

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ
AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

Yusifova Ülkər Elçin qızı
Dadaşova Leyla Tərlan qızı
Məmmədzadə Nərminə Fizuli qızı
Axundzadə Fuad Babək oğlu

**DATA MİNING YOLU İLƏ SOSIAL ŞƏBƏKƏLƏRDƏ SAXTA
HESABLARLIN AŞKARLANMASI**

mövzusunda
MAGİSTRİK DİSSERTASİYASI

İxtisas: 060632 - “İnformasiya tenologiyaları və sistemləri mühəndisliyi”

İxtisaslaşma: “İdarəetmədə informasiya sistemləri”

Elmi rəhbər:

t.e.n., dosent Z.Ə. Cəfərov

BAKİ – 2023

MÜNDƏRICAT

GİRİŞ	4
I FƏSİL. DATA MİNING NƏDİR?	7
1.1.Data mining haqqında ümumi məlumat.....	7
1.2.Data Miningin sosial mediada rolu	11
II FƏSİL. QRAFİK ANALİZ VƏ TƏSNİFAT ALQORİTMLƏRİ ƏSASINDA SAXTA HESABLARIN MÜƏYYƏN EDİLMƏSİ	16
2.1 Qrafik təhlili və oxşarlıq növləri.	16
2.2.Yenidən nümunə götürmə.Əsas komponent təhlili və Maşın Öyrənilməsinin tətbiqi.....	18
2.3.Təklif olunan metod.Qiymətləndirmə.Nəticəsi ...	21
III FƏSİL. SOSIAL MEDIADA SAXTA HESABLARINAŞKARLANMASI	32
3.1.Saxta hesab aşkarlanması üçün giriş parametrləri	32
3.2.METODLAR VƏ MATERİALLAR	34
3.3. Metrik dəyərlərinə görə nöqtələrin təyin edilməsi meyarları	41
IV FƏSİL. SOSIAL MEDIADA SAXTA HESABLARIN AŞKARLANMASI.....	48
METODLARI.....	48
4.1.Əsas komponentlərin analizi	48
4.3.Fəaliyyət və Qiymətləndirmə.....	59
NƏTİCƏ	67

İxtisarlarm siyahısı

OSN-Online Social Media(Online Sosial Şəbəkə)

SMOTE-Synthetic Minority Oversampling Technique(Sintetik Azlığın cox Nümunə Alma Texnikası)

PCA-Principal Component Analysis (Əsas Komponent Təhlili)

SVM -Support Vector Machine (Dəstək vektor maşını)

TP- True Positive (Həqiqi Müsbət)

FP-False Positive (Yalan Müsbət)

TN-True Negative (Həqiqi Mənfi)

FN- False Negative (Yalan Mənfi)

SM-Similarity Matrix (Oxşarlıq Matriksi)

AUC-Area Under Curve (Əyri Altındakı Sahədən)

TNR- True Negative Rate (Həqiqi Mənfi Nisbət)

FPR- False Positive Rate (Yanlış Müsbət Nisbət)

TPR- True Positive Rate (Həqiqi Müsbət Nisbət)

FNR-False Negative Rate (Yanlış Mənfi Nisbət)

FG-Friendship Graph(Dostluq Qrafiki)

CF- Common Friends (Ortaq Dostlar)

TF- Total Friends (Ümumi Dostlar)

JS- Jaccard Similarity(Jaccard Oxsarlıg)

IP -Internet Protocol(İnternet Protokolu)

GİRİŞ

Sosial media platformaları öz platformalarında saxta hesabların müəyyən edilməsində və onlarla mübarizədə mühüm rol oynayır. İstifadəçiləri üçün təhlükəsiz və etibarlı mühit təmin edərək, bu hesabları aşkar etmək və silmək üçün müxtəlif strategiya və üsullardan istifadə edirlər. Sosial media platformalarının saxta hesabları necə müəyyənləşdirdiyinə dair qısa məlumatı təqdim edirik:

Süni İntellekt və Maşın Öyrənmə: Sosial media platformaları böyük miqdarda istifadəçi məlumatlarını təhlil etmək və saxta hesabları göstərən nümunələri müəyyən etmək üçün süni intellekt (AI) və maşın öyrənmə alqoritmlərindən istifadə edir. Bu alqoritmlər geniş verilənlər bazasından öyrənir və şübhəli davranışı, məzmunu və ya hesab xüsusiyyətlərini aşkar edə bilir.

İstifadəçi Davranışının Təhlili: Platformalar saxta hesabları göstərə biləcək anomaliyaları müəyyən etmək üçün istifadəçi davranış nümunələrini izləyir. Onlar məzmun yaratma sürəti, nişanlanma tezliyi və ya digər hesablarla qeyri-adi qarşılıqlı əlaqə nümunələri kimi fəaliyyətləri izləyirlər. Fəaliyyətdə və ya şübhəli nişanlanma təcrübələrində qəfil artımlar hesabı nəzərdən keçirmək üçün qeyd edə bilər.

Hesabın Doğrulanması: Bəzi sosial media platformaları istifadəçilərin şəxsiyyətini təsdiqləmək üçün hesab yoxlama mexanizmlərini tətbiq edir. Doğrulama nişanları və ya işarələri ictimai xadimlərə, məşhurlara və ya yüksək profilli hesablara onların həqiqiliyini göstərən təyin edilir. Bu, istifadəçilərə həqiqi hesabları potensial saxta hesablardan ayırmağa kömək edir.

Məzmun və Link Analizi: Sosial media platformaları yazılar, şərhələr və keçidlər daxil olmaqla hesablar tərəfindən paylaşılan məzmunu təhlil etmək üçün alqoritmlərdən istifadə edir. Onlar spam və ya zərərli məzmunu, açar sözlərdən və ya hashtaglardan həddindən artıq istifadəni və ya məlum zərərli veb-saytlara keçidləri yoxlayırlar. Belə təhlil spamla məşğul olan və ya dezinformasiya yayan hesabları müəyyən etməyə kömək edir.

Şəbəkə təhlili: Platformalar hesablar arasındakı əlaqələri yoxlayır, əlaqələndirilmiş və ya avtomatlaşdırılmış davranışı göstərən nümunələri axtarır.

Hesablar şəbəkəsi məzmunun çarpaz təşviqi və ya süni gücləndirilməsi kimi şübhəli nümunələr nümayiş etdirirsə, əlavə araşdırma üçün bayraqlar qaldırır.

Hesabat Mexanizmləri: Sosial media platformaları istifadəçiləri saxta olduğundan və ya zərərli fəaliyyətlə məşğul olduqlarından şübhələndikləri hesabları bildirməyə təşviq edir. Bu hesabatlar platforma administratorlarına araşdırma aparmaq və müvafiq tədbirlər görmək üçün dəyərli məlumat verir.

Əməkdaşlıq və Məlumat Paylaşımı: Sosial media platformaları tez-tez saxta hesablar tərəfindən istifadə olunan yaranan təhdidlər və üsullardan xəbərdar olmaq üçün xarici təşkilatlar, təhlükəsizlik mütəxəssisləri və tədqiqatçılarla əməkdaşlıq edir. Onlar aşkarlama imkanlarını kollektiv şəkildə artırmaq üçün məlumat və ən yaxşı təcrübələri paylaşır.

Bu üsulların kombinasiyasından istifadə etməklə sosial media platformaları saxta hesabları tez bir zamanda müəyyən etməyə və silməyə çalışır. Müntəzəm monitoring, aktiv aşkarlama tədbirləri və istifadəçi iştirakı daha təhlükəsiz və etibarlı sosial media təcrübəsi yaratmağa kömək edir.

Qeyd etmək vacibdir ki, bu strategiyaların effektivliyi fərqli ola bilər və zaman keçdikcə saxta hesablar yaratmaq üçün yeni üsullar ortaya çıxa bilər. Beləliklə, sosial media platformaları pis niyyətli aktyorlar tərəfindən istifadə olunan inkişaf edən taktikalardan qabaq qalmaq üçün öz aşkarlama üsullarını davamlı olaraq təkmilləşdirir.

Sosial media son dövrlərdə ilk növbədə bir ünsiyyət, fikir mübadiləsi və məlumat əldə etmək vasitə olaraq geniş yayılıb, insanların sosial həyatı bu saytlarla daha çox əlaqələndirilir. Fərdlər və təşkilatlar sosial şəbəkələrdən öz fikirlərini ifadə etmək, məhsullarını reklam etmək, şirkət və təşkilatlarının gələcək siyasətlərini ifadə etmək üçün sosial mediadan istifadə edirlər

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ
AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

Yusifova Ülkər Elçin qızı

**DATA MİNING YOLU İLƏ SOSIAL ŞƏBƏKƏLƏRDƏ SAXTA
HESABLARLIN AŞKARLANMASI**

**mövzusunda
MAGİSTRİK DİSSERTASİYASI**

İxtisas: 060632 - "İnformasiya tenologiyaları və sistemləri mühəndisliyi"

İxtisaslaşma: "İdarəetmədə informasiya sistemləri"

Elmi rəhbər:

t.e.n., dosent Z.Ə. Cəfərov

BAKİ – 2023

I FƏSİL. DATA MİNING NƏDİR?

1.1.Data mining haqqında ümumi məlumat

Həm biznes, həm də akademiklər üçün məlumatların çıxarılması üçün bir neçə tərəf var. Data mining sadə təhlillə mümkün olmayan davranışları, nümunələri və meylləri aşkar etmək üçün avtomatik olaraq böyük həcmdə məlumat axtaracaq bir texnologiyadır. Data Mining bizneslərə proaktiv biliyə əsaslanan qərarlar qəbul etməyə imkan verir ki, bu da rəqiblərini qabaqlamaq üçün kifayətdir. Data Mining böyük verilənlər toplusundan məlumat çıxarmaq üçün istifadə edilir. Başqa sözlə, deyə bilərik ki, data mining verilənlərdən biliyin çıxarılması prosedurudur. Çıxarılan məlumat və ya biliklər aşağıdakılardan hər hansısa biri üçün istifadə edilə bilər:

- Bazar təhlili
- Fırıldaqçılığın aşkarlanması
- Müştərinin saxlanması
- İstehsal nəzarəti
- Elm Kəşfiyyatı

Bazar Təhlili və İdarəetmə:

Aşağıda verilənlərin istifadə olunduğu bazarın təhlilinin müxtəlif sahələri verilmişdir:

Müştəri profilinin yaradılması – Müştəri profilinin yaradılması məlumatların öyrənilməsinə insanların hansı növ məhsulları aldığını müəyyən etməyə kömək edir.

Müştəri Tələblərinin Müəyyənləşdirilməsi – Data mining müxtəlif müştərilər üçün ən yaxşı məhsulların müəyyən edilməsinə kömək edir. Yeni müştəriləri cəlb edə biləcək amilləri tapmaq üçün proqnozdan istifadə edir.

Çarpaz Bazar Təhlili – Data mining məhsul satışı arasında assosiasiya/korrelyasiya

həyata keçirir. Hədəf Marketing - Məlumatların öyrənilməsi maraqlar, xərcləmə vərdişləri, gəlir və s. kimi eyni xüsusiyyətləri paylaşan model müştərilərin qruplarını tapmağa kömək edir.

Müştərinin satınalma modelinin müəyyən edilməsi – Data mining müştərinin satınalma modelini müəyyən etməyə kömək edir.

Xülasə Məlumatın Təmin edilməsi – Data mining bizə müxtəlif çoxölçülü xülasə hesabatları təqdim edir.

Korporativ Təhlil və Risklərin İdarə Edilməsi

Data mining Korporativ Sektorun aşağıdakı sahələrində istifadə olunur :

Maliyyə Planlaşdırılması və Aktivlərin Qiymətləndirilməsi - Bu, aktivləri qiymətləndirmək üçün pul vəsaitlərinin hərəkətinin təhlili və proqnozlaşdırılmasını, şərti iddia təhlilini əhatə edir.

Resursların Planlaşdırılması - Bu, resursların və xərclərin ümumiləşdirilməsini və müqayisəsini əhatə edir.

Rəqabət - Bu, rəqiblərin və bazar istiqamətlərinin monitorinqini əhatə edir.

Fırıldaqlığın aşkarlanması

Data mining, fıırıldaqları aşkar etmək üçün kredit kartı xidmətləri və telekommunikasiya sahələrində də istifadə olunur. Fırııldaq telefon zənglərində o, zəngin təyinat yerini, zəngin müddətini, günün və ya həftənin vaxtını və s. tapmağa kömək edir. O, həmçinin gözlənilən normalardan kənara çıxan nümunələri təhlil edir. Əldə olunacaq məlumat növü əsasında Data Mining-də iştirak edən funksiyaların iki kateqoriyası mövcuddur:

- Təsvir funksiyası
- Sınıf/konsepsiya təsviri

Sınıf/Konsept siniflər və ya anlayışlarla əlaqələndiriləcək verilənlərə aiddir. Məsələn, bir şirkətdə satış üçün əşyaların siniflərinə kompüter və printerlər, müştərilər anlayışlarına isə böyük xərcləyənlər və büdcə xərcləyənlər daxildir. Bir sinfin və ya

konsepsiyanın bu cür təsvirləri sinif/konsept təsvirləri adlanır. Bu təsvirlər aşağıdakı iki yolla əldə edilə bilər.

Məlumatların xarakteristikası - Bu, öyrənilən sinifin məlumatlarının ümumiləşdirilməsinə aiddir. Tədqiq olunan bu sinif Hədəf Sinfi adlanır.

Məlumat ayrı-seçkiliyi - Bu, əvvəlcədən müəyyən edilmiş bəzi qrup və ya siniflə bir sinfin xəritələşdirilməsinə və ya təsnifatına aiddir.

Tez-tez Nümunələrlə data Mining

Tez-tez nümunələr əməliyyat məlumatlarında tez-tez baş verən nümunələrdir.

Tez-tez rast gəlinən nümunələrin siyahısı aşağıdakı kimidir.

Tez-tez Əşyalar çoxluğu - Bu, süd və çörək kimi tez-tez birlikdə görünən birsıra maddələrə aiddir. Yəni istifadəçi birini aldıqda digərini də alacaq deyə yan- yana qoymaqla təqdim olunur.

Tez-tez Ardıcılıq - Kameranın alınması kimi tez-tez baş verən nümunələr ardıcılığı yaddaş kartı ilə izlənilir.

Tez-tez Alt Struktur - Alt struktur element çoxluqları və ya alt ardıcılıqla birləşdirilə bilən qrafiklər, ağaclar və ya qəfəslər kimi müxtəlif struktur formalarına aiddir.

Assosiasiyanın Miningi

Birliklər pərakəndə satışda tez-tez birlikdə alınan nümunələri müəyyən etməyə üçün istifadə olunur. Bu proses verilənlər arasında əlaqənin aşkarlanması və assosiasiya qaydalarının müəyyən edilməsi prosesinə aiddir.

Məsələn, pərakəndə satıcı, südün 70% -nin çörəklə və yalnız 30% -də peçenyenin çörəklə satıldığını göstərən bir assosiasiya qaydası yaradır.

Korrelyasiya Miningi

Bu, bir-birinə müsbət, mənfi və ya heç bir təsir göstərmədiyini təhlil etmək üçün əlaqəli atribut-dəyər cütləri və ya iki element çoxluğu arasında maraqlı statistik korrelyasiyaları aşkar etmək üçün həyata keçirilən bir növ əlavə təhlildir.

Klasterlərin Miningi

Klaster oxşar növ obyektlər qrupuna aiddir. Klaster təhlili bir-birinə çox oxşar olan,

lakin digər çoxluqlardakı obyektlərdən çox fərqli olan obyektlər qrupunun formalaşmasına aiddir.

Təsnifat və proqnozlaşdırma

Təsnifat məlumat siniflərini və ya anlayışlarını təsvir edən bir model tapmaq prosesidir. Məqsəd sinif etiketi bilinməyən obyektlərin sinifini proqnozlaşdırmaq üçün bu modeldən istifadə etməkdir. Bu əldə edilmiş model təlim məlumat çoxluqlarının təhlilinə əsaslanır. Əldə edilmiş model aşağıdakı formalarda təqdim edilə bilər.

- Təsnifat qaydaları(Əgər-Onda prinsipi)
- Qərar Ağacları
- Riyazi düsturlar
- Neyron şəbəkələri

Bu proseslərdə iştirak edən funksiyaların siyahısı aşağıdakı kimidir.

Təsnifat - Sinif etiketi bilinməyən obyektlərin sinfini proqnozlaşdırır. Onun məqsədi məlumat siniflərini və ya konsepsiyalarını təsvir edən və fərqləndirən törəmə modeli tapmaqdır. Törəmə Model təlim məlumatlarının təhlili toplusuna, yəni sinif etiketi yaxşı məlum olan məlumat obyektinə əsaslanır.

Proqnoz - Sinif etiketlərindən daha çox çatışmayan və ya əlçatmaz ədədi məlumat dəyərlərini proqnozlaşdırmaq üçün istifadə olunur. Reqrəssiya təhlili ümumiyyətlə proqnozlaşdırmaq üçün istifadə olunur. Mövcud məlumatlar əsasında paylanma tendensiyalarını müəyyən etmək üçün də proqnozlaşdırmadan istifadə edilə bilər.

Outlier Analizi - Outliers mövcud məlumatların ümumi davranışına və ya modelinə uyğun gəlməyən məlumat obyektləri kimi müəyyən edilə bilər.

