

# Türkçe Metinlerde İroni Tespiti

Oğuzhan Dülger<sup>1</sup>

<sup>1</sup> Gazi Üniversitesi, Ankara, Türkiye  
oguzhan.dulger@gazi.edu.tr

**Özet.** İnternet hayatımıza girdiğinden beri web sitelerinde her gün milyonlarca yorum ve düşünce paylaşılmaktadır. Paylaşılan bu düşünceler duygu analizi alanındaki çalışmalara konu olmaktadır. İroni; söylenen sözün tersini kastetme niyetiyle yapılır ve internetteki metin kaynaklarında çok sık rastlanan bir söz sanatı olması sebebiyle duygu analizi çalışmalarında sınıflama algoritmasının yanlış sonuçlar üretmesine yol açmaktadır. Duygu analizi çalışması yapmadan önce metinde ironi varlığı tespit edilmelidir. Tespit edilen ironinin duygu analizi algoritmasına öznitelik olarak eklenmesi gerekmektedir. Bu nedenle yapılan bu çalışmada Türkçe metinlerde ironi tespiti yapılmaya çalışılmıştır. Metin madenciliği yöntemleri ve Türkçe cümleleri morfolojik açıdan inceleyen kütüphane yardımı ile ironi tespitinde yeni bir algoritma geliştirilmiştir. Bu algorithmada 13 farklı öznitelik kullanılarak sınıflama yapılmıştır. Çalışmanın sonucunda %88 oranında sınıflama başarısına ulaşılmıştır.

**Anahtar Kelimeler:** İroni sınıflama, Duygu analizi, Türkçe metin madenciliği, Doğal dil işleme.

## Irony Classification in Turkish Text

Oğuzhan Dülger<sup>1</sup>

<sup>1</sup> Gazi University, Ankara, Turkey  
oguzhan.dulger@gazi.edu.tr

**Abstract.** Millions of comments and thoughts have been shared every day in the websites since the Internet became a part of our life. These comments and thoughts are the subject of researches in the field of text mining. Irony is a language art that is common in the text resources on the Internet, it causes the classification algorithm to produce incorrect results in sentiment analysis studies. So, it is necessary to detect the irony to improve the sentiment analysis classification score. Classified irony result should be added to the sentiment analysis as a feature. In this study, irony is classified for Turkish sentiment analysis studies. A new algorithm for irony detection is developed with the help of a text mining methods and libraries which are analyzing Turkish texts morphologically and sentimentally. 13 different features were used and six different classification al-

gorithms was employed in the study. After all, %88 classification rate was achieved by this algorithm.

**Keywords:** Irony classification, Sentiment analysis, Turkish text mining, Natural language processing

## 1 Giriş

İnsanlar internetin yaygınlaşmasından ve neredeyse her eve girmesinden sonra düşüncelerini sosyal medya platformlarında, mikrobloglarda ve e-ticaret sitelerinde paylaşır oldu. Paylaşılan bu yorumlar büyük veri ve metin madenciliği çalışmaları için güzel bir kaynak anlamına gelmektedir. Bunun sonucunda analistler, politikacılar ve firmalar insanların belli bir konu hakkında düşüncelerini sosyal medya ve mikroblog sitelerinden öğrenmek istemektedirler. Ayrıca elektronik ticaret sitelerinde de ürünler hakkında girilen yorumlarda da binlerce eleştiri bulunmaktadır. Ticari kuruluşlar da ürünleri hakkındaki olumlu ve olumsuz yorumları öğrenmek istemektedirler. Bu yüzden duygu analizi araştırmaları sosyal medya, mikroblog ve e-ticaret siteleri üzerinden yapılan çalışmalarla gelişmektedir.

Duygu analizi metin madenciliği çalışmaları arasında yer alan konulardan biridir ve yazılı metinlerdeki düşünceyi kutupsallık bazında sınıfları bulmaya çalışır. Duygu analizi yaparken bazı nedenlerden dolayı sınıflama başarısı düşer. Bu nedenlerden biri de metin içinde bulunan ironidir. İroni özellikle sosyal medya paylaşımlarında ve e-ticaret sitelerinde sık karşılaşılan bir söz sanatıdır. Bunun sebebi bu platformlardaki dilin resmi dilden daha çok günlük konuşma diline benzemesidir.

İronilerin yapısal ve anlamsal kutupsallığı zıt olduğu için duygu analizi algoritmalarını yanıltır. Bu yüzden duygu analizi sınıflandırmasında başarı oranını arttırmak için metin içinde geçen ironinin mutlaka tespit edilmesi gerekmektedir [1-3]. Sarmiento ve arkadaşlarının (2009) yaptığı çalışmada negatif duygu sınıflamasının başarı oranı yaklaşık %90 iken, pozitif duygu sınıflaması %60'ın altında kaldığı görülmüştür [4]. Hatalı kayıtlar incelendiğinde yanlış sınıflanan metinlerin %35'inde ironi varlığı tespit edilmiştir. Maynard ve Greenwood (2014) ile Bouazizi ve Ohtsuki (2015) yaptığı çalışmalarında ironi sınıflamanın duygu analizine katkısı deneysel olarak kanıtlanmıştır ve sınıflama başarısının sayısal olarak arttığı gözlemlenmiştir [2] [3]. Bu sebeple duygu analizi yapmadan önce metinde ironi tespiti yapmak gereklidir.

