Spam Filtering Using a Logistic Regression Model Trained by an Artificial Bee Colony Algorithm Makale İncelemesi

Selin Cansu Akbaş
Ankara
Maltepe/Çankaya
191180005
selincansuakbas@gmail.com

Ümmü Nur Gülmez
Ankara
Maltepe/Çankaya
191180762
ummunurgulmez@gmail.com

Eylül Dalkıran
Ankara
Maltepe/Çankaya
191180030
dalkiraneylul@gmail.com

ÖZET

Bu makale, yapay arı koloni algoritması ile lojistik regresyon sınıflandırma modelini birleştiren yeni bir spam tespit yöntemini ele almaktadır. Önerilen model, mevcut spam tespit tekniklerinin düşük tespit oranları ve yüksek boyutlu verileri işlemedeki yetersizlik ve etkisizlik gibi kısıtlarını ele almayı amaçlamaktadır. Çalışma, Naive Bayes, Destek Vektör Makineleri (SVM) ve Lojistik Regresyon (LR) gibi spam tespitinde yaygın olarak kullanılan makine öğrenimi yöntemlerinin kapsamlı bir incelemesini içermektedir. Ayrıca Enron, CSDMC2010 ve TurkishEmail veri setleri dahil olmak üzere halka açık veri setleri üzerinde bu yöntemlerin performansı analiz edilmiştir. Önerilen modelin sınıflandırma doğruluğu açısından diğer spam tespit tekniklerinden daha iyi performans gösterdiği öne sürülmektedir.

Makale, veri temsili için ön işleme adımlarını, özellik seçiminin nasıl yapıldığını ve sınıflandırma için kullanılan makine öğrenimi algoritmalarını özetlemektedir. Çalışma, Multinomial ve Gaussian Naive Bayes dahil olmak üzere Naive Bayes sınıflandırıcıları tarafından yapılan sınıflandırmanın sonuçlarını sunar ve özellik vektör boyutunun ve yumuşatma parametrelerinin sınıflandırma doğruluğu üzerindeki etkisini vurgular. Ayrıca, çalışma, Lineer ve Radyal Baz Fonksiyonu (RBF) SVM dahil olmak üzere Destek Vektör Makinesi (SVM) sınıflandırıcıları tarafından yapılan sınıflandırmayı inceler ve kutu kısıtlama parametrelerinin ve gamma değerlerinin sınıflandırma doğruluğu üzerindeki etkisini araştırmaktadır.

Ayrıca makale gradyan iniş algoritması ile eğitimi yapılmış bir Lojistik Regresyon (LR) sınıflandırıcısı tarafından yapılan sınıflandırmayı değerlendirmekte; öğrenme hızı ve düzenleme parametrelerine odaklanmaktadır. Ayrıca LR'nin Yapay Arı Koloni (ABC) algoritması ile eğitilen yeni oluşturulmuş, önerilen model tarafından yapılan sınıflandırmanın sonuçlarını sunar. Bu sonuçlar modelin öğrenme hızına, düzenleme parametrelerine, özellik vektör boyutuna ve ABC algoritmasının kontrol parametrelerine duyarlılığını içermektedir. Ek olarak, çalışma, önerilen modelin Multinomial NB, Gaussian NB, Lineer ve RBF SVM ve LR gibi diğer yöntemlerle performans karşılaştırmasını TurkishEmail, CSDMC2010 ve Enron veri setleri üzerinde içerir.

Özetle, makale ABC algoritmasını lojistik regresyon ile birleştiren yeni bir spam tespit modelinin geliştirilmesine ilişkin bilgiler sağlamakta ve performansını birden fazla veri seti boyunca mevcut diğer yöntemlerle kapsamlı bir şekilde karşılaştırmaktadır. Önerilen model, yüksek boyutlu verileri işlemede ve sınıflandırma doğruluğu açısından diğer spam tespit tekniklerinden daha iyi sonuçlar göstererek umut verici sonuçlar ortaya koymaktadır.