Təkamül təhlili – Təkamül təhlili davranışı zamanla dəyişən obyektlərin təsviri və model qanunauyğunluqlarına və ya meyllərinə istinad edir.

1.2.Data Miningin sosial mediada rolu

Sosial media son dövrlərdə ilk növbədə bir ünsiyyət, fikir mübadiləsi və məlumat əldə etmək vasitə olaraq geniş yayılıb, insanların sosial həyatı bu saytlarladaha çox əlaqələndirilir. Fərdlər və təşkilatlar sosial şəbəkələrdən öz fikirlərini ifadə etmək, məhsullarını reklam etmək, şirkət və təşkilatlarının gələcək siyasətlərini ifadə etmək üçün sosial mediadan istifadə edirlər. OSN-lərin sürətlə artması və onun abunəçilərinin şəxsi məlumatlarının çoxluğu təcavüzkarları və saxtakarları şəxsi məlumatları oğurlamağa, yalan xəbərləri paylaşmağa və zərərli fəaliyyətləri yaymağa cəlb etdi, sosial media saxta hesablarından tez-tez müxtəlif kiberhücumlar, informasiya-psixoloji əməliyyatlar və müharibə zamanı sosial rəyin manipulyasiyası üçün istifadə olunmağa başlandı. Digər tərəfdən tədqiqatçılar hesabların xüsusiyyətlərinə və təsnifat alqoritmlərinə əsaslanaraq anormal fəaliyyətləri və saxta hesabları aşkar etmək üçün effektiv üsulları araşdırmağa başlayıblar.

Facebook, Twitter, RenRen, LinkedIn, Google+ və Tuenti kimi Onlayn Sosial Şəbəkələr (OSN-lər) son bir neçə ildə getdikcə populyarlaşır. İnsanlar bir-biri ilə əlaqə saxlamaq, xəbərləri bölüşmək, tədbirlər təşkil etmək və hətta öz elektron bizneslərini idarə etmək üçün bu tip OSN-lərdən istifadə edirlər.

2003-cü ildə Mark Zukerberq yeni konsepsiya üzərində iş başladı və nəticədə, Facebook kimi tanınan global sosial şəbəkəyə çevrildi. O vaxtdan bəri Facebook bütün dünyada genişlənərək 2018-ci ilin dekabrına olan məlumata görə aylıq 2,3 milyardan çox aktiv istifadəçiyə çatdı. Bunun kimi bir vasitə insanların bir-biri ilə qarşılıqlı əlaqəsini dəyişdirdi. Facebook ünsiyyət sərhədlərini aradan qaldırır ki, insanlar yüksək etibarlılıq və əlçatanlıqla həyat hadisələrini, hekayələrini və ya sosial fəaliyyətlərini paylaşsın və başqaları ilə asanlıqla əlaqə qura bilsinlər. Bu, çoxlu sayda qeydiyyatdan keçmiş istifadəçilərin bu şəbəkəyə abunə olması ilə nəticələndi. Amerika Pediatriya Akademiyasının məlumatına görə, Amerikada yenyetmələrin 84%-nin Facebook hesabı var və son rəsmi elanlara görə dünya üzrə cəmi 2,2 milyard istifadəçi var. Facebook üstünlük təşkil edən ünsiyyətdir. Bu qədər insan üçün platformada istifadəçilərin məxfiliyi fırılacaqçıların hədəfinə çevrilə bilər, məsələn,

dəyərli məlumatları oğurlamaq üçün qurbanı təqlid etmək üçün yalan məlumatlardan istifadə edərək saxta profillər yaratmaq və ya maliyyə üçün istifadəçinin kontaktlarından istifadə etməklə 2014-2018-ci illər arasında qeyri-kommersiya təşkilatları tərəfindən Facebook-da siyasi reklamların sponsorluğuna təxminən 2,53 milyon ABŞ dolları xərclənib. OSN-lərin açıq xarakteri və onun abunəçiləri üçün çoxlu sayda şəxsi məlumat onları Sybil hücumlarına qarşı həssas etmişdir.

2012-ci ildə Facebook öz platformasında yalan xəbərlərin dərc edilməsi, nifrət nitqi, sensasiyalı və qütbləşmə və digərləri daxil olmaqla sui-istifadə hallarını müşahidə etdi. Bununla birlikdə, onlayn Sosial Şəbəkə işləri (OSN) tədqiqatçıların böyük miqdarda məlumatların çıxarılması və təhlili üçün də marağını cəlb etdi. Tədqiqatçılar müştərilərin münasibətini proqnozlaşdıran ən təsirli koqnitiv xüsusiyyətləri müəyyən edərək, sosial mediaya əsaslanan onlayn brend cəmiyyətinə sadıqlığını proqnozlaşdırmaq, təhlil etmək və izah etmək üçün araşdırma aparıblar

Facebook icması illik 11% artımla 2,2 milyard aylıq aktiv istifadəçi və 1,4 milyard Gündəlik aktiv istifadəçi ilə böyüməyə davam edir . Təkcə 2018- ci ilin ikinci rübündə Facebook cəmi gəlirinin yalnız reklamlardan 13,0 milyard dollar olmaqla 13,2 milyard dollar olduğunu bildirdi.

Eynilə, 2018-ci ilin ikinci rübündə Twitter aylıq 335 milyon aktiv istifadəçi ilə təxminən bir milyard Twitter abunəçisinə çatdığını bildirdi . 2017-ci ildə twitter 2,44 milyard ABŞ dolları sabit gəlir artımı göstərmiş və əvvəlki illə müqayisədə 108 milyon ABŞ dolları az mənfəət əldə edib .

2015-ci ildə Facebook hesablamışdır ki, aylıq aktiv istifadəçilərinin 14 milyona yaxını əslində arzuolunmazdır, bu veb saytların xidmət şərtlərini pozaraq yaradılmış zərərli saxta hesabları təmsil edir .

Facebook, ilk dəfə 2018-ci ilin birinci rübündə 2017-ci ilin oktyabrından 2018-ci ilin martına qədər söylərini əhatə edən icma standartlarını tətbiq etmək üçün istifadə etdikləri daxili təlimatları göstərən bir hesabat paylaşdı, bu hesabat Facebook tərəfindən silinmiş arzuolunmaz məzmunun miqdarını göstərir və bir neçə kateqoriyanı əhatə edir: qrafik zorakılıq, terror təbliğatı, nifrət nitqi, spam və saxta hesablar və s .

Tərkibində spam olan 837 milyon yazı silinib, 583 milyona yaxın saxta hesab ləğv edilib, Facebook həmçinin qalan məzmun növlərini pozan 81 milyon arzuolunmaz məzmunu silib.

Bununla belə, Facebook-dan milyonlarla saxta hesabın qarşısını aldıqdan sonra belə, təxminən 88 milyon hesabın hələ də saxta olduğu təxmin edilir.

Statistikalar göstərir ki, ABŞ-da yetişkinlərin 40%-i və yeniyetmələrin 18%-i sosial mediada məhsul satmaq üçün saxta hesablardan və botlardan istifadə etməkdən çox narahatdırlar. Başqa bir misal, 2012-ci il ABŞ seçki kampaniyası zamanı rəqibi Romninin Twitter hesabında izləyicilərin sayında qəfil artım yaşandı. Onların böyük əksəriyyətinin sonradan saxta xalq olduğu iddia edildi .

2015-ci ilin dekabrında New Yorker-in müxbiri Adrian Chen qeyd etdi ki, o, izlədiyi bir çox rus hesablarının Trampı dəstəkləyən səylərə keçdiyini görüb, lakin onların əksəriyyəti idarə olunan trol hesablar olub. Eynilə, 2013-cü ilin fevralında keçirilən ümumi İtaliya seçkilərindən əvvəl onlayn bloqlar və qəzetlər əsas namizədlərin saxta izləyicilərinin ehtimal olunan faizi ilə bağlı statistik məlumat verdilər.

OSN-lərdə həmin təhdidedici hesabların aşkarlanması müxtəlif zərərli fəaliyyətlərdən qaçmaq, istifadəçi hesablarının təhlükəsizliyini təmin etmək və şəxsi məlumatı qorumaq üçün zəruri hala çevrilib.

Tədqiqatçılar saxta hesabları müəyyən etmək üçün avtomatlaşdırılmış aşkarlama alətləri hazırlamağa çalışırlar ki, bu da əl ilə aparıldığına görə çox əmək tələb edən və baha başa gəlir. Tədqiqatçıların cəhdlərinin nəticələri OSN operatoruna saxta hesabları səmərəli və effektiv şəkildə aşkar etməyə imkan verə bilər, bu, zəhlətökən spam mesajlarının və digər təhqiramiz məzmunun qarşısını almaqla istifadəçilərinin təcrübəsini artırır. OSN operatoru həmçinin öz istifadəçi ölçülərinin etibarlılığını artırır və üçüncü tərəflərə onun istifadəçi hesablarını nəzərdən keçirməyə imkan verə bilər.

Tədqiqatçılar son istifadəçilərdən funksiyalar çıxararaq, məsələn, yazıların

sayı, izləyicilərin sayı, profillər kimi istifadəçi səviyyəli fəaliyyətini təhlil edərək saxta hesabların müəyyən edilməsinə diqqət yetirirlər.

Onlar real/saxta hesabların təsnifatı üçün təlim keçmiş maşın öyrənmə texnikasını tətbiq edirlər. Başqa bir yanaşma, OSN-nin mahiyyətə qovşaqların(node) və kənarların(edge) toplusu kimi təqdim olunan qrafik kimimodelləşdirildiyi qrafik səviyyəli strukturdan istifadə etməkdir.

Hazırda informasiya-kommunikasiya texnologiyalarının davamlı təkmilləşməsi ilə istənilən münaqişə internetdə öz əksini tapır. Çox vaxt belə bir əks olunma rəqib tərəflərin qarşıdurmasının nəticələrinə təsir göstərir. Çoxmilyonlu auditoriyanın fəal iştirakı ictimai rəyi manipulyasiya etməyə və proseslərə əhəmiyyətli dərəcədə təsir göstərməyə imkan verir. Üstəlik, sosial media sosial mühəndislik kimi müxtəlif hücumlar üçün bir sahəyə çevrilir.

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ
AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

Dadaşova Leyla Tərlan qızı

**DATA MİNING YOLU İLƏ SOSİAL ŞƏBƏKƏLƏRDƏ SAXTA
HESABLARLIN AŞKARLANMASI**

**mövzusunda
MAGİSTRİK DİSSERTASİYASI**

İxtisas: 060632 - "İnformasiya tenologiyaları və sistemləri mühəndisliyi"

İxtisaslaşma: "İdarəetmədə informasiya sistemləri"

Elmi rəhbər:

t.e.n., dosent Z.Ə. Cəfərov

BAKİ – 2023

II FƏSİL. QRAFİK ANALİZ VƏ TƏSNİFAT ALQORİTMLƏRİ ƏSASINDA SAXTA HESABLARIN MÜƏYYƏN EDİLMƏSİ

2.1 Qrafik təhlili və oxşarlıq növləri.

Qrafik Təhlil bir çox tətbiqlərdə istifadə olunur, məsələn, SHAPE(fiqur) aşkar etmək üçün dövrə diaqramlarının göstərilməsi, təsvirin uyğunluğu və sosial şəbəkə təhlili. Sosial şəbəkə problemlərinin əksəriyyətini həll etmək üçün şəbəkələrin qrafiki təhlil edilir. Bundan sonra, qrafik oxşarlıq tədbirləri müxtəlif üsullardan istifadə etməklə qrafik təhlili problemlərinin mürəkkəbliyini azaldır. Bu qrafiklərdən bəziləri aşağıda təsvir edilmişdir.

Sosial şəbəkə $G = (E, N)$ qrafikə uyğunlaşdırılır, beləliklə N qovşağ çoxluğu sosial şəbəkə istifadəçilərini, $E \subset N \times N$ kənarlar çoxluğu isə əlaqələri təmsil edir. Bundan əlavə, nöqtə işarəsi qrafikdə müəyyən bir komponentə istinad etmək üçün istifadə edilmişdir. A G qrafiki üçün seyrək bitişiklik matrisini təmsil edir. Əgər (v, u) G -də kənardırsa, onda $A(v, u) = 1$. Əks halda, $A(v, u) = 0$. Dostluq qrafiki (FG): Sosial şəbəkə iş qrafiki G və v nəzərə alsaq, dostluq qrafiki birbaşa həmin əlaqəli qovşaqla olan və təpələri ehtiva edən təpədir:

$$FG(v).N = \{v\} \cup \{n \in G.N \mid n \neq v, \exists e \in G.E. e = \langle v, n \rangle\}$$

$$FG(v).E = \{\langle v, n \rangle \in G.E \mid n \in FG(v).N\}$$

$$\cup \{\langle n, n' \rangle \in G.E \mid n, n' \in FG(v).N\}$$

burada $FG(v).N$ və $FG(v).E$, v qovşağına birbaşa bağlı olan bütün təpələri və bu qovşaqlar arasındakı əlaqəni ehtiva edən təpəni göstərir.

Ortaq dostlar (CF): Sosial şəbəkələrdə oxşarlıq ölçülərindən biri paylaşılardostların sayıdır. Sosial şəbəkə G və iki v qovşağını nəzərə alsaq, uzunluğu u kimi, bu iki təpələr. qovşağ arasındakı ikisi həmin qovşaqların da göstərildiyi ümumi dostlarıdır.

$$CF(u, v) = |FG(v).N \cap FG(u).N|$$

Ümumi dostlar (TF): Bu, iki v və u qovşaqları arasında müxtəlif dostların sayını göstərir.

$$\text{Ümumi dostlar}(v, u) = |FG(v).N \cup FG(u).N|$$

Jaccard oxşarlığı (JS): Jaccard əmsalı nümunə çoxluqları arasındakı oxşarlığı təmsil edir və əslində iki qovşağın ümumi dostlarının bütün dostlarına nisbətini hesablamaq üçün istifadə olunur.

$$\text{Jaccard-coef}(v, u) = \frac{|FG(u).N \cap FG(v).N|}{|FG(u).N \cup FG(v).N|}$$

Kosinus oxşarlığı: Düyünlər arasında digər oxşarlıq ölçüsü kosinus oxşarlıq qrafikidir. Kosinus oxşarlığı əslində iki məhsul vektoru arasındakı oxşarlığı hesablayır.

$$\text{Cos}(v, u) = \frac{|FG(v).N \cap FG(u).N|}{\sqrt{|FG(v).N| \cdot |FG(u).N|}}$$

L1 norma oxşarlığı: Bu ölçü iki qovşağın üst-üstə düşən hissəsini ölçülərinə görə bölməklə əldə edilir.

$$\text{L1norm}(v, u) = |FG(u).N \cap FG(v).N|$$

Kənar çəki ölçüsü: Birincisi, kənar çəkilər

$$|FG(v).N| \cdot |FG(u).N|$$

iki kənarın hər biri üçün iki ayrı atribut kimi hesablanır.

$$W(u) = \frac{1}{\sqrt{1 + |FG(u).N|}}$$

On, u və v -nin iki tərəsi arasındakı kənarın çəkisi iki yolla hesablanmalıdır:

ümumi çəkilər: çəkilərin cəmi, aşağıda göstərildiyi kimi u və v üçün müəyyənləndirilmiş iki çəkinin cəminə bərabərdir.

$$W(v.u)=w(v) + w(u).$$

çəki əmsalı: bu parametr yuxarıda təsvir edilmiş iki çəkinin vurulmasıdır.

$$W(v.u) = w(v) * w(u)$$

2.2.Yenidən nümunə götürmə.Əsas komponent təhlili və Maşın Öyrənilməsinin tətbiqi.

Verilənlərin təsnifatındakı problemlərdən biri, bəzi siniflərdəki elementlərin digər siniflərdən daha çox olduğu məlumatların balanssız paylanmasıdır. Bu problem digərlərindən daha çox iki sinifli tətbiqlərdə yaranır; bu o deməkdir ki, birsinif digər sinifdən daha çox elementə malikdir.

Yenidən nümunə götürmə yanaşması məlumatların işlənməsi yolu ilə təlim nümunələri çoxluqlarının paylanmasının dəyişdirilməsi deməkdir. Verilənlər toplusunu balanslaşdırmaqla sinif səmərəliliyini artırmaq üçün bir neçə yanaşma var. Yenidən nümunə götürmə, məlumatı az seçmə yanaşmasından istifadə etməklə çoxluq sinfinin nümunələrini silmək və ya balanslaşdırmaq üçün həddindən artıq seçmə istifadə edərək azlıq sinfinin nümunələrini artırmaqla məlumat sinfinin paylanmasını balanslaşdırma bilər. Ona görə də, azlıq sinfi elementləri arasındakı xüsusiyyətlərin oxşarlığına əsaslanan süni məlumatların Sintetik Azlıqların Həddindən artıq Nümunə Alma Texnikasını (SMOTE) yaradan azlıq sinfi süni nümunəsi kimi tanınan başqa bir yanaşmadır. Təklif olunan modeldə qovşaqların oxşarlıq xüsusiyyətinin istifadəsi və məlumatı silmək istəməməsi səbəbindən SMOTE metodundan istifadə olunur. Bütün həddən artıq seçmə yanaşmalarında əsas məlumatlardan azlıq sinfi nümunələrinin təkrarlanması səbəbindən bu, səs-küy məlumatlarını və emal vaxtını artırır və

həddindən artıq yüklənməyə səmərəliliyin azalması ilə nəticələnə bilər.

