TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ TABANLI TWITTER SOSYAL BOT TESPİT SİSTEMLERİNİN PERFORMANSLARININ DEĞERLENDİRİLMESİ

YÜKSEK LİSANS TEZİ

Muhammet Buğra TORUSDAĞ

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ali Aydın SELÇUK

ARALIK 2020

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, alıntı yapılan kaynaklara eksiksiz atıf yapıldığını, referansların tam olarak belirtildiğini ve ayrıca bu tezin TOBB ETÜ Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlandığını bildiririm.

.

Muhammet Buğra TORUSDAĞ

ÖZET

Yüksek Lisans Tezi

MAKİNE ÖĞRENMESİ TABANLI TWITTER SOSYAL BOT TESPİT SİSTEMLERİNİN PERFORMANSLARININ DEĞERLENDİRİLMESİ

Muhammet Buğra Torusdağ

TOBB Ekonomi ve Teknoloji Üniveritesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Ali Aydın Selçuk

Eş Danışman: Mücahid Kutlu

Tarih: Aralık 2020

Twitter gibi sosyal medya platformları insanların rahat bir şekilde iletişim kurabilmesi için oldukça etkili mecralardır. Bu platformlar hayatı kolaylaştırmak adına birçok avantaj sağlamasına rağmen, insanların kandırılması, yanlış bilgi yayılarak insanların yanlış yönlendirilmesi, manipüle edilmesi ve sözlü taciz gibi birçok soruna da neden olmaktadır.. Özellikle sosyal botlar, bahsedilen zararlı içeriklerin hızlı bir şekilde yayılması ve daha görünür hale gelebilmesi adına sürekli olarak içerek paylaşarak bu aktivitelerin gerçekleştirilmesini daha kolay bir hale getirmektedir. Bu durumu engelleyebilmek adına sosyal bot tespit sistemleri geliştirilmiştir. Buna rağmen, geliştirilen sistemlerin performansları, veri setlerinin sınırlı sayıda ve türde bot hesap bulundurmasından dolayı tam doğru bir şekilde değerlendirilememektedir. Bundan dolayı, bu tezde yapılan çalışmalarda bot tespit sistemlerinin performanslarının doğru bir şekilde değerlendirilebilmesi ve sosyal bot tespiti probleminin çözülüp çözülemediği araştırılmaktadır. Yapılan deneyler ile, 4 farklı bot tespit sisteminin performansları farklı deney düzenekleri üzerinde karşılaştırılarak en yüksek performansa sahip model bulunmaya çalışılmıştır. Kullanılan modellerin orijinal çalışmalarında raporlanan skorların çok yüksek olduğu görülmesine rağmen, modeller farklı test setlerinde düşük performanslar

göstermişlerdir. Buna rağmen, performansı en yüksek olan modelin Botometer olduğu anlaşıldığından, Botometer'ın performansının daha detaylı incelenmesi gereksinimi ortaya çıkmıştır. Farklı bir bakış açısıyla deneyler gerçekleştirilerek Botometer'ın performansı 5 farklı bot senaryosu kullanılarak değerlendirilmiştir. Bu deneyler sonucunda, Botometer'ın yalnızca 1 senaryo dışında tüm senaryolarda kötü performans sergilediği görüldüğünden, sosyal bot tespiti probleminin hala araştırılmaya açık bir problem olduğunu anlaşılmaktadır.

Anahtar Kelimeler: Twitter, Makine öğrenmesi, Sosyal bot tespiti, Performans değerlendirme, Botometer, Bot senaryoları.

ABSTRACT

Master of Science

EVALUATION OF MACHINE LEARNING BASED TWITTER SOCIAL BOT DETECTION SYSTEMS

Muhammet Buğra Torusdağ

TOBB University of Economics and Technology
Institute of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Prof. Ali Aydın Selçuk

Co-Supervisor: Mücahid Kutlu

Date: December 2020

Social media platforms such as Twitter, provide an incredibly effective way to communicate with people. While these platforms have many benefits, they can also be used for deceiving people, spreading misinformation, manipulation and verbal harassment. Social bots are usually employed for these kind of activities to artificially increase the amount of a particular post. To mitigate the effects of social bots, many bot detection systems are developed. However, the evaluation of these methods are challenging due to lack limited available datasets and the variety of bots people might develop. Therefore, in this thesis, it has been investigated whether the performance of bot detection systems can be accurately evaluated and the social bot detection problem is solved. The experiments carried out, the performances of 4 different bot detection systems are compared on different experimental setups to find which model has the highest performance. Although it was observed that the scores in the original studies where the models were very high, the models showed poor performance in different test sets. However, since it is understood that the model with the highest performance is the Botometer, a more detailed examination of the performance of the Botometer should be done. Experiments were carried out with a

different perspective and the performance of the Botometer is evaluated using 5 different bot scenarios. As a result of these experiments, Botometer shows low performance in all scenarios except one and the problem of social bot detection is still an open problem to investigate.

Keywords: Twitter, Machine learning, Social bot detection, Evaluation, Botometer, Bot scenarios.

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren hocalarım Prof. Dr. Ali Aydın SELÇUK ve Dr. Mücahid KUTLU'ya, kıymetli tecrübelerinden faydalandığım TOBB Ekonomi ve Teknoloji Üniversitesi Bilgisayar Mühendisliği Bölümü öğretim üyelerine, eğitimim boyunca bana burs veren TOBB Ekonomi ve Teknoloji Üniversitesi'ne ve destekleriyle her zaman yanımda olan aileme ve arkadaşlarıma çok teşekkür ederim.

İÇİNDEKİLER

ŞEKİL LİSTESİ

Sayfa

ÇİZELGE LİSTESİ

Sayfa

KISALTMALAR

ACC : Doğruluk (Accuracy)
PRE : Kesinlik (Precision)
REC : Duyarlılık (Recall)
F1 : PRE ve REC'in harmonik ortalaması
AUC : Eğrinin Altındaki alan (Area Under the Curve)
OSA : Online Sosyal Ağ
API : Programlanabilir Uygulama Arayüzü (Application Programmable Interface)
SVM : Destek Vektör Makinesi (Support Vector Machine)
LSTM : Uzun Kısa Süreli Bellek **(**Long Short-Term Memory)
CAP : Bütüncül Otomasyon Olasılığı (Complete Automation Probability)
LCS : En Uzun Ortak Alt Dizi (Longest Common Subsequence)

RESİM LİSTESİ

Sayfa

1. GİRİŞ

Twitter ve Facebook gibi Online Sosyal Ağ (OSA) platformları birçok insan tarafından sıklıkla kullanıldığından günlük yaşamın ayrılmaz bir parçası haline gelmiştir. OSAlar insanlarla iletişim kurmak, herhangi bir konudaki düşüncelerin özgür bir biçimde edilebilmesi ve anlık olarak dünyadaki tüm olayların takip edilebilmesi gibi birçok özelliğe sahip olduklarından son derece kullanışlıdırlar. Sahip olduğu avantajların yanı sıra, OSAlar yanlış bilgi paylaşımı ile insanların fikirlerinin manipüle edilmesi (politik görüşler, borsa hisseleri üzerinde manipülasyon vb.), siber zorbalık (tehdit, aşağılayıcı ifade kullanımı vb.) zararlı yazılımların paylaşılması ile özel bilgilerin çalınması (kredi kartı bilgileri, hesap şifreleri vb.) gibi insanlara hem maddi hem de manevi yönden zarar verebilecek paylaşımların yapıldığı platformlar haline dönüşmüştür.

OSAlar'da zararlı paylaşımlar yapılması yalnızca gerçek insan hesapları tarafından gerçekleştirilmemektedir. Sosyal botlar adı verilen ve otomatik bir şekilde insana benzer şekilde zararlı aktivitelerde bulunması için programlanmış bilgisayar yazılımları, değişik şekillerde mesajlar paylaşarak (Twitter için hasthtag, link ve multimedya içeren paylaşımlar gibi) ve platformlara özgü etkileşimlerde bulunarak (Twitter için diğer kullanıcıların takip edilmesi, tweetlerinin retweet edilmesi veya beğenilmesi gibi) varolan zararlı içerik paylaşımı problemini farklı bir boyuta getirmektedirler. Suni bir şekilde içeriklerin artırılmaya çalışılmasının farklı amaçları bulunmaktadır. Örnek olarak Twitter üzerinde belirli bir konuyu gündeme getirebilmek amacıyla aynı hashtagin farklı hesaplar tarafından defalarca paylaşılması konuyu trend konular arasına sokarak dünya üzerinde daha fazla konuşulur bir hale getirebilir. Ayrıca, gerçek bir hesap (örneğin, bir ünlü veya politikacı) birçok bot hesabın takibiyle birlikte daha fazla ün ve politik etki alanına sahip olabilir. Ek olarak, sosyal botlar politik bir duruşa sahip olacak şekilde tasarlanarak (bir siyasi parti destekçisi olmak gibi), popüleritesini artırmak istedikleri hesaplar ile etkileşeme geçebilir ve bu hesapların paylaşımlarının daha fazla etkileşim ile daha ulaşılabilir bir hale gelmesini sağlayabilirler. Bu şekilde

paylaşımlar bir doğruluğa sahip olmasalar dahi insanlar yanlış yönlendirilerek siyasi görüşleri manipüle edilebilir.

Sosyal botların günlük yaşamlarımıza potansiyel olarak verebileceği zararlara karşın OSA platformları bot olarak tespit ettikleri hesapları teste tabi tutmak (CAPTCHA gibi) veya silmek gibi eylemlerde bulunmaktadırlar. Örneğin, Twitter yaptıkları paylaşımların propaganda içerdiğini tespit ettiği binlerce sosyal botu silebilmektedir. Buna rağmen, bot hesapların tespit işlemi manuel olarak zaman alıcı ve karar verilmesi zor bir işlem olduğundan, tespit edilememiş veya insanlar tarafından bot olduğu tespit edilmiş binlerce hesap aktif olarak propaganda yapmakta ve insanları manipüle edebilmektedir. Ek olarak, Twitter kurallar ve kısıtlamalar ile bot hesapların faaliyetlerinin önüne geçmeye çalışmaktadır. Örneğin, aynı tweetin gün içerisinde birden fazla kere atılmasını engelleme kuralı ile otomatik bir şekilde aynı içeriğin paylaşımının önüne geçilmeye çalışılmıştır. Buna karşın içerikte ufak değişiklikler bile bu kuralın atlatılmasını sağladığından bot hesapların kural bazlı bir şekilde tespitinin zor olduğu ve çözümün yeni bir bakış açısıyla bulunması gerektiği anlaşılmıştır. Bu sebeplerden dolayı, manuel ve kural bazlı tespitin sahip olduğu zorluklar, araştırmacıların farklı bir bakış açısıyla birlikte bot hesapların otomatik bir şekilde tespit edilmesine yönelik çalışmalarda bulunmalarına neden olmuştur.

Turing Test, bir insanı bir bilgisayardan ayırt etmek amacıyla iki tarafa da sorular sorarak gerçekleştirilir ve temel olarak, otomatize edilmiş bot tespit sistemlerinin mantığında da Turing Test vardır. Bu sistemlerde kullanıcılara Turing Test'teki gibi sorular sormak yerine, kullanıcının OSA platformunda gerçekleştirdiği aktiviteler, yaptığı paylaşımlar ve platform içerisinde nasıl bir sosyal ağ oluşturduğuna bakılarak bir karar işlemi uygulanır.

Başarılı bir şekilde gerçekleştirilmesi ve başarılı sonuçlar alınması zor olarak gözüken Turing Test, bot tespit sistemlerinin makine öğrenmesi veya derin öğrenme yöntemleriyle otomatize edilmesiyle birlikte araştırmalarda çok yüksek başarı skorlarına ulaşabilmektedir. Örneğin, Kudugunta ve Ferrara makalesinde [15] Area Under Curve skoru 0.99 olarak rapor edilmektedir, AUC skoru Davis'in [11] modelinde ise 0.95 olarak gösterilmektedir. Ayrıca, Cresci 0.95'ten yüksek F-Measure skoru rapor ederek çok yüksek başarı skorlarına ulaşabilmektedir. Geliştirilen modellerden elde edilen skorlar bot tespit problemini neredeyse tamamen

çözülmüş olarak gösterse de OSA kullanıcıları hala sosyal botların yaptığı yanıltıcı ve zararlı aktivitelere maruz kalmaktadırlar. Son yapılan araştırmalarda ise Twitter üzerinde birçok aktif botun bulunduğu ve Covid-19 ile alakalı yapılan paylaşım yapanların yarısının sosyal botlar olduğunu ortaya koymaktadır [21].

Bot tespit problemini neredeyse çözülmüş olarak gösterecek kadar yüksek olan skorlar, modellerin aşırı öğrenme problemiyle karşılaştığının ve modeller eğitilirken kullanılan veri setlerinin kısıtlı ve az sayıda tipte sosyal bota sahip olduğunun bir göstergesi olabilir. Bundan dolayı Yang et. al. [10] makalesinde çeşitli veri setleri ile yapılan deneyler ve Bölüm 6'daki deneyler ile Yang'ın modeline [11] yapılan senaryo bazlı incelemeler literatürde bulunan en başarılı bot tespit sistemlerinin bile en basit bot hesapları bile tespit edemediğini göstermektedir.

