

Sosial Media Resurslarının Pasiyent-dərman Seqmentində Rəylərin Sentiment Analizi

Məsumə Məmmədova¹, Zərifə Cabrayilova², Nərgiz Şıxəliyeva³

ETN İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹mmg51@mail.ru, ²djabrailova.z@mail.ru, ³shikhaliyeva.nara@gmail.com

Xülasə— Məqalədə sosial medianın tibb seqmentində toplanan dərmanlarla bağlı pasiyent rəylərinin sentiment analizi əsasında pasiyent-xəstəlik-dərman əlaqələrinin aşkarlanması və xəstəliklər üzrə dərmanların effektivliyinin təyini məsələsinə baxılmışdır. Dərmanlar haqqında pasiyent rəylərinin analizini aparmaq üçün Kaggle platformasının *drugsComTest_raw.csv* tibbi məlumat bazasından istifadə edilmiş, leksikon əsaslı sentiment analiz, statistik metodlar və maşın təlimi alqoritmləri tətbiq edilmişdir. Rəylərdə aşkarlanmış pasiyent-xəstəlik, xəstəlik-dərman əlaqələri üzrə rəylərin sentiment analizi aparılmış, pasiyent-dərman seqmentində toplanan rəylər əsasında: 1) ən çox hansı xəstəliklərlə bağlı dərmanlara müraciət edildiyi; 2) hər bir xəstəliklə bağlı daha çox hansı dərmana üstünlük verildiyi kimi məsələlərin həll imkanı göstərilmişdir. Bu məsələlərin integrasiyasından alınan nəticələr əsasında müxtəlif tibbi vəziyyətlər (xəstəliklər) üçün dərmanların effektivliyi ilə bağlı ictimai rəy identifikasiya edilmişdir.

Açar sözlər— tibbi sosial media; sentiment analiz; maşın təlimi; dərmanlar haqqında.

I. GİRİŞ

Son illərdə rəqəmsal texnologiyaların sürətli inkişafı və sosial media platformalarının geniş yayılması səhiyyə sahəsində yeni məlumat mənbələrinin meydana çıxmasına səbəb olmuşdur. Övvəllər pasiyentlərin sağlamlıqla bağlı təcrübələri əsasən həkim-pasiyent görüşlərində, rəsmi sorğularda və ya klinik tədqiqatlar vasitəsilə qeydə alınırıdysa, bu gün artıq pasiyentlər öz gündəlik sağlamlıq vəziyyətlərini, dərmanlara qarşı reaksiyalarını, xəstəliklərlə bağlı düşüncələrini və s. könüllü şəkildə sosial mediada paylaşırlar. Bu platformalardan olan X, Facebook, Reddit, Instagram, DailyStrength və digər forumlar pasiyentlərin xəstəlik simptomları, diaqnoz təcrübələri, müalicə metodları və dərman istifadəsinə dair fikirlərini real zamanda paylaşmağa imkan verir [1, 2]. Bu platformalar pasiyentlər üçün təkcə sosial dəstək məkanı deyil, eyni zamanda digər oxşar problemlərlə qarşılaşan şəxslərlə əlaqə qurmaq və alternativ müalicə yolları haqqında məlumat toplamaq üçün də effektiv vasitədir. Belə rəylər tibb sahəsi üçün dəyərli məlumat mənbəyi olmaqla yanaşı, xəstəliklər, dərman reaksiyaları və müalicə nəticələri haqqında qeyri-rəsmi, lakin əhəmiyyətli məlumatların toplanmasına şərait yaradır [3].

[4-6]-də ictimai rəyi nəzərə almaqla daha obyektiv və şəffaf qərarların qəbulu üçün tibbi sosial medianın həkim-pasiyent, tibb müəssisəsi-pasiyent seqmentində toplanmış informasiyaya istinad edilməklə həkim, tibb müəssisəsinin fəaliyyətinin

qiymətləndirilməsi məsələlərinin həllinə baxılmışdır. Son dövrlərdə dərmanların effektivliyinin qiymətləndirilməsi, səhiyyənin bu seqmentində daha şəffaf qərarların qəbuluna həsr edilmiş tədqiqatlarda ictimai rəyin nəzərə alınması məqsədilə pasiyent rəylərinin analizinə diqqət artmışdır [7-9].

