SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
VOL. 3, NO. 3, DECEMBER 2020
DOI: 10.35377/saucis.03.03.735463

On Term Weighting for Spam SMS Filtering

Turgut Doğan[1]
[1]Department of Computer Engineering, Trakya University, Edirne
turgutdogan@trakya.edu.tr

Received 11 May 2020; Revised 4 November 2020; Accepted 14 November 2020; Published 30 December 2020

Abstract

Due to rapid development of the technology, the usage of mobile telephones and short message services (SMS) have become widespread. Thus, the number of spam SMS messages has dramatically increased and the significance of identifying and filtering of suchlike messages raised. Moreover, since they have also risk to steal users' personal information; the problem of identifying and filtering of Spam SMS messages stays popular in terms of also information and data security. In this study, the classification performances of five different term weighting methods on three different datasets containing SMS messages categorized as Spam and legitimate are compared by using two classifiers for corresponding problem. The results obtained showed that reasonable weighting of SMS contents plays an important role in identifying of spam SMS messages. On the other hand, it can be expressed that real classification potential of term weighting schemes reflected betterly the with feature vectors created by using fifty and higher number of terms on especially Turkish and English SMS message datasets. In addition, it has been observed that value ranges of the classification results of obtained from term weighting methods on Turkish SMS message dataset is wider for than ones obtained in English SMS message datasets.

Keywords: term weighting, SMS collection, spam SMS detection, information security

İstenmeyen SMS Filtrelemede Terim Ağırlıklandırma

Öz

Teknolojideki hızlı gelişmeler, mobil telefonların sayısını arttırmış ve kısa mesaj hizmetlerinin (SMS) kullanımını yaygın hale getirmiştir. Bu durum, istenmeyen SMS sayılarını da dramatik bir biçimde arttırmış ve bu tip mesajların belirlenmesi veya filtrelenmesinin önemini arttırmıştır. Ayrıca, kullanıcıların kişisel bilgilerini çalma riski de taşıyabilecekleri için, istenmeyen SMS'lerin filtrelenmesi problemi günümüzde bilgi ve veri güvenliği açısından da popülerliğini korumaktadır. Bu çalışmada, bu probleme yönelik olarak, istenmeyen ve meşru olarak iki sınıfa kategorilendirilmiş SMS mesajlarını içeren üç farklı SMS mesaj veri seti üzerinde 5 farklı popüler terim ağırlıklandırma yönteminin sınıflandırma performansları iki popüler sınıflandırıcı yardımıyla kıyaslanmıştır. Elde edilen sonuçlar; istenmeyen SMS belirleme performansında; SMS içeriklerinin makul bir biçimde ağırlıklandırılmasının önemli bir rol oynadığını göstermiştir. Diğer taraftan, özellikle Türkçe ve İngilizce SMS mesaj verisetleri üzerinde terim ağırlıklandırma şemalarının sahip oldukları potansiyel sınıflandırma performanslarınının elli ve üzeri terim kullanılarak yapılan deneylerde daha iyi yansıtılabildiği ifade edilebilir. Ayrıca Türkçe SMS mesaj veri seti üzerinde terim ağırlıklandırma yöntemlerinden elde edilen sınıflandırma sonuçlarının değer aralıklarının, İngilizce SMS mesaj verisetlerinde elde edilenlere nazaran daha geniş olduğu da gözlenmiştir.

Anahtar Kelimeler: terim ağırlıklandırma, SMS veriseti, istenmeyen SMS filtreleme, bilgi güvenliği

1. Giriş

Teknolojinin gelişimindeki başdöndürücü hız, dünya çapındaki mobil telefon kullanıcılarının iletişim aracı olarak kısa mesaj servislerini (SMS) kullanmaya yönelik ilgisini her geçen gün arttırmaktadır. Bu artış, elektronik posta servislerinde olduğu gibi SMS hizmetlerinde de çeşitli finansal veya kişisel amaçlar taşıyan, kullanıcıların genellikle istenmeyen olarak ifade ettiği SMS mesajlarının sayılarını da kaçınılmaz bir biçimde arttırmıştır. Özellikle son yıllarda, mobil telefonlara ulaştırılan SMS

mesajlarının çoğunluğunu, giyim veya gıda sektörüne ait çeşitli mağazaların indirim duyuruları, internet veya iletişim hizmeti sağlayıcılarının yeni tarife/paket bilgilendirmeleri, bankaların kredi olanakları ile ilgili bilgilendirmeleri gibi kullanıcıyı rahatsız eden türde istenmeyen mesajlar oluşturmaktadır. İstenmeyen SMS mesajlarını bir bölümünü de mobil telefon kullanıcılarının kişisel verilerini maddi veya manevi kazanç sağlamak adına çalmaya çalışan SMS mesajları oluşturmaktadır. Mobil cihazlar kullanıcı adı, şifre ve kredi kartı detayları gibi hassas bilgileri içerdiklerinden; kötü niyetli kişiler iletişim kurmanın en ucuz yollarından biri olan SMS'i kimlik avı saldırılarını gerçekleştirmek için de kullanabilmektedir. SMS tabanlı kimlik avı (Smishing), kullanıcıya yolladığı SMS mesajındaki linki tıklamasını sağlayarak, mobil cihazındaki hassas bilgileri çalmaya çalışan ve günümüzde halen popülerliğini koruyan bir kimlik avı metodudur. Bu tip istenmeyen SMS mesajlarının filtrelenmesi veya engellenmesi kullanıcıların mobil telefonlarındaki bilgilerinin ve verilerinin güvenliğini garanti altına alma açısından önem taşımaktadır.

