

## SUN'IY INTELLEKT YORDAMIDA SPAMNI ANIQLASH JARAYONI

**Normatov Ibroximali Xolmamatovich**

Mirzo Ulug'bek nomidagi O'zbekiston milliy universiteti "Axborot hafsizligi"  
kafedrası professori, fizika-matematika fanlari doktori

**Atajonov Muzaffar Ne'matjon o'g'li**

Jaloliddin Manguberdi nomidagi harbiy-akademik litsey aniq fanlar qism boshlig'i  
E-mail: [muzaffar19910627@gmail.com](mailto:muzaffar19910627@gmail.com)

### ANNOTATSIYA

Ushbu maqolada sun'iy intellektni machine learning (mashinani o'rganish) usuli yordamida ma'lumotlarni spamdan himoya qilish jarayonlari keltirilgan bo'lib, shu bilan birga machine learning algoritmlari tahlil qilingan. Maqola so'nggida machine learning yordamida spamlarni aniqlashning bir necha usullaridan namunalar keltirilgan hamda ularni taqqoslab baholangan.

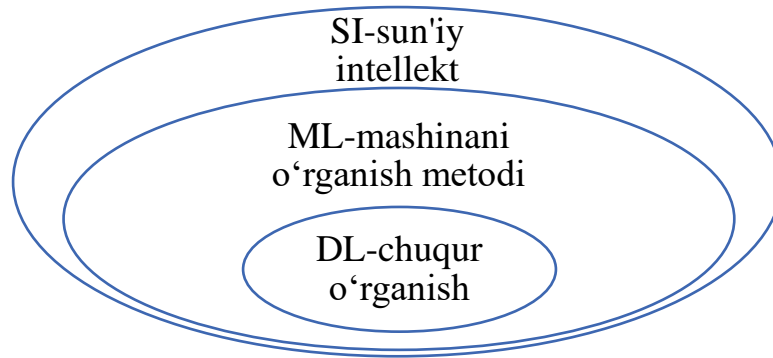
**Kalit so'zlar:** Sun'iy intellekt, machine learning, preprocessing, model choice, decision trees, random forests, support vector machines, neural networks.

### KIRISH

Bugungi kunda kelib barcha soxalarda sun'iy intellektga bo'lgan talab ortib bormoqda. Sun'iy intellekt fan va texnika, qishloq ho'jaligi va boshqa barcha fan atrmoqlarida actual yo'nalishlardan biri bo'lib, IT tarmoqlarining barcha murakkab tizimdagi masalalarni ochib berishda, jumladan sun'iy intellekt yordamida davlat va jamiyatning standart ma'lumotlar xavfsizligini ta'minlashda muhim rol o'ynamoqda. Aktual masalalrdan kelib chiqqan holda ushbu maqolada sun'iy intellekt yordamida spamlarni aniqlash hamda ma'lumotlarni spamlardan himoyalash jarayonlari keltirilgan.

### ASOSIY QISM

Uchta atamani qaraymiz. Bular mashinali o'rganish ya'ni machine learning (ML), sun'iy intellekt ya'ni artificial intelligence (SI) va chuqur o'rganish ya'ni deep learning (DL). Bularni bir biridan farqini quyidagi Venn diagrammasi yordamida tushintiramiz:



**1-rasm. MLni SI va DLdan farqi**

1-rasmdan ko‘rinib turibdiki ML metodi sun‘iy intellektning bir qismi ekan. O‘z navbatida DL ham sun‘iy intellektni bir qismi va ML ni ham aynan bir qismi ekan.

Sun‘iy intellekt (SI) – hisoblash mashinalari (HM) ning insonlarga intellektli bo‘lib ko‘rinadigan narsalarni qilishiga imkon beradigan kontseptsiyalar haqidagi fan.

Mashinali o‘qitish – bu kompyuter uchun dasturlashtirilmagan narsalarni chiqarish yoki bajarish qobiliyati. Mashinali o‘qitish kelajak haqida bashorat qilishni ta’kidlasada, sun‘iy intellekt odatda qaror qabul qilish uchun kompyuterlarni dasturlashga e’tibor beradi. Agar siz odamga o‘xshash xatti-harakatlarni o‘z ichiga olgan aqlli dasturdan foydalansangiz, bu sun‘iy intellekt bo‘lishi mumkin. Biroq, agar parametrlar ma’lumotlardan avtomatik ravishda o‘rganilmasa (yoki olinmasa), bu mashinani o‘rganish emas.

