CERT üçün insident verilənlərinin intellektual analizi

Yadigar İmamverdiyev[1], Elşən Bağırov[2]
[1,2]AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
[1]yadigar@lan.ab.az, [2]elsenbagirov1995@gmail.com

***Xülasə*— Bu işdə insidentlərin emal mərhələləri və bilet açma mərhələsi analiz edilmiş, real insident biletləri üzərində verilənlərin intellektual analizi metodları tətbiq edilmişdir. Eksperimentlər IBM-in insident biletləri bazası üzərində aparılmışdır.**

Açar sözlər—bilet analizi, insident analizi, CERT komandaları, təsadüfi meşə, bilet açma

I. GİRİŞ

İnformasiya texnologiyalarının sürətli inkişafı və geniş tətbiqi yeni növ təhdidlərin meydana çıxmasına səbəb olur. İnformasiya təhlükəsizliyi insidentlərinin vaxtında və effektiv şəkildə cavablandırılması təşkilatlar üçün olduqca mürəkkəb və kompleks bir problemdir [1]. Böyük təşkilatlar kiber-hücumlara tez-tez məruz qalır və onların nəticəsi olaraq, belə insidentlərlə tez-tez qarşılaşır. Odur ki, bu təşkilatlar haqlı olaraq insidentlərlə mübarizə məqsədilə təşkilat daxilində CERT- komandaları (Computer Emergency Response Team, Kompüter İnsidentlərini Cavablandırma Komandaları) yaradırlar. İnsident baş verən anda bu komandalar bir neçə emal prosedurunu ardıcıl olaraq yerinə yetirməklə təşkilatın informasiya aktivlərini risk altında qalmaqdan xilas etmiş olur.

CERT komandalarının qarşısında insidentlərlə mübarizədə böyük məsuliyyət durur. Hücumlar nəticəsində biznes verilənləri tez-tez risk altında qalır. Həmin hücumlara cəld və effektiv cavab vermək bəzən çətin ola bilər. Təşkilatın tələblərinə görə bu komandalar bir neçə seqmentdə insident idarəetmə funksiyaları və fəaliyyətlərini yerinə yetirirlər [2].

Sistemin avtomatik olaraq monitorinq edilməsi zamanı yalançı siqnalların olması qaçılmazdır. Bəzi təşkilatlar gündəlik biznes fəaliyyətləri zamanı baş verən insidentlərin həllində koordinasiya, problem aradan qaldırma (troubleshooting), yalançı siqnalların aşkarlanması, insidentlərin qeydiyyata alınmasında bir vasitə kimi İnsident Bilet Sistemindən (İBS) istifadə edirlər. Yaxşı strukturlaşdırılmış İBS insidentlərin idarəetmə prosesinə müsbət təsir göstərə bilər [3]. Buna baxmayaraq, bir çox hallarda lazımsız insident biletləri də ola bilər. Hər bir insident üçün ayrıca biletin açılması vaxt və səmərəlilik baxımından əhəmiyyətsiz ola bilər. Buna görə də böyük həcmdə insident biletləri üzərində verilənlərin intellektual anlizi metodlarının tətbiq edilməsi zərurəti yaranır.

Tədqiqat aparılan bazada siniflərin sayı məlum olduğundan bir neçə metod vasitəsilə insident biletlərinin klassifikasiyası aparılmış, tətbiq edilən metodların bir neçə kriteriyalar üzrə qiymətləndirilməsi aparılmışdır.

II. ƏLAQƏDAR TƏDQİQATLAR

Sistemin avtomatik monitorinq edilməsi üsulu müəssisələrdə problemlərin aşkarlanması üçün olduqca etibarlı və effektivdir. Monitorinq sisteminin iş qabiliyyəti onun sistem administratorları tərəfindən necə konfiqurasiya edilməsindən asılıdır. Yanlış konfiqurasiya nəticəsində yalançı siqnallar (false positives, false negatives) yaranır ki, bunun da nəticəsində sistemə ciddi ziyan gələ bilər [4].

L. Tang və b. insan tərəfindən yaradılmış insident biletlərindən istifadə edərək yalançı siqnalları aşkarlamaq qabiliyyətinə malik olan avtomatik bir yanaşma təklif etmişlər [4]. Bu yanaşma insident biletlərinin təsvir hissəsini analiz etmək üçün mətnlərin klassifikasiyası modelinin tətbiq edilməsindən ibarətdir. Bu isə nəticə etibarilə sistem administratorlarına səhv konfiqurasiyaları düzəltməyə imkan verməklə bərabər, gələcəkdə baş verə biləcək yalançı siqnalları minimuma endirmiş olur [4].

