

Information technologies and multimedia Informacinės technologijos ir multimedija

INCIDENTŲ VALDYMO MODELIO TOBULINIMAS, TAIKANT MAŠININIO MOKYMOŠI METODUS

Roman JEVSEJEV ¹, Mindaugas BEREIŠA²

¹Vilniaus Gedimino technikos universitetas, Vilnius, Lietuva

²UAB „LTG Link“, Vilnius, Lietuva

- gauta 2024 m. birželio 3 d.
- priimta 2024 m. birželio 4 d.

Santrauka. IT infrastruktūros techninis palaikymas yra esminis organizacijos veiklos aspektas, kurio sudėtingiausia užduotis yra užtikrinti veikimo tęstinumą. Kokybiškas palaikymas garantuoja aukštą IT efektyvumą, tačiau sudėtingi incidentai sumažina palaikymo kokybę ir reikalauja veiksmingo valdymo. Incidentų valdymas apima konfigūracijų procesus ir techninių sprendimų kontrolę. Siekiant pagerinti techninį palaikymą, būtina laikytis tiek kiekybinių, tiek kokybinių standartų ir atsižvelgti į sistemų specifiką. Pagal paslaugų lygio sutartis (SLA) svarbus incidentų sprendimo laikas. „Service Desk“ įrankiai, taikant mašininio mokymosi metodus, gali padėti optimizuoti šiuos procesus. Naudotojų neteisingai klasifikuotos užklausos lemia papildomą IT komandos darbą ir vilkina incidentų sprendimą. „K-means“ klasterizacijos, „Random Forest“ regresijos ir klasifikacijos mašininio mokymosi metodai gali optimizuoti incidentų valdymą ir pagreitinti sprendimo laiką. Tyrimo tikslas yra analizuoti „Service Desk“ incidentų duomenis, siekiant modeliuoti sprendimų laiką ir pagerinti incidentų valdymą.

Reikšminiai žodžiai: IT infrastruktūra, techninis palaikymas, incidentų valdymas, incidentų sprendimo laikas, *Service Desk*, mašininio mokymosi metodai, užklausos klasifikavimas.

✉ Autorius susirašinėti. El. paštas roman.jevsejev@gmail.com

1. Įvadas

Vienas svarbiausių IT infrastruktūros veiklos aspektų organizacijoje yra informacinių technologijų infrastruktūros techninis palaikymas, kuomet sudėtingiausia užduotis – užtikrinti infrastruktūros veikimo tęstinumą (Zuev et al., 2018). Kokybiškas techninis palaikymas užtikrina aukštą IT veiklos efektyvumo rodiklį (Bartolini et al., 2009). Nestandartinis įvykis sumažina techninio palaikymo kokybę ir vadinamas incidentu (Revina et al., 2021). Incidentų valdymas apima įvairių incidentų konfigūracijų valdymo procesus, techninių sprendimų sekos kontrolę (Costa et al., 2019).

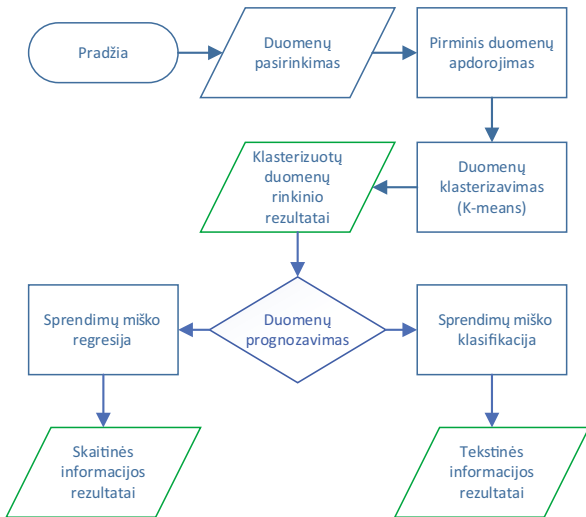
Siekiant kelti IT techninio palaikymo lygį, procesai turi atitikti tiek kiekybinius, tiek kokybinius standartus, atsižvelgiant į skirtingų sistemų charakteristikas ir konfigūracijas. Incidentų sprendimo laikas – tai laikotarpis nuo pranešimo apie incidentą iki jo sprendimo pabaigos ir yra vienas iš nustatytų matavimo standartų, kuris yra įtrauktas į Paslaugų lygio sutarties (SLA) protokolus (Eckerson, 2010). IT infrastruktūros veiklai valdyti yra naudojami įvairūs „Service Desk“ įrankiai, kurių efektyvumui didinti taikomi mašininio mokymosi metodai. „Service Desk“ struktūra priklauso nuo organizacijos poreikių. Įvykus incidentui, naudotojas regis-