İroni tespiti konusunda yapılan çalışma sayısı fazla değildir. Bu çalışmalar İngilizce başta olmak üzere Portekizce, Hintçe gibi çeşitli dillerde yapılmıştır. Türkçe dilinde de şimdiye kadar yapılan bir tane çalışma vardır.

Sarmiento ve arkadaşlarının (2009) çalışmasında Portekizce dilinde politik düşünceler üzerinde duygu analizi yapılmıştır [4]. Bu çalışmada öznitelik olarak küçültme sözcükleri, işaret zamirleri, ünlem işareti, fiil morfolojisi, olumsuz anlama gelmeyen sıfatlar, fazla noktalama, tırnak işareti, emojiler ve gülme ünlemleri kullanılmıştır.

Reyes ve arkadaşlarının (2013) çalışmasında yeni öznitelikler olarak kullanılan zamansal dengesizlik, bağlamsal dengesizlik ve duygusal polarite öznitelikleri göze çarpmaktadır [5].

Riloff ve arkadaşlarının (2013) yaptığı çalışmada ironinin negatif bir durumu pozitif bir şekilde ifade etme özelliğinden yola çıkarak sınıflama yapılmaya çalışılmıştır [6]. Örneğin “telefonum çalındığı için çok mutluyum” cümlesinde gibi pozitif eylem ve negatif durum birlikteliği aranmaktadır.

Vanin ve arkadaşlarının (2013) çalışmasında Twitterdan Portekizce dilinde ironi sınıflaması yapılmıştır [7]. Bu çalışmada gülme ifadeleri, emojiler, kesin ifadeler (“sadece”, ”iyi de”...), noktalama işaretleri ile sıfat, belirteç, zarf ve isimlerden oluşan bazı desenler kullanılmıştır. Sınıflama sonucunda emoji, gülme ifadeleri, tırnak ve ünlem işaretlerinin ironiyi en çok ortaya koyan desenler olduğu ortaya çıkmıştır.

Buschmeier ve arkadaşlarının (2014) makalesinde Amazon yorumlarında ironi tespiti yapılmıştır [8]. Bu çalışmada duygusal polarite dengesizliği, abartma, alıntı işareti, noktalama işareti desenleri, bazı ünlemler, emojiler ve ironi belirteci olabilecek kelimeler paketi üzerinde durulmuştur. Bu çalışmanın sonucunda duygusal kutupsallık özneliğinin önemi anlaşılmıştır.

Barbieri ve Saggion (2014) yaptıkları çalışmada Twitter’den ironi sınıflandırması yapmışlardır [9]. Bu çalışmada frekans, yazılı-sözlü, yoğunluk ve eş anlam başlıkları altında ilk defa kullanılan 4 farklı öznitelik bulunmaktadır. Frekans; ironinin beklenmediklik ve sürpriz içeren bir söz sanatı olmasından yola çıkmaktadır. Frekans ortalaması; bir tweetteki tüm frekansların aritmetik ortalamasıdır ve bir tweetin frekans stilini tespit etmek için kullanılır. En nadir kelime; dengesizlik yaratabilecek kelimeyi yakalamak için tespit edilen en nadir kelimenin frekans değeridir. Frekans farkı; bu iki değer arasındaki farkı sonucundaki dengesizlik değeri olarak hesaplanmakta ve sürpriz olma ihtimali de buradan tespit edilmektedir. Yazılı-sözlü özneliğinde ise ironinin yazılı ifadelerden daha çok sözlü ifadelerde ve günlük konuşma dilinde daha çok kullanıldığından faydalanılmaktadır. Yoğunluk özneliğinde sıfat ve zarfların kullanım yoğunluğu fazla ise ironi olma olasılığının arttığı öne sürülmektedir. Eş anlam özneliğinde sık ve nadir kullanılan eş anlamlı kelimelerin kullanımı ölçülmüştür.

Rajadesingan ve arkadaşlarının (2015) çalışmasında ironi tespiti için Twitter’daki metinler kullanılmıştır [10]. Bu çalışmada diğer çalışmalardan farklı olarak iki öznitelik üzerinde durulmuştur. Birincisi kişinin tweet atarken gösterdiği psikolojik ve davranışsal özelliklerinden çıkarımlarda bulunulmuştur. İkincisi bu çıkarımlar ile sınıflandırma yapılırken kişinin geçmiş ve şimdiki tweetleri arasında farklar da tespit edilmeye çalışılmıştır. Tweetlerde çıkarılmaya çalışılan öznitelikler hem incelenen tweet içinde hem de geçmiş tweetler ile karşılaştırılarak hesaplanmıştır.