Y. Li və T. Li tərəfindən İT xidmət infrastrukturunda yaradılmış insident biletlərinin bir sinfi üçün orta sərfiyyat vaxtının hesablanması üçün iki mərhələli yanaşma təklif edilmişdir [5]. Birinci mərhələdə metaverilənlər modeli əsasında hər bir bilet üçün sərfiyyat vaxtı hesablanmış, emal üçün prioritetlik qaydaları verilmişdir. İkinci mərhələdə isə maksimum həqiqətəoxşarlıq metodundan istifadə edərək insident biletləri sinfi üçün orta sərfiyyat vaxtı qiymətləndirilmişdir [5].

III. İNSİDENTLƏRİN EMAL EDİLMƏ MƏRHƏLƏLƏRİ VƏ PROSESLƏRİ

CERT komanda heyyətinin əsas missiyası təhlükəsizlik insidentlərinin kənarlaşdırılması, analizi, identifikasiyası və qarşısının alınması üçün plan, prosedur və siyasətin hazırlanması və buna mərhələlər şəklində riayət edilməsidir.

İnsidentlərin cavablandırılması üçün bir neçə mərhələlər ardıcıl şəkildə yerinə yetirilir. ISO/IEC 27035 İnformasiya təhlükəsizliyi insidentlərinin idarə edilməsi üzrə beynəlxalq standartında bu mərhələlər aşağıdakı kimi verilmişdir [6]:

1. Planlaşdırma və hazırlıq;
2. Aşkarlama və hesabat;
3. Qiymətləndirmə və qərar qəbuletmə;
4. Cavablandırma;
5. Dərs çıxarma.

DOI: 10.25045/NCInfoSec.2017.30

A. Planlaşdırma və hazırlıq

Planlaşdırma və hazırlıq mərhələsi insidentlərin cavablandırılması prosesində ilk mərhələ və digər mərhələlərə nisbətən ən vacib hazırlıq mərhələsi hesab olunur. Bu mərhələdə insidentlərin emaletmə prosesinə mane ola biləcək istənilən potensial problemləri aradan qaldırmağa kömək edən bir neçə əsas elementlər hazır vəziyyətə gətirilir. Belə ki, insidentlərin emalında koordinasiya vasitələri (komanda üzvləri üçün əlaqə məlumatları, insidenti bildirmə mexanizmi və s.), insident analizi üçün aparat və proqram təminatları, insident analizi resursları və təşkilatın tələbinə uyğun olaraq digər ilkin hazırlıq işləri yerinə yetirilir [7]. Əsas hazırlıq işi insident idarəetmə planının qurulmasından ibarətdir.

B. Aşkarlama və hesabat

Bu mərhələ insident idarəetmə prosesinin ilkin fəaliyyət mərhələsidir. İlk olaraq informasiya təhlükəsizliyi hadisəsinin baş verməsi haqqında bildirişlər müxtəlif yollarla toplanılır. Aşkar olunan hadisə insident olduğu halda onu insident izləmə sisteminə bildirmək tələb olunur. İnformasiya təhükəsizliyi insidentlərinin idarəetmə bazasındakı toplanmış bütün informasiyalar tam olaraq qeydə alınır [6].

C. Qiymətləndirmə və qərar qəbuletmə

Qeyd edilən insidentlər bu məhələdə qiymətləndirilir və hansı növ cavablandırmanın tələb olunduğuna, hansı növ informasiya təhlükəsizliyi insidenti olduğuna qərar verilir. Eyni zamanda, informasiya təhlükəsizliyi boşluqları da qiymətləndirilərək kimə, necə və hansı prioritetlilik əsasında yönləndirildiyinə də qərar verilir.

Yekun olaraq informasiya təhlükəsizliyi insidentlərinin idarəetmə bazasındakı bütün qiymətləndirmə nəticələri və qərarlar tam qeydə alınır [6].

D. Cavablandırma

Bu mərhələdə insidentin prioritetlilik dərəcəsinə baxılaraq cavablandırma prosesi reallaşdırılır. Tələb olunduğu halda həmin informasiya təhlükəsizliyi insidenti üzrə təhqiqat aparılması və eskalasiya prosesi yerinə yetirilir. İnsidentə səbəb olan boşluqlar qiymətləndirilir [6].