truoja pranešimą rankiniu būdu pildydamas incidento tipo, skubos, incidento apibūdinimo laukus, įkeldamas incidentą patvirtinančius dokumentus arba kitą incidentą apibūdinančią informaciją. Pasitaiko atvejų, kuomet registruojant pranešimą naudotojo pasirinktas incidento tipas lemia netinkamą užklausos klasifikavimą (Agarwal et al., 2012). Netinkamai klasifikavus užklausas naudotojai nukreipiami pas netinkamos srities specialistą, o tai suteikia papildomo darbo IT komandos darbuotojams, kuriems tenka iš naujo nukreipti naudotojus. Dėl šios priežasties yra neracionaliai išnaudojami organizacijos žmogiškieji išteklių ir vilkinamas incidento sprendimas (Xu et al., 2018). „Service Desk“ sistemos turi daug struktūrizuotų laukų, kurie nesugeba atskirti naudotojo pateikto užklausos apibūdinimo natūralia kalba (Dasgupta et al., 2014). IT specialistas privalo suprasti incidento apibūdinimą, siekiant pritaikyti tinkamą techninį sprendimą per trumpiausią terminą (Agarwal et al., 2017). Altintas ir Tantug (2014) siūlo „Service Desk“ problemai spręsti ir sistemai patobulinti taikyti duomenų klasterizavimo (angl. *K-means*), sprendimų miško regresijos (angl. *Random Forest Regression*) ir sprendimų miško klasifikacijos (angl. *Random Forest Classifier*) mašininio mokymosi metodus.

Tyrimo tikslas yra analizuoti „Service Desk“ sistemos incidentų duomenis ir, naudojant mašininio mokymosi metodus, apskaičiuoti incidentų sprendimo laiką, atsižvelgiant į darbuotoją, kuris sprendžia incidentą. Remiantis gautais rezultatais, modelis turi nuspręsti, kas galėtų išspręsti incidentą. Tai leis sukurti sprendimą, kuris padės optimizuoti incidentų valdymo procesą „Service Desk“ sistemose.

2. Tyrimo metodologija

Tyrimas įgyvendinamas eksperimentiniu būdu. Tyrimo eigos struktūrinė schema pateikta 1 paveiksle. Ji apima šiuos žingsnius: duomenų rinkinio pasirinkimas, pirminis duomenų apdorojimas, duomenų klasterizavimo metodo taikymas, duomenų prognozavimas, taikant sprendimų miško regresijos metodą skaitinei informacijai gauti ir sprendimų miško klasifikacijos metodą.



1 paveikslas. Tyrimo eigos struktūrinė schema

2.1. Tyrimo metodologijos apžvalga

Mašininiam apmokymui yra būtinas pirminis duomenų apdorojimas. Modelio kokybė tiesiogiai priklauso nuo duomenų kokybės, todėl teisingas duomenų apdorojimas atlieka lemiamą vaidmenį, siekiant norimų rezultatų.

Pirminis duomenų apdorojimas yra raktas į sėkmingą ir patikimą mašininio mokymo modelių kūrimą (Ng, 2018). Taikant šį metodą, duomenys paverčiami skaitmenine reikšme, kuomet kiekvienas parametras įgyja unikalų skaičių ir suprantamą mašininiam apmokymui formatą.