Taşlıoğlu’nun (2014) yüksek lisans tezinde Türkçe metinler için ironi sınıflandırması yapılmıştır ve veri kaynağı olarak da Twitter kullanılmıştır [11]. Parantezli ünlem işareti, ünlem işareti, soru işareti, alıntı işareti, emojiler, anlamsal skor farklılığı, küçültme ekleri, ünlemler, büyük harfli kelimeler öznitelik olarak kullanılmıştır. İroni tespitine en çok katkı sağlayan öznitelikler parantezli ünlem işareti, ünlem işareti ve soru işareti olarak hesaplanmıştır.

Bahsedilen çalışmalardan yola çıkarak şu çıkarımda bulunabilir. Tam tersi olduğu durumlar olsa da ironi çoğu zaman yapısal olarak olumlu gözükür ama anlam olarak

olumsuzdur. Bu yüzden duygusal sınıflandırma algoritmaları ironiyi olumlu olarak sınıflar ve sınıflamanın başarı oranı düşer. İroninin tespiti duygu analizi çalışmalarına ciddi oranda katkı sağladığı görülmüştür.

İroni tespiti içinduygu analizi çalışmalarında olduğu gibi özniteliklerin iyi belirlenmesi gerekmektedir. Metinde ironi olduğuna dair bir ipucu yakalamak gerekmektedir. Bunun için bazı noktalama işaretleri, sıfatlar, karakterler veya desenler bulmak öznitelik çıkarımı için ilk yapılması gereken işlerden biridir. Özniteliklerin bazıları dil bağımsız olsa da çoğu öznitelik çalışma yapılan dilin özelliklerinden etkilenmektedir. Bu özniteliklerin belirlenmesi için dilin kullanım tarzının, sözcük anlamlarının ve dil bilgisi kurallarının iyi bilinmesi gerekmektedir.

Bu çalışmanın ikinci bölümünde ironi tanımlanıp çeşitlerinden bahsedilmiştir. Üçüncü bölümde yapılan çalışmada kullanılan yöntem ve materyalden bahsedilmiştir. Dördüncü bölümde ise çalışmanın test ve sonuçları paylaşılmıştır. Son bölümde ise çalışma genel çerçevede değerlendirilmiş ve sonraki çalışmalar için önerilerde bulunulmuştur.

## 2 İroni ve Çeşitleri

İroni; sözlük anlamı olarak “Söylenen sözün tersini kastederek kişiyle veya olayla alay etme” demektir [12]. Günlük hayatta en çok kullanılan söz sanatlarından biridir. Özellikle günlük konuşmanın yaygın olduğu durumlarda konuşmacılar dikkat çekmek, iğnelemek, espri yapmak veya alay etmek için kullanırlar.

### 2.1 İroni Türleri

İroni genel olarak sözlü, durumsal ve dramatik olmak üzere üçe ayrılır [11].

**Sözlü İroni.** Konuşmacının niyetinin tam tersini söylediği veya bir durumdan yakındığı durumlarda yapılır. Bu konuşmacı tarafından kasıtlı bir şekilde üretilir. Günlük konuşmalarda, mikrobloglarda geçen yazılı ifadelerde en çok karşılaşılan ironi kategorisidir.

Ör: Sınava geç kaldım, aferin bana!

**Durumsal İroni.** Bir olayın beklentinin tam tersi şekilde gerçekleşmesi demektir. Burada sözlü veya yazılı bir ifadeden ziyade bir olay üzerinde ironi olur.

Ör: Bir trafik polisinin ceza yazmak için durdurduğu arabanın şoförünün çok yakın bir arkadaşının çıkması.

**Dramatik İroni.** Genelde film senaryolarında, romanlarda, tiyatrolarda karşılaşılan bir durumdur. İzleyicilerin veya okuyucuların bildiği ama karakterlerin bilmediği durumlarda gerilimin artması veya komik olaylar oluşturmak için oluşturulan durum-

lardır. Yazarlar veya senaristler bu ironi tarzını ilgi çekmek ve heyecanı üst düzeyde tutmak için kullanırlar.

Ör: Bir filmde maktul olan bir karakter kendisini odada bekleyen bir katilden haberi yokken, seyirciler bu durumdan haberdardır. Maktul her şeyden habersiz odaya girerken seyirci gerilim yaşar ve katil maktülü odada öldürür.

Sözlü ironi daha çok eğlenceli olma eğilimindedir. Durumsal ironi komik veya trajik olabilirken, dramatik ironi genellikle trajiktir.

## 2.2 İroninin Kullanım Amacı

İroni kullanıma amacına göre ince espri, sızlanma ve kaçınma olarak üç gruba ayrılır [13].

**İnce espri olarak ironi.** Bu tarz ironi komik olmak için yapılır. Konuşma sırasında abartma, tonlamanın farklılaşması ironinin anlaşılmasına yardımcı olur. Bunun yazıya dökülmüş hali ise büyük harfler, emoji, noktalama işaretleri, duygusal polarite şeklinde düşünülebilir ve bize ironi olduğuyla alakalı ipucu verebilir.

