

# Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi

İbrahim Onuralp Yiğit

Türk Telekom, İstanbul, Türkiye  
ibrahimonuralp.yigit@turktelekom.com.tr

**Özet.** Çağrı merkezlerine gelen çağrıların sesden metne dönüştürülerek kayıt altına alınabilmesi, görüşme metinleri üzerinde metin madenciliği tekniklerinin uygulanmasına imkan vermektedir. Bu çalışmada, sesden metne dönüştürülmüş görüşme kayıtlarından görüşmenin içeriğini duygu yönünden (olumlu/olumsuz) değerlendirebilen, müşteri memnuniyetini ve müşteri temsilcisi performansını öngörebilen, konu ve kelime ağaçlarını otomatik olarak çıkartabilen yeniden kullanılabilir bir yazılım çerçevesi anlatılacaktır. Bu çalışma sonucunda geliştirilen yazılım çerçevesinin Türk Telekom bünyesindeki çağrı merkezi uygulamalarında kullanılması hedeflenmektedir.

**Anahtar Kelimeler:** Yazılım Çerçevesi, Yazılım Bileşenleri, Yeniden Kullanım, Metin Madenciliği, Makine Öğrenmesi, Duygu Analizi, Konu Modelleme, Bilgi Çıkarımı.

## Call Center Text Mining Framework

İbrahim Onuralp Yiğit

Turk Telekom, Istanbul, Turkey  
ibrahimonuralp.yigit@turktelekom.com.tr

**Abstract.** In these days, the ability to convert call records from voice to text allows the application of text mining techniques on the call center text data. This study proposes a reusable software framework that automatically extracts subject and vocabulary from the conversation records that have been translated into sentences and whose contents are evaluated emotionally (positive / negative), customer satisfaction and customer representative performance. In consequence of this study, it is aimed to use the developed software framework for call center applications within Türk Telekom.

**Keywords:** Software Framework, Software Components, Software Reuse, Text Mining, Machine Learning, Sentiment Analysis, Topic Modeling, Information Retrieval.

## 1 Giriş

Şirketler, çağrı merkezlerine gelen çağrılar sayesinde müşterilerin şikâyetlerini, görüşlerini ve isteklerini öğrenebilmekte ve müşterilerden gelen geri bildirimlere göre ürünlerini ve/veya hizmetlerini iyileştirebilmektedirler. Bundan dolayı çağrı merkezleri şirketler için en önemli iletişim kanallarının başında gelmektedir.

Şirketler müşterilerin geri bildirimlerine ulaşmak için gelen çağrıları incelemesi gerekmektedir. Çağrı merkezlerine gelen çağrılar gün içerisinde yüz binleri bulmaktadır. Mevcut durumda çağrı merkezi yetkililerinin tek tek tüm çağrıları incelemesi mümkün olmadığı için büyük bir çağrı kümesi rastgele çağrılar seçilmektedir. Bundan dolayı şirketlerin tam anlamıyla müşteri geri bildirimlere ulaştıklarını söylemek oldukça güçtür.

Günümüzde, sesten metne dönüşüm teknolojileri giderek daha başarılı bir hale gelmesi ve çağrı merkezlerine gelen çağrıların sesten metne dönüştürülüp saklanabilmesi müşterilerin yaşadıkları sorunlarla ilgili bilgilere erişilebilmesini kolaylaştırmaktadır. Bu çalışma kapsamında müşteri memnuniyeti ve hizmet kalitesinin artırılması için metin madenciliği teknikleri kullanılarak Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi geliştirilmiştir. Geliştirilen yazılım çerçevesi kullanılarak çağrı merkezi görüşme kaydı metinleri üzerinden görüşmenin içeriğinin duygu yönünden (olumlu/olumsuz) değerlendirilmesi, müşteri memnuniyetinin ve müşteri temsilcisinin performansının ölçülmesi, konu ve kelime ağaçlarının otomatik olarak çıkartılması amaçlanmaktadır.

Bildirinin bundan sonraki bölümleri şu şekilde düzenlenmiştir: İkinci bölümde literatürde bulunan benzer çalışmalar hakkında bilgi verilmiştir. Üçüncü bölümde geliştirilen Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi ile ilgili detaylar paylaşılmıştır. Dördüncü bölümde geliştirilen yazılım çerçevesini kullanılarak yapılan örnek çalışmaya ve bu çalışmanın sonuçlarına yer verilmiştir. Son bölümde yapılan çalışma sonucunda geline noktanın değerlendirmesi yapılmış ve gelecek dönemde yapılacak çalışmalardan bahsedilmiştir.

