ELEKTRON XATLARDA SPAMNI ANIQLASH UCHUN SUN’IY INTELLEKT USULLARINI SOLISHTIRUVCHI TADQIQOT

Raximov Bobomurod Ramazonovich
TATU Mustaqil izlanuvchisi
Tel:+998997601554

KIRISH

Elektron pochta tizimlari zamonaviy kommunikatsiya va biznes jarayonlarining ajralmas qismiga aylangan. Dunyo bo‘yicha milliardlab foydalanuvchilar kundalik ishlarida elektron xatlardan foydalanadi, shu bilan birga ularning samaradorligi va xavfsizligi elektron xatlar orqali uzatiladigan ma’lumotlarning sifati va ishonchliligiga bog‘liq bo‘ladi. Biroq, so‘nggi yillarda elektron xatlar orqali yuboriladigan spam xabarlar sonining keskin oshishi tizimlarning ishlashini sekinlashtiradi, foydalanuvchi tajribasini pasaytiradi va kiberxavfsizlikka tahdid soladi. Spam xabarlar nafaqat reklamani o‘z ichiga oladi, balki phishing va zararli dasturlarni tarqatish orqali moliyaviy va ma’lumot xavfsizligiga zarar yetkazishi mumkin. Shu sababli elektron xatlarni samarali tasniflash va spamni avtomatik aniqlash masalasi bugungi kunda axborot texnologiyalari sohasida dolzarb ilmiy muammo hisoblanadi.

Spamni aniqlash masalasida an’anaviy qoidalarga asoslangan filtrlar ilgari foydali bo‘lsa-da, ular murakkab va doimiy o‘zgaruvchan xabar strukturalarini samarali qayta ishlay olmaydi. Shu bilan bog‘liq ravishda, so‘nggi yillarda sun’iy intellekt (SI) va mashinani o‘rganish algoritmlari elektron xatlarni avtomatik tasniflashda asosiy vositalardan biriga aylangan. SI usullari xat matnidagi lingvistik va statistik belgilarni tahlil qilish, xabarlarni spam yoki no-spam sifatida ajratish hamda tizimlarning o‘z-o‘zini moslashtirish qobiliyatini oshirish imkonini beradi[1]. Ularning yordamida katta hajmdagi elektron pochta ma’lumotlarini real vaqt rejimida qayta ishlash va yangi spam namunalari paydo bo‘lganda tizimni yangilash imkoniyati yuzaga keladi.

Bugungi kunda mashinani o‘rganishning turli algoritmlari, jumladan Naive Bayes, Support Vector Machines (SVM), Random Forest va sun’iy neyron tarmoqlar, elektron xatlarni tasniflashda keng qo‘llanilmoqda. Har bir usulning afzalliklari va kamchiliklari mavjud bo‘lib, ularni samaradorlik, aniqlik va ishlash tezligi nuqtai nazaridan solishtirish muhim ahamiyatga ega. Tadqiqotlar shuni ko‘rsatadiki, sun’iy neyron tarmoqlari va SVM algoritmlari kompleks matnli xabarlarni aniqlashda yuqori aniqlikka erishsa, Naive Bayes algoritmi oddiy tuzilishga ega xabarlarni tezkor qayta ishlashda samarali natija beradi [2].

Ushbu tadqiqotning maqsadi; elektron xatlarda spamni aniqlash uchun turli sun’iy intellekt usullarini solishtirish va ularning samaradorligini ilmiy jihatdan baholashdan iborat. Tadqiqot natijalari elektron pochta tizimlarini optimallashtirish, xavfsizlikni ta’minlash va foydalanuvchi tajribasini oshirishga xizmat qilishi kutilmoqda.

TADQIQOT METODOLOGIYASI

Ushbu tadqiqotning asosiy maqsadi — elektron xatlarda spamni aniqlashda turli sun’iy intellekt (SI) usullarining samaradorligini solishtirish. Tadqiqot jarayoni bir nechta bosqichlarda amalga oshirildi: ma’lumotlarni yig‘ish, oldindan qayta ishlash, xususiyatlarni tanlash, klassifikatsiya modellari yaratish, natijalarni baholash va solishtirish.

