

## АВТОМАТИЗАЦИЯ И УПРАВЛЕНИЕ INFORMATION SYSTEM AND TECHNOLOGIES

УДК 004

DOI: 10.18413/2518-1092-2022-7-2-0-6

Кнаан А.-Р.  
Польщиков К.А.

**АНАЛИЗ РЕАКЦИИ ОБЩЕСТВЕННОСТИ  
НА ЭЛЕКТРОННОЕ ОБУЧЕНИЕ В ТВИТТЕРЕ**

Белгородский государственный национальный исследовательский университет,  
ул. Победы д. 85, г. Белгород, 308015, Россия  
*e-mail: Kanan.rashed@gmail.com*

### Аннотация

За последние несколько десятилетий из-за взрывного роста социальных сетей, онлайн-ресурсов и сайтов микроблогов, таких как Twitter был приток пользовательского контента. Данные, полученные из этих ресурсов, являются богатым источником информации для интеллектуального анализа данных. Анализ тональности – это актуальная и важная область исследований, которая пытается определить полярность текста. Решающее значение приобрело определение чувств по поводу текущих событий в мире. Статья посвящена интеллектуальному анализу данных в Twitter и определению мнений в отношении электронного обучения. Основное внимание уделяется выявлению настроений из текстов, связанных с электронным обучением, которые публикуются в Twitter. Было извлечено около 3000 твитов и обнаружена полярность этих твитов, а затем визуализированы полученные данные.

**Ключевые слова:** анализ тональности текста; электронное обучение; анализ данных; визуализация; Orange Data Mining

**Для цитирования:** Кнаан А.-Р., Польщиков К.А. Анализ реакции общественности на электронное обучение в Твиттере // Научный результат. Информационные технологии. – Т.7, №2, 2022. – С. 48-54. DOI: 10.18413/2518-1092-2022-7-2-0-6

Кнаан А.-Р.  
Polshchikov K.A.

**ANALYSIS OF PUBLIC REACTION  
TO E-LEARNING ON TWITTER**

Belgorod State National Research University, 85 Pobedy St., Belgorod, 308015, Russia  
*e-mail: Kanan.rashed@gmail.com*

### Abstract

Over the past few decades, due to the explosive growth of social media, online resources, and microblogging sites such as Twitter. There was an influx of user-generated content. The data obtained from these resources is a rich source of information for data mining. Sentiment analysis is a current and important area of research that attempts to determine the polarity of a text. The definition of feelings about current events in the world has become crucial. This article focuses on data mining on Twitter and defining opinions regarding e-learning. The focus is on identifying sentiment from e-learning-related texts that are shared on Twitter. About 3,000 tweets were extracted and the polarity of those tweets was detected, and then visualize the resulting data.

**Keywords:** text sentiment analysis; e-learning; data analysis; visualization; Orange Data Mining

**For citation:** Knaan A.-R., Polshchikov K.A. Analysis of public reaction to e-learning on Twitter // Research result. Information technologies. – Т.7, №2, 2022. – P. 48-54. DOI: 10.18413/2518-1092-2022-7-2-0-6

### ВВЕДЕНИЕ

В наше время электронное обучение и дистанционное обучение стали неотъемлемой частью образовательного процесса, а образовательные платформы стали местом, на которое все люди во всем мире обращают внимание, чтобы обогатить и развить их навыки. В последние годы мы стали

свидетелями быстрого роста этой отрасли благодаря легкому доступу к таким платформам и низкой стоимости по сравнению с традиционным образованием. Эта область открыла нам новые возможности для изучения и лучшего понимания, поскольку здесь есть проблемы, которые препятствуют прогрессу и развитию работы этих платформ, и важно понимать мнения пользователей и студентов о доступных материалах и курсах.

