

Научно-исследовательский журнал «Modern Economy Success»

<https://mes-journal.ru>

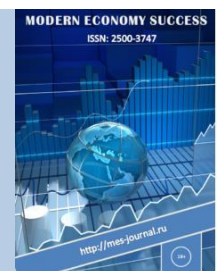
2025, № 1 / 2025, Iss. 1 <https://mes-journal.ru/archives/category/publications>

Научная статья / Original article

Шифр научной специальности: 5.2.2. Математические, статистические и инструментальные методы в экономике (экономические науки)

УДК 004

DOI: 10.58224/2500-3747-2025-1-227-236



¹ Копнин А.А.,

¹ Уральский государственный экономический университет

Технология применения методов машинного обучения для анализа текстовой информации

Аннотация: в статье рассматриваются ключевые аспекты реализации алгоритма K-means для кластеризации отзывов. Представлены теоретические основы кластерного анализа, особенности выбора метрик расстояния и подходы к определению оптимального числа кластеров. На основе текстовых данных отзывов пассажиров выполнена предварительная обработка данных, включая очистку текста, нормализацию и построение матрицы документов-термов. Проведена кластеризация с использованием алгоритма K-means, визуализированы и проанализированы результаты.

Ключевые слова: машинное обучение, кластерный анализ, текстовые данные, алгоритм K-means, обработка текста, анализ отзывов, визуализация кластеров

Для цитирования: Копнин А.А. Технология применения методов машинного обучения для анализа текстовой информации // Modern Economy Success. 2025. № 1. С. 227 – 236. DOI: 10.58224/2500-3747-2025-1-227-236

Поступила в редакцию: 19 сентября 2024 г.; Одобрена после рецензирования: 21 ноября 2024 г.; Принята к публикации: 9 января 2025 г.

¹ Kopnin A.A.,

¹ Ural State University of Economics

Technology of using machine learning methods for analyzing text information

Abstract: the article examines key aspects of implementing the K-means algorithm for clustering reviews. The theoretical foundations of cluster analysis, the specifics of selecting distance metrics, and approaches to determining the optimal number of clusters are presented. Based on textual data from passenger reviews, preliminary data processing was performed, including text cleaning, normalization, and the construction of a document-term matrix. Clustering was carried out using the K-means algorithm, and the results were visualized and analyzed.

Keywords: machine learning, cluster analysis, text data, K-means algorithm, text processing, review analysis, cluster visualization

For citation: Kopnin A.A. Technology of using machine learning methods for analyzing text information. Modern Economy Success. 2025. 1. P. 227 – 236. DOI: 10.58224/2500-3747-2025-1-227-236

The article was submitted: September 19, 2024; Approved after reviewing: November 21, 2024; Accepted for publication: January 9, 2025.

Введение

В эпоху цифровой трансформации общества анализ данных стал ключевым инструментом в различных сферах: от научных исследований до биз-

неса [5, 11, 12]. Огромные объемы данных, которые ежедневно генерируются в мировом масштабе, вызвали необходимость в создании новых технологий их обработки, анализа и интерпретации.

Машинное обучение (Machine Learning), в свою очередь, стало сферой, связанной с разработкой новых математических подходов к анализу данных [4, 8, 10].

Независимо от практической задачи процесс изучения и работы с данными при построении моделей, а также процесс ее обучения можно разбить на стандартные этапы, представленные на рис. 1.



Рис. 1. Обобщённые этапы реализации методов машинного обучения.

Fig. 1. Generalized stages of implementation of machine learning methods.

По подходу реализации методов машинного обучения можно выделить основные категории, каждая из которых имеет свои особенности [9]:

- обучение с учителем. Модель обучается на примерах с известными ответами для того, чтобы научиться предсказывать ответы на новые данные;
- обучение без учителя (Unsupervised Learning), где данные не содержат явных ответов, и задача модели – найти структуру в данных;
- частичное обучение (Semi-supervised Learning) Комбинирует элементы обучения с учителем и без учителя. Используется, когда имеется много неаннотированных данных и небольшое количество аннотированных данных;

– обучение с подкреплением (Reinforcement Learning) метод, при котором модели даются оценки её действий вместо явных ответов. Модель обучается через систему наград и штрафов, стремясь максимизировать награду в долгосрочной перспективе;

– глубокое обучение (Deep Learning) подкатегория машинного обучения, основанная на искусственных нейронных сетях с множеством слоёв.

