

ИНФОРМАТИКА

*Вестник Сыктывкарского университета.  
Серия 1: Математика. Механика. Информатика.  
Выпуск 2 (35). 2020*

УДК 004

**РАЗРАБОТКА СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ  
РЕШЕНИЙ НА БАЗЕ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ  
LOGINOM ДЛЯ ЭФФЕКТИВНОГО УПРАВЛЕНИЯ  
МАРКЕТИНГОВЫМИ КАМПАНИЯМИ**

*Д. А. Невекин, Н. Ю. Прокопенко*

В статье описывается разработка программного продукта для автоматизации процесса проверки корректности ссылок на электронные библиотечные системы в рабочих программах учебных дисциплин.

*Ключевые слова:* рабочая программа учебной дисциплины, электронные библиотечные системы, автоматизация, алгоритм.

**Введение**

Одним из главных факторов роста любой торговой компании является повышение эффективности управления внутренними ресурсами. Одним из таких ресурсов считается клиентская база организации. Потенциал клиентской базы в большей степени раскрывается благодаря применению клиентской аналитики, решению задачи сегментации клиентов и последующему осуществлению маркетинговых коммуникаций. Необходимыми условиями эффективного проведения анализа клиентской базы являются высокие стандарты качества данных, чёткая постановка задачи, а также успешно подобранные алгоритмы и модели для решения задачи. Всем этим требованиям удовлетворяют современные системы поддержки принятия решений (СППР), где используются разные методы анализа клиентских баз данных: информационный поиск, интеллектуальный анализ данных, поиск знаний в базах данных, рассуждение на основе прецедентов, имитационное моделирование, эволюционные вычисления и генетические алгоритмы, нейронные сети, ситуационный анализ, когнитивное моделирование и др.

Принципы построения рекомендательных систем поддержки решений на основе технологий хранилищ данных (Data Warehousing), оперативного анализа (OLAP) и машинных методов извлечения знаний (Data Mining) для аналитической поддержки процессов принятия решений рассматриваются многими авторами, так, в работе [1] рассматриваются проблемы и методы решения сбора и интеграции и расширенной аналитики больших данных, об успешном применении методов машинного обучения пишут авторы в работе [2], с 2009 г. по настоящее время издается журнал Decision Support Systems [3].

Опыт российских и зарубежных торговых компаний показывает, что применение маркетинговых коммуникаций в бизнесе позволяет в значительной степени повысить лояльность клиентов, а также привлечь новых клиентов, вследствие чего увеличить объёмы продаж. Поэтому в данной работе акцент смещён на решение конкретной бизнес-задачи — сегментации клиентской базы.

В настоящее время каждая торговая компания, как российская, так и зарубежная, понимает, насколько важно основательно и комплексно подходить к повышению качества клиентской базы и повышению эффективности маркетинговых коммуникаций. Поэтому использование существующих и разработка новых соответствующих инструментов, систем и методов является актуальным направлением как с научной, так и практической точки зрения.

В данной работе рассматриваются методы сегментации клиентских баз данных с использованием возможности аналитической платформы Loginom.

### **1. Обзор современного программного обеспечения для поддержки принятия решений в бизнес-аналитике**

Современные системы управления технологическими процессами требуют обработки и анализа больших объёмов информации. Данная проблема давно стала критической в областях, непосредственно связанных с аналитической обработкой данных (Data Mining, искусственный интеллект, системы поддержки принятия решений, техническое зрение и другие).

Но даже самые мощные технологии извлечения закономерностей и машинного обучения, такие как Knowledge discovery in databases (KDD) и Data mining, не представляют особой ценности без инструментальной поддержки в виде соответствующего программного обеспечения.

Рынок программного обеспечения KDD и Data mining делится на несколько сегментов [4; 5]:

— статистические пакеты с возможностями Data mining;

- настольные Data mining пакеты;
- СУБД с набором алгоритмов Data mining;
- аналитические платформы.

Статистические пакеты с возможностями Data mining и настольные Data mining-пакеты ориентированы в основном на профессиональных пользователей. Их отличительными особенностями можно назвать следующие:

- слабая интеграция с промышленными источниками данных;
- бедные средства очистки, предобработки и трансформации данных;
- отсутствие гибких возможностей консолидации информации, например в специализированном хранилище данных;
- поточная (конвейерная) обработка новых данных затруднительна или реализуется встроенными языками программирования и требует высокой квалификации;
- из-за использования пакетов на локальных рабочих станциях обработка больших объемов данных затруднена.

Настольные Data mining-пакеты могут быть ориентированы на решение как всех классов задач Data mining, так и только какого-либо одного, например кластеризация или классификация. Вместе с тем эти пакеты предоставляют богатые возможности в плане алгоритмов Data mining, что достаточно для решения каких-либо исследовательских задач. Однако создание эффективных прикладных решений промышленного уровня с помощью таких пакетов затруднено. Поэтому в бизнесе, как правило, используются СУБД с элементами Data mining и аналитические платформы.

Важной отличительной особенностью аналитических платформ от СУБД с набором алгоритмов Data Mining является наличие развитых графических инструментов для визуального моделирования. Такая среда моделирования позволяет аналитику создавать законченные аналитические решения от извлечения данных.

Тенденции последних лет в развитии аналитических систем заключаются в интеграции средств аналитической обработки, алгоритмов извлечения знаний, управления метаданными и визуализации результатов на одной программной аналитической платформе. Эти аналитические системы созданы, чтобы помочь в планировании и управлении различными бизнес-процессами и в решении некоторых бизнес-задач. К подобным информационно-аналитическим инструментам можно отнести аналитическую платформу Loginom (разработка компании «ООО Аналитические технологии Loginom Company»), реализующую методы

продвинутую аналитику с использованием подхода «low-code» (минимум кода), визуального проектирования и объектно-ориентированного подхода.

## **2. Библиотека компонентов базовой сегментации**

Сегментация рынка заключается в разделении рынков на четкие группы покупателей (рыночные сегменты), которые могут требовать разные продукты и к которым необходимо прилагать разные маркетинговые усилия.

Сегмент — это группа потребителей, характеризующаяся однотипной реакцией на предлагаемый продукт и на набор маркетинговых стимулов.

Нужно отметить, что в мировой практике используются два принципиальных подхода к маркетинговому сегментированию.

1. В рамках первого метода, именуемого «a priori», предварительно известны признаки сегментирования, численность сегментов, их количество, характеристики, карта интересов. То есть подразумевается, что сегментные группы в данном методе уже сформированы. Метод «a priori» используют в тех случаях, когда сегментирование не является частью текущего исследования, а служит вспомогательным базисом при решении других маркетинговых задач. Иногда этот метод применяют при очень четкой определенности сегментов рынка, когда вариантность сегментов рынка невысока. «A priori» допустим и при формировании нового продукта, ориентированного на известный сегмент рынка.

2. В рамках второго метода, именуемого «post hoc (cluster based)», подразумевается неопределенность признаков сегментирования и сущности самих сегментов. Исследователь предварительно выбирает ряд интерактивных по отношению к респонденту (метод подразумевает проведение опроса) переменных, и далее в зависимости от высказанного отношения к определенной группе переменных респонденты относятся к соответствующему сегменту. При этом карта интересов, выявленная в процессе последующего анализа, рассматривается как вторичная. Этот метод применяют при сегментировании потребительских рынков, сегментная структура которых не определена в отношении продаваемого продукта.

В разработанной системе поддержки принятия решений методы сегментации реализованы в виде библиотеки компонентов с помощью аналитической платформы Loginom (рис. 1).

Правильное проектирование компонента — нетривиальная задача. необходимо разобраться в предметной области тех методов и алгоритмов, которые закладываются в компоненты и их семантику, нужно

