

NAIV BAYES YORDAMIDA FILM SHARHLARINI SENTIMENT TAHLILI

¹ Mamasoliyev Asilbek, O'zbekiston¹ Farg'ona davlat texnika universiteti,
kompyuter injenering IT – servis yo'nalishi
talabasi asilbekmi80@gmail.com

Annotatsiya. Ushbu maqolada Naiv Bayes klassifikatori yordamida film sharhlarini sentiment tahlil qilish jarayoni yoritiladi. Sentiment tahlili matnni ijobiy yoki salbiy kayfiyatga ajratishdan iborat bo'lib, tabiiy tilni qayta ishlash (NLP) va mashinali o'qitishning eng mashhur qo'llanmalardan biridir. Maqolada Naiv Bayes algoritmining nazariy asosi, ma'lumotlarni tayyorlash bosqichlari, modelni o'qitish jarayoni va baholash usullari batafsil bayon qilinadi. [1] Shuningdek, eksperiment natijalari va amaliy qo'llanilish bo'yicha tavsiyalar beriladi.

Kalit so'zlar: *Naiv Bayes, sentiment tahlili, mashinali o'qitish, NLP, film sharhlari, klassifikatsiya, Bag-of-Words, TF-IDF.*

Annotation. This article highlights the process of sentiment analysis of movie reviews using the Naive Bayes classifier. Sentiment analysis involves classifying text as positive or negative and is one of the most popular applications of natural language processing (NLP) and machine learning. The article provides a detailed explanation of the theoretical basis of the Naive Bayes algorithm, data preparation stages, model training process, and evaluation methods. [1] It also presents experimental results and offers practical recommendations.

Keywords: Naive Bayes, sentiment analysis, machine learning, NLP, movie reviews, TF-IDF classification, Bag-of-Words

Kirish. Hozirgi kunda internetda foydalanuvchilar tomonidan yaratilgan matnlar miqdori keskin oshgan. Bloglar, ijtimoiy tarmoqlar, yangiliklar platformalari va film sharhlari kabi manbalarda katta miqdorda matnli ma'lumot to'planadi. Bu ma'lumotlardan samarali foydalanish uchun sentiment tahlili kabi avtomatlashtirilgan

texnikalar zarur. Sentiment tahlili matnning emotsional rangini aniqlash orqali biznes, marketing, media tahlili va ko'plab boshqa sohalarda muhim ahamiyat kasb etadi. [2] Sentiment tahlilining eng mashhur usullaridan biri — Naiv Bayes klassifikatori bo'lib, u soddaligi, tezligi va samaradorligi bilan ajralib turadi. Ayniqsa, matn klassifikatsiyasida, statistik xususiyatlari tufayli keng qo'llaniladi. Ushbu maqolada film sharhlari misolida Naiv Bayes algoritmidan foydalanib sentiment tahlilini amalga oshirishning amaliy jarayoni ko'rib chiqiladi. [3] Naiv Bayes algoritmining nazariy asoslari: Naiv Bayes algoritmi Bayes teoremasiga asoslanadi. Bayes teoremasi quyidagicha ifodalanadi:

$$P(C|X) = (P(X|C) * P(C)) / P(X)$$

Bu yerda:

C — klass (ijobiy yoki salbiy);

X — matndagi xususiyatlar to'plami;

$P(C|X)$ — X xususiyatlariga ko'ra C klassga tegishlilikini ehtimoli;

$P(X|C)$ — C klass uchun xususiyatlarning sharti ehtimoli,

$P(C)$ — klassning umumiy ehtimoli.

Naiv Bayesning asosiy "naivligi" shundan iboratki, u xususiyatlarning bir-biriga bog'liq emas deb hisoblaydi. Matn klassifikatsiyasida bu taxmin har doim to'g'ri bo'lmasa-da, amalda juda samarali natijalar beradi.[4] Naiv Bayesning eng keng tarqalgan turlari: Multinomial Naive Bayes, Bernoulli Naive Bayes, Gaussian Naive Bayes. Matn tahlilida, ayniqsa, film sharhlarida Multinomial Naive Bayes eng keng ishlatiladi, chunki u so'z chastotalari bilan ishlaydi. Ma'lumotlarni tayyorlash jarayoni Naiv Bayes algoritmini qo'llashdan oldin ma'lumotlarni to'g'ri tayyorlash juda muhim. Jarayon quyidagi bosqichlardan iborat.[5] 1. Ma'lumotlarni to'plash; Film sharhlari odatda IMDB, Rotten Tomatoes yoki boshqa ochiq manbalardan olinadi. Ushbu sharhlar ijobiy va salbiy kategoriyalarga ajratilgan bo'ladi. 2. Ma'lumotlarni tozalash: Matnni tozalash jarayoni quyidagi amallarni o'z ichiga oladi: Katta harflarni kichik harfga o'tkazish, punktuatsiyalarni olib tashlash, stop-so'zlarni chiqarib tashlash (masalan: "va", "bu", "ham"), Lemmatizatsiya yoki stemming. 3.

