Bilimsel Araştırma Teknileri ve Etik Dersi Dönem Ödevi

12 Aralık 2022

İçindekiler

Şekil Listesi

Tablo Listesi

Özet

Bu rapor bilimsel araştırma teknikleri ve etik dersi kapsamında 5 bildiri, 5 makale 5 tez'i incelemek için yazılmıştır. Çalışmalar incelenirken ders kapsamında öğrenilen kriterlere göre eleştiriler gerçekleştirilmiştir. Yapılan eleştiriler resmi olarak bir anlam ifade etmemektedir, dersi öğrenme amacı ile gerçekleştirilmiş öğrenme yolunda hazırlanan çalışmadır. Bu çalışma ile bilimsel makale hazırlamak isteyen şahsımın tecrübesini kazandırmak amaçlanmıştır. Ders kapsamında verilen görevin tamamlanması hedeflenmiştir. Bu çalışmanın sonucunda bilimsel makaleler incelenerek bilimsel makaleyi nasıl yazarım tecrübesi kısmen edinilmiştir.

Değerlendirmeler

1 Tezler

1.1 Şifreli Ağ Trafiğinin İçerik Açısından Sınıflandırılması

Bozkır [1] tarafından gerçekleştirilen çalışmada problem şifreli ağ trafiğinin sınıflandırılması olarak verilmiştir. Çalışma kapsamı GBTree LightGBM XGBOOST ile trafiği sınıflandırmaa olarak belirlenmiştir. Çalışmada makine öğrenimi ile sınıflandırma yapmak amaçlanmıştır. Hipotez olarak en iyi algoritmanın seçimi seçilmiş ve hipotez sonuç uyumludur. Çalışmada Apache Spark Machine learning ve MLFlow yöntem algoritma yöntemi verilmiştir.
Çalışmanın başlığı (Şifreli Ağ Trafiğinin İçerik Açısından Sınıflandırılması) içeriği anlatan net bir ifade olmuştur.

Çalışmanın özet kısmında problem, kapsam, amaç verilmiştir ancak çıkarım/katkı ve gök gürültülü cümle verilmemiştir.

- problem: Şifreli ağ trafiğnin sınıflandırılması,
- Kapsam: GBTree LightGBM XGBOOST ile trafiği sınıflandırma,
- Amaç: Makine öğrenimi ile sınıflandırma yapmak,
- Çıkarım/Katkı: bulunamamıştır,
- Gök gürültülü cümle: bulunamamıştır.

Çalışmanın 4 anahtar kelime kullanılmıştır. Kullanılan veri seti ve yöntemler verilebilirdi.

Çalışamın giriş bölümünde

- Problemin önemi: net olarak verilmemiştir.
- Literatür tarama kısmında geçmişte yapılan çalışmalar anlatılmış.
- Çalışmanın sonucu direkt olarak belirtilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez belirtilmiştir ve hipotezin ispatı algoritmalar içerisinden iyi sonuç(0.9994) XGBoost algoritması ile elde edildiği belirtilerek yapılmıştır. Kullanılan yöntem olarak Apache Spark Machine learning ve MLFlow yöntem algroitma olarak sunulmuş ve veri seti olarak UNB ISCX VPN-nonVPN 2016 veri kümesi(YouTube, Hangout, Spotify, Facebook, Gmail) belirtilmiştir. Daha önceki çalışmalarla ilgili kıyaslama bulunamamıştır. Üstünlük ve kısıtlar bölümünde algoritmaların üstünlükleri ve zayıflıkları gösterilmiştir ancak açık değildir. Çıktıların kritik edilmesinde olumlu özelliklerden bahsedilmiştir.
Çalışmanın sonuç bölümü basit ve açıktır, tekrarlar bulunmaktadır ve geçmiş zaman kullanılmamıştır, sonuç net(XGBoost algoritmasının her iki görev kapsamında ortalama başarısının en iyi sonuca ulaştığı görülmektedir) olarak verilmiş ve yöntemden bahsedilmiştir.
Çalışmada 35+ kaynak kullanılmış güncel bulunmakadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.

Şekil 1: Bilimsel Araştırma Kalitesi

1.2 Yaşanan Terör Olaylarını İçeren Büyük Verinin Makine Öğrenmesi Teknikleri ile Analizi

Karabay [5] tarafından gerçekleştirilen çalışmada problem Terör olaylarının tespitinin zorluğu olarak verilmiştir. Çalışma kapsamı Global Terrorism Database (GTD) veri kümesi olarak belirlenmiştir. Çalışmada bir terör olayının hangi örgüt tarafından gerçekleştirildiğini tahmin eden bir uygulama geliştirmek amaçlanmıştır. Hipotez teror olaylarının tahmınnde makine ogrenmesinin etkili olması ve hipotez sonuç uyumludur. Çalışmada yöntem Machine learning olarak verilmiştir.
Çalışmanın başlığı (Yaşanan Terör Olaylarını İçeren Büyük Verinin Makine Öğrenmesi Teknikleri ile Analizi) içeriği anlatan net bir ifade olmuştur.

Çalışmanın özet kısmında problem, çıkarım/katkı, kapsam, amaç ve gök gürültülü cümle verilmiştir.

- problem: Terör olayının hangi örgüt tarafından gerçekleştirildiğini tahmini.
- Kapsam: Global Terrorism Database (GTD) veri kümesi.
- Amaç: Bir terör olayının hangi örgüt tarafından gerçekleştirildiğini tahmin eden bir uygulama geliştirmek.
- Çıkarım/Katkı: Yaşanan bir terör olayının hangi örgüt tarafından gerçekleştirildiğini tahmin eden biruygulama geliştirilmiştir.
- Gök gürültülü cümle: Uygulanan algoritmalarda en yüksek doğruluk oranı K-En Yakın Komşu (KNN) algoritması ile %98,2 olarak bulunmuştur

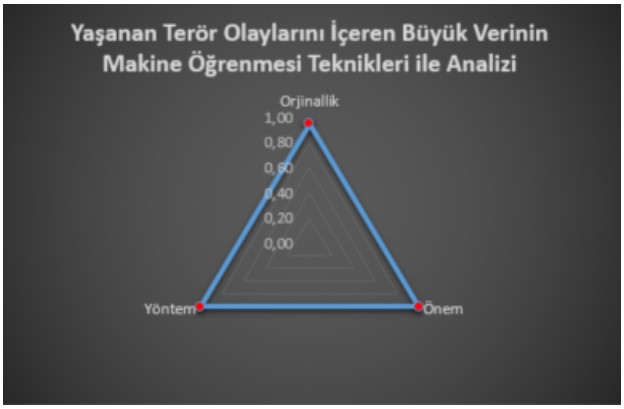
Çalışmanın 4 anahtar kelime kullanılmıştır. Kullanılan veri seti verilse idi faydalı olurdu.

Çalışamın giriş bölümünde

- Problemin önemi: Terör olayları yüzünden ülkelerin ekonomileri, insanların psikolojisi zarar görmektedir. Yaşanan terör olayları veya terör saldırıları binlerce insanın hayatının kaybetmesine ve kalıcı hasar bırakacak yaralanmalara sebebiyet vermiştir
- Literatür tarama kısmında geçmişte yapılan çalışmalar anlatılmış ve teknolojik dezavantajardan bahsedilmiş.
- Çalışmanın sonucu direkt olarak belirtilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez belirtilmiştir ve hipotezin ispatı yapılmıştır. Kullanılan yöntem olarak ve veri seti net değildir. Daha önceki çalışmalarla ilgili kıyaslama bulunamamıştır. Üstünlük ve kısıtlar bölümünde AAlgoritmaların üstünlükleri ve zayıflıkları gösterilmiş. Çıktıların kritik edilmesinde, 10 farklı teror orgutunun faaliyetlerini data set ile incelenmiş 6 farklı machine learning algoritmasının sonuçları karşılaştırılmış.