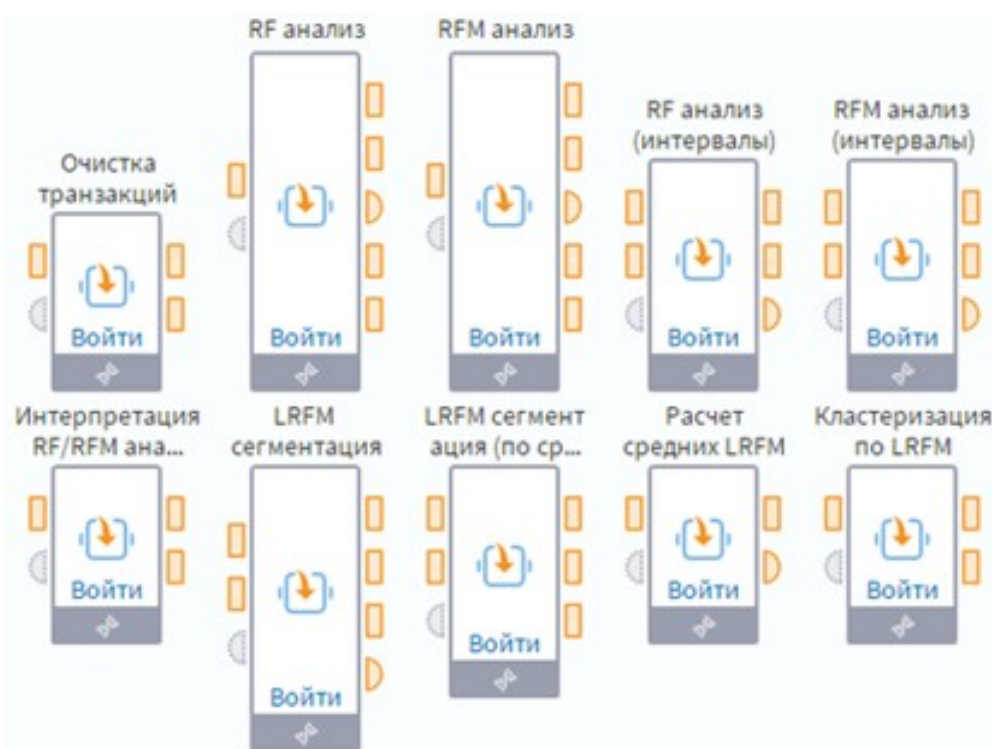


Рис. 1. Разработанные компоненты библиотеки базовой сегментации

исследовать кейсы использования компонентов и сформировать требования к наборам данных, гибкость и вариативность в применении компонентов закладывается через параметризацию, архитектура компонентов должна исключать дублирование логики обработки, вместо них применяется наследование или ссылки на фрагменты обработки, и нужно корректно реализовать алгоритм в Loginom.

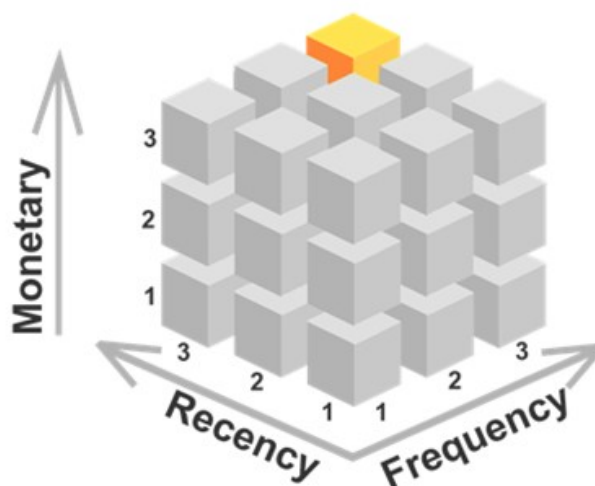
### 3. Описание методов сегментации

Одним из наиболее известных методов базовой сегментации является RFM-анализ. Он широко применяется в клиентской аналитике. Это техника сегментации клиентов, опирающаяся на их поведение.

Основными являются показатели R (recency) и F (frequency) — давность и частота. Показатель M (monetary) — денежная стоимость клиента, обычно выражающаяся как сумма всех покупок клиента или среднее значение суммы покупок. Важность показателей ранжируется согласно последовательности: Давность, Частота, Сумма продаж. Стоит отметить, что показатель Monetary необязательно привязывать к деньгам. Это может быть любой критерий, наложенный на основные показатели. Монетизацию не всегда удастся свести к покупкам. Например, для

какого-нибудь информационного сайта можно взять время просмотра страницы пользователем.