Matnni vektorlashtirish: kompyuter matnni to'g'ridan-to'g'ri tushunmaydi, shuning uchun uni sonli shaklga o'tkazish kerak. Buning uchun quyidagi metodlardan foydalaniladi: Bag-of-Words (BoW) modeli, TF-IDF (Term Frequency – Inverse Document Frequency), TF-IDF modelida kam uchraydigan, lekin ahamiyatli so'zlar yuqori baholanadi. Bu sentiment tahlilida samarali natijalar beradi. Modelni o'qitish: Naiv Bayes modeli quyidagi bosqichlarda o'qitiladi: 1. o'quv va test to'plamlariga bo'lish, odatda 70% ma'lumot o'quv uchun, 30% test uchun ajratiladi. [6]

2. Modelni o'qitish: Multinomial Naive Bayes modeli BoW yoki TF-IDF vektorlari yordamida o'qitiladi.

3. Modelni baholash

Baholash uchun quyidagi metrikalardan foydalaniladi:

Aniqlik (Accuracy);

To'g'rilik (Precision);

Qamrov (Recall);

F1-score;

Test ma'lumotlarida yuqori aniqlik Naiv Bayes algoritmining matn tahlilida samarali ekanligini ko'rsatadi. Matnni tozalashning kengaytirilgan texnikalari:

Film sharhlarida ko'pincha kinoya, hazil, o'xshatishlar, metaforalar va kontekstga boy gaplar uchraydi. Shu sababli tozalash jarayoni oddiy bo'lmasligi mumkin.

Kengaytirilgan metodlar: emoji va maxsus belgilarni semantik kategoriya sifatida kodlash (“😊” → ijobiy).

URL, hashtag va mentionlarni alohida token sifatida ko'rib chiqish. Bigram, trigram kabi N-gramlardan foydalanish orqali kontekstni hisobga olish. Lemmatizatsiyani yanada aniqroq vositalar bilan bajarish. Masalan, “not good” bigram sifatida salbiy ma'noni anglatadi, lekin unigramlar asosida bu aniqlanmasligi mumkin. Naiv Bayesning kuchli tomonlari:

Juda tez ishlaydi — katta datasetlarda darhol natija beradi.

Resurs talab qilmaydi — hatto oddiy kompyuterda ham osongina ishlaydi.

Oson tushuniladi — matematik asoslari sodda.

Kam ma'lumot bilan ham yaxshi natija beradi.

Kamchiliklari:

So'zlar mustaqil deb faraz qilinadi — bu har doim to'g'ri emas.

Murakkab kontekstni tushuna olmaydi.

Ironiya, kinoya, metafora kabi murakkab tuzilmalar uni chalg'itadi.

Masalan, "Film shu qadar zo'r ediki, uxlab qoldim" — aslida salbiy gap, lekin ijobiy so'zlar Naiv Bayesni chalkashtirishi mumkin.

Film sharhlarida sentiment tahlilining amaliy qiyinchiliklari

Film sharhlari ko'pincha quyidagi xususiyatlarga ega:

Subyektiv fikrlar.

Kontekstga bog'liq iboralar.

San'atga oid maxsus terminlar.

Kinoya yoki hazil.

Shuning uchun model doim ham aniqlik bilan ishlay olmaydi. Aynan shu sababli ko'plab tadqiqotchilar Naiv Bayesni boshqa algoritmlar bilan birlashtiradi.

Ma'lumotlar balansining ahamiyati: Agar datasetda ijobiy sharhlar salbiy sharhlarga nisbatan ko'proq bo'lsa, model ijobiy klassga og'ishadi. Buning oldini olish uchun:

Undersampling, oversampling, SMOTE texnikalari qo'llaniladi. Real hayotda qo'llanilishi: Film sharhlari sentiment tahlili quyidagilarda ishlatiladi: Kino industriasida avtomatik reyting yaratish. Tavsiya tizimlarini kuchaytirish (Netflix, Amazon Prime). Auditoriya kayfiyatini tahlil qilish va bashoratlash. Marketingda brend obro'sini kuzatib borish. Ijtimoiy tarmoqlarda kontent monitoringi.

Eksperimentlar natijalari

Turli ilmiy ishlarda:

TF-IDF + Multinomial Naive Bayes 88–92% aniqlikka erishgani kuzatilgan.

Bigramlar qo'shilganda natijalar 5–7% oshgan.

Lemmatizatsiya stemmingga qaraganda yaxshiroq ko'rsatkich bergan.

Bu Naiv Bayesning sodda bo'lsa-da, juda kuchli ekanini ko'rsatadi.

Эксперимент natijalari. Turli tadqiqotlarda Naiv Bayes modeli film sharhlari sentiment tahlilida 80–90% aniqlikka erishishi ko'rsatilgan. [7] TF-IDF bilan birgalikda ishlatilganda aniqlik yanada ortadi. Shuningdek, kichik va o'rta hajmdagi datasetlar uchun Naiv Bayes juda tez ishlaydi. Bu esa uni real vaqtli sentiment tahlili tizimlarida qo'llash imkonini beradi.

Xulosa. Naiv Bayes algoritmi soddaligi, tezligi va yuqori aniqligi sababli film sharhlari sentiment tahlilida keng qo'llaniladi. Matnni tozalash va to'g'ri vektorlashtirish model natijalariga sezilarli ta'sir ko'rsatadi. Ushbu maqolada Naiv Bayes asoslari, amaliy qo'llanilishi, modelni o'qitish jarayoni va baholash natijalari batafsil tahlil qilindi. Kelajakda sentiment tahlilini yanada kuchaytirish uchun Naiv Bayesni boshqa mashinali o'qitish usullari bilan birlashtirish yoki neyron tarmoqlar bilan solishtirish mumkin. [8]

Foydalanilgan adabiyotlar:

- 1.Manning C., Raghavan P., Schütze H. Introduction to Information Retrieval.
- 2.Liu B. Sentiment Analysis and Opinion Mining.
- 3.Bird S., Klein E., Loper E. Natural Language Processing with Python.
- 4.McCallum A., Nigam K. A Comparison of Event Models for Naive Bayes Text Classification.
- 5.Jurafsky D., Martin J. Speech and Language Processing.
- 6.Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques.
- 7.Pang B., Lee L. Thumbs Up? Sentiment Classification Using Machine Learning Techniques.
- 8.Goodfellow I., Bengio Y., Courville A. Deep Learning.