Anahtar Kelimeler - Spam Tespiti, Spam Filtreleme, Lojistik Regresyon, E-Posta

1. GİRİŞ

Son yıllarda, internet teknolojisinin de etkisiyle günlük iletişimlerimiz değişti. Elektronik posta (e-posta) kavramı günümüzde iletişim aracı olarak yaygın bir şekilde kullanılmaktadır. Bu teknoloji, birçok kişiye aynı anda kolay ve uygun maliyetli bir şekilde ulaşmayı mümkün kılar. Ancak, birçok kullanıcı istekleri dışında e-postalar alır. Spam posta, genellikle reklam amacıyla binlerce alıcıya gönderilen bu iletileri tanımlamak için kullanılan bir terimdir. E-posta spam'leri, internet kullanıcılarının karşılaştığı en can sıkıcı sorunlardan biridir ve bu istenmeyen iletiler, alıcıları rahatsız ederek zamanlarını boşa harcamalarına neden olmaktadır. Spam algılama sistemleri, bu tür istenmeyen postaları filtrelemek ve kullanıcılara daha temiz bir iletişim deneyimi sunmak için yaygın olarak kullanılmaktadır. Ancak mevcut spam algılama teknikleri, düşük tespit oranları ve yüksek boyutlu verilerle etkin bir şekilde başa çıkma zorluğu gibi sorunlarla karşılaşmaktadır.q

Bu çalışmada, e-posta spamlerini tespit etmek için daha etkili bir yöntem önerilmektedir. Yeni yaklaşım, yapay arı kolonisi algoritması ile lojistik regresyon sınıflandırma modelini birleştirmektedir. Bu kombinasyon, yüksek boyutlu verileri işleme kapasitesi ile bilinen spam tespit zorluklarına karşı çıkma amacını taşımaktadır. Yapay arı kolonisi algoritması, doğadan ilham alarak geliştirilmiş bir optimizasyon algoritmasıdır ve bu çalışmada, spam tespiti için özelleştirilerek kullanılmaktadır. Lojistik regresyon modeli ise, bu algoritma ile eğitilerek spam ve spam olmayan e-postaları sınıflandırmak üzere tasarlanmıştır. Geliştirilen bu yaklaşımın temel amacı, spam filtreleme sistemlerinin daha etkili hale getirilmesidir. Hızla gelişen spam tekniklerine karşı dirençli bir model elde edilerek, geleneksel spam filtreleme yöntemlerinin ötesinde bir performans sağlanması hedeflenmektedir [1].

Bu makalede [2], Naive Bayes (NB), Destek Vektör Makineleri (SVM) ve Lojistik Regresyon (LR) gibi üç yaygın makine öğrenimi yöntemi incelenmiştir.

Naive Bayes (NB): NB sınıflandırıcıları, her bir özelliğin birbirinden bağımsız olduğu naif bir varsayıma dayanır. Bu sınıflandırıcılar, özellikle spam filtreleme gibi uygulamalarda tercih edilir, çünkü basitlikleri, hızlı eğitim süreçleri ve düşük hesaplama karmaşıklığı ile bilinir.

Destek Vektör Makineleri (SVM): SVM, veri noktalarını bir hiper düzlemle bölen ve sınıfları ayırmaya çalışan bir öğrenme algoritmasıdır. Marjinal yaklaşımıyla bilinen SVM, metin sınıflandırma, görüntü tanıma ve biyomedikal uygulamalarda etkili bir performans sergiler.

Lojistik Regresyon (LR): LR, sınıflandırma problemleri için kullanılan bir istatistiksel modeldir. Giriş verilerini bir logit fonksiyonu aracılığıyla sınıflara atar ve çıktıları olasılık değerleri olarak sunar. İki sınıflı sınıflandırma problemlerinde sıkça tercih edilen bir yöntemdir ve tıbbi teşhis, pazarlama analitiği gibi alanlarda başarıyla kullanılır.