При структурировании информации и определении наиболее часто встречающихся методов реализации машинного обучения можно составить схему, представленную на рис. 2.

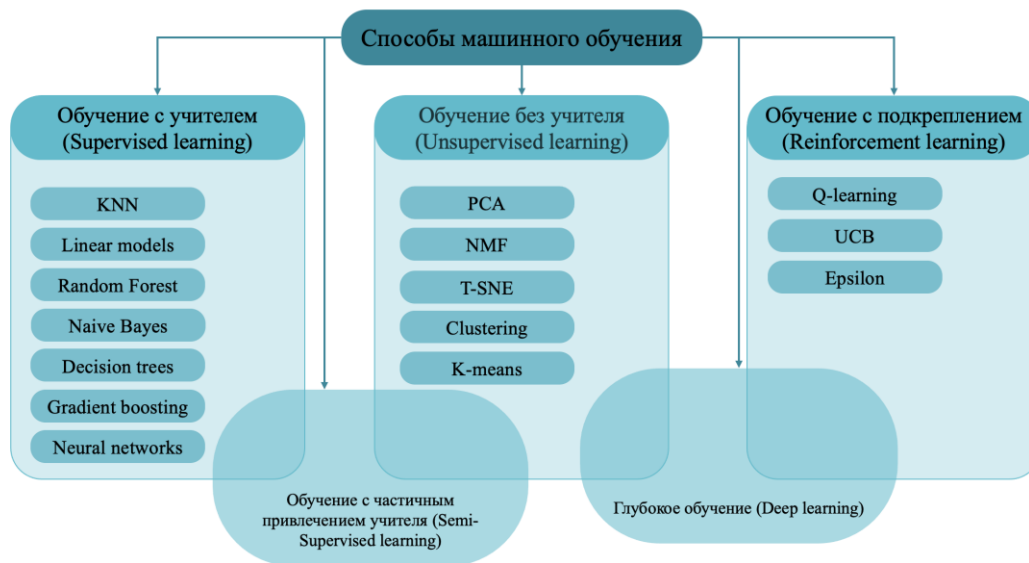


Рис. 2. Способы и примеры методов машинного обучения [6].
Fig. 2. Methods and examples of machine learning methods [6].

Каждый метод имеет свои уникальные возможности применения в зависимости от типа и объема доступных данных, а также от специфических задач, которые необходимо решать.

Материалы и методы исследований

Процесс реализации методы K-means базируется на основных тапах выполнения, которые включают в себя [2, 3]:

- процесс инициализации, связан с выбором количество кластеров k , которое необходимо идентифицировать в данных, в большинстве случаев данный шаг часто основан на априорном знании или эвристиках, но если нет понимания правильного значения k , то можно использовать, например, метод локтя или силуэтный анализ для оценки оптимального значения;

- суть процесса выбора начальных центроидов заключается в том, что случайным образом отбирается значение k точек из данных в качестве начальных центроидов;

- в рамках процесса расчета расстояния необходимо для каждой точки в наборе данных вычислить расстояние до каждого из отобранных ранее центроидов, чаще всего используется евклидово расстояние, но в зависимости от природы данных и задачи могут быть использованы другие метрики, такие как манхэттенское расстояние;

- для идентификации кластеров необходимо каждой точке каждой точке назначит кластер на основе ближайшего центроида;

- донастройка и перерасчет центроидов основывается на расчете среднего значения всех точек внутри каждого кластера и использование данного значения в качестве новых центроидов;

- повторение шагов от отбора центроидов до их перерасчетов пока центроиды не перестанут изменяться (или изменения будут незначительными);

- при оценке результатов, после завершения алгоритма необходимо понять качество кластеров, используя различные метрики (например, силуэтный коэффициент).