**Sızlanma olarak ironi.** Kişi kızgınlığını ironi olarak ifade edebilir. Negatif durumu abartma veya çok pozitif kelimelerle, ayrıca zamansal polarite ile ifade ironi yapabilir.

**Kaçınma olarak ironi.** Kişi eğer net cevap vermektan kaçınmak istiyorsa ironi yapabilir. Bu durumda karmaşık cümleler üretir, sık kullanılan kelimeler ve nadir görülen ifadelerle başvurur.

İroni günlük hayatta sözlü ve yazılı ifadelerde oldukça sık kullanılır. Sözlü halde ironiyi ayırt etmek yazılı ifadelerde geçenlere göre biraz daha kolaydır çünkü konuşan ses tonu, mimik ve gülmesiyle ironiyi belli edebilir. Bu çalışmanın konusu yazılı ifadelerde ironi tespitinde bulunmaktadır. İnsanların yazılı bir metindeki ironiyi algılayabilmesi bile kolay değilken bunu bir sınıflandırıcı ile bilgisayara yaptırmak hayli zor bir süreçtir. Hele metin madenciliği uygulamaları zor olan Türkçe gibi bir dil üzerinde yine sınıflaması zor olan bir söz sanatını sınıflandırmak hayli güç olacaktır.

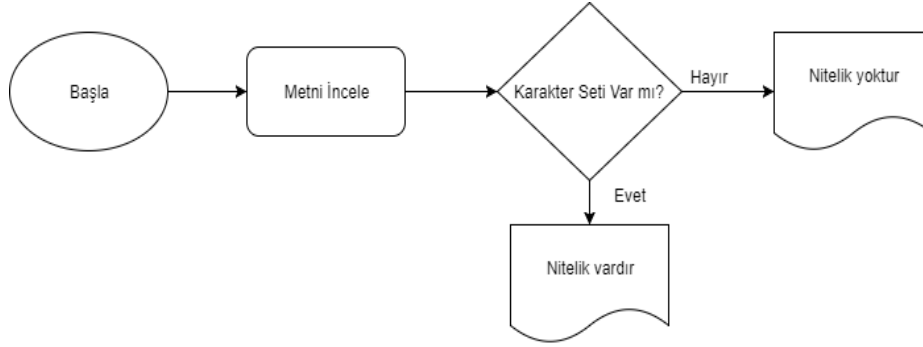
## 3 Çalışmanın Yöntemi

İroni, mantığı itibarıyla söylenen bir ifadenin zıt anlamının kastedilmesi olarak ifade edilebilir. Sözlü ifadelerde insanlar ironi yaptığını mimikleri, yüz ifadeleri ve ses tonuyla karşı tarafa belli edebilir. Yazılı ifadelerde ise böyle bir durum söz konusu olmadığından cümlede ironiyi yakalayabilmek için bazı ipuçlarına ihtiyaç duyulmak-

tadır. Bu ipuçlarından bazıları dilden dile değişmekte, bazıları ise birbirine benzemektedir. Bunun için çalışma yapılacak dilin iyi incelenmesi ve ironi desenlerinin çıkarılması gerekmektedir. Bu ipuçlarından yola çıkarak sürdürülecek olan çalışmamızda öznitelikler belirlenecek ve Türkçe diline uyarlanacaktır. Bu öznitelikler önceden yapılan çalışmalara benzer olmakla beraber Türkçeye uyarlanması ve dilimize özgü desenlerin çıkarılması açısından da özgün olacaktır. Sınıflama için kullanılacak öznitelikler ve detayları şu şekildedir:

### 3.1 Desen Tabanlı Öznitelikler

Bu öznitelik çeşidinde aranan desenin veya karakterin cümlede geçip geçmediğine bakılır. Şekil 1.'de görüldüğü gibi eğer aranan karakter seti cümlede geçiyor ise özniteliği sağlıyor demektir.



Şekil 1. Desen tabanlı öznitelikler için algoritma şeması

**Ünlem İşareti.** Ünlem işareti ironi içeren cümlelerde en çok kullanılan karakterlerden biridir. Yazarlar ifadenin ironi içeren bir cümle olduğunu belirtmek ve bu konuda ipucu vermek için cümle sonuna veya anlam kutupsallığını sağlayan kelimenin sonuna ünlem işareti koyar.

*Örnek.* Ben böyle uslu bir çocuk görmedim!

*Örnek.* Ben böyle uslu (!) bir çocuk görmedim.

*Örnek.* Ben böyle uslu bir çocuk görmedim!!!

Örneklerde görüldüğü gibi cümlede bahsi geçen çocuk şımarıktır ama bu cümlede ironi olduğu için çocuktan tam tersi bir şekilde uslu diye bahsedilmektedir. Cümledeki ironi varlığı ise ünlem işaretlerinden anlaşılmaktadır.