Her bir yöntem, kendine özgü avantajlara sahip olup, problem bağlamına göre seçilmelidir. Bu yöntemler, makine öğrenimi alanında geniş bir uygulama yelpazesi sunar ve farklı problemlere çözüm getirebilirler. Bu sınıflandırıcıların genelde kullanımı basit ve genelleştirilebilir olsa da, bir dizi kısıtlama ile karşılaşırlar. Boyutlanabilirlik sorunları, yüksek hesaplama maliyetleri, yanlış sınıflandırma oranları, özellik ağırlıklarına hassasiyet, gerçek uygulamalarda düşük işlem hızları ve aşırı uydurma veya yerel minimumlarda takılma gibi sorunlar bu kısıtlamalara örnektir. Bu sınıflandırıcıların performansı, belirli parametrelere ve problem türüne bağlıdır.

2. YÖNTEMLER VE MATERYALLER

Deneyler

Bu çalışmada, spam e-posta filtrelemesi için kullanılan yöntemlerin performansını değerlendirmek amacıyla çeşitli deneyler yapılmıştır. Bu deneyler, farklı veri setleri üzerinde gerçekleştirilmiş ve spam filtreleme algoritmalarının etkinliği çeşitli ölçütler kullanılarak değerlendirilmiştir

Veri Setlerinin Kaynakları ve İçeriği

Bu çalışmada kullanılan veri setleri, spam e-posta filtrelemesi için özel olarak toplanmış ve yaygın olarak kullanılan, kamuoyuna açık veri kaynaklarından elde edilmiştir. Bu veri setleri, internet üzerindeki çeşitli forumlardan, e-posta hizmet sağlayıcılarından ve açık veri platformlarından toplanmıştır. Her bir veri seti, geniş bir e-posta çeşitliliğini kapsar şekilde tasarlanmıştır ve hem spam hem de spam olmayan (legitimate) e-postaları içerir. Spam kategorisindeki e-postalar, ticari reklamlar, istenmeyen toplu gönderimler ve potansiyel olarak zararlı içerikleri içerirken, spam olmayan e-postalar günlük iletişim, iş yazışmaları ve abonelik bilgileri gibi gerçek kullanıcı içeriklerinden oluşmaktadır.

Veri Setlerinin Özellikleri ve Seçim Kriterleri

Seçilen veri setleri, spam filtreleme tekniklerinin çeşitli koşullar altında nasıl performans gösterdiğini değerlendirmek için özenle seçilmiştir. Bu veri setleri, farklı dillerde yazılmış e-postaları, çeşitli biçim ve formatlardaki içerikleri ve değişik gönderici profillerini içerecek şekilde tasarlanmıştır. Bu çeşitlilik, modelin gerçek dünya koşullarına daha iyi uyum sağlamasını ve geniş bir senaryo yelpazesinde etkili olmasını sağlamaktadır. Ayrıca, veri setlerinin büyüklüğü ve karmaşıklığı, modelin yüksek boyutlu verilerle başa çıkma yeteneğini test etmek için önemli bir faktördür.

Veri Temsili İçin Ön İşleme

Spam filtreleme modellerini eğitmek ve test etmek için verilerin uygun bir biçimde temsil edilmesi gerekmektedir. Bu amaçla, ham e-posta verileri üzerinde çeşitli ön işleme adımları gerçekleştirilmiştir. Bu adımlar arasında tokenizasyon, gereksiz karakterlerin temizlenmesi, durdurma kelimelerinin (stop words) çıkarılması ve köklenme (stemming) gibi işlemler bulunmaktadır. Bu ön işleme adımları, veri setinin model tarafından daha etkili bir şekilde işlenmesini ve daha iyi sonuçlar elde edilmesini sağlamak için önemlidir.

