

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПРИНЯТИЕ РЕШЕНИЙ ARTIFICIAL INTELLIGENCE AND DECISION MAKING

УДК 004.838.2

DOI: 10.18413/2518-1092-2021-6-1-0-7

Самигулин Т.Р.
Джурабаев А.Э.У.

АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТА МЕТОДАМИ МАШИННОГО
ОБУЧЕНИЯ

Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Кронверкский пр., д. 49, г. Санкт-Петербург, 197101, Россия

e-mail: timursamigulin98@gmail.com, anvar19971403@gmail.com

Аннотация

В современном мире накоплено гигантское количество информации. Одной из наиболее распространенных форм хранения информации являются тексты на естественном языке. Из-за необходимости анализа больших массивов текстовых данных получило развитие направление обработки текста на естественном языке (NLP – Natural Language Processing). Анализ тональности в тексте является одним из основных направлений раздела обработки текста на естественном языке. В статье рассматриваются основные методы и подходы к задаче анализа тональности текста на естественном языке. Дается краткая характеристика используемым на практике методам традиционного машинного обучения, а также методам глубокого обучения. По результатам данной статьи определены наиболее результативные методы анализа тональности.

Ключевые слова: анализ текста на естественном языке, анализ тональности, машинное обучение, глубокое обучение, искусственные нейронные сети.

Для цитирования: Самигулин Т.Р., Джурабаев А.Э.У. Анализ тональности текста методами машинного обучения // Научный результат. Информационные технологии. – Т.6, №1, 2021. – С. 55-62. DOI: 10.18413/2518-1092-2021-6-1-0-7

Samigulin T.R.
Djurabaev A.E.U.

SENTIMENT ANALYSIS OF TEXT BY MACHINE LEARNING
METHODS

Saint Petersburg National Research University of Information Technologies, Mechanics and Optics,
49 Kronverkskiy prospekt, St. Petersburg, 197101, Russia

e-mail: timursamigulin98@gmail.com, anvar19971403@gmail.com

Abstract

There is a gigantic amount of information in the world. One of the most common forms of information storage is natural language texts. For the analysis of gigantic arrays of text data, the direction of natural language processing has been developed. Sentiment analysis in text is one of the main areas of the natural language processing section. The article discusses the main methods and approaches to the problem of analyzing the sentiment of a text in a natural language. A brief description of the traditional machine learning methods and deep learning methods used in practice is given. Based on the results of this article, the most effective methods of sentiment analysis have been identified.

Keywords: natural language texts analysis, sentiment analysis, machine learning, deep learning, artificial neural networks.

For citation: Samigulin T.R., Djurabaev A.E.U. Sentiment analysis of text by machine learning methods // Research result. Information technologies. – Т.6, №1, 2021. – P. 55-62.
DOI: 10.18413/2518-1092-2021-6-1-0-7

ВВЕДЕНИЕ

Анализ тональности текста – это подраздел обработки естественного языка (NLP) целью которого является классификация текста по тональности. Тональность — это мнение, отношение и эмоции автора по отношению к объекту, о котором говорится в тексте. В качестве объектов могут выступать объекты реального мира, люди, события или процессы. Обычно используется бинарная классификация, выявление в тексте положительных и отрицательных оттенков. Но также может добавляться нейтральный класс или стоять более сложная задача, допустим выявление оценок, которые поставит пользователь: «Отлично», «Хорошо», «Плохо» и другие.

Анализ тональности текста имеет широкий спектр применений в современном мире. С его помощью можно выявлять отношение пользователей к продукту [8], применять данный анализ для политических [9], социологических [10], экономических [11], маркетинговых исследований, строить рекомендательные и обучающие системы [1].