Çalışmanın sonuç bölümü basit ve açıktır, tekrarlar bulunmaktadır, alıntı yapılmıştır ve geçmiş zaman kullanılmamıştır, sonuç net olarak verilmiştir ve yöntemden bahsedilmiştir.
Çalışmada 45 kaynak kullanılmış güncel bulunmakadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 2: Bilimsel Araştırma Kalitesi

1.3 Apache Spark ve Makine Öğrenmesi Algoritmalari ile Ağ Saldirisi Tespiti

Kurt [6] tarafından gerçekleştirilen çalışmada problem ağ verilerinin büyümesinin sonucu ağ trafiğindeki saldırılarının tespitinin zorlaşması olarak verilmiştir. Çalışma kapsamı Apache Spark kullanılarak KDD Cup'99 veri seti olarak belirlenmiştir. Çalışmada makine öğrenmesi algoritmalarının aynı ağ verileri üzerindeki performanslarını karşılaştırarak geliştirilmekte olan saldırı tespit sistemlerine referans kaynak oluşturmak amaçlanmıştır. Hipotez gerçekleştirilmek istenen amacı bir modele dönüştürmektir ve hipotez sonuç uyumludur. Çalışmada yöntem Lojistik Regresyon , Rastgele Orman , Naive Bayes, Destek Vektör Makineleri KDD Cup'99 olarak verilmiştir.
Çalışmanın başlığı (Apache Spark ve Makine Öğrenmesi Algoritmalari ile Ağ Saldirisi Tespiti) içeriği anlatan ve gerek siz teknik ifade içeren bir başlık olmuştur.

Çalışmanın özet kısmında problem, çıkarım/katkı, kapsam ve amaç verilmiştir, gök gürültülü cümle bulunamamıştır.

- problem: Ağ verilerinin büyümesinin sonucu ağ trafiğindeki saldırılarının tespitinin zorlaşması.
- Kapsam: Apache Spark kullanılarak KDD Cup'99
- Amaç: Makine öğrenmesi algoritmalarının aynı ağ verileri üzerindeki performanslarını karĢılaĢtırarak geliĢtirilmekte olan saldırı tespit sistemlerine referans kaynak oluşturmaktır.
- Çıkarım/Katkı: bulunamadı. Elde edilen sonuçlar, Apache Spark teknolojisinin büyük ağ verileri üzerindeki saldırıları tespit etmede giderek daha etkili olduğunu göstermiştir.
- Gök gürültülü cümle: bulunamadı

Çalışmanın 5 anahtar kelime kullanılmıştır.

Çalışamın giriş bölümünde

- Problemin önemi: Saldırganların ve saldırı teknikelrinin çoğalması ile klasik yöntemlerin yeterli gelemeyeceği davranış bazlı tespit sistemlerinin önermi vurgulanmıştır
- Literatür tarama kısmında geçmiş çalışmalar ve saldırı tespit sistemlerinin bilgileri verilmiş. Bunların başarı durumları genel ifadelerle geçilmiş.
- Çalışmanın sonucu direkt olarak belirtilmiştir.

- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez belirtilmiştir ve hipotezin ispatı yapılmıştır. Kullanılan yöntem olarak ve veri seti net değildir. Daha önceki çalışmalarla ilgili tekbir çalışma kıyaslanmış. 2019 ocakta biten bu çalışma için 2016 ve 2017 yılında 3 farklı çalışma daha var bunlarla kıyas yapılmamıştır. Üstünlük ve kısıtlar bölümünde Apache sparkı üstünlük olarak vermiştir, veri setinin ise eski olduğu belirtilerek güncel olmakta kısıtlar yaşadıkları belirtilmiş. Çıktıların kritik edilmesinde, bazı atak tiplerini doğru tahmin ederken örneğin rootkit gibi atakların tespitinde bu modelin başarızı olduğu belirtilerek objektif bir bakış açısı yakalanmış.
Çalışmanın sonuç bölümü basit ve açıktır, tekrarlar bulunmaktadır ve geçmiş zaman kullanılmamıştır, sonuç net olarak verilmiş ve yöntemden bahsedilmiştir. Olumsuz olarak tablolar ayrıntılı olarak burada verilmiştir.
Çalışmada 43 kaynak kullanılmış güncel ve güncel olmayan kaynaklar bulunmakadır. https://www.mustafaorhan.com/genel nedir/ erişilemeyen bir sayfa bulunmaktadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.

Şekil 3: Bilimsel Araştırma Kalitesi

1.4 Machine Learning Algorithms Implementation and Evaluation on Apache Spark Pyspark

İnanır [15] tarafından gerçekleştirilen çalışmada problem net olarak verilmemekle birlikte iade ya da geri dönen ürünlerin oluşturduğu kayıplar olarak belirlenebilir. Çalışma kapsamı 1968 örneği 8 özelliği bir sınıf öz niteliği olan veri seti olarak belirlenmiştir. Çalışmada geri dönen veya iptal edilen ürünlerden kaynaklanan finansal ve operasyonel kayıpların ön bilgisini perakende firmalarına verebilmek amaçlanmıştır. Hipotez net olarak görülememiştir ancak başlık hipotez yerine geçebilir ve hipotez sonuç uyumludur. Çalışmada yöntem Machine learning Logistic Regression, Decision Tree, Random Forest, gradient boosted tree, Naïve Bayes apache spark olarak verilmiştir.
Çalışmanın başlığı (Machine Learning Algorithms Implementation and Evaluation on Apache Spark Pyspark) içeriği tam olarak anlatmayan ve gerek siz teknik ifade içeren bir başlık olmuştur.

Çalışmanın özet kısmında problem, çıkarım/katkı ve gök gürültülü cümle bulunamamıştır, kapsam ve amaç verilmiştir.

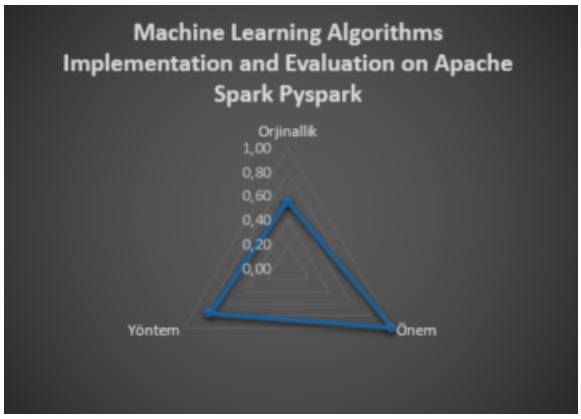
- problem: Büyük Verinin yaygın kullanımı, somut bir şekilde birçok güvenlik sorunlarını beraberinde getirmektedir.
- Kapsam: 1968 örnekli 8 özellikli ve 1 sınıf özelliği bulunan veri seti.
- Amaç: Geri dönen veya iptal edilen ürünlerden kaynaklanan finansal ve operasyonel kayıpların ön bilgisini perakende firmalarına verebilmektir.
- Çıkarım/Katkı: bulunamadı.
- Gök gürültülü cümle: bulunamadı

Çalışmanın 3 anahtar kelime kullanılmıştır. Cost , retail gibi ifadeleri içeren anahtar kelimeler de eklenmeliydi.