Bu çalışmamızda, ilk olarak botların aktivitelerine karşı geliştirilmiş olan ve yüksek başarı skorları elde etmiş makine öğrenmesi temelli bot tespit sistemleri için, etraflı bir şekilde ve daha önce yapılmamış boyutta (çok model ve çok veri seti), deneyler yapılarak aşırı öğrenme ve yetersiz veri seti problemlerinin varolup olmadığı anlaşılmaya çalışılmıştır. Bu deneyleri gerçekleştirmek adına, 4 farklı model kullanılmıştır. Cresci et. al. [13], Efthimion et. al. [14], Kudugunta and Ferrara [15] ve Botometer [11]. Bu 4 model üzerinde, bot tespit sistemi araştırmacıları tarafından sıklıkla kullanılan ve kullanımı herkese açık olan 20 farklı veri seti kullanılarak ve değişik deney düzenekleri ile deneyler gerçekleştirilmiştir. İlk olarak modeller, orijinal makalelerinde kullanılan veri setleri ile deneylere tabi tutulmuştur. İkinci olarak ise modeller daha önce hiç görmedikleri veri setleri ile test edilerek performansları ölçülmüştür. Son olarak ise Yang [10] makalesinde Botometer için yapılan deneyin aynısı Efthimion et. al. ve Kudugunta ve Ferrara modellerine [14,15] de uygulanarak tüm modeller ile Botometer'ın performansları karşılaştırılmıştır. Yapılan karşılaştırmalar sonucunda en yüksek performansa sahip olan modelin Botometer olduğu anlaşıldığından, modelin gerçekten iyi bir performansa sahip olup olmadığının anlaşılması için senaryo bazlı bir şekilde deneyler gerçekleştirilmiştir. 5 farklı bot senaryosu oluşturularak Botometer'ın bu senaryolara karşı performansları ölçülmüştür. *Hayran botları* sadece bir kişinin tweetlerini retweet edecek şekilde tasarlanmıştır. *Trend konu botları* günlerce aynı hashtagi paylaşarak bir konuyu daha popular bir hale getirmeye çalışan botlardır. *Propaganda botları* aynı mesajı sürekli

paylaşarak bir görüşü popular hale getirmeye çalışan botlardır. *Reklam botları* ise bir web sitesine ait linkleri paylaşarak bir ürünün reklamını yapmak amacıyla kullanılan botlardır. Tüm bot senaryoları tasarlanırken, botların otomatize davranışlarının insanlar tarafından açık ve kolay anlaşılır bir biçimde olmasına dikkat edilmştir. Bunun nedeni, mevcut bot tespit sistemlerinin en basit durumlarda bile yetersiz kaldığını göstermektir.

Yapılan deneyler sonucunda elde edilen çıkarımlara göre, tüm modeller kendi kullandıkları veri setleri ile test edildiklerinde çok yüksek başarı skorlarına ulaşabilmektedirler ancak eğitimi aynı veri setleri ile gerçekleştirip daha önce görmedikleri veri setleri ile yapılan testlerde performansların önemli bir ölçüde düştüğü görülmektedir. Bu çıkarımlara bakıldığında, veri setlerinin ne kadar kısıtlı oldukları ve modellerin ise ufak değişikliklere karşı dahi dayanaklı olmadıkları sonucuna varılmaktadır. Tüm modelleri karşılaştırmak için yapılan ve hem eğitim hem de test veri setlerinin değiştirildiği deney sonucunda ise en iyi performans gösteren modelin Botometer olduğu görülebilmektedir. Mevcut modeller arasından en iyi olan ve son teknoloji ürünü olarak adlandırılan Botometer'ın ise kırılganlığı ölçmek adına yapılan senaryo bazlı deneylerde ise Botometer'ın hayran botları hariç hiçbir bot senaryosunda tespit işlemini gerçekleştirmediği görülmektedir. Sonuç olarak, küçük bir eforla bile en gelişmiş bot tespit sistemlerinin aldatılabileceği deneylerimiz sonucunda gösterilmiş, daha başarılı ve kırılganlığı düşük modeller oluşturabilmek adına daha sofistike veri setlerine ihtiyaç olduğu ortaya koyulmuştur.

Yaptığımız çalışmalar sonucunda elde etmiş olduğumuz çıkarımlar ile literatüre katkı sağladığımız konular şunlardır:

- 4 farklı bot tespit sisteminin performansları değişik veri setleri üzerinde karşılaştırıldı. Tüm modellerimiz üzerinde farklı deney düzenekleri ile deneyler gerçekleştirildi. İlk defa bu kadar kapsamlı ve bilgimize göre daha önce karşılaştırılmamış modeller ile deneyler gerçekleştirildi.

- Değerlendirdiğimiz modellerden Cresci et. al., Efthimion et. al. ve Kudungunta ve Ferrara'ya daha önce test edilmedikleri veri setleri üzerindeki performansları ölçüldü.

- Modellerin ve gerçekleştirdiğimiz deneylerin kodları paylaşılarak deneylerin tekrar gerçekleştirilebilmesi imkan sağlanmıştır.

- Senaryo bazlı olarak deneyler gerçekleştirilerek en başarılı bulunan modelin performans ölçümü gerçekleştirilmiştir.

Bu tezin organizasyonu şu şekildedir: Bölüm 2 mevcut bot tespit sistemleri üzerine yapılmış literatür araştırmalarını, Bölüm 3 tezimizde kullanılan 4 farklı bot tespit sistemlerininin detaylarını, Bölüm 4 yapılan çalışmalar boyunca kullanılan veri setlerinin detaylarını, Bölüm 5 kullanılan modellerin nasıl gerçeklendiğini, ne tarz deneyler yapıldığını ve deneylerin sonuçlarını, Bölüm 6 Botometer için yapılmış olan senaryo bazlı deneylerin düzeneklerini ve sonuçlarını, Bölüm 7 ise tez sonucunda elde edilmiş çıkarımları içermektedir.

2. LİTERATÜR ARAŞTIRMASI

Siber güvenlik ve yapay zeka araştırmacıları, OSA paylaşımlarının neden olabileceği zararlar sonucunda ortaya çıkabilecek maddi kayıpların önüne geçebilmenin yanı sıra, manevi yönden de insanların korunması gerektiğini düşünerek son yıllarda sosyal siber güvenlik alanına ilgi göstermektedirler. Gösterilen bu ilgi özellikle sosyal botların tespiti için farklı yaklaşımlarla sistemler oluşturulması ve bu sayede OSA paylaşımlarının sebep olduğu zararların önüne geçilmesi için yapılan çalışmaların artmasını sağlamıştır [22,23]. Bot tespiti alanıında yapılan araştırmalara bakıldığında sistemlerin üç farklı yaklaşımla tasarlandığı anlaşılmaktadır.

2.1. Ağ Tabanlı Bot Tespit Sistemleri

Bot hesaplar arasındaki ilişkiler sonucunda oluşan yapılar bir çizge ile ifade edilerek, bu çizgeler sayesinde ağ tabanlı bot tespit sistenleri oluşturulur. *SybilRank* [1] adı verilen ve sosyal grafların kullanıldığı bir model sayesinde, OSA kullanıcılarının bot olup olmadığı sınıflandırır. Ölçeklendirilebilir bir sistem olan SybilRank yüz milyonlarca graf düğümü ile etkili bir sınıflandırma işlemini gerçekleşirebilmektedir. Yapılan bir çalışmada [2] uluslararası alanda bot hesaplara karşı oluşturulan savunma sistemlerinin zafiyetleri incelenerek, yeni bir beyaz listeleme yöntemi ile bot hesapların filtrelenmesini sağlayacak bir sistem tasarlanmıştır. Başka bir makalede ise bir organizasyona ait tüm hesapların aktiviteleri görüntülenerek, hedeflenen sosyal botların tespiti gerçekleştirilmiştir [3]. *Souche* [4] adında bir sistem geliştirilerek online platformlarda bulunan gerçek hesapların tespiti gerçekleştirilmiştir. Online platformları kullanan gerçek hesaplar ile bir kayıt ağacı gerçeklenmiştir. Bu ağaç sayesinde diğer gerçek hesapların %85 olasılıkla tespit edilebilmesi sağlanmıştır.

2.2 Kitle Kaynak Kullanımı Tabanlı Bot Tespit Sistemleri

Sosyal bot tespiti otomatize olarak yapılması istenen bir işlem olmasına rağmen, araştırmacıların yaptıkları kitle kaynak çalışmaları ile insanların, OSAlardaki hesapları bot olarak tespit edip veri setlerinin etiketlenmesi sağlanır. Ayrıca,

insanların farklı türlerdeki bot hesaplar ile gerçek hesapları ayırt edip edemediğine dair bir çalışma gerçekleştirilmiştir [5]. Elde edilen sonuçlara göre insanlar geleneksel bot türlerini %90'ın üzerinde bir isabet ile tespit edebilmiştir. Buna rağmen, yeni nesil sosyal botlar üzerindeki performanslarına bakıldığında insanlar yalnızca %23 gibi bir doğruluk olasılığı ile bot hesapları tespit edilebilmiştir. Wang et. al.'da [6] ise aynı süreç iki farklı grup insan kullanılarak gerçekleştirilmiştir. Bu işle uğraşan uzmanlar ve uzman olmayan insanlardan rastgele biçimde seçilmiş insanlardan oluşan bu gruplarla yapılan deney sonucunda ise uzmanların, uzman olmayanlardan sadece çok küçük bir farkla daha başarılı olabildikleri görülmüştür.

2.3 Makine Öğrenmesi Tabanlı Bot Tespit Sistemleri

OSAlardaki bot hesap tespitinin otomatize bir şekilde gerçekleştirilebilmesi için, makine öğrenmesi uygun bir yol olarak görülmektedir. Bu yüzden, bu alanda hem denetimli hem de denetimsiz öğrenme ile yapılmış birçok çalışma bulunmaktadır.

Cresci et. al. [7] makalesi denetimsiz bir öğrenme modeli uygulayarak, kullanıcıların aktivitelerini bir DNA zincirine benzetmekte ve zincirler arasındaki benzerlik ile bot tespit işlemini gerçekleştirmektedir. Alvasi et. al.[8] ise Naïve Bayes ve J48 algoritmalarını kullanarak gözetimli bir bot tespit sistemi oluşturulmuştur. Yang et. al. [2] makalesinde ise bot tespit sistemlerinin nasıl aşılacağına dair taktiklerden ve önerdikleri yeni öznitelikler ile sistemlerin nasıl daha başarılı bir şekilde oluşturulabileceği anlatılmıştır.

Makine öğrenmesi tabanlı sistemlerin paylaşıldığı makalelerde [9,10] modellerin çok yüksek performanslara ulaşıldığı belirtilmesine rağmen bu performansların yalnızca belirli veri setleri için bu seviyede olduğu, veri setleri farklılaştıkça performansların önemli bir ölçüde düştüğü görülmektedir. Bunun nedeni ise, makine öğrenmesi modellerinin daha önce görmedikleri tarzlardaki botlara karşı nasıl bir sınıflandırma yapacaklarını bilememeleridir. Yang et. al. [10] makalesi ayrıca Botometer modeline ait ölçeklendirme probleminden de bahsetmektedir. Model, Twitter API kullanımı ve kendi sorgu limitleri dolayısıyla oluşan büyük ölçeklerdeki bot tespit etme işlemlerinde yeterince hızlı bir şekilde performans sergileyememektedir. Ayrıca çağımızdaki mevcut bilgisayarların performansları da tespit işlemini yavaşlatan etkenler arasında görülmektedir.

Botometer [11] makine öğrenmesi tabanlı olup, bir Twitter hesabını bot olup olmama durumuna göre sınıflandıran bir bot tespit sistemidir. Botometer bot tespiti konusunda son teknoloji ürünü olarak görülmektedir. Bunun nedeni ise Twitter ile alakalı çok sayıda parametrelenin gözden geçirilip çok sayıda özniteliğe sahip olmuş olmasıdır ancak en büyük özelliği bir internet sitesi aracılığla tüm dünyaya bir hesabın bot olup olmadığını kontrol etme şansı vermesi ve kullanuşlı bir API'ye sahip olmasıdır. Bot tespiti konusunda tek API'ye sahip olan sistem Botometer olmamasına rağmen kısıtlamalara maruz kalmadan kullanılabilir bir durumdadır. Debot [12] denetimsiz öğrenme ile bot tespiti yapan bir modeldir ve Botometer gibi bir API'ye sahiptir fakat sahip olduğu günlük 50 sorgu limiti (Botometer için 17280) ve artık aktif olarak kullanılmayışı Botometer'ı kullanımı kolay bir son teknoloji ürünü haline getirmektedir. Ayrıca API'nin aktif olmayışı ve Chavoshi et. al. [12] tarafından artık modele ait anahtarlar dağıtılmaması Botometer'I tek alternatif durumuna getirmektedir.

3. DEĞERLENDİRİLEN BOT TESPİT SİSTEMLERİ

Kapsamlı ve doğru sonuçlar verebilecek şekilde deneyler yapabilmek adına 4 farklı türdeki sosyal bot tespit sistemi kullanılarak deneyler gerçekleştirilmiştir. Çizelge 3.1'de modellere ait bilgiler verilirken, alt başlıklarda modellere ait bilgiler detaylı bir şekilde belirtilmiştir.

Çizelge 3.1 Değerlendirilen Modeller

MODEL	**YILI**	**ÖĞRENME TÜRÜ**	**ALGORİTMA**
Cresci et. al. [13]	2017	Denetimsiz	Kümeleme
Efthimion et. al. [14]	2018	Denetimli	Ölçeklendirilmiş SVM
Kudugunta and Ferrara [15]	2018	Denetimli	Adaboost & LSTM
Botometer [11]	2020	Denetimli	Random Forest

3.1 Model-1: Cresci et. al. [13]

Denetimsiz bir öğrenme modeli olması dolayısıyla, Model-1 diğer bot tespit modellerinden farklı bir yaklaşıma sahiptir. Veri setlerinde bulunan Twitter kullanıcılarına ait profil bilgileri ve tweetleri ayrı ayrı kullanmaktansa, bütüncül bir yaklaşım sergileyerek Twitter profillerini bir bütün olarak ele almıştır. Kullanılan veri setlerindeki tüm kullanıcıların profillerinin zaman çizelgeleri alınarak DNA zincirlerine çevrilmiş ve zincirler arasındaki benzerlikler göz önüne alınarak bir hesabın bot olup olmadığına karar verilmiştir.