Pasiyentlərin müxtəlif platformalarda dərman istifadəsi ilə bağlı təcrübələr, həmçinin müalicə üsullarının effektivliyi və yan təsirləri haqqında yazılı şəkildə paylaşıdığı əhəmiyyətli məlumatlar çoxsaylı və müxtəlif formatlarda olduğundan onların təhlili mürəkkəb və çətindir. Bununla belə bu rəylərin içində strukturlaşdırılmamış məlumat və emosional ifadələr mövcud olduğundan, onların düzgün şəkildə emalı xüsusi yanaşmalar tələb edir. Bu məqsədlə təbii dilin emalı (*ing. natural language processing (NLP)*) və maşın təlimi (*ing. machine learning (ML)*) texnologiyalarından istifadə edilir [10]. NLP metodları, mətni analiz edərək məzmunu çıxarmağa və strukturlaşdırmağa imkan verir. Maşın təlimi alqoritmləri düzgün və avtomatlaşdırılmış nəticələr əldə etməyə kömək edir. Bu texnologiyalar vasitəsilə istifadəçilərin rəylərindəki emosiyaları və fikirləri təhlil etmək, rəylərin müsbət, mənfi və neytral olduğunu müəyyənləşdirmək mümkündür. Bu proses ictimai rəyin daha aydın və dəqiq şəkildə anlaşılmasına kömək edir.

II. MƏSƏLƏNİN QOYULUŞU

Hazırkı tədqiqatın əsas məqsədi pasiyent rəylərinin sentiment analiz əsasında dərmanların effektivliyinin müəyyənləşdirilməsindən ibarətdir. Bu məqsədlə pasiyentlərin dərmanların effektivliyi, yan təsirləri və istifadəsi ilə bağlı sosial mediada paylaşıdığı rəyləri (mətn tipli şərhləri) analiz olunur.

Məqsədə nail olmaq üçün aşağıdakı məsələlərin həlli nəzərdə tutulur:

1. Pasiyent-dərman seqmentində olan rəylərin toplanması və ilkin emalı.
2. Pasiyent rəylərinin sentiment analizi və nəticələrin dəqiqliyinin yoxlanılması.
3. Pasiyent-xəstəlik əlaqələrinə görə müxtəlif tibbi vəziyyətlər üzrə rəylərin nizamlanması.
4. Xəstəlik-dərman əlaqələrinə görə tibbi vəziyyətlər üzrə dərmanların nizamlanması.
5. Eyni bir tibbi vəziyyət üzrə dərmanların reytinginin təyini.

III. MƏSƏLƏNİN HƏLLİ

A. Pasiyent-dərman seqmentində olan rəylərin toplanması və ilkin emalı

Pasiyentlərin dərmanlara bağlı rəylərinin analizi üçün *Kaggle* şirkətinin *drugsComTest_raw.csv* adlı açıq məlumat bazası seçilmişdir [11]. Məlumat bazası aşağıdakı əsas sütunlarından ibarətdir (şəkil 1):

- dərmanın adı: pasiyentlərin müxtəlif tibbi vəziyyətlərdə, xəstəliklərdə istifadə etdiyi dərmanlardır. Hər bir vəziyyət, xəstəlik üçün çoxlu sayda dərman istifadə oluna bilər;
- vəziyyət: dərmanın istifadə edildiyi tibbi vəziyyəti (xəstəliyi) təsvir edir.
- rəy: pasiyentlərin dərmanla bağlı mətn tipli şərhərini əhatə edir;
- istifadə sayı: dərmanın nə qədər istifadəçi tərəfindən istifadə olduğunu göstərir.

Bu mərhələdə xam mətn məlumatlarının analizi üçün verilənlərin ilkin emalı həyata keçirilir, məlumatlar istifadə üçün yararlı formaya salınır. Bu mərhələdə aşağıdakı əməliyyatlar yerinə yetirilir [12]:

1. mətnin kiçik hərflərə çevrilməsi: mətnə yer alan bütün böyük hərflər kiçik hərflərə çevrilir, beləliklə "Quick" və "quick" kimi fərqli yazılımların eyni sözü ifadə etməsi təmin edilir;
2. Dürğu işarələrinin silinməsi: mətnin içindəki nöqtələr, vergüllər və digər dürğu işarələri təhlil zamanı maneə yaratmamaq məqsədilə mətnədən çıxarılır;

