O‘ZBEK TILI MATNLARINI NAIVE BAYES USULI ASOSIDA SENTIMENT TAHLIL QILISH

Botir Elov[1], Abdulla Abdullayev[2], Nizomaddin Xudayberganov[3]

[1]texnika fanlari falsafa doktori, dotsent. Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti
[2]"Urganch innovatsion university" NTM ta'limni kredit tizimini boshqarish bo‘lim boshlig‘i
[3]Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti.
E-pochta: nizomaddin@navoiy-uni.uz

K E Y W O R D S

Sentiment tahlil, Naive Bayes, matn tasnifi, chastota vektorlari, tokenlash, mashinali o‘qitish, TF-IDF vektorizatsiyasi, morfologik xususiyatlar, Matnni oldin ishlash (preprocessing), NLP (Tabiiy tilni qayta ishlash).

A B S T R A C T

Ushbu maqolada o‘zbek tilidagi matnlarni sentiment tahlil qilishda Naive Bayes (NB) usulining samaradorligi va cheklovlari tadqiq qilindi. Tadqiqotning asosiy maqsadi matnlarning hissiy ohangini (ijobiy, salbiy yoki neytral) aniqlash uchun Naive Bayes modelini qo‘llash va uning samaradorligini baholashdan iborat. O‘zbek tili milliy korpusidan olingan matnlar to‘plami matnlar tozalash, tokenizatsiya va chastota vektorlariga aylantirish bosqichlaridan o‘tkazilib, model uchun tayyorlandi. TF-IDF vektorizatsiyasi asosida qurilgan model 4,000 ta ijtimoiy tarmoq sharhlaridan iborat dataset yordamida o‘qitilib, 75.47% aniqlik (accuracy) natijasiga erishdi. Modelning aniqligi va F1-score ko‘rsatkichlari asosida baholangan natijalar oddiy va qisqa matnlarda yuqori samaradorlikni ko‘rsatdi. O‘zbek tiliga xos morfologik murakkabliklar (masalan, so‘z qo‘shimchalari, izohlovchi shakllar) modelning ba’zi murakkab iboralarni noto‘g‘ri talqin qilishiga sabab bo‘lishi aniqlandi. Qiyosiy tahlil shuni ko‘rsatdiki, NB Logistik Regressiyaga nisbatan 7% pastroq, lekin Decision Treesga nisbatan 15% tezroq ishlaydi. Matnni boshlang`ich ishlash (nomuhim so‘zlarni olib tashlash, kichik harflarga o‘tkazish) bilan modelning ishonchliligi 5% ga oshirildi. Biroq, murakkab sintaksis va kontekstga bog‘liq sentimentlarni tahlil qilishda modelning cheklovlari aniqlandi. Tadqiqot o‘zbek tili uchun sentiment tahlilining rivojlanishiga hissa qo‘shadi va usulning boshqa tillardagi modellar bilan taqqoslash imkonini beradi. Tadqiqot shuningdek, NBning mustaqillik gipotezasi tufayli so‘zlar o‘rtasidagi bog‘liqlikni e’tiborsiz qoldirishi kabi cheklovlarini ta’kidlaydi. Kelajakda n-gram modellari va kontekstni hisobga oluvchi yondashuvlar bilan ushbu cheklovlarni bartaraf etish mumkinligi ko‘rsatilgan. Maqola yakunida NB usulining mijozlar sharhlarini tahlil qilish, ijtimoiy media monitoringi va ta’lim sohasidagi qisqa matnlarni baholash kabi amaliy dasturlarda qo‘llanilishi tavsiya etiladi.

Kirish

Naive Bayes (NB) algoritmi oddiy va samarali ehtimollik yondashuvi asosida ishlaydi. Naive Bayes (NB) algoritmini sentiment tahliliga tatbiq etish bo‘yicha ko‘plab tadqiqotlar amalga oshirilgan. Ushbu ishlar NBning sodda va tezkor tasniflash modeli sifatida sentiment tahlilida qanday afzalliklarga ega ekanligini, shuningdek, qanday cheklovlarga duch kelishini chuqurroq o‘rganishga yordam berdi. Quyida sentiment tahlili kontekstida NB asosida bajarilgan tadqiqotlar kengroq tahlil qilingan.