Cresci et. al. [13] DNA zincirlerine dönüştürme işlemi için iki farklı strateji kullanmaktadır. Önerilen ilk stratejiye göre her bir tweet tipi bir nükleotite çevrilmiştir:

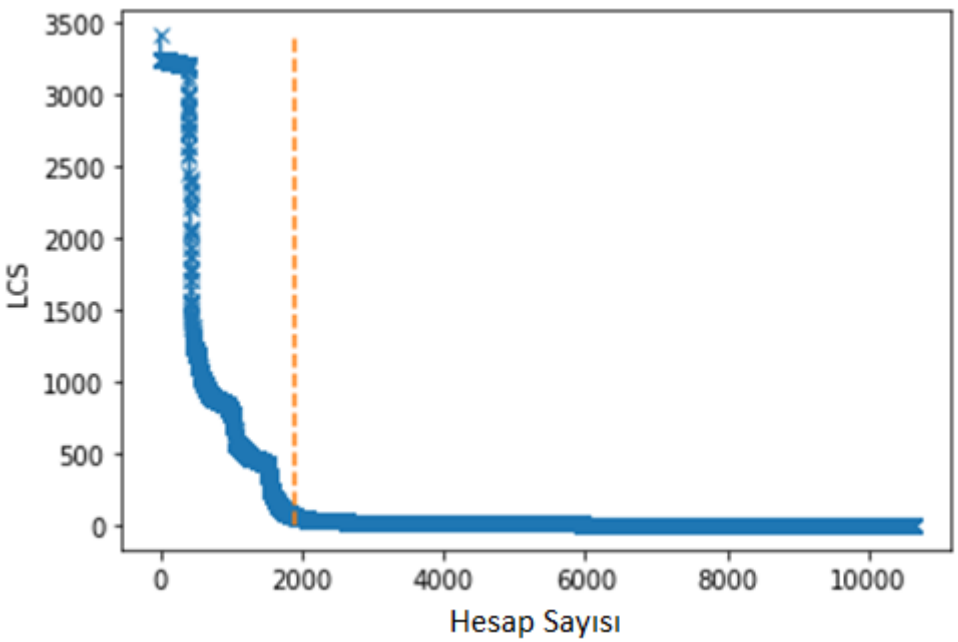
- Normal tweetler → A
- Retweetler → C

- Cevap tweetler → T

Diğer strateji ise tweetlerin içeriklerine bakarak DNA zincirlerinin oluşturulmasıdır. 6 türde oluşturulan tweet içerikleri şunlardır:

- Normal tweetler → N
- URL bulunduran tweetler → A
- # (hashtag) bulunduran tweetler → T
- @ (mention) içeren tweetler → C
- Medya içeren tweetler → G
- Birden fazla tweet türünün içeriğini barındıran tweetler → X

Yazarların iki stratejiyi de uygulayarak modelleri oluşturulması sonucunda ilk stratejinin ikinciye nazaran daha yüksek performans sergilediği görülmüştür.



Şekil 3.1 Cresci et. al. karar noktasının belirlenmesi [13]

İlk strateji ile model oluşturularak elde edilen DNA zincirlerinin arasındaki benzerlik zincirler arasındaki en uzun ortak alt diziye (longest common subsequence (LCS)) bakılarak elde edilmiştir. Elde edilen LCS değerleri sonrasında hesaplar benzerlik

skorlarına göre kümelenerek bir grafik oluşturulur. Bu grafik sonucunda elde edilen görüntüdeki kırılma noktası bir karar noktası olarak belirlenerek bir çizgi çizilir ve çizginin solunda kalan hesaplar bot olarak etiketlenirler.

3.2 Model-2: Efthimion et. al. [14]

Model-2 ölçeklendirilmiş SVM algoritmasını ve Twitter kullanıcılarının profil parametreleri öznitelik olarak kullanarak sosyal bot tespiti yapmaya çalışan bir sisteme sahiptir. 10 öznitelik kullanılan sistemde öznitelikler, tweet ve takipçi sayıları ve kullanıcı biyografisindeki farklı parametreler ile oluşturulmuştur. Öznitelikler aşağıda belirtildiği gibidir:

- Hesapta kullanılan dilin İngilizce olması
- Profil resminin bulunması
- Hesabın bir ekran adına sahip olması
- Hesabın 30'dan fazla takipçiye sahip olması
- Hesabın konumunun bilinmesi
- Biyografide link verilmesi
- Hesaptan 50'den fazla tweet atılması
- Takip edilen sayısının takipçinin 2 katından fazla olması
- 1000'den fazla takip edilen hesap bulunması
- Hiç tweet atılmamış olması

Profil bazlı tweetler haricinde, model tweet bazlı bir öznitelik olarak bir kullanıcının tweetleri arasındaki Levenshtein uzaklığa bakarak uzaklığın 30'dan yüksek olduğu hesaplar için özniteliği bot olarak etiketlemektedir. Farklı bir yaklaşım olmasına rağmen mevcut tekolojideki bilgisayarlar için her kullanıcının tweetleri arasındaki Levenshtein uzaklığı hesaplamak maliyetli olacağından bu öznitelik Efthimion et. al.'da [14] devre dışı bırakılmıştır. Aynı şekilde bu tezde deneyler gerçekleştirilirken yazılan kodlarda da Levenshtein uzaklığı devre dışı bırakılmıştır.

3.3 Model-3: Kudugunta ve Ferrara

Model-3, iki farklı sınıflandırma modeline sahiptir: 1) kullanıcı seviyesinde sınıflandırma ile bir Twitter hesabının bir bot olup olmadığının tespiti, 2) tweet seviyesinde sınıflandırma ile bir kullanıcıya ait bir tweete bakılarak bot olup olmadığının tespiti, bu sayede yalnızca bir tweetin bile bir hesabın bot durumu ile alakalı çok fazla bilgi verdiği ve tweet seviyesinde bir sınıflanırmanın başarılı bir şekilde çalıştığı gösterilmeye çalışılmıştır. Kullanıcı seviyesinde bir sınıflandırmada ise az sayıda öznitelik ile yüksek başarı skorlarına ulaşılabileceği kanıtlanmaya çalışılmıştır. Model oluşturulurken kullanılan öznitelikler şunlardır:

- Atılan tweet sayısı
- Takipçi sayısı
- Takip edilen sayısı
- Favori tweet sayısı
- Listelenen tweet sayısı
- Varsayılan profile sahip olması
- Hesabın konumunun bilinmesi
- Hesabın arka plan resmi kullanması
- Twitter onaylı bir hesap olması
- Korumaya alınmış bir hesap olması

Ayrıca SMOTE tekniği kullanılarak aşırı örnekleme yapılmış ve böylelikle daha fazla etiketli örneğe sahip olunmuştur. Birçok farklı makine öğrenmesi ve derin öğrenme algoritmaları model için uygulanmış ve kullanıcı seviyesinde sınıflandırma modeli için en iyi performansı Adaboost algoritmasının gösterdiği görülmüştür.

3.4 Model-4: Botometer

Botometer (eski adı BotOrNot) makine öğrenmesi tabanlı bir sosyal bot tespit sistemidir. Rastgele Orman (Random Forest) algoritmasını kullanan model, sahip olduğu 1000'den fazla öznitelik ile hesapları sınıflandırmaktadır. Modelin sahip olduğu öznitelikler 6 gruba ayrılır:

- Sosyal etkileşim tabanlı öznitelikler: retweetler, mention içeren tweetler, hashtag içeren tweetler ve hashtag birliktelikleri vb.
- Kullanıcı profilindeki parametrelere bağlı öznitelikler: dil, konum, hesabın yaratılma zamanı vb.
- Sosyal ağ tabanlı öznitelikler: takipçi sayısı, takip edilen sayısı vb.
- Zamanla ilişkili öznitelikler: tweet atma sıklığı, atılan tweetler arasındaki zaman farkları gibi.
- Dil özellikleri özelindeki öznitelikler: cümlenin ögeleri gibi.
- Duygu tabanlı öznitelikler: tweetlerde bulunan mutluluk belirten kelimeler, emojiler vb. İçeriklere bakılarak oluşturulan özniteliklerdir. Yalnızca İngilizce içerikler paylaşan hesaplara uygulanır.

Botometer, İngilizce ve Evrensel olmak üzere iki farklı Kapsamlı Otomasyon Olasılığı (CAP) puanı vermektedir. İngilizce puanı, dil bazlı özniteliklerinde modele eklenmesini ve CAP skorlarının bu şekilde elde edilmesini sağlar. Evrensel puan ise, dil tabanlı olmayan, profildeki parametreleri ve ağ yapısı tabanlı öznitelikler gibi özniteliklere bakılarak oluşturulur. Yang et. al.'da [11] ise Botometer'ın başarım olarak 0.95 AUC skoruna sahip olduğu raporlanmıştır.

4. VERİ SETLERİ

Veri setleri değerlendirilecek bot tespit modellerin performanslarının isabetli bir şekilde değerlendirilmesi ve karşılaştırması yönünden büyük bir önem taşımaktadır. Ayrıca bir nevi gerçek dünya simülasyonunda deneylerin gerçekleştirilmesi için etiketlenmiş veri setlerinin bulunması önemli bir husustur. Bu yüzden, veri setlerinin büyüklükleri, etiketlerin dağılımı ve farklı bot türlerinin bulunması gibi faktörler veri setleri ile güvenilir bir şekilde deney gerçekleştirmek için önemlidirler. Buna rağmen, yüksek kaliteye sahip bir veri seti oluşturma çalışmaları zorlayıcı olmaktadır. Bunun nedeni ise gerçek dünyada çok farklı türlerde bot hesapların bulunması ve sürekli olarak yeni türlerde sosyal botların geliştirilmesidir. Ek olarak etiketleme süreci de kendine ait zorluklar barındırmaktadır. Etiketlenme işleminin manuel olarak yapılmak zorunda olmasının ve yalnızca kullanıcı profillerine bakılarak bir hesabın bot olup olmadığına karar verilmesi ve kararın değerlendiren kişinin yapacağı tahmine kalması sorunlar oluşturabilmektedir. Bu sebeplerden dolayı, deneylerde kullanılabilecek güvenilir bir veri seti oluşturmak adına farklı boyutlarda ve farklı bot türlerine sahip 20 farklı veri seti toplanmıştır. Çizelge 4.1'de veri setleri detaylı bir şekilde gösterilmiştir.

Çizelge 4.1'deki satır 1 ile 10 arasındaki veri setleri Botometer'ın hem eğitimi hem de test işlemleri için kullanılmıştır. **Botwiki** veri seti botwiki.org sitesinden alınmış aktif botlardan oluşmaktadır. **Celebrity** veri seti ünlülerin Twitter hesaplarının toplanması ile oluşturulmuştur. **Cresci Stock** veri seti özel olarak seçilmiş bazı cashtaglerin (parasal ifadeler içeren hashtagler) takip edilmesi sonucu borsa hisselerinde manipülasyon yapmak isteyen botlar tespit edilerek oluşturulmuştur. Hesaplar manuel olarak etiketlenmişlerdir. **Cresci Rtbust** bazı tweetlerin görünürlüğünü ve bazı hesapların popüleritesini artırmak amacıyla retweet yapan İtalyan botlar ile oluşturulmuştur. Hem gerçek hesaplar hem de bot hesaplar manuel olarak etiketlenmiştir. **Pronbots** veri seti, tweet atarak dolandırıcılık yapan botlardan oluşmaktadır. Bu bot hesaplarla alakalı nasıl toplandıklarına dair bir bilgi bulunmazken, yine bu botlar değişik çalışmalarda kullanılmaktadır (örneğin [16]).

Çizelge 4.1 Deneylerde Kullanılan Veri Setleri

Satır	Veri Setinin Adı	Bot Sayısı	Gerçek Hesap Sasyısı	İçerik	Kullanan Çalışmalar
1	Botwiki [10]	540	0	Profil	[11]
2	Celebrity [16]	0	19997	Profil	[11]
3	Cresci Stock [17]	7102	6174	Profil	[11]
4	Cresci Rtbust [18]	353	340	Profil	[11]
5	Pronbots [16]	17882	0	Profil	[11]
6	Botometer Feedback [16]	139	380	Profil	[11]
7	Gilani [19]	1090	1413	Profil	[11]
8	Vendor Purchased [16]	1087	0	Profil	[11]
9	Verified [10]	0	1987	Profil	[11]
10	Political Bots [16]	62	0	Profil	[11]
11	Fake Followers [5]	3351	0	Profil & Tweetler	[11, 14]
12	Genuine Accounts [5]	0	1083*	Profil & Tweetler	[11, 13, 14, 15]
13	Social Spambots 1 [5]	991	0	Profil & Tweetler	[11, 13, 14, 15]
14	Social Spambots 2 [5]	3457	0	Profil & Tweetler	[11, 14, 15]
15	Social Spambots 3 [5]	464	0	Profil & Tweetler	[11, 13, 14, 15]
16	Traditional Spambots 1 [5]	1000	0	Profil & Tweetler	[11, 14]
17	Traditional Spambots 2 [5]	100	0	Profil	[11, 14]
18	Traditional Spambots 3 [5]	403	0	Profil	[11, 14]
19	Traditional Spambots 4 [5]	1128	0	Profil	[11]
20	NBC Russian Bots [20]	453	0	Profil & Tweetler	[14]

Botometer Feedback veri seti Botometer'ı kullanan kişiler tarafından verilen geri bildirimler sonucunda elde edilmiş botlardan oluşmaktadır. Etiketleme işlemi ise Yang et. al.'ın [16] yazarlarından bir tarafından yapılmıştır. **Gilani** veri setindeki

Twitter hesapları manuel olarak etikenlenmiş ve takipçi sayılarına göre gruplanmışlardır [19]. Hesaplar takipçi sayılarına göre 4 gruba şu şekilde ayrılmıştır: ünlü olma durumu (9M'den fazla takipçi gibi), çok popüler olma durumu (900 bin ile 1.1 milyon arasında takipçi gibi), orta seviye bir tanınırlık (90 bin ile 110 bin arasında takipçi sayısı gibi) ve düşük popülerlik (900 ile 1100 arasında takipçi sayısı gibi). Bu işlemden sonra ise, tüm hesaplar etiketleme işlemini yapacak 4 kişi tarafından manuel olarak ortaklaşa kararlar alınarak etikenlenmiştir. **Vendor-Purchased** veri seti farklı şirketlerden sahte takipçiler satın alınarak oluşturulmuştur. **Verified** Twitter tarafından doğruluğu kanıtlanmış hesaplardan oluşmaktadır. Bu veri seti Botwiki ve Vendor-Purchased gibi sadece botlardan oluşan veri setlerini dengelemek için oluşturulmuştur. **Political Bots** bir Twitter kullanıcısı (@josh_emerson) tarafından paylaşılmış, politik içerik paylaşımı yapan botlardan oluşan bir veri setidir. **Fake Followers** veri seti takipçi sayısının artırılması adına kullanılan botlardan oluşmaktadır. Bu botlar Cresci et. al. tarafından satın alınmıştır [5].