Günümüzde, kullanıcılar, kendilerine yollanan istenmeyen SMS mesajlarının yol açtığı rahatsızlıkları sahip oldukları akıllı telefonların çeşitli işlevleri aracılığıyla tamamen olmasa da kısmen azaltabilmektedir. Bu işlevler arasında SMS mesajları ile gelen kodları akıllı telefonun mesaj engelleme fonksiyonunda belirterek benzer koda sahip mesajların alınmasına engel olmak ya da sıklıkla istenmeyen mesaj yollayan bir kişi veya numaradan gelen mesajlar için bildirimleri gizleyerek istenmeyen veya bilinmeyen klasörüne düşmesini sağlamak sayılabilir. Bunların dışında kara-liste ve beyaz-liste adı verilen listeler oluşturularak, istenmeyen SMS mesajların yollayan göndericilerin numaraları kara listeye aktarılarak SMS yollaması engellenebilmektedir. Veya istenmeyen SMS mesajlarında sıklıkla yer alan anahtar sözcük listeleri oluşturularak, gelen mesajlarda bu sözcük veya sözcükler mevcutsa söz konusu SMS mesajı istenmeyen SMS olarak filtrelenmektedir. Her ne kadar bunlar istenmeyen mesaj filtreleme için bir çözüm olsa da aslında tek başına yeterli değildir. Çünkü aynı kişi veya numaradan yollanan mesajların tamamı istenmeyen türden olmayabilir. Bazen bankamız tarafından yollanan kredi alternatifleri ile ilgili bilgiler içeren ve istenmeyen türden gibi görünen bir SMS mesajı aslında istenmeyen değil de ilgilenebileceğimiz türden bir içeriğe sahip olabilir. Bu içerik kullandığımız banka kartı için bir şifre veya söz konusu bankanın mobil uygulamasına giriş için tek kullanımlık bir şifre de olabilir. Böyle bir durumda istenmeyen SMS mesajlarını engellemeye çalışırken meşru olan bir SMS mesajına erişememe problemi söz konusu olacaktır. Ayrıca anahtar sözcük listelerine takılmamak için göndericiler bu tip sözcükleri SMS içinde eksik veya yanlış biçimde yazıp filtrelenmekten kurtulabilmektedir. Bunun haricinde, anahtar sözcük listesini sürekli güncellemek de gerektiğinden bu durum sistem daha fazla sistem kaynağı tüketmek anlamına da gelmektedir. Bu nedenle söz konusu SMS mesajlarının içeriklerini daha akıllı bir biçimde işleyebilecek, sınıflandırabilecek ve filtreleyebilecek içerik tabanlı yöntemlere ihtiyaç duyulmuştur.

Literatürde, istenmeyen e-posta belirlemeye yönelik araştırmalar [1-3] kadar çok sayıda olmasa da, istenmeyen SMS mesajlarını etkin bir biçimde belirlemeye veya filtrelemeye yönelik araştırmalar [4-6] da son yıllarda artış göstermiştir. Bu bağlamda, Delany ve arkadaşları istenmeyen SMS filtrelemedeki çalışmaları ve bu alandaki yeni gelişmeleri gözden geçirmiştir [7]. Hidalgo ve arkadaşları, istenmeyen e-posta filtreleme için yaygın bir biçimde kullanılan Bayesian filtreleme tekniklerininin istenmeyen SMS filtreleme için de etkin bir biçimde kullanılabileğini göstermişlerdir [8]. Cormack ve arkadaşları, istenmeyen SMS filtreleme performansının arttırılmasının, e-posta filtreleme metotlarının SMS mesajlarını daha etkin öznitelik gösterimleriyle uyarlanmasına bağlı olduğu hipotezini çeşitli deneylerle desteklemiştir [9]. Almedia ve arkadaşları istenmeyen SMS mesajları içeren geniş bir SMS veriseti üzerinde yaptıkları çalışmada, çeşitli makine öğrenmesi yöntemlerini karşılaştırmış ve Destek Vektör Makinesi (SVM) yaklaşımının diğer yöntemlere nazaran istenmeyen SMS filtrelemede daha başarılı olduğunu belirtmişlerdir [10]. Nuruzzaman ve arkadaşları, mobil telefon üzerinde çalışabilen bir filtreleme sistemi önermiştir [11]. Araştırmacılar, önerilen yaklaşımın eğitim, filtreleme ve güncelleme süreçlerini bağımsız mobil telefon üzerinde gerçekleştirebildiği ve makul doğruluk, minimum bellek tüketimi ve kabul edilebilir bir işlem zamanına sahip olduğunu belirtmişlerdir. Junaid ve Farooq, istenmeyen SMS mesajı içeriğinde yer alıp meşru olan SMS mesajı içeriğinde yer almayan ayırt edici ve özgün öznitelikleri belirlemeye odaklanan bir yaklaşım geliştirmiştir [12]. Söz konusu yaklaşımın gözetimli sınıflandırıcı sisteminde söz konusu ayırt edici öznitelikler işlenerek istenmeyen SMS saptama oranının %89'un üzerine ulaştığını belirtmişlerdir. Uysal ve arkadaşları, istenmeyen SMS

filtreleme için Bilgi Kazancı (Information Gain) ve Ki-Kare (Chi-Squared) öznitelik seçim yöntemlerinden faydalanılarak SMS mesajlarını temsil eden ayırt edici öznitelikleri iki farklı Bayes sınıflandırıcıda kullanan bir şema önermiştir [13]. Söz konusu çalışmada, önerilen filtreleme şemasını kullanan Android tabanlı cep telefonları için gerçek zamanlı bir mobil uygulama da tanıtılmıştır. Yoon ve arkadaşları, mobil iletişimde istenmeyen SMS filtreleme probleminin çözümüne yönelik olarak içerik-tabanlı filtreleme yapan ve göndericinin tepkiye cevap zorluğunu sınayan hibrit bir yaklaşım önermiştir [14]. Söz konusu yaklaşımda, içerik tabanlı filtreleme aşamasında belirsiz olarak sınıflandırılan bir mesaj göndericisine geri yollanarak sınanmakta, hem mesajın istenmeyen mesaj olup olmadığı hem de göndericinin otomatik istenmeyen SMS oluşturucu olup olmadığı kontrol edildiği ifade edilmiştir. Najadat ve arkadaşları Kısa Mesaj Servislerinde uygulanan istenmeyen SMS filtrelemesinin metin sınıflandırma yöntemlerinden birini kullanması gerektiğini belirtikleri araştırmalarında, 12 farklı SMS sınıflandırıcısını incelemiş ve en yüksek doğruluk performansına %98.6 ile SVM sınıflandırıcının ulaştığı belirtilmiştir [15]. Shafi'I ve arkadaşları spam algılama teknikleri, mobil telefonlarda istenmeyen SMS mesajlarının filtrelenmesi ve azaltılması ile ilgili mevcut olan yöntemleri, zorlukları ve gelecekte ne tür çalışmalar yapılabileceğini anlatan bir derleme çalışması gerçekleştirmişlerdir [4]. Kawade ve arkadaşları açık kaynak yazılım Python üzerinde içerik tabanlı makine öğrenmesi yaklaşımını kullarak istenmeyen SMS mesajlarını sınıflandırdığı çalışmasında, istenmeyen SMS mesajlarını doğru olarak belirleme oranının %98 olarak elde edildiğini belirtmiştir [5]. Lee ve arkadaşları istenmeyen SMS filtreleme için derin öğrenme ve kelime gömme yaklaşımlarının ikili sınıflandırma için kullanıldığı çalışmada derin öğrenme yaklaşımının geleneksel SVM algoritmasından daha başarılı olduğu vurgulanmıştır [16]. Jain ve arkadaşları makine öğrenmesi tekniklerini kullanarak Smishing SMS mesajlarını ve istenmeyen SMS mesajlarını belirleyen ve filtreleyen yeni bir yaklaşım önermiştir [17]. Çalışmada, önerilen yaklaşımın yapay sinir ağı sınıflandırıcıda istenmeyen SMS mesajlarını %94.9 doğrulukla algılayabildiğini, Smishing mesajlarını ise %96 doğrulukla filtreleyebildiği ifade edilmiştir.