Важно заметить, что алгоритм k-means всегда будет разбивать данные на заданное количество кластеров k , даже если естественное число кластеров отличается от k . Тогда, если будет указано $k = 3$ для данных, где реально существует только два кластера, то алгоритм k-means все равно разделит данные на три кластера. Это может привести к ситуации, когда один из «искусственных» кластеров будет размещаться между двумя естественными кластерами или даже разделять естественный кластер на части.

Для реализации метода K-means использован датасет «отзывы пассажиров British Airways (2016-2023)». British Airways, одна из ведущих мировых авиакомпаний, на протяжении десятилетий является синонимом совершенства и надежности. Обладая богатой историей и стремлением обеспечить исключительный клиентский опыт, British Airways продолжает оставаться предпочтительным выбором для путешественников по всему миру.

Цель собранного датасета состояла в том, чтобы собрать ценную информацию из отзывов клиентов и использовать подходы, основанные на данных, для повышения качества услуг и удовлетворенности клиентов авиакомпании [1]. Рассмотрим набор данных подробнее, состоящий из следующих столбцов, каждый из которых содержит важную информацию, извлеченную из обзоров:

- отзывы (Reviews), содержит текстовые отзывы и отзывы, предоставленные клиентами после их опыта работы с British Airways;
- дата (Date), день, когда отзыв был опубликован клиентом, предлагая ценную временную информацию;
- звезды (Stars), рейтинг, данный путешественником, как правило, по шкале от 1 до 5 звезд, что отражает их общую удовлетворенность услугами авиакомпании;
- тип путешественника (Type of Traveler), где классифицируется тип путешественника, который оставил отзыв, различая различные демографические данные о путешествиях, такие как деловые путешественники, семьи или одинокие искатели приключений;
- тип сиденья (Type of Seat), предоставляющий информацию о типе сиденья, которое путешественник испытал во время полета,

включая эконом-, премиум-эконом, бизнес или первый класс;

- страна (Country), описывает страну происхождения клиента, что позволяет провести региональный анализ и понять предпочтения клиента;
- рекомендации (Recommended), является бинарным индикатором, который отражает, порекомендует ли путешественник British Airways на основе своего опыта;
- маршрут (Route), представляет собой информацию о конкретном маршруте или рейсе, который совершают пассажиры, предлагая контекст для их обзоров и опыта.

Результаты и обсуждения

Первостепенно для реализации алгоритма K-means необходимо установить и подключить соответствующие пакеты, как представлено на рис. 3, где пакет `tm` предназначен для обработки и анализа текста, `SnowballC` позволит произвести стемминга текста (приведение слов к корневой форме), `factoextra` направлен на визуализацию результатов кластеризации, `dplyr` для удобной работы с данными и манипуляций с ними [7].

```
2 # Загрузка библиотек
3 library(tm)
4 library(SnowballC)
5 library(factoextra)
6 library(dplyr)
```

Рис. 3. Установка необходимых библиотек.
Fig. 3. Installing the required libraries.

Для того, чтобы реализовать кластеризацию методом K-means необходимо подготовить текстовые данные, а именно реализуем пользовательскую функцию `clean_text`, в которой произведем удаление всех неалфавитных символов, кроме пробелов, URL-адресов, заменим множественных

пробелов одним, а также почистим лишние пробелы в начале и конце строки. Для того, чтобы применить функцию очистки ко всем отзывам в столбце «reviews» необходимо выполнить операцию, представленную на рис. 4 в строке 20.

```
12 # Очистка текстовых данных
13 clean_text <- function(text) {
14   text <- gsub("[^[:alnum:][:space:]]", "", text) # Удаление всех неалфавитных символов,
15   text <- gsub("http\\S+\\S*", "", text) # Удаление URL-адресов
16   text <- gsub("\\s+", " ", text) # Замена множественных пробелов одним
17   text <- trimws(text) # Удаление пробелов в начале и конце строки
18   return(text)
19 }
20 reviews_data$reviews <- sapply(reviews_data$reviews, clean_text)
```

Рис. 4. Очистка и подготовка текстовых данных.

