

IJTIMOIY TARMOQDAGI MATNLI MA'LUMOTLARNI TAHLIL QILISH  
MODELLARI

*Yunusov Azizjon Abdunazar o'g'li*

*Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti  
Dasturiy Injineriingi fakulteti bitiruchisi*

Ijtimoiy tarmoqlar, insonlar o'rtasidagi muloqotni yanada kengaytirib, katta hajmdagi matnli ma'lumotlar oqimini yaratdi. Ushbu ma'lumotlarni tahlil qilish va tushunish zamonaviy texnologiyalar yordamida amalga oshiriladi. Ushbu tezisda ijtimoiy tarmoqlardagi matnli ma'lumotlarni tahlil qilish modellarini ishlab chiqish bosqichlari va ulardan foydalanish haqida batafsil bayon qilinadi.

Model ma'lumotlar to'plamini tayyorlashdan boshlanadi, shu jumladan ma'lumotlarni yig'ish, etiketlash, va tozalash. Trening uchun vakolatli va muvozanatli ma'lumotlar to'plami talab qilinadi, chunki bu ko'rinmas ma'lumotlardan yaxshi umumlashtira oladigan hissiyotlarni tahlil qilish modeli yaratishga yordam beradi. Dastlabki ishlov berish kichik harflar, tokenizatsiya, stopword olib tashlash va maxsus belgilarni ajratish kabi texnikalar yordamida amalga oshiriladi, bu esa standartlashtirilgan va toza matn ma'lumotlarini taqdim etish uchun ishlatiladi .

Ijtimoiy tarmoq ma'lumotlari odatda murakkab va turli manbalardan keladi. Ushbu ma'lumotlarni to'plash uchun Python dasturlash tilidagi BeautifulSoup va Scrapy kabi vositalardan foydalanish mumkin. Ma'lumotlarni tozalash bosqichida esa matnli ma'lumotlardan nozik belgilar, emodji va boshqa keraksiz ma'lumotlar olib tashlanadi. Ushbu jarayon matnni yanada tahlil qilish uchun to'g'ri formatga keltiradi .

Ma'lumotlarni etiketlash uchun qo'lda va avtomatik usullar qo'llaniladi. Qo'lda etiketlash ko'proq aniqlikni ta'minlasa-da, ko'p vaqt talab etadi. Shu sababli, avtomatik etiketlash algoritmlaridan, masalan, Natural Language Toolkit (NLTK) va spaCy kabi kutubxonalaridan foydalanish mumkin. Ushbu kutubxonalar matnlarni avtomatik ravishda toifalashda yordam beradi va tezlikni oshiradi .

Ma'lumotlar to'plami tayyor bo'lgach, turli modellardan foydalanish mumkin. Eng keng tarqalgan modellar qatoriga Naive Bayes, Support Vector Machines (SVM), va Deep Learning modellaridan LSTM va Transformerlar kiradi. Ushbu modellar ma'lumotlarni tahlil qilishda yuqori aniqlik va samaradorlikni ta'minlaydi. Jumladan, LSTM va Transformer modellarini qo'llash, uzun muddatli bog'lanishlar va kontekstual ma'noni yaxshiroq tushunish imkonini beradi .

Modelni o'qitish jarayonida, ma'lumotlar trening va sinov to'plamlariga bo'linadi. Trening to'plami modelni o'qitish uchun ishlatiladi, sinov to'plami esa uning samaradorligini tekshirish uchun qo'llaniladi. Ushbu bosqichda cross-validation kabi texnikalardan foydalanib, modelning umumiy lashtirish qobiliyatini baholash mumkin. Modelni baholash uchun aniqlik (accuracy), aniqlik (precision), xotira (recall), va F1-score kabi o'lchovlardan foydalaniladi .

Natijalarni tahlil qilish uchun matn tahlilining turli xil metrikalarini qo'llash kerak. Ushbu metrikalar yordamida modelning qanchalik yaxshi ishlashini baholash mumkin. Natijalarni yanada aniqroq ko'rsatish uchun turli xil grafiklar va jadvallar qo'llanishi mumkin.

1-jadval: Turli modellarning aniqlik darajalari

Model	Aniqlik (%)	Aniqchilik (%)	Xotira (%)	F1-score (%)
Naive Bayes	85	82	78	80
SVM	88	85	83	84

Model	Aniqlik (%)	Aniqchilik (%)	Xotira (%)	F1-score (%)
LSTM	90	88	87	87.5
Transformer (BERT)	92	91	89	90

Jadvalda keltirilgan modellarni o'zaro taqqoslash uchun turli manbalardan[7,8] olingan empirik natijalar umumlashtirilib, moslashtirildi.

Ushbu tezisda ijtimoiy tarmoqdagi matnli ma'lumotlarni tahlil qilish modellarining asosiy bosqichlari va ulardan foydalanish usullari bayon qilindi. Ma'lumotlar to'plamini tayyorlash, tozalash, etiketlash, model yaratish, o'qitish va natijalarni tahlil qilish jarayonlari ko'rib chiqildi. Ushbu texnikalar va modellar yordamida ijtimoiy tarmoqdagi matnli ma'lumotlarni samarali tahlil qilish mumkin.

#### References:

1. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
2. Russell, M. A. (2013). Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. O'Reilly Media.
3. Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing. Pearson.
4. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
5. Chollet, F. (2017). Deep Learning with Python. Manning Publications.
6. Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.
7. Naive Bayes va Support Vector Machines (SVM) modellarining natijalari Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective.
8. LSTM va Transformer (BERT) modellarining natijalari Chollet, F. (2017). Deep Learning with Python va Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing.