ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ. ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

DOI: 10.6060/ivecofin.2026682.783
УДК: 658.562:004.67

ИНСТРУМЕНТЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ПРОЦЕССАХ УПРАВЛЕНИЯ КАЧЕСТВОМ ТЕКСТИЛЬНОГО ПРЕДПРИЯТИЯ

А.А. Шибаева, А.А. Хомякова, В.А. Хомяков

Александра Александровна Шибаева*, Анна Александровна Хомякова (ORCID 0000-0002-2085-2935), Ивановский государственный химико-технологический университет, пр. Шереметевский, 7, Иваново, 153000, Россия
E-mail: alexandra.splav@gmail.com*, khaa@isuct.ru
Владислав Андреевич Хомяков (ORCID 0009-0002-7218-4550)
Московский физико-технический институт (национальный исследовательский университет»), Институтский пер., 9, г. Долгопрудный, Московская область, 141701, Россия
E-mail: khomiakov.va@phystech.edu

В условиях активной цифровизации промышленности приобретает актуальность интеграция инструментов интеллектуального анализа данных в системы менеджмента качества. Работа посвящена адаптации методов кластеризации для изучения базы дефектов текстильного предприятия. Авторы обосновывают необходимость цифровой трансформации процессов контроля качества для обеспечения прослеживаемости и перехода от пассивной фиксации брака к предиктивному управлению. Проведен анализ лучших практик крупных российских промышленных компаний в области применения технологий искусственного интеллекта для контроля качества продукции. Объектом исследования выступает действующее текстильное предприятие, на материалах которого выполнено функциональное моделирование бизнес-процессов управления качеством. Проведен аудит и предварительная обработка производственного массива данных объемом более 51 тысячи записей, содержащего сведения о типах и видах брака, номенклатуре тканей и сортности готовой продукции. Разведочный анализ данных позволил выявить структуру распределения дефектов по фабрикам и видам ткани. Применены и сопоставлены три алгоритма кластеризации, предназначенные для работы с разными типами данных, а именно k-means, k-prototypes и иерархическая кластеризация с метрикой Гауэра. Оценка качества разбиения по коэффициенту силуэта показала, что алгоритм k-prototypes обеспечивает наилучшие результаты для смешанных данных (0,47 против 0,25 у остальных методов). На основе выделенных кластеров сформулированы адресные рекомендации для отдела контроля качества, направленные на снижение доли брака по конкретным видам тканей и типам оборудования. Предложенная методика может быть масштабирована на предприятия смежных отраслей легкой промышленности.

Ключевые слова: управление качеством, интеллектуальный анализ данных, кластеризация, алгоритм K-Prototypes, смешанные данные, цифровизация производства, коэффициент силуэта.

DATA MINING TOOLS IN QUALITY MANAGEMENT OF A TEXTILE ENTERPRISE

A.A. Shibaeva, A.A. Khomyakova, V.A. Khomyakov

Aleksandra A. Shibaeva*, Anna A. Khomyakova (ORCID 0000-0002-2085-2935)
Ivanovo State University of Chemistry and Technology, Sheremetevsky Ave., 7, Ivanovo, 153000, Russia
E-mail: alexandra.splav@gmail.com*, khaa@isuct.ru
Vladislav A. Khomyakov (ORCID 0009-0002-7218-4550)
Moscow Institute of Physics and Technology (National Research University), Institutsky Lane, 9, Dolgoprudny, Moscow Region, 141701, Russia

In the context of active industrial digitalization, integrating data mining tools into quality management systems is becoming increasingly relevant. This study focuses on adapting clustering methods to ana-

lyze the defect database of a textile enterprise. The authors argue for digital transformation of quality control processes to ensure traceability and enable a shift from passive defect recording to predictive management. Best practices of major Russian industrial companies in applying artificial intelligence technologies to product quality control are reviewed. The object of the study is an operating textile enterprise, whose operations served as the basis for functional modeling of quality management business processes. An audit and preprocessing of a production dataset comprising over 51,000 records have been carried out, covering defect types and categories, fabric nomenclature, and finished product grading. Exploratory data analysis reveales the distribution patterns of defects across factories and fabric types. Three clustering algorithms designed for different data types are applied and compared, namely k-means, k-prototypes, and hierarchical clustering with the Gower distance metric. Cluster quality evaluation using the silhouette coefficient shows that k-prototypes yield the best results for mixed data (0.47 versus 0.25 for the other methods). Based on the identified clusters, targeted recommendations for the quality control department have been formulated, aimed at reducing defect rates for specific fabric types and equipment categories. The proposed methodology can be scaled to enterprises in related sectors of the light industry.

Keywords: quality management, data mining, clustering, k-prototypes algorithm, mixed data, production digitalization, silhouette coefficient.

Для цитирования:
Шибаева А.А., Хомякова А.А., Хомяков В.А. Инструменты интеллектуального анализа данных в процессах управления качеством текстильного предприятия. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2026. №02(68). С. 85-98. DOI: 10.6060/ivecofin.2026682.783

For citation:
Shibaeva A.A., Khomyakova A.A., Khomyakov V.A. Data mining tools in quality management of a textile enterprise. *Ivecofin.* 2026. N 02(68). P. 85-98. DOI: 10.6060/ivecofin. 2026682.783 (in Russian)

ВВЕДЕНИЕ

Промышленные производства переживают период активной трансформации, связанной с внедрением инструментов искусственного интеллекта, обработки больших массивов информации и создания цифровых двойников. На этом фоне цифровизация систем менеджмента качества (рис. 1) становится не факультативным направлением, а условием сохранения конкурентоспособности. Система должна обеспечить сквозную прослеживаемость, управление данными и получение информации в режиме реального времени для обеспечения полной прозрачности и контроля. В современных реалиях это становится обязательным условием для непрерывного улучшения процессов контроля качества организации.

Одним из ключевых направлений развития системы менеджмента качества отечественных компаний является развитие систем интеллектуального анализа данных. Как показало изучение лучших практик управления качеством активнее всего технологии искусственного интеллекта осваивают в отраслях металлургии, добычи и переработки нефти, энергетики [1]. Так компания «Газпромнефть-Снабжение» внедрила в работу систему искусственного интеллекта для проведения входного контроля материально-технических ресурсов [2]. Система сопоставляет внедряемое в производство оборудование с цифровыми моделями и подтверждает соответствие деталей заявленным показателям. Результаты проведенных проверок автоматически загружаются в систему для формирования актов входного контроля. Компания «РТ-Техприемка» внедрила программно-аппаратный комплекс на основе искусственного интеллекта, нейронных сетей и системы машинного зрения для контроля качества стальных изделий для вертолетов [3]. Программный комплекс на основании заданных характеристик выявляет дефекты изделия, в том числе микродефекты размером 300мкм, и классифицирует продукцию. Дочерняя компания холдинга СИБУР «Нижнекамскнефтехим» использует специализированную систему, которая позволяет в реальном времени по текущим значениям техпараметров спрогнозировать качество продукции и, соответственно, внести необходимые корректировки в процесс [4]. Также компания внедрила инструменты предиктивной диагностики оборудования на основе интеллектуального анализа данных, которые предупреждают об отклонении от рабочих режимов предотвращая аварии и внеплановые остановки производства, из-за которых происходят ключевые экономические потери и потери в качестве. Компания «Северсталь» реализует проект по цифровизации контроля качества продукции, внедряя систему сквозной автоматической аттестации металлопроката [5]. Система

на основе прогнозной аналитики рассчитывает вероятности возникновения дефектов на новых единицах продукции на основании технологических параметров. Система автоаттестации обладает инструментами для поиска первопричин отклонений по качеству, анализа трендов и построения моделей, обеспечивая 100% контроль продукции и минимизируя воздействие человеческого фактора.

Рисунок 1. Направления цифровизации систем менеджмента качества (составлено авторами)
Figure 1. Directions of quality management system digitalization (compiled by the authors)

Цель проводимого нами исследования – разработка рекомендаций по совершенствованию системы управления качеством текстильного предприятия на основе внедрения инструментов интеллектуального анализа данных. Для текстильной отрасли, где значительная часть операций контроля качества до сих пор опирается на визуальный осмотр и бумажные журналы, переход к цифровым инструментам анализа дефектов способен существенно сократить долю брака и снизить издержки [6].