## 2 Benzer Çalışmalar

Duygu analizi, metin içerisinden görüş, duygu ve tutum gibi öznel bilgilerinin çıkarılması için doğal dil işleme, metin madenciliği gibi yapay zekâ tekniklerinin kullanıldığı bir araştırma alanıdır [1]. Duygu analizi alanında makine öğrenmesi ve sözlük tabanlı yöntemlerle birçok akademik çalışma yapılmıştır. Makine öğrenmesi tabanlı yöntemler kullanılarak duygu analizi gerçekleştirilirken, duygu yönünden olumlu/olumsuz olarak etiketlenmiş veri kümesi üzerinde makine öğrenmesi algoritmaları uygulanarak sınıflandırma modeli oluşturulmaktadır. Ardından oluşturulan model yeni örneklerin sınıflandırılmasında kullanılmaktadır. Sözlük tabanlı yöntemler ise daha önceden oluşturulmuş duygu sözcükleri içeren bir duygu sözlüğüne dayalı olarak duygu analizini gerçekleştirmektedir. Metinde geçen sözcük ve cümlelerin anlamsal yönelimlerine dayalı bir hesaplama gerçekleştirilerek duygusal sınıflandırma yapılmaktadır.

Duygu analizi alanında makine öğrenmesine dayalı yöntemler kullanılarak yapılan temel çalışmalardan biri Pang ve diğerleri [2] tarafından gerçekleştirilmiştir. Bu çalışmada, çeşitli makine öğrenmesi yöntemleri (SVM, Naïve Bayes ve maksimum entropi) IMDB platformundaki film yorumlarını içeren veri kümesine uygulanarak film yorumları olumlu-olumsuz olarak sınıflandırılmıştır. Uygulanan yöntemlerin başarımları ve performansları kıyaslanmıştır. En yüksek doğru sınıflandırma oranının %82,9 ile destek vektör makinesi (SVM) yöntemi uygulandığında elde edildiği görülmüştür. Türkçe için benzer yüksek lisans tez çalışmaları gerçekleştirilmiştir. Eroğlu çalışmasında destek vektör makineleri yöntemini ve N-gram modelini kullanarak %85 başarı ile Beyazperde platformundaki film yorumları olumlu-olumsuz olarak sınıflandırılmıştır [3]. Akbaş ise çalışmasında Türkçe tweetler üzerinde kişiler ve başlıklar hakkındaki alt konuları çıkarıp bu konuları duygu kutbuna göre gruplayabilmiştir [8].

Vural ve diğerleri [4] çalışmalarında Türkçe film yorumlarını sözlük tabanlı yöntemle duygu analizi yapmışlardır. SentiStrength Kütüphanesini Türkçe'ye çevirerek film yorumlarını olumlu-olumsuz olarak sınıflandırmışlardır. Eroğlu'nun çalışmasında kullandığı Beyazperde platformundan alınan veri kümesine kendi yöntemlerini uygulayarak %76 başarı elde etmişlerdir.

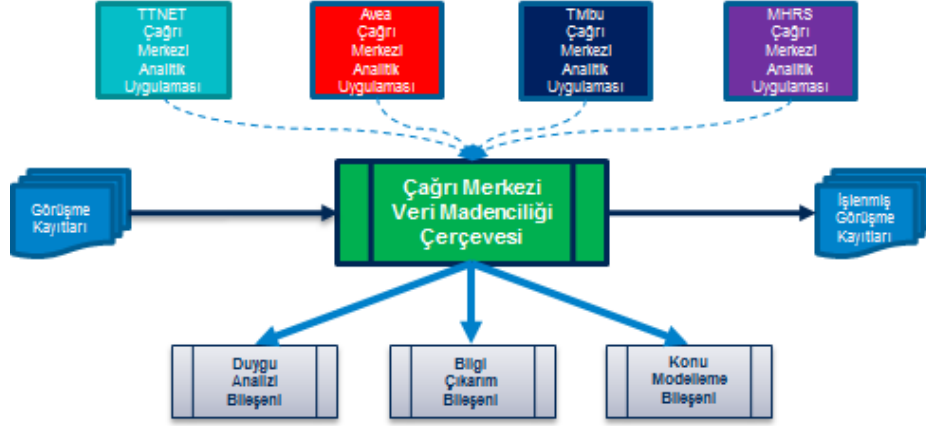
Ergün ve arkadaşları [5] yaptıkları çalışmada, sözcük tabanlı yöntemlerden yararlanarak internetteki tüketici yorumları üzerinden duygu analizi yapmışlardır. Tüketici yorumlarının metinleri içerisinde ürün özelliğini gösteren kelimeler ve olumlu-olumsuz anlamlı sıfatlar belirlenmiştir. Ürünü niteleyen sıfatlar temsil ettikleri memnuniyet derecelerine göre puanlanmıştır. Niteleyen ve nitelenen kelimelerin tespiti için açık kaynak kodlu Zemberek Kütüphanesi [6] kullanılmış ve Türkçe dilbilgisine göre ağaç yapısı oluşturulmuştur. Ağaç yapısı üzerinde Derinlik Öncelikli Arama algoritması kullanılarak ürünün herhangi bir özelliğine ilişkin memnuniyet derecesini ifade eden sayısal bilgiler hesaplanmıştır.