Çalışamın giriş bölümünde

- Problemin önemi:belirtilmemiştir.
- Literatür tarama kısmında geçmiş çalışmalardan bahsedilmiştir.
- Çalışmanın sonucu direkt olarak belirtilmemiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez net olarak belirtilmesede hipotezi çalışmadan anladığımız hiteze göre modelin çalıştığı ve hipotezin ispatı yapılmıştır. Kullanılan yöntem olarak ve veri seti net değildir. Daha önceki çalışmalarla ilgili tespit doğruşu karşılaştırmaları verilmemiştir. Üstünlük ve kısıtlar bölümünde mevcut problemlerden bahsedilmemiştir. Çıktılar kritik edilerek çalışma tamamlanmıştır.
Çalışmanın sonuç bölümü basit ve açıktır, tekrarlar bulunmaktadır ve geçmiş zaman kullanılmamıştır, sonuç net olarak verilmiş ve yöntemden bahsedilmiştir.
Çalışmada 50+ kaynak kullanılmış güncel ve güncel olmayan kaynaklar bulunmakadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 4: Bilimsel Araştırma Kalitesi

1.5 Büyük verilere yetkisiz erişimlerin tespit edilmesi ve engellenmesi tekniklerinin incelenmesi ve uygulaması

Toy [13] tarafından gerçekleştirilen çalışmada problem olarak büyük Verinin yaygın kullanımı, somut bir şekilde birçok güvenlik sorunlarını beraberinde getirmesi net olarak verilmiştir. Çalışma kapsamı Hadoop, Spark ve Storm belirlenmiştir. Çalılmada büyük verinin önemine dikkat çekmek ve öneriler ortaya koymak amaçlanmıştır. Hipotez net olarak görülememiştir ancak başlık hipotez yerine geçebilir ve hipotez sonuç uyumludur. Çalışmada yöntem net değildir.
Çalışmanın başlığı (Büyük verilere yetkisiz erişimlerin tespit edilmesi ve engellenmesi tekniklerinin incelenmesi ve uygulaması) içeriği özetleyen bir başlıktır.

Çalışmanın özet kısmında çıkarım katkı kısmı bulunamamıştır, problem, kapsam, amaç ve gök gürültülü cümle verilmiştir.

- problem: Büyük Verinin yaygın kullanımı, somut bir şekilde birçok güvenlik sorunlarını beraberinde getirmektedir.
- Kapsam: Hadoop, Spark ve Storm Platformlarının en popüler olduğu görülerek, bu üç platformun benzer yönleri ve farklı tarafları
- Amaç: hayatımızı büyük ya da küçük oranda etkilemekte olan Büyük Verinin güvenlik boyutunu inceleyerek, bu konunun önemine dikkat çekmek ve öneriler ortaya koymaktır

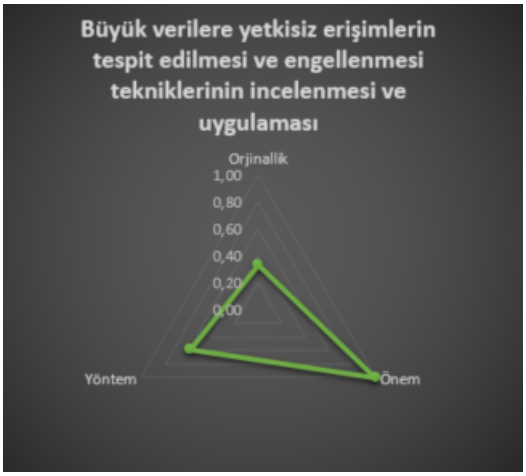
- Çıkarım/Katkı: bulunamadı.
- Gök gürültülü cümle: üyük Veri güvenliğinin ihmal edildiği takdirde "Büyük Verinin Büyük Probleme" dönüşebileceği değerlendirilmektedir. Adli Bilişim alanı için Büyük Veri göz korkutucu güçlükler sunmaktadır.

Çalışmanın 6 anahtar kelime kullanılmıştır. Yetkisiz erişim anahtar kelimelere eklenmeliydi.

Çalışamın giriş bölümünde

- Problemin önemi: Büyük verinin doğası gereği, birbiri ile ilişkili bilgileri araştırmak samanlıkta iğne aramak gibidir. Günümüzde mevcut olan geleneksel Adli Bilişim araçları Büyük Veriyi işlemek için yeterli düzeyde değildir. İrlanda Ulusal Polisi Garda Siochana Müfettişliği tarafından hazırlanan bir rapora göre ele geçirilen cihazlarla ilgili adli soruşturmalarda dört yıla varan gecikmeler yaşanması olarak verilmiştir.
- Literatür tarama kısmı bulunamamıştır.
- Çalışmanın sonuç olup tespit edilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde öne sürülen hipotezin ispatı açık olarak verilmemiştir. Kullanılan yöntem olarak ve veri seti net değildir. Daha önceki çalışmalarla ilgili tespit doğruşu karşılaştırmaları verilmemiştir. Üstünlük ve kısıtlar bölümünde mevcut problemlerden bahsedilmemiştir. Çıktılar kritik edilerek çalışma tamamlanmıştır.
Çalışmanın sonuç bölümü basit ve açık değildir tekrarlar bulunmaktadır. ve geçmiş zaman kullanılmamıştır, yöntemden bahsedilmiştir.
Çalışmada 15 kaynak kullanılmış ve 30+ üniversite ve organizayon web sayfası bulunmaktadır, güncel kaynaklar kullanılmıştır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 5: Bilimsel Araştırma Kalitesi

2 Bildiriler

2.1 Big data and data science: what should we teach?

Song ve diğerleri[11] tarafından gerçekleştirilen çalışmada problem büyük veri öğretimi verilen bölümlerde verilmesi gereken belirlenmesini net olarak verilmiştir. Çalışma kapsamı büyük veri, ABD'De bu bilimi öğreten okullar olarak belirlenmiştir. Fen eğitimin başarılı olması için çeşitli yaklaşımları önermeyi amaçlanmıştır. Hipotez olarak neler öğrenilemesi gerekiyor ortaya atmışladır ve hipotez sonuç uyumluluğu bulunmaktadır. Çalışmada yöntem olarak geniş bir keşif çalışması kullanılmıştır.

Çalışmanın başlığı (Big data and data science: what should we teach?) içeriği tam olarak özetleyen bir başlık olmuştur.

Çalışmanın özet kısmında Gök gürültülü cümle bulunamamıştır, problem, kapsam amaç ve çıkarım/katkı verilmiştir.

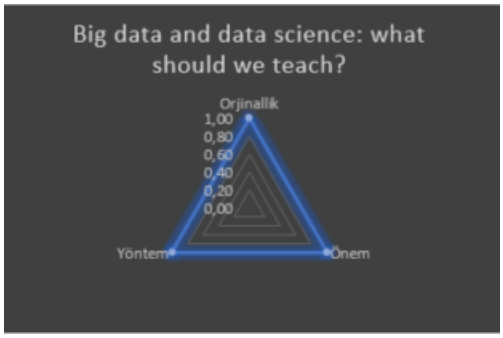
- problem: Büyük verinin öğrenimindeki öğrenilmesi gerekenler
- Kapsam: büyük veri, ABD'De bu bilimi öğreten okullar
- Amaç: Fen eğitimin başarılı olması için çeşitli yaklaşımları önermek,
- Çıkarım/Katkı: Big Data için neyi öğrenmeliyizi vermektedir.
- Gök gürültülü cümle: büyü veri eğitimi eğitim için anahtar keşme ve geleceğin veri bilimcilerini en iyi şekilde eğitmek için strateji ve yaklaşımlara ihtiyacımızın olması.

Çalışmanın 5 anahtar kelime kullanılmıştır.

Çalışamın giriş bölümünde

- Problemin önemi: büyük verinin yorumlanması anlaşılması işlenmesi ve bunları bilgiye ve bilgi birikimine döndürülmesi bir zorluk olarak görülmektedir.
- Literatür taramalarında genel bilgileri verilmiştir.
- Çalışmanın sonuç olup tespit edilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde öne sürülen hipotezin ispatı sonuçlar ve grafik ile gerçekleştirilmiştir. Kullanılan yöntem olarak veri bilimi eğitimi verilen programları müfredatları karşılaştırılmıştır. Daha önceki çalışmalarla ilgili tespit doğruşu karşılaştırmaları verilmemiştir, üstünlük ve kısıtlar bölümünde mevcut problemlerden bahsedilmiştir. Çıktılar kritik edilerek çalışma tamamlanmıştır.
Çalışmanın sonuç bölümü basit ve açıktır tekrar yapılmamıştır ve geçmiş zaman kullanılmıştır, yöntemden bahsedilmiştir.
Çalışmada 15 kaynak kullanılmış ve 30+ üniversite ve organizayon web sayfası bulunmaktadır, güncel kaynaklar kullanılmıştır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 6: Bilimsel Araştırma Kalitesi

2.2 Big Data for Remote Sensing: Challenges and Opportunities

Chi ve diğerleri[2] tarafından gerçekleştirilen çalışmada problem gemilerden yağ sızması ve terörist ataklar net olarak verilmiştir. Çalışma kapsamı büyük veride uzaktan algılama olarak net verilmiştir. Dağıtık olarak çalışan bir içerik çekme sistemi amaçlanmıştır. Hipotez denizde yağ sızdırılmasının ve terörist atakların big datada tespiti ortaya atılmıştır ve hipotez sonuç uyumluluğu bulunmaktadır. Çalışmada yöntem Support Vector Machine kullanılarak sağlanmıştır.