**Soru İşareti.** Soru işareti de ünlem işareti ile beraber kullanıldığında ironi için ipucu olabilir. Bu kullanım ünlem işareti özniteliği kadar çok yaygın olarak kullanılmamaktadır.

*Örnek.* Ben böyle uslu bir çocuk görmedim ?!

**Tırnak İşareti.** Tırnak işareti, cümlelerde bazı kelimeleri vurgulamak için kullanabiliyor. Cümlelerde ironiyi sağlayan kelimeler de tırnak işareti içinde kullanılıp bize sınıflama için ipucu verebiliyor.

*Örnek.* Ben böyle “uslu” bir çocuk görmedim.

**Büyük Harf.** İnternet yorumlarında vurgulanmak istenen bazı kelimeler büyük harfle yazılabilir. İroniyi yapan kişi ironiyi sağlayan kelimeleri büyük harfle vurgulamayı tercih edebiliyor.

*Örnek.* Ben böyle USLU bir çocuk görmedim.

**Ünlem İşareti Üç Nokta.** Ünlem işareti üç nokta karakteri “sözün bittiği yer” anlamında kullanılabilir ve bu durum ironiye işaret olabilir. Cümle ilk bakışta olumlu gözükse de ironik olması sebebiyle olumsuzdur.

*Örnek.* İnanılmaz zevkli bir maç izledik!...

**Emojiler.** Günlük yazı dili ve internet yorumları duygu yoğunluğunu göstermek için pozitif ve negatif emoji içerebilirler. İroni cümleleri de olumlu veya olumsuz duygu yoğunluğu fazla olan yorumlardır. Bu nedenle emojiler ironi sinyali verebilir.

*Örnek.* Böylesine güzel fırtınalı bir hava beklemiyorduk :(

*Örnek.* Ay ne kadar çirkin bir bebek :) Yerim ben bunu.

**Ünlem.** İnternetteki duygu içeren yorumlarda emojiler bulunduğu gibi bazı ünlemlerin de çokça yer aldığı görülmüştür. “haha”, “asdf”, “jkl”, “yeter”, “yuh”, “oley”, “bravo”, “zaa” gibi ünlemler Twitter’da ve bazı mikroblog sitelerinde en çok kullanılan ünlemlerden bazılarıdır.

*Örnek.* Baya ucuz bir ceketmiş, kredi çekip aldım sonunda asdfasdfg.

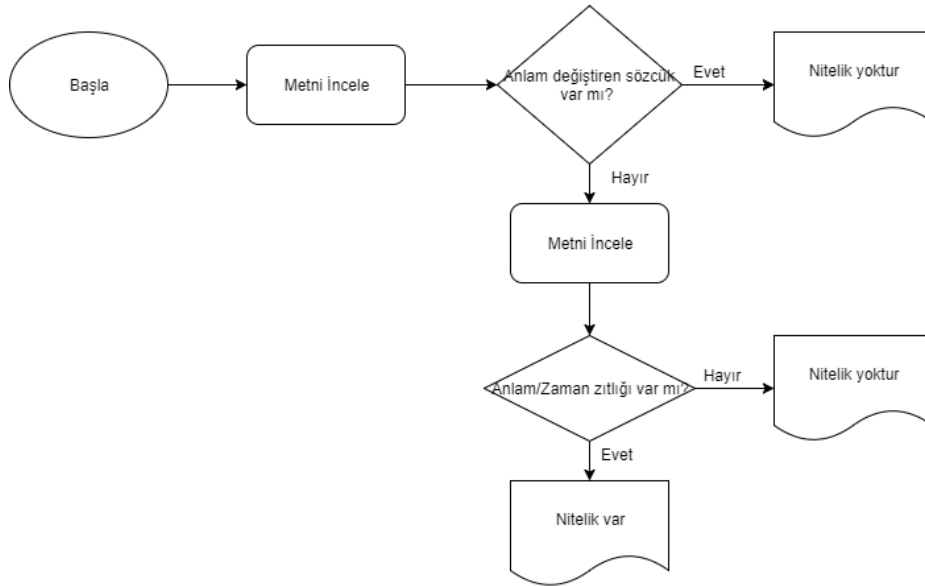
**Abartı.** İroni içeren cümleler olumlu veya olumsuz olarak duyguların keskin ve güçlü ifade edildiği cümlelerdir. Bu yüzden duygular aşırı ve abartı bir şekilde ifade edilebilir.

*Örnek.* Bu olağanüstü (!) manzaraya bak. Her yer çöple dolu.

**Duygusal Anlam.** Metnin duygusal olarak olumlu veya olumsuz olması tek başına ironi varlığı için yeterli olmuyor olsa da, diğer özniteliklerle beraber değerlendirilmesi ve sınıflamada öznitelik olarak kullanılması önemli olabilir.