SMOTE algoritmi təsadüfi olaraq azlıq sinfinin elementlərini yarada bilər müəyyən qayda əsasında və yeni təlim addımları yaratmaq üçün bu yeni nümunə elementlərini orijinal verilənlər toplusu ilə birləşdirir. Bu yanaşma yeni azlıqlar sinfi maddələrinin istehsalı üçün istifadə edilə bilər. Azlıq siniflərində fərqli nümunələr həddən artıq seçmə prosesində fərqli rollara malikdir və bu marjinal nümunələr azlıq sinfinin mərkəzində olan maddələrdən daha çox rol oynayır. Azlıq sinfinin marjasında əldə edilən nümunələr azlıq sinfi prototipləri üçün mövzunun tanınması qərarını və təsnifat dərəcəsini yaxşılaşdırmağa bilər.

Əsas Komponent Təhlili

Əsas komponent analizinin (PCA) əsas ideyası çoxvariantlı klassik metodlardan biridir və bəlkə də ən qədim və ən populyarıdır. Çoxlu məlumatların təhlili məlumatların təhlilində əyləncəli bir rol oynayır. Müşahidə edilməli olan çoxlu verilənlər bazasında çoxlu rejimlər və ya dəyişənlər var. Hər bir verilən çoxluğunda n dəyişən varsa, hər bir dəyişənin bir neçə ölçüsü ola bilər. Çoxölçülü məkanı qavramaq çox vaxt çətin olduğu üçün əsas komponentlərin təhlili metodu kombinasiya indeksi və oxşar müşahidələrin təsnifatı əsasında bütün müşahidələrin ölçülərini azaldır. PCA metodu xətti cəbri tətbiqin ən qiymətli nəticələrindən biridir və bütün analitik formalarda çox istifadə olunur, çünki mürəkkəb verilənlər toplusundan müvafiq məlumatı çıxarmaq üçün asan və qeyri-parametrik üsuldur. Bu üsulda çoxsaylı fəzada dəyişənlər bir-biri ilə əlaqəsi olmayan komponentlər toplusuna cəmlənir, onların hər biri əsas dəyişənlərin xətti kombinasiyasıdır. Alınankorrelyasiya olunmayan komponentlər xüsusi kovariasiya matrislərindən və ya əsas dəyişənlərin korrelyasiya matrislərindən alınan əsas komponentlərdir. Bu metod əsasən dəyişənlərin sayının azaldılması və dəyişənlər arasında əlaqə strukturunun tapılmasının əsas komponentlərini təhlil etmək üçün istifadə olunur.

Əsas komponentlər bütün verilənlər bazasında ən böyük fərqə malikdir və onlardan heç bir asılılıq yoxdur. PCA metodunda ən vacib məsələlərdən biri əsas komponentlərin sayının seçilməsidir. Formal və qeyri-rəsmi kateqoriyalara bölünə

bilən əsas komponentlərin sayını seçmək üçün bir neçə meyar təklif edilmişdir. Qeyri-rəsmi yanaşmada, ilk növbədə, məlumat və arzu olunan nəticələr üçün uyğun olan müvafiqdəqiqlik müəyyən edilir və sonra ümumi faiz əsasında variasiyaların ümumi sayı seçilir, ən yüksək dəqiqlik isə 80-90% arasında hesab olunur. Fərdi kompüterlərin sayını seçmək üçün istifadə olunan başqa bir üsul, Kaiser qaydası adlanan fərdi kompüter seçimi üçün birdən yüksək Eigen dəyərlərindən istifadə edən formal qrup metodlarının bir hissəsidir.

Maşın Öyrənmə.

Maşın öyrənmə üsullarının əksəriyyəti maşın öyrənmə alqoritmlərindən istifadə edərək təsnifatçıları öyrədir. Təsnifatçılar atribut oxşarlığı, şəbəkə dostu oxşarlığı və IP ünvan analizi kimi müxtəlif sosial şəbəkə atributlarına əsaslanır. Təklif olunan modeldə istifadə olunan bir sıra alqoritmlər olan maşın öyrənməsinin təsnifatçıları aşağıda təqdim olunur.

Statistik öyrənmə nəzəriyyəsi.

SVM eyni zamanda empirik səhvi və şagirdin mürəkkəbliyini minimuma endirən və təsnifat və reqressiya tapşırıqlarında yaxşı ümumiləşdirmə performansına nail olan struktur riskinin minimuma endirilməsi prinsipini həyata keçirir. Təsnifat üçün SVM-nin məqsədi ən böyük marja ilə optimal hiperplan qurmaqdır. Ümumiyyətlə, marja nə qədər böyükdürsə, təsnifatçının ümumiləşdirmə xətası da bir o qədər aşağı olur.

Bu məqalədə SVM təlimdə xətti və Gauss nüvəsi ilə istifadə edilmişdir. Gauss məlumat nöqtələri ətrafında normal əyrilərdən istifadə edir və bu məlumat nöqtələrini belə cəmləyir ki, qərar sərhəddi cəminin 0,5-dən yuxarı olduğu əyrilər kimi bir növ topologiya şərti ilə müəyyən edilə bilsin.

Logistik reqressiya.

Xüsusiyyət girişləri kimi $x^{(i)}$ və etikətlər kimi $y^{(i)} \in \{0, 1\}$ $i=1$ m təlim nümunələri nəzərə alınmaqla çoxluğu $S = \{(x^{(i)}, y^{(i)})\}_{i=1}^m$, logistik reqressiya göstərilə bilər:

$$P(y=1|x) = \frac{1}{1 + \exp(-\alpha^T x)}$$

Burada $\alpha \in \mathbb{R}^n$ model parametrləridir

Normallaşdırma olmadan, logistik reqressiya maksimum ehtimal kriteriyasından istifadə edərək parametrləri tapmağa çalışır, tənzimləmə ilə isə əlaqələr arasında mübadilə olur və modeldəki dəyişənlər daha azdır.

2.3. Təklif olunan metod. Qiymətləndirmə. Nəticəsi

Təklif olunan Metod

Saxta hesabın aşkarlanması probleminin xüsusiyyətlərinə əsaslanaraq, təklif etdiyimiz metod bu bölmədə təqdim edilmişdir. Əvvəlcə sosial şəbəkələr qrafikinin bitişiklik matrisi hesablanmışdır. O, şəbəkə dostunun qovşaqlar (sosial şəbəkə istifadəçiləri) arasında oxşarlıq ölçüləri hesablanmışdır. Bundan sonra ümumi dostların oxşarlığı, Jaccard oxşarlığı, kosinus oxşarlığı və digər ölçülər kimi müəyyən edilmiş ölçülərin hər biri üçün oxşarlıq matrisi hesablanmışdır. Bu addımın sonunda qovşaqlar arasında oxşarlığı təmsil edən bir neçə matris göstərilmişdir.

Nəzərə alsaq ki, belə hallarda verilənlər balanslaşdırılmır və həmçinin məlumatların təxminən 98-99%-i eyni çoxluq sinfinə (normal istifadəçilər) aiddir və belə məlumatlar üzərində iş azlıqlar sinfinin (saxta) dəqiqləşdirilməsinin məlumatsızlığına səbəb olur. istifadəçilər) və təsnifatların ümumi dəqiqliyinin artması, bütün məlumatların etikətlənməsi normal olaraq etikətləndi. Bu problemi həll etmək üçün məlumatları balanslaşdırmaq üçün SMOTE istifadə edilmişdir.

Bu oxşarlıq matrislərinin hər birinə SMOTE tətbiq edildikdən və məlumatları balanslaşdırdıqdan sonra, nəticədə hər biri yeddi nöqtəli oxşarlıq ölçülərini göstərən yeddi oxşarlıq matrisi var idi. Əvvəlki mərhələdə yeni oxşarlıq xüsusiyyətləri çıxarıldı. On, PCA metodundan istifadə etməklə, bu matrislərin hər birindən ən yüksək dispersiyaya malik on birinci sütun seçildi ki, yeni xassə matrisi formalaşsın. On, məlumat etikətləri tətbiq edildi və təsnifatçıya göndərildi. Təsnifat mərhələsində qovşaqlar xətti SVM alqoritmi, Orta Qauss DVM və logistik reqressiyadan istifadə etməklə, sonra isə normal və saxta istifadəçiləri ayırmaqla təsnifləşdirilib; siyahı növbəti mərhələyə göndərildi.

Süni saxta (SMOTE) istifadəçinin yaradılması üsulu aşağıda göstərilmişdir: arr1 nümunəsini nəzərdən keçirək və arr2 onun ən yaxın qonşusu olsun.

Arr1, K-ən yaxın qonşuların müəyyən edildiyi nümunədir.

Arr2 onun K-yaxın qonşularından biridir. Arr1 = (0,0045, 0,0014, 0,0145, 0,0046)

Arr2 = (0.003, 0.0004, -0.0135, 0.0057)

Deyək ki:

$$f1.1=0,0045$$

$$f1.2=0,0014$$

$$f2.1=0,003 \quad f2.1-f1.1= -0,0015$$

$$f2.2=0,0004$$

$$f2.2-f1.2= -0,001 \quad f1.3=0,0145$$

$$f2.3= -0,0135$$

$$f2.3-f1.3= -0,028$$

$$f1.4=0,0046$$

$$f2.4=0,0057$$

$$f2.4-f1.4=0,0011$$

Yeni nümunələr kimi yaradılacaq:

$$(f1', f2', f3', f4') = \text{Arr1} + \text{rand}(0-1) * (-0.0015, -0.001, -0.028, 0.0011)$$

Rand (0-1) 0 ilə 1 arasında təsadüfi bir ədəd yaradır. Arr1 və Arr2 v1 və v2dörd qovşağı arasında oxşarlığı ifadə edir.

Qiymətləndirmə

Təklif olunan metodun effektivliyi maşın öyrənmə üsulları ilə həyata keçirilir. Təsnifçi 10 dəfə çarpaz doğrulama yolu ilə öyrədilir; sonra effektivlik göstəriciləri hesablanır.

Bu hissədə birinci, çarpaz doğrulama texnikası müəyyən edilmiş və təsnifatçının performansının əsas ölçüləri və qiymətləndirilməsi təqdim edilmişdir.

Çarpaz Doğrulama. Çarpaz doğrulama proqnozlaşdırıcı modelləri qiymətləndirmək üçün istifadə edilən bir texnikadır. Bu texnikada orijinal nümunələr iki kateqoriyaya bölünür: model təlimi üçün təlim çoxluğu və qiymətləndirmə üçün test çoxluğu. Orijinal nümunə təsadüfi olaraq bərabər ölçülü kalt nümunəyə bölünür.

Bu alt nümunələrdən biri modeli sınaqdan keçirmək üçün qiymətləndirici məlumat kimi, qalanları, $k-1$ alt nümunələri isə təlim məlumatları kimi qəbul edilir. Çarpaz doğrulama prosesi k alt nümunə üçün k dəfə təkrarlanır, hər dəfə onlardan biri üçün qiymətləndirmə məlumatları kimi. Bu metodun birinci üstünlüyü ondan ibarətdir ki, bütün nümunələr həm təlim, həm də doğrulama prosesi üçün istifadə olunur, ikincisi isə hər bir nümunənin yalnız bir dəfə yoxlama üçün istifadə olunmasıdır.

Qiymətləndirmə Metrikləri. Qiymətləndirmə qarışıqlıq matrisinə və əlaqəli ölçülərə əsaslanır. Qarışıqlıq matrisində TP, FP, TN və FN dəyişənləri aşağıdakılara istinad edir:

Həqiqi müsbət (TP): saxta qovşaqların sayı saxta qovşaqlar kimi bəslənən identifikasiya,

Yanlış müsbət (FP): saxta qovşaqlar kimi müəyyən edilən normal qovşaqların sayı,

Həqiqi mənfi (TN): normal qovşaqların sayı normal düyünlər kimi müəyyən edilir,

Yalan mənfi (FN): saxta qovşaqların sayı normal düyünlər kimi müəyyən edilir.

Input

İşarəli node ilə qrafik $G(N, E)$.

Output

Saxta node (hesab) siyahısı

Prosedur:

$A \leftarrow$ bitişiklik matrisi (G) Bütün

oxşarlıqlar üçün

$SM \leftarrow$ oxşarlıq matrisini hesablayın

SM←PCA Alqoritmi ilə xüsusiyyət çıxarılması (SM)

End

SM←bütün oxşarlıq matrisində on sütun seçin (SM)

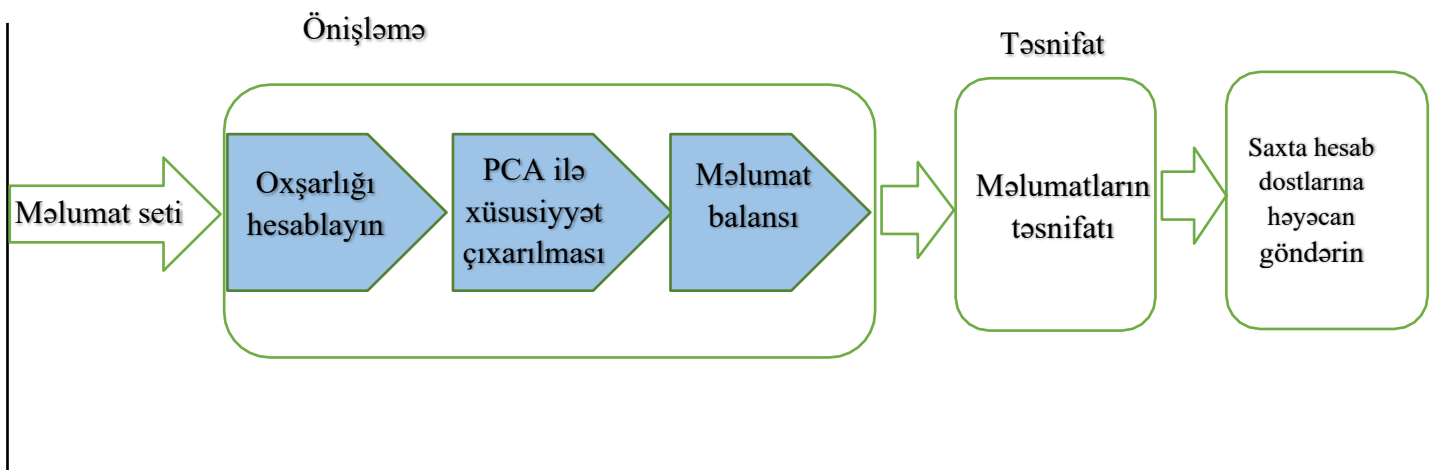
SM←balans SMOTE aktiv (SM)

Təlim təsnifatı

Saxta qovşaqların siyahısını qaytarın

Dostlar saxta hesablar üçün həyəcan göndərin.

Aşkarlama yanaşmasının diaqramı



Şək.2.1 Aşkarlama yanaşmasının diaqramı

Klassifikatoru qiymətləndirmək üçün dəqiqlik və əyri altındakı sahədən (AUC) istifadə olunur. AUC ikili təsnifatçılar üçün performans göstəriciləridir; bu AUC birinə nə qədər yaxın olarsa, təsnifatın son performansını bir o qədər əlverişli olacaqdır. ROC əyriələrini AUC ilə müqayisə edərək, əyrinin şimal-qərb küncündə nə dərəcədə yuxarı olduğunu göstərir. Aşağıda təqdim olunan göstəricilər ROC-u hesablamaq üçün istifadə olunur.

Həqiqi mənfi nisbət (TNR) = $TN / (TN + FP)$.

Yanlış müsbət nisbət (FPR) = $FP / (FP + TN)$.

Həqiqi müsbət nisbət (TPR) = $TP / (TP+FN)$.

Yanlış mənfi nisbət (FNR) = $FN / (FN+TP)$.

Performansı qiymətləndirmək üçün istifadə edilən başqa bir ölçüdür: Dəqiqlik = $(TP+TN) / (TP+FP+TN+FN)$.

Təklif olunan Modelin Performansı. Təklif olunan metodu qiymətləndirmək üçün Twitter verilənlər bazasından - real dünya etiketli verilənlər toplusundan istifadə edilmişdir.

Bu araşdırmanın nəticələrini dəstəkləmək üçün istifadə edilən Twitter məlumatları

GitHubsaytında(<https://translate.google.com/?hl=ru&sl=en&tl=az&text=repository&op=translate>) yerləşdirilib .

Bu verilənlər bazasında 5.384.162 istifadəçi var, onların arasında 16.011.445 keçid var. Bu verilənlər bazasından 1000 qovşaq əldə edilib ki, onlardan 990-ı normal, 10-u isə saxtadır.