Eliaçık ve Erdoğan [7] çalışmalarında makine öğrenmesi tabanlı bir yöntemle Twitter'daki Türkçe finans iletilerinin duygu kutbunu (pozitif, negatif) belirlemişlerdir. Çalışma kapsamında finans konusunda uzman kişiler tarafından 1501 negatif, 907 pozitif Türkçe kısa iletilerin bulunduğu veri kümesi oluşturulmuştur. Özellik çıkartmak için unigram ve bigram, özellik seçmek için PMI (Point-wise Mutual Information) yöntemi kullanılmıştır. Özellik çıkarma ve seçme işlemlerinden sonra 35030 özellikten oluşan özellik kümesi elde edilmiştir. Bu çalışmada, destek vektör makinesi tabanlı bir duygu sınıflandırma yöntemi kullanılarak %73,63 başarı oranı ile duygu kutbu belirlenmiştir.

Literatürde yapılan benzer çalışmalar duygu kutbuna göre kategorize etme ve sınıflandırmayla ilgilidir. Bu çalışmadaysa çağrı merkezlerindeki görüşmelerin duygu yönünden (olumlu/olumsuz) değerlendirilmesi yanında müşteri memnuniyeti ve temsilci performansının ölçülmesi, görüşmelerde geçen konuların tespit edilmesi hedeflenmiştir. Ayrıca bu işlevlerin yeniden kullanılabilir bir yazılım çerçevesi tarafından sunulması çalışmamızda önerilmektedir.

### 3 Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi

Bu çalışma kapsamında ses kayıtlarına ilişkin sesten dönüştürülmüş metinlerin sınıflandırılması ve içeriklerinin değerlendirilmesi için Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi geliştirilmiştir<sup>1</sup>.



Şekil 1. Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi

Bu çalışma kapsamında ses kaydı metinlerinin üzerinde metin madenciliği teknikleri uygulanarak metinlere ilişkin yeni özellikler elde edilmektedir. Metinlerden çıkarılan yeni özelliklerle beraber gözetimli makine öğrenmesi algoritmaları kullanılarak görüşmenin havasının, müşteri memnuniyetinin ve temsilci performansının tahmin edilmesi için sınıflandırma ve regresyon modelleri oluşturulmuştur. Sınıflandırma ve regresyon modellerini oluşturup çıktılarını elde etmek için Çağrı Merkezi Metin Madenciliği Yaklaşımı [10] uygulanmıştır. Ayrıca sorun/şikayet başlıklarının belirlenmesi için konu ve kelime ağaçlarının otomatik olarak çıkarılmıştır.

Bu bölümün alt başlıklarında Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi ile sınıflandırma ve regresyon modellerine eklenen özelliklere ve sorun/şikayet başlıklarının belirlenmesine ilişkin ayrıntılı bilgilere yer verilmiştir.

#### 3.1 Ses Kayıtlarının Meta Özelliklerinin Eklenmesi

Veri kümesinde yer alan ses kayıtlarının meta özellikleri müşteriyle, müşteri temsilcisiyle veya doğrudan görüşmenin kendisiyle ilgili tanımlayıcı istatistiksel özelliklerdir. Ses kayıtlarının meta özellikleri sesin analizi sonucunda ortaya çıkmıştır. Çalışma kapsamında meta özelliklerin tamamı gözden geçirilmiş ve tahmin modellerinde kullanılmıştır.

<sup>1</sup> Bu çalışmada kullanılan sesten metne dönüştürülmüş veriler Türk Telekom Grubu şirketlerinden AssisTT A.Ş. tarafından sağlanmıştır.

maya uygun olanlar belirlenmiştir. Görüşmenin analizi sonucunda çıkarılan meta özelliklere konuşma süresi, konuşmaların üst üste gelme süreleri, monotonluk, sınırlılık yerleri ve sınırlılık yüzdeleri örnek olarak verilebilir.

Veri kümesindeki bütün özellikler tahmin modellerinde kullanılması için uygun değildir. Özelliklerin bir kısmı sayısal değerler içermektedir ve doğrudan sınıflandırma modeline girecek durumdadırlar. Sayısal olmayan özellikler önışlemeden geçirilerek sayısal hale getirilmiştir. Böylece sayısal olmayan özellikler de tahmin modellerine girebilecek duruma getirilmiştir.

### 3.2 Ses Kayıtlarının Metin Tabanlı Özelliklerinin Eklenmesi

Çalışma kapsamında veri kümesindeki meta özelliklerin yanı sıra konuşmanın müşteri ve temsilci metinlerinden çıkarılan özelliklerle birlikte özellik kümesi zenginleştirilmiştir ve tahmin doğruluğu artırılmaya çalışılmıştır. Metin tabanlı özellikler çıkarılırken üç farklı metin madenciliği tekniği uygulanmıştır. Uygulanan metin madenciliği tekniklerine ilişkin ayrıntılar aşağıda verilmiştir. Bu başlık altında verilen metin tabanlı özellikler örnek olmakla birlikte eldeki veriye göre daha farklı özellikler eklenebilecekler veya bu özelliklerden bazıları kullanılamayabilecektir.