### 3.2 Polarite Tabanlı Öznitelikler

İroni cümle içindeki zıtlıklardan ve polariteden beslenir. Bir cümle anlam ve zaman olarak polarite içeriyorsa ve bu zıtlık anlam değiştiren bir kelime ile desteklenmiyorsa, bu durum ironi varlığı için güzel bir ipucu olabilir. Şekil 2.'de bu durumun algoritması basit bir şekilde gösterilmiştir.



Şekil 2. Duygusal polarite ve zamansal polarite öznitelikleri için algoritma şeması

**Duygusal Polarite.** İroni söylenen sözün tersinin kastedilmesi amacıyla yapıldığı için bazı ironi örneklerinde zıt duygusal polariteye sahip iki kelime bir arada bulunabiliyor. Bu cümlelerde polariteyi sağlayan iki kelimeyi ayıran ve anlam değiştiren bir bağlaç veya fiil yok ise metin ironi içerebilir gözüyle bakılır.



*Örnek.* Öyle sakın ses çıkarıyordun ki kulağım sağır olacaktı!

Verilen örnekte “sakın” olumlu polariteye, “sağır olma” olumsuz polariteye sahiptir. Cümlede anlam değiştiren bir bağlaç veya fiil de yoktur. Bu yüzden cümle ironi içerir.

*Örnek.* Öyle sakın ses çıkarıyordun ama yine de kulağım sağır olacaktı.

Bu cümlede “ama yine de” bağlacı anlam değiştirdiği için ironi yoktur.

**Zamansal Polarite.** Beklentinin tam tersinin gerçekleştiği bazı durumlarda da ironi yapılabilir. Geçmiş ve şimdiki zaman arasında eylemin beklenenin aksine geliştiği durumda zamansal polarite olabilir.

*Örnek.* Karşısındaki için adalet diyemeyen adam şimdi gelmiş adaletten bahsediyor!

Bahsedilen özneliklerin çıkarımı için Java dilinde kod yazılmıştır. Özneliklerin çıkarımı için Türkçe kelimelerin duygusal puanlarına ihtiyaç duyulmuştur. Bu nedenle SentiTurkNet sözlüğü kullanılmıştır [14]. Ayrıca öznelik çıkarımı için Türkçe dilini morfolojik olarak analiz eden Zemberek adlı kütüphaneden faydalanılmıştır [15]. Özneliklerin çıkarılmasından sonra sınıflama adımı için Weka adlı veri madenciliği programı kullanılmıştır [16].

## 4 Test Sonuçları ve Yorumlar

Veri madenciliğinde test verisinin sayısı ve özneliği çok önemlidir. İnternette ironi içeren cümle arama zorluğu nedeniyle ve bu konuda hazır bir veri seti olmadığı için fazla sayıda test verisi bulunamamıştır. Bulunabilen test verisi 72’si ironi içeren ve 72’si ironi içermeyen Türkçe metin olmak üzere toplamda 144 tanedir. Örnekler 10 katlı çapraz doğrulama yöntemiyle teste dahil edilmiştir. Bu veri setleri Twitter, e-ticaret siteleri ve çeşitli mikrobloglardan toplanmış olup manuel olarak ironi içerip içermemesi yönünden etiketlenmiştir. Türkçe morfolojik olarak incelenmesi zor bir dil olduğundan hazırlanan kodun öznelik çıkarımı hataları kontrol edilip elle düzeltilmiştir. Bu sayede daha başarılı öznelik çıkarımı ve daha başarılı sınıflama sonucu elde edilebilmektedir.

### 4.1 Sınıflandırma Başarısı Ölçümleri

Sınıflandırma başarısını ölçerken Tablo 1.’de gösterildiği gibi 4 tane terim kullanılır. TP; sınıf ataması doğru yapılan örnek sayısı, FP; sınıf ataması yanlış yapılan örnek sayısı, FN; sınıf ataması yanlış reddedilen örnek sayısı ve TN; sınıf ataması doğru

reddedilen örnek sayısı demektir. Ayrıca toplam örnek sayısı; bu dört terimin toplamı anlamına gelmektedir.

**Tablo 1.** Sınıflandırma Başarısı Terimleri

	Örnek Sınıf Etiketleri EVET	Örnek Sınıf Etiketleri HAYIR
Sınıflandırıcı Kararı EVET	TP	FP
Sınıflandırıcı Kararı HAYIR	FN	TN

Sınıflama algoritmalarının başarısını ölçen en temel parametreler doğruluk, hassasiyet (recall), kesinlik (precision) ve F-ölçümüdür. Denklem 1’de görüldüğü gibi hassasiyet; bulunan doğru sonuçların, bulunması gerekene oranıdır. Denklem 2’deki eşitliğe göre ise kesinlik; bulunan doğru sonuçların, bulunan tüm sonuçlara oranı olarak ifade edilir. F ölçümü ise bu iki değerlerin harmonik ortalamasıdır ve Denklem 3’de gösterilmiştir. Son olarak sınıflama başarısı ise doğru sınıflanan örneklerin tüm örneklere oranı demektir ve Denklem 4’teki gibi hesaplanmaktadır.