Çizelge 4.1'deki 11-15 ve 17-19 satırlarında bulunan veri setleri Cresci et. al. [5] tarafından oluşturulmuştur. **Genuine Accounts** veri seti insanlar tarafından kontrol edilen gerçek hesapları içermektedir. Bu veri seti oluşturulurken, rastgele olacak şekilde Twitter hesapları seçilerek sorular sorulmuş ve doğal dil metodları ile hesapların insanlar tarafından yönetilip yönetilmediği anlaşılmaya çalışılmıştır. Social Spambots veri setleri (satır 13,14,15) kompleks tweet yapısına ve sahte bilgilerle doldurulmuş profillere sahip sofistike botlardan oluşmaktadır. **Social Spambots 1** İtalya'daki seçimlerden birine katılacak bir politikacının propagandasını yapabilmek amacıyla kullanılmış botlardan oluşmaktadır. **Social Spambots 2** bir mobil uygulamanın reklamını yapmak amacıyla bir hashtag aracılığıyla içerik paylaşan botlardan oluşmaktadır. **Social Spambots 3** sürekli Amazon ürünlerininin bulunduğu linkler paylaşarak reklam yapan botları içermektedir. Traditional Spambots (satır 16-19) basit bir profil yapısına sahip ve tespit edilmesi Social Spambots'tan daha kolay olan botlardan oluşmaktadır. **Traditional Spambots 1** URL paylaşarak Twitter kullanıcılarını zararlı içereklere yönlendirmek isteyen botlardan oluşmaktadır. Yang et. al [2] tarafından eğitim seti olarak tasarlanmıştır. **Traditional Spambots 2** attığı tweetlere birkaç kullanıcı adı ve insanları kandırmak amacıyla içeriğinde para önerilen URL'ler ekleyerek paylaşım yapan botlardan

oluşmaktadır. **Traditional Spambots 3,4** ise basit bir profil yapısına sahip ve sürekli olarak Twitter kullanıcılarıyla iş teklifleri paylaşan botlar içermektedir.

Son olarak, **NBC Russian Bots** veri seti NBC News tarafından paylaşılan ve 2016 yılında yapılan Amerika Başkanlık Seçimleri boyunca insanların görüşlerini manipüle edebilmek adına zararlı aktivelerde bulunup içerikler paylaşan Rusya bağlantılı Twitter hesaplarından oluşmaktadır.

Uzun bir tablo ile gösterilebilecek sayıda veri setimiz olmasına rağmen, her veri seti için verinin farklılaştığı görülmektedir. Örnek olarak, satır 1-9 ve 17-19'daki veri setleri yalnızca kullanıcıların profil bilgilerinden oluşurken, kullanıcılara ait tweetler bulunmamaktadır. Bu durum, tüm veri setlerini bütün modeller için kullanmamızı ve tüm deneyleri gerçekleştirebilme kapasitemizi düşürmüştür.

5. MODELLERİ DEĞERLENDİRME DENEYLERİ

Bu bölümde performans değerlendirmesi yaptığımız 4 farklı modelin nasıl gerçeklendiğinden (Bölüm 5.1), performansları değerlendirmek için nasıl deney düzenekleri kurduğumuzdan (Bölüm 5.2) ve bu deneyler sonucunda elde ettiğimiz sonuçlardan (Bölüm 5.3) bahsedilmiştir.

5.1 Modellerin Gerçeklenmesi

Kullanılan her model farklı bir yapıya sahip olduğundan, her biri için farklı süreçler izlenmiştir. Makalelerde modellere ait parametreler ile alakalı yetersiz bilgiden kaynaklı ortaya çıkabilecek problemlerin önüne geçebilmek adına, yazarlar tarafından kodları kısmen de olsa paylaşılmış makalelerin seçilmesine karar verilmiştir.

5.1.1 Model-1

Cresci et. al. [13] makalesi göz önüne alınarak gerçeklenen modeldir. Modeli gerçekleştirmek adına yazarlarla iletişime geçerek kodu paylaşıp paylaşamayacakları soruldu. Yazarlar büyük bir iyilik göstererek modeli gerçekleştirmek adına kolaylık sağlayan ve veri setlerinde bulunan hesaplar arasındaki LCS'yi hesaplayabileceğimiz kodu paylaştılar.

Model, hesaplara ait tweetlerin alınarak DNA zincirlerine dönüştürülmesi ve sonrasında LCS'ye bakılmasıyla oluşturulmaktadır. Bundan dolayı, içeriğinde hesaplara ait tweetlerin bulunduğu veri setleri (Çizelge 4.1'de satır 11-15 ve 20) ile bu modele ait deneyler gerçekleştirilebilmektedir.

Yazarların paylaştığı kod, DNA zincirlerine çevrilmiş kullanıcı tweetleri arasındaki LCS değerlerini bulmakta ve ortaya bir LCS grafiği çıkarmaktadır. Bu grafik ile bot-gerçek hesap ayrımını doğru bir şekilde yapabilecek bir karar çizgisi belirlemek gerekmektedir. Bu çizgi yazarların da yaptığı gibi LCS değerlerinin önemli ölçüde düşmeye başladığı bir nokta olarak belirlenmelidir. Kod çizgi çizilecek noktayı otomatik olarak bulabilecek şekilde tasarlanmasına rağmen yapılan deneylerde

otomatik olarak bulunamadığından, çizgiler manuel olarak belirlenmiştir. Bu işlem LCS değerlerinin önemli ölçüde düşmeye başladığı birçok noktadan çizgiler çizilerek yapılmış ve en yüksek performansın gösterildiği noktadan çizilen çizgi referans alınarak sonuçlar paylaşılmıştır.

5.1.2 Model-2

Modele ait kod, Efthimion et. al. [14] makalesinin yazarları tarafından makale içerisinde paylaşılmıştır. Bu yüzden deneyler paylaşılan kod dönüştürülüp kullanılarak gerçekleştirilmiştir. Kodu dönüştürme işleminin amacı, kodu modüler hale getirerek farklı deney düzeneklerinde kolaylıkla kullanılabilir bir kod ortaya konabilmesidir. Öznitelikler daha toplu ve kompleks hale getirilerek kodun daha kolay anlaşılabilmesi sağlanmış, modelin performansı ile alakalı bilgi içermeyen algoritmalar, tablolar ve grafikler silinmiştir. Modüler hale getirilen kod paylaştığımız Github (https://github.com/bugratorusdag/Bot-Detection-Systems adresine ulaşabilmek için bugratorusdag@gmail.com adresine e-mail atılabilir.) veri havuzunda bulunmaktadır. Modelin makalesinde ve orjinal kodda bulunan hiper parametreler olduğu gibi bırakılarak kodun orjinaline uygun bir şekilde çalışması sağlanmıştır.

5.1.3 Model-3

Kudungunta ve Ferrara [15] makalesindeki iki farklı sınıflandırma modeli içinde kodlar yukarıda belirtilen Github linki aracılığı paylaşılmıştır. Çalışmalarımızda gerçekleştirdiğimiz deneylerde tüm modellerin sınıflandırma işlemini aynı şekilde yapmasını sağlamaya çalışıldığından, yalnızca makalede paylaşılan kullanıcı seviyesinde sınıflandırma modelinin paylaşıldığı kod parçası çalışmalar için kullanılmıştır. Tüm hiper parametreler orjinal koddaki gibi bırakılmasına rağmen kodu daha modüler hale getirmek amacıyla çalışmalar yapılmıştır. Paylaşılan kodda yalnızca en yüksek performansın elde edildiği algoritma, özniteliklerin toplu bir şekilde gösterimi ve her deney için farklı bir hücre paylaşılarak tüm deneylerin tüm veri setleri ile hızlı bir şekilde yapılabilmesi sağlanmıştır.

5.1.4 Model-4

Botometer [11], ticari bir ürün olmasından ötürü sosyal bot tespiti için hızlı ve rahat bir şekilde kullanılabilecek bir model olarak görülmektedir. Bir API aracılığıyla kolaylıkla kullanılabilecek bir şekilde tasarlanmış olan Botometer bu sayede her Twitter hesabını kolaylıkla analiz edebilecek bir yapıdadır. Her ne kadar teker teker hızlı ve rahat bir şekilde tespit işleminin sonuçlarına ulaşılsa da toplu bir şekilde sosyal bot tespiti yapılmak istenen deneylerde, Botometer kullanıcılarına bazı zorluklar çıkabilmektedir.

İlk olarak, Botometer ticari bir ürün yapısına sahip olduğundan, kodunun yazarlar tarafından paylaşılma imkanı bulunmamaktadır. Bundan dolayı, yapılacak çalışmalar API'lerin zaman ve sayı kısıtlarılarına (API kullanımındaki gecikmeler ve günlük belirli bir sayıda Twitter hesabının bot durumunun kontrol edilebilmesi) takılarak yapılmak durumundadır.

İkinci olarak ise, modelin kodunun yazılarak gerçeklenmesinin çok zor olmasıdır. Botometer hem binden fazla özniteliğe sahip olduğundan ve özniteliklerin tamamının makalesinde detaylıca bahsedilmemesinden hem de değişen bir ticari ürün olduğundan özniteliklerin sürekli bir değişim halinde olması dolayısıyla kodunun tam doğru bir şekilde yazılması zor gözükmektedir.. Bu durumda deneyler yalnızca API aracılığı ile gerçekleştirilebilecek durumdadır.

Üçüncü olarak, Botometer API yalnızca aktif olan Twitter hesapları için kullanılabilir bir durumdadır ancak literatürde kullanılan ve bizim deneylerimizde kullandığımız veri setlerindeki birçok hesap silinmiş veya Twitter tarafından askıya alınmış hesaplardan oluşmaktadır. Bunun yanında kullandığımız veri setlerinde bulunan hesapların zaman içerisinde yeni içerikler üreterek veya var olan içerikleri silerek yeni bir profile dönüşerek Botometer API tarafından farklı bot skorları elde edilmesine neden olacaktır. Çalışmamızda bu sorunu aşabilmek adına alternatif bir çözüm düşünülmüştür. Bu çözüm yeni bir Twitter hesabı açılıp kullanılarak, hesabın bilgilerinin API aracılığıyla veri setlerindeki hesapların bilgilerine dönüştürülmesidir fakat veri setlerinde bulunan bilgilerin eksikliği (retweet yapılan tweetin sahibiyle alakalı bilgiler, kullanılan renkler ve hesap dili gibi) ve Botometer'ın özniteliklerinde meydana gelen değişikliklerden kaynaklanan hangi bilginin gerekli olup olmadığı

problemi bu çözümün kullanılmasını zorlaştırmaktadır. Bu sebeplerden ötürü, deney düzeneklerimiz ve veri setlerimiz Botometer'ın kullanımına imkan tanımamaktadır. Bunun yerine, Botometer'ın [10] birçok veri seti kullanılarak yazarları tarafından yapılan deneylerin sonuçlarını tezde rapor ederek ve kullanılan deney düzeneğini Efthimion et. al. ve Kudugunta ve Ferrara modelleri üzerinde gerçekleştirerek Botometer'ın performansını karşılaştırma işlemi gerçekleştirilmiştir.

5.2 Deney Düzenekleri

Bölüm 5'te gerçekleştirilen tüm modellerin karşılaştırılması amacıyla 3 farklı deney düzeneği tasarlanmıştır: 1) kullandığımız modellerin makalelerinde geçen veri setleri ile yapılan deneyler, 2) orjinal makalede geçmemesine rağmen modellerin farklı veri setlerindeki performanslarını görmek amacıyla yapılan deneyler, 3) Botometer ile diğer modellerin performanslarının karşılatırılabilmesi için Yang et. al. [10] tarafından gerçekleştirilen Botometer'a uygulanan deneylerin diğer modellere uygulanması.

Yapılan her deney, o deney düzeneği için uygun olan tüm modeller ve veri setleri kullanılarak gerçekleştirilmiştir. Model-1 tweet verisini kullanarak çalışan bir model olduğundan, Çizelge 4.1'deki satır 1-10, 17-19 veri setleri bu model için kullanılamaz. Bundan dolayı, Model-1 için ikinci ve üçüncü deney düzenekleri kullanılarak deneyler gerçekleştirilmesi mümkün olmamaktadır. Aynı şekilde, yukarıda bahsedilen nedenlerden dolayı, Botometer deneyleri gerçekleştirmek adına kullanılamamaktadır. Bu nedenle, ilk iki deney düzeneği için Botometer'ı diğer modeller ile karşılaştırmak mümkün değildir. Model-2 ve Model-3 ise bütün deney düzeneklerinde kullanılmıştır. Modellerin performansları değerlendirilmek yerine ise, deneyler sonucunda her modelin elde etmiş olduğu kesinlik, doğruluk, duyarlılık, F1 ölçütü ve eğrinin altındaki alan değerleri raporlanmıştır.

5.3 Deney Sonuçları

Bu bölümde her bir deney düzeneği için elde edilmiş performans ölçütleri raporlanmıştır. Ayrıca her bir deney düzeneği için gerekli olacak detaylar da bu bölümde belirtilmiştir.

5.3.1 Orjinal makalede kullanılan veri setleri ile yapılan deneyler

Bu deney düzeneği, Model-1, Model-2 ve Model-3'ün gerçeklendiği makalelerde kullanılan veri setlerinin, aynı düzenekle Model-4 hariç tüm modeller için kullanılması ile gerçeklenmiştir. Bu deney düzeneği sayesinde, her bir modelin diğer modellerin veri setleri kullanıldığında nasıl performanslara sahip olduğu görülmüş ve raporlanmıştır. Ayrıca bu deney düzeneği ile, her bir modelin gerçeklendiği makalelerde raporlanan performans skorlarının, deney sonucunda elde edilen skorlarla karşılaştırılması sağlanmıştır. Bunun yanında, her üç modelin de makalelerinde kullanılmış olan veri setleri benzerlik gösterdiğinden, bu düzenek eğitim ve test veri setlerindeki küçük değişimlerin performansa olan etkileri hakkında da bilgi vermektedir.