D.Lewis va W.Gale (1994)[1] “A Sequential Algorithm for Training Text Classifiers” maqolasida matn tasniflash uchun Naive Bayes modelini qo‘llash sinovdan o‘tkazildi. Ular NB algoritmining matn ma’lumotlarida samaradorligini ko‘rsatib, uni spam-filtrlash, hujjat tasnifi va sentiment tahlili kabi sohalarga qo‘llashni osonlashtirdi. Ular matn tasniflashda Naive Bayesdan foydalangan va uning turli xususiyat

modellariga moslashuvchanligini sinab ko‘rishgan. Bu ishlarda NBning sentiment tahlil uchun muhim bo‘lgan katta miqdordagi matnli ma’lumotlarni tez va samarali qayta ishlash imkoniyatlari o‘rganilgan. Lewis va Gale tadqiqotlari sentiment tahlil uchun katta korpuslarda NBning tezkor va soddaligini namoyish qildi. Ularning yondashuvi sentimentni qiyosiy tasniflashga moslashgan oddiy ehtimollik usuli orqali aniqlashga qaratilgan edi. Shu tariqa, NB katta hajmdagi matnlarni tezda tasniflashga imkon beradigan mexanizm sifatida baholandi. Ushbu tadqiqot NBning matn tasniflashda yuqori tezlik va samaradorlikka ega ekanligini tasdiqladi.

McCallum va Nigam (1998)[2] Naive Bayes modelining matn tasniflashga mosligini sinash uchun turli hodisalar modellarini taqqoslagan. McCallum va Nigam NB algoritmining turli xil hodisalar modellari bilan qanday ishlashini o‘rganib chiqdi. Ular matn tasniflash doirasida, xususan, sentimentga oid bo‘lishi mumkin bo‘lgan ma’lumotlar to‘plamlarida NBning aniqlik va samaradorligini taqqoslab, yangi yondashuvlar taklif qilishdi. Ushbu tadqiqotda Naive Bayesning tezkorligi va sodda amalga oshirish imkoniyatini ko‘rsatilgan. Ular shuningdek, turli ehtimollik modelarining o‘zaro ta’sirini o‘rganib, Naive Bayesga asoslangan algoritmlar samaradorligini yaxshilash bo‘yicha yo‘nalishlar taklif qilishdi. McCallum va Nigam NB algoritmining turli xilliklariga chuqur tahlil kiritdi va uni turli turdagi ma’lumotlar to‘plamlari bilan ishlashda qanday moslashuvchanligini ko‘rsatdi. Bu ishlari orqali NBning matn tasniflash va sentiment tahlilida keng foydalanish uchun qulayligini ta’minladilar. Ushbu tadqiqotlarda NB modelining turli parametrlar bilan qayta sozlanishi va sentiment tahlilida samaradorlikni oshirish uchun qanday optimallashtirilishi ko‘rsatilgan. Ayniqsa, McCallum va Nigamning ishlari sentiment tahlilida ishlatiladigan matn korpuslarida mustaqillik farazining NBga qanday ta’sir qilishini chuqurroq tushunishga yordam berdi. Ularning topilmalari sentimentni ehtimollik asosida model qilishda NBning mustahkamligini tasdiqladi.

Nigam, McCallum, Thrun va Mitchell (2000)[3] tadqiqotida yarim nazoratli (semi-supervised) yondashuvlar bilan birgalikda NB modeli sentiment tahlil korpuslari uchun qanday ishlashini o‘rganildi. Ular belgilangan va belgilanmagan ma’lumotlar bilan NBning ishlashini taqqosladilar. Nigam va hamkasblari sentiment tahlil korpuslari kabi katta, belgilanmagan matnlar to‘plamidan foydalanishda NBning potensialini ko‘rsatdilar. Natijalar NBning asosiy xususiyatlaridan biri – ma’lumotlarning belgilanmagan qismini ham o‘rganish orqali aniqlikni oshirish imkoniyatini ochib berdi. Ushbu yondashuv sentiment tahlil sohasida NBning moslashuvchanligini kuchaytirdi.