Суть метода состоит в том, чтобы разделить клиентскую базу на  $N$  количество сегментов (рис. 2). В каждом отдельно взятом сегменте клиенты должны иметь схожие характеристики и отличаться от клиентов, принадлежащих другому сегменту. Сами сегменты должны быть примерно равные по объему входящих в них клиентов. Количество сегментов формируется путем разбиения показателей  $R$ ,  $F$  и  $M$  на любое количество групп (рекомендуется не разбивать показатель не более чем на 5 групп). Например, если мы разделим каждый показатель  $R$ ,  $F$  и  $M$  на 5 групп, то в итоге получим 125 конечных сегментов ( $5 \times 5 \times 5$ ).



**Рис. 2.** Графическое представление параметров  $R$ ,  $F$  и  $M$

RFM-метод широко применяется:

- для оценки привлекательности клиента;
- контроля изменения в поведении клиента;
- оценки миграции между группами (сегментами);
- оценки эффективности программы лояльности;
- выбора стратегии работы с каждым клиентом.

К инструментам сегментирования относят многочисленные модификации RFM-модели (например, RFD — Recency, Frequency, Duration; RFMTC — Recency, Frequency, Monetary Value, Time, Churn; LRFM — Length, Recency, Frequency, Monetary; и другие), а также деревья решений, карты Кохонена, нейронные сети, кластерный анализ.

Так, например, Чанг и Цай [6] предложили использовать LRFM-модель (LRFM — аббревиатура слов Length — длина, Recency — но-

визна, Frequency — частота, Monetary — вложения). Одна из причин заключается в том, что у RFM-модели есть проблема, связанная со сложностью различать клиентов, имеющих долгосрочные отношения, и клиентов, имеющих краткосрочные отношения с компанией, а лояльность и прибыльность клиента в значительной степени зависит от качества и продолжительности отношений между ним и компанией. В связи с этим для выявления наиболее лояльных клиентов необходимо учитывать эту продолжительность отношений с клиентом, то есть L, при этом L определяется как количество периодов времени (например, дней) от первой до последней покупки, записанных в базе данных.

Когда идентифицированы различные комбинации LRFM, клиенты могут быть классифицированы по соответствующим группам, таким как основные клиенты, потенциальные клиенты, потерянные клиенты, новые клиенты и потребители, потребляющие ресурсы. Важно, что различные маркетинговые стратегии могут быть разработаны для удовлетворения различных потребностей клиентов.

Основная особенность Data Mining — это сочетание широкого математического инструментария (от классического статистического анализа до новых методов машинного обучения) и последних достижений в сфере информационных технологий. В технологии Data Mining объединились строго формализованные методы и методы неформального анализа, то есть количественный и качественный анализ данных. Методы интеллектуального анализа данных, в частности деревья решений и кластерный анализ, могут применяться для сегментирования клиентских баз [7].

Деревья решений удобны для использования при описании портрета покупателя. Этот инструмент предиктивной аналитики достаточно хорошо изучен и применяется в различных прикладных областях, в том числе и в клиентской аналитике. В настоящее время сфера применения деревьев решений широка, и задачи, решаемые с помощью этого аппарата, можно выделить в следующие группы: описание данных, классификация, регрессия. Но на практике использование деревьев решений зачастую бывает проблематичным из-за сложности получения оптимального дерева решений в условиях, когда объёмы исчисляются миллионами записей в базах данных.

Кластерный анализ — способ классификации многомерных параметров, при котором образуются так называемые сегменты (кластеры, таксоны), содержащие в себе схожие между собой объекты. Этот метод сегментации не требует априорных предположений о параметрах анализа, а значит, им можно анализировать любые объекты по любым показате-

лям. Кластерный анализ применяется для решения широкого спектра задач, но чаще всего речь идет именно о задаче сегментации. Есть много исследований, посвященных проблеме применения различных методов кластеризации клиентских СУБД. Например, в статье [8] на основе анализа предпочтений предлагается новая метрика и описывается алгоритм иерархической кластеризации для выполнения сегментации рынка, в статье [9] описаны результаты применения иерархических и k-средних методов кластеризации для сегментирования клиентов торговых центров в Индии.

Сам принцип кластерного анализа предполагает построение некоторого метрического пространства, где каждый объект удален на определенное расстояние от остальных. Матрица расстояний и матрица сходства и дает основания для выделения кластеров. Каждый кластер имеет следующие характеристики: центр кластера, радиус кластера, средне-квадратическое отклонение, размер кластера.