В настоящее время в основном используются следующие подходы для выявления эмоциональной окраски текста:

1) Лингвистический подход или анализ, основанный на правилах и словарях. Данный подход основан на использовании словарей с заранее подготовленными вручную шаблонами эмоционально важных слов и словосочетаний с их эмоциональными оценками. При использовании данного подхода в тексте ищутся пересечения со словарем. Затем по сумме оценок найденных пересечений определяется тональность заданного текста. Данный подход показывает хорошие результаты для некоторых областей. Основной недостаток данного подхода в большой сложности подготовки словарей, надо хорошо знать предметную область, для которой составляется словарь. Второй недостаток — это плохая масштабируемость, нельзя использовать один и тот же словарь для разных предметных областей. Одинаковые термины в различных областях могут вносить разный вес в степень эмоциональной окраски [2].

2) Подход, основанный на использовании методов машинного обучения. Суть данного подхода в том, что вначале на заранее размеченных данных обучается классификатор, который потом используется для классификации новых текстов. В нашей статье мы подробнее рассмотрим методы именно данного подхода [5].

3) Гибридный подход, сочетающий в себе подходы как на основе правил и словарей, так и на основе машинного обучения. Ряд исследований показывает, что с помощью данного подхода можно добиться улучшения качества классификации, но такой подход является самым трудоемким и затратным по времени [12].

Цель данной статьи – исследовать современные методы машинного обучения для задачи определения тональности в тексте на естественном языке на основе публикаций за последние несколько лет. Будут рассматриваться, как и традиционные методы машинного обучения, так и популярные сейчас методы глубокого обучения. Результаты данного исследования планируется использовать для дальнейших исследований в области определения тональности текста.

ОСНОВНАЯ ЧАСТЬ

Далее будут представлены основные методы традиционного машинного обучения, а также методы глубокого обучения, в частности, искусственные нейронные сети, которые в последние годы очень популярны и вытесняют более традиционные методы. Дана их краткая характеристика и проанализирована их эффективность для задачи анализа тональности.

Наивный байесовский классификатор – является вероятностным классификатором. Наивная байесовская модель вычисляет условную вероятность класса на основе распределения слов в документе. Один из самых простых используемых классификаторов. Основан на теореме Байеса с предположением о том, что все признаки являются независимыми, благодаря чему и

получил название наивный байесовский классификатор [13]. Но обычно в текстовых документах, предположение о независимости не подтверждается, что делает его слабоэффективным. Тем не менее несмотря на всю простоту и ограничение на независимость, байесовский классификатор может показывать хорошие результаты при классификации текста. В данном исследовании с помощью наивного байесовского классификатора на различных данных получают точность от 55% до 79% [21]. Одним из преимуществ является малое количество данных необходимых для обучения и простота реализации [3].

Метод максимума энтропии – также как и наивный байесовский классификатор является вероятностным классификатором. Данный метод основан на принципе максимальной энтропии, что наиболее характерным распределением вероятностей неопределенной среды, являются распределения, которые максимизируют выбранную меру неопределенности при заданной информации о поведении среды. В отличие от наивного байесовского классификатора метод максимума энтропии не делает предположения о независимости признаков, что позволяет добиться лучших результатов. Также как и у наивного байесовского классификатора преимуществами являются простота реализации и малое количество данных необходимых для обучения [14].

Деревья решений – представляют из себя древовидную структуру, где на ветках записаны атрибуты, от которых зависит распределение вероятностей классов, а на листьях значения вероятностей классов. Данный метод просто в интерпретации и требует минимальной предобработки данных. Но сами по себе деревья решений используются редко, так как они легко переобучаемы и слишком зависимы от обучающих данных. При небольших изменениях в обучающей выборке мы получаем кардинально разные результаты на тестовых данных. Чаще применяются ансамбли решающих деревьев, которые решают данные проблемы. Примеры таких ансамблей: случайный лес или градиентный бустинг [15].

Случайный лес – ансамбль решающих деревьев. В данном методе строиться очень много решающих деревьев большой глубины на разных обучающих данных. Деревья строятся до тех пор, пока в каждом листе не окажется очень мало объектов, то есть они сильно переобучены. Затем все деревья объединяются, и мы получаем эффективный классификатор, у которого отсутствуют недостатки решающих деревьев. Но это вызывает некоторые проблемы, если признаков очень много, то этот подход работает не очень хорошо: деревья будут очень глубокими, на их построение будет уходить слишком много времени.