E. Dərs çıxarma

Tələb olunarsa bu mərhələdə daha dərin təhqiqat işləri aparılır. İnformasiya təhlükəsizliyi insidentlərinin idarəetmə siyasətini məhz dərsçıxarmanın nəticəsi olaraq göstərə bilərik.

İnformasiya təhlükəsizliyi insident, hadisə və boşluq bazasının yenilənməsi, nəticələrin etibarlı təşkilatlarla paylaşılması və kommunikasiya, proses, prosedur və bildiriş formatının necə effektiv olmasına baxış və digər məsələlər bu mərhələdə yerinə yetirilir [6].

IV. BİLET BAZASININ İNFORMASİYA MODELİ

İnformasiya təhlükəsizliyi insident biletlərində onların növünə görə müxtəlif informasiya yazılır. Biletlər aşağıdakı kimi qruplara ayrılır [8]:

- Resurs biletləri;
- Xidməti biletlər.

Resurs tipli biletlər sistemin monitorinq edilməsi ilə yaradılır. Xidməti biletlər isə müəyyən təcrübəyə malik olan məsuliyyətli şəxs tərəfindən açılır [8]. İnformasiya təhlükəsizliyi insidentləri biletlərdə strukturlaşdırılmış və strukturlaşdırılmamış şəkildə saxlanılır. Hər bir biletdə isə insidentə xas olan aşağıdakı xarakteristikalar əks oluna bilər [9]:

- Bilet nömrəsi: bileti identifikasiya edən unikal açar;
- Bileti təsdiqləyən şəxsin identifikasiya nömrəsi;
- Biletin emalına təhkim olunmuş şəxsin identifikasiya nömrəsi;
- Biletin açılma vaxtı: hadisənin baş verməsi təsdiqlənən vaxt;
- Biletin həll olunma vaxtı: insidentin çözülməsi vaxtı;
- Biletin bağlanma vaxtı: insidentin bağlanma vaxtı;
- Biletin statusu: biletin indiki vəziyyətini göstərir. Bilet bağlanmış, bağlana bilinməmiş və ya baxılma statusunda ola bilər;
- Biletin həll olunma müddəti: biletin açılması və bağlanması arasındakı zaman intervalı;
- Prioritetlilik: Biletin hansı ardıcıllıqla emal olunması;
- İnsidentin emalına təhkim olunmuş səlahiyyətli şəxs;
- Təsvir: hadisə sözlərlə təsvir edilir.

Əlavə olaraq bir neçə biletlər üzərində çoxlu sayda metrikaları hesablamaq mümkündür. Məsələn, bir neçə bilet üzərində sıxlıq metrikasını tətbiq etmək olar. Belə ki, sıxlıq dedikdə müəyyən zaman periodu ərzində yazılan biletlərin sayı ola bilər [9].

Şəkil 1-də tipik bir insident bileti əks olunmuşdur.

Incident ID:	INC1
Severity:	High
Status:	Closed
Open Time:	7/16/2010 6:55:20 AM
Close Time:	7/17/2010 8:31:47 AM
Assignee Name:	John Doe
Assignment Group:	Account Management
Description:	The USER xxx has a successful *login* into the hub after *registration*, but he is *unable to access* SAP. Every time when he clicks on Sap work place, the screen goes blank!
Resolution:	Fixed USER xxx permission to access SAP.

Şəkil 1. İnsident bileti

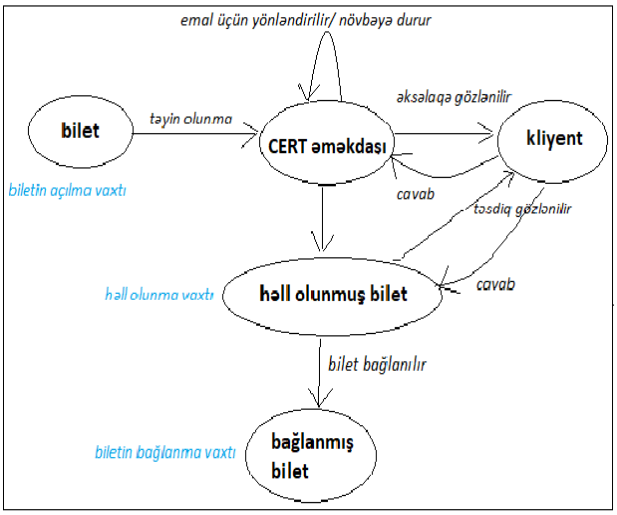
A. Bilet açma

Yeni bir informasiya təhlükəsizliyi insidenti baş verən anda offline rejimdə insident bileti açılır (ing. ticketing). Belə biletlər administratorlar, yardım masası (helpdesk) və ya son istifadəçilər tərəfindən əl ilə yaradıla bilər. Biletin açılması istənilən növ sistem problemlərindən qaynaqlana bilər.