МАТЕРИАЛЫ И МЕТОДЫ

В ходе работы было произведено функциональное моделирование процесса управления качеством предприятия ООО «Ткань» (название предприятия изменено в целях защиты коммерческой информации). Процесс управления качеством (рис. 2) включает несколько ключевых этапов: планирование, обеспечение, контроль и улучшение качества.

На этапе планирования качества определяются стратегические цели и задачи предприятия в области качества, разрабатываются соответствующие процессы и методы их реализации, также определяются целевые показатели качества. На этапе обеспечения качества технолог предприятия создает технологические карты и технические условия, определяет параметры процессов и оборудования для соответствия заданным показателям качества. Контроль качества включает комплекс мероприятий по выявлению дефектов, несоответствий и отклонений от заданных параметров на всех этапах жизненного цикла продукции. Он предполагает систематический сбор данных о процессах изготовления продукции и документирова-

ние этих данных и всех несоответствий для дальнейшего анализа. На этапе улучшения полученные данные сравниваются с запланированными показателями качества, выявляются причины несоответствий и брака. Итогом этого процесса является создание плана мероприятий по повышению качества, который затем используется на этапе планирования. Это обеспечивает непрерывное улучшение системы управления качеством.

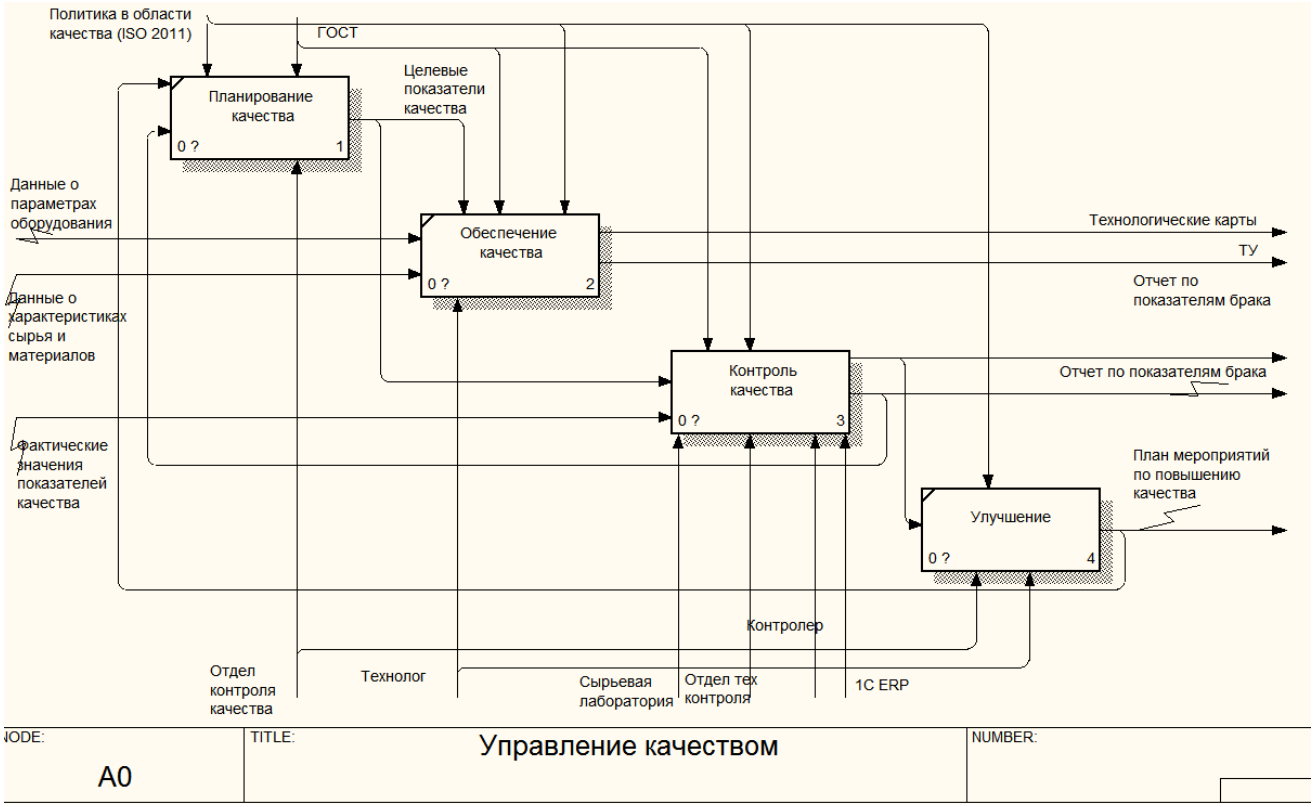

Рисунок 2. Функциональная модель процесса управления качеством текстильного предприятия (модель AS IS)

Figure 2. Functional model of the textile enterprise quality management process (AS IS model)

С точки зрения внедрения методов интеллектуального анализа данных в условиях исследуемого предприятия наиболее перспективным является процесс контроля качества (рис. 3), так как на этом этапе предприятие получает наибольший объем разнообразных данных пригодных для последующего анализа. В функциональной модели представлены четыре основных этапа: входной контроль сырья, контроль ткацкого и отделочного производства, а также контроль качества готовой продукции. На этапе входного контроля сырьевая лаборатория проверяет соответствие сырья и материалов стандартам и политике качества, фиксируя результаты в бумажных журналах. В процессе контроля ткацкого производства отдел техконтроля проверяет параметры работы оборудования, также фиксируя отклонения в журналах. Контроль отделочного производства включает проверку параметров оборудования с заданной периодичностью, результаты и отклонения фиксируются в журналах. На конечном этапе в складальном цехе проводится 100%-ный визуальный контроль качества готовой ткани с определением сортности и видов брака, информация о дефектах заносится в ERP систему для дальнейшего анализа. В дальнейшем по результатам работы складального цеха можно сформировать отчет в системе 1С:ERP.

На выходе этапов контроля качества мы получаем информацию об отклонении фактических показателей качества от заданных и отчет в электронном виде по уровню и видам брака готовой продукции, который впоследствии можем использовать для применения методов ИАД.

В ходе проводимого исследования нами был проведен аудит данных о процессах качества, были собраны данные о процессах контроля качества, о выпускаемом продукте и данные о дефектах. Был произведен разведочный анализ данных, очистка данных, обработка пропусков, работа с дубликатами, идентификация и работа с выбросами. В итоге был получен чистый набор данных пригодный для анализа (рис. 4).