Tinkamai paruošus mašininio mokymo duomenis, taikomas duomenų klasterizavimo metodas. Šio metodo algoritmas yra vienas iš populiariausių, kadangi leidžia efektyviai sugrupuoti duomenis į skirtingus klasterius (Arthur & Vassilvitskii, 2007). Duomenų klasterizavimo metodo algoritmas yra paprastas, bet galingas klasterizavimo metodas, kuris naudojamas įvairiose srityse, įskaitant duomenų analizę, matematinę statistiką ir mašininį mokymą (Jain et al., 1999).

Duomenų klasterizavimo algoritmas prasideda nuo pradinių klasterių centrų atsitiktinio pasirinkimo. Kiekvienas duomenų taškas yra priskiriamas artimiausiam klasteriui pagal atstumą iki klasterio centro. Kitas žingsnis yra klasterių centrų atnaujinimas, kai jie yra perskaičiuojami kaip visų priskirtų duomenų taškų vidurkiai. Šie žingsniai kartojami tol, kol klasterių centrų pozicijos stabilizuojasi. Taikant šį metodą, pradiniai „Service Desk“ duomenys paverčiami į statistinių duomenų grupes, tarp kurių nustatoma tarpusavio priklausomybė.

Sprendimų miško algoritmas plačiai aprašytas Breiman (1996). Įprastai mokymo rinkinio dydis priklauso nuo duomenų sudėtingumo, apimties ir tikslumo. Dažniausiai naudojama mokymo rinkinio dalis yra apie 70–80 % visų turimų duomenų. Likę 20–30 % duomenų naudojami modelio tikslumui įvertinti ir vadinami testavimo rinkiniu. Todėl 70–80 % duomenų yra naudojami sprendimų miško algoritmui mokytis, o likę 20–30 % yra naudojami modelio tikslumui įvertinti. Algoritmas naudojamas rezultatui numatyti. Jei skaitiniams duomenims analizuoti naudojamas sprendimų miško regresijos algoritmas, tai tekstinei informacijai analizuoti naudojamas sprendimų miško klasifikacijos algoritmas (Paramesh & Shreedhara, 2019).

2.2. Duomenų pasirinkimas ir valymas

Duomenų rinkinys, naudojamas šiame tyrime, pagrįstas Lietuvos Respublikoje esančios organizacijos „Service Desk“ duomenų struktūra. Duomenų rinkinį sudaro 5000 (penki tūkstančiai) incidentų įrašų, kuriuos registravo 203 (du šimtai trys) naudotojai. Šiuos incidentus per įvairų laiką sprendė 7 (septyni) IT darbuotojai. Informacija pateikta 1 lentelėje.

Taikomo duomenų rinkinio struktūra:

- Naudotojo ID (USER_ID) – numeris, priskirtas naudotojui, kuris užregistravo incidentą;
- Nešiojamojo kompiuterio modelis (HW_MODEL) – skaičius, kuris žymi nešiojamojo kompiuterio modelį;
- Incidento aprašymas (ISSUE) – incidentų aprašymai;
- Techninis sprendimas (SOLUTION) – techniniai sprendimai;
- Sprendimo laikas minutėmis (TIME_MIN) – laikas minutėmis, per kurį incidentas buvo išspręstas;
- IT darbuotojo ID (WORKER) – IT darbuotojo, kuris sprendė konkretų incidento atvejį, numeris.

Nors duomenų rinkinys atitinka tikros organizacijos struktūrą, jame gali būti triukšmingų ir nereikalingų duomenų. Norint pritaikyti šį modelį skirtingų įmonių poreikiams, atsižvelgiant į kiekvienos įmonės specifiką, būtina atlikti papildomą duomenų analizę ir valymą, siekiant gauti švarų duomenų rinkinį.

Nesuderinti duomenys gali turėti neigiamą poveikį modelio tikslumui. Duomenų valymas yra svarbus procesas, kuris apima kelis esminius žingsnius, siekiant išvalyti duomenų rinkinį nuo netikslumų, triukšmo ir kitų problemų, kad būtų galima duomenis naudoti analizei ar modeliavimui.