Bu verilənlər bazası normal və saxta qovşaqlar arasında 1:100 nisbətinə malikdir. Bu məlumatda qovşaqlar arasında mövcud əlaqələr haqqında məlumat göstərilir, sonra qrafikin qonşuluq matrisi əldə edilir və qovşaqlar arasında oxşarlıq ölçüləri hesablanır. Yeni xüsusiyyətlər PCA texnikasından istifadə edərək çıxarılır. Bundan sonra SMOTE vasitəsilə süni məlumatlar yaradılıb. SMOTE tətbiq etməklə məlumatların paylanması dəyişdirilir. Bu o deməkdir ki, 99% normal istifadəçilər və 1% saxta istifadəçilər 75% normal və 25% saxta istifadəçilərə dəyişdirilir və bu balanslaşdırılmış məlumatlar növbəti mərhələyə göndərilir. Modelin performansını qiymətləndirmək üçün FPR, TPR, dəqiqlik və AUC hesablamaq üçün çarpaz doğrulama texnikasından istifadə edilmişdir. Klassifikatorların nəticələrinin müqayisəsi göstərdi ki, bəzi təsnifatçılar daha dəqiq, bəziləri isə daha yüksək səviyyəli AUC-yə malikdir. Bu alt səviyyə birinə yaxınlaşdıqca, performans dəqiqliyi bir o qədər yüksək olar. Cədvəldə üç alqoritm üçün sınaq TPR, FPR, dəqiqlik və AUC(əyri altındakı sahə) göstərilir. Orta Gauss teoreminə əsasən SVM kimi qeyri-

xətti metodların istifadəsi, verilənləri daha yüksək ölçülü xüsusiyyət fəzalarına uyğunlaşdırdıqlarına görə, məlumatları fərqləndirmək qabiliyyətinə malikdir və bu metoddan istifadə edərək modelin performansını ilə nəticələnir.

Cədvəl 2.1 Təsnifatçının performans müqayisəsi

AUC	Dəqiqlik	FPR	TPR	Alqoritm
98%	95.8%	4%	96%	Xətti SVM
1%	97.6%	2%	97%	Orta Qauss SVM
96%	96.6%	3%	94%	Logistik reqressiya

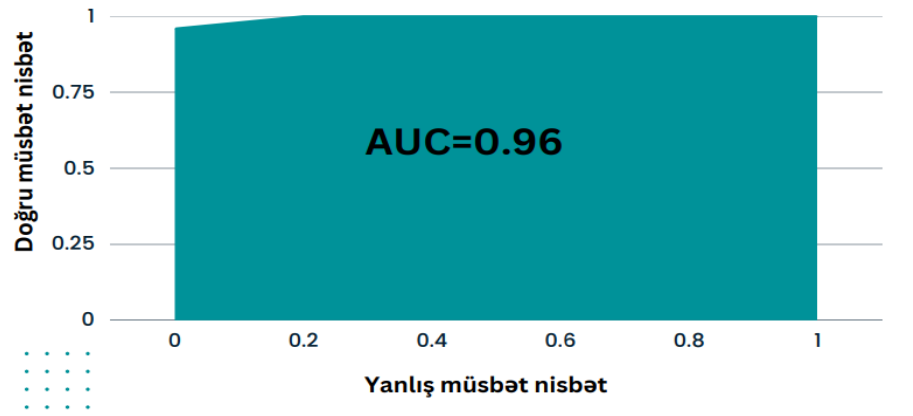
Cədvəl 2.2 İki halda Orta Qauss DVM-nin müqayisəsinin nəticələri, balans/balansməlumatları.

Alqoritm	TPR	FPR	Dəqiqlik	AUC	Balans/balanssızlıqverilən çoxluğu
Orta Qauss SVM	97%	%	97.6%	1	Balans
Orta Qauss SVM	0		99%	47%	Balanssızlıq

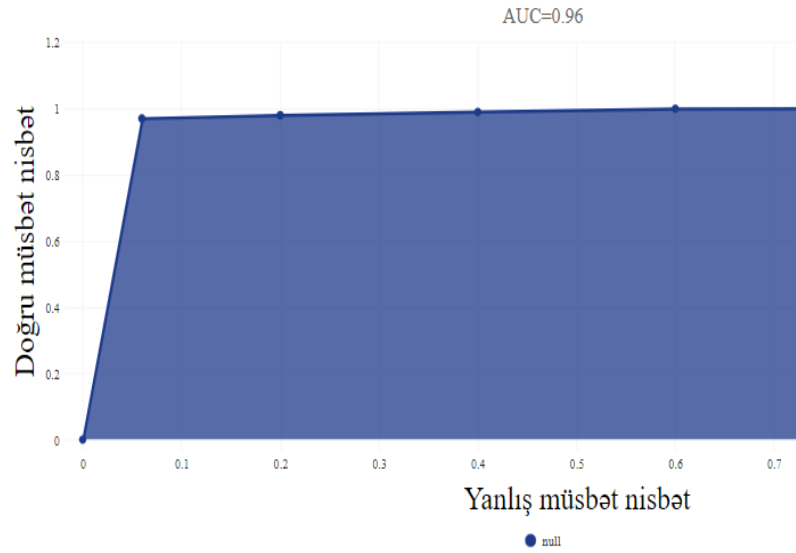
Bu nəticələri <https://www.canva.com/graphs/> vasitəsilə göstərdik və müqayisə etdik.



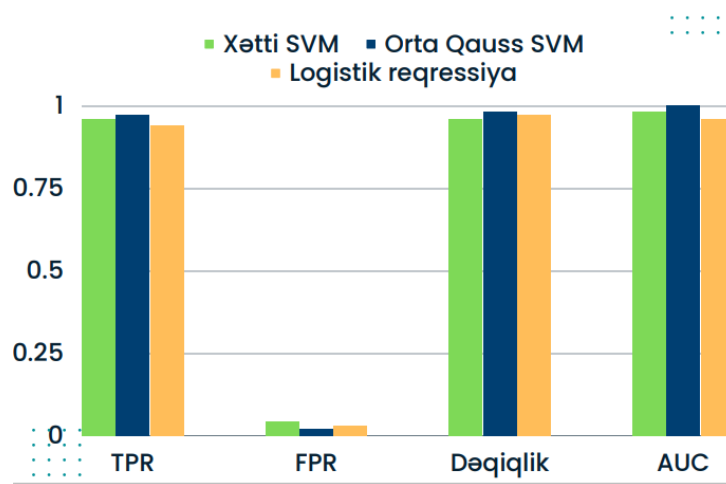
Şək. 2.1 Orta Qauss SVM təsnifatçısı üçün AUC diaqramı



Şək. 2.2 SVM Təsnifatçısı üçün AUC diaqramı



Şək. 2.3 Logistic reqressiya təsnifatçısı üçün AUC diaqramı



Şək. 2.4 Klassifikatorların performansının müqayisəsi

Nəticə Cədvəl və Şəkildə əldə edilən qiymətlərə əsasən, balanssız məlumatlarazlıq sinfindən verilən məlumatların təsnifatçı tərəfindən nəzərə alınmamasına, hamısının normal etiketlə proqnozlaşdırılmasına və heç bir düyünün saxta etiketlə etiketlənməməsinə səbəb olmuşdur. Bu hərəkət yalnız bütün sistemin dəqiqliyini artırır.

Sosial şəbəkələrdə saxta hesabları müəyyən etmək üçün istifadəçinin dostlarının oxşarlığına əsaslanan üsul təqdim edilib. Bu üsulda əvvəlcə şəbəkə qrafikinin bitişiklik matrisindən dost oxşarlıq meyarları hesablanmış və PCA metodundan yeni xüsusiyyətlər çıxarılmışdır. Növbəti mərhələdə məlumatlar SMOTE istifadə edərək balanslaşdırılmış və təsnifatçıya göndərilmişdir. Çarpaz doğrulama texnikasından istifadə edərək, təsnifatçı təlim və sınaqdan keçirildi ki, buda göstərdi ki, Orta Gauss SVM təsnifatçısı $AUC = 1$ -ə malikdir.

Medium Gaussian metodu ilə istifadəçi dostu şəbəkə strukturu təhlil edilmiş və hesablama oxşarlığı və təsnifat alqoritmləri ilə saxta istifadəçilər proqnozlaşdırılmışdır. Bu üsulda saxta hesablar şəbəkədə işləməlidir ki, dostunun şəbəkələrini təhlil edərək onları qanuni və ya saxta hesab etmək mümkün olsun. Bu təklif olunan metodun zəif tərəfidir. Gələcək tədqiqatlarda yeni üsul təqdim olunacaq; istifadəçinin şəbəkədəki hər hansı fəaliyyətindən əvvəl və ya qeydiyyat zamanı qanuni və ya saxta hesabı tanıya bilər.

Sosial media çoxlu sayda agentdən (subyektlər - fərdlər, icmalar, fərdlər qrupları və ya təşkilatlar) və onun üzərində müəyyən edilmiş münasibətlər toplusundan ibarətdir.

Bu gün dünyanın ən populyar sosial mediası olan Facebook-un istifadəçilərinin sayı bir neçə milyard nəfərə çatır və onun saxladığı məlumatların həcmi təxminən 600 petabayt təşkil edir.

Facebook istifadəçiləri hər gün 5 milyarddan çox nəşr hazırlayırlar ki, bu da istənilən ölkə əhalisinin sosial vəziyyətini araşdırmaq üçün əla bazadır. Buna görə də belə nəticəyə gəlmək olar ki, sosial media çoxlu sayda saxta hesablar üçün əla bazadır

və onların informasiya xarakterli müharibə zamanı digər istifadəçilərə təsiredir.

Hesab adətən daxil olduqda istifadəçini müəyyən etmək üçün lazım olan məlumatları, avtorizasiya və uçot üçün məlumatları ehtiva edir. Bu istifadəçi ID və paroldur. Parol və ya onun analoqu, bir qayda olaraq, istifadəçi təhlükəsizliyi üçün şifrələnmiş və ya səhifələnmiş formada saxlanılır. Bununla belə, sirr deyil ki, Facebook öz serverlərində adi istifadəçiyə görünməyən çoxlu hesab məlumatlarını saxlayır. Belə məlumatlara aşağıdakılar daxildir:

- İstifadəçinin geolokasiya məlumatları;
- Şərhlər və istifadəçi yazıları haqqında məlumatlar;
- Bu və ya digər istifadəçi tərəfindən fotosəkildə istifadəçi işarələri haqqında məlumatlar;
- Şəbəkə siqnalının gücü;
- Əməliyyat sistemi;
- Brauzer;
- Xidmət provayderi / internet provayderi;
- Üçüncü tərəf kukiləri (axtarış və alış-veriş daxil olmaqla);
- İstifadəçi tərəfindən gizli yazılar;
- Hesabdan silinmiş hesab məlumatları;
- “Bəyən” işarələri;
- Və daha çox məlumat;

Beləliklə, sosial mediada hesabın ümumi strukturu həm istifadəçinin gördüyü hesabın görünən məzmununu, həm də sosial mediada saxlanılan adi istifadəçi tərəfindən görünməyən məzmunu əks etdirən sxem kimi təqdim olunur.

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ
AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

Məmmədzadə Nərminə Fizuli qızı

**DATA MİNING YOLU İLƏ SOSIAL ŞƏBƏKƏLƏRDƏ SAXTA
HESABLARIN AŞKARLANMASI**

mövzusunda

MAGİSTRİK DİSSERTASİYASI

İxtisas: 060632 - “İnformasiya tenologiyaları və sistemləri mühəndisliyi”

İxtisaslaşma: “İdarəetmədə informasiya sistemləri”

Elmi rəhbər:

t.e.n., dosent Z.Ə. Cəfərov

BAKİ – 2023

olmaması və hər cür maneələr virtual məkanda uğurlu aksiyalar üçün əlverişli zəmin yaradır. Bütün bunlar sosial medianın informasiya qarşılıqlı əlaqəsinə meydanına çevrilməsindən xəbər verir.

Saxta hesabın aşkarlanması üçün müxtəlif parametrlər təklif olunur:

Cədvəl 3.1 Saxta aşkarlama üçün daxil edilən məlumat

Hesabların statik parametrləri	Hesabların dinamik parametrləri
--------------------------------	---------------------------------

Əsas parametrlər:	Dəyişən parametrlər:	Davranış parametrləri:
Avatar	şərhlər	dəyişikliklərin tezliyi
Ad	bəyənmələr	sosial şəbəkələrin qrafiki
ad günü.	şəkillər	Və s.
və s	dostlar	
	maraqlar	
	İP adresləri	
	Və s	

Statik parametrlər adətən ad təhlili üçün istifadə olunur, məsələn, profil şəkli (avatar), ad, doğum tarixi, dostların sayı, şəkillər, bəyənmələr və s. Onlardan bəziləri dəyişkəndir, məsələn: dostların sayı və onların vaxt intervalında dəyişdiyi dinamik analiz üçün istifadə edilə bilər. Bəzi yanaşmalara sosial media bağlantılarının qrafik qurulması və ya hesablardakı dəyişikliklərin onlayn təhlili üçün istifadə olunan davranış parametrləri daxildir. Daxil edilən parametrlərə görə əsasında saxta hesabın aşkarlanması üçün müxtəlif üsullar təklif olunur:

- Hesab parametrlərinin xülasəsi (reyting);

- Vaxt intervalı üçün hesab təhlili;
- Maşın öyrənmə alqoritmləri;
- Qrafikə əsaslanan texnikalar;
- Dig Data yanaşmalar.

Saxta hesabların aşkarlanması problemi aktual olaraq qalır və bu tədqiqatın məqsədləri sosial mediada real və saxta hesabların parametrlərini təhlil etmək və saxta hesabların aşkarlanmasının yüksək dəqiqliyi ilə əsas maşın öyrənməsi əsaslı klassifikatoru müəyyən etməkdir.

3.2.METODLAR VƏ MATERİALLAR

Sosial media çoxlu sayda agentdən (subyektlər - fərdlər, icmalar, fərdlər qrupları və ya təşkilatlar) və onun üzərində müəyyən edilmiş münasibətlər toplusundan ibarətdir.

Bu gün dünyanın ən populyar sosial mediası olan Facebook-un istifadəçilərinin sayı bir neçə milyard nəfərə çatır və onun saxladığı məlumatların həcmi təxminən 600 petabayt təşkil edir.

Facebook istifadəçiləri hər gün 5 milyardan çox nəşr hazırlayırlar ki, bu da istənilən ölkə əhalisinin sosial vəziyyətini araşdırmaq üçün əla bazadır. Buna görə də belə nəticəyə gəlmək olar ki, sosial media çoxlu sayda saxta hesablar üçün əla bazadır və onların informasiya xarakterli müharibə zamanı digər istifadəçilərə təsiredir

Hesab adətən daxil olduqda istifadəçini müəyyən etmək üçün lazım olan məlumatları, avtorizasiya və uçot üçün məlumatları ehtiva edir. Bu istifadəçi ID və paroldur. Parol və ya onun analoqu, bir qayda olaraq, istifadəçi təhlükəsizliyi üçün şifrələnmiş və ya səhifələnmiş formada saxlanılır. Bununla belə, sirr deyil ki, Facebook öz serverlərində adi istifadəçiyə görünməyən çoxlu hesab məlumatlarını saxlayır. Belə məlumatlara aşağıdakılar daxildir:

- İstifadəçinin geolokasiya məlumatları;
- Şərhlər və istifadəçi yazıları haqqında məlumatlar;
- Bu və ya digər istifadəçi tərəfindən fotosəkildə istifadəçi işarələri haqqında məlumatlar;
- Şəbəkə siqnalının gücü;
- Əməliyyat sistemi;
- Brauzer;
- Xidmət provayderi / internet provayderi;
- Üçüncü tərəf kukiləri (axtarış və alış-veriş daxil olmaqla);
- İstifadəçi tərəfindən gizli yazılar;
- Hesabdan silinmiş hesab məlumatları;
- “Bəyən” işarələri;
- Və daha çox məlumat;

Beləliklə, sosial mediada hesabın ümumi strukturu həm istifadəçinin gördüyü hesabın görünən məzmununu, həm də sosial mediada saxlanılan adi istifadəçi tərəfindən görünməyən məzmunu əks etdirən sxem kimi təqdim olunur.

Cədvəl 3.2 Sosial mediada hesab məlumatlarının ümumistrukturu

Şəkil (avatar)	İstifadəçi adı
İstifadəçi fotoları	İstifadəçi statusu
İstifadəçi maraqları	Şəxsi məlumat.İstifadəçi haqqında
Multimedia	İstifadəçi postları

Cədvəl 3.3 Təsnifatçının performans müqayisəsi

Geolokasiya	3cü şəxs cookie-lər	Gizli məlumatlar
Etiketlər	Əməliyyat sistemi	Brauzer
Silinmiş məlumatlar	Postlar və şərhələr	Və s.

Əhəmiyyətli odur ki, müasir sosial medianın imkanı olduğu üçün hesabın bütün məlumatları üçüncü tərəf istifadəçilər tərəfindən görünür.

Facebook-un yaradıcısı Mark Zukerberq hesabların təxminən yarısının (və bu, 1 milyardan çox hesabın) saxta olduğunu bildirib. Ona görə də sosial mediada saxta hesabların aşkarlanması hazırda aktual məsələdir.

Sosial mediada məlumatların təhlili üçün ölçülərin formalaşdırılmasında siz bəyənmlər, şərhələr, nəşrlərdəki repostlar, istifadəçi identifikasiyası məlumatları (cins, yaş, yaşayış şəhəri, iş/təlim yeri, ünvan və s.), haqqında məlumatlardan istifadə edə bilərsiniz. İstifadəçinin dostları, onların icması, istifadəçinin sosial mediada qalma statistikasını, istifadəçilər arasında əlaqələr və s.

Bəyənmə işarələrinə görə bəyənmə sayına və onları istifadəçi səhifəsində kimin qoyduğuna bölünür. Öz növbəsində həm dostlar, həm də yad adamlar bəyənmə alır. Bəyənmələrin sayı da profili müəyyənləşdirmək üçün vacibdir. Müəyyən bir yazıda istifadəçinin dostlarının sayından çox bəyənmə varsa, bu istifadəçinin bu bəyənmələri qeyri-adi şəkildə aldığını göstərə bilər. Səhifədə ləqəbin olmaması istifadəçinin “təcrid olunmasını” göstərir ki, bu da onun saxta olduğunu göstərə bilər.