TAHLIL VA NATIJALAR

Ushbu bo‘limda elektron xatlarda spamni aniqlash uchun turli sun’iy intellekt (SI) usullarini solishtirish natijalari tahlil qilinadi. Oldindan qayta ishlangan ma’lumotlar to‘plami ustida mashinani o‘rganish modellari sinovdan o‘tkazildi va ularning samaradorlik ko‘rsatkichlari (aniqlik, F1 skor, precision va recall) asosida solishtirildi. Mazkur tahlil joriy ilmiy tadqiqotlar va nashrlangan maqolalar natijalaridan olinadi.

So‘nggi yillarda e mail spam deteksiyasi bo‘yicha ishlab chiqilgan tadqiqotlar ko‘rsatadiki, klassik mashinani o‘rganish algoritmlari va zamonaviy chuqur o‘rganish modellari ikkalasi ham samarali natijalar beradi, biroq ularning kuchli va zaif tomonlari farqlanadi. Misol uchun, Support Vector Machine (SVM) klassifikatori va Logistic Regression klassik usullari, ko‘p hollarda juda yuqori aniqlik darajasiga erishadi. 5 000 dan ortiq xatlarni o‘z ichiga olgan Kaggle ma’lumotlar to‘plamida SVM 99,0 % aniqlik va 0,97 F1 skor bilan eng yuqori ko‘rsatkichlardan birini namoyon etgan, Logistic Regression esa 98,4 % aniqlikka erishganligi qayd etilgan. Shuningdek, BERT kabi transformer asosidagi chuqur o‘rganish modellari kontekstual tahlil imkoniyati tufayli 98,8 % gacha aniqlikka ega bo‘lgan, bu spam va no spam xabarlarni ajratishda yuqori sifatni ko‘rsatadi.

Bundan tashqari, boshqa maqolalarda Random Forest klassifikatori ham o‘tkazilgan solishtirishlarda yuqori natijalar ko‘rsatgani qayd etilgan. Masalan, UCI Spambase ma’lumotlar bazasida Random Forest eng yuqori aniqlik — 94,57 % bilan boshqa klassik modellardan ustun bo‘lgan.[3]

Bu shuni ko‘rsatadiki, daraxt asosidagi ansambl metodlar (Random Forest) o‘rtacha va katta hajmdagi ma’lumotlarda murakkab qaror chegaralarini aniqlashda samarali ishlaydi.

Tadqiqotlarda odatda quyidagi performance metrikalari ishlatiladi: aniqlik (accuracy), F1 skor, precision va recall. Aniqlik umumiy tasniflashning to‘g‘ri bo‘lgan ulushini o‘lchaydi. F1 skor esa precision va recall o‘rtasidagi balansni hisobga olgan umumiy ko‘rsatkich bo‘lib, ayniqsa imbalanced datasetlarda juda muhimdir. Misol uchun, yuqoridagi Kaggle dataset bo‘yicha SVM hamda transformer modellari ham 0,97 dan yuqori F1 skorlarga yetgan[4].

Shi bilan birga, boshqa tadqiqotlarda Random Forest, Logistic Regression va Naive Bayes singari klassik modellarning kombinatsiyasi 97 % gacha aniqlik va 97,5 % F1 skor ko‘rsatganligi aniqlangan[5]. Bu esa klassik metodlarning hali ham sodda va tushunarli bo‘lishiga qaramay, aniq natijalar bera olishini tasdiqlaydi. Shuningdek, ba’zi tadqiqotlarda Gibrid usullar (ensemble technique) qo‘llanilib, stacking yondashuvi bilan aniqlik 98,8 % atrofida bo‘lgan natija qayd etilgan. Ushbu usullar turli modellarning prognozlarini birlashtirib, yakuniy qarorni optimallashtiradi.