Özellik Seçimi

Spam filtreleme modellerinin başarısında önemli bir rol oynayan özellik seçimi bu çalışmada dikkatle gerçekleştirilmiştir. Seçilen özellikler, modelin eğitimi ve testi sırasında kullanılmıştır.

Logistic Regression (LR) Modeli

Bu çalışmada, spam filtreleme için bir logistic regression modeli kullanılmıştır. Logistic regression, veri setindeki her bir özelliğin (feature) ağırlığını belirleyerek, e-postaları spam ve spam olmayan olarak sınıflandırmak için bir logistic aktivasyon fonksiyonu kullanır. Bu model, e-posta sınıflandırılmasında etkili sonuçlar göstermiş ve diğer yöntemlerle karşılaştırıldığında rekabetçi sonuçlar üretmiştir.

Modelin Matematiksel Temeli

Logistic regression modeli, veri noktalarının ait olduğu sınıfın olasılığını tahmin etmek için bir lojistik fonksiyon kullanır. Bu fonksiyon, girdi özelliklerinin doğrusal bir kombinasyonunu alır ve çıktıyı [0,1] aralığında bir olasılık değerine dönüştürür. Bu olasılık, belirli bir e-postanın spam olma ihtimalini temsil eder. Modelin öğrenme süreci sırasında, gerçek etiketlerle tahmin edilen etiketler arasındaki farkı en aza indirecek şekilde özellik ağırlıkları ayarlanır.

Eğitim Süreci ve Optimizasyon

Logistic regression modelinin eğitimi, veri setinin bir kısmını kullanarak gerçekleştirilir. Bu süreçte, modelin ağırlıkları, verilen eğitim veri seti üzerindeki performansını en iyi hale getirecek şekilde ayarlanır. Eğitim sürecinde yaygın olarak kullanılan bir yöntem olan gradyan inişi (gradient descent), modelin hata oranını minimize edecek şekilde ağırlıkları iteratif olarak günceller. Bu optimizasyon süreci, modelin veri setindeki karmaşık ilişkileri öğrenmesini ve daha doğru sınıflandırmalar yapmasını sağlar.

Artificial Bee Colony (ABC) Algoritması

Bu çalışmada, logistic regression modelinin eğitimi için yapay arı kolonisi (ABC) algoritması kullanılmıştır. ABC algoritması, doğadan ilham alınarak oluşturulmuş bir swarm zeka algoritmasıdır ve bal arılarının yiyecek arama davranışını simüle eder. ABC algoritması, çok modlu ve yüksek boyutlu problemlerde iyi performans gösterir ve spam sınıflandırması için dengeli bir keşif ve sömürü yeteneğine sahiptir.

Modelin Uygulanması ve Değerlendirilmesi

Logistic regresyon modeli, spam ve spam olmayan e-postaları ayırt etmek için eğitildikten sonra, test veri seti üzerinde değerlendirilir. Bu değerlendirme süreci, modelin gerçek dünya verileri üzerindeki performansını test etmek için kritik öneme sahiptir. Modelin başarısı, doğruluk, hassasiyet, duyarlılık ve F1 skoru gibi çeşitli metrikler kullanılarak ölçülür. Bu metrikler, modelin hem spam hem de spam olmayan e-postaları ne kadar etkili bir şekilde sınıflandırıldığını gösterir.