Ma’lumotlarni qayta ishlashda ishlatiladigan elektron pochta spamini aniqlashning keng tarqalgan usullari:

1. Preprocessing: bu ma’lumotlarni tozalash va normallashtirish, to‘xtash so‘zlarini yo‘q qilish va ma’lumotlarning o‘lchovlilikini kamaytirish va modellarning aniqligini oshirish uchun so‘zlarni to‘xtatish kabi ishlarni o‘z ichiga oladi.

2. Identifying and extracting pertinent elements: Ma’lumotlardan jo‘natuvchining elektron pochta manzili, mavzu satri va xabar mazmuni kabi tegishli elementlarni aniqlash va ajratib olish. Keyinchalik, ushbu atributlar yordamida mashinani o‘rganish modellarini o‘rgatish mumkin.

3. Model training (modelni o‘rgatish): Ushbu bosqichda decision trees (qaror daraxtlari), random forests (tasodifiy o‘rmonlar), support vector machines (qo‘llab-quvvatlovchi vektor mashinalari) yoki neural networks (neyron tarmoqlar) kabi mashinalarni o‘rganish modellari olingan xususiyatlardan foydalangan holda o‘qitiladi. Odatda, modellarni o‘rgatish uchun spam yoki haqiqiy deb tasniflangan elektron pochta xabarlarining katta ma’lumotlar to‘plamidan foydalaniladi.

4. Modelni baholash: Ushbu bosqichda trening davomida ishlatilmagan boshqa ma’lumotlar to‘plamidan etiketlangan elektron pochta xabarlarini o‘qitilgan

modellarning aniqligini sinash uchun ishlatiladi. Bu modellar elektron pochta xabarlarini haqiqiy yoki spam sifatida to'g'ri aniqlashiga ishonch hosil qiladi.

Umuman olganda, ma'lumotlarni qayta ishlash elektron pochta orqali spamni aniqlashning muhim bosqichidir, chunki bu aniqlikni yaratishga imkon beradi.

Quyidagi jadval orqali mavjud bo'lgan ba'zi texnik va usullarni ko'rsatamiz:

1-jadval.

Existing Techniques	Existing Methods
<ul style="list-style-type: none"><li>• <input type="checkbox"/> Syntactical</li><li>• <input type="checkbox"/> ELM - Extreme Learning Machine</li><li>• <input type="checkbox"/> SLFN - Single Hidden FeedForward Neural Networks</li><li>• <input type="checkbox"/> SVM - Support Vector Machine</li><li>• <input type="checkbox"/> BRR- Burst Review Ratio</li></ul>	<ul style="list-style-type: none"><li>• <input type="checkbox"/> Texical</li><li>• <input type="checkbox"/> Amalgam</li><li>• <input type="checkbox"/> Stylometric</li><li>• <input type="checkbox"/> N-grams model</li><li>• <input type="checkbox"/> RAVP - Ratio of Amazon Verified Purchase</li><li>• <input type="checkbox"/> RD - Rating Deviation</li><li>• <input type="checkbox"/> RCS - Review Content Similarity</li></ul>

Spam tasnifiga kiruvchi xabarlar belgilari va xususiyatlarni topish uchun elektron pochta orqali spamni aniqlashda ma'lumotlarni yig'ish elektron pochta xabarlariga ulangan ma'lumotlarni olish va tahlil qilish lozim. Yuboruvchining elektron pochta manzili, xabarning mazmuni, mavzu satri va har qanday havolalar yoki qo'shimchalar ushbu ma'lumotlarga kiritilishi mumkin.

Mashinani o'rganish algoritmlari nonspam spam-xabarlarini muvaffaqiyatli aniqlash uchun har bir xabar turining xususiyatlarini tushunish uchun spam va nonspam elektron pochta xabarlarining ulkan ma'lumotlar to'plamida tez-tez o'qitiladi. Ma'lumot yig'ish jarayoni algoritmlarga spam va qonuniy elektron pochta xabarlarini aniqroq farqlash va yangi spam-strategiyalarga moslashish imkonini berganligi sababli, spamni aniq va muvaffaqiyatli aniqlash tizimlarini yaratish juda muhimdir.