Социальные сети помогли пролить свет на образовательный опыт студентов, поскольку они предоставили им широкое пространство для выражения своего мнения и проблем. Эти мнения являются ключевым фактором при оценке качества образовательного процесса и помогают в принятии решений и предоставлении необходимых услуг для повышения качества образовательного процесса и возможности перенять опыт студентов, который побуждает лиц, принимающих решения, улучшать свои услуги и, таким образом, подталкивать студентов к прогрессу и успеху.

Проведенные исследования выявили важную роль, которую Twitter может играть в оценке мнения аудитории во многих областях. Например, Twitter использовался для предварительного просмотра опыта MOOC. В работе [1] автор собрал в общей сложности 1090 сообщений на Twitter по теме Coursera и изучил стили обучения учащихся и инструкции учителей «как есть», что объясняется в различных твитах. Исследование показало, что тренеры MOOC широко использовали инструменты социальных сетей для продвижения своего обучения и обеспечения успеха онлайн-образования. В работе [2] автор проанализировал реакцию общественности на вспышку COVID-19 в Твиттере, используя методы анализа закономерностей. В работе [3] автор предложил алгоритм, позволяющий узнать общественное мнение граждан Индии о налоговой системе, одобренной правительством Индии, поскольку из Twitter было собрано 190 твитов о налоговой системе на английском языке, а затем применен лексический подход для определения чувств, которые каждый твит несет. В работе [4] автор предложил лексический подход для добычи неструктурированных данных из Twitter, чтобы установить семантическую тенденцию взаимодействий и мнений в режиме реального времени путем извлечения настроений аудитории из комментариев, связанных с новостными статьями BBC, публикуемыми в Twitter.

### **МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ**

В этой работе мы использовали методологию изучения реакции людей на Twitter для электронного обучения, согласно которой 3000 твитов, содержащих ключевые слова, связанные с образованием, были собраны из Twitter с использованием инструмента интеллектуального анализа данных “Orange” [5], а затем эти данные были подготовлены и очищены перед их анализом с использованием лексикон “Vader” [6], а затем выполнено фотографирование и определение полярности данных с помощью программного обеспечения Power IB [7]. На рисунке 1 показан процесс получения и анализа данных из Twitter.

