Автоматизация CRM-сегментации с использованием ML-прогнозирования и LLM-генерации контента

Макарчук Татьяна Анатольевна
к.пед.н., доцент, доцент кафедры информатики, Санкт-Петербургский государственный экономический университет, tmakarchuk@mail.ru,

Михайлов Артем Новрузович
магистр по направлению подготовки «Прикладная информатика», руководитель группы стратегии, ООО «КОКОС GROUP», borsic@inbox.ru,

Ставер Егор Юрьевич
магистр по направлению подготовки «Прикладная информатика», аналитик, ООО «Вейм», lysenok.1998@mail.ru

В статье рассматривается подход к автоматизации сегментации клиентских баз для CRM-маркетинга на основе интеллектуальной веб-платформы. Актуальность исследования обусловлена ростом объемов клиентских данных и необходимостью перехода от ручных, экспертно-зависимых методов формирования сегментов к управляемым аналитическим процессам. Цель работы – разработка и апробация системы интеллектуальной сегментации, объединяющей модули загрузки и подготовки данных, модуль сегментации клиентской базы и модуль машинного обучения, модуль генерации контента с использованием искусственного интеллекта. В статье представлены архитектурные решения, обоснован выбор LLM-модели, описана реализация ключевых модулей платформы, а также приведены результаты апробации на данных компании, оказывающей маркетинговые услуги. Показано, что внедрение предложенного решения позволяет сократить время подготовки кампаний до 75%, увеличить количество тестируемых гипотез в 3–4 раза. Результаты подтверждают практическую значимость интеллектуальной сегментации для повышения эффективности маркетинговых коммуникаций.

Ключевые слова: CRM-маркетинг, сегментация клиентов, машинное обучение, предиктивная аналитика, генеративный искусственный интеллект, CDP-платформа, автоматизация маркетинга.

Введение

В условиях цифровой трансформации компании переходят от массовых обезличенных рассылок к персонализированным кампаниям на основе поведенческой аналитики клиентов [1]. Сегментация клиентской базы выступает ключевым инструментом этого перехода, позволяя выделять однородные группы потребителей для дифференцированных коммуникационных стратегий [2]. Несмотря на наличие современных CRM/CDP-платформ, до 60–80% рабочего времени специалистов в сфере CRM-маркетинга по-прежнему уходит на ручной отбор аудиторий и подготовку сегментов [3].

Актуальность исследования определяется тремя основными факторами. Во-первых, экспоненциальный рост объемов клиентских данных требует автоматизации процессов их обработки и анализа. Во-вторых, уход с российского рынка крупных зарубежных MarTech-платформ после 2022 года формирует потребность в отечественных решениях, обеспечивающих технологический суверенитет в области управления клиентскими данными. В-третьих, существующие инструменты не предоставляют единого контура, объединяющего функциональность CDP, ML-прогнозирование и генерацию контента, что вынуждает компании использовать разрозненные решения и увеличивает риски ошибок.

Проблемой исследования выступает высокая доля ручного труда при подготовке сегментов для российских CRM-кампаний, отсутствие прозрачного прогноза ключевых метрик (OR, CTR, CR) и ограниченные возможности A/B-тестирования. Имеющиеся в мировой практике платформы в области CRM-сегментации (Salesforce Einstein, HubSpot), автоматизирующие процессы подготовки сегмента, часто не доступны на российском рынке. В существующей практике подготовка сегмента для одной кампании может занимать до нескольких рабочих дней и включает выгрузку данных из CDP, очистку в Excel/SQL, субъективный выбор критериев и ручную оценку качества. Существующие исследования в области CRM-сегментации сосредоточены либо на применении отдельных методов машинного обучения [3, 6, 7, 8, 10], либо на использовании CDP-платформ для управления клиентскими данными [2, 4, 5].

В последние годы активно развивается направление применения больших языковых моделей (LLM) в маркетинге, в частности, для автоматизации создания контента [9]. При этом вопросы интеграции предиктивного моделирования, автоматизированного контроля качества данных и генерации контента в рамках единой веб-платформы в условиях импортозамещения остаются недостаточно проработанными, что обуславливает актуальность настоящего исследования.

Целью исследования является повышение эффективности маркетинговых коммуникаций и снижение трудоемкости подготовки рассылок за счет разработки платформы интеллектуальной сегментации аудитории, обеспечивающей автоматизацию отбора аудиторий, ML-прогнозирование метрик и генерацию контента.