İstifadəçinin səhifəsindəki şəxsi məlumatlar profil və ya profilin həqiqiliyi haqqında çox şey göstərir. Əlavə təhlil üçün şəxsi məlumatlar istifadəçi adlarına, məlumatların miqdarına, ziddiyyətli məlumatlara və şəxsi məlumatlara bölünür.

Doğum tarixində səhifənin etibarlılığını göstərən əlamətlər var. Çox vaxt saxta

hesabların istifadəçiləri hesab yaratarkən diqqət yetirmirlər, doğum gününü standart olaraq daxil edirlər (adətən 1 yanvar). İstifadəçinin yaşının şübhəli olması və ya səhifədəki digər tarixlərlə üst-üstə düşməməsi də mümkündür. Məsələn, istifadəçinin 15 yaşı var, lakin səhifədəki digər məlumatlarda onun universiteti 10 iləvvəl bitirdiyi göstərilir.

İstifadəçi adını araşdırmaq çətindir, çünki eyni ad və soyadlı çoxlu insan var. Ancaq məşhur insanların adı ilə üst-üstə düşməsi üçün adı yoxlamağa dəyər. İstifadəçi adının bu istifadəçinin ölkəsinin standart adlarına aid olub-olmamasına diqqət yetirməlisiniz.

Profildə şəxsi məlumatların olmaması və maraqlar və istifadəçi qrupları haqqında əhəmiyyətsiz məlumatlar istifadəçinin digər istifadəçilər tərəfindən müəyyən olunmaq istəmədiyini göstərir və buna görə də bu saxtakarlığın əlamətidir. Yazıların və statusların redaktə/əlavə edilməsi tezliyi istifadəçinin aktivliyini göstərir. Yazılar/statuslar nadir hallarda və ya çox tez-tez əlavə edilirsə - bu saxtakarlığın əlamətlərindən biridir. Əgər istifadəçi çoxdan post/status əlavə edibsə və onu uzun müddət yeniləməyibsə, çox güman ki, bu hesab saxtadır. Səhifədəki ziddiyyətli məlumatlar saxtakarlığın ən etibarlı göstəricilərindən biridir; lakin kompleks təhlil tələb edir. Məsələn, istifadəçinin yazılarındakı məlumatlar profildə göstərilən məlumatlarla uyğun gəlmir və ya istifadəçi onun bəyan etdiyi maraqlarına cavab verməyən qruplardadır

Şəxsi məlumatlara e-poçt və mobil telefon nömrəsi daxildir. İstifadəçilər saxta hesablardan və xüsusi yaradılmış tanıtım profillərindən fərqli olaraq bu cür məlumatları nadir hallarda ictimaiyyətə açıqlayırlar.

İstifadəçinin səhifəsində vahid obyekt kimi status və yazılar yalnız profilin yerləşdirilməsi ilə fərqləndiyindən aşağıdakı xüsusiyyətlər əsasında təhlil edilir: redaktə/əlavə və şərh tezliyi. Status və yazılar bəzən reklam kimi istifadə olunur.

Yazıların və statusların redaktə/əlavə edilməsi tezliyi istifadəçinin aktivliyini göstərir. Yazılar/statuslar nadir hallarda və ya çox tez-tez əlavə edilirsə - bu saxtakarlığın əlamətlərindən biridir. Əgər istifadəçi çoxdan post/status əlavə edibsə və

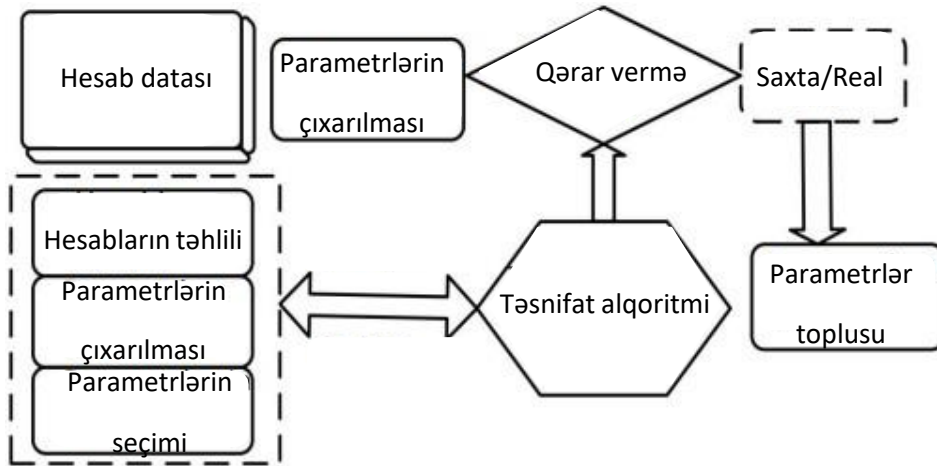
onu uzun müddət yeniləməyibsə, çox güman ki, bu hesab saxtadır

Şərhlərin sayı həm də profilin özünün aktivliyini göstərir. Onların yoxluğu və ya həddindən artıq sayı ən çox rast gəlinir. Həm istifadəçinin dostları, həm də yad adamlar şərhlərə yazı yaza bilirlər.

İstifadəçinin dostları saxta müəyyənləşmədə mühüm rol oynayırlar, çünki onlar həm profilin sosial mediada aktivliyini, həm də istifadəçinin maraq dairəsini göstərirlər. Saxta istifadəçinin dostlarının sayından asılıdır və onu təhlil etmək çətinidir, çünki nəticə çıxarmaq üçün dostların özlərini təhlil etmək lazımdır. Məsələn, istifadəçinin dost siyahısında saxtakarlıq varsa, çox güman ki, istifadəçinin özü də saxtadır. Əgər istifadəçinin dostları yoxdursa, onun profilindən ünsiyyət üçündəyil, başqa məqsədlər üçün istifadə olunma ehtimalı yüksəkdir. Profil yaratdıqdan sonra qısa müddət ərzində çoxlu dost qazanmaq da şübhə doğurur, ona görə də çox güman ki, belə profil saxtadır.

İstifadəçi fotolarının təhlili hesabın saxtılığının öyrənilməsində mühüm və eyni zamanda, ən çətin rol oynayır. Birincisi, həm avatarda, həm də albomlarda fotosəkillərin olmaması artıq bu hesabın saxta olduğunu göstərir. İkincisi, səhifədə təqdim olunan fotosəkillərin İnternetdəki digər şəkillərlə və ya digər profillərin fotosəkilləri ilə üst-üstə düşməsi üçün təhlil edilməlidir, çünki istifadəçi məşhurların, heyvanların və ya digər obyektlərin şəkillərini yükləyə və istifadə edə bilər. Fotoların sayı da vacib göstəricidir, çünki həddindən artıq və ya az miqdarda fotosəkillər onların saxtılığını və ya istifadəçinin hərəkətsizliyini göstərir.

Təbii ki, bu meyarlar təkbaşına hesabın “saxtılığına” işarə edə bilməz, çünki sadəcə onların assosiasiyasını təhlil etmək hesabın həqiqiliyini şübhə altına ala bilər. Hesab statusunun daha dəqiq tərfi üçün mümkün qədər çox meyardan istifadə edərək təhlildən istifadə etməlisiniz.



Şəkil 2.1 Saxtakarlıq aşkarlamaq sisteminin ümumi quruluşu

Parametrlərin təhlili, çıxarılması və seçilməsi modulu təsnifat modulu üçün məlumat toplusunun toplanmasının öyrədilməsi və yoxlanılması üçün istifadə olunur. Qərar vermə sistemi saxta hesabı aşkar etmək və sonrakı təhlil və öyrənmə üçün parametrlər toplamaq üçün öyrənilmiş təsnifatçıdan istifadə edir. Saxta aşkarlama sisteminin ümumiləşdirilmiş strukturu şəkildə göstərilmişdir.

Sosial şəbəkələrdə informasiyanın təhlili üçün təsnifat vəzifələrini həll edən çoxlu üsullar mövcuddur.

- Bayes klassifikatoru (ən yaxın qonşular üsulu, xətti Fisher diskriminantı);
- Neyron şəbəkəsi (perseptron);
- Xətti ayırıcı (logistik reqressiya, xətti Fisher diskriminantı);
- İnduksiya (qərar ağacı, test alqoritmi);
- Ölçünün azaldılması (əsas komponent metodu, müstəqil komponentmetodu);
- Modelin seçimi (empirik riskin minimuma endirilməsi, genetik alqoritm,

modellərin öz-özünə təşkili).

Facebook-dan əldə edilən məlumatların sonrakı tədqiqi və təhlili üçün dəstəklənən vektor maşını (SVM) metoduna əsaslanan neyron şəbəkəsindən istifadə olunur. SVM-nin əsas vəzifəsi məlumatları iki sinfə bölən ən düzgün xəttivə ya hipertəpəni tapmaqdır. SVM, giriş məlumatlarını qəbul edən və müvafiq bölmə xətti siniflərini qaytaran bir alqoritmdir.

SVM alqoritmi elə qurulmuşdur ki, o, ayırma xəttinə ən yaxın olan müstəvi nöqtələri axtarır. Bu nöqtələrə istinad vektorları deyilir. Sonra alqoritm istinad vektorları ilə bölmə müstəvisi arasındakı məsafəni hesablayır. Bu məsafə interval adlanır. Alqoritmın əsas məqsədi boşluq məsafəsini maksimuma çatdırmaqdır.

Hipermüstəvi hiperplan hesab olunur, bunun üçün bu boşluq maksimumdur. Hiperplane n -ölçülü Evklid fəzasındakı $(n-1)$ ölçülü alt fəzadır və məkanı iki ayrı hissəyə bölür.

İki ölçülü məlumat çoxluq üçün bölücü xətt hiperplandır. Sadə dillə desək, n ölçülü fəza üçün bu məkanı iki hissəyə bölən $(n-1)$ ölçülü hipermüstəvi mövcuddur.

Neyron şəbəkəsi əmsalların və çəkirlərin müəyyən dəyərlərini saxlayan n sayda neyrondan ibarətdir. Ümumiyyətlə, məlumatların daxil edilməsi üçün riyazi hesablama modelidir

Neyronların giriş təbəqəsinə daxil olur, onları gizli təbəqələrdə emal edir və çıxış qatını əldə edir.

Sistemin əsas xüsusiyyəti açıq şəkildə proqramlaşdırılmamış, öz-özünə öyrənmədir.

Neyronlar, ən sadə hesablama vahidləri kimi, qəbul edilən məlumatın sadə hesablamalarını yerinə yetirir və onu daha da ötürür, beləliklə, hər bir neyron giriş signalının və çıxış signalının iki əsas parametrinə malikdir və beləliklə, ilkin məlumatları ötürür. Təsnifat problemi vəziyyətində çıxış təbəqəsindəki neyronların sayı modelin hesablanması nəticələrinin aid edilməli olduğu siniflərin sayından

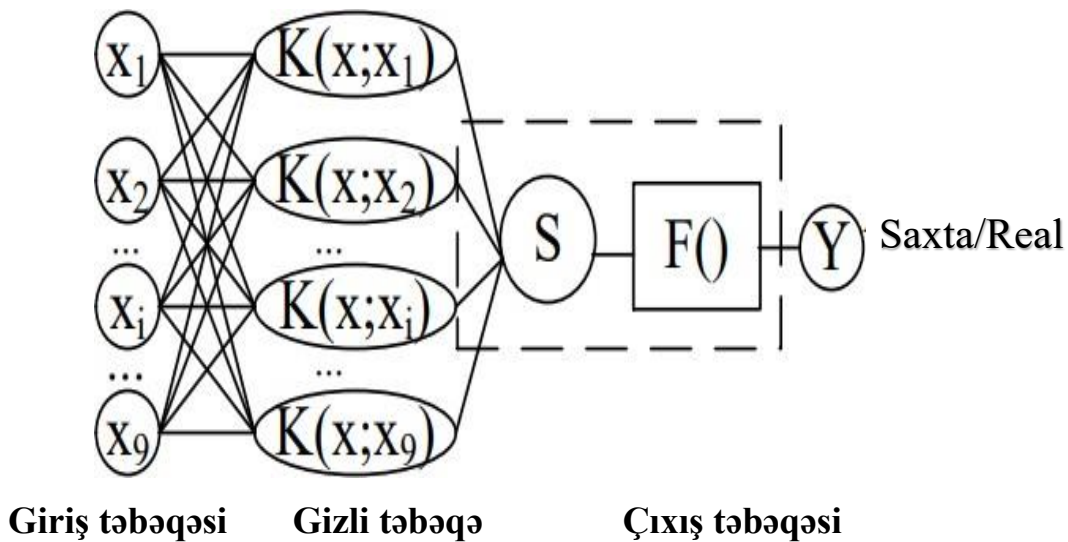
asılıdır. Bu halda, iki sinif var: hesab saxta və ya doğru ola bilər və çıxış signalı

«0»və ya «1» qiymətini alır.

Giriş vektoru giriş qatına verilir, onun hər bir elementi neyronun müəyyən çəkili ilə vurularaq neyronlar arasında qurulan əlaqələrə uyğun olaraq növbəti təbəqəyə ötürülür.

Bu halda, təsnifatda yoxlanılan hesab parametrlərinin hər biri üçün cavabdeh olan 9 giriş ($x_1, x_2, x_3, \dots, x_9$) və «Real» və ya «Saxta» qiymətini alan bir çıxış Y var.

Növbəti təbəqədə, giriş qiymətlərinin və sabitin cəmi S -dən əvvəl normallaşdırılmış signalı qaytaran və onu şəbəkə üzərindən daha da ötürən müəyyən aktivləşdirmə funksiyası $F()$ tətbiq olunur.



Şəkil 3.2 SVM təsnifatçısının quruluşu

3.3. Metrik dəyərlərinə görə nöqtələrin təyin edilməsi meyarları

Məlumatların təsnifatı problemini həll etmək üçün şəbəkə öyrənilməsi aşağıdakı kimidir: təsnifat "giriş - çıxış" cütlərində öyrənilir, burada çıxış obyektə aid edilməli olan düzgün sinifdir. Klassifikatorun öyrənilməsi neyron çəkiliyərinin elə kalibrlənməsi zamanı başa çatır ki, şəbəkə hər hansı bir giriş dəyərini verərkən düzgün cavablar verəcək.

Klassifikatorda giriş və çıxış təbəqələri seçilir. Neyroşəbəkə məlumatı giriş reseptorları tərəfindən qəbul edir və sonra bu məlumatı özündən keçirərək onu çevirərək çıxış dəyərlərini yaradır. Neyron şəbəkə daxilindəki qovşaqlar gizli adlanır, çünki onların çıxış və giriş qovşaqları kimi xarici dünya ilə əlaqəsi yoxdur. Onlar yalnız əvvəlki təbəqələr aktivləşdirildikdə çağırılır.

Göstərici kimi reyting sistemi seçilir. Təhlil zamanı hər bir göstərici müəyyənləşir

0-dan 3-ə qədər şərtlərdən asılı olaraq balların sayı Hər bir parametrlər üçün alınan qiymətlər təsnifatçının girişinə qidalanan verilənlər massivində formalaşır və bu da öz növbəsində onları təhlil edir və nəticəni verir. -dən asılı olaraq vəziyyətdə, hesabın vəziyyətini dəqiq müəyyən etmək üçün mütəxəssislərin iştirakı ilə və histoqramda olan məlumatları nəzərə alaraq əlavə araşdırmalara ehtiyac var.

Cədvəl 3.6 Metrik dəyərlərə görə nöqtələrin təyin edilməsi meyarları

Parametr	0	1	2	3
Dostların sayı	30<Dostlarının sayı<500	500<Dostlarının sayı<2000	0<Dostların sayı<30& Dostların sayı >2000	Dostlarınınsayı=0
Avatarda Foto	Şəkil mövcuddur	-	-	Şəkil mövcud deyil
Şəkillərin sayı	30<Şəkillərinin sayı<500	Şəkil sayı>500	0<Şəkillərin sayı<100	Şəkillərinin sayı=0
Profil şəkili	Şəkil mövcuddur		Şəkil mövcud deyil.	-
Postların sayı	-	Post mövcuddur	-	Post mövcud deyil
Ad günü	1932<Ad günü<2009	2009<Ad günü	Ad günü<1932	Ad günü göstərilmir
Şəxsi məlumat	5 sahə doldurulub	3-4 sahə doldurulub	1-2 sahə doldurulub	0 sahə doldurulub
Yeniləmələr	Yeniləmə<14gün	14gün<Yeniləmə<1 ay	1ay<Yeniləmə <1 il	Yeniləmə >1 il
İstifadəçi adı	İstifadəçi adı klassik ölkə istifadəçisinin ilə uyğunlaşır	-	İstifadəçi adı klassik ölkə istifadəçisinin ilə uyğunlaşmır.	-

Bu məqsədlə sütunların hər birinin yoxlanılması və onların təsir səviyyəsinin müəyyən edilməsi əldə edilir. Ümumilikdə dörd təsir səviyyəsi var (0-dan 3-ə qədər), bu göstəricilərin hər birinin hesabın saxta olması ilə bağlı nəticəyə nə qədər təsir etdiyini göstərir. Təsir səviyyələri aşağıdakı mənalara ifadə edir:

0– Təsir yoxdur;

1-Zəif təsir.

2-Əhəmiyyətli Təsir.

3-Kritik Təsir.