Компонент «Кластеризация по LRFM» (рис. 3) производит кластеризацию объектов на основе алгоритмов k-means и g-means [5]. Основное отличие одного алгоритма от другого в том, известно ли заранее количество кластеров. Если количество кластеров известно, то применяется алгоритм k-means, в противном случае — g-means, который определит это количество автоматически в рамках заданного интервала.

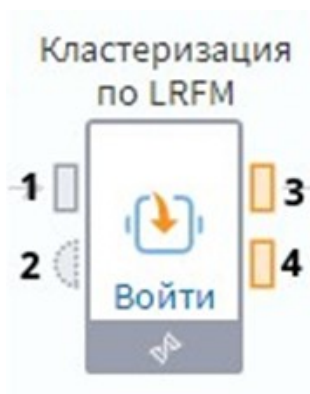


Рис. 3. Компонент «Кластеризация по LRFM»

Вход «Входной набор данных» (порт 1) включает следующие параметры:

- 1) ID клиента (*ClientID*);
- 2) Длина отношений (*Lenght*);
- 3) Дата последней покупки (*Recency*);



- 4) Количество (*Frequency*) — количество покупок, визитов;
- 5) Сумма (*Monetary*).

Вход «Переменные для кластеризации» (порт 2) включает:

- 1) Автоопределение числа кластеров (*AutoSeg*) — переменная логического типа, которая отвечает за то, будет ли количество кластеров задано переменной *CountSeg* или будет рассчитано автоматически с помощью алгоритмов системы;
- 2) Минимальное количество кластеров (*MinCount*) — по умолчанию равно 1;
- 3) Максимальное количество кластеров (*MaxCount*) — по умолчанию равно 10;
- 4) Порог разделения кластеров (*SeparationSeg*) — значение порога разделения кластеров (должно находиться в интервале от 0,1 до 5), причём чем больше порог разделения, тем больше кластеров будет сгенерировано при кластеризации (если автоопределение — true);
- 5) Количество кластеров (нет автоопределения) (*CountSeg*) — ввести нужное количество кластеров. Должно быть больше 2, по умолчанию значение "4".

Выход «Разбиение на кластеры» (порт 3) — таблица, состоящая из полей:

- Номер кластера — каждому объекту присвоен номер того кластера, в который он входит;
- Расстояние до центра кластера — положение объекта относительно центра кластера;
- Остальные поля — это поля исходного набора данных (значения не изменяются).

Выход «Центры кластеров» (порт 4) содержит следующие характеристики:

- Центр кластера — среднее значение переменных объектов, входящих в кластер. Результат — таблица, количество записей которой соответствует числу кластеров, то есть данные сгруппированы по кластерам. Состоит из полей:

- Номер кластера — перечислены номера сформированных кластеров;
- Поля исходного набора данных, в ячейках которых рассчитано среднее значение параметров.

### **Заключение**

Компоненты базовой сегментации клиентской базы, реализованные на базе аналитической платформы Logipom, были протестированы на реальных данных сети магазинов формата DIY (база данных включала

600000 записей о клиентах). При тестировании библиотеки сбоев и ошибок в работе компонентов обнаружено не было. Клиентская база была разделена на пять конечных групп: лучшие, перспективные, норма, новые, потерянные.

Разработанная система поддержки принятия решений в виде библиотеки компонентов на базе АП Loginom позволяет повысить эффективность маркетинговых кампаний за счет точного сегментирования клиентских баз и вследствие этого сокращения затрат на незаинтересованных клиентов компании. Созданная библиотека компонентов методов базовой сегментации поможет аналитикам, использующим систему, адаптировать, совершенствовать и эффективно проводить маркетинговые кампании.

## Список литературы

1. **Kitchens B., Dobolyi D., Li J., Abbasi A.** Advanced customer analytics: Strategic value through integration of relationship-oriented big data // *Taylor & Francis, Journal of Management*. 2018. Vol. 35. Issue 2. Pp. 540–574.
2. **Mbugua A. W., Omondi A. O.** An Application of association rule learning in recommender systems for e-Commerce and its effect on marketing // *IEEE Pan African Conference on Science, Computing, and Telecommunications*. 2017.
3. Decision support systems / Elsevier Science Publishing Company, Inc. ISSN: 0167-9236 eISSN: 1873-579.
4. **Кацко И. А., Паклин Н. Б.** Практикум по анализу данных на компьютере. М.: КолосС, 2009. 278 с.
5. **Паклин Н. Б., Орешков В. И.** Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2013. 624 с.
6. Group RFM analysis as a novel framework to discover better customer consumption behavior / Hui-Chu Chang, Hsiao-Ping Tsai. Taiwan: Elsevier, 2011. 15 p.
7. **Барсегян А. А., Куприянов М. С., Елизаров С. И.** Анализ данных и процессов : учеб. пособие. 3-е изд., перераб и доп. СПб.: БХВ-Петербург, 2009. 512 с.