Научная новизна исследования заключается в разработке архитектуры и программной реализации цифровой российской платформы, которая впервые в условиях импортозамещения объединяет автоматический контроль качества клиентских данных с формированием DQ-отчетов, прогнозирование временного интервала до следующей покупки на основе ансамблевых ML-моделей, генерацию персонализированного контента с использованием LLM Qwen3 14B.

Практическая значимость. Предложенная платформа может быть внедрена в компаниях, предоставляющим маркетинговые услуги, для автоматизации CRM-сегментации. Ожидаемый экономический эффект оценивается в сокращении времени подготовки сегментов до 75%, росту числа кампаний в 3–4 раза и повышению точности таргетинга. Предложенная платформа ориентирована на российский рынок MarTech-решений в условиях импортозамещения.

Эволюция подходов к сегментации клиентских баз

Сегментация клиентской базы традиционно реализуется на основе демографических, поведенческих и психографических признаков [4]. Классическим инструментом выступает RFM-анализ [5], группирующий клиентов по давности (Recency), частоте (Frequency) и денежному объему (Monetary) покупок. В цифровой экономике такие методы имеют существенные ограничения: они опираются на ограниченный набор признаков, статичны и не учитывают динамику поведения клиента.

С развитием машинного обучения возникли более сложные подходы. Обучение без учителя, в частности алгоритмы кластеризации (k-means, DBSCAN, иерархическая кластеризация), позволяет выявлять скрытые группы клиентов на основе множества признаков [6]. Обучение с учителем применяется для предиктивной сегментации, в частности, прогнозирования вероятности отклика, оттока или совершения покупки. Современные исследования демонстрируют эффективность ансамблевых моделей (XGBoost, Random Forest) и глубокого обучения для повышения точности сегментации [7].

Важным трендом в области клиентской аналитики является интеграция методов объяснимого искусственного интеллекта (Explainable AI, XAI) [8]. В частности, подходы SHAP (Shapley Additive exPlanations) и LIME (Local Interpretable Model-agnostic Explanations) позволяют не только строить прогнозы, но и интерпретировать факторы, влияющие на решение модели [9]. Это обеспечивает формирование прозрачных и проверяемых аналитических выводов, что критически важно для принятия обоснованных стратегических решений в CRM-маркетинге.

Цифровые платформы управления клиентскими данными

Ключевым элементом современной MarTech-экосистемы выступают платформы класса CDP (Customer Data Platform), обеспечивающие консолидацию данных из различных источников и формирование единого клиентского профиля. CDP интегрируется с операционными системами, сайтами, мобильными приложениями, CRM и ERP-системами, предоставляя данные для аналитики и активации в маркетинговых каналах.

На российском рынке значимую роль играет платформа Mindbox, предоставляющая инструментарий для сегментации, триггерных коммуникаций и персонализации. Однако типовой сценарий использования CDP часто остается частично ручным: сегменты выгружаются в табличные форматы и дорабатываются аналитиками, что увеличивает трудоемкость и снижает воспроизводимость процессов.

Анализ процесса CRM-сегментации и постановка требований к платформе

В традиционной маркетинговой компании (или отдела маркетинга компании) фактический процесс подготовки сегментов начинается с формирования контент-плана стратегом на основе акций клиента, сезонных факторов и иных оснований для коммуникаций. Далее стратег соотносит каждую коммуникацию с сегментом аудитории в описательной форме («активные клиенты с покупками в категории X», «клиенты со снижением частоты заказов»), после чего переводит это описание в no-code фильтр внутри CDP-платформы. Полученная выборка проверяется, при необходимости корректируется, и только затем передается в контур подготовки коммуникации. Такой подход имеет ряд существенных ограничений:

зависимость от субъективной экспертной логики, т.к. выбор сегментов строится на представлениях стратега, а не на автоматически сформированных рекомендациях;

высокая трудоемкость ручной настройки фильтров специалистом в конкретных условиях с итеративной корректировкой результата;

недостаточная воспроизводимость результата, в связи с тем, что один и тот же сегмент может быть по-разному описан разными специалистами;

слабая формализация контроля качества данных при фрагментарной проверки корректности данных;

ограниченность предиктивного компонента в условиях построения сегментов на исторических данных без прогнозирования будущего поведения.

ИТ-ландшафт маркетинговой компании часто представлен разрозненностью используемых решений: CRM/CDP-платформа, табличные инструменты (Excel, SQL), отдельные скрипты для обработки данных не образуют единого технологического контура.