Bu çalışmada, iki farklı dile ait SMS mesaj filtreleme başarımı üzerinde literatürdeki popüler terim ağırlıklandırma yöntemlerinin etkisi araştırılmıştır. Bu araştırmanın odak noktası az sayıda özniteliğe sahip olan SMS mesaj verilerinin içeriklerinin metin sınıflandırma mekanizması ile işlenmesi ve çeşitli ağırlıklandırma yöntemleri vasıtasıyla ağırlıklandırılmasının doğru olarak sınıflandırılmasını ne derece sağlayabildiği ve dolayısıyla da istenmeyen SMS mesaj filtrelemeye nasıl katkı sağlayabildiğini analiz etmektir. Bir diğer odak noktası da popüler terim ağırlıklandırma şemalarının bu problemin üstesinden gelme kaabiliyeti açısından performanslarını görmektir. Çalışmada 3 farklı SMS mesaj veri seti, 5 farklı terim ağırlıklandırma şeması ve 2 farklı sınıflandırıcı kullanılarak sınıflandırma performansları Makro-F1 cinsinden hesaplanmış ve ilerleyen bölümlerde değerlendirilmiştir.

2. Deneysel Çalışma

Deneysel bölümde alt bölümlerde bilgileri detaylı olarak verilen 3 farklı SMS mesaj veri koleksiyonu üzerinde 5 popüler terim ağırlıklandırma şemasının istenmeyen SMS sınıflandırma performansları 2 farklı sınıflandırıcı kullanılarak hesaplanmış ve kıyaslanmıştır.

2.1. Veri Setleri

British İngilizce SMS Mesaj Veriseti (British English SMS Dataset): 425 adet istenmeyen, 450 adet meşru olmak üzere iki sınıfa ait SMS mesajlarından oluşan bu veriseti Birmingham üniversitesinde gerçekleştirilen bir doktora tez çalışması için oluşturulmuş olup, eklemeli olmayan bir dil karakteristiği gösteren İngilizce için SMS filtreleme çalışmalarında yaygın olarak kullanılmaktadır [11]. Deneysel bölümde, eğitim aşaması için söz konusu verisetinin %70'i (612 SMS mesajı), test aşaması için ise %30'u (263 SMS mesajı) kullanılmıştır.

Türkçe SMS Mesaj Koleksiyonu (Turkish SMS Collection): Sondan eklemeli dillerden Türkçe olarak yazılmış 420 adet istenmeyen ve 430 adet meşru olmak üzere iki sınıfa ait toplamda 850 adet SMS mesajından oluşan bu veriseti, SMS filtreleme çalışamalarının farklı yapıya sahip dillerdeki başarısını

analiz etmek için oluşturulmuştur [18]. 595 SMS mesajının eğitim için, 255 SMS mesajının ise test için kullanıldığı bu verisetinde de eğitim ve test için kullanılan mesaj yüzdeleri sırasıyla %70 ve %30'dur.

İngilizce SMS Mesaj Koleksiyonu (English SMS Collection): İngilizce için bir başka SMS mesaj veri seti olan bu koleksiyonda 4827 meşru ve 747 istenmeyen SMS mesajı mevcuttur [13]. Toplamda 5500'den fazla SMS mesajı içeren bu setin seçilmesinin sebebi, terim ağırlıklandırma şemalarının dengesiz yapıdaki verisetlerindeki SMS sınıflandırma başarımlarını da analiz edebilmektir. Bu veriseti için ise deneysel kısımda 3900 SMS mesajı eğitim için, 1674 SMS mesajı ise test için kullanılmıştır.

2.2. Ön İşleme ve Öznitelik Seçimi

Ön işleme aşamasında üç veri setinden de çıkarılan öznitelikler, sırasıyla dizgelere ayrılmış, küçük harfe dönüştürülmüş, içlerinden her dökümanda sıklıkla geçme ihtimali olan "ve", "veya" gibi durak terimler çıkarılmış ve son olarak köklerine indirgenmiştir.

Yukarıda belirtilen ön işlemlerden geçirildikten sonra elde edilen öznitelikler arasından tekrar tekrar geçenler filtrelenerek, kelime çantası yaklaşımında her özniteliğin yalnızca bir defa temsil edilmesi sağlanmıştır. British İngilizce, Türkçe ve İngilizce SMS mesaj verisetleri için çıkarılan benzersiz öznitelik sayıları sırasıyla 1829, 2142 ve 5172'dir. Öznitelik seçim sürecinde ise, metin sınıflandırma çalışmalarında yaygın olarak kullanılan istatistiksel bir yöntem olan Ki-Kare (Chi-Squared) öznitelik seçim yöntemi kullanılmıştır [19]. Çalışmada Ki-Kare öznitelik seçim yöntemini kullanmaktaki amaç, sınıflandırma başarımının öznitelik boyutu ile ilgisini detaylı bir biçimde analiz etmektir. British İngilizce SMS, Türkçe SMS ve İngilizce SMS mesaj koleksiyonları için sırasıyla 10 ile 1500, 10 ile 2000 ve 10 ile 3000 arasında öznitelik ile istenmeyen SMS sınıflandırma deneyleri gerçekleştirilmiştir.