3. DENEY SONUÇLARI

Çalışmanın deneysel bulguları, Naive Bayes (NB), Destek Vektör Makineleri (SVM) ve Lojistik Regresyon (LR) gibi çeşitli makine öğrenimi algoritmalarının, TürkEmail ve CSDMC2010 veri setleri kullanılarak spam filtrelemedeki performansının değerlendirilmesini içerir. Deneyler, farklı sınıflandırıcıların eğitimi ve test edilmesini, sınıflandırıcıların çeşitli parametrelere duyarlılığının analiz edilmesini ve önerilen modelin sınıflandırma doğruluklarının diğer mevcut yöntemlerle karşılaştırılmasını içermektedir. Önerilen model, her iki veri setinde de diğer algoritmaları geride bırakarak yüksek sınıflandırma doğruluğu elde etmiştir. TürkEmail veri setinde %99.25 ve CSDMC2010 veri setinde %98.70 başarı oranına ulaşmıştır. Çalışma ayrıca, önerilen modelin performansını önceki çalışmalarda kullanılan son teknoloji yöntemlerle karşılaştırarak spam filtrelemedeki etkinliğini ortaya koymuştur. Ayrıca, çalışma, özellikle özel ön işleme teknikleri gerektiren Türkçe veri setleri olmak üzere spam filtrelemede dil farklılıklarını ve karmaşıklığını dikkate almanın önemini vurgulamıştır. Makaledeki nicel sonuçlar, çeşitli sınıflandırıcıların TürkEmail ve CSDMC2010 veri setleri üzerindeki sınıflandırma doğruluklarını içerir. Örneğin, Gaussian NB sınıflandırıcısı, TürkEmail veri setinde 500 özellik vektör boyutu için %95.38 ve 1000 özellik vektör boyutu için %96.63 sınıflandırma doğruluğu elde etmiştir. Ayrıca, SVM sınıflandırıcısı, RBF çekirdeği ile CSDMC2010 veri setinde farklı parametre ayarları için %68.21 ile %98.91 arasında doğruluklar elde etmiştir. Makale ek olarak ANOVA testlerinin sonuçlarını sunar ve her iki veri seti için de farklı gruplar ve parametreler arasında sınıflandırma doğruluklarında istatistiksel olarak anlamlı farklılıklar olduğunu gösterir. Dahası, önerilen model, her iki veri setinde de diğer algoritmaları geride bırakarak yüksek sınıflandırma doğruluğu elde etmiştir. TürkEmail veri setinde %99.25 ve CSDMC2010 veri setinde %98.70 başarı oranına ulaşmıştır. Bu sayısal sonuçlar, çeşitli sınıflandırıcıların ve önerilen modelin spam tespiti bağlamındaki performansına dair bilgiler sağlar.

4. SONUÇLAR

Ana Bulguların Özeti

Bu çalışma, e-posta spam tespiti için geliştirilen yenilikçi bir yaklaşımın etkinliğini değerlendirmiştir. Yapay arı kolonisi algoritması ile güçlendirilmiş lojistik regresyon modeli, spam filtreleme alanında önemli bir ilerleme sağlamıştır. Bu model, yüksek boyutlu ve karmaşık veri setleri üzerinde yüksek doğruluk oranları ve düşük sahte pozitif oranları ile dikkat çekici sonuçlar elde etmiştir. Geleneksel spam filtreleme yöntemlerine kıyasla, önerilen modelin daha hızlı ve etkin bir şekilde spam tespiti yapabildiği gözlemlenmiştir.

Yapay Arı Kolonisi Algoritmasının Rolü

Yapay Arı Kolonisi Algoritması, lojistik regresyon modelinin eğitimi sırasında önemli bir rol oynamıştır. Bu algoritma, modelin ağırlıklandırılmış özelliklerini optimize ederek sınıflandırma kabiliyetini artırmıştır. Algoritmanın doğa esinli yapısı, karmaşık ve çok boyutlu veri setlerinde etkinlik sağlamak için kritik olmuştur. Bu özellik, özellikle yüksek boyutlu veri setlerinde spam tespitinde geleneksel yöntemlere göre büyük bir avantaj sağlamıştır, çünkü bu algoritma modelin daha kapsamlı ve doğru bir şekilde eğitilmesine imkan tanımıştır. Bu entegrasyon, modelin spam e-postaları daha doğru ve hızlı bir şekilde tespit etmesine olanak tanıyarak spam filtreleme tekniklerinde önemli bir ilerleme olarak kabul edilmektedir. Bu yaklaşımın geliştirilmesi, spam tespit teknolojilerinin daha da gelişmesini sağlayarak, spam e-postaların sürekli evrimine karşı daha etkili bir çözüm sunabilir.