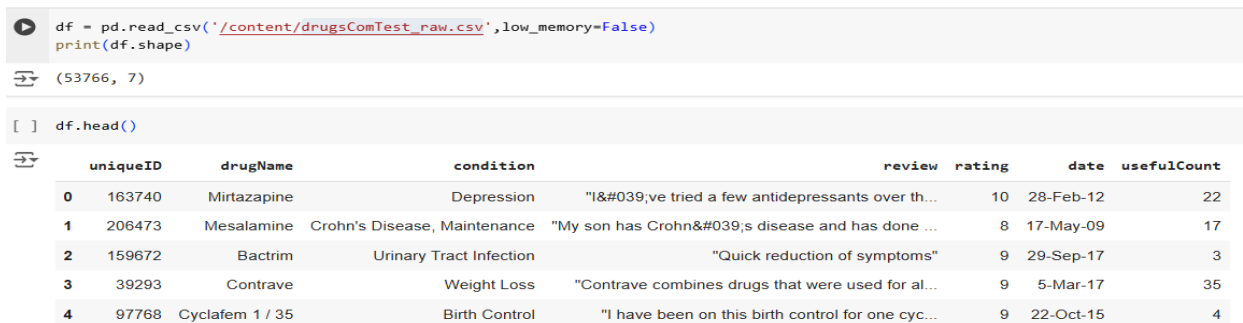
3. Rəqəmlərin silinməsi: məndəki rəqəmlər silinir, çünki təhlil söz əsaslı verilənlərə əsaslanır.

4. Stop sözlərinin silinməsi: bu mərhələdə çox istifadə olunan, lakin sentiment analizinə faydası olmayan "and", "the", "is" kimi sözlər məndən çıxarılır. Bu, mətnin daha yığcam və təhlilə uyğun hala gəlməsinə kömək edir;

5. Lemmatizasiya: sözlər kök formasına salınır, məsələn, "running" və "runs" sözləri "run" şəklində normallaşdırılır. Bu, təhlil prosesində eyni kökə sahib müxtəlif formaların eyni anlayış kimi qəbul edilməsini təmin edir.

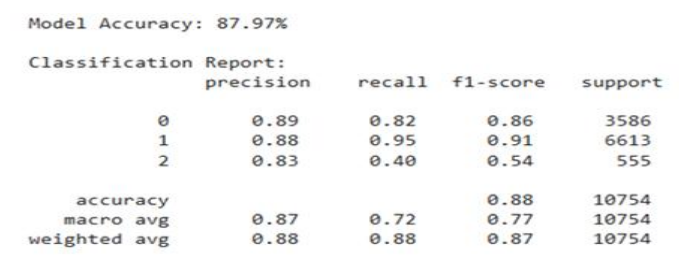
B. Pasiyent rəylərinin sentiment analizi və nəticələrin dəqiqliyinin yoxlanılması

Bu mərhələdə leksikon əsaslı sentiment analiz yanaşması tətbiq edilib. Bu yanaşmada hər bir rəyin tonallığı (müsbət, mənfi və ya neytral) *TextBlob* kitabxasından istifadə edilməklə müəyyən edilir. *TextBlob*-un sentiment analiz funksiyası, məndəki emosional ifadələri təhlil edərək həmin rəyin polyarlıq (müsbət və ya mənfi istiqamətdə) dərəcini hesablayır. Polyarlıq dərəcəsinə əsasən rəyin müsbət, mənfi və ya neytral olduğu müəyyən edilir. Sonra məlumatlar train və test hissələrinə ayrılır və mətnlər TF-IDF ilə vektorlaşdırılır. *Logistik Regressiya* modeli təlim edilir və test verilənləri üzərində proqnozlaşdırma aparılır. Modelin dəqiqliyi və qiymətləndirilməsi ilə nəticələr dəyərləndirilir (şəkil 2). Bu mərhələ, mətn tipli rəylərin sentiment analizi modelinin yaradılmasını və onların qiymətləndirilməsini təmin edir.



	uniqueID	drugName	condition	review	rating	date	usefulCount
0	163740	Mirtazapine	Depression	"I've tried a few antidepressants over th...	10	28-Feb-12	22
1	206473	Mesalamine	Crohn's Disease, Maintenance	"My son has Crohn's disease and has done ...	8	17-May-09	17
2	159672	Bactrim	Urinary Tract Infection	"Quick reduction of symptoms"	9	29-Sep-17	3
3	39293	Contrave	Weight Loss	"Contrave combines drugs that were used for al...	9	5-Mar-17	35
4	97768	Cyclafem 1 / 35	Birth Control	"I have been on this birth control for one cyc...	9	22-Oct-15	4