**Duygu analizi ile yeni özellikler elde etme.** Duygu analizi modelleri serbest metin verileri üzerinde oluşturulmaktadır ve metinlerin duygu yönünü tahmin etmede kullanılmaktadır. Bu modeller kullanılarak bir cümlenin duygu yönü pozitif/negatif ekseninde tahmin edilebilmekte ve bu tahminin kuvvet derecesini gösteren skorlar elde edilebilmektedir. Duygu analizi modelleri gözetimli [11][12][13] veya gözetimsiz [13][14][15] sınıflandırma modelleri olarak oluşturulabilmektedir. Duygu analizi modelleri oluşturulduktan sonra ses kaydı metinlerindeki cümlelerin duygu yönü ölçülerek her ses kaydına ait çeşitli istatistikler çıkarılmaktadır. Çıkarılan bu istatistikler elde edilmeye çalışılan özellik kümesinde birer özellik olarak değerlendirilmektedir.

Ses kaydı metinlerindeki bütün cümleler duygu analizi modelinden geçirilmektedir ve bu modelin verdiği skorlar (0 ile 1 arasında) kullanılarak yeni özellikler hesaplanmaktadır. Yeni özellikler hem müşteri hem de temsilci metinleri için ayrı ayrı çıkarılmaktadır. Aşağıda duygu analizi ile elde edilen özellikler açıklamalarıyla birlikte verilmektedir.

- **Negatif/Pozitif yüzdesi:** Negatif/Pozitif olarak tahmin edilen cümlelerin sayısının metin içinde geçen toplam cümlelerin sayısına oranıyla hesaplanmaktadır.
- **Ortalama negatif/pozitif skoru:** Negatif/Pozitif olarak tahmin edilen cümlelerin skorlarının metin içinde geçen tüm cümleler bazında ortalaması alınarak hesaplanmaktadır.
- **Toplam negatif/pozitif skoru:** Negatif/Pozitif olarak tahmin edilen cümlelerin skorlarının toplamı alınarak hesaplanmaktadır.

**Ayırt edici kelimeler ile yeni özellikler elde etme.** Ses kaydı metinleri için özellik kümesi oluşturma sürecinde kullanılan diğer bir metot ayırt edici kelimeleri tespit etme

yöntemidir. Bu metodun uygulanabilmesi için ses kaydı metinlerinin olumlu/olumsuz olarak etiketlenmiş olması gerekmektedir. Etiketlenmiş ses kaydı metinleri üzerinde WLLR (Weighted Log Likelihood Ratio) [16][17] tekniği uygulanarak her bir etiket sınıfı için ayırt edici kelimeler tespit edilebilir ve daha sonra bu kelimeler her sınıf için oluşturulacak tahmin modelleri için birer özellik olarak kullanılabilir.

**Metin üzerinden önceden belirlenmiş kurallar ile yeni özellikler elde etme.** Ses kaydı metinleri üzerinden önceden belirlenmiş kurallara göre yeni özellikler çıkarılmaktadır. Aşağıda tahmin modellerinde bulunan özellikler açıklamalarıyla birlikte verilmektedir.

- **Kelime sayısı:** Ses dosyasındaki toplam kelime sayısını yansıtmaktadır.
- **Temsilcinin görüşmeyi olumlu sonlandırması:** Temsilcinin görüşmenin sonuna doğru müşterinin sorununa yardımcı olacak veya yönlendirici şekilde davranıp davranmadığına, son cümlelerde bazı kelimelerin söylenip söylenilmediğine bakılır.
- **Müşterinin görüşmeyi olumlu sonlandırması:** Konuşmanın sonuna doğru müşterinin olumlu cümleler söyleyip söylemediğinin tespiti için kullanılır. Müşteri konuşmayı kapatırken naziklik içeren bazı kelimeleri söyleyip söylemediğine bakılır.
- **Görüşmenin olumlu sonlandırması:** Yukarıdaki iki özellik dikkate alınarak konuşmanın sonuna doğru hem müşterinin hem de temsilcinin davranışı ölçülür.
- **Negatiflik/Pozitiflik:** Konuşmadaki olumsuzluk/olumluluk sayısını belirten özelliktir. Önceden belirlenen kötü/iyi kelime listesi kullanılarak belirlenir. Görüşmedeki herhangi bir cümlede kötü/iyi kelimelerden birisi geçiyorsa bu özelliğin değeri bir artırılır.
- **Kibarlık:** Temsilcinin müşteriye karşı ne kadar kibar davrandığına, kibarlık ifade eden kelimeleri kullanıp kullanmadığına bakılır.
- **Telefonun kapanması:** Müşterinin konuşmayı olağan bir biçimde sonlandırıp sonlandırmadığına bakılır.