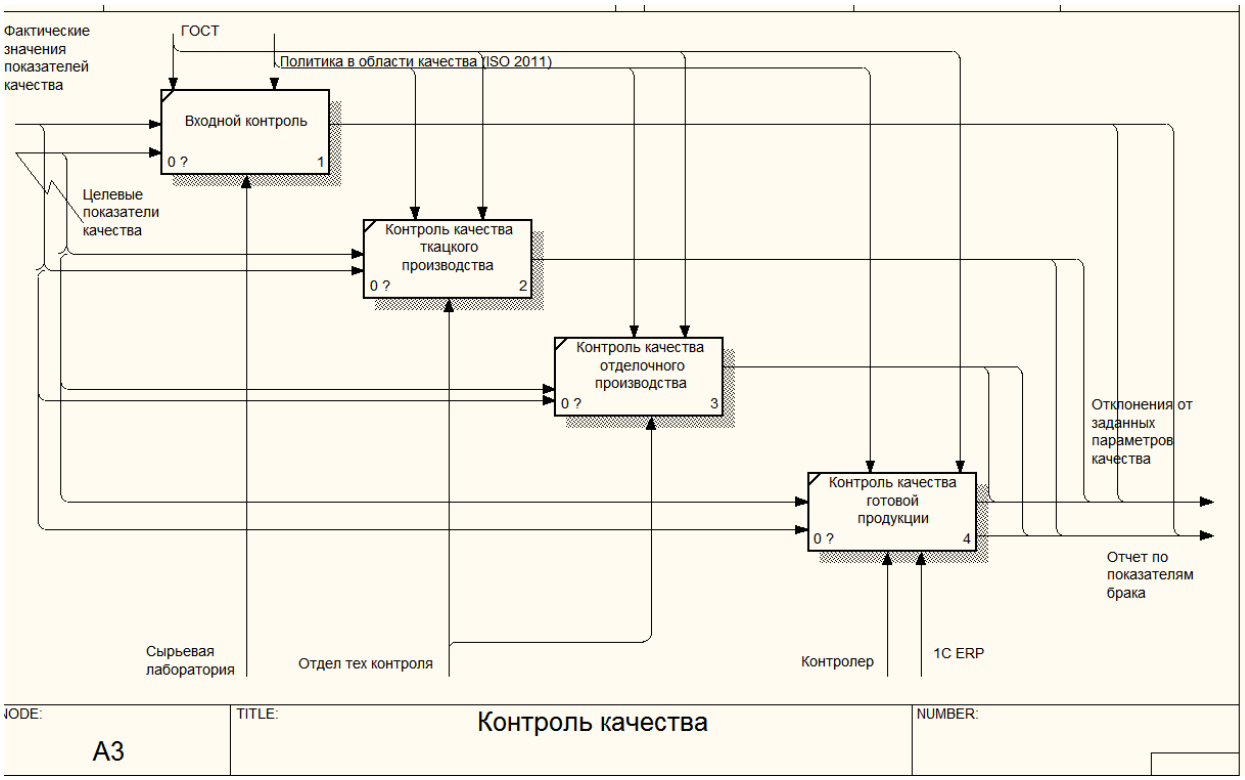

Рисунок 3. Функциональная модель процесса контроля качества текстильного предприятия (модель AS IS)

Figure 3. Functional model of the textile enterprise quality control process (AS IS model)

```
df.head()
```

	Тип брака	Вид брака	Номенклатура ткани	Ткацкая фабрика №	Номер куска	Длина, м	1 сорт	2 сорт	м/л	в/л	перепр
0	ткацкий брак	Бахрома на кромке	Бязь 262, 220, набивная мерный лоскут	2	19870611	3,2	NaN	NaN	3,2	NaN	NaN
1	ткацкий брак	Бахрома на кромке	Вафельная 1151, 45, отбелённая мерный лоскут	1	19807795	16,5	NaN	NaN	16,5	NaN	NaN
2	ткацкий брак	Бахрома на кромке	Перкаль, 155, гладкокрашеная мерный лоскут	2	19865139	15,0	NaN	NaN	15,0	NaN	NaN
3	ткацкий брак	Бахрома на кромке	Перкаль, 155, набивная мерный лоскут	2	19865099	8,2	NaN	NaN	8,2	NaN	NaN
4	ткацкий брак	Бахрома на кромке	Перкаль, 227, отбелённая мерный лоскут	2	19815045	9,2	NaN	NaN	9,2	NaN	NaN

Рисунок 4. Фрагмент исследуемого набора данных
Figure 4. Fragment of the studied dataset

Приведем краткое описание переменных (признаков):

«Тип брака» - указывается принадлежность брака, отдел в котором этот брак имел место, прядильный, ткацкий или отделочный,

«Вид брака» - определяется в соответствии с ГОСТ или внутренним регламентом предприятия,

«Номенклатура» - указывается номенклатура ткани. Например, вид ткани перкаль/ бязь/ вафельная и т.д., ширина ткани в готовом виде 155/220/ 227 и т.д., и вид отделки отбеленная/набивная/гладкокрашеная,

«Длина» - указано количеством ткани в метрах:

- «1 сорт» - ткань 1 сорта согласно ГОСТ не имеет дефектов и брака,

- «2 сорт» - ткань 2 сорта может иметь незначительные визуальные дефекты согласно ГОСТ,

- «м/л» - мерный лоскут, ткань с незначительными дефектами, которая потом идет в реализацию,

- «в/л» - весовой лоскут, ткань с дефектами, которая потом идет в реализацию на вес,

- «перепр.» - «переправа» означает, что ткань имела исправимые дефекты, например, намины или недостаточную ширину и была отправлена обратно в производство на исправление.

Полученный набор данных имеет 51996 строк, из них строки с 12 по 42192 относятся к сортовой ткани, не имеющей брака. При анализе

брака готовой ткани эти строки не будут являться информативными.

РЕЗУЛЬТАТЫ

Как видно на рис. 5 в исследуемом наборе данных преобладает с большим перевесом брак возникающий на этапе отделки ткани, остальные типы брака примерно равны по частоте упоминаний.

В видах брака (рис.6) преобладают грязные помарки, разделка, помарки, утолщенная нить и останов машины. Но просто частота появления определенного дефекта для нас мало информативна, следует построить более информативные графики.

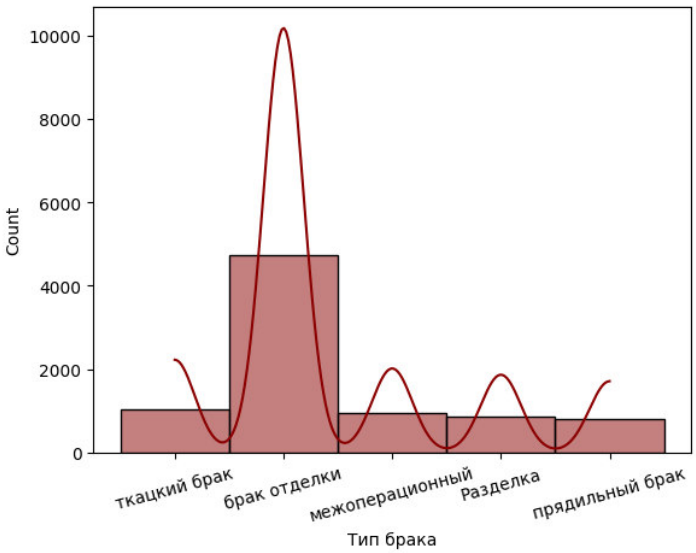

Рисунок 5. Распределение по типу брака
Figure 5. Distribution by defect type

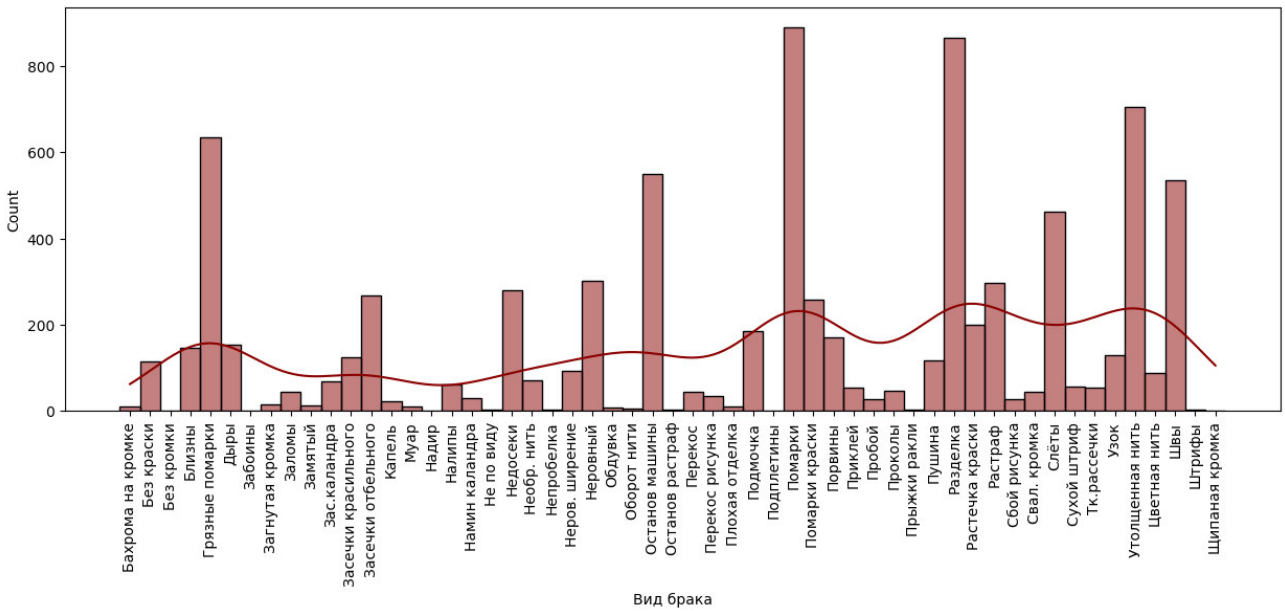

Рисунок 6. Распределение видов брака
Figure 6. Distribution of defect categories

Сравнительная гистограмма (рис. 7) отражает, как распределяются типы дефектов в зависимости от того на какой фабрике произведена ткань. На построенной диаграмме видно, что для тканей с ткацкой фабрики №1 в большей степени характерны ткацкие и прядильные браки, а для ткани с фабрики №2 характерны браки процесса отделки. Если учесть, что на второй фабрике вырабатывают в основном широкие ткани, то можно сделать вывод, что у предприятия преобладают проблемы с отделочным оборудование по ширине 220.