1 lentelė. Mašininio mokymo duomenų rinkinys

USER_ID	HW_MODEL	ISSUE	SOLUTION	TIME_MIN	WORKER
U1	NB_MODEL_1	ISSUE_1	SOLUTION_1	TIME_1	W1
U2	NB_MODEL_2	ISSUE_2	SOLUTION_2	TIME_2	W2
U3	NB_MODEL_3	ISSUE_3	SOLUTION_3	TIME_3	W3
...
U203	NB_MODEL_10	ISSUE_X	SOLUTION_Y	TIME_Z	W7

Pirmiausia identifikuojami pasikartojantys duomenų įrašai „ISSUE“ ir „SOLUTION“ stulpeliuose. Tai daroma lyginant visas eilutes arba stulpelius ir, jei aptinkami tokie pasikartojimai, vienas iš jų pašalinamas iš duomenų rinkinio.

Antrasis žingsnis yra skirtas atrasti ir ištaisyti bet kokias klaidas, kurios yra duomenų rinkinyje. Tai gali būti neteisingos reikšmės ar reikšmės, kurių trūksta.

3. Modelio sudarymas

Duomenų paruošimas vykdomas nuskaitant duomenis iš „Excel“ failo, naudojant „NumPy“ ir „Pandas“ bibliotekas. Duomenys struktūrizuojami ir šiuo formatu galima lengvai manipuliuoti. Duomenų valymo metu šalinami dublikatai, triukšmingi duomenys ir kitokie triukšmingi įrašai. Šis procesas užtikrina, kad duomenys būtų tikslūs ir tinkami tolesniam analizės etapui.

Duomenų normalizavimo metu reikšmių matavimo vienetai suvienodinami. Tai užtikrina teisingą duomenų interpretaciją. Duomenys paverčiami į skaitinį formatą ir tampa suprantami modeliui. Pavyzdys pateiktas 2 paveiksle.

	USER_ID_U10	USER_ID_U100	USER_ID_U101	USER_ID_U102
index				
Application freezes frequently.	0.000000	0.0	0.000000	0.014286
Browser is redirecting to unknown websites.	0.014493	0.0	0.000000	0.014493
Cannot access VPN.	0.014493	0.0	0.000000	0.014493
Cannot access certain websites.	0.000000	0.0	0.000000	0.000000
Cannot access email account.	0.000000	0.0	0.000000	0.000000
...

2 paveikslas. Mašininiam mokymui paruošto duomenų rinkinio fragmentas

3.1. Duomenų klasterizavimas

Duomenys importuojami iš „Excel“ failo, naudojant „Pandas“ biblioteką, į „DataFrame“ struktūrą, kuri leidžia atlikti įvairias duomenų manipuliacijas ir analizę. Segmentavimas vyksta pagal „ISSUE“ stulpelį. Kiekvienam klasteriui apskaičiuotos šios agreguotos metrikos: vidutinis incidento sprendimo laikas (*Avg_Time*), maksimalus incidento sprendimo laikas (*Max_Time*), minimalus incidento sprendimo laikas (*Min_Time*), standartinis nuokrypis (*STD_Time*) ir medianinis laikas (*Median_Time*). Taip

pat iš stulpelių „WORKER“, „SOLUTION“, „HW_MODEL“ ir „USER_ID“ sudarytas unikalių reikšmių sąrašas. Duomenys išsaugomi naujame „Excel“ faile ir vizualizuojami.

Kategorinės reikšmės „WORKER“, „SOLUTION“, „HW_MODEL“ ir „USER_ID“ programuojamos naudojant „OneHot Encoding“ funkciją, konvertuojant kategorines reikšmes į dvejetainius kintamuosius, kurie yra tinkami klasterizacijos algoritmams.

Tiek kategoriniai, tiek skaitiniai požymiai normalizuojami naudojant „MaxAbsScaler“ funkciją. Normalizavimas atliekamas siekiant užtikrinti, kad visi požymiai turėtų panašų mastelį, taip išvengiant situacijos, kai vienas požymis dominuoja prieš kitus dėl savo didesnio mastelio.