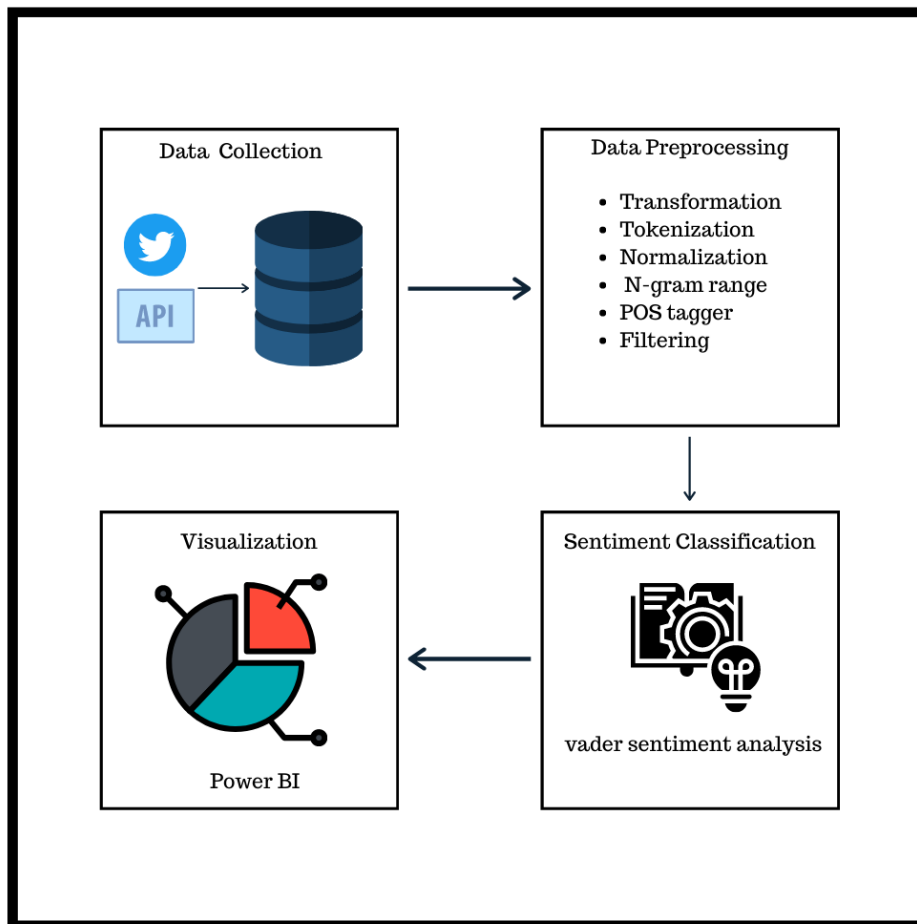


Рис. 1. Процесс получения и анализа данных из Twitter  
Fig. 1. The process of obtaining and analyzing data from Twitter

### СБОР ДАННЫХ

Данные были получены из Twitter через Twitter API [8], и в них были введены ключ и код клиента, которые используются для идентификации пользователя и получения твитов путем поиска по определенным ключевым словам. В нашем случае мы использовали экстрактор данных “Orange” для извлечения данных из Twitter. Были введены следующие ключевые слова для запроса:

Mooc  
online course  
digital learning  
e-learning  
learn online  
online learning

Данные, которые не будут обрабатываться, отфильтровали следующим образом:

1. Твиты написаны только на английском языке.
2. Исключение ретвитов.

1	Content	Date	Language	Location	Number of	Number of	In Reply	T	Author Na	Author De	Author Sta	Author Fav	Author Fri	Author Fol	Author Lis	Author Ver	Author
2	Our rates	2021-04-19 19:en			0	0		Suregrade	Your succe	1723	0	170	62	1	False	@Timelywriters01	
3	Our rates	2021-04-19 19:en			0	0		Excel Onlir	Hire a tear	5427	834	604	504	0	False	@online_excel	
4	@Reallifec	2021-04-19 19:en			0	0	@Reallifec	Coach Car	Father to t	1918	3425	350	236	4	False	@kciz1e	
5	Our rates	2021-04-19 19:en			0	0		Superior V	We offer r	3923	406	131	68	0	False	@SuperiorWrite13	
6	ATTENTIO	2021-04-19 19:en			3	1		zoe	♡(di) good c	47751	192804	5964	7133	100	False	@POTTEROMENS	
7	We help er	2021-04-19 19:en			0	0		Superior V	We offer r	3923	406	131	68	0	False	@SuperiorWrite13	
8	We help er	2021-04-19 19:en			0	0		Perfecte	We offer A	5458	14743	313	698	0	False	@perfectdwrite4	
9	Online Lea	2021-04-19 19:en			0	0		Regina Nic	Empowerii	2691	5599	474	74	0	False	@ImReginaNicole	
10	We are avi	2021-04-19 19:en			0	0		Superior V	We offer r	3923	406	131	68	0	False	@SuperiorWrite13	
11	@Thee_m	2021-04-19 19:en	KE		0	0	@Thee_m	The Overtl	MANCH	255042	104	444	1760	37	False	@Carytosh	
12	Our rates	2021-04-19 19:en			0	0		W R I T E	Y PROFESSIC	9164	222	128	3965	2	False	@writegram	
13	Our rates	2021-04-19 19:en			0	0		W R I T E	v we write;E	20064	1945	398	4927	3	False	@write_wave	
14	Our rates	2021-04-19 19:en			0	0		Superior V	We offer r	3923	406	131	68	0	False	@SuperiorWrite13	
15	@pr_marg	2021-04-19 19:en			0	0	@pr_marg	Donna Russell		23	73	22	2	0	False	@DonnaRu87737275	
16	My Trade	2021-04-19 19:en			0	0		DenDon60	Son, Broth	3	2	10	1	0	False	@DDon604	
17	never thou	2021-04-19 19:en			0	0		loz	21	9622	14032	619	402	0	False	@lozaxk	
18	We are avi	2021-04-19 19:en			0	0		Superior V	We offer r	3923	406	131	68	0	False	@SuperiorWrite13	
19	@edsbs @	2021-04-19 19:en			0	0	@edsbs	Dee Seiffe	Retired frc	31999	63216	1414	307	0	False	@deeseiffer	
20	We are avi	2021-04-19 19:en			0	0		Purity Writ	The profici	39	14	4	2	0	False	@PurityWriter	
21	Our rates	2021-04-19 19:en			0	0		KAIZEN WI	WE @kaiz	9842	3531	5000	645	1	False	@KaizenEssays	
22	If you wan	2021-04-19 19:en			0	0		TwentyOn	We believ	536	11	84	46	1	False	@TwentyOneLead	
23	Learn how	2021-04-19 19:en			0	0		MediaWisi	We're a nc	3280	1231	392	15657	252	True	@mediawise	
24	@naomirv	2021-04-19 19:en			1	0	@naomirv	Steven Wil	If you're g	11739	6619	347	239	0	False	@Best_Wilson777	
25	We are avi	2021-04-19 19:en			0	0		EssayWrit	We are ha	473	3970	5001	1045	0	False	@StevenWilson777	
26	@RogersH	2021-04-19 19:en			0	0	@RogersH	Tkerr		1	0	4	0	0	False	@yeadytrk	
27	@Ofstedn	2021-04-19 19:en			0	0	@Ofstedn	Amy Joyn	Mum of 2	50	127	65	27	0	False	@ajaynes75	
28	more in th	2021-04-19 19:en			0	0	@ChildMe	Child Ment	We conn	76	79	24	16	0	False	@ChildMentorship	
29	We ensure	2021-04-19 19:en			0	0		PAPERS W	Competiti	5921	2245	1387	478	2	False	@paperwritingpr	