Lojistik Regresyon Modelinin Başarısı

Lojistik regresyon modeli, basit yapısı ve hızlı sınıflandırma yeteneği ile ön plana çıkmıştır. Model, eğitim sürecinde yapay arı kolonisi algoritması tarafından desteklenerek, spam ve spam olmayan e-postaları etkin bir şekilde ayırt etmiştir. Modelin yüksek performansı, gerçek zamanlı uygulamalar için ideal bir çözüm sunmuş ve spam filtreleme sistemlerinin genel etkinliğini artırmıştır.

Uygulamada Karşılaşılan Zorluklar ve Gelecek Çalışmalar

Bu çalışmada, modelin eğitimi ve testi sırasında bazı zorluklarla karşılaşılmıştır. Özellikle, modelin farklı türde spam e-postalara adaptasyonu ve yeni spam tekniklerine karşı direnci, gelecekteki çalışmaların odak noktası olmalıdır. Ayrıca, modelin daha geniş veri setleri ve gerçek dünya senaryolarında test edilmesi, onun genel uygulanabilirliğini ve dayanıklılığını artıracaktır.

Sonuçların Değerlendirilmesi

Sonuç olarak, bu çalışma, spam filtreleme tekniklerinde önemli bir gelişme sunmuş ve yapay arı kolonisi algoritması ile güçlendirilmiş lojistik regresyon modelinin, spam tespitinde etkili bir çözüm olduğunu göstermiştir. Ancak, spam e-postaların sürekli evrimi ve yeni türlerinin ortaya çıkması, bu alanda sürekli araştırma ve gelişmeye ihtiyaç duyulduğunu ortaya koymaktadır. Gelecekte, bu modelin daha da geliştirilmesi ve yeni teknolojilerle entegre edilmesi, spam tespitinde daha üstün sonuçlar elde edilmesini sağlayacaktır.

5. GELECEK ÇALIŞMALAR

Daha Güçlü Derin Öğrenme Modelleri

Gelecekteki spam filtreleme çalışmalarında, özellikle daha güçlü derin öğrenme modellerine odaklanmak, önemli bir araştırma alanı olacaktır. Derin öğrenme, karmaşık veri yapılarını analiz etme ve öğrenme kapasitesi ile bilinir, ancak spam tespiti için bu modellerin daha da geliştirilmesi gerekmektedir. Bu bağlamda, derin öğrenme modellerinin daha geniş ve çeşitli veri setleri ile eğitilmesi, özellikle spam e-posta verilerine odaklanarak model performansını artırmak için stratejilerin uygulanması önemli bir adım olacaktır. Bu gelişmiş modeller, spam filtreleme sistemlerinin daha hassas, etkili ve geniş kapsamlı hale gelmesine olanak tanıyabilir. Lojistik regresyon modeli derin öğrenme modelleriyle karşılaştırılabilir. Özellikle, derin sinir ağlarının karmaşıklık ve öğrenme kapasitesi avantajları göz önüne alındığında, bu modellerin performansını değerlendirmek önemli olabilir.

Kullanıcı Geri Bildirimine Dayalı Güncellemeler

Kullanıcı geri bildirimlerini dikkate alan kişiselleştirilmiş spam filtreleme stratejileri geliştirilebilir. Kullanıcı geri bildirimleri model güncellemelerine dahil edilebilir. Kullanıcıların pozitif veya negatif sonuçları rapor etmelerine izin veren bir geri bildirim mekanizması eklenebilir ve bu geri bildirimleri de modelin sürekli olarak geliştirilmesi için kullanılabilir. Kullanıcıların spam tanımları üzerinde daha fazla kontrol sahibi olmalarına izin veren bir model güncelleme arayüzü eklenebilir. Bu da kullanıcıların spam filtreleme modelini kendi tercihlerine göre daha hassas bir şekilde ayarlamalarına olanak tanıyabilir.