2.3. Öznitelik/Terim Ağırlıklandırma

Terim ağırlıklandırma, metin sınıflandırma sürecinde metin dokümanları ile içerdikleri öznitelikler/terimler arasındaki ilişkilerin birtakım terim ağırlıklandırma yöntemleri aracılığıyla hesaplandığı ve çoğunlukla kelime çantası yaklaşımı kullanılarak sayısal hale dönüştürüldüğü süreç olarak ifade edilebilir. Bu çalışmada, dokümanlar SMS mesajları olduğundan, seçilen terimlerin ilgili SMS mesajlarının kategorisini ayırt etme derecelerini gösteren ağırlık değerleri tek tek hesaplanmıştır. Ağırlık hesabı için beş farklı terim ağırlıklandırma yöntemi kullanılmış olup, aşağıda her bir yöntemin ağırlıklandırma stratejisinden ve formülünden bahsedilmiştir.

TF-IDF (Terim Frekansı & Ters Doküman Frekansı, Term Frequency & Inverse Document Frequency): TF-IDF, en temel ve popüler terim ağırlıklandırma yöntemlerinden biridir. Terimlerin her bir dokümandaki Terim Frekansı (TF) değerleri ile tüm koleksiyondaki Ters Doküman Frekansı (IDF) değerlerinin çarpımına dayanır [20]. TF-IDF yöntemine göre, tüm koleksiyonda diğer terimlere göre daha az dokümanda/mesajda geçen herhangi bir terim, diğerlerine göre daha yüksek IDF değerine sahiptir. Dolayısıyla, söz konusu terim eşit terim frekansına sahip diğer terimlere nazaran daha yüksek TF-IDF skoru ile ağırlıklandırılır. TF-IDF terim ağırlıklandırma şeması ile herhangi bir t_i teriminin ağırlık hesabı Eşitlik-1'deki gibidir.

$$W_{TF-IDF}(t_i) = TF(t_i, m_k) * \log(\frac{M}{m(t_i)}) \quad (1)$$

Eşitlikte yer alan M ifadesi SMS koleksiyonunda yer alan toplam SMS sayısını, $m(t_i)$ ise t_i teriminin geçtiği toplam SMS mesajı sayısını göstermektedir. Ayrıca diğer terim ağırlıklandırma yöntemlerinin de ağırlıklandırma formülünde yer alan $TF(t_i,m_k)$ ifadesi, t_i teriminin k nolu m mesajındaki frekansını temsil etmektedir.

TF-PB (Olasılık Dağılımlarına Bağlı Terim Ağırlıklandırma, Term Weighting based on Probability Distributions): TF-PB, terim ağrlıklarını hesaplarken ikili (binary) sınıflandırma yaklaşımını esas alan iki farklı oranın çarpımına dayanır [21]. Bu oranlardan biri terimin sınıflar-arası dağılımını diğeri ise sınıf-içi dağılımını ifade etmektedir. İkili sınıflandırma yaklaşımında, bir t_i

teriminin yer aldığı pozitif ve negatif sınıflara (C_j) ait SMS mesajı sayıları sırasıyla a_{ij} ve c_{ij}, söz konusu t_i teriminin yer almadığı pozitif ve negatif sınıflara ait SMS mesajı sayıları ise sırasıyla b_{ij} ve d_{ij} ile ifade ettiğimizi varsayalım. Böyle bir durumda, TF-PB ile terim ağırlıklandırma formülü Eşitlik-2'deki hali alır.

$$W_{TF-PB}(t_i) = TF(t_i, m_k) * \max_{j=1}^{C} \left\{ \log\left(1 + \frac{a_{ij}}{b_{ij}} * \frac{a_{ij}}{c_{ij}} \right) \right\} \tag{2}$$

Eşitlikte yer alan C ifadesi toplam sınıf sayısını, max ifadesi ise, ağırlık ataması yapılırken pozitif ve negatif sınıf için hesaplanan parantez içindeki ağırlık değerlerinden maksimum olanının baz alınacağını ifade etmektedir.

TF-DFS (Terim Frekansı & Ayırt Edici Öznitelik Seçici, Term Frequency & Distinguishing Feature Selector): DFS, ayırt edici özniteliklerin bulunması için önerilmiş olan olasılıksal bir öznitelik seçim yöntemidir [22]. Bu yöntemin TF-DFS adıyla terim ağırlıklandırmaya ilk uyarlanışı; terim frekans faktörünün çeşitli terim ağırlıklandırma şemalarının sınıflandırma performanslarına etkisinin analiz edildiği çalışma ile gerçekleşmiştir. Söz konusu çalışmada TF-DFS çoğu terim ağırlıklandırma şemasından daha üstün bir performans sergilemiştir. Bir t_i terimin TF-DFS ile ağırlıklandırma hesabı aşağıdaki eşitlikteki gibi gerçekleştirilir.

$$W_{TF-\mathrm{DFS}}(t_i) = TF(t_i, m_k) * \sum_{j=1}^{C} \left(\frac{\left(\frac{a_{ij}}{a_{ij} + c_{ij}} \right)}{\left(\frac{b_{ij}}{a_{ij} + b_{ij}} \right) + \left(\frac{c_{ij}}{c_{ij} + d_{ij}} \right) + 1} \right) \tag{3}$$

TF-RF (İlgi Frekansına Bağlı Terim Ağırlıklandırma, Term Weighting based on Relevance Factor): TF-RF ile terim ağırlıklandırma süreci, terimin sınıflar arası dağılımına odaklıdır [23]. Bu da ikili sınıflandırma yaklaşımında söz konusu terimin geçtiği pozitif ve negatif sınıflara ait SMS mesajlarının oranına (a_{ij}/c_{ij}) dayanmaktadır. TF-RF ile terim ağırlığı hesaplama işlemi aşağıdaki formüle göre gerçekleştirilir

$$W_{TF-RF}(t_i) = TF(t_i, m_k) * \max_{j=1}^{C} \left\{ \log\left(2 + \frac{a_{ij}}{\max(1, c_{ij})} \right) \right\} \tag{4}$$

Eşitliğin paydasında yer alan max ifadesi c_{ij} değerinin sıfır olması durumunda sıfıra bölme durumundan kaçınmak için yer almaktadır. Başka bir deyişle, eğer c_{ij} sıfıra eşit olduğunda, payda 1 olarak kabul edilecektir.