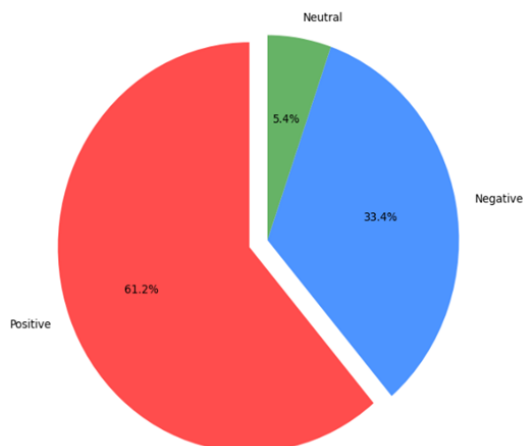
Şəkil 1. Sentiment analiz üçün Kaggle şirkətinin drugsComTest_raw.csv adlı açıq məlumat bazasına giriş



Model Accuracy: 87.97%				
Classification Report:				
	precision	recall	f1-score	support
0	0.89	0.82	0.86	3586
1	0.88	0.95	0.91	6613
2	0.83	0.40	0.54	555
accuracy			0.88	10754
macro avg	0.87	0.72	0.77	10754
weighted avg	0.88	0.88	0.87	10754

Şəkil 2. Logistik Regressiya modelinin dəqiqlik nəticəsi.

Aparılan analiz nəticəsində ümumilikdə 53766 rəydən 32920 pozitiv, 17956 neqativ və 2890 neytral olduğu müəyyən olunmuşdur. Nəticələrin faizlə ifadəsinin vizual görüntüsü Şəkil 3-də verilmişdir.



Şəkil 3. Kaggle şirkətinin drugsComTest_raw.csv adlı açıq məlumat bazasındakı pasiyentlərin dərmanlarla bağlı rəylərinin sentiment analizindən alınan nəticələrin vizual görüntüsü.

C. Pasiyent-xəstəlik əlaqələrinə görə müxtəlif tibbi vəziyyətlər üzrə rəylərin nizamlanması

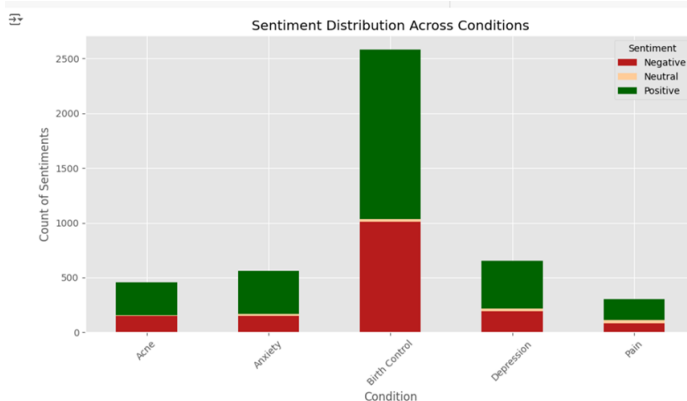
Bu mərhələdə dərman istifadə edən pasiyentlərin rəylərinin hansı tibbi vəziyyətlə (xəstəliklə) bağlı olduğu araşdırılmış, tədqiqat pasiyent-xəstəlik seqmentində davam etdirilmişdir. Bu mərhələdə ən çox rəy toplayan 5 tibbi vəziyyət seçilmiş (Sızanaq, Narahatlıq, Doğum Nəzarəti, Depressiya və Ağrı), onlar üzrə rəylərin SA yerinə yetirilmişdir. Nəticələrin mənfi, neytral və müsbət kateqoriyalara paylanması cədvəl 1-də və bu nəticələrin vizual görüntüsü şəkil 4-də verilmişdir.