Çoklu Dil Desteği

Kültürel farklılıkları ele almak ve spam filtreleme modelini daha geniş bir kullanıcı kitlesine uyarlamak için önemli bir çalışma alanıdır. Dil tabanlı özellikler eklenerek, modelin çoklu dilde daha etkili olması sağlanabilir. Kültürel veri setleri kullanılarak, farklı kültürlerden gelen spam desenleri üzerinde modelin performansı artırılabilir.

Ansambl Modelleri ve Stacking Teknikleri

Birden çok modelin birleştirilerek daha güçlü bir spam filtreleme sistemi oluşturmasını hedefler. Farklı öğrenme algoritmalarını içeren ansambl modelleri, model çeşitliliğini artırabilir ve spam tespitindeki doğruluğu optimize edebilir. Ansambl modelleri, birden çok zayıf veya temel modelin birleştirilerek daha güçlü bir modelin elde edilmesini hedefler. Bu modeller, farklı öğrenme algoritmalarını veya aynı algoritmayı farklı alt özellik setleriyle kullanarak model çeşitliliğini artırır. Ansambl modelleri genellikle daha yüksek doğruluk, daha iyi genelleme ve daha dirençli modeller elde etmek amacıyla kullanılır. Stacking, farklı temel modellerin tahminlerini bir üst düzey modelin girdisi olarak kullanarak daha güçlü bir model elde etme stratejisidir. Temel modeller, genellikle farklı öğrenme algoritmalarını veya parametre ayarlarını içerir. Stacking, farklı modellerin birleştirilerek birbirlerinin zayıflıklarını dengeleyerek daha genel bir çıkarım elde etmeye olanak tanır. Ancak, dikkatlice ayarlanmış ve optimize edilmiş olması gereken bir tekniktir. Ayrıca, daha karmaşık bir model olması nedeniyle eğitim süreci daha uzun sürebilir.

Doğal Dil İşleme ve Semantik Analiz

Doğal Dil İşleme (NLP) tekniklerinin spam içeriğini daha etkili bir şekilde anlamak için kullanılabilir. Semantik analiz ve kelime bağlamını değerlendiren modeller geliştirilebilir.

Gizlilik ve Güvenlik Önlemleri

Spam filtreleme modelinin güvenilirliğini sağlamak açısından kritiktir. E-posta içeriğinin gizliliğini korumak için şifreleme ve diğer güvenlik önlemleri güçlendirilebilir. Aynı zamanda, modelin güvenlik açıklarına karşı direncini artırmak ve olası saldırılara karşı koruma sağlamak için güvenlik analizleri gerçekleştirilebilir.

Bu öneriler, "Spam Filtering Using a Logistic Regression Model Trained by an Artificial Bee Colony Algorithm" başlıklı makalenin gelecekteki araştırmalarına ışık tutacak şekilde genişletilebilir. Her bir öneri, spam filtreleme sistemlerinin daha etkili, güvenilir ve kullanıcı dostu olması yolunda önemli katkılar sağlayabilir.

6. REFERANSLAR

[1] Özgür, L., Güngör, T., & Gürgen, F. (2004). Adaptive anti-spam filtering for agglutinative languages: a special case for Turkish. *Pattern Recognition Letters*, *25*(16), 1819-1831.

[2] Dedeturk, B. K., & Akay, B. (2020). Spam filtering using a logistic regression model trained by an artificial bee colony algorithm. *Applied Soft Computing*, *91*, 106229.

Sunum Video Linki

https://drive.google.com/file/d/1sfIDW-ejgvjymS9HZHR6oZYtfIfeFpW3/view?usp=sharing