TF-IGM (Terim Frekansı & Ters Yerçekimi Momenti, Term Frequency & Inverse Gravity Moment): TF-IGM, Ters Yerçekimi Momentine dayalı olarak ağırlıklandırma yapan yakın zamanda terim ağırlıklandırma için önerilmiş istatistiksel bir modeldir [24]. Ağırlık hesabını ikili sınıflandırma yaklaşımıyla değil, çoklu-sınıflandırma yaklaşımıyla gerçekleştirir. Yani herhangi bir terimin ağırlığına, her bir sınıftaki doküman frekansları hesaba katılarak tek seferde global olarak ulaşılır. TF-IGM ile terim ağırlıklandırma formülüzasyonu aşağıda yer alan Eşitlik-5'teki gibidir.

$$W_{TF-\mathrm{IGM}}(t_i) = TF(t_i, m_k) * \left(1 + \lambda * \overbrace{\left[\frac{f_{i1}}{\sum_{r=1}^{C} f_{ir} * r} \right]}^{IGM(t_i)} \right) \tag{5}$$

Eşitlikte yer alan f_{ir} ifadesi, t_i teriminin r sırasıyla büyükten küçüğe sıralanmış vaziyette olmak üzere pozitif ve negatif sınıflardaki doküman frekanslarını ifade etmektedir. Bu çalışmadaki veri setlerinde toplamda 2 sınıf bulunduğundan r değeri 1 ve 2 olarak payda kısmında hesaplamalara dahil olmuştur. λ ifadesi ise verisetinin dengeli veya dengesiz bir yapıya sahip olma durumu için formülde yer alan, 5.0-9.0 değer aralığına sahip olan ve varsayılan değeri 7.0 olarak belirlenmiş ayarlanabilir bir katsayıyı temsil etmektedir.

2.4. Sınıflandırma ve Değerlendirme

Bu çalışmada, sınıflandırma sürecinde içerik sınıflandırma açısından yaygın olarak tercih edilen Destek Vektör Makineleri (SVM) ve K-En Yakın Komşu (KNN) sınıflandırıcıları kullanılmıştır. SVM sınıflandırıcı doğrusal veya doğrusal olmayan bir hiperdüzlem oluşturarak pozitif örnekleri negatif örneklerden ayırmak için karar sınırını belirleyen, hem ikili hem de çok-sınıflı sınıflandırmaya uygun popüler bir makine öğrenmesi algoritmasıdır [25]. SVM çok yüksek boyutlu sınıflandırma çalışmaları için dahi tutarlı bir biçimde sınıflandırma yapabilme kabiliyetine sahiptir. KNN sınıflandırıcı ise nispeten daha basit bir çalışma yapısına sahip olan ve sınıflandırma problemlerinde yaygın olarak kullanılan bir sınıflandırma algoritmasıdır [25]. Algoritması, en basit anlatımıyla test aşamasında gelen herhangi bir SMS mesajınının sınıfını belirlerken, kendisine en benzer k adet komşusunun sınıfını baz almaktadır ve hangi sınıf daha çoğunlukta ise, söz konusu test mesajı o sınıfa atanır. Deneysel bölümde SVM sınıflandırıcı varsayılan parametrelerle çalıştırılmış olup, çok-sınıflı sınıflandırmayı destekleyen LibSVM paketi kullanılarak deneyler gerçekleştirilmiştir [26]. KNN sınıflandırıcı için en iyi sınıflandırma performansını veren k değerleri dataset bazında belirlenmiş ve deneysel sonuçlar bölümünde gösterilmiştir.

Sınıflandırma sonuçlarının değerlendirilmesinde literatürde yaygın olarak kullanılan değerlendirme metriği *Makro-F_1* ölçüm metriği tercih edilmiştir. *Makro-F_1* ölçütü Eşitlik-6'daki gibi hesaplanmaktadır.

$$Macro - F_1 = \frac{\sum_{k=1}^{C} F_k}{C} \quad F_k = \frac{2 * p_k * r_k}{p_k + r_k} \tag{6}$$

Eşitlikteki p_k ve r_k ifadeleleri sırasıyla, k.ncı sınıf için kesinlik (precision), hatırlama (recall) değerlerini ifade etmektedir. Bu çalışmada toplamda iki sınıf yer aldığından, formüldeki C değeri 2'dir. Söz konusu p_k ve r_k değerleri ise aşağıdaki gibi hesaplanmaktadır.

$$p_k = \frac{tp_k}{tp_k + fp_k} \quad r_k = \frac{tp_k}{tp_k + fn_k} \tag{7}$$

Eşitlik-7'de yer alan tp_k ifadesi, gerçekte k sınıfına ait olan ve doğru olarak sınıflandırılmış mesaj sayısını, fp_k gerçekte k sınıfına ait olan ve yanlış olarak sınıflandırılmış mesaj sayısını, son olarak fn_k ise gerçeke k sınıfına ait olmadığı halde yanlış olarak sınıflandırılmış mesaj sayısını ifade etmektedir. Makro-F_1 ölçütünde koleksiyon içerisinde yer alan her sınıf için F ölçümü gerçekleştirilip ortalaması alınmaktadır. Bu nedenle dengesiz metin veya SMS koleksiyonlarının yer aldığı sınıflandırma problemlerinde başarım ölçümü için daha adil bir seçim olarak nitelendirilebilir.