### 3.3 Sorun/Şikayet Başlıklarının Belirlenmesi

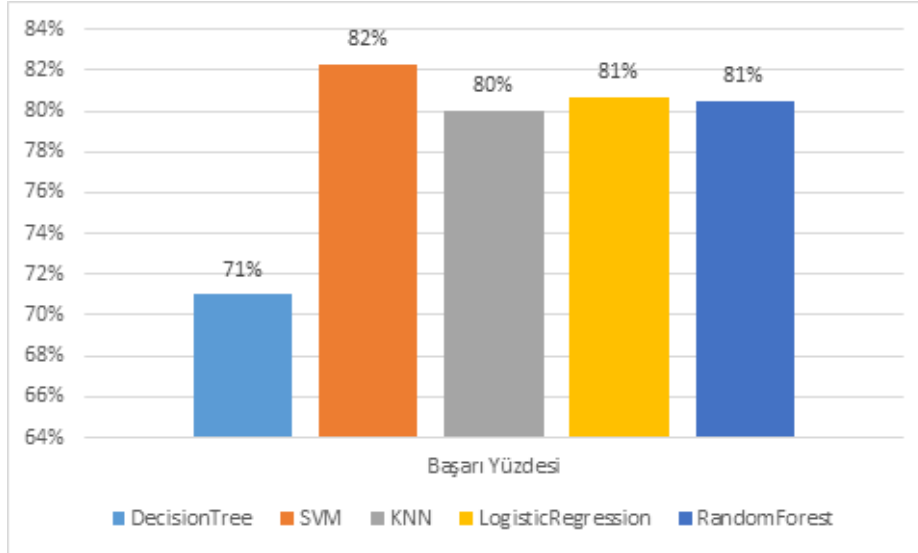
Çağrı merkezlerinde elde edilen veri kümelerinin büyüklüğü oldukça fazladır. Buna rağmen bu veri genellikle etiketlenmemiş haldedir. Diğer yandan ses kayıtlarında geçen müşteri sorunlarının sınıflandırılması ve daha önceden belirlenmiş sorun başlıklarıyla ilişkilendirilmesi oldukça önemlidir. Hali hazırda hangi konuşmanın hangi sorun başlığıyla ilişkili olduğunu belirlemek için her bir sorun başlığı için kelime ağaçları oluşturulması gerekmektedir. Hem etiketsiz veri üzerinde çalışması hem de belirli bir doküman kümesi üzerinde konu başlıklarını gözetimsiz bir şekilde tespit etmesi nedeniyle konu modelleme [19] tekniğinin bu veri seti üzerine uygulanmasına karar verilmiştir. Konu modelleme tekniğinin uygulanması için Mallet [20] kütüphanesinden yararlanılmıştır. Bu tekniğin uygulanması neticesinde bütün doküman kümesinde yer alabilecek konu başlıklarının tespiti, bu konu başlıklarını temsil edecek kelime listeleri ve her bir dokümanın her bir konu başlığıyla ilişkili olma oranları elde edilmektedir. Belirlenen konu başlıkları çağrı merkezi alanındaki sorun başlıklarıyla eşleştirilebilir ve böylelikle sorun başlıklarıyla ilgili kelime ağaçları otomatik olarak elde edilmiş olunur. Diğer

yandan her bir ses kaydının daha çok hangi sorunla ilgili konuşmayı içerdiği tespiti yapılabilir.

#### 4 Örnek Çalışma

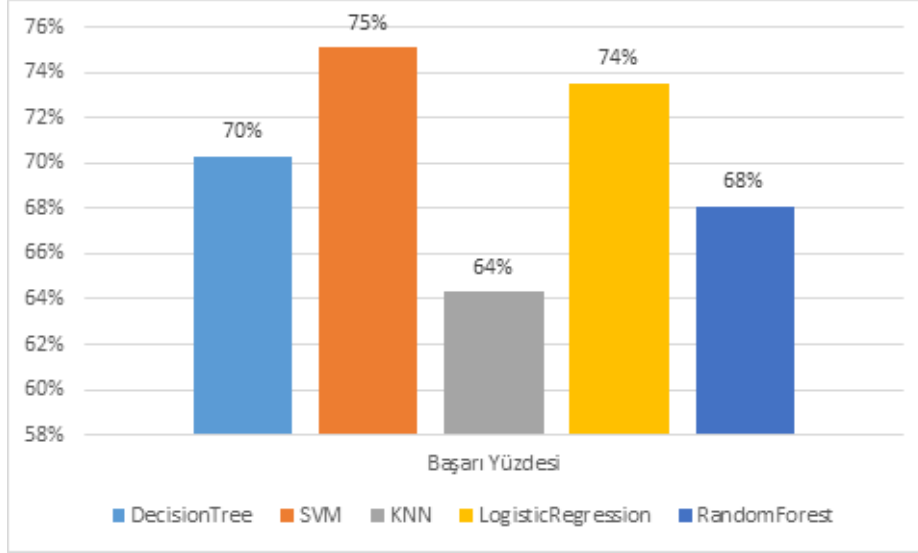
Tahmin modellerinin oluşturulması için ilk olarak ses kayıt dosyalarının etiketlemesi ve metin madenciliği teknikleri kullanılarak özellik vektörlerine dönüştürülmesi işlemleri tamamlanmıştır. Özellik vektörlerine dönüştürme sürecinde bir önceki bölümde anlatılan özellik çıkarma yöntemleri ve yine aynı bölümde bahsedilen özellikler kullanılmıştır. Etiketleme işlemi ise görüşmenin havası olumlu/olumsuz, müşteri memnuniyeti ve temsilci performansı ise 1 ile 5 arasında olacak şekilde derecelendirerek tamamlanmıştır. Etiketleme çalışmaları sonucunda 400 tane ses kaydı olumlu/olumsuz olarak etiketlenmiştir. Bu örneklerden 92 tanesi olumsuz, kalan 308 tanesi ise olumlu olarak tespit edilmiştir. Bu aşamada iki farklı deney gerçekleştirilmiştir. Her iki deneyde de çapraz doğrulama (cross validation) tekniği kullanılarak eğitim ve test veri kümeleri belirlenmiştir.