На гистограмме распределения вида дефектов в зависимости от ткацкой фабрики (рис. 8), видно, что для тканей с фабрики №2 характерными

видами брака являются грязные помарки, останов машины и швы. А для тканей с фабрики №1 характерны недосеки, слеты и утолщенная нить. Очевидно, что отделу управления качеством стоит обратить внимание на настройку ткацкого оборудования и качество пряжи на ткацкой фабрике №1.

В тканях 2-го сорта преобладают ткани с ткацкой фабрики №1 и прядильными и ткацкими браками (рис.9). Следует учесть, что 2-й сорт имеет мелкие дефекты и продается с небольшим дисконтом относительно тканей 1 сорта. Ткани 2-го сорта, возможно, не приносят значительных экономических потерь, поскольку продаются с небольшим дисконтом. Но так как дефекты незначительные, их легко устранить, улучшив обслуживание и настройку ткацких станков фабрики №1, что даст существенный экономический эффект. В категории брака «мерный лоскут» (рис. 10), суммарно по длине преобладают браки отделочной фабрики характерные для тканей с ткацкой фабрики №2. Отдел управления качеством должен явно обратить внимание на качество работы отделочного оборудования, суммарно более 35 км ткани ушло в категорию «мерный лоскут» по вине отделки, что при средней стоимости погонного метра ткани означает ощутимые финансовые потери. Также нужно отметить, что количество критических ткацких и прядильных браков у тканей с первой и второй фабрик примерно одинаково.

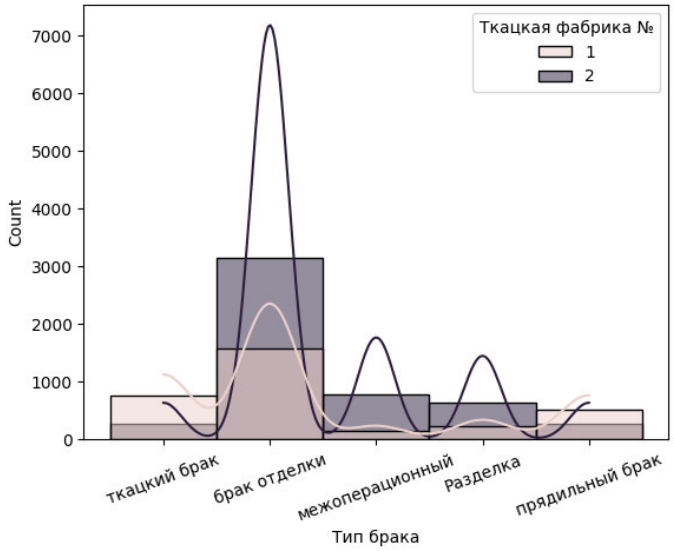

Рисунок 7. Распределение типов дефектов в зависимости от ткацкой фабрики
Figure 7. Distribution of defect types by weaving factory

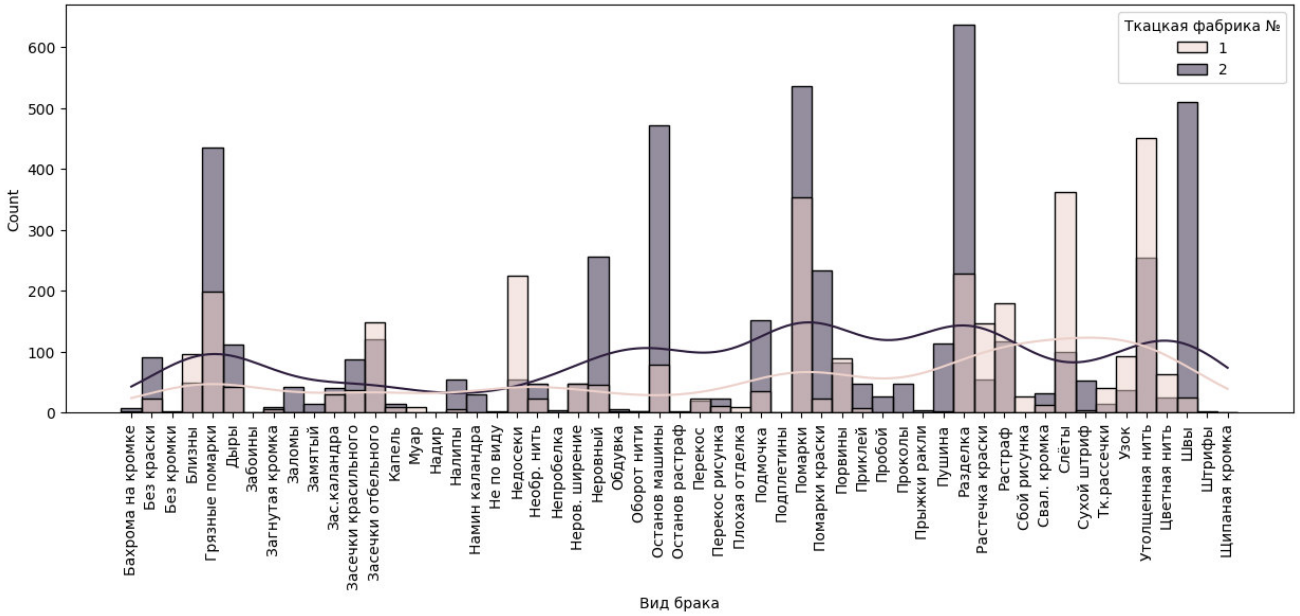

Рисунок 8. Распределение видов дефектов в зависимости от ткацкой фабрики
Figure 8. Distribution of defect categories by weaving factory

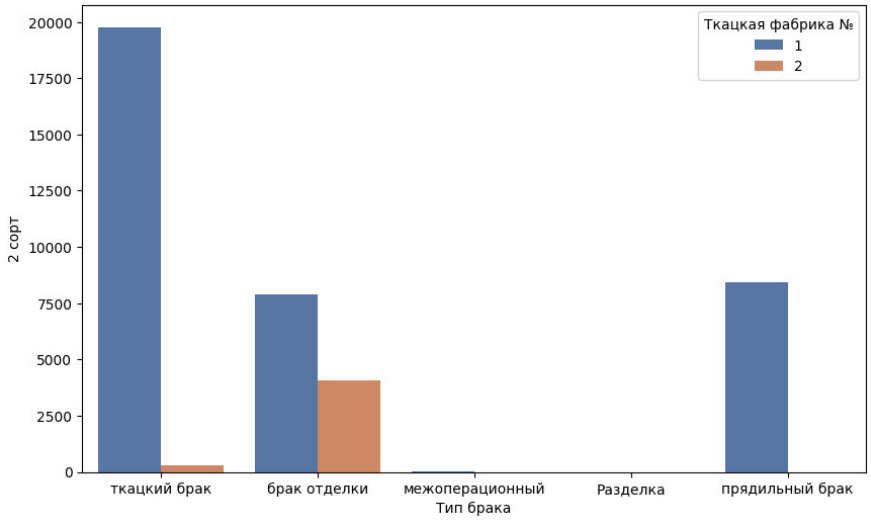

Рисунок 9. Суммарная длина ткани 2-го по типу дефекта
Figure 9. Total length of second-grade fabric by defect type