3. Deneysel Sonuçlar

3.1. British İngilizce SMS Mesaj Veriseti Üzerindeki Sınıflandırma Sonuçları

British İngilizce SMS Mesaj veriseti üzerinde toplamda 5 farklı terim ağırlıklandırma yönteminden KNN ve SVM sınıflandırıcılar kullanılarak elde edilen *Makro-F_1* sonuçları Tablo 1 ve Tablo 2'de sırasıyla verilmiştir. Tablolarda, ilgili sınıflandırıcı için söz konusu terim ağırlıklandırma şemalarından elde edilen en yüksek *Makro-F_1* değeri kalın biçimde ifade edilmiştir.

Tablo 1 British İngilizce SMS Mesaj Verisetinde KNN (k=3) Sınıflandırıcı ile Elde Edilen *Makro-F_1* Sonuçları

Öznitelik Sayısı	**TF-IDF**	**TF-PB**	**TF-DFS**	**TF-RF**	**TF-IGM**
10	90.84	88.49	90.84	88.49	90.84
50	93.91	88.17	93.53	88.17	93.91
100	94.29	87.43	93.14	87.43	94.29
300	90.49	87.41	93.53	87.80	93.53
500	92.00	87.02	93.53	87.42	**94.67**
1000	91.60	84.69	92.38	85.13	93.14
1500	90.87	84.69	92.01	85.13	91.63

Tablo 2 British İngilizce SMS Mesaj Verisetinde SVM Sınıflandırıcı ile Elde Edilen *Makro-F_1* Sonuçları

Öznitelik Sayısı	**TF-IDF**	**TF-PB**	**TF-DFS**	**TF-RF**	**TF-IGM**
10	89.65	86.07	88.87	86.87	90.04
50	93.51	86.07	90.79	87.66	91.97
100	93.88	86.07	92.73	88.45	90.84
300	90.44	86.07	**94.64**	89.27	91.97
500	91.58	86.07	94.26	87.68	92.35
1000	90.78	86.07	93.48	88.45	93.49
1500	91.55	86.07	92.70	88.45	93.09

Tablolar incelendiğinde, British İngilizce SMS Mesaj veriseti üzerinde TF-IGM, TF-DFS ve TF-IDF terim ağırlıklandırma yöntemlerinin her iki sınıflandırıcı ile de diğer yöntemlerden nispeten daha üstün performans sergilediğini söylemek münkündür. KNN sınıflandırıcı için en yüksek sınıflandırma değeri TF-IGM ile SVM için ise TF-DFS ile elde edilmiştir. Genel olarak, TF-RF ile TF-PB'nin sınıflandırma performansları her iki sınıflandırıcı üzerinde de diğer şemalara nazaran daha düşüktür. Ayrıca TF-PB terim ağırlıklandırma şemasının SVM sınıflandırıcıdaki performansı tüm öznitelik boyutlarında da değişim göstermemiştir. Bunun sebebi için, sınıflandırma işlevlerindeki boyut artışından SVM sınıflandırıcının KNN sınıflandırıcıya nazaran daha az etkilenmesinden kaynaklı olduğu yorumu yapılabilir. Ayrıca bu verisetinde üzerinde elde edilen tüm sınıflandırma performansları yaklaşık % 84-94 bandında elde edilmiştir.

3.2. Türkçe SMS Mesaj Veriseti Üzerindeki Sınıflandırma Sonuçları

Türkçe SMS Mesaj veriseti üzerinde toplamda 5 farklı terim ağırlıklandırma yönteminden KNN ve SVM sınıflandırıcılar kullanılarak elde edilen *Makro-F_1* sonuçları Tablo 3 ve Tablo 4'te sırasıyla verilmiştir.

Tablo 3 Türkçe SMS Mesaj Verisetinde KNN (k=1) Sınıflandırıcı ile Elde Edilen *Makro-F_1* Sonuçları

Öznitelik Sayısı	**TF-IDF**	**TF-PB**	**TF-DFS**	**TF-RF**	**TF-IGM**
10	89.40	79.26	89.40	79.26	89.40
50	94.11	85.83	93.72	85.03	93.32
100	**95.69**	86.63	93.72	86.22	94.11
300	93.72	89.39	94.11	90.17	95.29
500	92.15	89.39	94.51	90.17	93.73
1000	90.98	91.36	90.98	90.95	92.55
1500	89.02	91.36	91.76	90.95	92.55
2000	91.37	91.36	92.94	90.95	94.12

Tablo 3 ve 4 incelendiğinde, öznitelik sayısı 10 ile 50 iken her bir terim ağırlıklandırma yönteminden elde edilen sınıflandırma performansları arasındaki farkların yüksek olduğu görülmektedir. Bu durum için, söz konusu ağırlıklandırma yöntemleri vasıtasıyla ağırlıklandırılmış doküman terim matrisinin yöntemlerin sahip olduğu potansiyelin 10 öznitelik ile yeterince yansıtılamadığı değerlendirmesi yapılabilir.

Tablo 4 Türkçe SMS Mesaj Verisetinde SVM Sınıflandırıcı ile Elde Edilen *Makro-F_1* Sonuçları

Öznitelik Sayısı	**TF-IDF**	**TF-PB**	**TF-DFS**	**TF-RF**	**TF-IGM**
10	85.61	79.26	85.61	79.26	86.59
50	92.11	82.99	91.75	82.99	92.11
100	**95.29**	82.99	94.10	86.61	93.31
300	94.50	81.34	94.50	86.54	93.70
500	93.71	81.34	95.29	86.54	93.31
1000	92.91	81.34	94.49	88.57	93.70
1500	93.30	81.34	94.88	88.57	93.70
2000	93.70	81.34	94.89	88.57	92.91