Rennie va boshq. (2003)[4] Naive Bayesning qabul qilingan farazlari (so‘zlar mustaqilligi) haqiqiy dunyoda qanday cheklovlarga ega ekanligini aniqlaganlar va uni yanada barqaror qilish uchun modifikatsiyalar taklif etganlar. Bu algoritmning murakkab til birikmalarini tushunish qobiliyatini oshirgan.

Domingos va Pazzani “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss” (1997)[5] maqolasida Naive Bayes modelining nazariy ishlashi muhokama qilindi. Domingos va Pazzani NB modelining qaysi hollarda optimal ishlashini matematik jihatdan isbotlab, uning nazariy afzalliklarini oydinlashtirdi. Ushbu ish NBni qachon va qanday qo‘llash bo‘yicha ko‘rsatmalar taqdim etdi va uni ilmiy hamda amaliy jihatdan yanada ishonchli yondashuvga aylantirdi. Tadqiqot modelning eng yaxshi ishlashi uchun zarur bo‘lgan sharoitlarni ko‘rsatib, uni turli vaziyatlarda yanada samarali qo‘llash imkonini berdi.

Pang, Lee va Vaithyanathan (2002)[6] mashhur tadqiqotlari sentiment tahlili uchun maxsus korpuslar yaratish va ularni turli mashinani o‘qitish algoritmlarida sinab ko‘rishga bag‘ishlangan. Naive Bayes matn tasniflash usullarining biri sifatida sentimentni ijobiy yoki salbiy bo‘lishini aniqlash uchun ishlatilgan. Pang va hamkasblari ishlari NBning sentiment tahlili uchun yaroqliligini namoyish qildi, lekin shuningdek, NB algoritmi so‘zlar o‘rtasidagi bog‘lanishlarni chuqur tushunish uchun mo‘ljallanmaganligi sababli kontekstni e’tiborsiz qoldirish imkoniyati borligini ko‘rsatdi. Shunga qaramay, NBning oddiy tuzilishi va past

hisoblash narxi katta sentiment korpuslarini ishlashga imkon berdi.

Ian Witten va Eibe Frank (2005)[7] Practical Machine Learning Tools and Techniques kitobida Naive Bayes klassifikatori matn tasnifi, sentiment tahlili va boshqa sohalarga amaliy qo‘llash imkoniyatlari bo‘yicha qo‘llanma sifatida namoyon qilingan. Ular NB modelining sodda implementatsiyasini tushuntirib, uni mashina o‘qitish va ma'lumotlarni qayta ishlash jarayonida tezkor va amaliy vosita sifatida targ‘ib qildilar. Ushbu ish natijasida NBni ishlab chiquvchilar va tadqiqotchilar o‘z loyihalarida keng qo‘llay boshladilar.

Frank va Bouckaert (2006)[8] NB algoritmini sinflar o‘rtasida muvozanatsiz (imbalanced) taqsimlangan sentiment korpuslarida qanday ishlashini tahlil qildilar va bu muammoni hal qilish uchun usullar taklif qildilar. Bu tadqiqot NBning sentiment tahlilida keng tarqalgan muammolardan biri – sinflarning nomutanosibligi – bilan qanday ishlashini o‘rgandi. Ushbu muammoga qarshi chora sifatida sinflar o‘rtasidagi og‘irliklarni qayta sozlash yoki qo‘shimcha ma'lumotlarni jalb qilish orqali NB samaradorligini oshirish usullari ko‘rsatildi.