CƏDVƏL 1. BEŞ TIBBI VƏZİYYƏT ÜZRƏ İSTİFADƏ EDİLƏN DƏRMANLARLA BAĞLI PASİYENT RƏYLƏRİNİN SENTIMENT ANALİZİNİN NƏTİCƏLƏRİ

Vəziyyət	Pozitiv	Neqativ	Neytral	Umumi
Doğuma Nəzarət	5696	3835	117	9648
Depressiya	2043	936	116	3095
Ağrı	1369	537	194	2100
Narahatlıq	1290	539	79	1908
Sızanaq	1210	608	29	1847

D. Xəstəlik-dərman əlaqələrinə görə tibbi vəziyyətlər üzrə dərmanların nizamlanması

Bu mərhələdə baxılan 5 tibbi vəziyyət (xəstəlik) üzrə istifadə olunan dərmanlar haqqında rəylərin sentiment analizi aparılmış və hər bir vəziyyət üzrə pozitiv rəylər sayına görə dərmanların nizamlanmış siyahısı alınmışdır (cədvəl 2).



Şəkil 4. Dərmanlarla bağlı pasiyent rəylərinin beş tibbi vəziyyət üzrə sentiment analizin vizual görüntüsü.

CƏDVƏL 2. BEŞ TIBBI VƏZİYYƏT ÜZRƏ İSTİFADƏ EDİLƏN DƏRMANLARLA BAĞLI PASİYENT RƏYLƏRİNİN SENTIMENT ANALİZİNƏ ƏSASƏN İLK ÜÇLÜKDƏ OLAN DƏRMANLAR

Vəziyyət	Dərman adı	Pozitiv (%)	Neytral (%)	Neqativ (%)
Sızanaq	Accutane	74.57	0	25.43
	Doxycycline	66.42	3.47	30.11
	Isotretinoin	64.36	0.93	34.71
Narahatlıq	Bupropion	75	0	25
	Gabapentin	74.67	1.33	24
	Escitalopram	72.65	1.68	25.67
Doğuma Nəzarət	Etonogestrel	64.53	0.75	34.72
	Ethinyl estradiol / norethindrone	61.11	0.86	38.03
	Levonorgestrel	57.83	1.55	40.62
Depressiya	Bupropion	71.46	3.35	25.19
	Sertraline	70.77	2.93	26.30
	Escitalopram	67.88	3.85	28.27
Ağrı	Oxycodone	64.88	12.70	22.42
	Tramadol	63.79	8.60	27.61
	Gabapentin	60.53	10.53	28.95

E. Xəstəlik-dərman əlaqələrinə görə tibbi vəziyyətlər üzrə dərmanların nizamlanması

C və D mərhələlərində alınan nəticələrin inteqrasiyası əsasında dərmanların effektivliyi ilə bağlı ictimai rəy aşağıdakı kimi identifikasiya olunur:

1. Sızanaq: Sızanaq üçün dərmanlar arasında, *Accutane* dərmanı ən çox müsbət rəy sayına və ən az neqativ rəy faizinə malikdir, bu baxımdan effektivlik baxımından digər dərmanları qabaqlayır.

2. Narahatlıq: Narahatlıqla bağlı dərmanlar arasında *Bupropion* ən çox müsbət rəyi toplayan dərmandır, bu dərman haqqında heç bir neytral rəy olmadığı məlum olmuşdur.

3. Doğum Nəzarəti: Doğum nəzarəti üçün dərmanlar arasında *Etonogestrel* yüksək müsbət rəy, ən az neytral və neqativ rəy faizinə sahibdir.

4. Depressiya: Depressiya üçün dərmanlar arasında, *Bupropion* ən yaxşı nəticələrə, istifadəçilər tərəfindən daha çox müsbət rəy və ən az neqativ rəy faizinə sahibdir. Pasiyentlərin rəylərinin pozitivlik faizlərinə görə *Sertraline* və *Bupropion* dərmanlarının nəticələri bir-birinə çox yaxın olsa da, ikinci daha çox neqativ rəyə malikdir.

5. Ağrı: Ağrı müalicəsi üçün dərmanlar arasında *Oxycodone* ən yüksək müsbət rəy faizinə və ən az neqativ rəy faizinə malik dərmandır. Bu vəziyyət üzrə *Gabepentin* ən çox müsbət rəy toplayanların sırasında olsa da, ən çox neqativ rəy faizinə sahib dərmandır.

Beləliklə, tibbi sosial medianın pasiyent-dərman seqmentində toplanan rəylər əsasında pasiyentlərin ən çox hansı xəstəliklərlə bağlı dərmanlara müraciət etdikləri, hər bir xəstəliklə bağlı daha çox hansı dərmana üstünlük verildiyi kimi məsələlərin həll imkanı göstərilmişdir. Bu məsələlərin integrasiyasından alınan nəticələr əsasında müxtəlif tibbi xəstəliklər üzrə dərmanların effektivliyi ilə bağlı ictimai rəy identifikasiya edilmişdir.