$$\text{Hassasiyet (recall)} = TP / ( TP + FN ) \quad (1)$$

$$\text{Kesinlik (precision)} = TP / ( TP + FP ) \quad (2)$$

$$F \text{ ölçümü} = 2 \times \text{Hassasiyet} \times \text{Kesinlik} / (\text{Hassasiyet} + \text{Kesinlik}) \quad (3)$$

$$\text{Sınıflama Başarısı} = ( TP + TN ) / \text{Örnek Sayısı} \quad (4)$$

## 4.2 Test Sonuçları

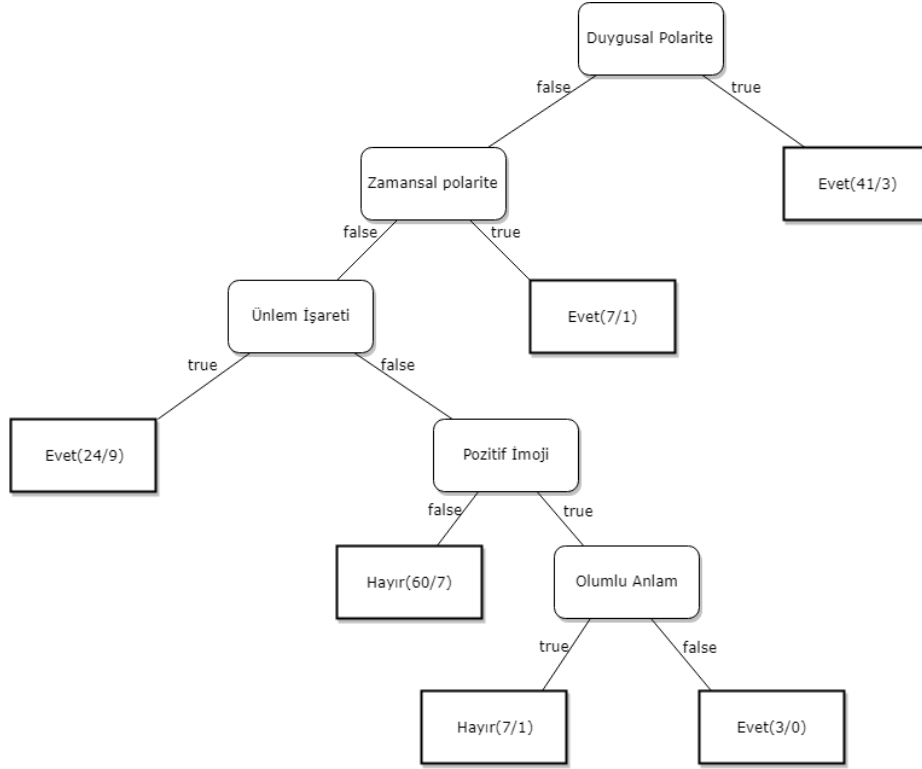
Metinlerin öznetelikleri Weka programı yardımıyla çeşitli sınıflama algoritmaları tarafından sınıflandırılmıştır. Sınıflama algoritması olarak C4.5 (Karar Ağacı), Naive Bayes, Lojistik Regresyon, Karar Tablosu, Rastgele Orman ve Multilayer Perceptron (Çok Katmanlı Algılayıcı) kullanılmıştır. Hesaplamalar sonucu çıkan sınıflama başarısı sonuçları Tablo 2.’de gösterilmiştir.

**Tablo 2.** Test Sonuçları

Algoritmalar	Hassasiyet	Kesinlik	F ölçümü	Sınıflama Başarısı
<b>C4.5 (Karar Ağacı)</b>	0,856	0,854	0,854	85.42 %
<b>Naive Bayes</b>	0,806	0,799	0,797	79.86 %
<b>Lojistik Regresyon</b>	0,841	0,840	0,840	84.03 %
<b>Karar Tablosu</b>	0,819	0,819	0,819	81.94 %
<b>Rastgele Orman</b>	0,882	0,882	0,882	88.19 %
<b>Multilayer Perceptron</b>	0,886	0,882	0,882	88.19 %

Tablo.2'deki sonuçlara göre tüm algoritmalar yaklaşık olarak %80 ile %90 arasında başarı oranına ulaşmıştır. Rastgele Orman ve Multilayer Perceptron (Çok Katmanlı Algılayıcı) algoritmaları %88'lik oran ile en yüksek başarı oranına sahip algoritmalar olmuştur. Bunun yanında hassasiyet, kesinlik ve F ölçümü değerleri de sınıflama başarısı oranlarına benzer bir şekilde çıkmıştır.

Şekil 3.'de gösterilen C4.5 algoritmasının karar ağacı şemasına göre bazı özniteliklerin sınıflamaya katkısı gösterilmiştir. Duygusal polarite özniteliği diğer özniteliklere göre ironi sınıflaması için daha fazla katkı sağlamıştır. Ayrıca ünlem işareti, zamansal polarite öznitelikleri de sınıflamaya yüksek oranda katkıda bulunmuştur. Ünlem işaretinin olmadığı ve pozitif duygunun yoğun olduğu cümlelerde ise ironi olasılığı düşük çıkmıştır.