Deney düzeneği, Çizelge 4.1'te bulunan 4 farklı veri seti varyasyonu kullanılarak gerçekleştirilmiştir:

- [13] makalesinde kullanılan tüm veri setlerinin kombinasyonu (Çizelge 4.1 satır 12, 13 ve 15)
- [14] makalesinde kullanılan tüm veri setlerinin kombinasyonu (Çizelge 4.1 satır 12-15)
- [15] makalesinde kullanılan tüm veri setlerinin kombinasyonu (Çizege 4.1 satır 11-16 ve 20)
- [13], [14] ve [15] makalelerinde kullanılan tüm veri setlerinin kombinasyonu (Çizelge 4.1 satır)

Deney düzeneğindeki tüm veri seti varyasyonlarıyla deneyler gerçekleştirilirken verinin, %80'i eğitim, %20'si ise test için ayrılmıştır.

Model-1 Twitter hesaplarının sınıflandırılmasını özdenetimsiz bir şekilde kümeleme yaparak gerçekleştirmektedir. Bu yüzden, Model-1 için bir veri seti kullanılacakken tüm hesaplar kümeleme işlemi için kullanılıp model oluşturulurken, tüm veri seti direkt olarak test için kullanılmış gibi performans skorları hesaplanır. Hesaplamalar sonucu elde edilen skorlar Çizelge 4.1'de belirtilen makalede bulunmaktadır. Model-1'in Çizelge 3.1'deki son veri seti ile performansı ölçülememiştir. Bunun nedeni ise,

Çizelge-5.1 Model-1, Model-2 ve Model-3'e ait makalelerde kullanılan veri setleri ile yapılan deneyler

	Model 1 [13]			Model 2 [14]			Model 3 [15]		
Veri Seti	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
[13] makalesindeki veri seti	0.44	0.18	0.21	0.92	0.85	0.91	0.98	0.99	0.98
[14] makalesindeki veri seti	0.38	0.41	0.27	0.94	0.95	0.94	0.99	0.99	0.99
[15] makalesindeki veri seti	0.31	0.007	0.01	0.96	0.97	0.94	0.98	0.97	0.98
[13,14,15] makalelerindeki veri setleri	–	–	–	0.94	0.96	0.95	0.98	0.96	0.98

son veri setinde bulunan Traditional Spambots 2 ve 3'ün Model-1 tarafıdan kullanılması gerekecek tweet içeriğine sahip olmamasıdır.

Farklı veri seti kombinasyonlarıyla yapılan deeneyler sonucunda çeşitli gözlemler yapılmıştır. Bu gözlemler neticesinde Model-2 ve Model-3'ün sınıflandırma konusunda çok yüksek performans skorlarına ulaştığı görülmüştür. Özellikle Model-3 elde ettiği performans skorları ile diğer modellerin önüne geçmiş ve makalelerindeki sonuçlarla tutarlılık gösterdiği anlaşılmıştır. Model-1'in performansı diğer modellere göre önemli ölçüde daha düşüktür. Cresci et. al. [13] makalesi Model-1 için 0.95 F1 skoru elde edilmiştir. Buna rağmen, yaptığımız deneylerde veri setlerinin %20'lik kısımlarını test seti olarak alırken, Model-1 performans skorlarını hesaplarken tüm veri setini test için kullanmaktadır. Deneylerin aynı düzenekte gerçekleştirilebilmesi adına Model-2 ve Model-3'te aynı eğitim ve test seti ile deneyler gerçekleştirilmiş, Model-1 için eğitim işlemi gerekmediğinden tüm veri seti ile test işlemi yapılmış ancak Model-2 ve Model-3'te kullanılan test seti ile aynı olan kısım alınarak skorlar hesaplanmıştır. Veri setlerinin tamamını kullanarak deneyler yapılmış ve 1. 2. ve 3. veri setleri için Model-1 sırasıyla 0.95, 0.45 ve 0.31 F1 skoru elde etmiştir. Elde edilen 1. veri seti skoru [13] makalesinde elde edilmiş F1 skoru ile uyumluluk göstermekte ve böylelikle raporlanmış bulguların doğruluğu gösterilmeye çalışılmıştır. Elde edilen sonuçlara göre Model-1 veri setlerindeki ufak değişikliklerde dahi yüksek performans kayıplarına uğramaktadır. Bu durum da Model-1 için önerilmiş olan benzer aktivitelere sahip olan sosyal bot aktivitelerinin tespiti yaklaşımının tutmadığını göstermektedir.

5.3.2 Farklı veri setleri ile yapılan deneyler

Botometer dışındaki tüm modellerimiz makalelerinde performanslarını rapolandıkları veri setleri, Çizelge 4.1'deki satır 11 ile 20 arasındaki veri setlerinin farklı kombinasyonlarından oluşmaktadır ancak satır 1 ile 10 arasıındaki veri setleri bu modellerde kullanılmamıştır. Bundan dolayı, bu deney düzeneğinde satır 11 ile 20 arasındaki veri setlerinin tamamı ile modeller eğitilerek, sonrasında daha önce hiç görmedikleri veri setlerindeki performanslarını ölçmek adına satır 1 ile 10 arasındaki veri setleri ile test edilmişlerdir. Bu deney düzeneğinde ise satır 1 ile 10 arasındaki veri setlerine ait tweet bilgileri bulunmadığından Model-1 kullanılarak deneyler

Çizelge 5.2 Model-2 ve Model-3'ün eğitim verisi sabit tutularak farklı Çizelge 4.1'deki veri setlerinin farklı kombinasyonları üzerinde performanslarının değerlendirilmesi

	Model-2					Model-3				
Test Veri Seti	**ACC**	**PRE**	**REC**	**F1**	**AUC**	**ACC**	**PRE**	**REC**	**F1**	**AUC**
Botwiki & Botometer Feedback (B+BF)	0.57	0.79	0.49	0.61	0.61	0.24	0.14	0.27	0.18	0.25
Celebrity & Pronbots (C+P)	0.83	0.89	0.72	0.79	0.82	0.88	0.96	0.8	0.87	0.88
Cresci Stock	0.65	0.67	0.68	0.68	0.65	0.59	0.79	0.16	0.27	0.56
Cresci Rtbust	0.56	0.66	0.29	0.41	0.57	0.46	0.34	0.12	0.18	0.45
Gilani	0.58	0.6	0.15	0.24	0.54	0.37	0.41	0.28	0.34	0.38
Vendor Purchased & Verified & Political Bots (VP+V+PB)	0.78	0.88	0.45	0.6	0.71	0.84	0.8	0.995	0.89	0.77

gerçekleştirilememiştir. Bundan dolayı, yalnızca Model-2 ve Model-3'e ait performans skorları raporlanmıştır.

Çizelge 4.1'de bulunan bazı veri setleri yalnızca bot hesapları bulundururken bazıları ise yalnızca gerçek hesapları bulundurmaktadır. Bot ve gerçek Twitter hesapları arasında sayı konusunda bir dengesizlik olduğundan bu problemi gidermek ve sayıları dengelemek adına bazı veri setleri birleştirilerek deneyler gerçekleştirilmiştir. Çizelge 4.1'de birleştirilmiş veri setleri gösterilmektedir.

Gözlemlerimiz neticesinde elde edilen performans değerleri Çizelge 5.2'de gösterilmektedir. İlk olarak, sonuçları Çizelge 3.1'de paylaşılan bir önceki deney düzeneği ile kıyaslandığında iki modelin de elde edilen performans skorları daha düşüktür. Örneğin, Çizelge 3.1'de Model-2'nin ACC değerleri 0.92 ile 0.96 arasındadır ve Model-3 de 0.98 ile 0.99 ACC değerleri ile neredeyse kusursuz bir sosyal bot tespiti performansına erişmiş gibi gözükmektedir. Buna rağmen, her iki model de Çizelge 4.1'de paylaşılan sonuçlara bakıldığında hiçbir veri seti veya performans kriterinde 0.9 başarıma ulaşamamaktadır. Ayrıca modellerin ACC ve F1 skorlarına bakıldığında ise sonuçların çok daha düşük olduğu görülmektedir. Örneğin, Model-2 ve Model-3 B+BF veri setine bakıldığında sırasıyla 0.57 ve 0.24 ACC skorlarına ulaşabilmişlerdir. Bu durum, her iki modelin de aşırı öğrenme problemine sahip olduğunu ve daha önce görmedikleri veri setlerinde çok düşük başarı skorlarına ulaşacaklarını göstermektedir. İkinci olarak, her iki modelin de performanslarının C+P ve VP+V+PB veri setlerinde diğer veri setlerine daha yüksek başarı skorları elde ettiği görülmektedir. Bu durum, her iki modelin de yalnızca çok sınırlı sayıda bot türüne sahip veri setlerinde eğitildiğini ve yalnızca eğitildikleri bot türlerini gördüklerinde başarılı olabildiklerini göstermektedir. Bu yüzden, eğitim yapılacak veri setleri seçilirken veri setlerindeki bot hesap türlerinin fazla ve çok çeşitli olmasına dikkat edilmelidir. Son olarak ise, iki model arasında bir kıyaslama yapıldığında bir modelin diğerine göre bir üstünlüğü bulunmadığı görülmektedir. F1 skorları referans olarak alındığında Model-3'ün Model-2'ye karşı 3 veri setinde (C+P, Gilani ve VP+V+PB) üstünlük elde ettiği görülürken, Model-2'nin de Model-3'e karşı diğer 3 veri setinde üstünlük sağladığı görülmektedir. Buna karşın, veri setlerinin boyutları göz ardı edildiğinde Model-2'nin Model-3'e göre daha iyi performans sergilediği görülmektedir.

5.3.3 Botometer ile karşılaştırma

Bölüm 5.1'de açıklandığı üzere, veri setlerindeki ve Botometer'a ait kısıtlar deneylerimizi Botometer'ı kullanarak gerçekleştirmemize imkan tanımamaktadır. Çizelge-5.3 Farklı eğitim ve test veri setlerinin kombinasyonu kullanılarak Botometer ile Model-2 ve Model-3'ün karşılaştırılması. AUC skorları gösterilmektedir. Botometer için skorlar [10] makalesinden alınmıştır. Her senayordaki en iyi performans griye boyanarak gösterilniştir.

	TEST VERİ SETLERİ									
EĞİTİM VERİ SETLERİ	VERİ SETLERİ (Çizelge-4'teki Satır Numaraları)	Modeller	1,9	2,5	3	4	6,10	7	8,9	11-19
	1,9	Model-2		0.82	0.67	0.5	0.45	0.54	0.68	0.88
		Model-3	-	0.91	0.54	0.58	0.53	0.55	0.89	0.77
		Botometer		1	0.64	0.71	0.5	0.61	0.99	0.87
	2,5	Model-2	0.65		0.6	0.54	0.49	0.5	0.67	0.47
		Model-3	0.74	-	0.62	0.61	0.39	0.49	0.81	0.38
		Botometer	0.97		0.67	0.58	0.4	0.46	0.89	0.65
	3	Model-2	0.65	0.56		0.72	0.47	0.5	0.58	0.5
		Model-3	0.48	0.74	-	0.59	0.48	0.49	0.53	0.5
		Botometer	0.97	0.96		0.66	0.51	0.65	0.94	0.74
	4	Model-2	0.67	0.82	0.68		0.49	0.49	0.72	0.5
		Model-3	0.88	0.87	0.64	-	0.54	0.46	0.79	0.73
		Botometer	0.97	0.9	0.7		0.6	0.57	0.83	0.68
	6,10	Model-2	0.43	0.45	0.49	0.47		0.61	0.56	0.35
		Model-3	0.42	0.42	0.47	0.47	-	0.64	0.62	0.69
		Botometer	0.94	0.96	0.54	0.5		0.73	0.95	0.93
	7	Model-2	0.81	0.51	0.53	0.43	0.59		0.56	0.53
		Model-3	0.68	0.58	0.42	0.52	0.57	-	0.5	0.39
		Botometer	0.94	0.84	0.45	0.45	0.7		0.72	0.88
	8,9	Model-2	0.65	0.84	0.7	0.73	0.5	0.54		0.8
		Model-3	0.95	0.91	0.52	0.52	0.58	0.51	-	0.51
		Botometer	1	0.99	0.64	0.71	0.57	0.53		0.82
	11-19	Model-2	0.88	0.84	0.59	0.54	0.58	0.57	0.75	
		Model-3	0.81	0.48	0.5	0.51	0.52	0.51	0.76	-
		Botometer	0.97	0.64	0.58	0.57	0.72	0.67	0.95	

Bundan dolayı, Botometer'ı seçtiğimiz diğer modellerin göstereceği performanslarla karşılaştırmak adına [13] makalesinde bulunan ve Şekil 2'de gösterilen deney düzeneğinin aynısını Model-2 ve Model-3 ile gerçekleştirerek performans karşılaştırılmasının bu şekilde yapılması sağlanmıştır. Özellikle, Çizelge 4.1'de

bulunan 19 veri seti kullanılarak 8 farklı veri seti oluşturulmuş ve sonrasında oluşturulan her bir veri seti ile modeller eğitilmiş ve kalan 7 veri seti kullanılarak eğitilmiş modellerin testleri sonucunda her bir senaryo için tüm modellere ait AUC skorları hesaplanmıştır. Bu sayede, eğitim ve test için farklı veri setleri kullanılarak modellerin performanslarının ölçülmesi sağlanmıştır. Elde edilen sonuçlar Çizelge 5.3'te gösterilmiştir.

Botometer, Model-2 ve Model-3 sırasıyla 40, 11 ve 5 senaryoda (56 senaryo arasından) en yüksek AUC performansına ulaşan modeller olmuşlardır. Bu sonucun ortaya çıkmasının en büyük nedenlerinden birisi Botometer'ın diğer modellere oranla çok daha fazla özniteliği modeline eklemiş olması olabilir. Botometer Twitter kullanıcılarının ağ yapıları ve tweet içerikleri gibi değişkenleri 6 farklı grup öznitelik ile sosyal botları sınıflandırmaya çalışırken, Model-2 ve Model-3 yalnızca Twitter kullanıcılarının profillerindeki değişkenleri öznitelik olarak kabul etmektedir. Bu sayede, farklı bot türlerine ait özelliklerin de göz önünde bulundurulmasıyla daha yüksek performans skorları elde edilebilmektedir.