Градиентный бустинг – метод машинного обучения для классификации и регрессии. Этот метод строит модель поэтапно, улучшая на каждом следующем этапе модель, которая получилась на прошлом этапе. В качестве базовых алгоритмов используются очень простые модели, например неглубокие решающие деревья. При использовании градиентного бустинга решающие деревья, в отличие от случайного леса, имеют очень небольшую глубину. Но это тоже проблема. Каждое дерево может учесть лишь небольшое подмножество признаков, в то время как зачастую ответ зависит от комбинации большого количества слов в тексте. Поэтому для хорошей работы градиентного бустинга нужно будет использовать очень много деревьев, но даже в этом случае нет гарантии, что полученное качество будет приемлемым.

Логическая регрессия – является методом линейного классификатора, оценивающий вероятность принадлежности объектов к классу путем сравнения с логической кривой по значениям множества признаков. Используется как для задач регрессии, так и для классификации. На практике часто рассматривается логическая регрессия с регуляризацией. Регуляризация заключается в том, что модель начинает штрафовать за очень большие веса, что не дает модели переобучиться. Логическая регрессия один из самых популярных методов классификации и обученная модель показывает очень хорошие результаты. Из недостатков можно выделить, что необходима качественная предобработка признаков и их отбор [14].

Метод опорных векторов (SVM) – набор линейных алгоритмов машинного обучения для задач регрессии и классификации. Цель метода заключается в нахождении среди всех возможных гиперплоскостей пространства, отделяющих два класса обучающих примеров друг от друга, такой

гиперплоскости, расстояния от которой до ближайших векторов обоих классов равны (оптимальная разделяющая гиперплоскость) [16]. Является одним из наиболее эффективных методов классификации. Данный метод часто применяется в задачах классификации текстов и показывает хорошие результаты [17]. Линейные модели хорошо масштабируются, могут работать с большим количеством признаков, на очень больших выборках.

В исследовании методов машинного обучения в задаче автоматического определения тональности текстов на естественном языке применялись основные традиционные методы машинного обучения [4]. В качестве входных данных использовались следующие англоязычные корпуса текстов:

- 1) Корпус отзывов о фильмах, входящий в состав библиотеке NLTK, 2000 текстов, в среднем 3500 символов в тексте.
- 2) Корпус из лексического семантического тезауруса SentiWordNet, 2000 текстов, в среднем 150 символов в тексте.

В качестве метрики эффективности метода использовалась AUC – площадь под ROC-кривой (кривой ошибок). Авторы обучили несколько моделей, подбирая разные параметры, чтобы добиться лучших результатов. Перед этим была произведена предобработка данных и отбор признаки. Ниже представлены наилучшие результаты для различных методов традиционного машинного обучения.

- Логическая регрессия: обучающая выборка – 0.93445, тестовая выборка – 0.93445.
- Дерево принятий решений: обучающая выборка – 0.68204, тестовая выборка – 0.6500.
- Случайный лес: обучающая выборка – 0.90799, тестовая выборка – 0.84000.
- Метод опорных векторов: обучающая выборка – 0.89416, тестовая выборка – 0.86167.

Еще одно исследование в качестве входных данных использует наборы данных содержащие отзывы о товарах с интернет-магазина [22]. Метрики оценивания ассигасу (точность) – соотношение правильно предсказанных объектов к общему количеству объектов в наборе данных. В данном исследовании получились следующие результаты:

- Метод максимума энтропии – точность 72.60%
- Случайный лес – точность 88.39%
- Наивный байесовский классификатор – точность 75.50%
- Метод опорных векторов – точность 91.15%

Как можно заметить среди методов традиционного машинного обучения наилучшие результаты показывают линейные модели: логическая регрессия и метод опорных векторов. Также неплохо себя показывает случайных лес. Но эффективность работы традиционных методов сильно зависит от объема и качества обучающих данных. Кроме того, точность во многом зависит от выбора признаков, что достаточно трудоемкая задача. Качество анализа тональности текста, в которых не соблюдаются правила и грамматика языка, например сообщения в социальных сетях, зачастую оказывается довольно низким. В связи с этим, можно сделать вывод, что необходима предобработка данных и тщательный отбор признаков.