Анализ процесса CRM-сегментации позволил сформировать ключевые функциональные требования к цифровой платформе для автоматизации CRM-сегментации с использованием ML-прогнозирования и LLM-генерации контента:

мультипроектный режим работы с ролевой моделью доступа;

подключение источников данных (интеграция с CDP и ручная загрузка CSV/Excel);

автоматический контроль качества данных (Data Quality) с формированием DQ-отчетов;

конфигурация логики сегментации через профиль проекта и контент-план;

формирование аудитории по нескольким сценариям (тематическая, активностная, гибридная сегментация);

ML-ядро для прогнозирования вероятностей реакции клиента;

рекомендательный модуль для формирования A/B-вариантов стратегий;

экспорт сегментов с историей версий и анализом результатов кампаний.

Среди нефункциональных требований важно выделить: производительность (приемлемое время выполнения операций при больших объемах данных); масштабируемость (рост объема базы без переписывания логики); надежность (идемпотентность задач, ретраи при сбоях, регулярные бэкапы); безопасность (RBAC, шифрование секретов, аудит доступа, маскирование персональных данных); наблюдаемость (централизованные логи и метрики, дашборды состояния системы); объяснимость (понятные сообщения об ошибках, пояснения к сегментам и рекомендациям).

Архитектура и технологический стек платформы интеллектуальной сегментации клиентов

Разрабатываемая платформа реализует полный цикл интеллектуальной CRM-сегментации и построена по многоуровневому принципу:

клиентский уровень;

серверный уровень;

уровень хранения данных;

интеграционный уровень.

Клиентский уровень представлен веб-интерфейсом на React, TypeScript, Vite, Tailwind CSS, Zustand, Axios, React Router. Выбор React обусловлен его пригодностью для построения интерактивных интерфейсов, TypeScript – необходимостью типизации.

Серверный уровень является backend на Python с использованием FastAPI. Выбор Python обусловлен его широкой применимостью в задачах обработки данных и машинного обучения, FastAPI обеспечивает удобный механизм построения REST API и высокую скорость разработки.

Уровень хранения данных представлен реляционной базой данных для структурированных сущностей и файловое хранилище для датасетов и ML-артефактов. Интеграционный уровень обеспечивает получения данных из внешних систем (CDP, табличные источники).

Разработка первой работающей версии платформы интеллектуальной сегментации клиентов выполнялась с использованием инфраструктуры MVP на Yandex Cloud с 4 vCPU, 8 ГБ ОЗУ, 50 ГБ дискового пространства, ОС Ubuntu 24.04 LTS. Генерация на основе LLM выполняется на отдельной рабочей станции с GPU NVIDIA GeForce RTX 5070 Ti (12 ГБ VRAM).

Структура платформы представлена четырьмя взаимосвязанными блоками: модуль загрузки и подготовки данных, модуль контроля качества данных, модуль сегментации клиентской базы на основе машинного обучения, модуль генеративного ИИ. Совокупность этих блоков формирует единый технологический пайплайн, в рамках которого исходные данные преобразуются в практические аналитические результаты, пригодные для использования в CRM-маркетинге.

Модуль загрузки и подготовки данных является исходной точкой работы платформы. Процесс загрузки реализует два сценария: импорт из файлов CSV/Excel и подключение к внешнему источнику данных (CDP). После загрузки выполняется валидация формата, сохранение файла, определение структуры колонок и создание записи о датасете. Важной составляющей является маппинг полей для сопоставления колонок входного датасета с логическими сущностями CRM-аналитики (идентификатор клиента, дата заказа, сумма и т.д.). В процессе предобработки данных выполняется фильтрация некорректных записей, обработка пропусков, проверка идентификаторов клиентов, приведение дат к единому формату, устранение дубликатов и формирование согласованного временного ряда клиентских действий. Для заказов и событий из CRM/CDP-среды дополнительно может требоваться дедупликация и учет версии записи, если одно и то же событие фиксируется в системе несколько раз.

На рисунке 1 представлен фрагмент окна разрабатываемой платформы для подключения к данным на платформе Mindbox через протокол Delta Sharing.