8. **Liu J., Liao X., Huang W., Liao X.** Market segmentation: A multiple criteria approach combining preference analysis and segmentation decision. Elsevier: Omega, Volume 83, 2019. Pp. 1–13.
9. **Chandan Parsad, Sai Vijay T., Harvinder Singh, Sanjeev Prashar.** Segmenting Indian shoppers on mall attractiveness factors // *International Journal of Services Technology and Management. Vol. 25. 2019.*

### Summary

**Nevekin D. A., Prokopenko N. U.** Development of a decision support system on the basis of analytical platform Loginom for effective management of marketing and customer analytics

The article shows the need to use specialized software and modern intelligent methods in purpose to solve client database segmentation problems. Data mining techniques, particularly cluster analysis, enable to divide all customers into appropriate number of segmentation in accordance with their similarities. Market segmentation is very useful for identification of customers' different needs.

*Keywords: decision support system; analytical platform Loginom, customer analytics, segmentation decision, intellectual methods of data analysis, Data Mining.*

### References

1. **Kitchens B., Dobolyi D., Li J., Abbasi A.** Advanced customer analytics: Strategic value through integration of relationship-oriented big data, *Taylor & Francis, Journal of Managemen*, 2018, Vol. 35, Issue 2, pp. 540–574.
2. **Mbugua A. W., Omondi A. O.** An Application of association rule learning in recommender systems for e-Commerce and its effect on marketing, *IEEE Pan African Conference on Science, Computing, and Telecommunications*, 2017.
3. *Decision support systems*, Elsevier Science Publishing Company, Inc. ISSN: 0167-9236eISSN: 1873-579.
4. **Katsko I. A., Paklin N. B.** *Praktikum po analizu dannykh na komp'yutere* (Workshop on data analysis on a computer), M.: KolosS, 2009, 278 p.

5. **Paklin N. B., Oreshkov V. I.** *Biznes-analitika: ot dannykh k znaniyam* (Business analytics: from data to knowledge), St. Petersburg: Peter, 2013, 624 p.
6. *Group RFM analysis as a novel framework to discover better customer consumption behavior*, Hui-Chu Chang, Hsiao-Ping Tsai, Taiwan: Elsevier, 2011, 15 p.
7. **Barseghyan A. A., Kupriyanov M. S., Elizarov S. I.** *Analiz dannykh i protsessov: ucheb. posobiye* (Analysis of data and processes: textbook. allowance), 3rd ed., Revised and add, SPb.: BHV-Petersburg, 2009, 512 p.
8. **Liu J., Liao X., Huang W., Liao X.** Market segmentation: A multiple criteria approach combining preference analysis and segmentation decision, Elsevier: Omega, Volume 83, 2019, pp. 1–13.
9. **Chandan Parsad, Sai Vijay T., Harvinder Singh, Sanjeev Prashar.** Segmenting Indian shoppers on mall attractiveness factors, *International Journal of Services Technology and Management*, Vol. 25, 2019.

**Для цитирования:** Невекин Д. А., Прокопенко Н. Ю. Разработка системы поддержки принятия решений на базе аналитической платформы Loginom для эффективного управления маркетинговыми кампаниями // *Вестник Сыктывкарского университета. Сер. 1: Математика. Механика. Информатика. 2020. Вып. 2 (35). С. 37–48.*

**For citation:** Nevekin D. A., Prokopenko N. U. Development of a decision support system on the basis of analytical platform Loginom for effective management of marketing and customer analytics, *Bulletin of Syktyvkar University. Series 1: Mathematics. Mechanics. Informatics*, 2020, 2 (35), pp. 37–48.

Нижегородский государственный  
архитектурно-строительный университет

Поступила 14.03.2020