Çalışmanın başlığı (Big Data for Remote Sensing: Challenges and Opportunities) içeriğe bakaran çok genel kalmıştır.

Çalışmanın özet kısmında Gök gürültülü cümle bulunamamıştır, problem, kapsam amaç ve çıkarım/katkı verilmiştir.

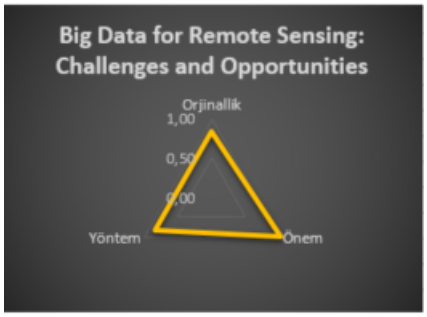
- problem: Terörist ataklarını ve denizde yağ sızdırma problemini tespiti,
- Kapsam: Remote sensing big data,
- Amaç: Denizde yağ sızdıra ve terorist arak tespiti
- Çıkarım/Katkı: önemli zorluk ve fırsatları göz önüne sermek olarak net olarak verilmiştir.
- Gök gürültülü cümle: bulunamamıştır.

Çalışmanın 6 anahtar kelime kullanılmıştır.

Çalışamın giriş bölümünde

- Problemin önemi: Denizlere yağ sızdırılmasının ve terörist atakaların olumsuz etkileri anlatılmıştır.
- Literatür taramalarında genel bilgileri verilmiş ancak detaya inilebilirdi.
- Çalışmanın sonuç olup olmadığı tespit edilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde öne sürülen hipotezin ispatı sonuçlar ve grafik ile gerçekleştirilmiştir. Kullanılan yöntem olarak support vector machine (SVM) belirtilmiştir. Daha önceki çalışmalarla ilgili tespit doğruşu karşılaştırmaları verilmemiştir, üstünlük ve kısıtlar bölümünde mevcut problemlerden bahsedilmiştir. Çıktılar kritik edilerek çalışma tamamlanmıştır.
Çalışmanın sonuç bölümü basit ve açık değildir tekrar yapılmıştır ve geçmiş zaman kullanılmamıştır, yöntemden bahsedilmiştir.
Çalışmada 56 kaynak kullanılmış ve güncel kaynaklar kullanılmıştır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 7: Bilimsel Araştırma Kalitesi

2.3 Language Based Web Crawlıng on Big Data

Girgin ve diğerleri[4] tarafından gerçekleştirilen çalışmada problem dolaylı bir biçimde anlatılmıştır net olarak verilmemiştir. Çalışma kapsamı Türkçe içerikli internet sayfasından oluşan büyük veri olarak net verilmiştir. Dağıtık olarak çalışan bir içerik çekme sistemi amaçlanmıştır. Hipotez net olarak verilmemiştir bu yüzden hipotez sonuç uyumluluğu uygulanamamıştır. Çalışmada yöntem hadoop altyapısı kullanılarak sağlanmıştır.
Çalışmanın başlığı tükçe ve ingilizce olarak verilmiş olup tam olarak çalışmayı vermemektedir. Türkçe başlıkta(BÜYÜK VERİDE DİL ODAKLI İÇERİK ÇEKME) çekilen içerik nereden geleceği belli değilken ingilizce olanda belirgindir.

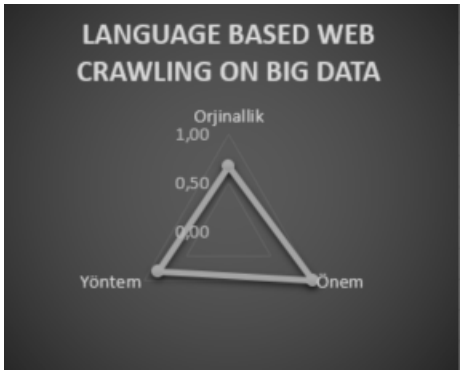
Çalışmanın 3 anahtar kelime kullanılmıştır.
Çalışmanın özet kısmında Çıkarım/Katkı ve Gök gürültülü cümle net bir şekilde verilmemiştir sadece Problem, kapsam, amaç net bir şekilde verilmiştir.

- Problem: Çevirim içi verilere kolay ve hızlı erişim ihitacı özetten çıkarabilir.
- Kapsam: 4729 internet sayfasını içeren veri seti üzerinde çalışma yapılacacağı bilgisi alınmıştır.
- Amaç: Hadoop üzerinde da˘gıtık olarak çalı¸san bir içerik çekme sistemi
- Çıkarım/katkı: bulunmamaktadır.
- Gök gürültülü Cümle: bulunmamaktadır.

. Çalışamın giriş bölümünde

- Problemin önemi: Bazı uygulamalar için sadec belli bir dildeki ya da belli konulardaki verilere ihtiyaç duyulmasının yanı sıra, disk alanı ve işlemci gibi kaynakların sınırlı olması problem olarak gösterilmiştir.
- Literatür taramalarında geçmiş çalışmalar yüzeysel olarak anlatılmıştır.
- Çalışmanın sonuç olduğu açıkça yazılmıştır.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

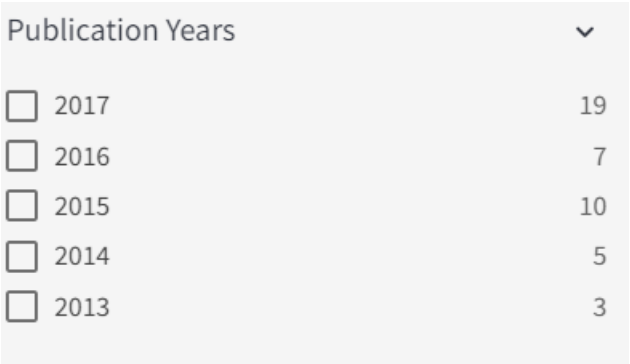
Çalışma metodolojisinde öne sürülen hipotezin ispatı sonuçlar ve grafik ile gerçekleştirilmiştir. Kullanılan yöntem olarak Hadoop, HBase ve Zookeper belirtilmiştir. Daha önceki çalışmalarla ilgili tespit doğruşu karşılaştırmaları verilmemiştir, üstünlük ve kısıtlar net olarak verilmemiştir. Çıktılar kritik edilerek çalışma tamamlanmıştır.
Çalışmanın sonuç bölümü basit ve açık olarak tekrar yapılmadan geçmiş zamanda ve yöntem belirtilerek verilmiştir. Elde edilen sonuç %98 doğruluk değeri net bir biçimde paylaşılmıştır.
Çalışmada 15 kaynak kullanılmış ve bunların 7'sinin web teknoloji sayfası olduğu görülmüştür. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 8: Bilimsel Araştırma Kalitesi