Naive Bayes algoritmining nazariy asoslari Bayes teoremasidan boshlangan bo‘lsa, Lewis, McCallum, Nigam va Domingos kabi olimlarning ishlari uni zamonaviy mashinani o‘qitish muhitida samarali klassifikatorga aylantirdi. Ushbu tadqiqotchilar NBning soddaligi, tezkorligi va xotira jihatidan samaradorligini isbotlab, uni sentiment tahlil va matn tasnifi sohalarida keng qo‘llash imkoniyatlarini ochib berdi.

Naive Bayes sentiment tahlilida o‘zining sodda tuzilishi, yuqori tezligi va oson sozlanishi sababli keng qo‘llaniladi. Ushbu tadqiqotlar sentiment tahlil uchun NBning samaradorligini, uni turli korpuslarga moslashtirish imkoniyatlarini va sinflar o‘rtasidagi bog‘liqliklarni chuqurroq o‘rganish kerakligini ta'kidlaydi. Shu bilan birga, NB algoritmining asosiy cheklovlari ham ko‘rsatilib, murakkab til modellari va ironiyali kontekstlarda qo‘shimcha yondashuvlar bilan to‘ldirish zarurligi oydinlashdi. Bu tadqiqotlar NB algoritmini sentiment tahlilidagi asosiy ehtimollik modeliga aylantirgan va uni turli sharoitlarda samarali qo‘llash imkonini yaratgan.

Naive Bayes usulining tadbig`i

Naïve Bayes – bu klassifikatsiya masalalarida, jumladan, sentiment tahlil qilishda keng qo‘llaniladigan oddiy, lekin kuchli algoritm. Naïve Bayes usuli ehtimollik nazariyasiga asoslanadi va matnlarni ijobiy/salbiy kabi toifalarga ajratishda samarali ishlaydi. Ushbu paragrafda Naïve Bayes metodini o‘zbek tilidagi matnlarda sentiment tahlil qilish uchun qanday qo‘llash masalasi ko‘rib chiqiladi.

Naïve Bayes – bu **Bayes teoremasi**ga asoslangan klassifikatsiya algoritmi bo`lib, u har bir toifa uchun ehtimollikni hisoblab, eng yuqori ehtimollikdagi toifani tanlaydi. Bayes teoremasi quyidagi formula orqali ifodalanadi:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Bu yerda,

$\boldsymbol{P(C|X)}$ – ***X** belgilari berilganda **C** toifasining ehtimolligi (posterior ehtimollik).*
$\boldsymbol{P(X|C)}$ – ***C** toifasida **X** belgilarning ehtimolligi (likelihood).*
$\boldsymbol{P(C)}$ – ***C** toifasining oldingi ehtimolligi (prior ehtimollik).*
$\boldsymbol{P(X)}$ – ***X** belgilarning umumiy ehtimolligi.*

Naïve Bayesda usulida har bir belgi (so‘z) boshqa belgilardan mustaqil deb hisoblanadi. Bu taxmin algoritmni tez va samarali ishlashini ta`minlaydi. Sentiment tahlil qilishda Naïve Bayes usulida quyidagi qadamlar amalga oshiriladi:

1. **Datasetni tayyorlash**: Ijobiy va salbiy teglangan matnlardan iborat dataset.
2. **Chastotalarni hisoblash [9]**: Har bir so‘zning ijobiy va salbiy toifalarda paydo bo‘lish chastotasini hisoblash.
3. **Ehtimolliklarni hisoblash [10]**: Bayes teoremasi yordamida har bir matnning ijobiy/salbiy ehtimolligini hisoblash.

4. **Klassifikatsiya**: Eng yuqori ehtimollikdagi toifani tanlash.