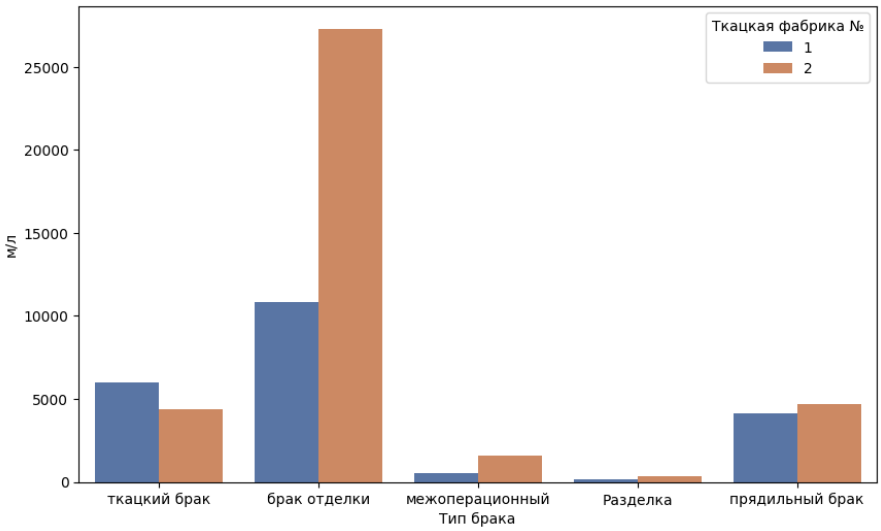

Рисунок 10. Диаграмма суммарная длина ткани «мерный лоскут» по типу дефекта
Figure 10. Total length of cut remnant fabric by defect type

Для поиска скрытых закономерностей в процессе исследования была применена кластеризация [7]. Она используется для автоматического разделения набора данных на группы (кластеры) схожих объектов, где элементы внутри группы максимально похожи, а в разных – отличаются. Кластерный анализ предъявляет следующие требования к исходным данным: показатели не должны коррелировать между собой; показатели должны быть безразмерными; распределение показателей должно быть близко к нормальному; показатели должны отвечать требованию «устойчивости», под которой понимается отсутствие влияния на их значения случайных факторов; выборка должна быть однородна, не содержать «выбросов». Очевидно, что на практике эти требования в полной мере никогда не выполняются [8], поэтому в ходе проводимого нами исследования были подобраны методы кластеризации, показывающие лучшие результаты на наших данных по следующим критериям:

1) совместимость с типом данных: в нашем случае важна работа со смешанными данными (категориальные + числовые) без потери информации;
2) интерпретируемость результатов: возможность объяснить почему объект попал в конкретный кластер;
3) устойчивость: показывает, насколько различными получаются результирующие разбиения на группы после многократного применения алгоритмов кластеризации для одних и тех же данных [9]. Этот критерий обеспечивает корректную работу при доминировании одного кластера (по результатам проведенного нами исследования такой кластер был выделен; "брак отделки" ~5000 из ~9000);

4) чувствительность к гиперпараметрам: критерий учитывает насколько сложно подобрать параметры и насколько полученный результат стабилен.

Таким образом были применены три алгоритма кластеризации: k-prototypes, иерархическая кластеризация и k-means (в роли базового бэнчмарка). Рассмотрим каждый по отдельности.

K-means (метод k-средних) - является одним из самых популярных алгоритмов машинного обучения без учителя, предназначенный для кластеризации [10]. Алгоритм разбивает набор данных на определенное количество групп. Главная цель алгоритма минимизировать суммарное квадратичное отклонение точек от центров их кластеров. На каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем объекты снова разбиваются на кластеры в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения внутрикластерного расстояния [11]. Основным преимуществом алгоритма K-means является простота реализации и скорость работы, алгоритм работает быстро даже на больших объемах данных. Однако, метод требует максимально равномерно распределённые данные и самое главное – однотипные. Эти условия делают данный метод наименее подходящим для решения задачи кластеризации исследуемых данных по дефектам.

k-prototypes - метод кластеризации, объединяющий алгоритмы K-means, который работает для числовых данных, и K-modes, предназначенный для категориальных. Он эффективно обрабатывает смешанные типы данных, используя евклидово расстояние для числовых признаков и совпадения для категориальных. Это улучшение позволяет кластеризовать наборы данных, содержащие одновременно числовые и категориальные признаки, что невозможно для обычного K-means [12]. Преимуществом алгоритма является то, что он не требует предварительного кодирования, например, One-Hot Encoding, для категориальных признаков, что сохраняет смысл данных и хорошо работает на больших наборах данных, так как является методом разделения. Расстояние между объектом X_i прототипом кластера Q_j:

$$D(X_i, Q_j) = \sum_{l=1}^{p} (x_{il} - q_{il})^2 + \gamma \sum_{l=p+1}^{m} \delta(x_{il} - q_{il}), \quad (1)$$

$$\delta(x, q) = \begin{cases} 0, если\ x = q \\ 1, если\ x \neq q \end{cases}$$

где p – количество числовых атрибутов;

m – общее количество атрибутов;

γ – весовой параметр (гиперпараметр), балансирующий влияние категориальных признаков;

$\delta(x, q)$ - функция несовпадения.

Метод реализует нативную поддержку через Евклидово расстояние для чисел («Длина», «Сорт») и расстояние Хэмминга для категорий («Вид брака», «Фабрика»), что обеспечивает соответствие критерию совместимости с типом анализируемых данных. Центроиды - объекты, определяющие кластер, для числовых данных интерпретируются через среднее значение, для категориальных – через модальное значение. Остальные объекты являются максимально схожими к ним. В результате описать кластер достаточно легко (например, «Фабрика 2 + Перкаль + Длина 15м»). Однако, метод k-prototypes склонен создавать большие кластеры вокруг мажоритарного класса, что снижает его устойчивость к дисбалансу. Для исключения этого недостатка можно увеличить вес γ для категориальных признаков, чтобы редкие виды брака не поглощались. При этом следует отметить чувствительность результата к гиперпараметрам, поэтому γ следует подбирать аккуратно.

Иерархическая кластеризация - метод машинного обучения без учителя, который строит дерево вложенных кластеров для структурирования данных. В данном исследовании использовался агломеративный метод иерархической кластеризации, в котором каждая точка сначала считается отдельным кластером. Затем алгоритм постепенно объединяет самые похожие пары в группы, пока не получится один, включающий в себя все группы большой кластер. Этот метод не требует заранее заданное количество кластеров, а для объединения используются метрики близости [13]. Для двух объектов i и j сходство S_{ij} вычисляется как взвешенное среднее частичных сходств по всем признакам k:

$$S_{ij} = \frac{\sum_{k=1}^{p} w_k \cdot \delta_{ijk} \cdot s_{ijk}}{\sum_{k=1}^{p} w_k \cdot \delta_{ijk}}, \quad (2)$$

где p – общее количество атрибутов;

w_k - вес k-го признака (обычно 1, если не задано иное);

δ_{ijk} - индикатор доступности данных (1 если значение известно для обоих объектов, 0 если есть пропуск);

s_{ijk} - частичное сходство по k-му признаку (нормированное от 0 до 1).

Таким образом, совместимость с типом данных обеспечивается нативной поддержкой через матрицу расстояний. Для этого используется метрика близости Gower Distance, разработанная

для вычисления расстояния между объектами, описанными смешанными типами данных. Интерпретируемость результатов применения метода обеспечивается посредством построения дендрограммы – визуального дерева слияний. Метод нативно минимизирует дисперсию внутри кластеров, что помогает выделять малые группы: можно увидеть на дендрограмме как малые ветви. Это повышает устойчивость метода к дисбалансу [14]. Дендрограмма позволяет визуально выбрать порог чувствительности к гиперпараметрам: distance_metric: метрика, по которой мы строим матрицу расстояний, в нашем случае мы взяли её как gower distance, которая хорошо учитывает смешанные типы данных; linkage – метрика, определяющая как вычисляется расстояние между двумя кластерами (множествами точек); cut_treshold – константа, определяющая максимальное расстояние, на котором можно объединять кластеры.