Birinci deneyde Karar Ağacı (Decision Tree), Destek Vektör Makineleri (SVM), K En Yakın Komşu (KNN), Lojistik Regresyon (Logistic Regression), Rasgele Orman (Random Forest) algoritmaları denenerek görüşmenin havasını tespit etmek için en başarılı sınıflandırma algoritması belirlenmeye çalışılmıştır. Şekil 2'de denenen algoritmaların sınıflandırma başarılarıyla ilgili sonuçlar gösterilmektedir. Bu sonuçlara göre %82 doğruluk ile az bir farkla en başarılı sınıflandırma SVM algoritması olmuştur.



Şekil 2. Dengesiz veri kümesi başarı yüzdeleri

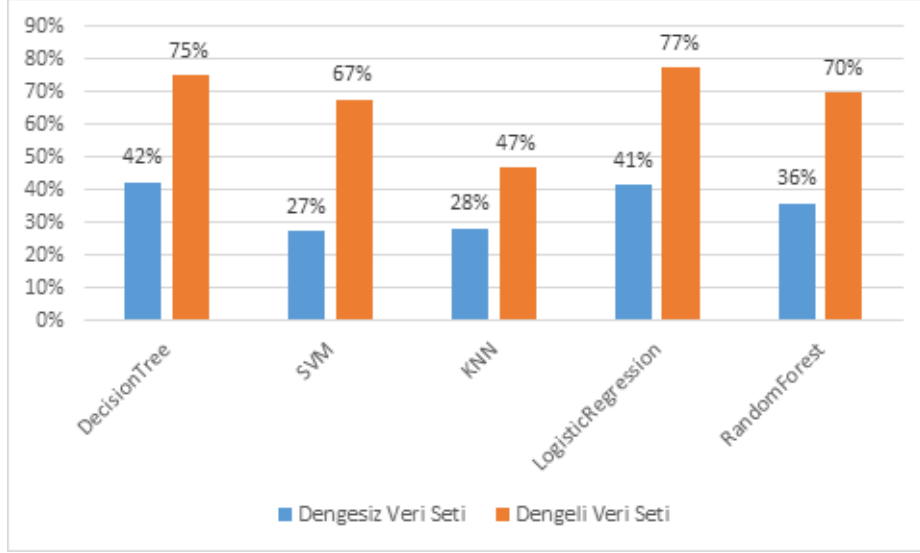
İkinci deneyde olumlu ve olumsuz sayısının dengeli/eşit olacak şekilde veri kümesi düzenlenmiştir. Veri kümesi düzenlikten sonra sınıflandırma algoritmaları tekrar karşılaştırılmıştır. Karşılaştırma sonuçları incelendiğinde SVM ve Lojistik Regresyon metodlarının diğerlerine göre daha başarılı sonuçlar elde ettiği gözlemlenmiştir.



Şekil 3. Dengeli veri kümesi başarı yüzdeleri

Şekil 4'deki grafikte algoritmaların olumsuz görüşmelerin tahminine ilişkin dengeli ve dengesiz veri kümeleriyle yapılan deneylerdeki başarı yüzdeleri yer almaktadır. Dengeli veri kümesi kullanılarak oluşturulan sınıflandırma modellerinde genel tahmin başarısı açısından düşüş görülse bile olumsuz görüşmelerin tahmin etmedeki başarı yüzdesinde kayda değer artış olduğu görülmektedir. Olumsuz görüşmeleri sınıflandırmada en başarılı algoritmaların Karar Ağacı ve Lojistik Regresyon olduğu tespit edilmiştir. Ayrıca, tüm sınıflandırma algoritmalarının dengeli veri kümeleriyle beraber kullanıldığında olumsuz görüşmeleri tahmin etmedeki başarı yüzdeleri dengesiz veri kümelerinin kullanıma göre artış göstermiştir. Örneğin, genel başarı açısından en iyi algoritma olan SVM sonuçları incelendiğinde dengesiz veri kümesi kullanıldığı zaman olumsuz görüşmeleri doğru tahmin etme başarısı %27 iken dengeli veri kümesi ile yapılan deneyde başarının %67'e çıktığı gözlemlenmiştir.





**Şekil 4.** Olumsuz görüşme tahmininde başarı yüzdeleri

Etiketleme sürecinde konuşmanın içeriği dikkate alınarak 1 ile 5 arasında derecelendirilen müşteri memnuniyeti ve temsil performansı değişkenlerinin tahmini için Lineer Regresyon algoritması kullanılarak regresyon modelleri oluşturulmuştur. Müşteri memnuniyeti ve temsilci performansı değişkenlerinin bağımlı, özellik kümesinde yer alan diğer özelliklerin bağımsız değişken olduğu regresyon modellerinde ilk 320 örnek öğrenim kümesi için son 80 örnek ise test kümesi için kullanılmıştır. Müşteri memnuniyeti ve temsilci performansı değişkenleri için bu deneyle ilgili sonuçlar aşağıdaki tabloda verilmiştir.