Çizelge 5.1'e bakıldığında modellere ait performansların veri setleri değiştikçe önemli bir ölçüde değiştiği gözlemlenmektedir. Botometer, Vendor Purchased ve Verified (Çizelge 5'teki satır 8 ve 9) veri setleri ile eğitildiğinde, Celebrity-Pronbots (Çizelge 4.1'teki satır 2 ve 5) veri setleri kullanılarak yapılan testte model 0.99 AUC skoru elde etmektedir. Bu çok yüksek skora rağmen, Celebrity-Pronbots veri setleri ile yapılan eğitim ve Botometer Feedback-Political Bots (Çizelge 4.1'teki satır 6 ve 10) veri setleri ile yapılan test sonucunda Botometer yalnızca 0.4 AUC skoruna ulabilmiştir. Bu durum, yüksek kaliteye ve çeşitliliğe sahip sosyal botlardan oluşan veri setlerinin önemini göstermektedir.

6. SENARYO BAZLI BOTOMETER DENEYLERİ

Bölüm 5'te farklı eğitim ve test verileri ile gerçekleştirilen deneyler sonucunda Botometer en başarılı model olarak gözükmektedir. Buna rağmen, farklı veri setlerinde elde edilen değişken performans değerleri nedeniyle, en başarılı olmasına rağmen Botometer üzerinde deneyler gerçekleştirerek gerçek performansının daha iyi bir şekilde anlaşılması gereksinimi ortaya çıkmıştır. Bu gereksinime binaen, bir bot tespit sisteminin hangi bot türleri üzerinde başarılı olabileceğini anlamak adına Botometer'ın farklı bot türleri ile performansı ölçülmüştür. Bundan dolayı bu bölümde, tasarlanan bot senaryoları ve sonrasında yapılan performans değerlendirmelerinden bahsedilmektedir.

6.1 Bot Senaryoları

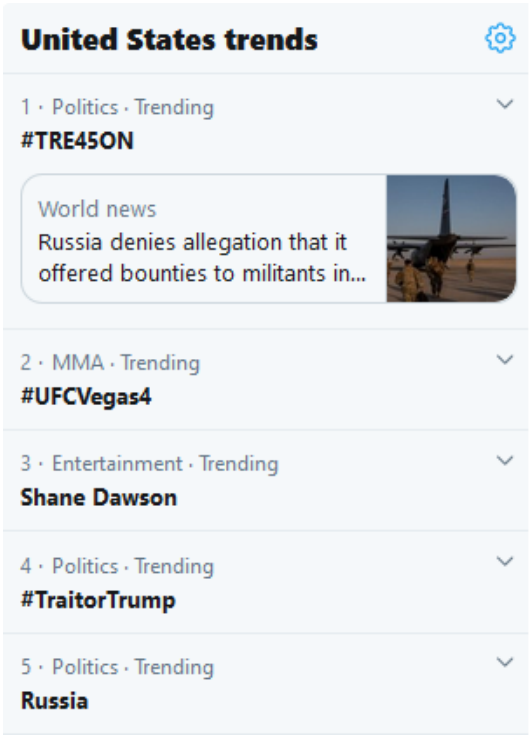
Bu bölümde, elde edilen performans değerleri sonucunda en başarılı model olan Botometer'ın senaryo bazlı olarak deneyler gerçekleştirerek performanslarının ölçülmesi amacıyla yaratılan bot senaryolarından bahsedilmektedir. Bot hesaplar özellikle Twitter'da bir hesaba veya içeriğe dair etkileşimi artırmak ve bilginin hızlı bir şekilde yayılmasını sağlamak adına birçok farklı eylemde bulunan hesaplardır. Bu durum göz önüne alınarak Twitter'da yapılan gözlemler sonucunda Botometer'ın performansının ölçülmesi adına 5 farklı bot senaryonu tasarlanmıştır. Mevcut bot tespit sistemlerinin veri seti farklılaştıkça performanslarının önemli ölçüde düşmeye meyilli olduğu yapılan deneylerle sabit olduğundan, en azından yaratılan bot senaryolarının insanlar tarafından kolaylıkla algılanabilecek şekilde tasarlanmıştır. Bunun yanında, Cresci et. al'da [13] bahsedilen ve Twitter gözlemlerimizle de örtüşen bot senaryoları bulunmaktadır. Twitter gözlemleri, mevcut veri setleri ve çalışmalar incelendiğinde oluşturulan 5 farklı bot senaryosu, bot tespit sistemlerinin zaafiyetlerin daha iyi anlaşılabilmesi açısından önem arzedecektir. Bot tespit sistemlerinin daha sofistike botlar üzerindeki performansları ise gelecek çalışmalara bırakılmıştır. Tasarlanan bot senaryoları takip eden bölümlerde anlatılmaktadır.

6.1.1 Hayran botları

Bu bot senaryosu, yalnızca bir Twitter hesabına odaklı bir bot hesabın, odaklandığı hesabın popüleritesini ve tweet içerikleri ile yapabileceği etkiyi artırmak amacıyla retweet işlemi gerçekleştiren bir bot türünden oluşmaktadır. Bu bot türünün odaklandığı hesap, bir politikacı veya reklam yapan bir şirket olabilir.

6.1.2 Trend konu botları

Hashtagler Twitter üzerinde sıklıkla tweetlere eklenerek belirli bir konu kakkında bilginin daha fazla insana ulaşmasını ve konu ile alakalı paylaşılan bütün içeriklerin bir araya toplanmasını sağlamak amacıyla kullanılırlar. Toplanan içerikler çok sayıda ise Twitter'da bulunan trend konular listesi aracılığla insanların kolaylıkla bilgiye ulaşması ve içeriği hızlı bir şekilde tüketmesi sağlanır. Bu konular anlık olarak paylaşıldığından, paylaşıldığı anda bölgesel veya dünya çapında konuşulan popüler etkinlikler veya Twitter üzerinde tartışma konusu olmuş içerikler bu liste içerisinde yer alır. Bu sebepten ötürü, bir konuyu trend konular listesine sokabilmek, bir mesajı olabildiğince çok sosyal medya kullanıcısına ulaşabilmek adına etkili bir yoldur.



Resim 6.1 Twitter trend konu listesinin bir zamandaki görünümü

Gerçek olmayan bir gündem oluşturmak veya varolan bir gündemi daha etkili bir şekilde görünür hale getirmek amacıyla sosyal botlar kullanılırlar. Aynı hashtag kullanılarak botlar aracılığıyla sürekli olarak tweet paylaşılması, paylaşılan hashtagin trend konular listesine girme ihtimali artırır. Bundan dolayı, gerçekleştirdiğimiz deneyi daha gerçekçi yapmak adına, tek aktivitesi her gün aynı hashtagi bir kere tweet atnak olan bir bot senaryosu oluşturulmıştur. Bot hesabın günde yalnızca 1 tweet atma nedeni ise Twitter'ın aynı tweeti günde yalnızca bir kere paylaşmaya izin vermesidir.

6.1.3 Rastgele kelimeler kullanan trend konu botları

Yukarıda da bahsedildiği üzere Twitter'ın koyduğu kısıtlamalar dolayısıyla aynı tweet günde yalnızca 1 kere paylaşılabilmektedir. Bu durum, bot hesapların bir hashtagi sürekli paylaşarak belirli bir konuyu trend konular listesine sokabilme ihtimalini önemli bir ölçüde düşürmektedir. Bunun nedeni ise, trend konu listesinin kısa süreleri içerisinde atılan aynı konudaki içereklere bakılarak oluşturulmasıdır. Buna rağmen, Twitter'ın uyguladığı bu kısıtlama kolaylıkla aşılabilmektedir. Atılan tweeetlerdeki hashtaglerin yanına rastgele bir kelime eklenmesi ve eklenen kelimenin her yeni tweette değiştirilmesi ile bu kısıtlama kolaylıkla atlatılabilir. Bu sayede, yalnızca bir bot hesapla bile bir gün içerisinde aynı hashtagi içeren birçok tweet atılabilir ve bir hashtagin trend konular listesine girme olasılığını artırılabilir.

6.1.4 Propaganda botları

Hashtagler trend konu listelerine sokulmak amacıyla kullanılsa da trend konu listeleri yalnızca hashtaglerden oluşmamaktadır. Sıklıkla kullanılan söz öbekleri de sıklıkla trend konu listelerinde yer alabilmektedirler. Resim 6.1'de rastgele bir anda ekran görüntüsü alınmı Twitter trend konu listesi görülmektedir. Resim 6.1'den görülebileceği üzere konuların bir kısmı hashtaglerden oluşurken diğer kısmı ise birkaç kelimeden oluşan konulardır. Twitter'ın sahip olduğu bu özellik kullanılarak, botlar tarafından ürünler için reklamlar yapılması, yanlış bilginin yayılması ve daha başka birçok konuda insanlar yanlış yönlendirilebilirler.

Bu bot senaryosunda, bot sürekli aynı içeriği paylaşacak şekilde tweetler atmaktadır. Twitter'ın aynı tweetin günde yalnızca bir kere atılabilmesi kuralını aşmak amacıyla, simüle edilmiş şekilde atılan her tweetin içeriği sabit tutulurken tweetin sonuna bir karakter (noktalama işareti, emoji vb.) eklenerek bu kısıt aşılabilmektedir. Böylelikle her bir bot hesap tarafından içeriği aynı birçok tweet atılarak trend konular listesine tweetin içeriğinin girebilmesi amaçlanır.

6.1.5 Reklam botları

Sosyal medya birçok hesap tarafından ürün reklamı yaparak çok sayıda insana erişmek amacıyla kullanılmaktadır. Bu hesaplar, tweet içeriklerinde fotoğraf, video ya da web sayfalarına ait linkler paylaşarak insanların dikkatini çekmek ve kendi web sayfalarına yönlendirerek bir pazarlama stratejisi uygulamaktadırlar. Ek olarak, paylaştıkları tweetlere bir veya birden fazla hashtag ya da kelime öbekleri ekleyerek reklamlarının trend konular listesinde görüntülenmesini sağlarlar. Bu sayede, trend konuları takip eden birçok kişi tarafından reklamları görüntülenerek pazarlama stratejileri başarılı olmuş olur.

6.2 Deney Düzeneği

Twitter üzerindeki bot hesaplar genel olarak politikacılar tarafından rakiplerine karşı avantaj sağlamak amacıyla kullanılmaktadır. Bundan dolayı, senaryo bazlı deneylerde Amerika Birleşik Devletleri'nde seçimler için yarışan politikacılar ana tema olarak kullanılmıştır. 6.1.1'de açıklanan Hayran Botları yalnızca Amerikan Başkanı Donald Trump'ın tweetlerini retweetleyecek şekilde kullanılmıştır. 6.1.2'de açıklanmış olan Trend Konu Botları için ise Donald Trump'ın seçim çalışmaları için kullanılmış olan popüler bir hashtag olan #MAGA (Make America Great Again) ile tweetler simüle edilmiştir. 6.1.3'teki Rastgele Kelime Kullanan Trend Konu Botları için yine #MAGA hashtagi kullanılarak, hashtagin yanına rastgele bir kelime eklenmiş ve böylelikle Twitter'ın günlük aynı tweetin yalnızca bir kere atılabilme kısıtı aşılmıştır. Rastgele eklenen kelimeler bot hesap olduğunun insanlar tarafından kolaylıkla anlaşılabilmesi amacıyla konuyla alakasız olması için meyve isimlerinden seçilmiştir ("Apple #MAGA" gibi). Propanganda botları 6.1.4'te bahsedildiği gibi kelime öbeklerinden oluştuğu için yine Donald Trump'ın "Make America Great

Again" sloganı kullanılarak gerçekleştirilmiştir. Twitter'ın günlük aynı tweeti bir kere atabilme kısıtını aşabilmek adına bu senaryoda rastgele kelimeler yerine noktalama işaretleri ve emojiler gibi tek karakterlik eklemelerle kısıt aşılmıştır. 6.1.5'te bahsedilen Reklam Botları ise tweetleri www.apple.com linkini ekleyerek atmaktadır. Atılan her bir tweete anlık olarak trend konu listesinden bir konu eklenerek kullanıcılar tarafından, yapılan reklamın daha görünür hale gelmesi sağlanmıştır. Her bir tweet içerisinde yalnızca bir trend konu olacak şekilde paylaşılmıştır.

Yapılan deneylerde Botometer kullanılmasının performans dışında en önemli nedeni rahatlıkla kullanılabilecek bir API'ye sahip olmasıdır. Bu API senaryo bazlı deneylerimiz gerçekleştirebilmek için uygun bir yapıdadır. Bu yapı, payload adı verilen, bir Twitter kullanıcısının tüm profil bilgilerini sabit bir format ile alınmasıyla sınıflandırma işlemini gerçekleştirmektedir. Payload API aracılığıyla bir kere alındığında içeriği değiştirilerek farklı bot senaryolarına uygun hale getirilebilmektedir. Bu sayede deneyler, sanal bir Twitter profili üretilerek ve bu profillerin CAP skorları hesaplanarak gerçekleştirilebilir. Burada gerçek hesaplar üzerinden tweet atmak yerine sanal bir Twitter hesabı oluşturulmasının nedeni ise en ufak bir Twitter trafiğinin bile oluşturulmasının önüne geçerek insanların az da olsa manipüle edilme olasılığının önüne geçilmesidir. Özellikle, ilk olarak bot senaryolarına uygun profillere sahip hesaplar bulunarak Botometer API ile hesaba ait payload elde edilmiştir. Daha sonrasında, bot senaryosuna uygun tweetler haricindeki bütün içerikler silinerek hesap yalnızca gerekli tweetlere sahip hale getirilmiştir. İçeriği değiştirilen hesabın profil resmi, biyografi bilgisi, takipçi sayısı, favori tweet sayısı vb. diğer bütün parametreleri sıfırlanarak hesabın yeni açılmış varsayılan bir Twitter hesabı haline getirilmesi sağlanmıştır. Sonrasında tweet içerikleri tam olarak bot senaryosuna uygun olmayan içerikler değiştirilerek deney yapılabilecek hale getirilmiştir. Örneğin Hayran Botları için bakıldığında Donald Trump destekçisi olarak gözüken gerçek bir hesabı ele alındı. Sonraki adımda hesapta bulunan Donald Trump'a ait retweet edilmiş tweetler haricindeki bütün tweetler silindi. Diğer bütün profil parametreleri sıfırlandı. Son olarak ise, hesabın içeriğinde önce 1 retweet olacak şekilde başlanarak bu sayı birer birer artırılarak 25 retweete kadar artırıldı. Her bir değer için CAP skorları hesaplanarak 6.3'te rapor edildi. Trend Konu Botları'na bakıldığında ise her gün #MAGA tweetini paylaşan bir hesap bulunarak

deneyler gerçekleştirilmiştir. Aynı şekilde #MAGA tweetleri haricinde kalan bütün tweetler silinmiş, hesap parametreleri sıfırlanmış ve tweet sayısı birer birer artırılarak skorlar elde edilmiştir. Rastegele Kelimeler Kullanan Trend Konu Botları, Propaganda Botları ve Reklam Botları için senaryolara birebir uygun hesaplar bulunamamıştır. Bu sebepten ötürü, bu senaryolar için istenen tweetler manuel bir şekilde atılıp sonrasında Botometer API ile CAP skorları elde edilmiştir. Gerçek tweet atma işlemi, bot senaryoları için Botometer API kullanılarak hashtaglerde yapılan değişiklik veya varolan hashtaglere kelime veya karakter ekleme gibi değişikliklerin skor üzerinde herhangi bir değişikliğe neden olmadığından gerçekleştirilmiş ve bir manipülasyona sebep olmaması açısından tweetler atıldıktan hemen sonra silinmiştir.

