİ 2020

KOCAELİ ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

ARAŞTIRMA PROBLEMLERİ

SORU DOKÜMANLARININ ANLAMSAL BENZERLİKLERİNE DAYALI DERİN ÖĞRENME TABANLI KÜMELEME ANALİZİ

ERAY YELMEN

Dr. Öğr. Üyesi Alpaslan Burak İNNER		
Danışman,	**Kocaeli Üniversitesi**	**.....**
Prof. Dr. Nevcihan DURU		
Jüri Üyesi,	**Kocaeli Üniversitesi**	**.....**
Dr.Öğr. Üyesi Ersin KAYA		
Jüri Üyesi,	**Konya Teknik Üniversitesi**	**.....**

Tezin Savunulduğu Tarih: 17.08.2020

ÖNSÖZ VE TEŞEKKÜR

Bu tez çalışması kapsamında, x-band radarlar tarafından tespit edilen deniz hedeflerinin otomatik olarak sınıflandırılması için özgün bir füzyon yöntemi geliştirilmiştir. Önerilen yöntem sayesinde başarım artışı ve otomatik sınıflandırma çalışması için SOA yaklaşımının kullanımı irdelenmiştir.
Tez çalışmamda desteğini esirgemeyen, çalışmalarıma yön veren, bana güvenen ve yüreklendiren danışmanım Prof. Dr. Nevcihan DURU"ya sonsuz teşekkürlerimi sunarım.
Akademik çalışmalarım sırasında, birçok aşamada beni destekleyen Bilgisayar Mühendisliği Bölümü araştırma görevlilerine ve Fen Bilimleri Enstitüsü çalışanlarına teşekkür ediyorum.
Doktora öğrenimim boyunca desteklerini esirgemeyen başta Osman KARABAYIR ve Mehmet Zahid KARTAL olmak üzere çok sevgili TÜBİTAK BİLGEM RAPSİM çalışanlarına sonsuz teşekkürler sunarım.
Hayatım boyunca bana güç veren en büyük destekçilerim, her aşamada sıkıntılarımı ve mutluluklarımı paylaşan sevgili babam Fahrettin BATI, annem Hacer BATI, eşim Özge BATI ve kardeşlerim Mehmet Emre BATI ile Fatma BATI"ya sonsuz teşekkürlerimi sunarım.
Son olarak, gelecekte kendilerine ilham kaynağı olmasını ümit ederek bu tezi çocuklarım Elif Nil BATI ve Ege BATI"ya ithaf ediyorum.

Temmuz − 2020 Eray YELMEN

Bu dokümandaki tüm bilgiler, etik ve akademik kurallar çerçevesinde elde edilip sunulmuştur.
Ayrıca yine bu kurallar çerçevesinde kendime ait olmayan ve kendimin üretmediği ve başka
kaynaklardan elde edilen bilgiler ve materyaller (text, resim, şekil, tablo vb.) gerekli şekilde
referans edilmiş ve dokümanda belirtilmiştir.
Öğrenci No: 150201387
Adı Soyadı: Eray YELMEN
İmza:……..

İÇİNDEKİLER

ŞEKİLLER DİZİNİ

TABLOLAR DİZİNİ

SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar

AUC : Area Under the Curve (Eğri Altında Kalan Alan)
BT : Bilgisayarlı Tomografi
CPTAC : Clinical Proteomic Tumor Analysis Consortium (Klinik Proteomik Tümör Analiz Konsorsiyumu)
DICOM : Digital Imaging and Communications in Medicine (Tıpta Dijital Görüntüleme ve İletişim
DMI : Depth of Myometrial Invasion (Miyometriyal invazyon derinliği)
EFB : Exclusive Feature Bundling (Özel Değişken Paketi)
GBM : Gradient Boosting Machines (Gradyan Artırma Makineleri)
GLCM : Grey Level Co-occurence Matrix (Gri Seviye Eş Oluşum Matrisi)
GLDM : Grey Level Dependence Matrix (Gri Seviye Bağımlılık Matrisi)
GLRLM : Grey Level Run Length Matrix (Gri Seviye Dizi Uzunluğu Matrisi)
GLSZM : Grey Level Size Zone Matrix (Gri Seviye Boyutu Bölge Matrisi)
KNN : K-Nearest Neighbours (K-En Yakın Komşu)
MCC : Matthews Correlation Coefficient (Matthews Korelasyon Katsayısı)
MLP : Multi Layer Perceptrons (Çok Katmanlı Algılayıcılar)
MRI : Magnetic Resonance Imaging (Manyetik Rezonans Görüntüleme)
MRMR : Minimum Redundancy Maximum Relevence (Minimum Fazlalık Maksimum Alaka)
NGTDM : Neighborhood Grey Tone Difference Matrix (Komşuluk Gri Ton Fark Matrisi)
PET : Positron Emission Tomography (Pozitron Emisyon Tomografi)
ROC : Receiver Operating Characteristic (Alıcı İşletim Karakteristiği)
ROI : Region of Interest (İlgili Bölge)
SBS : Step Backward Selection (Geri Yönlü Arama Seçimi)
SFS : Step Forward Selection (İleri Yönlü Arama Seçimi)
SVM : Support Vector Machines (Destek Vektör Makineleri) TCGA : The Cancer Genome Atlas (Kanser Genom Atlası)
TCIA : The Cancer Imaging Archive (Kanser Görüntüleme Arşivi)
UCEC : Uterine Corpus Endometrial Carcinoma (Rahim Yapısı Endometrial Karsinom)

SORU DOKÜMANLARININ ANLAMSAL BENZERLİKLERİNE DAYALI DERİN ÖĞRENME TABANLI KÜMELEME ANALİZİ

ÖZET

İnternet ortamında metinsel dokümanların miktarının büyük boyutlara ulaşması ile birlikte aranan doğru dokümana kolay ve hızlı bir şekilde ulaşmak zorlaşmıştır. Metin dokümanlarının benzerliklerine göre kümelenmesi manuel yöntemlerle oldukça zahmetlidir. Bu durumu otomatik hale getirerek kolaylaştırmak için gelişmiş yöntemlere ihtiyaç vardır. Belge kümelemede metin verileri yakınlık ve benzerlik ölçüsüne göre gruplandırılır. Kümelemede yüksek başarı elde etmek, belgelerin doğru bir şekilde keşfedilmesi için oldukça önemlidir.

Anahtar Kelimeler: Bulanık C-means, Derin Öğrenme, Doc2Vec, Doğal Dil İşleme, Doküman Kümeleme.