В последние годы методы глубокого обучения все более популярны. Глубокое обучение представляет собой набор алгоритмов машинного обучения, которые моделируют высокоуровневые абстракции в данных, используя архитектуры, состоящие из множества нелинейных трансформаций и выделяя из данных «скрытые признаки» [18].

Мы рассмотрим несколько искусственных нейронных сетей, которые являются алгоритмами глубокого обучения. Искусственная нейронная сеть – это математическая модель, построенная по принципу организации и функционирования биологических нейронных сетей. Искусственные нейронные сети широко используются при автоматической обработке текстов на естественном языке, в том числе и для определения тональности текстов. Основная трудность использования искусственных нейронных сетей заключается в необходимости их настройки: определении количества используемых скрытых слоев, функции активации для каждого узла, пороговой ошибки. Также для обучения качественной модели необходимо большое количество

обучающих данных, ресурсов и времени. Одно из ключевых преимуществ искусственных нейронных сетей над традиционными методами машинного обучения, нейронные сети способны отбирать признаки в данных без участия человека. В процессе обучения искусственная нейронная сеть выявляет сложные зависимости между входными данными и выходными. Это значит, что в случае успешного обучения сеть сможет вернуть верный результат на основании данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искаженных данных. Еще одним преимуществом искусственной нейронной сети является ее способность адаптироваться к разным вариантам постановки задачи с небольшими изменениями в системе, выполняющей анализ тональности [5]. Наиболее зарекомендовавшими себя в области анализа тональности текстов методами глубокого обучения являются сверточные и рекуррентные нейронные сети.

Сверточные нейронные сети (CNN). Изначально сверточные сети начали использовать для распознавания изображений, но после большого успеха в области компьютерного зрения, их стали пробовать применять и в других областях. В частности, их стали использовать для решения задач классификации текстов. В сверточных нейронных сетях используется операция свертки, когда каждый фрагмент данных умножается на матрицу (ядро) свертки поэлементно, после чего результат суммируется и записывается в аналогичную позицию выходных данных. Поскольку свертки происходят на соседних словах, модель может уловить отрицания или n-граммы, которые несут новую информацию о настроении. В данном исследовании показано, что сверточные сети могут демонстрировать высокие результаты при анализе тональности текста, превосходя другие алгоритмы на некоторых тестах [19].

Рекуррентные нейронные сети (RNN) широко распространены в задачах обработки текста, в том числе и для анализа тональности. Особенности рекуррентных нейронных сетей — это наличие обратных связей, связь от более удаленного элемента к менее удаленному. Это позволяет запоминать и воспроизводить последовательности реакций на один стимул. Значение весов сети зависит как от текущих, так и от предыдущих входных данных, благодаря чему вес каждого слова влияет на веса остальных слоев в предложении [6].

GRU (Gated Recurrent Unit – управляемые рекуррентные нейроны) и LSTM (Long Short-Term Memory – длительная кратковременная память) являются модификациями рекуррентных нейронных сетей. Они решают проблему исчезающего градиента, которой подвержена рекуррентная нейронная сеть.

Рекуррентные нейронные сети показывают наилучшие результаты во многих задачах, но процедура их обучения достаточно трудоемка [7].

Как уже было сказано выше, среди традиционных методов машинного обучения лучшие результаты показывают линейные модели. В этом исследовании проводят сравнение сверточных нейронных сетей и логической регрессии по определению тональности в сообщениях из твитера [20]. В качестве метрики использовалась ассигасу, соотношение правильно классифицированных объектов к общему количеству объектов в наборе данных. Получились следующие результаты:

- Логическая регрессия: обучающие данные – 84.7%, тестовые данные – 76.7%.
- Сверточная нейронная сети: обучающие данные – 82.9%, тестовые данные – 79.5%.