Požymiai sujungiami į vieną rinkinį. Naudojant „TruncatedSVD“ funkciją atliekamas dimensijų mažinimas, todėl duomenų dimensijos sumažėja iki dviejų komponentų, tokiu būdu siekiama, kad būtų lengviau vizualizuoti duomenis ir kad būtų sumažintas skaičiavimų sudėtingumas. Rezultatai pateikti 3 paveiksle.

Optimaliam klasterių kiekiui nustatyti naudojama „Silhouette score“ funkcija. Silueto rezultatas rodo kiekvieno duomenų taško priskyrimo tam tikram klasteriui tikslumą, lyginant su kitais klasteriais. Šio tyrimo atveju nustatytas aukščiausias 20 klasterių priskirtas silueto rezultatas pateiktas 4 paveiksle.

Nustačius optimalų klasterių kiekį, pakartotinai paleidžiamas *k-means* algoritmas su šiuo klasterių kiekiu. Algoritmas sugrupavo duomenų taškus taip, kad būtų minimizuotas atstumas tarp atitinkamų klasterių centrų. Kiekvienam duomenų taškui buvo priskirtas klasterio numeris, kuris pridėdamas prie pradinio „DataFrame“ leido toliau analizuoti duomenis, atsižvelgiant į jiems priskirtus klasterius.

3.2. Duomenų klasifikavimas

Kaip ir duomenų klasterizavimo metu, kategorinės „WORKER“ reikšmės programuojamos naudojant „OneHotEncoder“ funkciją. Tekstinės „ISSUE“ reikšmės programuojamos naudojant „TF-IDF“ (angl. *Term Frequency-Inverse Document Frequency*) vektorizaciją. Ši funkcija nustato, kiek svarbus yra tam tikras žodis, atsižvelgiant į jo dažnį visame duomenų rinkinyje. Taip gaunami svarbūs požymiai, kurie geriau reprezentuoja tekstinius duomenis.

Kategoriniai požymiai ir „TF-IDF“ vektorizuotos tekstinės reikšmės apjungiami į vieną reikšmių rinkinį, naudojant „hstack“ funkciją. Apjungtas požymių rinkinys (X) ir tikslinės

reikšmės (Y) padalinti į mokymo ir testavimo rinkinius, naudojant „train_test_split“ funkciją. Mokymo rinkinys sudarė 80 % visų duomenų, o testavimo rinkinys – likusius 20 %.

Laiko (TIME_minutes) reikšmėms numatyti naudojama sprendimų miško regresijos funkcija, kuri apdoroja tiek kategorines, tiek skaitines reikšmes. Modelis apmokytas su mokymo rinkiniu ir patikrintas su testavimo rinkiniu. Modelio tikslumas įvertintas, taikant vidutinės kvadratinės paklaidos (angl. Mean Squared Error) funkciją.

Darbuotojų (WORKER) ir incidentų (ISSUE) reikšmėms klasifikuoti panaudota sprendimų miško klasifikacijos funkcija, kuri apdoroja mišrius duomenis ir leidžia sukurti tikslius klasifikacijos modelius. Modelis apmokytas su mokymo rinkiniu ir patikrintas su testavimo rinkiniu. Tikslumas buvo įvertintas, naudojant tikslumo funkciją (angl. Accuracy Score), kuri parodo, kiek procentų atvejų modelis teisingai klasifikavo.

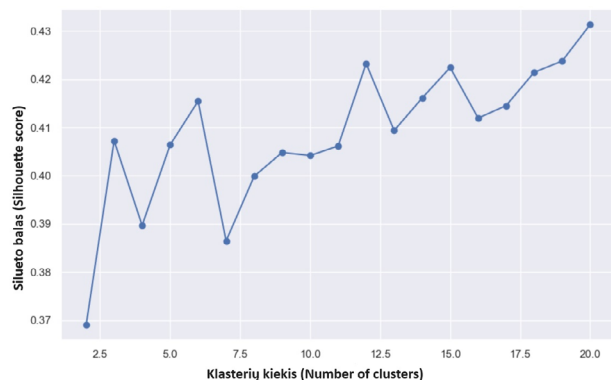
4. Rezultatų analizė

Eksperimento rezultatai rodo, kad naudojant pažangius mašininio mokymosi metodus galima efektyviai analizuoti ir prognozuoti incidentų sprendimo laiką bei klasifikuoti incidentus pagal jų pobūdį. Šie metodai padeda pagerinti darbo krūvio paskirstymą tarp IT darbuotojų, sumažinti incidentų sprendimo laiką ir padidinti bendrą paslaugų teikimo efektyvumą.