Məsələn, histqram sütunu 3-cü səviyyəyə çatırsa, bu o deməkdir ki, bu sütunu xarakterizə edən parametr hesab saxtakarlığına kritik təsir göstərir. Əks halda, sütun 0 və ya 1 səviyyəsindədirsə, bu, parametrin real hesaba xas olduğunu bildirir. Beləliklə, parametrlərin hər birinin səviyyəsinə əsaslanaraq, müəyyən bir hesabın saxtılığı və ya reallığı barədə nəticə çıxarıyıq.

Təsnifatlayıcı uyğunluğu (X, y) öyrətmək üçün “sklearn” kitabxanasından $(X -$ cədvəl kimi giriş 2D verilənlər toplusu, $y -$ vektor kimi hədəflər) istifadə olunur. Bunun üçün əvvəlcədən təlimdən və 60:40 nisbətində paylanmış doğrulama çoxluqndan ibarət verilənlər toplusu hazırlanır.

Cədvəl 3.7. Dataset.csv faylının strukturu

	A	B	C	D	E	F	G	H	I	J
1										
2	STATUS	P1	P2	P3	P4	P5	P6	P7	P8	P9
3	Real	1	3	3	3	0	2	0	3	0
4	Saxta	3	3	1	0	2	2	2	3	0
5	Saxta	3	3	3	3	0	2	3	3	0
6	Real	0	0	0	0	2	2	0	2	0
7	Real	2	0	2	0	2	2	0	0	0
8	Real	2	0	0	0	2	2	3	0	0
9	Saxta	3	3	0	3	3	0	3	3	1
10	Real	2	0	3	3	0	2	1	3	0
11	Real	0	0	3	0	1	2	2	2	0

Verilənlər toplusu Facebook-da real hesablar (P1-dən P9-a qədər sütunlar), eləcə də bu hesablara uyğun statuslar (STATUS sütunu) haqqında rəqəmsal məlumatlardan ibarətdir. Verilənlər toplusunda 100 mövcud hesab haqqında məlumat var.

Model təlimindən sonra, təlimin düzgünlüyünü və sonrakı nəticələrin etibarlılığını göstərən dəqiqlik parametri göstərilir

Proqnoz (X) metodu Facebook-dan gələn məlumatların işlənməsi üçün cavabdehdir.

“Facebook” sosial mediasında saxta hesabların aşkarlanması sistemi rahat və aydın qrafik istifadəçi interfeysinə malik olan, vahid hesabın parametr təsir histogramını və onun statusunu göstərən Python proqramlaşdırma dili ilə həyata keçirilir.

Qrafik istifadəçi interfeysi aşağıdakı elementlərdən ibarətdir:

- Sosial mediada istifadəçi identifikatorunun daxil edilməsi üçün “User ID” mətn sahəsi;
- Yoxlanılan hesabın statusunun göstərildiyi "Status" etiketi;
- Sosial medianın ölçülərinin və onların hesab statusuna təsirinin göstərildiyi histogram.

Proqram təminatı aşağıdakı parametrləri təhlil edir: dostların sayı, dostların sayı, avatarda mövcud olan fotosəkil, istifadəçinin fotosəkilləri, mövcud örtük şəkli, postların mövcudluğu, doğum tarixi, istifadəçi haqqında şəxsi məlumatlar, səhifənin yenilənməsi və istifadəçi adı. Beləliklə, təhlili bu parametrlər hesabların statusları haqqında nəticə çıxarmağa imkan verir.

Hazırlanmış proqram alətinin effektivliyi SVM tərəfindən saxta hesabların aşkarlanmasının dəqiqliyi ilə qiymətləndirilir. Proqram Facebook-dan məlumatları topladı və daha sonra toplama təcrübə üçün hazırlanır. 100 ümumi hesabdan 10-nun təhlilinin nəticələri aşağıdakı cədvəldə göstərilmişdir.

Cədvəl 3.8 İstifadəçi hesablarının analiz nəticəsi

İstifadəçi	Hesab statusu	Proqnozlaşdırılmış status	İstifadəçi	Hesab status	Proqnozlaşdırılmış status
Dadaşova Leyla	Real	Real	Rüstəmli Kənan	Saxta	Saxta
Axundzadə Fuad	Saxta	Saxta	Zeynalov Ümid	Real	Real
Narmina Məmməd	Real	Real	Nəsirli Əli	Real	Real
Yusifova Ülkər	Saxta	Saxta	İbrahimli Qoşqar	Saxta	Saxta
Ayazlı Sevinc	Real	Real	Heydərova Aitan	Real	Real

Təhlildən sonra düzgünlüyün yoxlanılması həyata keçirilir.

$$\frac{n-t}{n} * 100\% = \frac{100-3}{100} * 100\% = 97\%$$

Burada , t - hesabın vəziyyəti ilə proqramın nəticəsi arasındakı uyğunsuzluqların sayı;
n - yoxlanılan hesabların sayı.

Facebook-da saxta hesabları aşkarlayan sistemin düzgünlüyünü yoxlayan 3100 hesabın statusu səhv, 100 hesabdan 97-si düzgün aşkarlanıb.

Sosial mediada hesabların strukturu təhlil edilir və onlarda olan istifadəçi haqqında məlumatlar işıqlandırılır. Facebook-un əsas ölçüləri nəzərdən keçirilir və təhlil edilir.

Bu göstəricilərə əsasən saxta hesabı müəyyən etmək mümkündür. Aşağıdakı ölçülər təhlil edilir: bəyənmələr, dostlar, yazılar və statuslar, istifadəçi haqqında şəxsi məlumatlar və fotosəkillər, onların mümkün parametrləri və hesabın statusuna təsiri nəzərə alınmaqla. Hər bir metrik təhlilin rahatlığı üçün müvafiq kateqoriyalara təyin edilir.

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ
AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

Axundzadə Fuad Babək oğlu

**DATA MİNING YOLU İLƏ SOSIAL ŞƏBƏKƏLƏRDƏ SAXTA
HESABLARLIN AŞKARLANMASI**

**mövzusunda
MAGİSTRİK DİSSERTASİYASI**

İxtisas: 060632 - "İnformasiya tenologiyaları və sistemləri mühəndisliyi"

İxtisaslaşma: "İdarəetmədə informasiya sistemləri"

Elmi rəhbər:

t.e.n., dosent Z.Ə. Cəfərov

BAKİ – 2023

IV FƏSİL. SOSIAL MEDIADA SAXTA HESABLARIN AŞKARLANMASI

METODLARI

4.1. Əsas komponentlərin analizi

Saxta hesabların aşkarlanmasının vacibliyindən ilhamlanan tədqiqatçılar bu yaxınlarda saxta hesabların aşkarlanmasının effektiv mexanizmlərini araşdırmağa başlayıblar. Əksər aşkarlama mexanizmləri istifadəçi səviyyəsində fəaliyyətləri və ya qrafik səviyyəli strukturları təhlil edərək istifadəçi hesablarını real və ya saxta (zərərli, Sybil) kimi proqnozlaşdırmağa və təsnif etməyə çalışır. Aşağıdakı yarımbölmələrdə təsvir edilən saxta hesabları aşkar etməyə kömək edən bir neçə məlumat mədən metodologiyası və yanaşmalar mövcuddur:

1. Xüsusiyyət əsaslı aşkarlama

Bu yanaşma istifadəçi səviyyəsində fəaliyyətlərə və onun hesab təfərrüatlarına (istifadəçi qeydləri və profillər) əsaslanır. Unikal xüsusiyyətlər son istifadəçi fəaliyyətlərindən çıxarılır (məsələn, dostluq sorğularının tezliyi, qəbul edilən sorğuların bir hissəsi), sonra bu xüsusiyyətlər maşın öyrənmə üsullarından istifadə etməklə oflayn təlim keçmiş təsnifatlandırıcıya tətbiq edilir

METIS klasterləşdirmə alqoritmini həm sessiya, həm də klik funksiyaları ilə istifadə etmək, məsələn:

- Seans başına orta kliklər
- Orta seans uzunluğu
- Dost sorğuları göndərmək üçün istifadə edilən kliklərin faizi
- Fotoların ziyarəti
- Məzmunların paylaşımı

Müəlliflər məlumatları 3% yanlış müsbət nisbət və 1% yanlış mənfi nisbətlə təsnif edə bildilər

Müəlliflər saxta hesabları aşkar etmək üçün SVM klassifikatorunu öyrətmək üçün RenRen tərəfindən verilən əsas həqiqətdən istifadə etdilər.

Sadə funksiyalardan istifadə etməklə, məsələn:

- dostluq sorğularının tezliyi
- qəbul edilən sorğuların bir hissəsi

Müəlliflər 99% həqiqi müsbət nisbətə (TPR) və 0,7% yanlış müsbət nisbətə (FPR) malik bir təsnifat hazırlaya bildilər.

Tədqiqatçılar Twitter tərəfindən verilən əsas həqiqətdən istifadə etdilər; Məlumatlar iki əsas yanaşmadan istifadə etməklə işlənmişdir:

- Vahid təsnifat qaydaları
- Spamerlərin aşkarlanması üçün ədəbiyyatda təklif olunan xüsusiyyətlər çoxluqları

Stateofsearch.com qayda çoxluq və Socialbakers qayda çoxluq kimi bəzi xüsusiyyətlər əvvəlki işlərdən istifadə edilmişdir. Müəlliflər orijinal təlim çoxluğunun hesablarının 95%-dən çoxunu düzgün təsnif edə bilmişlər

2. Xüsusiyyətin azaldılması

Yüksək ölçülü verilənlər bir çox təsnifat alqoritmləri üçün yüksək hesablama dəyərində və yaddaşdan istifadəyə görə ciddi problem ola bilər. Digər tərəfdən, ölçü sahəsinin azaldılması səs-küylü (yəni əhəmiyyətsiz) və lazımsız xüsusiyyətləri aradan qaldıracaq və daha yaxşı təsnifat modelinə və sadə vizuallaşdırma texnikasına gətirib çıxaracaq. Xüsusiyyətlərin azaldılması üsullarını iki növə bölmək olar:

Yüksək ölçülü məkanda verilənlərin daha az ölçülü fəzaya çevrildiyi ölçülərin kiçilməsi

- Orijinal xüsusiyyətlər çoxluğunun daha sadə və daha sürətli modellər yaratmaq, modellərin performansını artırmaq və verilənləri daha yaxşı başa düşmək üçün funksiyaların seçilmiş alt çoxluqlarına bölündüyü funksiya alt çoxluq seçimi.

Əsas Komponent Analizi (PCA): . PCA üstünlük təşkil edən normal istifadəçi davranışını ən yaxşı izah edən xüsusiyyətləri (ölçüləri) müəyyən etmək üçün istifadə edilən bir texnikadır.

PCA yüksək ölçülü məlumatları verilənlərdə mümkün qədər çox dəyişkənliyi nəzərə alan üst N əsas komponentlərin aşağı ölçülü alt fəzasına (normal alt fəza adlanır) layihələndirir

Müəlliflər normal istifadəçi davranışının aşağı ölçülü olduğunu nümayiş etdirmək üçün üç sosial şəbəkənin real məlumatlarından istifadə etdilər (14K Facebook istifadəçisi, 92K Yelp istifadəçisi, 100K Twitter istifadəçisi)

Onlar həmçinin saxta, güzəştə gedən və sövdələşmiş istifadəçilər tərəfindən nümayiş etdirilən anomal davranışın geniş əsaslı həqiqət məlumatlarından istifadə edərək anomaliyaların aşkarlanması texnikasını qiymətləndirdilər. Bu yanaşma 0,3%-dən az yanlış pozitivlərlə 66%-dən çox aşkarlama dərəcəsinə nail olur (natamam davranışın 94%-dən çoxunu əhatə edir).

3.Neyron şəbəkəsi (NN) və Dəstək vektor maşını (SVM)

Müəlliflər PCA-dan istifadə edərək profil xüsusiyyətlərini çıxardılar və sonra qanuni profilləri aşkar etmək üçün Neyron Şəbəkələri və Dəstək Vektor maşınlarını tətbiq edirlər. PCA nəticələrinin alınmasının riyazi yolu kimi “Variansın maksimumlaşdırılması” seçilmişdir, bunlar:

3. Dillərin sayı.
4. Profil Xülasəsi
5. Əlaqələrin sayı
6. Bacarıqların sayı
7. LinkedIn Qruplarının sayı
8. Nəşrlərin sayı

Müəlliflər ləpə funksiyaları kimi elastik arxa dayaq aqasiyası (Rprop) və Cdəstək vektor təsnifatı və polinom nüvəsi (polydot) ilə SVM ilə Neyron Şəbəkələri tətbiq etdilər.

Bu sənədin tapıntıları göstərdi ki, PCA-nın ölçülərin azaldılması kimi istifadə edilməsi heç bir seçim etmədən bütün xüsusiyyətləri istifadə etməkdən daha yaxşı dəqiqlik nəticələri verir.

Xüsusiyyətlərə əsaslanan aşkarlama böyük OSN-lərə çatsa da, onu keçmək hələ də nisbətən asandır. Tədqiqatçılar aşkar ediblər ki, təcavüzkarlar öz saxtakarlıqlarının məzmununu və fəaliyyət modellərini mənfi şəkildə dəyişdirərək spam aşkarlama üsullarından yayınmaq üçün davranışlarını dəyişirdilər . Xüsusiyyətlərə əsaslanan aşkarlama heç bir rəsmi təhlükəsizlik zəmanəti vermir və çox vaxt praktikada yüksək yanlış müsbət nisbətə nəticələnir.

İlkin verilən məlumat çoxluqları

Bu araşdırmada araşdırmamızı həyata keçirmək üçün “MIB” verilənlər bazasından istifadə etdik, o, aşağıdakı kimi twitterdən toplanmış üç verilənlər çoxluğundan ibarətdir

Saxta izləyicilərin ilkin verilən çoxluqları

2013-cü ilin aprel ayında tədqiqatçılar aşağıdakı onlayn bazarlardan təxminən 3351 saxta Twitter hesabı satın aldılar: <http://fastfollowerz.com/> və <https://intertwitter.com/> saytıdan hesablar, 19 dollara, müvafiq olaraq \$14. Onlar iki verilənlər bazası yaratdılar:Fastfollowerz verilənlər bazası və Intertwitter verilənlər bazası.

Fastfollowerz verilənlər bazasında FSF etikətli 1169 saxta hesab var. İntertwitter verilənlər bazası 1337 saxta hesabdan ibarətdir.

Tədqiqatçılar etiraf etdilər ki, onların saxta izləyiciləri verilənlər bazası bütün mövcud saxta izləyicilər toplusunun yalnız illüstrativdir və hərtərəfli deyil. Müəllifləri Twitter spamerlərinin istifadə etdiyi yayınma taktikalarının təhlilini aparmışdılar. Onlar müşahidə etdilər ki, twitter spamçıları spam aşkarlama üsullarından yayınmaq üçün

davranışlarını dəyişdirir, ona görə də onlar spamgöndərənlərin aşkarlanmasını təkmilləşdirəcək və onların bundan yayınması daha çətin olacaq yeni funksiyalar dizayn etməyi təklif etdilər. Onlar yeni xüsusiyyətlərinidörd maşın öyrənmə təsnifatında birləşdirdilər və onların tətbiqini digər mövcud yanaşmalarla müqayisə etdilər. Aşağıdakı cədvəldə spam göndərənlərin aşkar edilməsində yüksək töhfə verən funksiyalardan səkkizi ətraflı təsvir edilmişdir.