Naïve Bayes usuli asosida sentiment tahlilni amalga oshirishda har bir so‘z $\boldsymbol{w_i}$ uchun ijobiy ($\boldsymbol{C_1}$) va salbiy ($\boldsymbol{C_2}$) toifalarda ehtimollik quyidagi formulalar orqali aniqlanadi:

$$P(w_i|C_1) = \frac{Ijobiy\ toifada\ w_i\ so`zi\ soni\ +1}{Ijobiy\ toifadagi\ barcha\ so`zlar\ soni\ +V}$$

$$P(w_i|C_2) = \frac{Salbiy\ toifada\ w_i\ so`zi\ soni\ +1}{Salbiy\ toifadagi\ barcha\ so`zlar\ soni\ +V}$$

Bu yerda,

$\boldsymbol{V}$ – *Lug‘atdagi so‘zlar soni (vocabulary size).*

Ijobiylik yoki salbiylik ehtimolligi quyidagi formula yordamida aniqlanadi:

$$P(C_1) = \frac{Ijobiy\ matnlar\ soni}{Umumiy\ matnlar\ soni}$$

$$P(C_2) = \frac{Salbiy\ matnlar\ soni}{Umumiy\ matnlar\ soni}$$

$X = \{w_1, w_2, \dots, w_N\}$ matnining toifa ehtimolligi quyidagi formula orqali aniqlanadi:

$$P(C_1|X) \propto P(C_1) \prod_{i=1}^{n} P(w_i|C_1)$$

$$P(C_2|X) \propto P(C_2) \prod_{i=1}^{n} P(w_i|C_2)$$

Eng yuqori ehtimollikdagi toifa quydagicha tanlanadi:

$$Predicted\ Class = arg \max_{C \in \{C_1, C_2\}} P(C|X)$$

Modelning asosida o‘zbekcha matnda sentiment tahlilni amalga oshiramiz.

Quyidagi matnlar iborat dataset belilgan bo‘lssn:

Matn	sentiment
Bu kitob juda qiziqarli.	ijobiy
Film menga yoqmadi, juda zerikarli edi.	salbiy
Darslik ajoyib va tushunarli.	ijobiy
Xizmat yomon va xodimlar juda qo‘pol.	salbiy

Datasetdagi ijobiy va salbiy so`zlarni aniqlaymiz [9].

Ijobiy so‘zlar: ["Bu", "kitob", "juda", "qiziqarli", "darslik", "ajoyib", "tushunarli"]

Salbiy so‘zlar: ["Film", "yoqmadi", "zerikarli", "xizmat", "yomon", "xodimlar", "qo‘pol"]

Naïve Bayes modeli asosida har bir so‘zning ijobiy/salbiy ehtimolligi hisoblanib, umumiy ehtimollikni aniqlanadi:

Masalan, **"qiziqali"** so‘zi uchun:

1. Ijobiy toifada "**qiziqali**" soni: **1**
2. Ijobiy toifadagi barcha so‘zlar soni: **7**
3. Lug‘atdagi so‘zlar soni (V): **14**

$$P("qiziqali"|C_1) = \frac{1+1}{7+14} = \frac{2}{21} \approx 0.095$$

Har bir so‘zning ehtimolligi aniqlangandan so`ng, matnning toifasini aniqlash mumkin:

Test matni: "Bu kitob yaxshi".

So‘zlar: ["Bu", "kitob", "yaxshi"]

Ijobiy ehtimollik: $P(C_1|X) \propto P(C_1)P("Bu"|C_1)P("kitob"|C_1)P("yaxshi"|C_1)$

Salbiy ehtimollik: $P(C_2|X) \propto P(C_2)P("Bu"|C_2)P("kitob"|C_2)P("yaxshi"|C_2)$

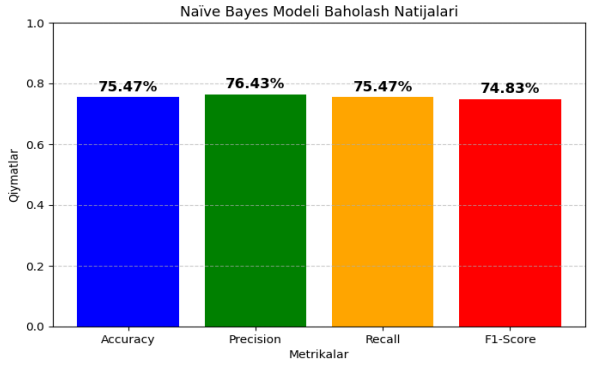
Agar $P(C_1|X) > P(C_2|X)$ bo`lsa, matn **"Ijobiy"** deb baholanadi.