**Tablo 1.** Regresyon modeli sonuçları

Değişken	Varyans	Ort. Mutlak Hata	Ort. Karesel Hata
Müşteri Memnuniyeti	-0,06	0,68	0,66
Temsilci Performansı	-0,01	1,12	1,81

Regresyon modelinin sonuçları incelendiğinde müşteri memnuniyetinin ve temsilci performansının düşük varyans değeri ile tahmin edilebildiği görülmektedir. Ortalama mutlak hata ve ortalama karesel hata metriklerine bakıldığında regresyon modelinin müşteri memnuniyetini temsilci performansına göre daha başarılı bir şekilde ölçebildiği gözlemlenmektedir.

Ses kayıtlarını içeriğinde geçen sorun/şikayet başlıklarının tespit edilmesi için Mallet kütüphanesinden faydalanarak konu modelleme tekniği uygulanmıştır. Bu yaklaşımı uygulamak için her bir ses kaydı bir doküman olarak düşünülmüştür. Farklı bir yaklaşım olarak müşterinin konuşma metni veya temsilcinin konuşma metni ayrı ayrı doküman olarak da düşünülebilir. Gözetimsiz olarak gerçekleşen ve etiketsiz 6000 ses kaydı üzerinde inşa edilen bu modelle ilgili konu başlıklarını temsil eden kelime listeleri ve bu konuların tahmini isimlendirmesi aşağıdaki şekilde gösterilmiştir.

Konu No	Konu Ağırığı	Anahtar Kelimeler	Tahmini Konu
0	0.08256	evet nokta internet yüz beş dms sıfır otuz bağlantı doksan diyor kablolu anda sorun bilgisayar yok bağlantısı sayfa kablolu dört yüz evet beş dört altı kırk otuz sıfır numarası seksen yirmiyetmiş elli hizmet doksan telefon	Anıza (Bağlantı sorunu)
1	0.49116	numara numarasını bey hat iptal için istiyorum imza gün gerekiyor işlemi türk buradan içerisinde başıru anda iptali yok	
2	0.15778	tekrar şimdi iptalini ettirmek etmek hattınız randevu nakil için dönüş bugün ilgili yirmi gün geri evet saat ayın günler kayıt içerisinde	İptal
3	0.31272	kaydınızı şikayet günü konuyla beş fatura ödeme hattınız dolayı faturanız otomatik aktif tane olarak ödenmemiş yirmi kapalı	Randevu
4	0.19142	ödedim ayın tarihi lira kontrol kredi gün faturam lira yirmi dört evet beş elli altı ücreti bin kontrol yüz olarak yetmiş internet fatura	Ödenmemiş Fatura
5	0.27288	taahhüt doksan ayın eee ikinci mail miyim olara k bina önceki evlenmeden fatura harfını annenizin için nokta öğrenebilir	Fatura Sorgulama
6	0.26349	harfi olduğunuz numarası evet güvenliğiniz soyadının kom nakil olarak solak evet mahallesi bina işlemi adres daire yirmi dört numarası işleminiz adresi	Bilgi Doğrulama
7	0.11445	eee ii mahalle ücreti devam kat evet bey ediyorum ilgili kontrol anda olarak devam iyi sağlıyorum iii günler teşekkür miyim	Nakil
8	0.83259	konuyla tarafınıza yardımcı alakalı işlem teşekkürler tivbu ister yok bilgi misiniz müzik almak türk tnet peki hakkında vermek özel devam onay	Diğer Hizmetler Bilgilendirme (Tivbu, TTNET Müzik vb.)
9	0.12605	sadece abonelik bilgilendirme ait yer anıza ilgili sorun problem kaydı/kaynaklı kontrol kaydınızı modem kopma ağk internet saat beş	
10	0.24884	gerekli teknik sinyal sürekli sıkıntı ekip şey yok böyle şimdi anlıyorum nasıl dedi şekilde değil dediğim bakın hanımefendi gün kayıt ilgili	Anıza (İnternet bağlantısının kopması)
11	0.3513	tekrar işte hiçbir şikayet müşterileri iyi günler iii teşekkür evet miyim hanım yardımcı ederim teşekkürler için kontrol ediyorum	Şikayet Kaydı
12	1.13075	akşamlar öğrenebilir hat nasıl isim bey rica modem internet yanyoru anda modemi ağk yok evet sinyal kontrol kapatıp işi direkt modeme	
13	0.21949	modemin adsl yanyıyor dsl yapıp kablo telefon internet türk ilgili dört için telefonu iii bey yok günler sabitiyi şimdi cep kırk tnet	Anıza (Modem arızası)
14	0.34213	telefonla internete nasıl evet kullanıo şifre adı modem yüz kablolu nokta tnet şifrenizi giriş kurulum şifreyi şimdi	
15	0.12703	yazıyor küçük doksan tekrar yeni dört dondurma hat gün hattınız aktif hale hattınızı günlük süre doksan devam altı internet tekrar	Kurulum
16	0.06078	itibarıyla içerisinde yirmi süresi bin dondumak elli gigabyte paket kullanım kampanya doksan dört geçiş modem megabite beş adli yetmiş	Üyelik Dondurma
17	0.13484	yirmi altı evet internet megabit limitsiz pakete olarak ilgili işlem için bilgi gerekiyor şekilde türk bununla üzerinden peki farklı çünkü sistem	Kampanya
18	0.43049	kaynaklı buradan eğer eee tekrar hanım iii evet şimdi şey eee yok anda değil için bey anladım şöyle peki çünkü öyle şekilde işte hemen	
19	1.05673	bakalım mesela	