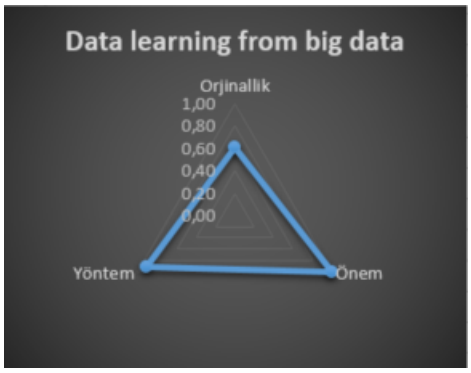
2.4 Data Learning From Big Data

Torrecilla ve diğerleri[12] büyük veride bazı durumları tartışmak ve "data learning" kavramını litaretüre katmak amaçlı çalışma gerçekleştirmişlerdir. Çalışma içerik açısından incelendiğinde problem net olarak verilmemiştir. Çalışma kapsamı yine net olarak verilmemiş olup "big data" olarak ele alınabileceği makale içerisinden çıkarılabilmektedir. Çalışmada hipotez "data learning" kavramını literatüre sokmaktır, yöntem ise bulunamamıştır. Hipotez'in sonuçla uygunluğu olarak bu kavramı destekleyecek bilgilere rastlanılmamıştır. Çalışma başlığı(Data learning from big data) anlam bakımından biraz garip gelmiştir, veri tek başına anlam ifade etmeyen birbiri arasında örüntüler kurulduğunda anlam ifade eden veriyi, veri öğrenme olarak hedef konulmuştur. Yine de literatür kontrol edildiğinde "data learning" tabirinin makalelerde kullanıldığı hatta hali hazırda zaten kullanıdığı görülmütür. Web of science'da "data learning" ve "big data" sözcükleri & işlememine tabi tutularak araştırma yapıldığında bu çalışmanın yapıldığı 2018 yılından önce zaten bu tabirin geçmiştede kullanıldığı bu makalenin literatüre yeni bir deyim kazandırmadığı ortaya çıkarmıştır ancak yinede bu terimin varlığından bahseden



Şekil 9: 2017 ve öncesinde data learning ibaresi geçen big data çalışma sayıları.

ilk makaledir. Makalenin özet kısmında problem tanımı görülmemiştir. Kapsamın big data olduğu özetin genelinden anlaşılmaktadır. Çalışmanın iki amacı bulunmaktadır birincisi big data ile yükselen bazı durumalarla ilgili olarak istatistiğin rolü ve "data learning" ifadesi literatüre sokmak. Bu bakımdan başlık tek bir amacı ihtiva ettiğinden eksik kalmıştır. "data learning" ifadesini literatüre sokmak gök gürültülü bir cümle olarak algılanmıştır. Çalışma başlık yapısı bakımından giriş gelişme sonuç şeklinde yapılmadığı görülmüştür bu okunurluğu zorlaştırmıştır. Giriş bölümünde konunun önemi big datanın son yıllarda sıkça konuşulduğu ve bu bilgilerin çeşitli unsuralarca kullanıldığından bahsedilerek vurgulanmıştır. Geçmiş çalışmalardan bahsedilerek konuya açıklık getirilmeye çalışılmıştır ve tarihsel bir anlatımda bulunmayarak akıcılık sağlanmıştır. çalışmanın bir sonuçmu ön sonuçmu olduğundan bahsedilmemiştir. Çalışmanın diğer çalışmalara kıyasla üstünlük ve kısıtlarından, çıktıların kritik edilmesi gibi bir alana rastlanılmamıştır. Sonuç kısmı ise ayrı başlıkta değil sadece bir paragrafta verilmiştir ve basit değildir. Sonucu kararsız ibarelerle inanca veya tahmine dayalı olarak belirtmişlerdir, bilimsel bir ifade ile sonuçlandırılmamıştır. Ayrıntı, tekrar olmaması faydalı olmuş ancak geçmiş zaman kullanılmaması, yöntemin yazılmaması eksi puan olarak görülmüştür. Çalışmada 26 kaynak kullanılmış, kaynakların güncel olduğu gözlemlenmiştir. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 10: Bilimsel Araştırma Kalitesi

2.5 S-DDoS: Apache spark based real-time DDoS detection system

Patil ve diğerleri[9] Hizmet durdurma saldırılarını problem olarak gösterip ve kapsam olarak belirleyerek bu problemi çözmeye yönelik apache spark ile gerçek zamanlı DDOS tespiti yapabilen S-DDOS sistemi yöntemini geliştirmeyi amaçlamışlaştır. Hipotez olarak yüksek hızlı dağıtık DDOS tespit sistemini öne sürmüşlerdir. Araştırmanın sonucu olarak apache spark ile bir ddos tespit sistemi gerçekleştirdikleri için hipotez tez sonuç uyumluluğu sağlanmıştır.
Çalışma başlığı(S-DDoS: Apache spark based real-time DDoS detection system) çalışmayı net bir şekilde özetleyen yapıdadır.

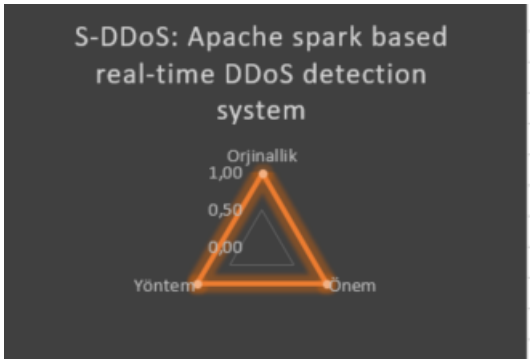
Çalışmanın özet kısmında Problem, kapsam, amaç, Çıkarım/Katkı ve Gök gürültülü cümle net bir şekilde verilmiştir.

- Problem: Daha önceki DDOS tespit sistemlerinde tespit yönteminde yavaş tepsit, gerçek zamanlı tepki vermenin olmaması ve bu yöntemlerin big data için uygun olmaması problemi ortaya atılmıştır.
- Kapsam: DDOS detection, Big Data ve apache spark.
- Amaç: Yeni bir gerçek zamanlı dağıtık sistem gerçekleştirmek-¿S-DDOS
- Çıkarım/katkı: Spark ile trafiği sınıflandırabilen ve atakları canlı olarak tespit edebilen sistem geliştirilmiştir.
- Gök gürültülü Cümle: The results show that the proposed S-DDoS detection system efficiently detects the DDoS attack from network traffic flows with higher detection accuracy (98Geliştirilen S-DDOS sistemi %98 doğrulukla atakları gerçek zamanlı olarak tespit ettiği belirtilmiştir.

Çalışmanın anahtar kelimeleri yeterli seviyededir.
Çalışamın giriş bölümünde

- Problemin önemi: teknolojideki büyük gelişmeler ile birlikte internetteki zaafiyetlerde büyümesi ve belkide en büyüğü ve en zorlusu Big data ortamında en iyi sınıflandırma ve gerçek zamanlı tepkime süresi ile DDOS trafiğini tanımlama olarak gösterilmiştir.
- Literatür taramalarında geçmiş sistemlerdeki teknikler karşılaştırılmıştır.
- Çalışmanın sonuç olduğu açıkça yazılmamış ancak okuma sonunda bir sonuç olduğu ortaya çıkmaktadır.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metordojisind öne sürülen hipotezin ispatı sonuçlar ve resimlerle gerçekleştirilmiştir. Kullanılan yöntem apache spark machin learning ve gerçek zamanlı veri seti olarak belirtilmiştir. Daha önceki çalışmalarla ilgili tespit doğruşu karşılaştırmaları verilmiş, gecikme ve ölçeklenebilirlik özelliklerinin üstünlükleri ve kısıtları ortaya konulmuştur. Çıktılar kritik edilerek çalışma tamamlanmıştır.
Çalışmanın sonuç bölümü basit ve açık olarak tekrar yapılmadan geçmiş zamanda ve yöntem belirtilerek verilmiştir. Elde edilen sonuç %98 doğruluk değeri net bir biçimde paylaşılmıştır.
Çalışmada 28 kaynak kullanılmış, kaynakların güncel olduğu gözlemlenmiştir. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 11: Bilimsel Araştırma Kalitesi

3 Makaleler

3.1 Yazılım Tanımlı Ağlar – YTA

Cicioğlu ve diğerleri [3] tarafından gerçekleştirilen çalışmada problem geleneksel ağ yaklaşımı hantal ve hata eğilimli olması ve verimli şekilde kullanılmasına da olanak tanımaması olarak verilmiştir. Çalışma kapsamı Yazılım Tanımlı Ağlar olarak belirlenmiştir. Çalışmada net olarak belirtilmemiş ama genel olarak yazılım tanımlı ağları incelemek amaç denilebilir. Hipotez olarak bir tespit edilememiş olup problemin giderilmesi denilebilir. Çalışmada geçmiş çalışmaları ve YTA'yı incelemek yöntemi verilmiştir.
Çalışmanın başlığı (Yazılım Tanımlı Ağlar) açık ve nettir.