Рис. 2. Список данных  
Fig. 2. Data list

### ПОДГОТОВКА ДАННЫХ

После процесса сканирования с помощью инструмента Orange был применен процесс очистки данных, поскольку извлекаемые данные могут содержать несоответствующие и раздражающие данные. Этот шаг помогает сократить устаревший словарный запас для достижения наилучших результатов [9]. Универсального списка таких слов не существует, поэтому в процессе очистки все неактуальные данные были удалены. Некоторые из часто используемых шагов предварительной обработки включают удаление стоп-слов, символа @, тегов POS, бессмысленного слова или символа и т.д. Этапы предварительной обработки следующие:

- Используйте строчные буквы для всех текстовых данных;
- Удалите гиперссылки, если они есть;
- Токенизация: RegExr;
- Удаление стоп-слова: Исключение исключенных слов, то есть слов, не влияющих на значение и общее восприятие текста, таких как предлоги.

В таблице 1 представлен список слов, удаленных из собранных твитов.

Таблица 1

Стоп-слова

Table 1

Stop words

E	16	3	9
Us	Also	next	10
2021	Would	19	8
Due	5	covid	6
2	dm	Covid 19	
1	via	4	

Список слов после процесса предварительной обработки данных показан на рисунке 3.



Рис. 3. Список слов  
Fig. 3. List of words

### ПРОЦЕСС ЭМОЦИОНАЛЬНОГО АНАЛИЗА

После процесса предварительной обработки данных данные были классифицированы на положительные, отрицательные или нейтральные. Сначала мы использовали “Vader”, чтобы найти полярность каждого слова в предложении, затем общая полярность была вычислена путем вычитания общей полярности всех положительных слов из общей полярности всех отрицательных слов.