Fig. 4. Cleaning and preparing text data.

Используем функцию `corpus` для создания корпуса текстов из отзывов, после применим последовательные трансформации для приведения текста к нижнему регистру, удаления пунктуации, чисел, стоп-слов и приведение слов к корневой форме (стемминг). Основываясь на предыдущих

действиях создадим матрицу документов-термов (DTM), где строки представляют документы (отзывы), а столбцы – термы (слова) и преобразуем DTM в `data frame` для дальнейшего анализа, как представлено на рис. 5.

```
22 # Создание корпуса текстов
23 corpus <- Corpus(VectorSource(reviews_data$reviews))
24 # Приведение текста к нижнему регистру, удаление пунктуации, чисел и стоп-слов
25 corpus <- tm_map(corpus, content_transformer(tolower))
26 corpus <- tm_map(corpus, removePunctuation)
27 corpus <- tm_map(corpus, removeNumbers)
28 corpus <- tm_map(corpus, removeWords, stopwords("english"))
29 corpus <- tm_map(corpus, stripWhitespace)
30 corpus <- tm_map(corpus, stemDocument)
31
32 # Создание матрицы документов-термов
33 dtm <- DocumentTermMatrix(corpus)
34 # Удаление редких слов (спарсити)
35 dtm <- removeSparseTerms(dtm, 0.99)
36 # Преобразование DTM в data frame
37 data_dtm <- as.data.frame(as.matrix(dtm))
38 # Нормализация данных
39 data_dtm <- scale(data_dtm)
40
```

Рис. 5. Создание корпуса текста и матрицы документов-термов.

Fig. 5. Creation of a text corpus and a matrix of document terms.

В строке 39 из рис. 5 произведена нормализация данных для приведения всех признаков к одному масштабу. Для реализации метода R-means

необходимо определить оптимальное количество кластеров с использованием метода Elbow:

```
41 # Определение оптимального количества кластеров с использованием метода Elbow
42 fviz_nbclust(data_dtm, kmeans, method = "wss")
```

Рис. 6. Определение оптимального количества кластеров.

Fig. 6. Determining the optimal number of clusters.

Теперь визуализируем полученное распределение используя график, где можно определить оптимальное количество кластеров по точке «излома», проанализировав график, можно сказать, что оптимальное количество кластеров составляет 3

или 4, так как на графике метода Elbow, представленного на рис. 7, видно, что при этих значениях происходит заметное снижение суммы квадратов внутри кластеров.