Для определения оптимального количества кластеров был применен метод локтя (рис. 11) [15].

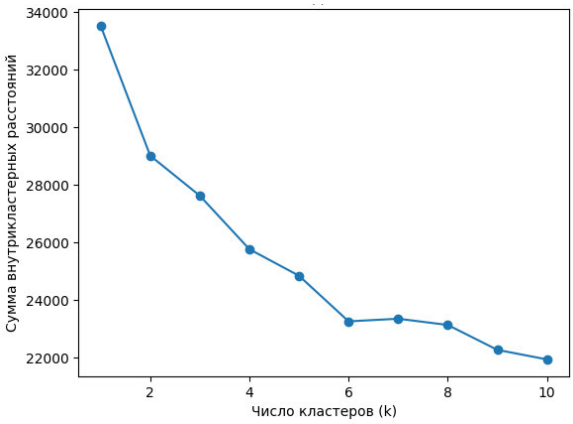

Рисунок 11. Определение количества кластеров методом локтя д ля исследуемого набора данных
Figure 11. Determining the number of clusters using the elbow method for the studied dataset

На данном графике локоть выражен не четко, определить его трудно, субъективно это будет точка от 6 до 9 кластеров. В таких случаях часто используют дополнительные метрики, например, коэффициент силуэта [16].

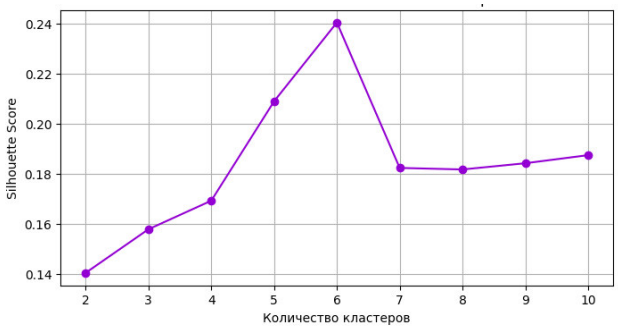

Рисунок 12 Определение количества кластеров методом силуэта
Figure 12 Determining the number of clusters using the silhouette method

Метод силуэта измеряет, насколько хорошо объект соответствует своему кластеру по сравнению с другими. Высокое среднее значение силуэта указывает на лучшее разделение [17, 18]. По графику (рис.12) четко видно, что оптимальное число кластеров равно 6.

Результат кластеризации при помощи алгоритма K-means представлен на рис.13.

Для определения оптимального количества кластеров для k-prototypes также применялся метод локтя. На рис 14 видно, что оптимальное количество кластеров для алгоритма K-prototypes – 5 кластеров.

Результат работы K- prototypes представлена на рис.15.

Результат применения агломеративного метода иерархической кластеризации представлен на рис. 16. Результат иерархической кластеризации для 6 кластеров представлен на рис. 17.

Как видим по результатам исследования алгоритмы k-means и иерархическая кластеризация разделили датасет на 6 кластеров, алгоритм k-prototypes – на 5. Явных ошибок/галлюцинирующих результате работы алгоритмов нет. Каждый метод показал, что невозможно, поделить данные на кластеры с одинаковым количеством элементов, но, видимо это особенность исследуемых данных. Все используемые методы определили основной кластер, количество элементов в котором более 3000 шт. и остальные кластеры в которых менее 1600 элементов. Минимальный кластер у методов k-means и иерархическая кластеризация состоит из 794 элементов, а у k-prototypes – 716.

cluster_label	Ткацкая фабрика №	Тип брака	Вид брака	Номенклатура ткани	
0	1	брак отделки	Помарки	Бязь 262, 150, набивная	1577
1	2	Разделка	Разделка	Перкаль, 223, набивная	866
2	2	брак отделки	Помарки	Перкаль, 223, набивная	3155
3	1	прядильный брак	Утолщенная нить	Бязь 262, 150, набивная	794
4	2	межоперационный	Швы	Перкаль, 223, набивная	802
5	1	ткацкий брак	Слёты	Полина, 80, набивная	1167

Silhouette Score Kmeans: 0.239

Рисунок 13. Результат кластеризации методом K-means
Figure 13. K-means clustering results

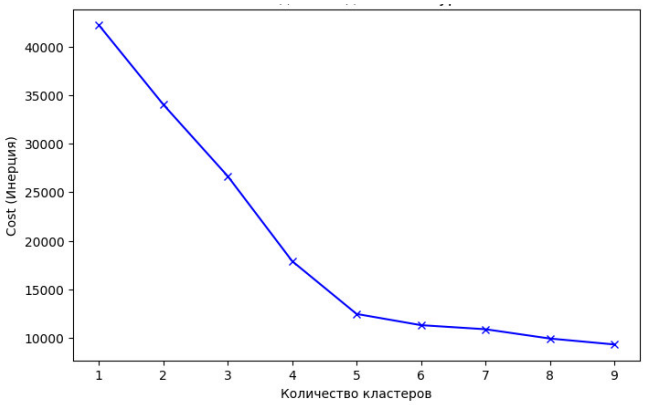

Рисунок 14. Определение количества кластеров методом локтя для K-prototypes
Figure 14. Determining the number of clusters using the elbow method for K-prototypes

cluster_KP	Ткацкая фабрика №	Тип брака	Вид брака	Номенклатура ткани	
0	1	брак отделки	Помарки	Бязь 262, 150, набивная	1460
1	2	межоперационный	Швы	Перкаль, 223, набивная	716
2	2	брак отделки	Помарки	Перкаль, 223, гладкокрашеная	3333
3	1	ткацкий брак	Слёты	Полина, 80, набивная	1000
4	1	прядильный брак	Утолщенная нить	Бязь 262, 150, набивная	1039

Silhouette Score k-prototype (Gower): 0.479

Рисунок 15. Результат кластеризации методом K-prototypes
Figure 15. K-prototypes clustering results

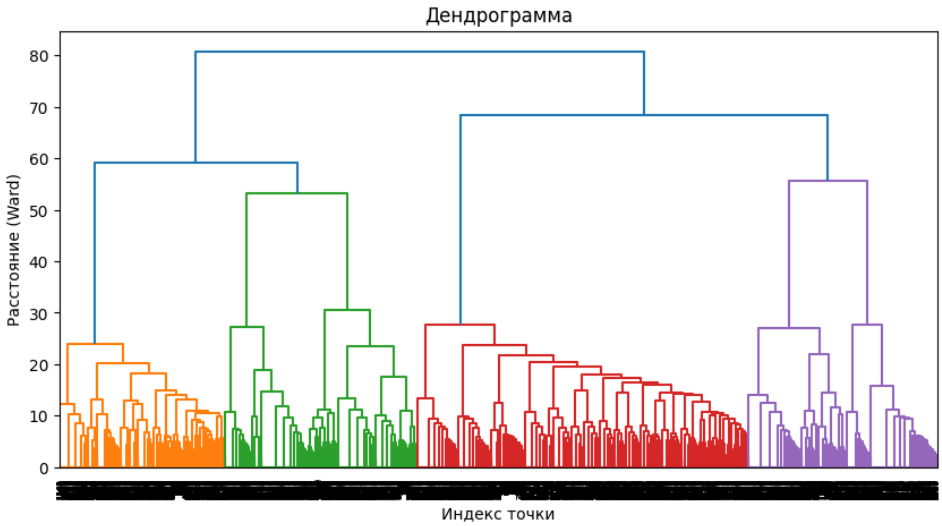

Рисунок 16. Дендрограмма для исследуемого набора данных
Figure 16. Dendrogram for the studied dataset

cluster_Dendro	Ткацкая фабрика №	Тип брака	Вид брака	Номенклатура ткани	
0	1	ткацкий брак	Слёты	Бязь 262, 150, набивная	1033
1	2	Разделка	Разделка	Перкаль, 223, набивная	866
2	2	брак отделки	Помарки	Перкаль, 223, набивная	3155
3	1	брак отделки	Помарки	Бязь 262, 150, набивная	1577
4	2	межоперационный	Швы	Перкаль, 223, набивная	936
5	1	прядильный брак	Утолщенная нить	Бязь 262, 150, набивная	794