Sun'iy intellekt va mashinani o'rganish (Machine learning) yordamida elektron pochta orqali spamni aniqlash modelini tayyorlashda odatda quyidagi qadamlar ishtirok etadi:

**1. Ma'lumotlarni yig'ish va tayyorlash (Data gathering and preparation):** modelni o'rgatish uchun spam va qonuniy elektron pochta xabarlarining katta va xilma-xil ma'lumotlar to'plami talab qilinadi. Ma'lumotlar to'plami yig'ilgandan so'ng, begona yoki ortiqcha ma'lumotlardan xalos bo'lish uchun uni tozalash va qayta ishlash kerak bo'lishi mumkin.

2. **Xususiyatlarni ajratib olish (Feature extraction):** ma'lumotlar to'plangan va tayyorlangandan so'ng, modelni o'rgatish uchun ishlatiladigan xususiyatlar har bir elektron pochta olinishi kerak. Bunga jo'natuvchining elektron pochta manzili, xabarning mazmuni va mavzu satri, xabarning asosiy qismi va birlashtirilgan har qanday fayl yoki havolalar kabi ma'lumotlar kirishi mumkin.

3. **Model tanlovi (Model choice):** elektron pochta spamini aniqlash uchun qaror daraxtlari (decision trees), tasodifiy o'rmonlar (random forests), qo'llab-quvvatlovchi vektorli mashinalar (support vector machines) yoki neyron tarmoqlar (neural networks) kabi turli xil mashinalarni o'rganish usullari va modellaridan foydalanish mumkin. Shaxsiy loyiha ehtiyojlari va ma'lumotlar to'plamining xususiyatlari qaysi model ishlatilishini aniqlaydi.

4. **Namunaviy trening (Model training):** elektron pochtalardan olingan xususiyatlarning etiketli ma'lumotlar to'plamidan foydalanib, tanlangan model keyinchalik o'qitiladi. Maqsad yangi, ko'rinmaydigan elektron pochta xabarlarini spam yoki qonuniy deb ajratish uchun modelni aniq o'rgatishdir.

5. **Modelni baholash (Model evaluation):** model o'qitilgandan so'ng, uning qanchalik yaxshi ishlashi va qanchalik aniqligini ko'rish uchun uni baholash kerak.

**Natija va xulosa:** Endi baholash parametrlarini o'rganamiz. Bular quyidagicha bo'ladi:

**True Negative Rate** = True Negative / True Negative + False Positive

**False Negative Rate** = False Negative / False Negative + True Positive

**False Positive Rate** = False Positive / False Positive + True Negative.

**Precision Formula:** True Positive / True Positive + False Positive

**Recall Formula:** True Positive / True Positive + False Negative

**F- Measure Formula** = Precision.Recall / Precision + Recall

**Accuracy Formula (aniqlik formulasi)** = True Positive Rate + True Negative Rate / True Positive Rate + True Negative Rate False Positive Rate + False Negative Rate.

### FOYDALANILGAN ADABIYOTLAR RO'YXATI: (REFERENCES)

1. **T.A. Xo'jaqulov, N.T.Malikova** «Sun'iy intellekt». Toshkent-2019. 4-8 betlar.
2. **I.N.Normatov, M.N.Atajonov** «Logistik regressiya usuli yordamida spam xabarlarini aniqlash». Raqamli texnologiyalar va sun'iy intellektni rivojlantirishning zamonaviy holati nomli ilmiy konferensiya, Guliston, 2022-yil.
3. **Dr. P.Selvarani, A. Kanagalashmi, M. Keerthika** «**A survey of email spam detection using artificial intelligence**». Journal of Emerging Technologies and Innovative Research, 2023-yil.
4. **M.N.Atajonov**, «Machine learning metodi yordamida elektron pochta spamini aniqlash». O'zbekiston Milliy universiteti talabalar va ilmiy-tadqiqotchilarining ilmiy konferensiyasi, Toshkent, 2022-yil.
5. [www.ziyonet.uz](http://www.ziyonet.uz)