Naudojant „K-means“ algoritmą ir „Silhouette score“ funkciją, buvo nustatyta, kad optimalus klasterių kiekis yra 20. Rezultatas pateiktas 3 paveiksle. Jis parodo, kad incidentų duomenys gali būti efektyviai suskirstyti į 20 klasterių, atsižvelgiant į jų charakteristikas.

Kiekvienam klasteriui apskaičiuotos šios agreguotos reikšmės: vidutinis incidento sprendimo laikas (Avg_Time), maksimalus incidento sprendimo laikas (Max_Time), minimalus incidento sprendimo laikas (Min_Time), standartinis nuokrypis (STD_Time) ir medianinis laikas (Median_Time). Taip pat iš stulpelių „WORKER“, „SOLUTION“, „HW_MODEL“ ir „USER_ID“ sudarytas unikalių reikšmių sąrašas. Duomenys pateikti 4 paveiksle.

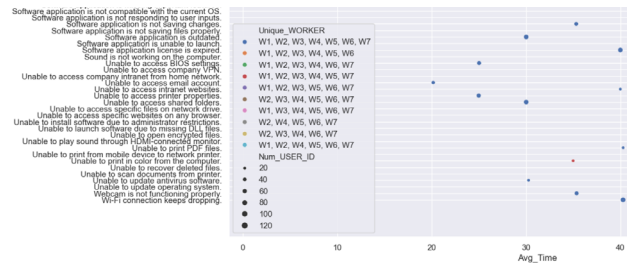
Sprendimų miško klasifikacijos modelis taip pat pasiekė aukštą tikslumo lygį. Modelis sugebėjo tiksliai klasifikuoti



3 paveikslas. Klasterių pokyčio grafikas

incidentus pagal jų aprašymus ir priskirti juos atitinkamiems IT darbuotojams. Aukštas tikslumo procentas rodo, kad modelis yra patikimas ir gali būti naudojamas realioje aplinkoje, siekiant efektyviau valdyti incidentų sprendimo procesą.

Naudojant sprendimų miško regresijos modelį, pasiekta aukšta tikslumo reikšmė, vertinant vidutinės absoliučios paklaidos (angl. Mean Absolute Error, MAE) metriką. Modelis sugebėjo tiksliai prognozuoti incidentų sprendimo laiką, remiantis mokymo duomenimis. Rezultatas pateiktas 6 paveiksle. Maža MAE reikšmė rodo, kad modelis yra efektyvus ir gali būti naudojamas numatant, kiek laiko užtruks naujų incidentų sprendimas.



4 paveikslas. Segmentuotų duomenų fragmentas

```
# New data example
new_issue = ["The computer crashes when trying to print."]
new_worker = ["W2"]

# Transforming the new data
new_worker_encoded = ohe.transform([new_worker])
new_issue_tfidf = tfidf.transform(new_issue)

# Combine the transformed features
new_X = hstack([new_worker_encoded, new_issue_tfidf])

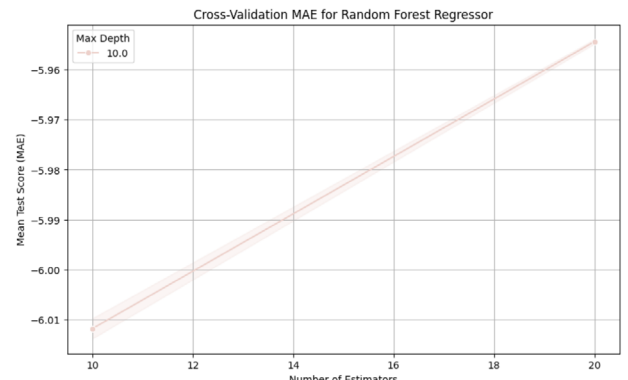
# Define and train the model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Make the prediction using the trained model
predicted_time = model.predict(new_X)

# Print or use the predicted time
print(predicted_time)
```