Her iki sınıflandırıcı için de en yüksek sınıflandırma performansını TF-IDF göstermiştir. Türkçe SMS Mesaj verisetinde, TF-IDF ile TF-DFS yüksek boyutlarda ve SVM sınıflandırıcı ile KNN sınıflandırıcıya nazaran daha başarılıdır. Benzer şekilde, TF-RF ile TF-PB terim ağırlıklandırma yöntemlerinin KNN sınıflandırcı ile sınıflandırma performansları SVM sınıflandırıcı ile elde edilen değerlerden daha yüksektir. Sınıflandırıcı bazında kıyaslama yapılırsa, TF-PB ve TF-RF'in KNN sınıflandırıcı ile elde edilen sınıflandırma başarımlarının SVM sınıflandırıcı ile elde edilen sınıflandırma başarımlarından genel olarak daha yüksek olduğunu söylemek mümkündür. Terim ağırlıklandırma yöntemlerinin performanslarının yüksek boyutlarda genel bir kıyaslaması yapılırsa, KNN sınıflandırıcı ile genel olarak TF-IGM'in daha üstün performanslara sahip olduğu, SVM sınıflandırıcı ile ise TF-DFS'in diğer yöntemlere göre net bir şekilde daha yüksek Makro-F1 değerlerine sahip olduğu görülmektedir. Türkçe SMS mesaj verisetinde KNN ve SVM sınıflandırıcılar üzerinde sırasıyla TF-RF ve TF-PB terim ağırlıklandırma şemalarının sınıflandırma performansları diğer 4 şemanın gerisinde kalmıştır. Ayrıca söz konusu veriseti üzerinde her iki sınıflandırıcı ile de tüm terim ağırlıklandırma şemalarından elde edilen *Makro-F_1* değerleri yaklaşık %79-95 aralığında hesaplanmıştır.

3.3. İngilizce SMS Mesaj Veriseti Üzerindeki Sınıflandırma Sonuçları

British İngilizce SMS Mesaj veriseti üzerinde toplamda 5 farklı terim ağırlıklandırma yönteminden KNN ve SVM sınıflandırıcılar kullanılarak elde edilen *Makro-F_1* sonuçları Tablo 5 ve Tablo 6'da sırasıyla verilmiştir.

Tablo 5 İngilizce SMS Mesaj Verisetinde KNN (k=1) Sınıflandırıcı ile Elde Edilen *Makro-F_1* Sonuçları

Öznitelik Sayısı	**TF-IDF**	**TF-PB**	**TF-DFS**	**TF-RF**	**TF-IGM**
10	85.47	85.50	85.47	85.50	85.64
50	93.16	91.86	93.42	92.15	93.58
100	93.37	**93.63**	92.79	93.27	93.56
300	90.89	92.61	91.68	92.07	92.50
500	90.28	92.88	90.36	91.43	91.53
1000	89.30	92.76	88.42	91.38	90.24
1500	91.83	92.50	91.71	91.23	92.44
2000	92.24	92.62	92.53	91.70	93.20
2500	92.62	92.50	92.91	91.55	93.20
3000	91.94	92.38	93.20	91.46	93.20

İngilizce SMS Mesaj verisetinde en yüksek Makro-F1 değeri, KNN sınıflandırıcı üzerinde TF-PB'ye, SVM sınıflandırıcı üzerinde ise TF-DFS'e aittir. SVM sınıflandırıcı tüm öznitelik boyutları hesaba katılarak terim ağırlıklandırma şemalarının performansları arasında bir kıyaslama yapılırsa; TF-DFS terim ağırlıklandırma yönteminin performanslarının diğerlerine nazaran genel olarak daha üstün olduğu, TF-PB'nin ise daha düşük olduğu yorumu yapılabilir. Söz konusu yöntemlerin KNN sınıflandırıcı ile ise performansların birbirlerine daha yakın olduğu söylenebilir. Bu verisetinde de sadece 10 öznitelik ile çalışmak terim ağırlıklandırma yöntemlerinin sahip oldukları gerçek sınıflandırma potansiyelini tam olarak yansıtamamıştır. Öznitelik sayısı 50'ye çıkarıldığında daha yüksek ve dolayısıyla da daha tutarlı sınıflandırma sonuçları elde edilmiştir. SVM sınıflandırıcı ile 500 öznitelikten sora TF-PB'nin Makro-F1 değerleri bu veri setinde de sabitlenmiş olup öte yandan KNN sınıflandırıcı ile ise TF-IGM ve TF-PB'nin performansları diğerlerine nazaran nispeten daha öne çıkmıştır. Söz konusu SMS mesaj

verisetinde tüm terim ağırlıklandırma şemalarından elde edilen sınıflandırma başarımlarının yüzdelik değerleri 85-95 bandında ölçülmüştür.

Tablo 6 İngilizce SMS Mesaj Verisetinde SVM Sınıflandırıcı ile Elde Edilen *Makro-F_1* Sonuçları

Öznitelik Sayısı	**TF-IDF**	**TF-PB**	**TF-DFS**	**TF-RF**	**TF-IGM**
10	85.77	84.73	85.77	84.73	85.77
50	92.82	90.36	92.92	92.41	92.26
100	93.82	90.70	93.18	93.11	93.96
300	94.07	89.86	94.93	94.16	93.79
500	93.65	90.32	94.89	94.21	92.30
1000	93.98	90.32	94.89	93.77	93.53
1500	94.31	90.32	95.72	93.47	94.85
2000	94.47	90.32	95.72	93.66	93.20
2500	94.75	90.32	95.48	92.822	92.55
3000	94.51	90.32	**95.76**	92.822	92.80