Как видно сверточная нейронная сеть показывает лучшие результаты. Но у линейных моделей есть преимущество над сверточными нейронными сетями в скорости обучения 45 секунд против 6 часов. Так что если время обучения не приоритет и в распоряжении есть большое количество обучающих данных, то лучше отдать предпочтение искусственным нейронным сетям.

Еще одно исследование, которое сравнивает уже рекуррентные нейронные сети с линейными моделями [7]. В качестве входных данных используется набор данных отзывов о машинах. В качестве метрики – ассигасу (точность).

- Логическая регрессия: точность 70.1%.
- RNN: точность 76%.
- LSTM: точность 77%

Сравнение различных архитектур рекуррентных и сверточных нейронных сетей. Оценка производится на наборе данных, содержащий обзоры фильмов, а в качестве метрики используется точность.

- Сверточная нейронная сеть: точность – 86.8%.
- Рекуррентная нейронная сеть: точность – 87.5%.

Среди методов глубокого обучения в задаче анализа тональности текста больше зарекомендовали себя рекуррентные нейронные сети. Но в некоторых задачах сверточные нейронные сети также могут показывать хорошие результаты, которые превосходят рекуррентные нейронные сети. Допустим исследование, где производится анализ тональности в коротких сообщениях [19]. В данном исследовании сравниваются различные архитектуры сверточных и рекуррентных нейронных сетей. Получились следующие результаты:

- Сверточная нейронная сеть: точность – 88.1%.
- Рекуррентная нейронная сеть: точность – 85.4%.

ЗАКЛЮЧЕНИЕ

В данной обзорной статье описаны основные методы используемые для анализа тональности текста. Представлены методы традиционного машинного обучения, используемые для анализа тональности. А также описаны методы глубокого обучения.

Результаты данного исследования будут использованы для другого исследования «Разработка программного модуля для распознавания психоэмоционального состояния на основе текстовых сообщений чата проводимого занятия». Следующей задачей стоит проверить на экспериментальных данных эффективность методов глубокого обучения и подобрать оптимальную архитектуру для задачи определения тональности текста. Также в качестве будущего исследования можно сделать сравнительный анализ различных методов традиционного машинного обучения и методов глубокого обучения на разных наборах данных.

Список литературы

1. Трофимова Е.В., Туральчук К.А. Разработка рекомендательной системы на основе анализа тональности текста // Актуальные проблемы гуманитарных и естественных наук. – 2015. – №1. – С. 93-94.
2. Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: сб. научных статей. Вып. 10 (17). – М.: Изд-во РГГУ, 2011. – С. 510–522.
3. Котельников Е.В., Клековкина М.В. Автоматический анализ тональности текстов на основе методов машинного обучения // Программный комитет конференции выражает искреннюю благодарность Российскому фонду фундаментальных исследований за финансовую поддержку, грант № 12-06-06045-г. – 2012. – С. 27.
4. Ермаков П.Д., Федянин Р.В. Исследование методов машинного обучения в задаче автоматического определения тональности текстов на естественном языке // Новые информационные технологии в автоматизированных системах. – 2015. – №18. URL: <https://cyberleninka.ru/article/n/issledovanie-metodov-mashinnogo-obucheniya-v-zadache-avtomaticheskogo-opredeleniya-tonalnosti-tekstov-na-estestvennom-yazyke>
5. Пескишева Т.А. Методы анализа тональности текстов на естественном языке // Общество. Наука. Инновации (НПК-2017). – 2017. – С. 1730-1742.
6. Кудинов М.С. Статистическое моделирование русского языка с помощью нейронных сетей: Дис. д-ра тех. наук. М., – 2016. – 106 с.
7. Тарасов Д. Глубокие рекуррентные нейронные сети для аспектно-ориентированного анализа тональности отзывов пользователей на различных языках // По материалам ежегодной Международной конференции «Диалог». – 2015. – №. 14. – С. 21.
8. Shelke N., Deshpande S., and Thakare V. “Domain independent approach for aspect-oriented sentiment analysis for product reviews” in Proc. 5th Int. Conf. Frontiers Intell. Comput., Theory Appl., Singapore, Mar. 2017, pp. 651–659.