Алгоритм, используемый для расчета полярности каждого твита, выглядит следующим образом:

- Если сумма  $> 0$ , твит положительный, в противном случае – отрицательный.
- Если Total = 0, твит нейтрален.

Power BI использовался для визуализации данных. Рисунок 4 показывает процент положительных, отрицательных и нейтральных твитов.

Было замечено, что процент положительных твитов больше, чем отрицательных твитов, как показано на рис. 4. Тем не менее, есть еще возможности для улучшения, поскольку 19.27% - отрицательные твиты (рис. 4).

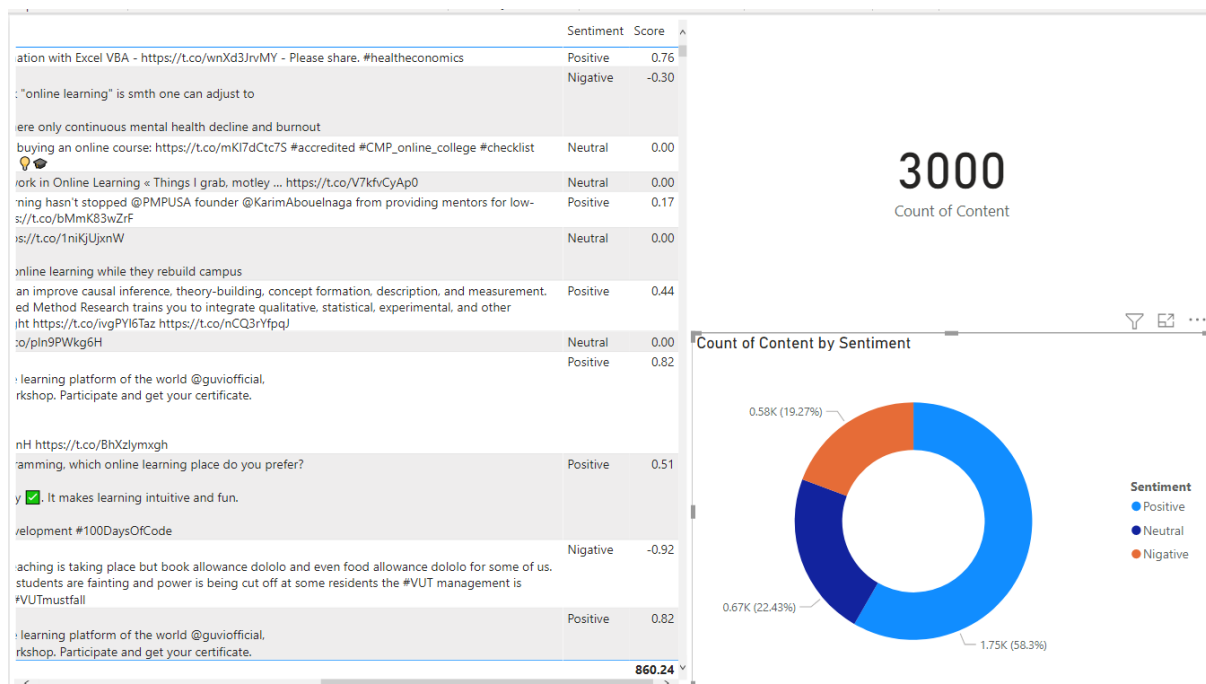


Рис. 4. Процент положительных, отрицательных и нейтральных твитов  
Fig. 4. Percentage of positive, negative and neutral tweets



данные Twitter. Будет интересно извлечь и проанализировать данные с других интернет-платформ (таких как Google Trend, Google News, Facebook и YouTube), чтобы оценить корреляцию с информацией в твитах, а также увидеть, совместимы ли результаты этого исследования с другими платформами. В-третьих, данные не собирались для целевых регионов или языков, кроме английского. Концентрация на определенных языках может быть интересным способом намеренного сбора твитов из определенных стран или регионов для более подробного анализа местной реакции людей в этих странах.