Təklif olunan alqoritm

Bu bölmə saxta twitter hesablarını proqnozlaşdırmaq üçün təklif olunan üsulları təqdim edir. Təklif olunan üsullar iki əsas hissəyə bölünür: xüsusiyyətlərin azaldılması və ağlabatan müddətdə yüksək təsnifat dəqiqliyinə nail olan yeni texnikanın işlənilib hazırlanması məqsədi daşıyan məlumatların təsnifatı:

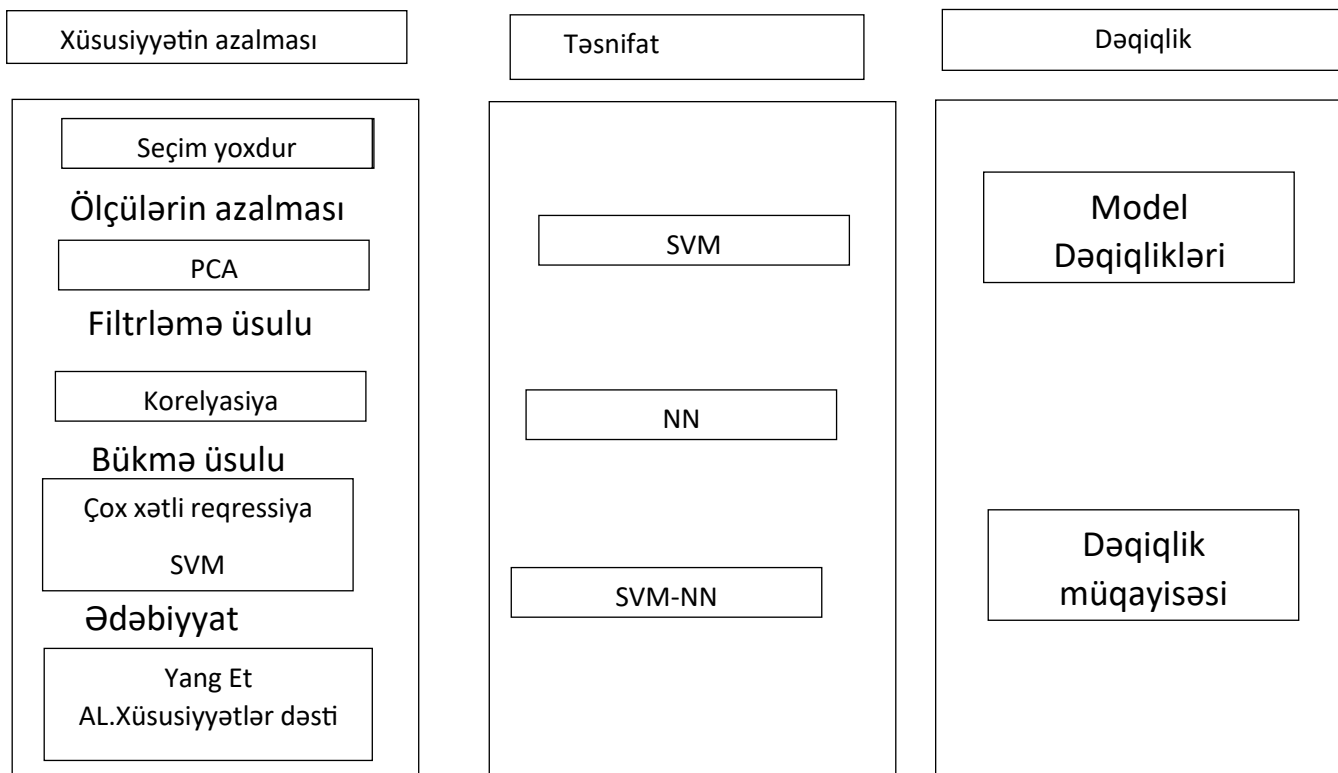
Məlumatların Əvvəlcədən Emalı

“MIB” verilənlər bazası xüsusiyyət vektorları iki növdə təqdim olunur:

- Kateqorik xüsusiyyətlər, məsələn, dil, profilin yan rəngi, tvitlər.
- Rəqəmsal xüsusiyyətlər, məsələn, dostların sayı, izləyicilərin defolt profili, profildən istifadənin fonu.

Cədvəl 4.1 Spam göndərənlərin aşkar edilməsi.

1	Hesab yaşı
2	İki istiqamətli əlaqə əmsalı
3	Qonşuların orta abunəçiləri
4	Qonşuların orta tvitləri
5	Orta qonşuların abunəçiləri abunəçilər
6	İzləmə dərəcəsi
7	API nisbəti
8	API URL nisbəti



Şəkil 4.1 Data mining üsulu ilə dəqiqlik dərəcəsini hesablamaq üçün dizayn yanaşma

Real və saxta hesabları ayırd etmək üçün xüsusiyyət etiketi əlavə edilib.

İlkin emal addımı 2-ci cədvəldə göstərilədiyi kimi Twitter-də istifadəçilərin davranışlarını təsvir edən 16 ədədi xüsusiyyət vektoru ilə nəticələndi.

Cədvəl 4.3 İstifadəçilərin davranışının təsviri

1	<i>Statusların sayı</i>	9	<i>Standar-profil-başlığı-urlesi</i>
2	<i>İzləyicilərin say</i>	10	<i>Profil-istifadə-arxafon-şəkli</i>
3	<i>Dostların say</i>	11	<i>Profil-mətn-rəng</i>
4	<i>Sevimlilər – sayı</i>	12	<i>Profil-yan panel-sərhəd-rəng</i>
5	<i>Siyahıda sayı</i>	13	<i>Profil-arxa plan-Qutusu</i>
6	<i>Geo-aktiv</i>	14	<i>Profil-yan panel-rəng-doldurma</i>
7	<i>Defolt profil</i>	15	<i>Profil-arxa-plan rəngi</i>
8	<i>Defolt profil şəkli</i>	16	<i>Profil-bağlantı-rəngi</i>

Xüsusiyyətin Azaldılması

Xüsusiyyətlərin azaldılması mərhələsində mining prosesində istifadə ediləcək ən perspektivli xüsusiyyət nümunələrinə qərar vermək prosesinə rəhbərlik etmək üçün dörd məlumat azaldılması texnologiyası tətbiq edilmişdir.

- PCA
- SpearmansRank-Sifariş Korrelyasiyası
- SVMdən istifadə edərək Qapaq Xüsusiyyətinin Seçimi
- Çox Xətti Reqressiya

Əsas Komponent Analizi "PCA". Bölmə 4.1-də qeyd etdiyimiz kimi, PCA xüsusiyyət vektor ölçülərini azaltmaq üçün istifadə edilən ölçü kiçilmə texnikasıdır, o, məlumatları və məlumatları ən yaxşı təsvir edən ən yaxşı N xüsusiyyətləri tapır.

Mining prosesinə təsir göstərməməsi üçün daha az çəki təyin edərək lazımsız xüsusiyyətləri aradan qaldıraraq, onun çox fərqliliyini əhatə edir. Bu işdə 16 PCA- dan 10 PCA seçilmişdir, bunlar məlumatların təxminən 92%-ni əhatə edir Spearmanın rütbə-sifariş korrelyasiyası. Spearman's RankOrder Correlation, xüsusiyyət seçimi filtrləmə üsullarından biridir. O, iki kəmiyyət dəyişəni X və Y arasındakı monoton əlaqənin gücünü və istiqamətini ölçür. Bu korrelyasiya ölçülərinin hər biri X və Y müstəqil olduqda tam olaraq sıfırdır və səviyyəni göstərmək üçün -1 ilə +1 arasında dəyişən dəyərlərə malikdir.

Münasiblik və artıqlığın təhlili texnikası.

Xüsusiyyət seçimi üçün uyğunluq və artıqlıq təhlili texnikasından istifadə olunur. Münasiblik təhlili addımında sinif dəyişəni ilə korrelyasiya səviyyəsi bizim təxmin edilən "0.8" həddimizdən aşağı olan bütün xüsusiyyət cütlərini aradan qaldırmaq üçün Spearmanın Rank-Order korrelyasiyasından istifadə edilmişdir.

Münasiblik təhlili addımında seçilmiş xüsusiyyətlər ehtiyat təhlili addımına giriş kimi istifadə olunur. İki xüsusiyyətin bir-biri üçün lazımsız olduğu, əgər onların dəyərləri tamamilə korrelyasiya olunursa, qəbul edilir, lakin əslində bir xüsusiyyət bir sıra

xüsusiyyətlər ilə əlaqələndirildikdə funksiyaların artıqlığını müəyyən etmək o qədər də asan deyil. Beləliklə, lazımsız xüsusiyyətləri aradan qaldırmaq üçün Markov blanket texnikasından istifadə etdik.

Cədvəl 4.4 Markov blanket texnikası 1

#	xüsusiyyət 1	xüsusiyyət 2
Çoxluq 1	2	1
Çoxluq 2	4	1
Çoxluq 3	10	7
Çoxluq 4	14	11
Çoxluq 5	16	12
Çoxluq 6	14	12
Çoxluq 7	15	12
Çoxluq 8	16	12
Çoxluq 9	15	14
Çoxluq10	16	14
Çoxluq11	16	15

ƏLAQƏLİ XÜSUSİYYƏTLƏRİN CÜTLƏRİ. KOMOFİSİENT >0,8

Markov yorğan texnikası. Bayes şəbəkəsindəki A, MB(A) node üçün Markov örtüyü A-nın valideynlərindən, uşaqlarından və uşaqlarının digər valideynlərindən ibarət qovşaqlar çoxluqdur . Markov təsadüfi sahəsində bir düyünün Markov örtüyü onun qonşu qovşaqlar çoxluqdur.

Cədvəl 4.5 Markov Blanket texnikası 2

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16
Çoxluq1	1	0	0	0	0	0	1	0	0	0	1	1	0	1	1	0
Çoxluq2																

İki seçilmiş xüsusiyyət çoxluq, burada töhfə verən funksiyalar 1 ilə təmsil olunur və nəzərə alınmayan xüsusiyyətlər 0 ümumi 16 xüsusiyyət ilə təmsil olunur.

SVM istifadə edərək Sargı Xüsusiyyətinin Seçimi. Tanınmış xüsusiyyət seçmə üsullarından biri müxtəlif xüsusiyyət alt çoxluqlarının öyrənmə modeli ilə seçildiyi və kvalifikasiya edildiyi sargı metodlarıdır. Ən yüksək proqnozlaşdırılan performans malik funksiyalar alt çoxluq seçiləcək. Çoxluğun bütün alt çoxluqları bit manipulyasiyasından istifadə etməklə tapıla bilər, verilmiş çoxluq üçün 2 alt çoxluq olacaq , burada n çoxluqda F xüsusiyyətlərin sayıdır. Çoxluğun bütün alt çoxluqları bit manipulyasiyasından istifadə etməklə tapıla bilər, verilmiş çoxluq üçün 2 alt çoxluq olacaq , burada n çoxluqda F xüsusiyyətlərin sayıdır. Məsələn, Cədvəldə göstərildiyi kimi {1, 2, alt 3} çoxluğu olacaq .

Cədvəl 4.6 Çoxluq

F1	F2
0	0
0	0
0	1
0	1
1	0
1	0
1	1



Xüsusiyyət alt çoxluğu
{0}
{3}
{2}
{2,3}
{1}
{1,3}
{1,2}

Bit manipulyasiyasından istifadə etməklə təqdim edilən $\{1,2,3\}$ çoxluq üçün bütün mövcud alt çoxluqlar , binar təsvirdə hər 1 həmin mövqedəki elementi göstərir.

Bu üsul adətən həmin xüsusi öyrənmə modeli üçün ən yaxşı performans xüsusiyyətlərini təmin edir, lakin böyük funksiya sahəsi üçün intensiv hesablama tələbləri tələb oluna bilər. Əsas məlumat çoxluqmızda 16 xüsusiyyət vektorumuzvar idi ki, bu da boş alt çoxluq olmadan $2^{16-1}=65,535$ xüsusiyyət alt çoxluğu deməkdir. Əsas məlumat çoxluq 70% təlim və 30% testə bölünüb. Sonra 65,535 xüsusiyyət alt çoxluqlarının hamısı Şəkil 4-də göstərildiyi kimi Dəstək vektor maşını “SVM” istifadə edərək öyrədilib və sınaqdan keçirilib.

Cədvəl 4.7 SVM ilə qarma texnikası istifadə edilən top 5 seçilmiş xüsusiyyətlərin alttoplumu

#	SET1	SET 2	SET 3	SET 4	SET 5
F1	0	0	0	0	0
F2	1	1	1	1	1
F3	1	1	1	1	1
F4	0	0	0	0	0
F5	1	1	1	0	1
F6	1	1	0	1	1
F7	1	1	1	1	1
F8	1	1	1	1	1
F9	1	1	1	1	1
F10	1	1	1	1	1
F11	1	1	1	1	1
F12	1	1	1	1	1
F13	1	0	1	1	1
F14	1	1	1	1	1
F15	1	1	1	1	1
F16	1	1	1	1	0

Sadə xətti reqressiyada iki dəyişən x və y var, burada y x -dən asılıdır və yatənlik 1-də göstəriləndiyi kimi x -dən təsirlənir.

$$y = a + bx \quad (1)$$

(burada, a sabitdir, b reqressiya əmsalındır)

$$y = a + bx_1 + cx_2 + dx_3 \quad (2)$$

Çox xətti reqressiyadan istifadə edilmişdir, çünki MIB verilənlər bazası 16 asılı dəyişənə və bir müstəqil dəyişənə malikdir. Çox xətti reqressiya problemi **multikollinearlıqdır**, burada model təkcə cavab dəyişəni ilə deyil, həm də bir-biri ilə

əlaqəli olan çoxsaylı amilləri ehtiva edir.

Multikollinearlıq əmsalların standart səhvlərini artırır ki, bu da bəzi dəyişənləri statistik cəhətdən əhəmiyyətsiz edir, əks halda onlar əhəmiyyətli olmalıdır. Lazımsız dəyişənlərin çıxarılması 16 proqnozlaşdırıcı əvəzinə 12 proqnozlaşdırıcı ilə nəticələndi.

Regressiya modelinin performansını iki atribut, R2 və P-dəyəri ilə ölçülə bilər.

R2-məlumatların uyğun regressiya xəttinə nə qədər yaxın olduğunu göstərən statistik ölçüdür.

Başqa sözlə, R kvadratı nə qədər yüksək olarsa, məlumat regressiya modeli ilə bir o qədər yaxşı uyğunlaşar. R2-nin 0,57%-dən çox olması müəyyən edilmişdir.

P-dəyəri: regressiya modelinin sıfır fərziyyəsini göstərir, p-dəyəri nə qədər kiçik olarsa, model verilənlərə bir o qədər yaxşı uyğun gəlir, 0,01-dən az olması təyin edilmişdir.

Bu addımın nəticəsi 106 uyğun alt qrup idi, ilk 5 xüsusiyyət alt çoxluğu Cədvəl 4-də verilmişdir.

4.3.Fəaliyyət və Qiymətləndirmə

Əvvəlcə üç fərqli təsnifat alqoritmi hazırlanmış və fərqli dörd xüsusiyyət çoxluqundan istifadə etməklə sınaqdan keçirilmişdir

Neyroşəbəkə təsnifat alqoritmi və SVM təsnifat alqoritmi bir çox sosial şəbəkə tədqiqatlarında mədən texnologiyalarının prinsipləri kimi istifadə edilmişdir, buna görə də onlar xüsusiyyətlər çoxluqlarında tətbiq edilmiş və təklif olunan SVN-NN alqoritmi ilə müqayisə edilmişdir.

Neyron şəbəkələri

Hal-hazırda, modelləri öyrətmək və əvvəllər öyrədilmiş modellər əsasında nəticələri proqnozlaşdırmaq üçün istifadə edilən bir çox neyron şəbəkəsi alqoritmləri mövcuddur. Əsas alqoritm kimi irəliyə doğru yayılma alqoritmi seçilmişdir. Proqnozlaşdırılan nəticələr faktiki qanuni dəyərlərlə (yəni hesabın real və ya saxta olması) müqayisə edilib və proqnozun düzgünlüyü aşağıdakı kimi hesablanıb:

$$\% \text{ Dəqiqlik} = \frac{\text{Bütün düzgün müəyyən edilmiş hesablar}}{\text{T hesabların ümumi sayı}} \times 100$$

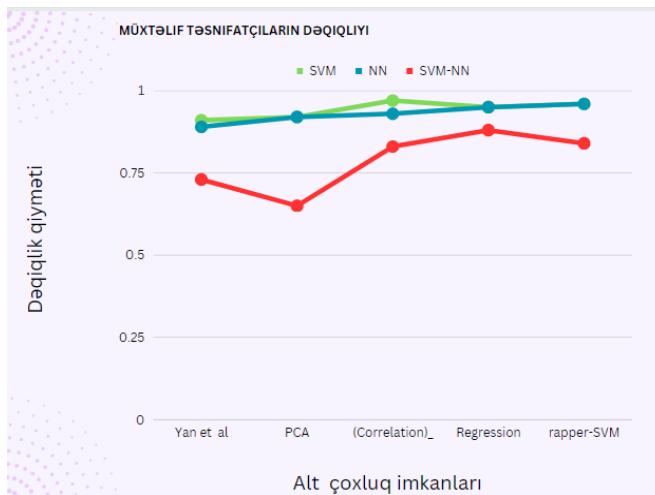
Yuxarıda qeyd edildiyi kimi ən yüksək dəqiqliyə malik xüsusiyyət alt çoxluqları aşağıdakı kimi vurğulanmışdır: spearman's rank-order Korrelyasiya ən yaxşı nümunəsi (1000001000110110), Çox xətti Reqressiya ən yaxşı nümunəsi (0110110111001111), Wrapper-SVM11 ən yaxşı nümunəsi (11011111011111). NN dəqiqlik nəticələri:

Cədvəl 4.8 Təklif olunan xüsusiyyətlər dəstində SVM, NN, SVM-NN tətbiqinin dəqiqlik nəticələri

Xüsusiyyətlər dəsti	SVM			NN			SVM-NN		
	Dəqiqlik	Yalnız müsbət	Yalnız neqativ	Dəqiqlik	Yalnız müsbət	Yalnız neqativ	Dəqiqlik	Yalnız müsbət	Yalnız neqativ
Yang et al.	0.886	0.111	0.001	0.737	0.059	0.203	0.912	0.086	0.001
PCA	0.914	0.039	0.046	0.653	0.278	0.067	0.922	0.033	0.043
Correlation	0.823	0.036	0.046	0.822	0.079	0.097	0.983	0.013	0.003
Regression	0.947	0.035	0.016	0.888	0.04	0.071	0.96	0.027	0.011
Wrapper-SVM	0.956	0.039	0.004	0.833	0.052	0.114	0.965	0.027	0.007

Cədvəldə göstərildiyi kimi, nəticələr göstərir ki, SVM təsnifatçısı Wrapper-SVM xüsusiyyət çoxluqndan istifadə edərkən ən yüksək dəqiqliyə malikdir və ən aşağı dəqiqlik Yang et al ilə olub. xüsusiyyət çoxluq. NN təsnifatçısı üçün dəqiqlik nəticələri SVM təsnifatını istifadə edən analoqlarından aşağı idi, reqressiya xüsusiyyətləri çoxluqndan ən yüksək dəqiqlik 0,888 və PCA xüsusiyyət çoxluqndan istifadə edərək ən aşağı dəqiqliklə

Hər üç təsnifat alqoritminin dəqiqlik nəticələrini müqayisə edərək aydın olduki, SVM-NN təsnifat alqoritmi Qrafik-də göstərildiyi kimi əvvəlki digər iki təsnifatçı ilə müqayisədə bütün xüsusiyyət alt çoxluqları üzrə ən yüksək təsnifat dəqiqliyi nəticələrinə malikdir, ən yüksək dəqiqliklə 0,983.