9. Sharma U., Datta R.K., and Pabreja K. “Sentiment analysis and prediction of election results 2018” in *Social Networking and Computational Intelligence*, R.K. Shukla, J. Agrawal, S. Sharma, N.S. Chaudhari, and K.K. Shukla, Eds. Singapore: Springer, 2020, pp. 727–739.
10. Iacus S., Porro G., Salini S., and Siletti E., “An Italian composite subjective well-being index: The voice of Twitter users from 2012 to 2017,” *Social Indicators Res.*, vol. 149, pp. 1–19, 2020.
11. Carosia A.E.O., Coelho G.P., and Silva A.E.A. “Analyzing the Brazilian financial market through portuguese sentiment analysis in social media” *Appl. Artif. Intell.*, Vol. – 34. – No. 1. – Pp. 1–19. Jan. 2020.
12. Kumar A., Srinivasan K., Cheng W.-H., and Zomaya A.Y. “Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data” *Inf. Process. Manage.*, vol. 57, no. 1, Jan. 2020, Art. no. 102141. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306457319306934>
13. Schütze H., Manning C.D., and Raghavan P., *Introduction to Information Retrieval*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.
14. Medhat W., Hassan A., Korashy H. Sentiment analysis algorithms and applications: A survey // *Ain Shams engineering journal*. – 2014. – Т. 5. – №. 4. – P. 1093-1113.
15. Quinlan J.R. “Induction of decision trees” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
16. Chin Chen Chien, Tseng You-De. Quality evaluation of product reviews using an information quality framework. *Decis Support Syst* 2011; 50: 755–68.
17. Vapnik V. *Statistical learning theory*. New York: Wiley, 1998.
18. Bengio Y. *Learning deep architectures for AI* // *Foundations and Trends in Machine Learning*, 2009.
19. Kim Y. Convolutional neural networks for sentence classification. // *In Proceedings of the 2014 Conference on EMNLP*. – 2014.
20. Zvonarev A. A Comparison of Machine Learning Methods of Sentiment Analysis Based on Russian Language Twitter Data. – 2019.
21. Hasan A. et al. Machine learning-based sentiment analysis for twitter accounts // *Mathematical and Computational Applications*. – 2018. – Т. 23. – №. 1. – P. 11.
22. Ahmad M. et al. Machine learning techniques for sentiment analysis: A review // *Int. J. Multidiscip. Sci. Eng.* – 2017. – Т. 8. – №. 3. – P. 27.
23. Wang X., Jiang W., Luo Z. Combination of convolutional and recurrent neural network for sentiment analysis of short texts // *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*. – 2016. – P. 2428-2437.

Reference

1. Trofimova E.V., Turalchuck K.A. Development of a recommendation system based on text sentiment analysis // *Aktual'nye problemy gumanitarnyh i estestvennyh nauk*. 2015. №1. P. 93-94.
2. Pazel'skaja A. G., Solov'ev A. N. Method for determining emotions in texts in Russian // *Computational Linguistics and Intelligent Technologies: collection of scientific articles*. Ed. 10 (17).: Publisher RGGU, 2011. P. 510–522.
3. Kotel'nikov E. V., Klekovkina M. V. Automatic sentiment analysis based on machine learning methods // *Programmnyj komitet konferencii vyrazhaet iskrennjuju blagodarnost' Rossijskomu fondu fundamental'nyh issledovanij za finansovuju podderzhku*, grant № 12-06-06045-y. – 2012. – P. 27.
4. Ermakov P. D., Fedjanin R. V. Research of machine learning methods in the problem of automatic sentiment determination of natural language texts // *Novye informacionnye tehnologii v avtomatizirovannyh sistemah*. 2015. №18. URL: <https://cyberleninka.ru/article/n/issledovanie-metodov-mashinnogo-obucheniya-v-zadache-avtomaticheskogo-opredeleniya-tonalnosti-tekstov-na-estestvennom-yazyke>
5. Peskischeva T.A. Methods for Sentiment Analysis of Natural Language Texts. // *Obshhestvo. Nauka. Innovacii (NPK-2017)*. 2017. P. 1730-1742.
6. Kudinov M.S. Statistical modeling of the Russian language using neural networks: dis. Dr. tech. sciences. 2016. 106 p.
7. Tarasov D. Deep recurrent neural networks for aspect-oriented analysis of the sentiment of user reviews in various languages // *Po materialam ezhegodnoj Mezhdunarodnoj konferencii «Dialog*. – 2015. – №. 14. – P. 21.
8. Shelke N., Deshpande S., and Thakare V. “Domain independent approach for aspect-oriented sentiment analysis for product reviews” in *Proc. 5th Int. Conf. Frontiers Intell. Comput., Theory Appl.*, Singapore, Mar. 2017, pp. 651–659.