Silhouette Score AgglomerativeClustering: 0.244

Рисунок 17. Результат иерархической кластеризации
Figure 17. Hierarchical clustering results

Алгоритмы дают одинаковые выводы о том, что основным видом брака является брак отделочной фабрики «Помарки», который составляет самый большой кластер по ткани Перкаль 223 со 2-й ткацкой фабрики. И кластер, идущий на втором месте по количеству элементов в работе всех трех алгоритмов, брак отделочной фабрики «Помарки» по ткани Бязь шириной 150см с 1-ой фабрики. Также алгоритмы выделили кластер, с примерно одинаковым количеством элементов для каждого метода, с ткацким браком «слеты» по ткани с 1-ой фабрики. Но алгоритмы k-prototypes и k-means рекомендовали обратить внимание на ткань Полина 80, а иерархическая кластеризация указала на ткань Бязь 150. Также все три метода совпали во мнениях при создании кластеров с прядильным браком «утолщенная нить» по ткани Бязь 150 и межоперационным браком «швы» по ткани Перкаль 223. Стоит отметить, что алгоритмы k-means и иерархическая кластеризация выделили отдельный кластер с количеством элементов 866 для брака Разделка по ткани Перкаль, k-prototypes такой кластер не создал вообще.

По итогам применения трех методов кластеризации можно сказать, что в целом результаты работы алгоритмов очень похожи и принципиальных различий в их работе очень мало. При небольших вычислительных мощностях можно рекомендовать использование метода k-means, который работает быстро и с большими объемами данных. Но в рассматриваемом наборе данных большое количество категориальных признаков, а метод k-means работает только с числовыми значениями. Использование кодирования часто «раздувает» пространство признаков и портит качество кластеризации. Метод k-prototypes работает и с числовыми и категориальными данными, но чем больше в данных уникальных категорий, тем сильнее падает скорость работы алгоритма. Недостатка методов k-means и k-prototypes является сложное определение количества кластеров перед запуском алгоритма. Преимуществом метода иерархической кластеризации является дендрограмма, по которой легко визуально определить количество кластеров. Алгоритм хорошо работает со сложными формами кластеров и на объемах данных до 20 тыс. строк и для требуемых предприятию нужд этого вполне достаточно. Для более точного выбора лучшего метода нужно провести оценку качества кластеризации [19-20].

Для оценки качества кластеризации была использована метрика качества кластеризации «Коэффициент Силуэта», которая измеряет, насколько похож объект на объекты своего кластера по сравнению с объектами других кластеров. Значения могут варьироваться от -1 до 1, где более высокие значения указывают на лучшую кластеризацию. Все коэффициенты качества кластеризации имеют положительные значения, что в среднем элементы кластеров находятся ближе к своим центрам, чем к границам чужих кластеров. То есть алгоритм нашел реальную структуру, а не просто разбил данные случайным образом. Значения silhouette score в районе 0,25 у алгоритмов K-means и иерархической кластеризации означают, что кластеризация имеет слабую структуру и, возможно, границы между кластерами не четкие или группы находятся слишком близко друг к другу. Коэффициент качества кластеризации равный 0,47 для K-Protoypes является достаточно хорошим результатом для реальных производственных данных, в которых есть шумы и информация собирается из разных отделов. Скорее всего, не все кластеры являются изолированными, некоторые соприкасаются границами или имеют сложную форму. Структура кластеризации есть, но она не идеальна. Возможно, некоторые записи похожи сразу на два кластера [21]. В целом группы выделены верно и данные кластеризации K-Protoypes можно использовать для выводов.

Опираясь на результаты кластеризации методом K-Prototypes, можно сделать ряд конкретных выводов. Отделу контроля качества на предприятии следует обратить внимание на определенные моменты для уменьшения количества дефектов готовой продукции:

- качество закупаемой пряжи для ткани Бязь 262, пряжа явно не подходит для создания ткани 1го сорта;
- наличие ткацкого брака Слеты по ткани Полина шириной 80, говорит о плохой настройке ткацких станков на ткацкой фабрике №1;
- брак отделки Помарки по ткани Бязь 262 набивная шириной 150 свидетельствует о недостаточном обслуживании печатных машин по ширине 150, плохая чистка или неправильная настройка;
-брак отделки Помарки по ткани Перкаль гладкокрашеной 223 говорит о плохом состоянии красильных машин по ширине 220 в красильном цехе, возможно несоблюдение технологических параметров или плохая подготовка красителя;
- межоперационный брак Швы по ткани Перкаль 223 свидетельствует о том, что на каком-то этапе технологического процесса был сбой оборудования и часть ткани пришлось рвать и сшивать в дальнейшем. Этот кластер по количеству элементов самый маленький, но все равно влияет на качество продукции.

ЗАКЛЮЧЕНИЕ

В условиях цифровизации промышленности традиционные методы сбора и обработки информации о дефектах, основанные на бумажных журналах и стандартной отчетности, перестают отвечать требованиям оперативного и предиктивного управления. Они позволяют фиксировать факт брака, но не раскрывают скрытые многомерные взаимосвязи между технологическими параметрами, типом сырья и возникающими дефектами.

В ходе работы авторами был выполнен полный цикл аналитического исследования: от функционального моделирования бизнес-процессов и аудита данных до применения алгоритмов машинного обучения и интерпретации полученных результатов. Главным научно-практическим итогом стала разработка и апробация методики сегментации базы дефектов, специально адаптированной для работы со смешанными данным, специально адаптированной для работы со смешанными данными (категориальными и числовыми), характерными для реальных производственных условий.

Проведенное исследование показало, что внедрение в процессы управления качеством методики сегментации базы дефектов на основе алгоритма k-prototypes позволит трансформировать процесс контроля качества из пассивного инструмента учета брака в активный инструмент поиска первопричин отклонений. Сформулированные на основе выделенных кластеров рекомендации для отдела контроля качества позволяют перейти от тотального поиска проблем к точечным и экономически обоснованным корректирующим действиям.

Таким образом, внедрение методов интеллектуального анализа данных, и в частности алгоритма k-prototypes, в систему менеджмента качества текстильного предприятия способствует повышению прозрачности производственных процессов, снижению доли брака и создает предпосылки для построения предиктивных моделей, позволяющих прогнозировать возникновение дефектов на основе технологических параметров ещё до этапа финишного контроля.

Авторы заявляют об отсутствии конфликта интересов.
The authors declare no conflict of interest.

ЛИТЕРАТУРА

1. Цифровизация промышленности. Обзор. https://www.tadviser.ru/index.php/Статья:Обзор_Цифровизация_ промышленности_2024.
2. «Газпромнефть-Снабжение» и «Инлайн Груп» тестируют нейросети для контроля качества деталей и оборудования. https://supply.gazprom-neft.ru/press-center/news/gazpromneft-snabzhenie-i-inlayn-grup-testiruyut-neyroseti-dlya-kontrolya-kachestva-detaley-i-oborudo.
3. РТ-Техприемка будет контролировать качество стали для авиации с помощью искусственного интеллекта. https://rttec.ru/media/news/256/.
4. ПМЭФ: цифровизация СИБУРа и искусственный интеллект. https://sibur.digital/202-pmef-tsifrovizatsiya-sibura-i-iskusstvennyy-intellekt.
5. «Северсталь» внедряет цифровую систему управления качеством. https://severstal.com/rus/media/archive/severstal-vnedryaet-tsifrovuyu-sistemu-upravleniya-kachestvom/.
6. **Онопюк Е.Ю., Шахова И.Ю.** Цифровизация в текстильной промышленности: новые возможности и вызовы. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2025. № 2(64). С. 38–43. DOI: 10.6060/ivecofin.2025642.720.
7. **Ксенофонтова О.Л., Смирнова Н.В., Котова А.В.** Применение методов интеллектуального анализа данных в эпидемиологии. *Современные наукоемкие технологии. Региональное приложение*. 2023. № 2(74). С. 88–93. DOI: 10.6060/snt.20237402.0009.
8. **Луценко Е.В., Коржаков В.Е.** Некоторые проблемы классического кластерного анализа. Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. 2011. № 2. С. 91–102.
9. **Орехов А.В.** Марковский момент остановки агломеративного процесса кластеризации в евклидовом пространстве. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*. 2019. Т. 15. № 1. С. 76–92. DOI: 10.21638/11702/spbu10.2019.106.