Biletlərin emal edilmə prioriteti müxtəlif cür ola bilər:

- Birinci gələn, birinci emal olunan (FİFO, First İn and First Out): Bu, ən sadə yanaşmadır;
- Vaciblik dərəcəsinə görə: yüksək vaciblik dərəcəsinə malik biletlər aşağı dərəcəyə malik olan biletlərdən daha tez emal edilir;
- Bərabər zaman paylanması ilə və s [8].

Şəkil 2-də insident biletinin həyat dövrü verilmişdir [3,5].



Şəkil 2. İnsident biletinin həyat dövrü

V. BİLET BAZASI VERİLƏNLƏRİNİN İNTELLEKTUAL ANALİZİ

Tədqiqat aparmaq məqsədilə IBM-in CSV fayl formatında 100 min sətirlik insident bilet bazasından istifadə edilmişdir. Tədqiqat zamanı bazanın 70%-dən öyrətmə məqsədilə, digər 30% həcmli baza isə test üçün istifadə edilmişdir.

Eksperiment apardığımız insident biletləri bazasının attributları aşağıdakılardan ibarətdir:

- Ticket number: biletin nömrəsi;
- Requestor: bileti təsdiqləyən şəxsin identifikasiya nömrəsi;
- Requestor Seniority: bileti təsdiqləyən şəxsin səlahiyyətlilik dərəcəsi;
- ITOwner: biletin emalına təhkim olunmuş şəxsin identifikasiya nömrəsi;
- Severity: biletin kritiklik dərəcəsi;
- Priority: biletin emal edilmə üçün prioritetlik dərəcəsi;
- DaysOpen: biletin emal olunma müddəti;
- Satisfaction: biletin emal edilməsinin nəticəsi;
- Ticket Type: biletin növü.

Klassifikasiya metodu olaraq “təsadüfi meşə” (Random forest), “sadəlövh Bayes” (Naive Bayes), “neyron şəbəkələr”, “k-ən yaxın qonşular” (kNN, k-nearest neighbours), “dayaq vektor maşınları” (SVM, support vector machines) klassifikatorlarından istifadə olunmuşdur.

Sinifləndirmə biletin növünə əsasən aparılmışdır ki, bazada bu attributun 2 qiyməti vardır. Beləliklə nəticədə iki sinif alınır: issue və request.

Ən yaxşı nəticə isə təsadüfi meşə alqoritminin tətbiqi ilə əldə edilmişdir.

A. Klassifikasiya nəticələrinin interpretasiyası

Klassifikasiya prosesinin nəticəsində xəta matrisi (confusion matrix) əldə olunur. Matrisin ölçüsü siniflərin sayına görə iki və daha artıq ola bilər. Aparılan tədqiqatda iki sinif olduğu üçün iki ölçülü matris alınmış və strukturu cədvəl 1-də göstərilmişdir:

CƏDVƏL 1. XƏTA MATRİSİNİN STRUKTURU

	Proqnoz verilən sinif		
Həqiqi sinif		Sinif=yes	Sinif=no
	Sinif=yes	TP	FN
	Sinif=no	FP	TN

Klassifikasiya alqoritmlərinin effektivliyini müqayisə etmək üçün aşağıdakı keyfiyyət göstəricilərindən istifadə olunur:

- Düzgün müsbət hallar (True positives, TP): Tədqiqat aparılan insident bazasında bilet iki sinifdən birinə aid edilə bilər (issue=yes, request=no). Əgər bilet insidenti bildirirsə “yes” sinfinə, əks halda “no” sinfinə aid edilir. Düzgün müsbət hallar dedikdə, həqiqi sinif “yes” olduqda proqnoz edilən sinifin də “yes” olduğu hal başa düşülür. Yəni, burada səhv siniflərdirmə yoxdur.
- Düzgün mənfi hallar (True negatives, TN): Düzgün müsbət hallara analoji olaraq, həqiqi sinif “no” olduqda, proqnoz edilən sinifin də “no” olduğu haldır. Eyni qaydada bu da səhv siniflərdirmə hesab edilmir.
- Səhv müsbət hallar (False Positives, FP): Həqiqi sinifin “no”, proqnoz verilən sinifin “yes” olduğu haldır. Bu tip səhvlərə I növ səhvlər də deyilir.
- Səhv mənfi hallar (False Negatives, FN): Həqiqi sinfin “yes” olduğu halda, “no” sinfinə aid edilməsi olaraq, II növ səhvlər kimi də adlandırılır.

Yuxarıda qeyd olunan dörd parametrin qiymətini bildikdən sonra doğruluq (accuracy), dəqiqlik (precision), tamlıq (recall), F-ölçü (F-measure) qiymətləndirmələrini də hesablaya bilərik.

Doğruluq: düzgün proqnoz edilən müşahidələrin sayının ümumi müşahidələrin sayına nisbəti ilə ifadə olunur. Burada, düzgün proqnoz edilən müşahidələr “yes” və “no” sinfindən ola bilər. Odur ki, düzgün proqnoz edilən müşahidələri tapmaq üçün TP və TN-lər cəmlənir. Ümumi müşahidələrin sayı isə aydındır ki, yuxarıda deyilən dörd parametrin cəmi ilə ifadə olunursa, doğruluq üçün aşağıdakı düstur doğrudur [10].

$$Doğruluq = \frac{TP+TN}{TP+TN+FP+FN}$$

Dəqiqlik: düzgün proqnoz edilən müsbət müşahidələrin sayının ümumi müsbət müşahidələrin sayına olan nisbəti ilə ölçülür.

$$Dəqiqlik = \frac{TP}{TP+FP}$$

Tamlıq: düzgün proqnoz edilən müsbət müşahidələrin sayının həqiqi sinifi yes olan ümumi müşahidələrin sayına olan nisbətidir.

$$Tamlıq = \frac{TP}{TP+FN}$$

F-ölçü: dəqiqlik və tamlıq qiymətlərinin çəkili ortası ilə ifadə olunur.

$$F\text{-}ölçü = \frac{2*dəqiqlik*tamlıq}{dəqiqlik+tamlıq}$$

B. Metodların müqayisəli təhlili

Tədqiqat aparılan metodların nəticələri aşağıdakı cədvəldə göstərilmişdir.

CƏDVƏL 2. METODLARIN QİYMƏTLƏNDİRİLMƏSİ

Metod	Dəqiqlik	Doğruluq	Tamlıq	F-ölçü
Naïve Bayes	0.765	0.751	0.962	0.852
Random Forest	0.986	0.988	0.999	0.992
kNN	0.750	0.748	0.996	0.855
Neyron Şəbəkəsi	0.836	0.848	0.992	0.907
SVM	0.867	0.390	0.222	0.353

Məlum olmuşdur ki, tətbiq edilən mövcud metodlar içərisində ən yaxşı nəticə verən alqoritm təsadüfi meşə alqoritmidir. Belə ki, 100 min insident bileti üzərində 70% öyrənmə üçün, 30% isə test üçün sınaqdan keçirilmişdir. Yəni, 30 min insident biletindən test məqsədilə istifadə olunmuşdur. Təsadüfi meşə alqoritminin reallaşdırılması nəticəsində aşağıdakı cədvəldə verilmiş xəta matrisi alınmışdır:

CƏDVƏL 3. TƏSADÜFİ MEŞƏ ALQORİTMİ ÜÇÜN XƏTA MATRİSİ

		Proqnoz verilən sinif	
		Sinif=yes	Sinif=no
Həqiqi sinif	Sinif=yes	7268	320
	Sinif=no	27	22385

C. Təsadüfi meşə alqoritmi

Alqoritm ilk dəfə Tin Kam Ho tərəfindən 1995-ci ildə irəli sürülmüşdür [11]. O, ilk dəfə təsadüfi altmeşə metodundan istifadə etmişdir. Daha geniş variantda isə Leo Breiman və Adele Cutler tərəfindən inkişaf etdirilmişdir [12].