4. Sonuç ve Öneriler

Bu çalışmada, İngilizce ve Türkçe dilleri için istenmeyen SMS mesajlarının belirlenmesi probleminin çözümünde popüler terim ağırlıklandırma yöntemlerinin etkileri ayrıntılı olarak analiz edilmiştir. Söz konusu etki analizinde SMS mesajlarının doğru olarak sınıflandırılmasında terim ağırlıklandırma şemalarının katkılarına odaklanılmıştır. Deneysel kısımda istenmeyen ve meşru olarak kategorize edilmiş SMS mesajlarını içeren Türkçe ve İngilizce dillerine ait üç farklı veri seti, beş farklı popüler terim ağırlıklandırma şeması ile SVM ve KNN sınıflandırıcı kullanılmış olup söz konusu terim ağırlıklandırma şemalarının sınıflandırma performansları *Makro-F_1* ölçütü cinsinden hesaplanmıştır. İstenmeyen SMS mesaj filtreleme problemi için bu çalışmada yararlanılan üç veri setinden elde edilen sonuçlar; TF-IGM, TF-DFS ve TF-IDF terim ağırlıklandırma şemalarının, TF-PB ve TF-RF terim ağırlıklandırma şemalarına nazaran nispeten daha başarılı olduğunu göstermiştir. Ancak büyük resme odaklanacak olunursa, kullanılan terim ağırlıklandırma yöntemlerinin hiçbirinin genel anlamdaki etkinliğinin her bir veri seti için de birbirinden açık ara üstün olmadığı yorumunu yapmak yanlış olmaz. Her bir terim ağırlıklandırma şemasından elde edilen sınıflandırma sonuçlarındaki farklar, istenmeyen SMS sınıflandırmada mesaj içeriklerinin uygun bir biçimde ağırlıklandırılmasının ne denli önemli olduğuınu da göstermektedir. Dil bazında bir kıyaslama yapılırsa, Türkçe SMS mesaj veri setinde elde edilen tüm sınıflandırma başarımlarının İngilizce SMS mesaj verisetlerinde elde edilenlere nazaran daha geniş değer aralığına sahip olduğu gözlenmiştir. Çalışmada elde edilen bulgular, Türkçe ve İngilizce dilleri gibi sırasıyla sondan eklemeli olma ve olmama özellikleri gösteren başka dillerde yapılacak çalışmalara ışık tutabilir.

Referanslar

[1] H. Faris, I. Aljarah, and B. Al-Shboul, "A hybrid approach based on particle swarm optimization and random forests for e-mail spam filtering," in *International Conference on Computational Collective Intelligence*, 2016: Springer, pp. 498-508.

[2] R. Varghese and K. Dhanya, "Efficient feature set for spam Email filtering," in *2017 IEEE 7th International Advance Computing Conference (IACC)*, 2017: IEEE, pp. 732-737.

[3] M. Diale, T. Celik, and C. Van Der Walt, "Unsupervised feature learning for spam email filtering," *Computers & Electrical Engineering*, vol. 74, pp. 89-104, 2019.

[4] M. A. Shafi'I et al., "A review on mobile SMS spam filtering techniques," *IEEE Access*, vol. 5, pp. 15650-15666, 2017.

[5] K. O. Kawade and K. S. Oza, "Content-based SMS spam filtering using machine learning technique," *International Journal of Computer Engineering and Applications*, vol. 7, p. 4, 2018.

[6] T. H. Apandi and C. A. Sugianto, "Analisis Komparasi Machine Learning Pada Data Spam Sms," *Jurnal TEDC*, vol. 12, no. 1, pp. 58-62, 2019.

[7] S. J. Delany, M. Buckley, and D. Greene, "SMS spam filtering: Methods and data," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9899-9908, 2012.

[8] J. M. Gómez Hidalgo, G. C. Bringas, E. P. Sánz, and F. C. García, "Content based SMS spam filtering," in *Proceedings of the 2006 ACM Symposium on Document Engineering*, 2006, pp. 107-114.

[9] G. V. Cormack, J. M. G. Hidalgo, and E. P. Sánz, "Feature engineering for mobile (SMS) spam filtering," in *Proceedings of the 30th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007, pp. 871-872.

[10] T. Almeida, J. M. G. Hidalgo, and T. P. Silva, "Towards sms spam filtering: Results under a new dataset," *International Journal of Information Security Science*, vol. 2, no. 1, pp. 1-18, 2013.

[11] M. T. Nuruzzaman, C. Lee, and D. Choi, "Independent and personal SMS spam filtering," in *2011 IEEE 11th International Conference on Computer and Information Technology, 2011: IEEE*, pp. 429-435.

[12] M. B. Junaid and M. Farooq, "Using evolutionary learning classifiers to do MobileSpam (SMS) filtering," in *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, 2011, pp. 1795-1802.

[13] A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal, "A novel framework for SMS spam filtering," in *2012 International Symposium on Innovations in Intelligent Systems and Applications*, 2012: IEEE, pp. 1-4.

[14] J. W. Yoon, H. Kim, and J. H. Huh, "Hybrid spam filtering for mobile communication," *Computers & Security*, vol. 29, no. 4, pp. 446-459, 2010.

[15] H. Najadat, N. Abdulla, R. Abooraig, and S. Nawasrah, "Mobile sms spam filtering based on mixing classifiers," *International Journal of Advanced Computing Research*, vol. 1, pp. 1-7, 2014.

[16] H.-Y. Lee and S.-S. Kang, "Word Embedding Method of SMS Messages for Spam Message Filtering," in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2019: IEEE, pp. 1-4.

[17] A. K. Jain, S. K. Yadav, and N. Choudhary, "A Novel Approach to Detect Spam and Smishing SMS using Machine Learning Techniques," *International Journal of E-Services and Mobile Applications (IJESMA)*, vol. 12, no. 1, pp. 21-38, 2020.

[18] A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal, "The impact of feature extraction and selection on SMS spam filtering," *Elektronika ir Elektrotechnika*, vol. 19, no. 5, pp. 67-73, 2013.

[19] Y.-T. Chen and M. C. Chen, "Using chi-square statistics to measure similarities for text categorization," *Expert systems with applications*, vol. 38, no. 4, pp. 3085-3090, 2011.

[20] K. Sparck Jones, "A Statistical Interpretation of Term Specificity and Its Application in Retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11-21, 2004

[21] Y. Liu, H. T. Loh, and A. Sun, "Imbalanced text classification: A term weighting approach," *Expert Systems with Applications*, vol. 36, no. 1, pp. 690-701, 2009

[22] A. K. Uysal and S. Gunal, "A novel probabilistic feature selection method for text classification," *Knowledge-Based Systems*, vol. 36, pp. 226-235, 2012

[23] M. Lan, C. L. Tan, J. Su, and Y. Lu, "Supervised and traditional term weighting methods for automatic text categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 721-735, 2009.

[24] K. Chen, Z. Zhang, J. Long, and H. Zhang, "Turning from TF-IDF to TF-IGM for term weighting in text classification," *Expert Systems with Applications*, vol. 66, pp. 245-260, 2016

[25] T. Dogan and A. K. Uysal, "Improved inverse gravity moment term weighting for text classification," *Expert Systems with Applications*, vol. 130, pp. 45-59, 2019.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.