Şekil 5. Konu modelleme sonucu tespit edilen konular ve kelimeler

## 5 Sonuç

Çağrı merkezi görüşme kayıtlarına ilişkin veriler günümüzde birçok kurum tarafından elde edilmekte ve sesten metne dönüştürülerek saklanmaktadır. Bu çalışma kapsamında geliştirilen Çağrı Merkezi Metin Madenciliği Çerçevesi, metin madenciliği ve makine öğrenmesi algoritmaları kullanılarak görüşme kaydı metinlerinin analiz edilmesini için bir altyapı sunmaktadır. Bu çalışma sonucunda ortaya çıkan yazılım çerçevesi kullanılarak Türk Telekom Çağrı Merkezine gelen çağruların görüşmenin havasının olumlu/olumsuz olması, müşteri memnuniyeti ve temsilci performansı açılarından değerlendirilebilecek bir platform geliştirilmesi amaçlanmaktadır. Önümüzdeki dönemde Türk Telekom bünyesinde geliştirilecek Ar-Ge proje sonunda kapsamlı bir ürün çıkarılması planlanmaktadır. Ayrıca çağrı merkezi verileri dışında farklı kaynaklardan (sosyal medya, e-posta, anketler vb.) elde edilecek müşteri görüşlerini içeren metin tabanlı verilerinin de geliştirilecek platforma entegre edilmesi değerlendirilecektir.

## Kaynakça

1. Onan, A., Korukoğlu, S.: Makine öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür araştırması. Pamukkale Univ Muh Bilim Derg, cilt 22, no. 2, pp. 111-122, 2016.
2. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. Proceedings of EMNLP, 2002.
3. Eroğlu, U.: Sentiment Analysis in Turkish. The Graduate School of Natural and Applied Sciences of Middle East Technical University, Ankara, 2009.
4. Vural, A., Cambazoglu, B., Senkul, P., Tokgoz, Z.: A Framework for Sentiment Analysis in Turkish: Application to Polarity Detection of Movie Reviews in Turkish. Computer and Information Sciences III, Springer, pp. 437-445, 2012.
5. Ergün, K., Kubat, C., Çağır, G., Cesur, R. : İnternet ortamındaki tüketici yorumlarından özet bilgi çıkarımı. SAÜ. Fen Bil. Der., cilt 17, no. 1, pp. 33-40, 2013.
6. Akın, M., Akın, A.: Türk Dilleri İçin Açık Kaynaklı Doğal Dil İşleme Kütüphanesi: Zemberek. Elektrik mühendisliği, cilt 431, pp. 38-44, 2007.
7. Eliaçık, A., Erdoğan, N.: Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi. UYMS, İzmir, 2015.
8. Akbaş, E.: Aspect Based Opinion Mining on Turkish Tweets. The Graduate School of Engineering and Science of Bilkent University, Ankara, 2012.
9. Caputo, B.: Using text mining to understand the call center customers' claims. WIT Transactions on Information and Communication Technologies, Vol 37, 2006.
10. Yiğit, İ., Ateş, A., Güvercin, M., Ferhatosmanoğlu, H., Gedik, B.: Çağrı Merkezi Metin Madenciliği Yaklaşımı. 25. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, 2017.
11. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 2009.
12. Maas, A.: Learning word vectors for sentiment analysis. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, Association for Computational Linguistics, 2011.
13. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and trends in information retrieval 2.1-2, 1-135, 2008.

14. Hu, X.: Unsupervised sentiment analysis with emotional signals. Proceedings of the 22nd international conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2013.
15. Turney, P.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, 2002.
16. Ng, V., Sajib, D., Arifin, S., Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. Proceedings of the COLING/ACL on Main conference poster sessions, Association for Computational Linguistics, 2006.
17. Nigam, K., Text classification from labeled and unlabeled documents using EM. Machine learning 39.2-3, 103-134, 2000.
18. Yiğit, İ., Dogru, A.: Yazılım Ürün Hatlarında Alana Özgü Bileşenleri Belirleme Yaklaşımı. Ulusal Yazılım Mühendisliği Sempozyumu, İzmir, Türkiye, 2015.
19. Blei, D.: Probabilistic topic models. Communications of the ACM 55.4 (2012): 77-84.
20. McCallum, A.: Mallet: A machine learning for language toolkit. 2002.