Qərar ağacları müxtəlif maşın təlimi məsələlərində istifadə olunan məşhur bir metoddur. Təsadüfi meşə alqoritmi və ya qərar ağacları meşəsi supervizorlu təlimə əsaslanan öyrənmə alqoritmi olaraq klassifikasiya, reqressiya və digər məsələlərdə geniş istifadə olunur.

Adından da göründüyü kimi, bu metod bir qərar ağacı yaratmaqla kifayətlənməyərək, çoxlu sayda, çoxdəyişənli ağac strukturunun yaradılmasına əsaslanır. Ən çox səs verilmiş (vote) qərar ağacları yekun ağacda nəzərə alınır. Qərar ağaclarının sayının artması dəqiqliyə müsbət təsir göstərir.

Metodun tətbiqi nəticəsində kənaraçıxmaların (outlier) sayı azalaraq klassifikasiya dəqiqliyi artmış olur. Buna görə də tətbiq sahəsi olduqca genişdir (bank sektoru, tibb və s.)

NƏTİCƏ

Məqalədə CERT komandalarında bilet açma mərhələsi üzərində dayanaraq IBM-in 100 000 sətirlik real insident biletləri bazası üzərində tədqiqatlar aparılmışdır. Baza əsasən kateqoriya tipli verilənlərdən ibarət olduğundan bu bazaya verilənlərin intellektual analizi üsullarından olan klassifikasiya metodları tətbiq edilmiş, sonda modellərin qiymətləndirilməsi aparılmışdır. Qiymətləndirmə nəticəsində bilet bazası üzərində ən yaxşı klassifikasiya nəticələrini verən modelin təsadüfi meşə alqoritmi olması qənaətinə gəlinmişdir.

İnsident idarəetmə prosesinin inkişafı üçün biletlərin korrelyasiyası, klasterizasiyası, proqnozlaşdırılması tətbiq edilə bilər. Strukturlaşdırılmamış insident biletlərinin təsvir hissəsinə mətnlərin intellektual analizinin tətbiqi daha əlverişli nəticələr verə bilər.

ƏDƏBİYYAT

[1] R.Əliquliyev, Y.İmamverdiyev, “İnformasiya təhlükəsizliyi insidentləri”, İnformasiya Texnologiyaları nəşriyyatı, 2012, 219 s.
[2] R. Ruefle et al. “Computer Security Incident Response Team Development and Evolution”, IEEE Security & Privacy, vol. 12, no. 5, pp. 16-26, 2014.
[3] S. Salah et al. “A model for İncident Tickets Correlation in Network”, Department of Signal Theory, Telematics and Communications, University of Granada, Spain, 2015, 35 p.
[4] L. Tang et al. “Identifying Missed Monitoring Alerts based on Unstructured Incident Tickets”, 9th CNISM and Worksops, 2013, pp. 143-146.
[5] Y. Li, T. Li, “A method of Effort Estimation for Incident Tickets in IT Services”, Proceedings of IEEE International Conference on Service Operations and Logistics, and Informatics, 2013, p. 311-316.
[6] ISO/IEC 27035, Information Technology – Security Techniques – Information security incident management, 2011.
[7] P. Chichonski et al. Computer Security Incident Handling Guide, NIST Special Publication 800-61, U.S. Department of Commerce, 2012, 79 p.
[8] P. Marcu et al. “Towards an Optimized Model of Incident Ticket Correlation”, 2009 IFIP/IEEE International Symposium on Integrated Network Management, 2009, p. 569-576.
[9] T. Li et al. “Incident Ticket Analytics for IT Application Management Services”, Proceedings of IEEE International Conference on Services Computing, 2014, p. 568-574.
[10] J. Davis, M. Goadrich, “The Relationship Between Precision-Recall and ROC Curves”, 2005.
[11] H. Kam, “Random Decision forests”, Proceedings of the 3rd International Conference on Document Analysis and Recognition, pp. 278-282 , 1995.
[12] Random forest, https://en.wikipedia.org.

MINING OF INCIDENT DATA FOR CERT

Yadigar Imamverdiyev[1], Elshan Baghirov[2]
[1,2]Institute of Information Technology of ANAS, Baku, Azerbaijan
[1]*yadigar@lan.ab.az*, [2]*elsenbagirov1995@gmail.com*

Abstract – In this work, phases of incident handling and ticketing were analyzed, and data mining methods were applied to real incident tickets database.

Keywords – ticket analysis, incident analysis, CERT teams, random forest, ticketing