Çalışmanın özet kısmında problem, kapsam, amaç verilmiştir ancak çıkarım/katkı ve gök gürültülü cümle verilmemiştir.

- problem: geleneksel ağ yaklaşımı hantal ve hata eğilimlidir ve verimli şekilde kullanılmasına da olanak tanımamaktadır
- Kapsam: YTA ile ilgili yapılan çalışmalar ve son gelişmeler
- Amaç: net olarak belirtilmemiş ama genel olarak yta 'yı incelenek amaç denilebilir
- Çıkarım/Katkı: bulunamamıştır,
- Gök gürültülü cümle: bulunamamıştır.

Çalışmanın 6 anahtar kelime kullanılmıştır. Aynı ifadeler tekrarlanmış.

Çalışamın giriş bölümünde

- Problemin önemi: Her geçen gün hızla yaygınlaşan ve kullanımı artan internet geleneksel ağ yetersiz kalmaktadır olarak verilmiştir.
- Literatür tarama kısmında IOT kullannan insanların sağlı , finansal durumları gibi kritik bilgileri elde etiğindenve insanların farkındalık durumundan bahsedilmiştir.
- Çalışmanın sonucu direkt olarak belirtilmemiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez belirtilmemiştir ve hipotez geçmiş çalışmalar ve mimari anlatılarak ispatlanmaya çalışılmış. Kullanılan yöntem YTA çalışmaları incelenmiş ve YTA mimarisi veri kümesi belirtilmiştir. Daha önceki çalışmalarla ilgili geleneksel ağ ile YTA arasındaki farklardan bahsedilmiştir. Üstünlük ve kısıtlar bölümünde Birden fazla çalışmalardaki özelliklerden bahsedilmiştir. Çıktıların kritik edilmesinde YTA geleneksel ağ yapısı için geniş çözüm yelpazesi sunmasına rağmen henüz gelişmesi ve geliştirilmesi gereken yeni bir yaklaşımdır yorumu ile yapılmıştır.
Çalışmanın sonuç bölümü basit ve açıktır, tekrarlar bulunmaktadır ve geçmiş zaman kullanılmıştır, sonuç net(YTA uzun süredir devam eden bu sorunlar için bir çözüm olmuştur) olarak verilmiş ve yöntemden bahsedilmiştir.
Çalışmada 21 kaynak kullanılmış güncel bulunmakadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.

3.2 Privacy of Big Data in the Internet of Things Era

Perera ve diğerleri [10] tarafından gerçekleştirilen çalışmada problem müşteri verilerinin mahremiyeti olarak verilmiştir. Çalışma kapsamı IOT'de mahremiyet olarak belirlenmiştir. Çalışmada yenilik ve araştırma fırsatları için IOT'de mahremiyetin temel zorluklarını tartışmak amaç olarak verilmiştir. Hipotez net olarak verilmemiştir. Çalışmada 5li bir model yöntemi verilmiştir.
Çalışmanın başlığı (Privacy of Big Data in the Internet of Things Era) açık ve nettir.

Çalışmanın özet kısmında problem, kapsam, amaç verilmiştir ancak çıkarım/katkı ve gök gürültülü cümle verilmemiştir.

Şekil 12: Bilimsel Araştırma Kalitesi

- Problem: Kullanıcı bilgileri IOT cihazlarında yerel olarak yada bulutta tutulmasının kullanıcı mahremiyeti meselelerine yol açabileceği olarak gösterilmetedir.
- Kapsam: IOT'de mahremiyet.
- Amaç: Yenilik ve araştırma fırsatları için IOT'de mahremiyetin temel zorluklarını tartışmak olarak verilmiştir.
- Çıkarım/Katkı: bulunamamıştır,
- Gök gürültülü cümle: bulunamamıştır.

Çalışmanın 0(sıfır) adet anahtar kelime kullanılmıştır.

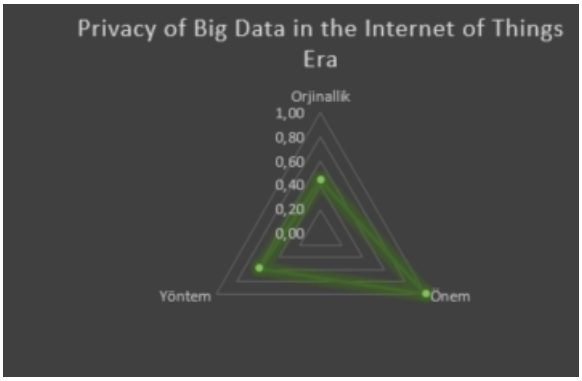
Çalışamın giriş bölümünde

- Problemin önemi: Her geçen gün hızla yaygınlaşan ve kullanımı artan internet geleneksel ağ yetersiz kalmaktadır olarak verilmiştir.
- Literatür tarama kısmında IOT kullannan insanların sağlı , finansal durumları gibi kritik bilgileri elde etiğindenve insanların farkındalık durumundan bahsedilmiştir.
- Çalışmanın sonucu direkt olarak belirtilmemiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez belirtilmemiştir ve hipotez sonuç uyumluluğuna yorum getirilememiştir. Kullanılan yöntem 5li bir model teklif ediliyor ve device manufacturers, "IoT cloud services and platform providers, third party application developers, Government and Regulatory bodies, and Individual Consumers" veri kümesi belirtilmiştir. Daha önceki çalışmalarla ilgili olarak yazar IOT uygulaması/cihazı sağlayan firmaların kullanıcı bilgilerini izin dahilinde sattığı veya koruduğu vb. durumlardan bahsetmiştir ve bir kıyaslama yapmamıştır. Üstünlük ve kısıtlar bölümünde Birden fazla çalışmalardaki özelliklerden bahsedilmemiştir. Çıktıların kritik edilmesi tespit edilmemiştir.
Çalışmanın sonuç bölümü basit ve açıktır, tekrarlar bulunmaktadır ve geçmiş zaman kullanılmıştır, sonuç verilmemiştir ve yöntemden bahsedilmemiştir.
Çalışmada 21 kaynak kullanılmış güncel bulunmakadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.

3.3 Detecting Web Attacks with end-to-end Deep Learning

Pan ve diğerleri [8] tarafından gerçekleştirilen çalışmada problem; saldırı tespit sistemlerinde el ile özellik belirlemenin zaman tüketici ve güvenlik alanında derin bilgi gerekemi gerektirmesi olarak belirlenmiştir. Web uygulama atakları kapsam olarak belirlenmiştir. Çalışmada Robust Software Modeling Tool'u ortaya koymak amaç olarak verilmiştir. Hipotez denetimsiz web atağı tespiti ve uçtan uca derin



Şekil 13: Bilimsel Araştırma Kalitesi

öğrenme olarak belirlenmiştir.. Çalışmada makine öğrenmesi ve derin öğrenme yöntemi kullanılacağı bilgisi iletilmiştir.
Çalışmanın başlığı (Detecting Web Attacks with end-to-end Deep Learning) açık ve nettir.

Çalışmanın özet kısmında problem, kapsam, amaç, çıkarım/katkı ve gök gürültülü cümle verilmiştir.