[54.325]

5 paveikslas. Incidento klasifikavimas



6 paveikslas. Vidutinė absoliuti paklaida

Modelis sėkmingai prognozavo, kad darbuotojas, kuris turėtų tvarkyti incidentą „the computer crashes when trying to print“ (liet. bandant spausdinti, kompiuteris išsijungia), yra W2 ir sprendimo laikas bus apytiksliai 54.325 minutės. Rezultatas pateiktas 5 paveiksle. Regresijos modelio vidutinė absoliuti paklaida (MAE) yra 0.436, o tai rodo, kad modelis prognozuoja laiko rezultatus vidutiniškai su 0.436 minučių absoliučia paklaida. Tai rodo, kad modelis gali būti naudingas realiam scenarijui, sprendžiant problemų paskirstymo ir laiko prognozavimo klausimus.

5. Išvados ir pasiūlymai

Eksperimento rezultatai rodo, kad sukurtas modelis gali būti sėkmingai taikomas realiose organizacijose, siekiant pagerinti IT paslaugų valdymo procesus. Norint pritaikyti šį modelį kitose įmonėse, būtina atlikti papildomą duomenų analizę ir valymą, atsižvelgiant į konkrečios organizacijos specifiką ir poreikius. Tai padės užtikrinti, kad modelis būtų tinkamai pritaikytas ir galėtų tiksliai prognozuoti ateities incidentus bei efektyviai valdyti incidentų sprendimo procesą.

Modelio pagrindu galima apskaičiuoti žmogiškųjų išteklių resursus ir įvertinti IT darbuotojų efektyvumą. Tai suteikia galimybę organizacijoms optimaliai paskirstyti darbo krūvį ir efektyviau planuoti resursus. Be to, modelį galima integruoti į API, pavyzdžiui, į IT pagalbos „chat“ sistemą, veikiančią „ChatGPT“ pagrindu. Ši integracija leistų realiu laiku analizuoti ir prognozuoti incidentų sprendimo laiką, taip pat automatiškai paskirstyti incidentus tinkamiems IT darbuotojams, pagerinant klientų aptarnavimo kokybę ir efektyvumą. Toks sprendimas ne tik pagerintų darbo organizavimą, bet ir padėtų operatyviai reaguoti į klientų poreikius, užtikrinant aukštą paslaugų kokybę.

Autorių indėlis

Eksperimentinį tyrimą atliko VJ Ignalinos atominės elektrinės vyresnysis informacinių technologijų sistemų administratorius Roman Jevsejev (toliau – R. J.) ir UAB „LTG Link“ traukinių techninės priežiūros inovacijų vadovas Mindaugas Bereiša (toliau – M. B.). R. J. inicijavo eksperimentinį tyrimą ir sukūrė tyrimo metodologijos aprašymą. R. J. ir M. B. atliko eksperimentinio tyrimo duomenų rinkimą ir analizę. M. B. praktiškai pritaikė mašininio mokymosi metodus ir teikė konsultaciją techniniais klausimais sudarant tyrimo modelį. R. J. interpretavo rezultatų duomenis ir parengė mokslinį straipsnį. Abu autoriai peržiūrėjo ir patvirtino galutinį mokslinio straipsnio variantą.

Atskleidimo pareiškimas

Autoriai patvirtina, kad neturi jokių konkuruojančių finansinių, profesinių ar asmeninių interesų, kurie galėtų paveikti pateikto mokslinio straipsnio turinį. Naudota tyrimo metodologija yra viešai prieinama. Autoriai suteikia teisę naudoti moksliniame straipsnyje pateiktus duomenis tik akademinės veiklos tikslams.