Şək. 4.2 Təsnifat alqoritminin dəqiqlik nəticələrinin müqayisəsi

Təhlillərimizin nəticələri göstərdi ki, “SVM-NN” digər iki təsnifatçı ilə müqayisədə bütün xüsusiyyət çoxluqları ilə daha yaxşı dəqiqlik nəticələrini arxivləşdirib, təsnifat dəqiqliyi 98% civarındadır. Qeyd olundu ki, NN alqoritmi SVM və SVM-NN ilə müqayisədə daha aşağı təsnifat dəqiqliyinə malikdir.

Bunun səbəbi, SVM alqoritmi optimallaşdırılmış funksiyanın qlobal minimumuna çatır NN isə gradient eniş texnikasından istifadə edir və SVM kimi “qlobal minimuma deyil” yerli minimuma çata bilər.

Həmçinin qeyd olundu ki, PCA tərəfindən təmin edilən xüsusiyyətlər çoxluqndən istifadə çox aşağı təsnifat dəqiqliyi ilə nəticələnir, korrelyasiya xüsusiyyətləri çoxluq isə yüksək təsnifat dəqiqliyi ilə nəticələnir. Bu, PCA-nın ölçüləri azaltması və orijinal xüsusiyyətlərin xətti birləşməsinə əsaslanaraq yeni xüsusiyyətlər bazası yaratması səbəbindən baş verdi. Lakin korrelyasiya və digər xüsusiyyət seçmə üsulları bütün xüsusiyyətlərin xətti birləşməsini deyil, orijinal xüsusiyyətlərin ən yaxşı çoxluqni seçir.

Başqa sözlə, xüsusiyyət seçimi ən effektiv orijinal xüsusiyyətləri seçin, lakin PCA xətti birləşməni edir orijinal xüsusiyyətləri hadisə onlar təsirli deyil Korrelyasiya xüsusiyyətləri çoxluq digər xüsusiyyət seçim texnikası çoxluqları arasında diqqətəlayiq dəqiqliyi qeyd edir, çünki korrelyasiya texnikası yalnız ən yaxşı xüsusiyyətləri seçmir, həm də artıqlığı aradan qaldırır.

Performans Metriklər:

Doğrulama prosesində ümumi ölçülər qrupu tətbiq oluna bilər. Bu işdə [19]-dan götürülmüş aşağıdakı ölçülərdən istifadə edilmişdir:

1.Dəqiqlik: Aşkarlama modelinin performansını ölçün $Dəqiqlik = (\text{düzgün proqnozlar}) / (\text{ümumi nümunələr})$.

2. Xatırla: həqiqi müsbət nisbət

$Xatırla = (\text{həqiqi müsbət proqnozlar}) / (\text{müsbət nümunələr})$,

3.Dəqiqlik: Müsbət predikasiyanın düzgün olması ehtimalını ölçün $Dəqiqlik = (\text{həqiqi müsbət proqnozlar}) / (\text{müsbət proqnozlar})$.

4.Spesifiklik: həqiqi mənfi dərəcələr $Spesifiklik = (\text{həqiqi mənfi proqnozlar}) / (\text{mənfi nümunələr})$.

k-NN qiymətləndiricisindən istifadə edərək çatışmayan dəyərləri idarə etdikdən sonra nəzarət edilən alqoritmlər 982 profilə tətbiq edildi.

Cədvəl 4.9 Nəzarət edilən alqoritmlərin nəticələrini göstərir.

Ölçülər	Dəqiqlik	Səhihlik	Geri çağırılma	Spesifiklik
İD3	0.9776	0.9872	0.9846	0.9502
SVM	0.9572	0.9780	0.9680	0.9154
K-NN ilə k-3	0.9145	0.9520	0.9398	0.8159

Yanaşmadan sonra nəzarət olunmayan alqoritmlər eyni verilənlər toplusundan istifadə etməklə tətbiq edilmişdir. Qeyd etmək vacibdir ki, təlim məlumatları etikətlənməmiş verilənlər toplusu idi və bu, qiymətləndirməni problemlə etdi, çünki modelin nəticələrinin mənalı şəkildə müqayisə oluna biləcəyi heç bir şey yox idi. Beləliklə, tətbiq olunan alqoritmin düzgünlüyünü qiymətləndirmək üçün sadə üsul yox idi. Klasterləşdirmə üsullarını qiymətləndirmək üçün biz çevik üsullarla istifadə oluna bilən RapidMiner-in xüsusi operatorlarından istifadə edərək qiymətləndirmə modeli yaratdıq. Məsələn, Map Clustering on Labels operatoru verilmiş klasterləri sinif etikətləri ilə tənzimləməklə klasterləşmə və proqnozlaşdırma proseslərini xəritələşdirir. Bu, məlumat çoxluğunu tənzimləməyə və modelimizi qiymətləndirməyə imkan verir.

Hər iki alqoritm ($k = 2$) və ya 2 klasterə (C_0, C_1) tətbiq edilmişdir, burada C_0 real profilləri, C_1 isə saxta profilləri təmsil edir. k -Means alqoritmi verilənlər toplusunu bölmək və 0,6731 dəqiqlik göstərdi, k -medoidlər isə Cədvəl 5-də göstərildiyi kimi 0,6701 dəqiqlik göstərdi.

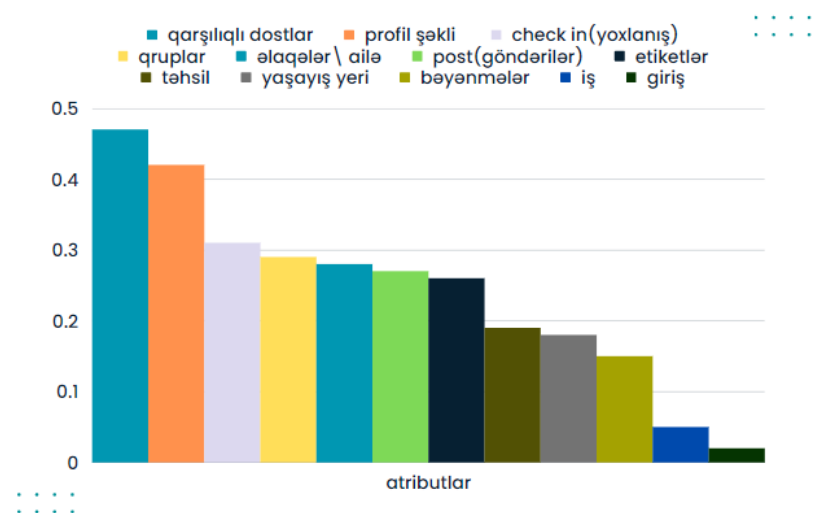
Təcrübələrdə istifadə üçün cəmi 906 profil toplanıb. Onlardan 125 profil əhəmiyyətsiz və ya dublikat olduğu üçün xaric edilib. Bu, 781 real profillə nəticələndi, bunlardan 19 profilin saxta olduğu aşkar edildi, buna görə də bunlar verilənlər bazasında saxta olaraq etiketləndi. Daha çox saxta profil toplamaq üçün daha 250 profil alınıb. Bununla belə, onlardan yalnız 182 profil hesab edilib, çünki bəzi profillər deaktiv və ya bloklanıb. Bunun nəticəsində 201 saxta profil yaranıb. Beləliklə, toplama prosesi nəticəsində ümumilikdə 982 profil (781 real və 201 saxta) əldə edilib. Bu cəmindən 86-nın (61 real və 25 saxta) bəzi atributlarında çatışmayandəyərlər var idi. Çatışmayan dəyərləri idarə etmək üçün məlumatların hesablanması üçün k -NN modeli istifadə edilmişdir. Nəhayət, təlim və sınaq məqsədləri üçün profilləri saxta və ya real olaraq etiketləmək üçün toplanmış məlumat çoxluğuna əl ilə etiketləmə tətbiq edildi.

Cədvəl 4.10 Nəzarət olunmayan alqoritmlərin nəticələri

Alqoritm	Gerçək vəziyyət		Real	Saxta	Klasterlər
	Təxmin edilən vəziyyətlər	Real Saxta	661	201	C_0
			120	0	C_1
	Dəqiqlik	Səhihlik	Xatırlamaq	Spesifiklik	
k-Əlaqələndirici	0.6731	0.7668	84.64%	0.0000	
k-Medoidlər	0.6701	0.8840	0.6735	0.6567	

Əvvəlki səhifədə göstərildiyi kimi, nəzarət edilən alqoritmlər nəzarətsiz alqoritmləri üstələyib. Ancaq bu nəticələri əsaslandırılmazdan əvvəl bəzi vacib məqamları qeyd etmək lazımdır. Birincisi, model qərar vermək üçün informativ atributlardan asılıdır. Bu atributlar Şəkil də təsvir edilmişdir. Göründüyü kimi,

“qarşılıqlı dostlar” atributu ən məlumatlandırıcı, “giriş” atributu isə ən az məlumatlandırıcıdır. İkincisi, qeyd edirikki, bəzi atributlar həm real, həm də saxta profillərdə eyni dəyərlərə malikdir. Məsələn, saxta profillərdə adətən sıfır etiket, sıfır yazı və yüksək bəyənmə aktivliyi olur.



Şək. 4.3 Atributların məlumat qazancı

Şəkildə iki sinifli etiketlərə (saxta və real) münasibətdə müdaxilə edilmiş atributlar üçün histoqram diaqramlarını təsvir edir. Beləliklə, müdaxilə edilmiş atributları düzgün idarə etməyə qadir olan alqoritm ən dəqiq qərarlar verəcəkdir.

Bununla belə, nəzarətsiz üsullar verilənlər toplusunu ayırmaq üçün sinif etiketləri olmayan bir vahid kimi bütün profillərlə məşğul olur. Bu, müdaxilənin aşkarlanmasını çətinləşdirir, çünki atributlar təkcə saxta və real profillərdə deyil, həm də eyni sinif profillərində etikətlənməlidir. Bu, aşkarlama modelini yayındırır və müdaxilə edilmiş atributları olan hər hansı profilə məhəl qoymadan oxşar atributları olan profilləri qruplar şəklində qruplaşdırır. Üstəlik, nəzarətsiz alqoritmlər aşağı dəqiqlik dərəcələrinə malik idi, çünki bu klasterləşdirmə üsulları verilənlər toplusunu vahid vahid kimi idarə edir və oxşar atributlara malik profilləri bir klasterdə qruplaşdırır. Bu səbəbdən, müdaxilə edilmiş atributlarla bağlı problemlər yaranırdı, burada bəzi informativ atributlar müxtəlif qruplara yığılmadı. Beləliklə, bu üsullar profilləri saxta və real olaraq düzgün qruplaşdırma bilmədi.

Nəzarət olunan alqoritmlər yüksək dəqiqlik dərəcəsinə malikdir, çünki onlar k-NN qiymətləndiricisindən istifadə edirlər və bu, çatışmayan dəyərlərin idarə edilməsində effektivliyini sübut etmişdir. K-NN qiymətləndiricisi k-ən çox oxşar profillərdən itkin dəyərlərə qiymətlər təyin etməklə məlumatları daxil edir. Bunu modelin bütün alqoritmlər üçün demək olar ki, eyni dəqiqliklə sabit performans nümayiş etdirdiyi təcrübədən görmək olar.

Qeyd edək ki, çatışmayan dəyərlərin əksəriyyəti ən yüksək müdaxilə faktoruna malik olan “qruplar” və “bəyənmələr” atributlarında olub. k-NN bu atributları hesablama prosesindən çıxarır ki, bu da dəqiqlik dərəcəsinə müsbət təsir göstərir.

Bu nümunə data-mining üsullarından istifadə edərək saxta Facebook profillərinin aşkar edilməsini nəzərdə tuturdu. 12 ayrı-seçkilik (davranış və qeyri-davranış) atributları ilə 5 nəzarət olunan və nəzarətsiz texnikadan istifadə edən bir model təklif edilmişdir. RapidMiner Studio 8.0.1 982 profilli (781 real və 201 saxta) verilənlər bazası əsasında modelin düzgünlüyünü qiymətləndirmək və eksperiment aparmaq üçün istifadə edilmişdir. Nəzarət edilən alqoritmlər nəzarətsiz alqoritmləri üstələdi və bütün təcrübələrdə yüksək və perspektivli dəqiqlik dərəcələri göstərdi. Daha dəqiq desək, ID3 qərar ağacı bütün alqoritmlər arasında ən yüksək dəqiqliyi nümayiş etdirdi və bütün nəzarət olunmayan alqoritmlər nisbətən oxşar aşağı dəqiqlik göstərdi. Bu yazının sonunda bu nəticələrin dərin izahı verilmişdir.

İSTİFADƏ EDİLMİŞ ƏDƏBİYYAT

1. Anomaly Detection Using Data Mining Techniques in Social Networking (SIRIPURI KIRAN)
2. Smith, AN, Fischer, E. & Yongjian, C., Brendlə əlaqəli İstifadəçi tərəfindən yaradılan Məzmun YouTube, Facebook və Twitter arasında necə fərqlənir?, İnteraktiv Marketing jurnalı, 26(2), səh. 102-113 , 2012.
3. Eugene, A., Castillo, C., Donato, D., Gionis, A. & Mishne, G., Finding High Quality Content in Social Media, 2008-ci il Veb Axtarış və Məlumat Mədəni üzrə Beynəlxalq Konfransın materiallarında , ACM, s. 183-194, 2008
O'Keefe, Schurgin, G. & Pearson, KC, The Impact on Social Media on Children, Adolescents, and Families, *Pediatrics*. 127(4), səh.800-804, 2011.
4. Voitovych, OP, Holovenko, VO (2016). Sosial şəbəkələrin müharibədə məlumat mənbəyi kimi tədqiqi. J. Rysiński (red.), *Inżynier XXI wieku projectujemy przyszłość* (səh. 111-119).
5. Mohammadrezaei, M., Shiri, ME, Rahmani, AM (2018). Qrafik Analiz və Təsnifat Alqoritmləri əsasında Sosial Şəbəkələrdə Saxta Hesabların Müəyyən edilməsi. Təhlükəsizlik və Kommunikasiya Şəbəkələri, 2018, 1- 8.
6. Lakhina, M. Crovella və C. Diot, “Diagnosing network-wide trafik anomaliyaları”, *ACM SIGCOMM Computer Communication Review*, cild. 34, yox. 4. ACM, 2004, səh. 219– 230.
7. Karel, H., Templ, M. & Filzmoser, P., Klassik və Güclü Metodlardan istifadə etməklə Kompozisiya Məlumatları üçün Eksik Dəyərlərin Təsdiqlənməsi , *Hesablama Statistika və Məlumat Təhlili*, 54(12), s.3095-3107, 2010.
8. Kontaxis, Georgios, Polakis, I., Ioannidis, S. & Markatos, EP, Detecting Social Network Profile Cloning, In *Yaygın Hesablama və Kommunikasiya Seminarlarında (PERCOM Workshops)*, 2011 IEEE Beynəlxalq Konfransı, səh. 295-300. IEEE, 2011.

NƏTİCƏ

Nəticə olaraq, sosial media platformaları öz platformalarında saxta hesabların müəyyən edilməsində və onlara ünvanlanmasında mühüm rol oynayır. Süni intellekt, maşın öyrənmə alqoritmləri, istifadəçi davranışı təhlili, hesab yoxlaması, məzmun və keçid təhlili, şəbəkə təhlili, hesabat mexanizmləri və əməkdaşlığın birləşməsindən istifadə etməklə bu platformalar istifadəçiləri üçün təhlükəsiz və etibarlı mühit yaratmağa çalışırlar.

Saxta hesabların aşkarlanması davamlı problemdir, çünki zərərli aktyorlar daim öz taktikalarını uyğunlaşdırırlar. Sosial media platformaları davamlı olaraq öz texnikalarını təkmilləşdirir və xarici təşkilatlar və tədqiqatçılarla əməkdaşlıq yolu ilə yaranan təhlükələrdən xəbərdar olur. Məqsəd istifadəçi məxfiliyini qorumaq, dezinformasiya ilə mübarizə aparmaq və onlayn müzakirələrin bütövlüyünü qorumaq üçün saxta hesabları operativ şəkildə müəyyən etmək və silməkdir.

Sosial media platformaları saxta hesablarla mübarizəyə ciddi səylər sərf etsə də, istifadəçi məlumatlılığı və iştirakı eyni dərəcədə vacibdir. İstifadəçilərə şübhəli hesablar barədə məlumat vermələri və məsuliyyətli onlayn davranışlarla məşğul olmaları tövsiyə olunur.

Sosial mediada saxta hesabların müəyyən edilməsi mürəkkəb bir işdir və platformalar inkişaf edən təhdidlərdən qabaq qalmaq üçün aşkarlama üsullarını təkmilləşdirməyə sadıqdırlar. Sosial media platformaları əməkdaşlıq və proaktiv yanaşmanı təşviq etməklə istifadəçilərə təhlükəsiz və orijinal onlayn təcrübə təqdim etməyi hədəfləyir.