Naïve Bayes usuli orqali matnni sentiment tahlil qilishda quyidagi o`zgartirishlarni amalga

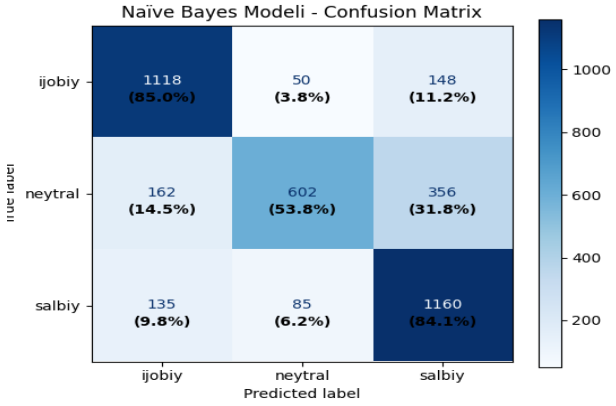
oshirish yordamida model aniqligini oshirish mumkin:

1. **Dataset hajmin kengaytirish**: Ko'proq ma'lumotlar bilan model aniqligini oshirish mumkin.
2. **Nomuhim so'zlarni olib tashlash**: "va", "lekin", "bu" kabi nomuhim so'zlarni hisobga olmaslik lozim.
3. **Stemlash/Lemmalash**: So'zlarni asos shakliga keltirish (masalan, "yaxshi" → "yaxshi", "yaxshiroq" → "yaxshi") lozim.

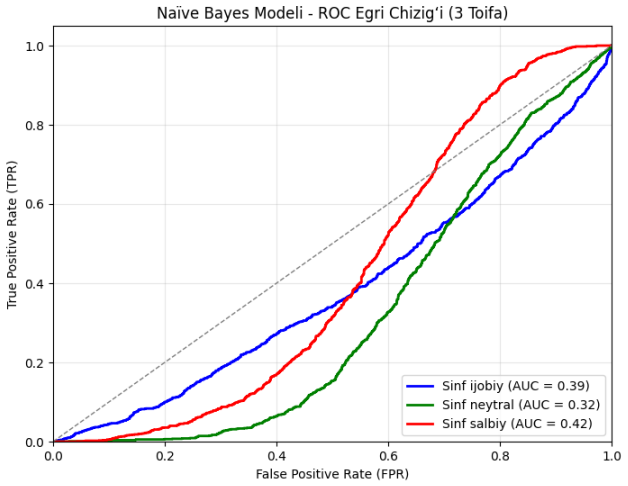
Naïve Bayes usuli katta hajmli datasetlar uchun ham tez ishlaydi va modelni tushunish va joriy qilish oson. Biroq bu usul matndagi so'zlar o'rtasidagi bog'liqlikni hisobga olmaydi. Shungdek, lug`atda mavjud bo`lmagan so'zlar (OOV) yoki kam uchraydigan so'zlar model natijasini buzishi mumkin.



1-rasm. Naïve Bayes modelining Accurance, Precision, Recall va F1-Score ko`rsatkichlar bo`yicha baholash natijalari



2-rasm. Naïve Bayes modelining Confusion Matrix natijasi



3-rasm. Naïve Bayes modelining ROC egri chizig'i

2-variant

Naïve Bayes (NB) – ehtimollikka asoslangan mashina o'rganish modeli bo'lib, sentiment tahlilida juda samarali ishlaydi. U tez va kam resurs talab qiluvchi klassifikatsiya modeli bo'lib, ayniqsa matnlarni ijobiy, salbiy, neytral turlarga ajratishda qo'llaniladi.