Naive Bayes modeli oddiyligi va tezkor ishlashi bilan ajralib turadi, shu bilan birga matnli ma’lumotlarni tezkor qayta ishlashda yaxshi natijalarga erishadi. Biroq murakkab semantik kontekstni chuqur o‘rgana olmaydi, bu holatda chuqur o‘rganish modellari ustun bo‘lishi mumkin. Aksariyat hollarda Naive Bayes tezkor prototip yondashuv sifatida foydalaniladi, lekin F1 skor va precision ko‘rsatkichlari ba’zan chuqur o‘rganish yondashuvlariga nisbatan past bo‘ladi.

Support Vector Machines (SVM) esa yuqori aniqlik va yomon tasniflangan klasslarni yaxshi ajratish imkoniyati bilan ajralib turadi. SVM ko‘pincha kichik va o‘rta hajmdagi datasetlarda eng yaxshi natija beradi, ammo ko‘p parametrli kontekstual matnlar bilan ishlashda chuqur o‘rganish modellari bilan raqobatlashishi qiyin bo‘ladi. Random Forest esa o‘zi ko‘p daraxtlar ansamblidan tashkil topgani uchun, shovqinga chidamliligi va xususiyatlarning ahamiyatini aniqlashda yaxshi yordam beradi, ammo katta datasetlarda hisoblash xarajatlari oshadi. Chuqur o‘rganish modellari, xususan transformer asosidagi modellari, xat matnining kontekstual tuzilishini chuqur o‘rganadi va murakkab semantik bog‘lanishlarni aniqlashda yuqori natijalarga olib keladi. Biroq ularning asosiy kamchiligi — yuqori hisoblash resurslari talabi bo‘lib, real vaqt tizimlarda joriy etish qiyin bo‘lishi mumkin.[6]

XULOSAVIY TAHLIL

Umuman olganda, tadqiqot natijalari shuni ko‘rsatadiki, sun’iy intellekt asosida elektron xatlarni spam va no spamga ajratish bugungi kunda juda aniq va samarali bajariladi. Tegishli SI yondashuvlarni qo‘llash orqali aniqlik 90 % dan yuqori ko‘rsatkichlar erishilgan, ba’zan 99 % darajagacha ham boradi

Klassik algoritmlar sodda va tezkor ishlashi bilan ajralib, kichik datasetlarda juda yaxshi natijalarga ega bo‘lishi mumkin. Shu bilan birga, ansambl va chuqur o‘rganish modellari murakkab va kontekstual xabarlarda yanada yuqori aniqlikka erishadi. Shuning uchun puxta tizimlar ko‘pincha bir nechta yondashuvni birlashtiradi, bu esa samaradorlikni yanada oshiradi.

Tadqiqot natijalari elektron pochta tizimlarida spam filtrini optimallashtirish, foydalanuvchi tajribasini yaxshilash va kiberxavfsizlikni oshirish bo‘yicha amaliy tavsiyalar ishlab chiqishda asosiy yo‘nalishlarni aniqlaydi.

FOYDALANILGAN ADABIYOTLAR:

1. Kumar S., Garg P. Spam Email Detection Using Machine Learning Techniques // International Journal of Computer Applications. – 2021. – Vol. 175, No. 30. – P. 25–33.

2. Ahmed M., Mahmood A. N., Hu J. A Survey of Network Anomaly Detection Techniques // Journal of Network and Computer Applications. – 2016. – Vol. 60. – P. 19–31.

3. IJICIS. Spam Detection in Emails Using Machine Learning Techniques: A Review. – 2023. – URL: https://ijicis.journals.ekb.eg/article_406689.html

4. IJCT Journal. Machine Learning for Email Spam Detection. – 2022. – URL: https://ijctjournal.org/machine-learning-email-spam-detection/?utm_source=chatgpt.com

5. ResearchGate. Spam Detection in Emails Using Machine Learning Techniques: A Review. – 2023. – URL: https://www.researchgate.net/publication/391203243_Spam_Detection_in_Emails_Using_Machine_Learning_Techniques_A_Review?utm_source=com

6. Ensemble Methods for Spam Detection in Emails. – 2023. – URL: https://link.springer.com/article/10.1007/s10207-023-00756-1?utm_source=.com

Global Science Publication