IV. NƏTİCƏ

Dərmanların effektivliyinin təyini üçün təklif edilən yanaşma əsasında onların mümkün yan təsirlərinin də qiymətləndirilməsi maraq kəsb edir. Belə ki, pasiyent rəyləri vasitəsilə təkcə dərmanların effektivliyi deyil, həmçinin onların yan təsirləri və pasiyentlər tərəfindən necə qəbul edildiyi ilə bağlı ətraflı məlumat əldə etmək mümkündür. Digər tərəfdən, müəyyən xəstəliklərdə istifadə olunan bir sıra dərmanların yüksək müsbət rəy faizinə malik olmasına baxmayaraq onunla bağlı neqativ rəylərdə (hətta az sayda olsa da) ciddi yan təsirlər haqqında məlumatlara rast gəlmək mümkündür. Bu səbəbdən tədqiqatın davamı olaraq neqativ rəylərin daha dərinlən təhlili və onlardakı məlumatlar əsasında dərmanların effektivliyinin qiymətləndirilməsi xüsusi əhəmiyyət kəsb edir.

İSTİNADLAR

- [1] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, "Utilizing social media data for pharmacovigilance: A review," *Journal of Biomedical Informatics*, 2015, vol. 54, pp. 202–212. <https://doi.org/10.1016/j.jbi.2015.02.004>
- [2] A. Nikfarjam, et al., "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features," *Journal of the American Medical Informatics Association*, 2015, vol. 22, no. 3, pp. 671–681. <https://doi.org/10.1093/jamia/ocu041>
- [3] D. Wessel and N. Pogrebnyakov, "Using Social Media as a Source of Real-World Data for Pharmaceutical Drug Development and Regulatory Decision Making," *Drug Saf*, 2024, vol. 47, pp. 495–511. <https://doi.org/10.1007/s40264-024-01409-5>
- [4] M. H. Mammadova and Z. G. Jabrayilova, *Electronic medicine: formation and scientific-theoretical problems*. Bakı: "Information Technologies" Publishing House, 2019. <https://ict.az/uploads/files/E-medicine-monograph-IIT-ANAS.pdf>
- [5] M. H. Mammadova, Z. G. Jabrayilova, and N. R. Shikhaliyeva, "Lexicon-based sentiment analysis of medical data," *Technology Transfer: Fundamental Principles and Innovative Technical Solutions*, pp. 7–10, 2022. <https://doi.org/10.21303/2585-6847.2022.002671>
- [6] M. H. Mammadova, Z. G. Jabrayilova, and A. M. Isayeva, "Conceptual approach to the use of information acquired in social media for medical decisions," *Online Journal of Communication and Media Technologies*, 2020, vol. 10, no. 2, p. e202007.
- [7] D. Cavalcanti and R. Prudêncio, "Aspect-based opinion mining in drug reviews," *Progress in Artificial Intelligence*. EPIA 2017, Lecture Notes in Computer Science, 2017, vol. 10423, pp. 815–827. Springer. https://doi.org/10.1007/978-3-319-65340-2_66
- [8] F. Gräßer, S. Kallumadi, H. Malberg, and S. Zaunseder, "Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning," In *Proceedings of the 2018 International Conference on Digital Health*, 2018, pp. 121–125.
- [9] B. Panda, C. R. Panigrahi, and B. Pati, "Exploratory data analysis and sentiment analysis of drug reviews," *Computación y Sistemas*, 2022, vol. 26, no. 3, pp. 1191–1199.
- [10] G. Alexander, M. Bahja, and G. F. Butt, "Automating large-scale health care service feedback analysis: sentiment analysis and topic modeling study," *JMIR Medical Informatics*, 2022, vol. 10, no. 4, p. e29385.
- [11] Kaggle.com, "UCI ML Drug Review dataset," <https://www.kaggle.com/datasets/jessicali9530/kuchackathon-winter-2018>
- [12] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. E. Barnes, and D. E. Brown, "Text Classification Algorithms: A Survey," *Information*, 2019, vol. 10, no. 4, p. 150. <https://doi.org/10.3390/info10040150>