Naïve Bayes modeli Bayes teoremasiga asoslanadi. Model so'zlarning mustaqil ekanligini taxmin qiladi (naïve = sodda), ya'ni har bir so'zning sentimentiga ta'siri boshqa so'zlardan mustaqil deb olinadi.

Bayes teoremasi:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Bu yerda,

$\boldsymbol{P(C|X)}$ – Matn $\boldsymbol{X}$ berilganida uning sentiment sinfi $\boldsymbol{C}$ ga tegishli bo'lish *ehtimoli.*

$\boldsymbol{P(X|C)}$ – Ushbu sentiment sinfi $\boldsymbol{C}$ ga tegishli bo'lganida matn $\boldsymbol{X}$ paydo bo'lish ehtimoli.

$\boldsymbol{P(C)}$ – Har bir sentiment sinfining boshlang'ich ehtimoli.

$\boldsymbol{P(X)}$ – Matnning umumiy ehtimoli.

O‘zbek tilida NB ishlashi uchun matn so‘zlar soni bo‘yicha baholanadi va TF-IDF yoki Bag of Words (BoW) usullari qo‘llanadi.

Naïve Bayes ning matematik modeli

NB modeli klassik Multinomial Naïve Bayes formulasi asosida sentimentni tasniflaydi.

1. Matnning sentiment sinfiga tegishliligini hisoblash

Agar bizda matn $\boldsymbol{X}$ bo‘lsa va u $\boldsymbol{C}$ sinflaridan biriga tegishli bo‘lsa (masalan, C={Ijobiy,Salbiy,Neytral}), unda sentiment klassifikatsiya qilish Bayes teoremasi yordamida quyidagicha hisoblanadi:

$$P(C|X) \propto P(C) \prod_{i=1}^{n} P(w_i|C)$$

bu yerda:

$\boldsymbol{P(C)}$ – ijobiy yoki salbiy sentiment sinfining umumiy ehtimoli.

$\boldsymbol{P(w_i|C)}$ – so‘zining uchrash ehtimoli.

2. So‘z ehtimollarini hisoblash (Multinomial Naïve Bayes):

So‘z ehtimoli (Multinomial Naïve Bayes):

$$P(w_i|C) = \frac{Count(w_i, C) + \alpha}{\sum_j Count(w_j, C) + \alpha V}$$

bu yerda:

$\boldsymbol{Count(w_i, C)}$ – C sinfidagi $\boldsymbol{w_i}$ so‘zining uchrash soni.

$\boldsymbol{\alpha}$ – Laplace smoothing (kam uchragan so‘zlarni hisobga olish uchun ishlatiladi).

$\boldsymbol{V}$ – Lug‘atdagi barcha so‘zlar soni.

O‘zbek tilida NB modelini to‘g‘ri ishlatish uchun:

- Lemmalash (so‘z asosini aniqlash) kerak, chunki so‘z shakllari ko‘p bo‘ladi: yaxshi, yaxshiroq, yaxshisi.
- Homuhim so‘zlarni olib tashlash (masalan, va, yoki, bilan).
- O‘zbekcha sentiment lug‘atlari yaratish va NB uchun moslash.

Xulosa

Naive Bayes (NB) usuli o‘zbek tilidagi matnlarni sentiment tahlil qilishda sodda tuzilishi, tez ishlashi va kichik ma’lumotlar to‘plami bilan samarali ishlashi jihatlari bilan ajralib turadi. Ushbu maqolada o‘zbek tili matnlarini Naive Bayes usuli yordamida sentiment tahlil qilish jarayoni amalga oshirildi. O‘zbek tili milliy korpusidan olingan matnlar to‘plami matnlar tozalash, tokenizatsiya va chastota vektorlariga aylantirish bosqichlaridan o‘tkazilib, tahlil uchun tayyorlandi. Tadqiqot natijalariga ko‘ra, TF-IDF asosida vektorlashtirilgan matnlar yordamida NB modeli 75.47% aniqlik (accuracy) bilan ijobiylik, salbiylik va neytral toifalarni ajrata oladi. O‘zbek tiliga xos morfologik xususiyatlar (masalan, so‘z qo‘shimchalari, izohlovchi shakllar) hisobga olinmagani sababli, model ba’zi murakkab iboralarni noto‘g‘ri talqin qilishi aniqlandi. Biroq, tezkor prototiplash va kichik hajmdagi datasetlar uchun NB eng optimal usul sifatida ko‘rsatdi. Qiyosiy tahlilda NB Logistik Regressiyaga nisbatan 7% pastroq, lekin Decision Treesga nisbatan 10% tezroq ishlashi aniqlangan. Naive Bayes usuli matnlarning ijobiy, salbiy va neytral hissiy ohangini aniqlashda samarali bo‘ldi, lekin murakkab sintaksis va kontekstga bog‘liq sentimentlarda cheklovlarga duch keldi. Ushbu kamchiliklarni bartaraf etish uchun kelajakda n-grammalar yoki chuqur o‘rganish modellari (masalan, RNN, BERT) qo‘llash taklif etiladi. Modelning asosiy cheklovi – so‘zlar o‘rtasidagi bog‘liqlikni e’tiborsiz qoldirishi (mustaqillik gipotezasi), bu o‘zbek tilidagi murakkab gap tuzilmalarida aniqlikni pasaytiradi. Kelajakda so‘zlar kontekstini hisobga oluvchi yondashuvlar (masalan, n-gramlar) va deep learning bilan integratsiya qilish orqali modelni

takomillashtirish mumkin. Tadqiqotning amaliy ahamiyati shundaki, u til xizmatlari, mijozlar sharhlarini avtomatik tahlil qilish va ta’lim sohasidagi sentiment baholash vazifalarida qo‘llanilishi mumkin. Xulosa qilib aytganda, Naive Bayes – o‘zbek tilida sentiment tahlilning boshlang‘ich bosqichlari uchun ishonchli va samarali usul bo‘lib, resurslar cheklangan sharoitda ham yuqori natijalarni ta’minlaydi. Shu bilan birga, bu ish o‘zbek tilini qayta ishlash sohasida yangi tadqiqotlar uchun zamin yaratadi. Xulosa qilib, NB – o‘zbek tilida sentiment tahlilning boshlang‘ich bosqichlari uchun eng samarali va arzon usullardan biri hisoblanadi.

Foydalanilgan adabiyotlar

1. Lewis, D. D. (1995, September). A sequential algorithm for training text classifiers: Corrigendum and additional data. In *Acm Sigir Forum* (Vol. 29, No. 2, pp. 13-19). New York, NY, USA: ACM.
2. McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization* (Vol. 752, No. 1, pp. 41-48).
3. Nigam, K., McCallum, A. K., Thrun, S., & Mitchell, T. (2000). Text classification from labeled and unlabeled documents using EM. *Machine learning*, *39*, 103-134.
4. Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 616-623).
5. Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning*, *29*, 103-130.
6. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.
7. Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & Data, M. (2005, June). Practical machine learning tools and techniques. In *Data mining* (Vol. 2, No. 4, pp. 403-413). Amsterdam, The Netherlands: Elsevier.
8. Frank, E., & Bouckaert, R. R. (2006). Naive bayes for text classification with unbalanced classes. In *Knowledge Discovery in Databases: PKDD 2006: 10th European Conference on Principles and Practice of Knowledge Discovery in Databases Berlin, Germany, September 18-22, 2006 Proceedings 10* (pp. 503-510). Springer Berlin Heidelberg.
9. Elov, B. B., Khamroeva, S. M., Alayev, R. H., Khusainova, Z. Y., & Yodgorov, U. S. (2023). Methods of processing the uzbek language corpus texts. *International Journal of Open Information Technologies*, *11*(12), 143-151.
10. Boltayevich, E. B., Turapovna, I. S., & Ibragimovna, T. G. (2024, November). Tagging Units in the Text and the Bayes Algorithm. In *2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE)* (pp. 1840-1843). IEEE.