REFERENCES

1. Digitalization of industry. Overview. https://www.tadviser.ru/index.php/Статья:Обзор_Цифровизация_промышленности_2024. (in Russian).
2. Gazpromneft-Snabzhenie and Inline Group test neural networks for quality control of parts and equipment. https://supply.gazprom-neft.ru/press-center/news/gazpromneft-snabzhenie-i-inlayn-grup-testiruyut-neyroseti-dlya-kontrolya-kachestva-detaley-i-oborudo. (in Russian).
3. RT-Tekhpriyomka will control steel quality for aviation using artificial intelligence. https://rttec.ru/media/news/256/. (in Russian).
4. SPIEF: digitalization of SIBUR and artificial intelligence. https://sibur.digital/202-pmef-tsifrovizatsiya-sibura-i-iskusstvennyy-intellekt. (in Russian).
5. Severstal introduces a digital quality management system. https://severstal.com/rus/media/archive/severstal-vnedryaet-tsifrovuyu-sistemu-upravleniya-kachestvom/. (in Russian).
6. **Onopyuk E.Yu., Shakhova I.Yu.** Digitalization in the textile industry: new opportunities and challenges. *Ivecofin*. 2025. N 2(64). P. 38–43. DOI: 10.6060/ivecofin.2025642.720. (in Russian).
7. **Ksenofontova O.L., Smirnova N.V., Kotova A.V.** Application of data mining methods in epidemiology. *Modern High Technologies. Regional Application*. 2023. N 2(74). P. 88–93. DOI: 10.6060/snt.20237402.0009. (in Russian).
8. **Lutsenko E.V., Korzhakov V.E.** Some problems of classical cluster analysis. *Bulletin of the Adyghe State University. Series 4: Natural-Mathematical and Technical Sciences*. 2011. N 2. P. 91–102. (in Russian).
9. **Orekhov A.V.** Markov stopping moment of the agglomerative clustering process in Euclidean space. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*. 2019. Vol. 15. N 1. P. 76–92. DOI: 10.21638/11702/spbu10.2019.106. (in Russian).
10. **Hartigan J.A., Wong M.A.** Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*. 1979. Vol. 28. N 1. P. 100–108.

10. **Hartigan J.A., Wong M.A.** Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*. 1979. Vol. 28. N 1. P. 100–108.
11. **Краковецкий А**. Кластеризация: алгоритмы k-means и c-means. https://habr.com/ru/articles/67078/.
12. Aprilliant A. The k-prototype as Clustering Algorithm for Mixed Data Type (Categorical and Numerical). https://towardsdatascience.com/the-k-prototype-as-clustering-algorithm-for-mixed-data-type-categorical-and-numerical-fe7c50538ebb/.
13. **Мошкин Л.С.** Использование аппроксимационно-оценочных критериев для модификации алгоритма HDBSCAN. *Процессы управления и устойчивость*. 2025. Т. 12. № 1. С. 352–359.
14. **Шаталова О.М., Касаткина Е.В., Лившиц В.Н.** Кластерный анализ и классификация промышленно ориентированных регионов РФ по экономической специализации. *Экономика и математические методы*. 2022. Т. 58. № 1. С. 80–91. DOI: 10.31857/S042473880018971-7.
15. **Прохоренков П.А., Регер Т.В., Гудкова Н.В.** Методы кластерного анализа в региональных исследованиях. *Фундаментальные исследования*. 2022. № 3. С. 100–106.
16. **Сивоголовко Е.В.** Методы оценки качества четкой кластеризации. *Компьютерные инструменты в образовании*. 2011. № 4. С. 14–31.
17. Kaufman L., Rousseeuw P.J. Finding Groups in Data. An Introduction to Cluster Analysis. Hoboken: Wiley. 2005. 342 p.
18. **Астраханцева И.А., Герасимов А.С., Астраханцев Р.Г.** Прогнозирование региональной инфляции с помощью алгоритмов машинного обучения. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2022. № 4(54). С. 6-13. DOI: 10.6060/ivecofin.2022544.620.
19. **Астраханцева И.А., Котенев Т.Е., Горев С.В., Астраханцев Р.Г., Грименицкий П.Н.** Интеграция методов линейной регрессии и случайного леса в системы интеллектуальной поддержки управления криогенными транспортными системами. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2024. № 03(61). С.81-90. DOI: 10.6060/ivecofin.2024613.692. EDN RGESSD
20. **Астраханцева И.А., Кутузова А.С., Астраханцев Р.Г.** Интеллектуальные методы обработки данных при прогнозировании оборота наличных денежных средств в банкоматах коммерческих банков. *Научные труды Вольного экономического общества России*. 2019. Т. 218. № 4. С. 481-488.
21. **Astrakhantsev R., Chuhno A., Dmukh A., Kurochkin A., Astrakhantseva I.** Differences with high probability and impossible differentials for the KB-256 cipher. *Journal of Computer Virology and Hacking Techniques*. 2024. Vol. 20. N 3. P. 525-531. DOI: 10.1007/s11416-024-00532-2.

11. Krakovetsky A. Clustering: k-means and c-means algorithms. https://habr.com/ru/articles/67078/. (in Russian).
12. Aprilliant A. The k-prototype as Clustering Algorithm for Mixed Data Type (Categorical and Numerical). https://towardsdatascience.com/the-k-prototype-as-clustering-algorithm-for-mixed-data-type-categorical-and-numerical-fe7c50538ebb/.
13. **Moshkin L.S.** Using approximation-evaluation criteria for modifying the HDBSCAN algorithm. *Control Processes and Stability*. 2025. Vol. 12. N 1. P. 352–359. (in Russian).
14. **Shatalova O.M., Kasatkina E.V., Livchits V.N.** Cluster analysis and classification of Russia's industrially oriented regions by economic specialization. *Economics and Mathematical Methods*. 2022. Vol. 58. N 1. P. 80–91. DOI: 10.31857/S042473880018971-7. (in Russian).
15. **Prokhorenkov P.A., Reger T.V., Gudkova N.V.** Cluster analysis methods in regional studies. *Fundamental Research*. 2022. N 3. P. 100–106. (in Russian).
16. **Sivogolovko E.V.** Methods for assessing the quality of hard clustering. *Computer Tools in Education*. 2011. N 4. P. 14–31. (in Russian).
17. Kaufman L., Rousseeuw P.J. Finding Groups in Data. An Introduction to Cluster Analysis. Hoboken: Wiley. 2005. 342 p.
18. **Astrakhantseva I.A., Gerasimov A.S., Astrakhantsev R.G.** Forecasting regional inflation using machine learning algorithms. *Ivecofin*. 2022. N 4(54). P. 6–13. DOI: 10.6060/ivecofin.2022544.620. (in Russian).
19. **Astrakhantseva I.A., Kotenev T.E., Gorev S.V., Astrakhantsev R.G., Grimenitsky P.N.** Integration of linear regression and random forest methods into intelligent support systems for managing cryogenic transport systems. *Ivekofin*. 2024. N 03(61). P. 81-90. DOI: 10.6060/ivecofin.2024613.692. EDN RGESSD. (in Russian).
20. **Astrakhantseva I.A., Kutuzova A.S., Astrakhantsev R.G.** Artificial intelligent methods in commercial banks ATM cash turnover forecast. *Scientific Works of the Free Economic Society of Russia*. 2019. Vol. 218. N 4. P. 481–488. (in Russian).
21. **Astrakhantsev R., Chuhno A., Dmukh A., Kurochkin A., Astrakhantseva I.** Differences with high probability and impossible differentials for the KB-256 cipher. *Journal of Computer Virology and Hacking Techniques*. 2024. Vol. 20. N 3. P. 525-531. DOI: 10.1007/s11416-024-00532-2.

Поступила в редакцию 15.01.2026
Принята к опубликованию 29.01.2026

Received 15.01.2026
Accepted 29.01.2026