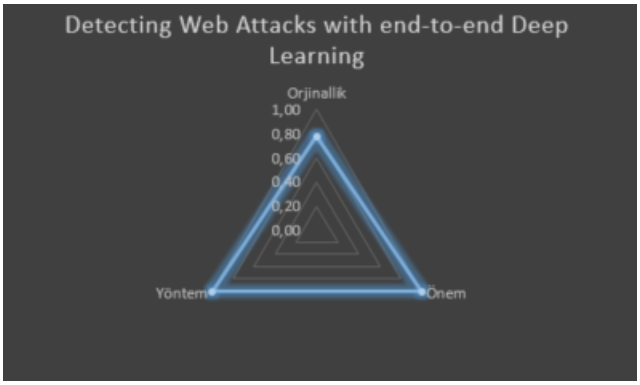
- Problem: Saldırı tespit sistemlerinde el ile özellik belirlemenin zaman tüketici ve güvenlik alanında derin bilgi gerekemi gerektirmesi.
- Kapsam: Web uygulama atakları
- Amaç: Robust Software Modeling Tool (RSMT)
- Çıkarım/Katkı: 3 katkıda bulunulmuştur. 1- denetimsiz/yarı denetimli yaklaşımın uygunluk hesabı 2- Uçtan uca derin öğrenmede çağrı grafiğini kodlamak ve yeniden yapılandırmak için RSMT'nin yığılmış bir gürültü giderici otomatik kodlayıcıyı nasıl eğittiğini tanımlamak 3- Her iki veri setinin sonuçlarını analiz etme
- Gök gürültülü cümle: Sonuçlarının önerilen yaklaşımın etkili ve doğru bir şekilde atatları tespit ettiğinin göstermesi.

Çalışmanın 3 adet anahtar kelime kullanılmıştır. WAS ve IDS eklenebilir.

Çalışamın giriş bölümünde

- Problemin önemi: Amerikkada gerçekleşen atakların firma başına 15M dolar olduğu belirtilmiş.
- Literatür tarama kısmında detayşı bir şekilde anlatılmıştır.
- Çalışmanın sonucu direkt olarak belirtilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez önerilen moderin tasarımı ile sonuçların uyumlu olduğu paylaşılmıştır. Kullanılan yöntem gerçek zamanllı uygulama çalıştırılmış ve 4M rastgele verileri içeren veri kümesi belirtilmiştir. Daha önceki çalışmalarla ilgili olarak Statik analiz, squence based Manual modelleme WEb applicatyion Firewall ve Machine learnin mekanizmaları ayrı ayrı anlatılmış kıyaslama yapılmamıştır ve sonuçlar karşılaştırılmamıştır. Üstünlük ve kısıtlar bölümünde fiyat performans ile öner çıktığı belirtilmiştir. Çıktıların kritik edilmesi için 4 farklı algoritma ile farklı atak tiplerine karşı sonuçlar nasıl paylaşılmıştır.
Çalışmanın sonuç bölümü basit ve açık değildir ve şuana kadar tecrübe edilen en uzun sonuç bölümüne sahiptir, tekrarlar bulunmaktadır ve geçmiş zaman kullanılmıştır, sonuç şüpheli bir şekilde verilmiştir ve yöntemden bahsedilmiştir.
Çalışmada 65 kaynak kullanılmış güncel bulunmakadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 14: Bilimsel Araştırma Kalitesi

3.4 Anomaly Detection in Software-Defined Networking Using Machine

Ceken ve diğerleri [14] tarafından gerçekleştirilen çalışmada problem; saldırganların SDN(YTA)'nın 3 ortamınada atak yapabileceği olarak belirlenmiştir. SDN ve makine öğrenmesi kapsam olarak belirlenmiştir. Çalışmada makine öğrenmesi ile anormal durumları tespit etmek amaç olarak verilmiştir. Hipotez anormal durumları tespit etmek olarak belirlenmiştir. Çalışmada makine öğrenmesi IDS modülü yöntem olarak belirlenmiştir.
Çalışmanın başlığı (Anomaly Detection in Software-Defined Networking Using Machine) açık ve nettir.
Çalışmanın özet kısmında problem, kapsam, amaç, çıkarım/katkı ve gök gürültülü cümle verilmiştir.

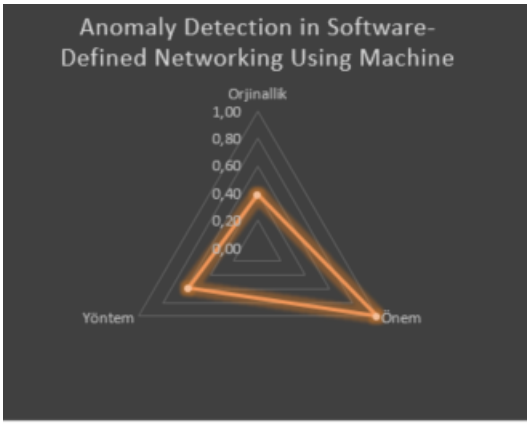
- Problem: Saldırganların SDN(YTA)'nın 3 ortamınada atak yapabilmesi.
- Kapsam: Makine öğrenmesi SDN
- Amaç: makine öğrenmesi ile anormal durumları tespit etmek
- Çıkarım/Katkı: Karar ağacı algoritmasının başarılı bir şekilde ataklara karşı çalıştırılabilmesi.
- Gök gürültülü cümle: SOnuçların karar ağacı algoritmasının DDOS ataklarına karşı başarılı birşekilde karşı çalışması.

Çalışmanın 3 adet anahtar kelime kullanılmıştır. Machine Learning ve SDN eklenebilir.

Çalışamın giriş bölümünde

- Problemin önemi: SDN'nin 3 modülünde ataklara karşı açık olduğu ve control plane'nin bu yapının beyni olduğu için tek noktada hatanın bullunduğundan bahsedilmiş.
- Literatür tarama kısmında çalışmalar anlatılmıştır.
- Çalışmanın sonucu direkt olarak belirtilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez önerilen hipotezde saldırıların tespiti var ilen burda sadece ping flooding saldırıların tespit edildiği sonucuna varılmıştır, hipotez tez uyumluluğu bu yüzden bulunmamaktadır. Kullanılan yöntem karar ağıcı mekanizması olup veri seti olarak gercek zamanlı ping paketleri kullanışmıştır. Daha önceki çalışmalarla ilgili sonuçlar karşılaştırılmamıştır. Üstünlük ve kısıtlar bölümüne rastlanılmamıtır. Sonuç karşılaştırması ve Çıktıların kritik edilmesi bölümü bulunmamaktadır.
Çalışmanın sonuç bölümü basit ve açık değildir, tekrarlar bulunmaktadır ve geçmiş zaman kullanılmamıştır, sonuç verilmiştir ve yöntemden bahsedilmiştir.
Çalışmada 16 kaynak kullanılmış güncel bulunmakadır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 15: Bilimsel Araştırma Kalitesi

3.5 Malicious URL filtering - A Big Data Aplication

Lin ve diğerleri [7] tarafından gerçekleştirilen çalışmada problem; zararlı web sitelerin kısa yaşam süresine sahip olması ve adresleirini sürekli olarak değiştirmesi olarak belirlenmiştir. Malicious URLs kapsam olarak belirlenmiştir. Çalışmada URL dizisine bağlı olarak yeni bir filtreleme sunmak amaç olarak verilmiştir. Hipotez 2 milyon URL'i 5 dakikadan az süre içerisinde zararlı olup olmadığını tespitini yapabilmek olarak belirlenmiştir. Çalışmada sözcüksel ve tanımlayıcı özellikler ve bunların yakıştırılması ile filtreleme modeli yöntemi verilmiştir.
Çalışmanın başlığı (Malicious URL filtering - A Big Data Aplication) açık ve nettir. Bir ifadesi kullanılması olumsuz olarak görülmüştür.

Çalışmanın özet kısmında problem, kapsam, amaç, çıkarım/katkı ve gök gürültülü cümle verilmiştir.