Yapılan deneylerde maksimum tweet sayısı 25 olarak belirlenmiştir. Bunun sebebi ise, aynı tweeti her gün düzenli bir şekilde en fazla 25 kez atmış bir Hayran Botu hesabının bulunabilmiş olmasıdır. Sonrasında her senaryo için 1 ile 25 arasında tüm tweet sayıları birer birer artırılmıştır.

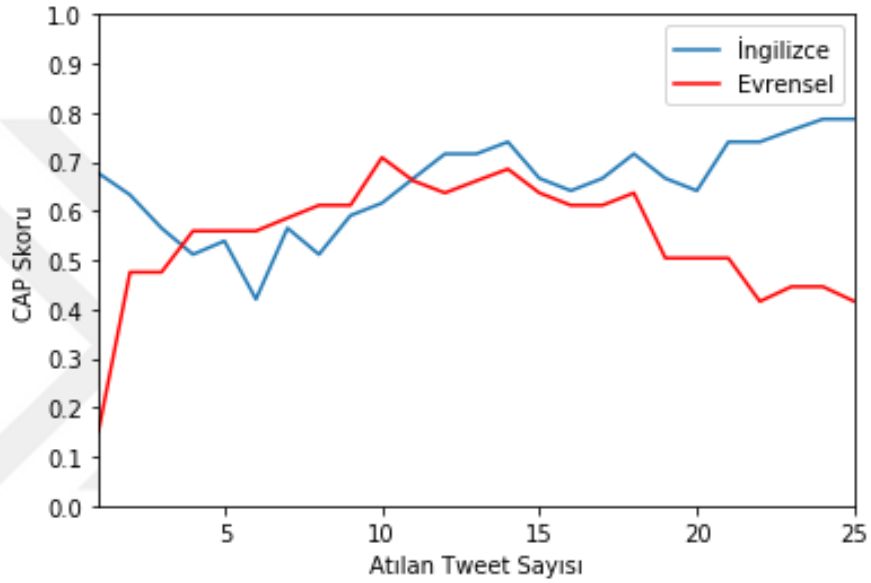
Tüm gerçekleştirilen deneylerde, farklı skorlar elde etmek için tweet sayıları değiştirilmiş ve her bir tweet sayısı için skorlar Botometer API ile alınmıştır. Botometer API döndürdüğü skorlar arasından İngilizce ve Evrensel CAP skorları elde edilerek Bölüm 6.3'te bulunan alt başlıklarda bu skorlar raporlanmıştır. İki farklı skor paylaşılmasına rağmen İngilizce skoru esas olarak alınmalıdır. Bunun nedeni ise, üzerinde çalışılan bütün bot hesaplarının içeriklerinin İngilizce olmasıdır. Skorlara bakıldığında %50 skorunun bir hesabın bot olup olmadığına karar verme konusunda bir karar noktası olabileceği söylenebilir.

6.3 Deney Sonuçları

Bu bölümde Bölüm 6.2'de bahsedilen 5 farklı bot senaryosu için elde edilmiş sonuçlar ve sonuçlara ait değerlendirmeler paylaşılmıştır. Sonuçlara genel olarak bakıldığında, Botometer'ın 5 bot senaryosu arasından 4'ünü tespit edemediği görülmektedir. Her bir senaryoya ait deney sonuçlarına ait detaylı bilgiler alt bölümlerde paylaşılmıştır.

6.3.1 Hayran botları

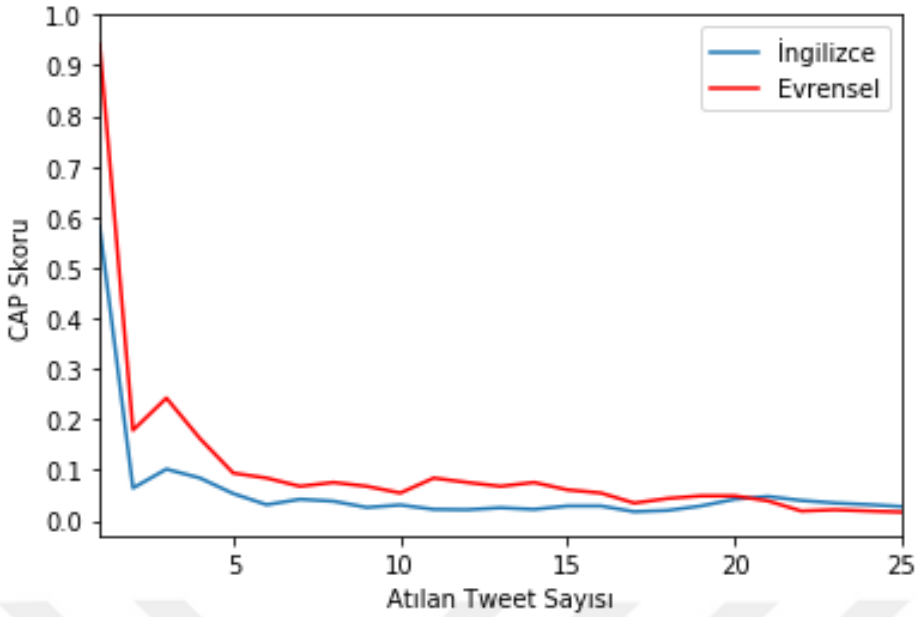
Şekil 6.1 Hayran Botları'na ait Botometer CAP skorlarını göstermektedir. Şekilden görülebileceği üzere Botometer birçok durumda %50'nin üzerinde bir İngilizce skoruna ulaşarak Hayran Botlarını tespit edebilmektedir. Buna rağmen model, 4 ile 6 arasında retweet yapıldığında bu bot senaryosunda başarılı olamamaktadır.Ayrıca ilginç bir şekilde 5 ile 25 retweet değerleri arasına bakıldığında İngilizce CAP skorlarının arttığını görürken, Evrensel skorun düştüğü gözlemlenmektedir.



Şekil 6.1 Hayran botları senaryosuna ait CAP skorları

6.3.2 Trend konu botları

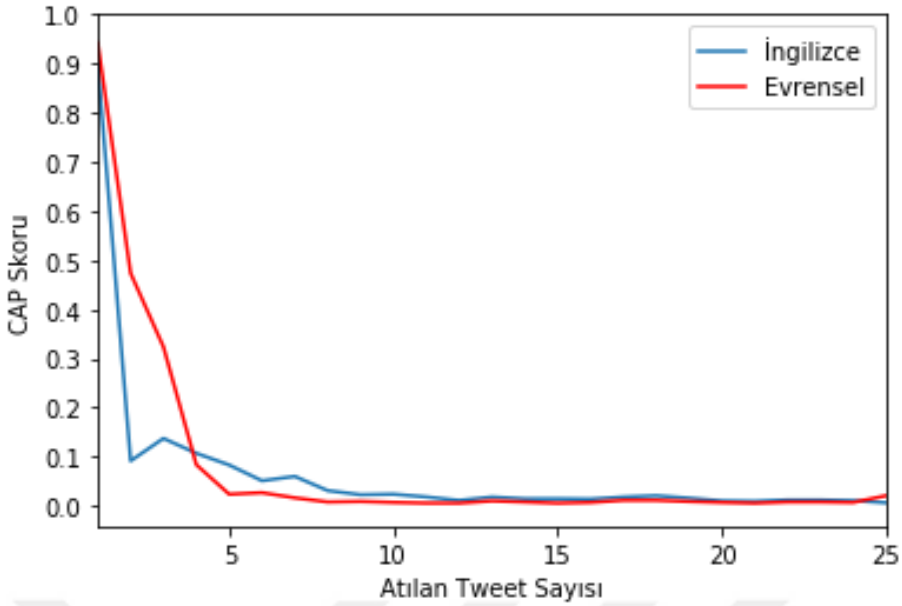
Şekil 6.2'de Trend Konu Botları'nın CAP skorları paylaşılmaktadır. İlginç bir şekilde, Botometer yalnızca bir hashtagli tweet atıldığında bot hesabı tespit edebilmektedir ancak tweet sayısının artışıyla beraber Botometer'ın CAP skorları önemli bir ölçüde azalmaya başlamaktadır. Hatta tweet sayısı artmaya başladıkça elde edilen İngilizce skorlarına bakıldığında, gerçek bir Twitter hesabı için elde edilmesi gereken skorlar bot hesap için elde edilebilmektedir. Botometer'ın bu bot senaryosuna karşı büyük bir zafiyeti olduğu anlaşılsa dahi Twitter'ın günlük tweet limiti bu senaryonun etkinliğini azaltmaktadır. Örneğin, bir hashtagi trend konular listesine sokmak için 10K tweet gerekliyse bu sayıya ancak 10K farklı hesap kullanılarak erişilebilecektir.



Şekil 6.2 Trend konu botları senaryosuna ait CAP skorları

6.3.3 Rastgele kelimelerle kullanılan trend konu botları

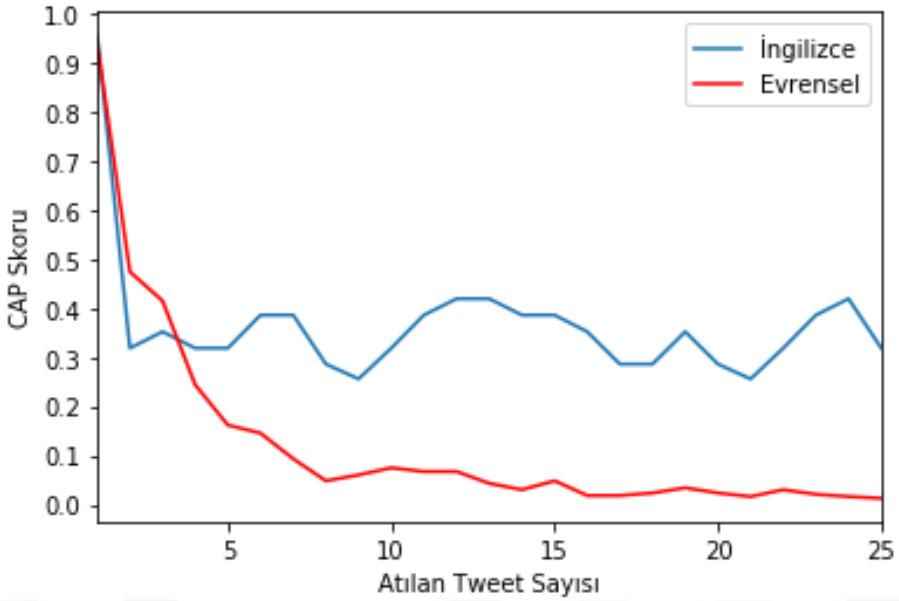
Şekil 6.3 rastgele kelimeler kullanarak Twitter'ın günlük tweet limitini aşmak amacıyla kullanılan bu bot senaryosuna ait CAP skorlarını göstermektedir. Şekil 6.2 ile karşılaştırıldığında, benzer sonuçlar elde edildiği görülebilmektedir. Her iki senaryoda da yalnızca 1 tweet atılmışken çok yüksek CAP skorları elde edilirken tweet sayısı arttıkça elde edilen skorlar önemli ölçüde düşüş göstermektedir. Tweetlere yalnızca bir karakter veya rastgele bir kelime eklenmesine rağmen bot tespit sisteminin düşük skorlar vermeye devam etmesi veri setlerinin yetersizliğinden kaynaklanmaktadır. Eğitim yapılan veri setleri içerisinde belirli bir bot türü bulunmuyorsa o bot türüne karşı makine öğrenmesi algoritmalarının başarı sağlaması mümkün olmayacaktır. Bu yüzden, bir Twitter hesabı kullanılarak bir gün içerisinde aynı tweet, sonuna rastgele kelimeler eklenerek paylaşımlar yapılabilir ve bu şekilde Botometer atlatılmış olur. Bu büyük zafiyet göz önüne alındığında, yalnızca çok az sayıda bot hesap kullanılarak ve sofistike içerikler üreten yapay zeka uygulamalarına ihtiyaç duymadan, istenilen içerikler trend konulara küçük bir çaba harcanarak dahi sokulabilirler.