### Список литературы

1. Xia Цзинфэн. (2014). Контент-анализ твитов об опыте MOOC. Вездесущее обучение [Текст]. 6. 1-14. 10.18848/1835-9795/CGP/v06i03/40251.
2. Нур, Салеха и Го, Йи и Хамад, Сайед и Шах, Сайед Хамад Хассан и Фурнье Вигер, Филипп и Наваз, М. Сакиб. (2020). Анализ реакции общественности на вспышку нового коронавируса (COVID-19) в Твиттере. Кибернет [Текст]. 10.1108/К-05-2020-0258.
3. Рао К.П., Конеру А. и Раджу Д.Н. (2019). Алгоритм ОЕFC — анализ настроений по системе налогообложения товаров и услуг в Индии. Когнитивная информатика и программные вычисления [Текст] (pp. 441-451).
4. Юнас, Фарах и Оуда, Маджди. (2021). Пространственный анализ настроений и восприятия статей BBC News с использованием интеллектуального анализа сообщений Twitter [текст]. 10.1007/978-3-030-55187-2\_27.
5. Orange data mining documentation [Электронный ресурс]. – Режим доступа: <https://orangedatamining.com/docs/>
6. VADER Sentiment Analysis documentation [Электронный ресурс]. – Режим доступа: <https://pypi.org/project/vaderSentiment/>
7. Power BI documentation [Электронный ресурс]. – Режим доступа: <https://docs.microsoft.com/en-us/power-bi/>
8. Twitter API [Электронный ресурс]. – Режим доступа: <https://developer.twitter.com/en/docs>
9. All you need to know about text preprocessing for NLP and Machine Learning [Электронный ресурс]. – Режим доступа: <https://www.kdnuggets.com/2019/04/text-preprocessing-nlp-machine-learning.html>

### References

1. Xia, Jingfeng. (2014). A Content Analysis of Tweets about MOOC Experiences. Ubiquitous Learning. 6. 1-14. 10.18848/1835-9795/CGP/v06i03/40251.
2. Noor, Saleha & Guo, Yi & Hamad, Syed & Shah, Syed Hamad Hassan & Fournier Viger, Philippe & Nawaz, M. Saqib. (2020). Analysis of public reactions to the novel Coronavirus (COVID-19) outbreak on Twitter. Kybernetes. 10.1108/К-05-2020-0258.
3. Rao, K. P., Koneru, A., & Raju, D. N. (2019). OEFC Algorithm—Sentiment Analysis on Goods and Service Tax System in India. Cognitive Informatics and Soft Computing (pp. 441-451).
4. Younas, Farah & Owda, Majdi. (2021). Spatial Sentiment and Perception Analysis of BBC News Articles Using Twitter Posts Mining. 10.1007/978-3-030-55187-2\_27.
5. Orange data mining documentation [Electronic resource]. – URL: <https://orangedatamining.com/docs/>
6. VADER Sentiment Analysis documentation [Electronic resource]. – URL: <https://pypi.org/project/vaderSentiment/>
7. Power BI documentation [Electronic resource]. – URL: <https://docs.microsoft.com/en-us/power-bi/>
8. Twitter API [Electronic resource]. – URL: <https://developer.twitter.com/en/docs>
9. All you need to know about text preprocessing for NLP and Machine Learning [Electronic resource]. – URL: <https://www.kdnuggets.com/2019/04/text-preprocessing-nlp-machine-learning.html>

**Кнаан Або-Рашед**, аспирант Белгородского государственного национального исследовательского университета  
**Польщиков Константин Александрович**, доктор технических наук, доцент, директор института инженерных и цифровых технологий

**Knaan Abo-Rashed**, postgraduate student, Belgorod State National Research University  
**Polshchikov Konstantin Alexandrovich**, Doctor of Technical Sciences, Associate Professor, Director of the Institute of Engineering and Digital Technologies