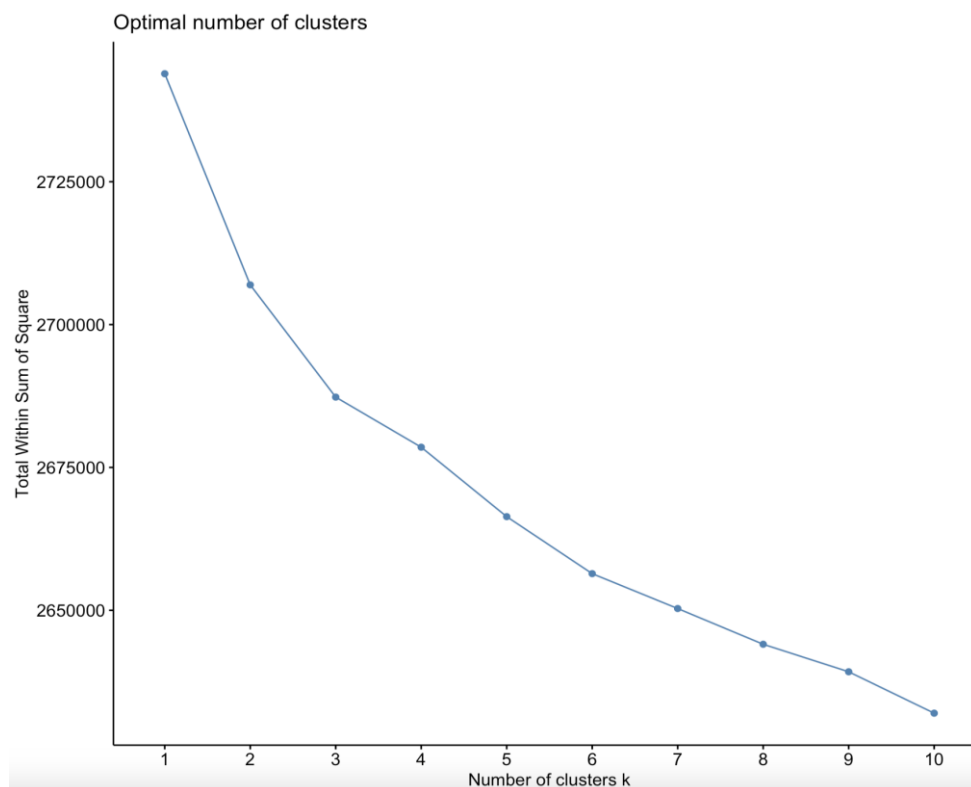


Рис. 7. График метода Elbow.
Fig. 7. Elbow method graph.

Теперь зная число кластеров, в рамках данного исследования будем использовать 4 кластера, можно реализовать K-means, для этого первоначально зафиксируем значения с использованием

функции «set.seed()», присвоим переменной k значение число кластеров и с использованием функции «kmeans» реализуем алгоритм, как представлено на рис. 8.

```
44 # Указание количества кластеров
45 set.seed(123)
46 k <- 4
47
48 # Применение K-means
49 kmeans_result <- kmeans(data_dtm, centers = k, nstart = 25)
```

Рис. 8. Применение K-means кластеризации.
Fig. 8. Application of K-means clustering.

Используя функцию fviz_cluster визуализируем результатов кластеризации в виде графика:

```
51 # Визуализация кластеров
52 fviz_cluster(kmeans_result, data = data_dtm, geom = "point", stand = FALSE, frame = FALSE) +
53   labs(title = "K-means Clustering of British Airways Reviews", x = "Dimension 1", y = "Dimension 2")
```

Рис. 9. Визуализация кластеров.
Fig. 9. Visualization of clusters.

Таким образом получим график, представленный на рис. 10.



Рис. 10. График кластеризации методом K-means для данных.

Fig. 10. K-means clustering plot for the data.

Анализ результатов, полученных в ходе реализации метода K-means

При анализе результатов, полученных в ходе кластеризации с применением алгоритма K-means, первоначально обратим внимание на характеристики полученных кластеров, так как первоначально данный метод машинного обучения был применен с целью оценки мнения потребителя,

что в дальнейшем напрямую влияет на развитие такого экономического предложения, как впечатление, поэтому рассмотрим кластеры, представленные в виде графика, а также исходной таблицы, но с учетом разделения на кластеры.

Для того, чтобы добавить данные к исходной таблице необходимо реализовать следующие функции:

```
56 # Добавление меток кластеров к исходным данным
57 reviews_data$cluster <- as.factor(kmeans_result$cluster)
58 # Просмотр данных с кластерами
59 head(reviews_data)
```

Рис. 11. Фрагмент кода добавления меток кластеров к датасету.

Fig. 11. Code fragment for adding cluster labels to the dataset.

При рассмотрении таблицы, представленной на рис. 12, распределения, по ключевым словам, а также диаграммы, обращая внимание на исходную

таблицу можно сделать описательную характеристику каждого из кластеров.