- Problem: zararlı web sitelerin kısa yaşam süresine sahip olması ve adresleirini sürekli olarak değiştirmesi
- Kapsam: Malicious URLs
- Amaç: makine öğrenmesi ile anormal durumları tespit etmek
- Çıkarım/Katkı: %75 iş kazancı, URL sorgusunu azaltmakta, hesaplama hızını azaltma olarak belirtilmiş.
- Gök gürültülü cümle: 2 milyon URL'i 5 dakikadan az süre içerisinde zararlı olup olmadığını tespitini yapabilmek ve %75 iş gücünden kazanç ve en az %90 başarım oranı verilmesidir.

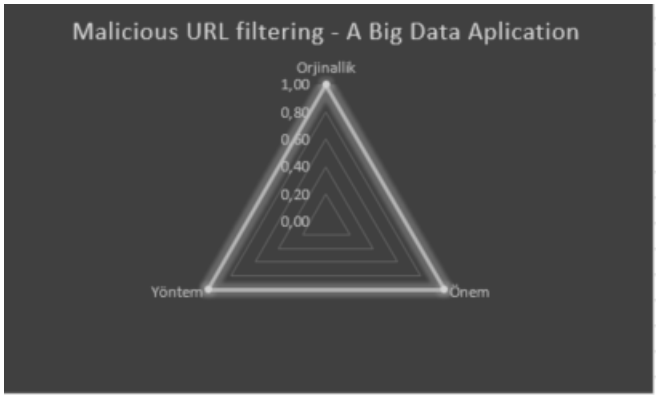
Çalışmanın 4 adet anahtar kelime kullanılmıştır. URL filter eklenebilirdi.

Çalışamın giriş bölümünde

- Problemin önemi: Zararlı URL ataklarının artışı ve bu zararlı URL'lerin yaşam ömrünün kısa olması ve tespitinin zorluğu.
- Literatür tarama kısmında çalışmalar başarılı bir şekilde anlatılmıştır.
- Çalışmanın sonucu direkt olarak belirtilmiştir.
- Tarihsel anlatım olmaması akıcılığı artırmıştır.

Çalışma metodolojisinde hipotez önerilen modele göre başarılı sonuca varılmıştır, hipotez tez uyumluluğu bulunmaktadır. Kullanılan yöntem öznitelik çıkartma birden fazla alogritma kullanmak olup veri seti "web reputation" servisleri kullanışmıştır. Daha önceki çalışmalarla ilgili sonuçlar karşılaştırılmamıştır. Üstünlük ve kısıtlar bölümüne rastlanılmamıtır. Sonuç karşılaştırması bulunmamaktadır.Çıktıların kritik edilmesi bölümü bulunmaktadır.
Çalışmanın sonuç bölümü basit ve açıktır, tekrarlar bulunmamaktadır ve geçmiş zaman kullanılmamıştır,

sonuç verilmiştir ve yöntemden bahsedilmiştir.
Çalışmada 21 kaynak kullanılmış güncel bulunmakadır bunların 8'i teknoloji firmalarının veya teknolojilerin web sayfasıdır. Çalışmanın bilimsel araştırma kalitesi aşağıdaki gibi resimlendirilmiştir.



Şekil 16: Bilimsel Araştırma Kalitesi

Sonuçlar Gerçekleştirilen bu çalışmada en fazla puanı aşağıdaki çalışmalar almıştır:

- S-DDoS: Apache spark based real-time DDoS detection system
- Big data and data science: what should we teach?
- Malicious URL filtering - A Big Data Aplication

en düşük puanı aşağıdaki çalışmalar almıştır

- Büyük verilere yetkisiz erişimlerin tespit edilmesi ve engellenmesi tekniklerinin incelenmesi ve uygulaması
- Anomaly Detection in Software-Defined Networking Using Machine
- Privacy of Big Data in the Internet of Things Era

$$Basarılıcalısmalar \geq Orjinalik * Önem * Yöntem = 0.8 * 0.8 * 0.8 = 0.512$$

kritirine göre değerlerdirilmiş olup aşağıda verilmiştir.

- S-DDoS: Apache spark based real-time DDoS detection system
- Big data and data science: what should we teach?
- Malicious URL filtering - A Big Data Aplication
- Yaşanan Terör Olaylarını İçeren Büyük Verinin Makine Öğrenmesi Teknikleri ile Analizi
- Şifreli Ağ Trafiğinin İçerik Açısından Sınıflandırılması
- Yazılım Tanımlı Ağlar – YTA
- Detecting Web Attacks with end-to-end Deep Learning
- Big Data for Remote Sensing: Challenges and Opportunities
- Language Based Web Crawlıng on Big Data

Kaynaklar

[1] Ramazan BOZKIR. Şİfrelİ aĞ trafİĞİnİn İÇerİk aÇisindan siniflandirilmasi. *https://tez.yok.gov.tr/UlusalTezMerkezi/*, 1, 1 2019.

[2] Mingmin Chi, Antonio Plaza, Jon Atli Benediktsson, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104:2207–2219, 11 2016.

[3] Murtaza Cicioğlu and Ali Çalhan. Şİfrelİ aĞ trafİĞİnİn İÇerİk aÇisindan siniflandirilmasi. *http://fbd.beun.edu.tr*, 1, 1 2017.

[4] Canari Girgin, Hayati Gonultac, F. Canan Pembe Muhtaroglu, Seniz Demir, Ahmet A. Akin, and Murat Obali. Büyük veride dil odakli içerik çekme. pages 1528–1531. IEEE Computer Society, 2014.

[5] Barış KARABAY. YaŞanan terÖr olaylarini İÇeren bÜyÜk verİnİn makİne ÖĞrenmesİ teknİklerİ İle analİzİ. *https://tez.yok.gov.tr/UlusalTezMerkezi/*, 1, 1 2019.

[6] Elif Merve KURT. Apache spark ve makİne ÖĞrenmesİ algorİtmalari İle aĞ saldirisi tespİtİ. *https://tez.yok.gov.tr/UlusalTezMerkezi/*, 1, 1 2019.

[7] Min Sheng Lin, Chien Yi Chiu, Yuh Jye Lee, and Hsing Kuo Pao. Malicious url filtering - a big data application. pages 589–596. IEEE Computer Society, 2013.

[8] Yao Pan, Fangzhou Sun, Zhongwei Teng, Jules White, Douglas C. Schmidt, Jacob Staples, and Lee Krause. Detecting web attacks with end-to-end deep learning. *Journal of Internet Services and Applications*, 10, 12 2019.

[9] Nilesh Vishwasrao Patil, C. Rama Krishna, and Krishan Kumar. S-ddos: Apache spark based real-time ddos detection system. volume 38, pages 6527–6535. IOS Press, 2020.

[10] Charith Perera, Rajiv Ranjan, Lizhe Wang, Samee U. Khan, and Albert Y. Zomaya. Big data privacy in the internet of things era. *IT Professional*, 17:32–39, 5 2015.

[11] Il Yeol Song and Yongjun Zhu. Big data and data science: what should we teach? *Expert Systems*, 33:364–373, 8 2016.

[12] José L. Torrecilla and Juan Romo. Data learning from big data. *Statistics and Probability Letters*, 136:15–19, 5 2018.

[13] Feridun TOY. Büyük verilere yetkisiz erişimlerin tespit edilmesi ve engellenmesi tekniklerinin incelenmesi ve uygulaması. *https://tez.yok.gov.tr/UlusalTezMerkezi/*, 1, 1 2019.

[14] Celal ÇEKEN and Soumaine BOUBA MAHAMAT. Anomaly detection in software-defined networking using machine. *Düzce University Journal of Science & Technology*, 1, 1 2019.

[15] MERT İNANIR. Machine learning algorithms implementation and evaluation on apache spark pyspark. *https://tez.yok.gov.tr/UlusalTezMerkezi/*, 1, 1 2021.