9. Sharma U., Datta R.K., and Pabreja K. “Sentiment analysis and prediction of election results 2018” in Social Networking and Computational Intelligence, R.K. Shukla, J. Agrawal, S. Sharma, N.S. Chaudhari, and K.K. Shukla, Eds. Singapore: Springer, 2020, pp. 727–739.
10. Iacus S., Porro G., Salini S., and Siletti E., “An Italian composite subjective well-being index: The voice of Twitter users from 2012 to 2017,” Social Indicators Res., vol. 149, pp. 1–19, 2020.
11. Carosia A.E.O., Coelho G.P., and Silva A.E.A. “Analyzing the Brazilian financial market through portuguese sentiment analysis in social media” Appl. Artif. Intell., Vol. – 34. – No. 1. – Pp. 1–19. Jan. 2020.
12. Kumar A., Srinivasan K., Cheng W.-H., and Zomaya A.Y. “Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data” Inf. Process. Manage., vol. 57, no. 1, Jan. 2020, Art. no. 102141. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0306457319306934>
13. Schütze H., Manning C.D., and Raghavan P., Introduction to Information Retrieval, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.
14. Medhat W., Hassan A., Korashy H. Sentiment analysis algorithms and applications: A survey // Ain Shams engineering journal. – 2014. – Т. 5. – №. 4. – P. 1093-1113.
15. Quinlan J.R. “Induction of decision trees” Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.
16. Chin Chen Chien, Tseng You-De. Quality evaluation of product reviews using an information quality framework. Decis Support Syst 2011; 50: 755–68.
17. Vapnik V. Statistical learning theory. New York: Wiley, 1998.
18. Bengio Y. Learning deep architectures for AI // Foundations and Trends in Machine Learning, 2009.
19. Kim Y. Convolutional neural networks for sentence classification. // In Proceedings of the 2014 Conference on EMNLP. – 2014.
20. Zvonarev A. A Comparison of Machine Learning Methods of Sentiment Analysis Based on Russian Language Twitter Data. – 2019.
21. Hasan A. et al. Machine learning-based sentiment analysis for twitter accounts //Mathematical and Computational Applications. – 2018. – Т. 23. – №. 1. – P. 11.
22. Ahmad M. et al. Machine learning techniques for sentiment analysis: A review //Int. J. Multidiscip. Sci. Eng. – 2017. – Т. 8. – №. 3. – P. 27.
23. Wang X., Jiang W., Luo Z. Combination of convolutional and recurrent neural network for sentiment analysis of short texts //Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers. – 2016. – P. 2428-2437.

Самигулин Тимур Русланович, инженер, магистрант 1 курса программы Финансовых технологий больших данных Факультета цифровых трансформаций
Джурабаев Анвар Эркин угли, магистрант 1 курса

Samigulin Timur Ruslanovich, engineer, 1st year undergraduate student of the Big Data Financial Technologies program of the Faculty of Digital Transformation
Djurabaev Anvar Erkin ugli, 1st year undergraduate student