| reviews | date | country | seat_type | recommended | stars | route | type_of_traveller | cluster |
|---|-----------------|----------------|-----------------|-------------|-------|-----------------------|-------------------|---------|
| 1 Trip Verified I had the most fantastic BA Flight today ... | 1st August 2023 | Hong Kong | Business Class | yes | 5 | Heathrow to Las Vegas | Family Leisure | 3 |
| 2 Trip Verified Couldnt book in online Arrived at check i... | 31st July 2023 | United Kingdom | Economy Class | no | 3 | Rome to Heathrow | Solo Leisure | 1 |
| 3 Trip Verified London Heathrow to Mumbai in a Boeing... | 31st July 2023 | Iceland | Business Class | yes | 3 | Gatwick to Venice | Solo Leisure | 1 |
| 4 Trip Verified Keflavik Iceland to London Heathrow on ... | 31st July 2023 | Iceland | Business Class | yes | 5 | London to Luanda | Couple Leisure | 1 |
| 5 Trip Verified Terrible Experience with British Airways I... | 29th July 2023 | Canada | Economy Class | no | 5 | Denver to Heathrow | Family Leisure | 2 |
| 6 Trip Verified An airline that lives in their past glory an... | 26th July 2023 | Qatar | Business Class | no | 3 | BKK to LHR | Business | 1 |
| 7 Trip Verified Checkin Desk rude and dismissive Flight... | 26th July 2023 | United Kingdom | Economy Class | no | 3 | London to Tampa | Couple Leisure | 1 |
| 8 Trip Verified I chose British Airways especially becaus... | 25th July 2023 | Spain | Economy Class | no | 1 | London to Sydney | Couple Leisure | 2 |
| 9 Not Verified I booked Premium Economy from INV to ... | 24th July 2023 | United Kingdom | Premium Economy | no | 9 | LHR to CPT | Couple Leisure | 1 |
| 10 Trip Verified A simple story with an unfortunate outco... | 21st July 2023 | Germany | Economy Class | no | 7 | BLR to LHR | Business | 2 |
| 11 Trip Verified Flight was delayed due to the inbound fli... | 21st July 2023 | United Kingdom | Economy Class | no | 3 | Heathrow to Las Vegas | Family Leisure | 1 |
| 12 Not Verified Fast and friendly check in total contrast t... | 20th July 2023 | United Kingdom | Business Class | no | 5 | Rome to Heathrow | Solo Leisure | 1 |
| 13 Trip Verified I dont understand why British Airways is ... | 20th July 2023 | United Kingdom | Economy Class | no | 3 | Gatwick to Venice | Solo Leisure | 1 |
| 14 Not Verified Im sure that BA have gradually made thei... | 20th July 2023 | United Kingdom | Economy Class | no | 3 | London to Luanda | Couple Leisure | 1 |
| 15 Trip Verified Customer Service does not exist One wo... | 18th July 2023 | United Kingdom | Business Class | no | 5 | Denver to Heathrow | Family Leisure | 2 |
| 16 Trip Verified Another really great pair of flights on ti... | 17th July 2023 | United Kingdom | Business Class | yes | 5 | BKK to LHR | Business | 1 |
| 17 Not Verified Our A380 developed a fault taxiing to th... | 17th July 2023 | United Kingdom | Business Class | no | 3 | London to Tampa | Couple Leisure | 2 |
| 18 Not Verified Horrible airline Does not care about thei... | 16th July 2023 | United Kingdom | Economy Class | no | 3 | London to Sydney | Couple Leisure | 1 |
| 19 Trip Verified My family and I have flown mostly on Bri... | 9th July 2023 | United Kingdom | Premium Economy | no | 1 | LHR to CPT | Couple Leisure | 3 |
| 20 Trip Verified This has been by far the worst service I h... | 9th July 2023 | United States | Economy Class | no | 9 | BLR to LHR | Business | 1 |
| 21 Trip Verified In Nov 2022 I booked and paid for a retu... | 8th July 2023 | United Kingdom | Economy Class | no | 7 | Heathrow to Las Vegas | Family Leisure | 2 |
| 22 Not Verified BA is not treating its premium economy ... | 6th July 2023 | United Kingdom | Premium Economy | no | 3 | Rome to Heathrow | Solo Leisure | 1 |

Рис. 12. Фрагмент датасета с добавленными к нему кластерами.
Fig. 12. A fragment of the dataset with clusters added to it.

В первый кластер попали такие ключевые слова, как «flight», «business», «class», «lounge», «service», «comfortable», «boarding», «delayed», «seating», как следствие можно сделать следующий вывод о данных, входящих в данный кластер, часто упоминают бизнес-класс, лаундж-зоны, комфортные условия и обслуживание, иногда могут быть упоминания о задержках и процессе посадки. Значит персоны, попадающие в первый кластер, с большей долей вероятности могут быть клиенты бизнес-класса с опытом как положительным, так и негативным, а значит для улучшения и персонализации предложения в дальнейшем стоит обратить на те негативные ключевые слова, которые формируют кластер с целью повышения качества впечатления.

Второй кластер зависит от таких ключевых слов, как «cancelled», «refund», «website», «cost», «booking», «compensation», «customer», «service», «arrangements», тогда кластер включает отзывы с негативным опытом, связанным с отменами рейсов, процессом возврата средств, проблемами с бронированием и компенсацией, то есть в большей долей вероятности это клиенты, которые столкнулись с проблемами в обслуживании и не получили удовлетворительного решения. Как следствие необходимо стремиться уменьшение объема вхождения в данный кластер так, как формируемый экономический продукт в виде впечатления оказывается негативным.

Третий кластер характеризуется такими ключевыми словами, как «crew», «meal», «entertainment», «friendly», «comfortable», «recommend», «suite»,

«food», «flight», а значит отзывы в этом кластере в основном положительные, фокусируются на дружелюбном экипаже, качестве питания, развлечениях на борту и общих комфортных условиях полета, поэтому в с большей вероятностью клиенты, входящие в данный кластер, довольны своим опытом и в дальнейшем рекомендуют авиакомпанию, а значит впечатление будет положительным.

Четвертый кластер связан с такими ключевыми словами, как «economy», «class», «space», «legroom», «seats», «cramped», «uncomfortable», «food», «drink», как следствие отзывы в этом кластере касаются эконом-класса, неудобных сидений и ограниченного пространства для ног, можно сказать, что клиенты, возможно, жалуются на нехватку комфорта и качество еды и напитков, что в контексте экономики впечатлений не является показателем хорошего качества формируемого опыта. Таким образом понимая критерии определённого клиента, можно предполагать в какой кластер потребитель может попасть и тем самым минимизировать возникающую проблему.

Выводы

В ходе исследования была продемонстрирована эффективность применения методов машинного обучения для анализа текстовых данных на примере алгоритма K-means. Кластеризация отзывов клиентов авиакомпании позволила выявить значимые паттерны, связанные с различными аспектами клиентского опыта, такими как качество обслуживания, комфорт, решение проблем, связанных с бронированием и возвратами, и общее восприятие услуг. Анализ ключевых слов внутри

каждого кластера позволил сделать выводы о проблемных зонах и сильных сторонах авиакомпании. Применение данного подхода дает возможность предприятиям целенаправленно улучшать слабые стороны своей деятельности, минимизировать негативные впечатления клиентов и создавать условия для формирования положительного опыта.

Для дальнейших исследований целесообразно рассмотреть использование других алгоритмов

кластеризации, таких как DBSCAN или иерархический кластерный анализ, а также интеграцию предиктивных моделей для прогнозирования клиентских предпочтений. Совмещение методов машинного обучения с инструментами анализа эмоциональной окраски текста может дополнительно улучшить интерпретацию данных и разработку персонализированных стратегий взаимодействия с клиентами.

Список источников

1. British Airways Passenger Reviews (2016-2023) // Kaggle. URL: <https://www.kaggle.com/datasets/praveensaik/british-airways-passenger-reviews-2016-2023> (дата обращения 20.07.2024)
2. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. М.: Питер, 2020. 336 с.
3. Бурков А. Машинное обучение без лишних слов. М.: Питер, 2020. 160 с.
4. Гачаев А.М., Датаев А.А., Мурадова П.Р. Исследование технологии компьютерной обработки информации в условиях больших данных // Журнал монетарной экономики и менеджмента. 2023. № 1. С. 90 – 94.
5. Батаева Б.С., Карпов Н.А. Влияние факторов корпоративного управления на уровень раскрытия ESG-информации российскими публичными компаниями // Управленец. 2023. Т. 14. № 3. С. 30 – 43.
6. Копнин А.А. Способы и методы машинного обучения: интеграция, взаимодействие и дополнение методов // Наука и бизнес: пути развития. 2023. № 12 (150). С. 21– 24.
7. Ланц Б. Машинное обучение на R: экспертные техники для прогностического анализа. М.: Питер, 2020. 464 с.
8. Назаров Д.М., Рыжкина Д.А. Интеллектуальные средства бизнес-аналитики: учебное пособие. М.: КНОРУС, 2022. 242 с.
9. Назаров Д.М. Анализ семантики понятий машинного обучения и Data science // Современные тренды управления и цифровая экономика: от регионального развития к глобальному экономическому росту: сборник статей V Международной научно-практической конференции, Екатеринбург, 10 августа 2023 года / Институт цифровой экономики и права. Екатеринбург: Общество с ограниченной ответственностью "Институт Цифровой Экономики и Права", 2023. С. 42 – 46.
10. Стефанова Н.Л. Методы математической обработки данных: учебник и практикум для вузов М.: Юрайт, 2024. 317 с.
11. Благинин В.А., Соколова Е.В., Адакава М.И. Достижения и тенденции в области нейротехнологий и искусственного интеллекта в Российской Федерации: комплексный наукометрический анализ // Цифровые модели и решения. 2023. Т. 2. № 4. С. 13 – 29.
12. Элежбиев Б.Э., Джабраилов И.А., Юсупова Р.В. Использование технологии компьютерных баз данных в целях повышения эффективности управления информацией современного предприятия // Журнал монетарной экономики и менеджмента. 2023. № 1. С. 156 – 159.

References

1. British Airways Passenger Reviews (2016-2023). Kaggle. URL: <https://www.kaggle.com/datasets/praveensaik/british-airways-passenger-reviews-2016-2023> (date of access 20.07.2024)
2. Brink H., Richards D., Feverolf M. Machine learning. Moscow: Piter, 2020. 336 p.
3. Burkov A. Machine learning without unnecessary words. Moscow: Piter, 2020. 160 p.
4. Gachayev A.M., Dataev A.A., Muradova P.R. Research of computer information processing technology in the context of big data. Journal of monetary economics and management. 2023. No. 1. P. 90 – 94.
5. Bataeva B.S., Karpov N.A. The Impact of Corporate Governance Factors on the Level of ESG Information Disclosure by Russian Public Companies. Manager. 2023. Vol. 14. No. 3. P. 30 – 43.

6. Kopnin A.A. Machine Learning Methods and Techniques: Integration, Interaction, and Complementation of Methods. Science and Business: Development Paths. 2023. No. 12 (150). P. 21 – 24.
7. Lanz B. Machine Learning on R: Expert Techniques for Predictive Analysis. Moscow: Piter, 2020. 464 p.
8. Nazarov D.M., Ryzhkina D.A. Intelligent Business Analytics Tools: A Tutorial. M.: KNORUS, 2022. 242 p.
9. Nazarov D.M. Analysis of the semantics of the concepts of machine learning and Data science. Modern management trends and the digital economy: from regional development to global economic growth: collection of articles from the V International scientific and practical conference, Yekaterinburg, August 10, 2023. Institute of Digital Economy and Law. Yekaterinburg: Limited Liability Company "Institute of Digital Economy and Law", 2023. P. 42 – 46.
10. Stefanova N.L. Methods of mathematical data processing: textbook and workshop for universities M.: Yurait, 2024. 317 p.
11. Blaginin V.A., Sokolova E.V., Adakava M.I. Achievements and trends in the field of neurotechnology and artificial intelligence in the Russian Federation: a comprehensive scientometric analysis. Digital models and solutions. 2023. Vol. 2. No. 4. P. 13 – 29.
12. Elezhibiev B.E., Dzhabrailov I.A., Yusupova R.V. Using computer database technology to improve the efficiency of information management of a modern enterprise. Journal of Monetary Economics and Management. 2023. No. 1. P. 156 – 159.

Информация об авторе

Копнин А.А., Уральский государственный экономический университет, Kopnin_aa@usue.ru

© Копнин А.А., 2025