Literatūra

- Agarwal, S., Aggarwal, V., Akula, A. R., Dasgupta, G. B., & Sridhara, G. (2017). Automatic problem extraction and analysis from unstructured text in IT tickets. *IBM Journal of Research and Development*, 61(1), 41–52. <https://doi.org/10.1147/JRD.2016.2629318>
- Agarwal, S., Sindhgatta, R., & Sengupta, B. (2012). SmartDispatch: Enabling efficient ticket dispatch in an IT service environment. In *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1393–1401), Beijing, China. <https://doi.org/10.1145/2339530.2339744>
- Altintas, M., & Tantug, A. C. (2014). Machine learning based ticket classification in issue tracking systems. In *Proceedings of the International Conference on Artificial Intelligence and Computer Science (AICS)* (pp. 195–207), Bandung, Indonesia.
- Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035). Society for Industrial and Applied Mathematics.
- Bartolini, C., Stefanelli, C., & Tortonesi, M. (2009). Business-impact analysis and simulation of critical incidents in IT service management. In *2009 IFIP/IEEE International Symposium on Integrated Network Management* (pp. 9–16), New York, NY, USA. <https://doi.org/10.1109/INM.2009.5188781>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/BF00058655>
- Costa, J., Pereira, R., & Ribeiro, R. (2019). ITSM automation-using machine learning to predict incident resolution category. In K. S. Soliman (Ed.), *Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision 2020* (pp. 5819–5830). International Business Information Management Association, IBIMA.
- Dasgupta, G. B., Nayak, T. K., Akula, A. R., Agarwal, S., & Nadgouda, S. J. (2014). Towards auto-remediation in services delivery: Context-based classification of noisy and unstructured tickets. In *Proceedings of the International Conference on Service-Oriented Computing (SOC)* (pp. 478–485). Springer. https://doi.org/10.1007/978-3-662-45391-9_39
- Eckerson, W. (2010). *Performance dashboards: Measuring, monitoring, and managing your business*. Wiley.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>
- Ng, A. (2018). *Machine learning yearning*. https://nessie.ilab.sztaki.hu/~kornai/2020/AdvancedMachineLearning/Ng_MachineLearningYearning.pdf
- Paramesh, S., & Shreedhara, K. (2019). IT help desk incident classification using classifier ensembles. *ICTACT Journal on Soft Computing*, 9(4), 1980–1987.
- Revina, A., Buza, K., & Meister, V. G. (2021). Designing explainable text classification pipelines: Insights from IT ticket complexity prediction case study. In *Interpretable artificial intelligence: A perspective of granular computing* (pp. 293–332). Springer. https://doi.org/10.1007/978-3-030-64949-4_10
- Zuev, D., Kalistratov, A., & Zuev, A. (2018). Machine learning in IT service management. *Procedia Computer Science*, 145, 675–679. <https://doi.org/10.1016/j.procs.2018.11.063>
- Xu, J., He, R., Zhou, W., & Li, T. (2018). Trouble ticket routing models and their applications. In *IEEE Transactions on Network and Service Management* (pp. 530–543). IEEE. <https://doi.org/10.1109/TNSM.2018.2790956>

IMPROVEMENT OF INCIDENT MANAGEMENT MODEL USING MACHINE LEARNING METHODS

R. Jevsejev, M. Bereiša

Abstract

Technical support of IT infrastructure is a crucial aspect of organizational operations, with the most challenging task being ensuring service continuity. Quality support guarantees high IT efficiency, but complex incidents reduce support quality and require effective management. Incident management includes configuration processes and control of technical solutions. To improve technical support, adhering to both quantitative and qualitative standards and considering system specifics is necessary. According to service level agreements (SLA), the resolution time of incidents is important. „Service Desk“ tools, applying machine learning methods, can help optimize these processes. Incorrectly classified user requests lead to additional work for the IT team and delay incident resolution. Machine learning methods, such as K-means clustering, Random Forest regression, and classification, can optimize incident management and speed up resolution time. The research analyzes „Service Desk“ incident data to model resolution times and improve incident management.

Keywords: IT infrastructure, technical support, incident management, incident resolution time, Service Desk, machine learning methods, request classification.