Şekil 3. C4.5 Karar Ağacının görsel ağaç yapısı

Taşlıoğlu'nun (2014) Türkçe için ironi tespiti çalışmasında biçimsel öznitelikler yani ünlem işareti, üç nokta, tırnak işareti gibi noktalama işaretleri ön plandadır. Kullanılan veri seti de yoğunluklu olarak noktalama işareti içeren ironi cümleleri içermekte-

dir. Dolayısıyla sınıflamaya katkı sağlayan öznitelikler bu noktalama işaretleri olmuştur. Bu çalışma ile karşılaştırılacak olursa; kullanılan test verisinde noktalama işaretleri ile beraber duygusal ve zamansal polarite içeren ironi cümleleri vardır. Bu yüzden sınıflamaya katkı sağlayan özniteliklerde de duygusal ve zamansal polarite öznitelikleri öne çıkmaktadır. Ayrıca Taşlıoğlu'nun (2014) çalışmasında sınıflama başarısı en yüksek olan algoritmalar Rastgele Orman ve K-en Yakın Komşu algoritmaları olmuştur. Yapılan bu çalışmada da Rastgele Orman algoritmasının başarı oranının yüksek çıkması ile bu algoritmanın az sayıda ve çok öznitelikli veri setlerinde başarılı olduğu sonucu çıkarılabilir.

## 5 Sonuç

Yapılan çalışmada Türkçe metinlerde ironi tespiti yapılmıştır. Altı farklı algoritmanın arasından Rastgele Orman ve Multilayer Perceptron (Çok Katmanlı Algılayıcı) algoritmaları %88'lik oran ile en yüksek sınıflama başarısına sahip algoritmalar olmuştur. Duygusal polarite, ünlem, zamansal polarite, pozitif emoji ve duygu öznitelikleri sınıflamaya en çok katkı sağlayan özniteliklerdir. Sonuç olarak Türkçe dilinde metin madenciliği uygulamaları zor bir çalışma alanıdır. Ancak metinlerdeki özniteliklerin iyi belirlenmesi, öznitelik çıkarımının doğru bir şekilde yapılması ve kaliteli veri seti ile başarı oranı artacaktır.

## Kaynaklar

1. Carvalho, P., Sarmiento, L., Silva, M. J., Oliveira, E. D.: Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy" ;-). Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, 53-56 (2009).
2. Maynard, D., Greenwood, M.A.: Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis. LREC (2014).
3. Bouazizi, M., Ohtsuki, T.: Opinion Mining in Twitter How to Make Use of Sarcasm to Enhance Sentiment Analysis. Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM 15 (2015).
4. Sarmiento, L., Carvalho, P., Silva, M. J., Oliveira, E. D.: Automatic creation of a reference corpus for political opinion mining in user-generated content. Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, 29-39 (2009).
5. Reyes, A., Rosso, P., Veale, T.: A multidimensional approach for detecting irony. Language Resources and Evaluation, 47(1), 239-268 (2013).
6. Riloff, E., Qadir, A., Surve, P., Silva, L., Gilbert, N., Huang, R.: Sarcasm as Contrast between a Positive Sentiment and Negative Situation. Conference on Empirical Methods in Natural Language Processing, 704-714 (2013).
7. Vanin, A. A., Freitas, L. A., Vieira, R., Bochernitsan, M.: Some clues on irony detection in tweets. Proceedings of the 22nd International Conference on World Wide Web, 635-636 (2013).
8. Buschmeier, K., Cimiano, P., Klinger, R.: An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews. The 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Baltimore (2014).

9. Barbieri, F., Saggion, H.: Modelling Irony in Twitter. Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics, 56-64 (2014).
10. Rajadesingan, A., Zafarani, R., Liu, H.: Sarcasm Detection on Twitter: A Behavioral Modeling Approach. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 97-106 (2015).
11. Taşlıođlu, H.: Irony Detection On Turkish Microblog Texts, Yüksek Lisans Tezi, Ortaođu Teknik Üniversitesi, Bilgisayar Mühendisliđi, Ankara (2014).
12. Türk Dil Kurumu Sitesi, <http://www.tdk.gov.tr/> . Erişim tarihi: 30.03.2018
13. Bouazizi, M., Otsuki, T.: A Pattern-Based Approach for Sarcasm Detection on Twitter. IEEE Access, 4, 5477-5488 (2016).
14. Dehkharghani, R., Saygin, Y., Yanikoglu, B., Oflazer, K.: SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. Language Resources and Evaluation, 1–19 (2016).
15. Akın, A. A., Akın M. D.: Zemberek, An Open Source Nlp Framework For Turkic Languages. Structure, 10, 1-5 (2007).
16. Weka İnternet Sitesi, <https://www.cs.waikato.ac.nz/ml/weka/>, Erişim tarihi: 31.03.2018