Şekil 6.3 Rastgele kelimelerle kullanılan trend konu botları senaryosuna ait CAP skorları

6.3.4 Propaganda botları

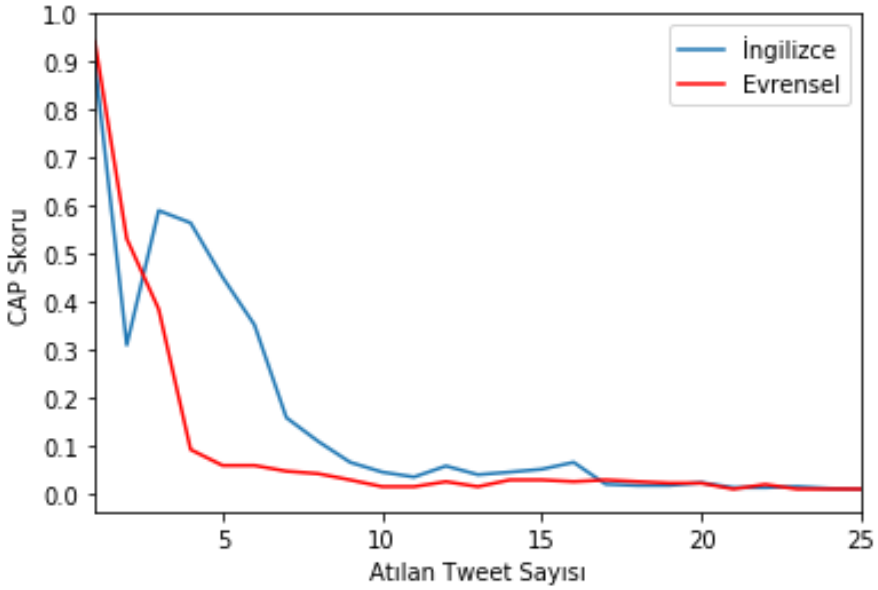
Şekil 6.4 Propaganda Botları'na ait CAP skorlarını göstermektedir. Şekle bakıldığında Evrensel CAP skorunun İngilizce skoruna göre çok daha düşük olduğu görülmektedir. Bunun nedeni sürekli olarak İngilizce bir içerik olan "MAKE AMERICA GREAT AGAIN" kelime öbeğininin kullanılmasıdır. Bu durum, kelime öbeğinin içerdiği kelimeler dolayısıyla duygu analizinin sonuç üzerinde önemli bir rolü bulunduğunu göstermektedir. Buna rağmen İngilizce CAP skorunun yalnızca 1 tweet atılması durumu haricinde 0.5 sınır değerinden düşük olduğu görülmektedir. Böylelikle Botometer'ın bu bot senaryosunda da başarılı olamadığı ve zafiyet gösterdiği anlaşılmaktadır.



Şekil 6.4 Propaganda botları senaryosuna ait CAP skorları

6.3.5 Reklam botları

Şekil 6.5 Reklam Botları için elde edilmiş CAP skorlarını göstermektedir. Önceki bot senaryolarına bakıldığında benzer şekilde bu senaryoda da Botometer'ın yalnızca 1 tweet atıldığı durumlarda çok yüksek CAP skorlarına ulaştığı görülmektedir. Buna rağmen, tweet sayısı 7'yi aştığında CAP skorlarının önemli bir ölçüde düştüğü gözlemlenmektedir. Böylelikle Botometer'ın birçok tweet sayısı için elde ettiği CAP skorlarının çok düşük olduğu görülmektedir. Bu durum, sosyal medya platformlarının reklam aktivitelerine karşı korunmasız olduğunu göstermektedir. Her ne kadar bot tespit sistemleri reklamlara karşı sosyal medya platformlarını koruyabilecek durumda olmasa da Twitter kuralları ve kısıtları reklam içeriklerine karşı hassasiyetle çalışmaktadır. Bu yüzden Twitter kurallarına dair yeni bot senaryoları da geliştirmeli ve gelecekte bu konu ile alakalı kapsamlı bir çalışma yapılmalıdır.



Şekil .5 Reklam botları senaryosuna ait CAP skorları

7. SONUÇLAR

Bu tezde yapılan çalışmalara, ilk olarak farklı özelliklere sahip olan ve farklı veri setleri kullanan 4 bot tespit sistemini (Cresci et. al., Efthimion et. al., Kudugunta ve Ferrara ve Botometer) değerlendirerek başlanmıştır. Veri setlerinde bulunan veri ve parametre eksiklikleri dolayısıyla, kullanılan tüm modeller tüm deney düzeneklerinde çalıştırılamamıştır. Bu yüzden, ilk olarak Çizelge 3.1'de belirtilen ilk 3 model kullanılarak, modellerin orjinal makalelerinde bahsedilen deney düzeneklerin aynı kurularak performansları ölçülmüştür. İkinci olarak, Efthimion et. al. modeli ve Kudungunta ve Ferrara modelleri kullanılarak, bu modeller daha önce görmedikleri veri setleri üzerinde test edilmişlerdir. Son olarak ise, son üç model kullanılıp, her bir model için eğitim ve test veri setleri değiştirilerek 56 senaryo için modellere ait performanslar elde edilmiş ve böylelikle modellerin karşılaştırılabilmesi sağlanmıştır.

Yapılan kapsamlı deneyler sonucunda, önemli bulgulara ulaşılmıştır. Öncelikle modellerin orjinal çalışmalarında elde edilen skorlar ile karşılaştırma yapıldığında elde edilen sonuçlar tutarlılık göstermekte ve modeller yüksek performans skorları ulaşmaktadırlar. Buna rağmen, modeller farklı veri setleri ile testlerin gerçekleştirildiği deney düzeneklerine bakıldığında performanslarını önemli bir ölçüde kaybetmektedirler. Bu durum, bir bot tespit sistemi geliştirilirken, daha başarılı sonıçlar elde etmek ve daha güvenilir değerlendirme yapabilmek adına, daha kapsamlı ve farklı bot türlerine sahip olan veri setlerinin kullanılması gerektiğini göstermektedir.

Yapılan deneylerde elde edilen sonuçlar neticesinde Botometer'ın en yüksek performans skorlarına ulaştığı görülmektedir. Elde edilen skorlar, Botometer'ın literatürde son teknoloji ürünü olarak görülmesi ve deneylerin kolaylıkla gerçekleştirilebildiği bir API'ye sahip olması, Botometer'ın performansının daha isabetli bir şekilde ölçülmesi adına yeni deneyler gerçekleştirme ihtiyacını doğurmuştur. Botometer ile yapılan deneylere bakıldığında ise, daha önce yapılmış

çalışmaların aksine yeni bir bakış açısı ortaya konmuştur. Literatürde bulunan veri setlerini kullanmak yerine, bot türleri tasarlanarak model performansı ölçülmeye çalışılmıştır. Toplamda 5 farklı bot senaryosu tasarlanarak, bu senaryolar ile, Twitter hesaplarının popüleritesini artırmaya çalışan ve içeriğinde hashtag, düz metin veya linkler bulunan tweetler paylaşarak paylaştıkları içeriği trend konu listelerine sokmaya çalışan botların model için bir zafiyet oluşturup oluşturmadığı anlaşılmaya çalışılmıştır. Bot senaryoları, bot tespit probleminde varolan problemin şiddetinin ne denli büyük olduğunun gösterilebilmesi adına, insanlar tarafından kolaylıkla bot oldukları kolaylıkla anlaşılabilecek bir şekilde tasarlanmışlardır. Bot senaryoları ile gerçekleştirilen deneyler sonucunda, Botometer'ın kullanılan bot senaryolarından Hayran Botları haricindeki 4 bot senayosunu tespit edemedği görülmüştür. Bundan dolayı, son teknoloji ürünü dahi olsa, makine öğrenmesi temelli bot tespit sistemlerinin kolaylıkla aldatılabildiği görülmekte ve yeni bakış açılarıyla acil bir şekilde yeni bot tespit sistemlerinin geliştirilmesine ihtiyaç duyulmaktadır.

Deneylerde kullanılan bot tespit sistemlerinin raporladıkları performans değerlerine bakıldığında sosyal bot tespiti probleminin neredeyse çözüldüğü sonucuna varılabilmektedir. Buna rağmen, deneylerimiz sonucunda elde edilen bulgulara bakıldığında bu problemin çözülmekten ziyade, henüz efektif bir çözüm yolu dahi bulunamadığı görülmektedir. Bu sebepten ötürü, daha çok veri ve bot türüne sahip veri setlerine ihtiyaç duyulmaktadır. Bunun nedeni ise, her bir sosyal bot bir programcı tarafından tasarlanmakta ve programcı değiştiğinde botun karakteristik özellikleri de değişebilmektedir ve bu sayede sosyal botlar daha akıllı hale gelebilmektedirler. Bu durum, siber güvenlik uzmanları ile hackerlar arasındaki mücadeleye benzeyen bir bitmeyen savaş olarak tanımlanabilir. Sosyal botlara karşı verilen bu savaşta, tespit sistemlerinin hangi noktalarda yetersiz kaldığını görebilmek adına daha fazla nicel analize ihtiyaç duyulmaktadır.

Mevcut bot tespit sistemlerinin düşük performanslara sahip olduğunun görülmesi üzerine, bu alanda problemlerin devam ettiği ve gelecekte birçok akademik çalışmanın yapılması gerektiği görülmektedir. Bundan dolayı, bu konuyla alakalı gelecekte yapacağımız çalışmalar öncelikli olarak daha sofistike botlar ile yapılacak deneyler sonucunda mevcut modellerin nasıl performanslar sergilediği üzerine olacaktır. Ek olarak bot tespitine etki edebilecek profil resimleri, biyografi bilgileri

ve tweet atma sıklığı gibi tüm faktörler göz önünde bulundurulacak ve bot senaryoları faktörler göz önünde bulundurulacak şekilde tasarlanacaktır. Yukarıda da bahsedildiği üzere, bot tespit probleminde en büyük sorunlardan biri veri seti olarak gözükmektedir. Bundan dolayı, gelecekte yapılacak çalışmalarda modeller tarafından kullanılabilecek farklı bot türlerine sahip yeni veri setlerinin oluşturulabilir. Bu sayede, bot tespit sistemlerinin performansları daha güvenilir bir şekilde test edilebilecektir. Son olarak ise, yeni bir bakış açısı ile oluşturulacak yeni bir bot tespit sistemi oluşturulabilir.

KAYNAKLAR

[1] **Cao, Q., Sirivianos, M., Yang, X., & Pregueiro, T.** (2012). Aiding the detection of fake accounts in large scale social online services. In *Presented as part of the 9th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 12)* (pp. 197-210).

[2] **Yang, C., Harkreader, R., & Gu, G.** (2013). Empirical evaluation and new design for fighting evolving twitter spammers. *IEEE Transactions on Information Forensics and Security*, *8*(8), 1280-1293.

[3] **Paradise, A., Puzis, R., & Shabtai, A.** (2014). Anti-reconnaissance tools: Detecting targeted socialbots. *IEEE Internet Computing*, *18*(5), 11-19.

[4] **Xie, Y., Yu, F., Ke, Q., Abadi, M., Gillum, E., Vitaldevaria, K., ... & Mao, Z. M.** (2012, October). Innocent by association: early recognition of legitimate users. In *Proceedings of the 2012 ACM conference on Computer and communications security* (pp. 353-364).

[5] **Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M.** (2017, April). The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th international conference on world wide web companion* (pp. 963-972).

[6] **Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M., Zheng, H., & Zhao, B. Y.** (2012). Social turing tests: Crowdsourcing sybil detection. *arXiv preprint arXiv:1205.3856*.

[7] **Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M.** (2016). DNA-inspired online behavioral modeling and its application to spambot detection. *IEEE Intelligent Systems*, *31*(5), 58-64.

[8] **Alvisi, L., Clement, A., Epasto, A., Lattanzi, S., & Panconesi, A.** (2013, May). Sok: The evolution of sybil defense via social networks. In *2013 ieee symposium on security and privacy* (pp. 382-396). IEEE.

[9] **Sayyadiharikandeh, M., Varol, O., Yang, K. C., Flammini, A., & Menczer, F.** (2020, October). Detection of Novel Social Bots by Ensembles of Specialized Classifiers. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 2725-2732).

[10] **Yang, K. C., Varol, O., Hui, P. M., & Menczer, F.** (2020, April). Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 01, pp. 1096-1103).

[11] **Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F.** (2016, April). Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web* (pp. 273-274).

[12] **Chavoshi, N., Hamooni, H., & Mueen, A.** (2016, December). Debot: Twitter bot detection via warped correlation. In *Icdm* (pp. 817-822).

[13] **Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M.** (2017). Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling. *IEEE Transactions on Dependable and Secure Computing*, *15*(4), 561-576.

[14] **Efthimion, P. G., Payne, S., & Proferes, N.** (2018). Supervised machine learning bot detection techniques to identify social twitter bots. *SMU Data Science Review*, *1*(2), 5.

[15] **Kudugunta, S., & Ferrara, E.** (2018). Deep neural networks for bot detection. *Information Sciences*, *467*, 312-322.

[16] **Yang, K. C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F.** (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, *1*(1), 48-61.

[17] **Cresci, S., Lillo, F., Regoli, D., Tardelli, S., & Tesconi, M.** (2018). $ FAKE: Evidence of spam and bot activity in stock microblogs on Twitter. In *12th International AAAI Conference on Web and Social Media, ICWSM 2018* (pp. 580-583). AAAI Press.

[18] **Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., & Tesconi, M.** (2019, June). Rtbust: Exploiting temporal patterns for botnet detection on twitter. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 183-192).

[19] **Gilani, Z., Farahbakhsh, R., Tyson, G., Wang, L., & Crowcroft, J.** (2017, July). Of bots and humans (on twitter). In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 349-354).

[20] **Popkin, B.,** (2018, February). Twitter deleteted 200,000 Russian troll tweets. Read them here. https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731, 15.12.2020

[21] **Allyn, B.,** (2020, May). Researchers: Nearly Half of Accounts Tweeting About Coronavirus Are Likely Bots. https://www.npr.org/sections/coronavirus-live-updates/2020/05/20/859814085/researchers-nearly-half-of-accounts-tweeting-about-coronavirus-are-likely-bots, 15.12.2020

[22] **Carley, K. M., Cervone, G., Agarwal, N., & Liu, H.** (2018, July). Social cyber-security. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (pp. 389-394). Springer, Cham.

[23] **Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A.** (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.

ÖZGEÇMİŞ

Ad-Soyad : Muhammet Buğra TORUSDAĞ

Uyruğu : T.C.

Doğum Tarihi ve Yeri : 09.05.1995 - Malatya

E-posta : bugratorusdag@gmail.com

ÖĞRENİM DURUMU:

- **Lisans** : 2018, TOBB Ekonomi ve Teknoloji Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği
- **Yüksek Lisans** : 2020, TOBB Ekonomi ve Teknoloji Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

Yıl	Yer	Görev
2020-halen	TURKCELL	Siber Güvenlik AR-GE Mühendisi
2018-2020	TOBB ETÜ	Tam Burslu Yüksek Lisans Öğrencisi

YABANCI DİL: İNGİLİZCE

TEZDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE PATENTLER:

- **Torusdağ, M. B.**, Kutlu, M., & Selçuk, A. A. (2020, September). Are We Secure from Bots? Investigating Vulnerabilities of Botometer. In *2020 5th International Conference on Computer Science and Engineering (UBMK)* (pp. 343-348). IEEE.