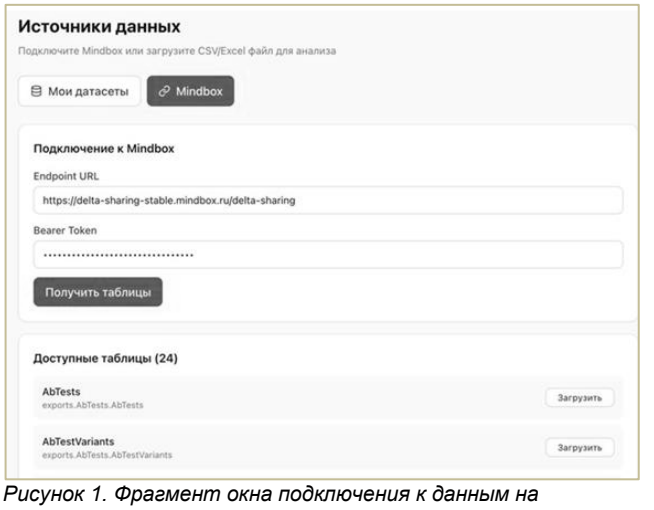


Рисунок 1. Фрагмент окна подключения к данным на платформе Mindbox через протокол Delta Sharing

В проектной реализации модуль загрузки и подготовки данных дополнен модулем контроля качества данных (Data Quality). Необходимость включения модуля Data Quality обусловлена тем, что качество сегментации и точность моделей машинного обучения напрямую зависят от полноты, корректности и согласованности исходных данных. После того как файл импортирован в систему или данные получены из внешнего источника, платформа формирует DQ-отчет, в котором отражаются ключевые характеристики набора: наличие обязательных полей, процент пропусков, наличие числового поля суммы, наличие дубликатов и другие параметры. В результате пользователь получает не просто набор данных, а формализованную оценку его пригодности для дальнейшего анализа.

Ядром интеллектуальной CRM-аналитики является модуль сегментации и ML-прогнозирования, реализующий сегментацию клиентов на нескольких уровнях:

правиловая и тематическая сегментация для отбора клиентов по заданным признакам (частота покупок, активность, категории интереса);

активностная сегментация на основе анализа взаимодействия клиента с брендом во времени (давность последней покупки, интенсивность заказов);

гибридная сегментация, как комбинация исторических признаков с результатами машинного обучения.

Ключевым компонентом является ML-модуль, решающий задачу предсказания следующей покупки клиента как задача регрессии: по истории заказов система прогнозирует количество дней до следующей покупки. Это имеет важное прикладное значение для CRM-маркетинга, позволяя точнее выбирать момент коммуникации и определять приоритетность сегментов.

Реализация ML-пайплайна включает:

загрузку и очистку истории заказов;

Feature engineering или формирование признакового пространства (количество заказов, средний интервал, средний чек, суммарные траты, характеристики скидок, временные признаки);

подготовку обучающей и тестовой выборок с временным разделением данных;

обучение модели с подбором гиперпараметров;

расчет метрик качества (MAE, RMSE, MAPE, R^2).

Для повышения качества прогнозирования используется ансамблевый подход, сочетающий несколько алгоритмов, что соответствует современным тенденциям в области клиентской аналитики [10]. Построенная модель используется для скоринга всей клиентской базы и формирования прикладных сегментов (частые, регулярные, редкие клиенты, группы с высокой/низкой ожидаемой активностью).

Визуальная демонстрация пользовательского интерфейса платформы представлена на рисунке 2 с демонстрацией экрана работы с ML-моделью и результатами сегментации клиентов.

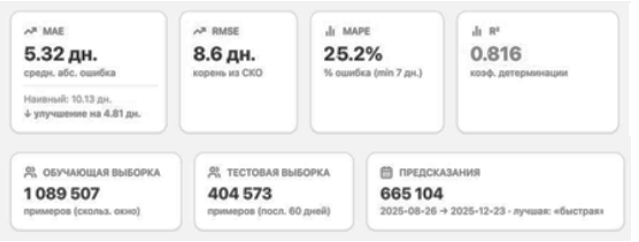


Рисунок 2. Фрагмент окна модуля ML-модели и результатов сегментации

Модуль генеративного ИИ ориентирован на поддержку коммуникационного контура CRM-маркетинга и решает задачи автоматизированной генерации черновиков маркетинговых сообщений (email, push), подготовку контент-планов для различных сегментов аудитории, формирование промптов и сценарных заготовок. С архитектурной точки зрения модуль генеративного ИИ целесообразно реализовывать как отдельный сервис или компонент, взаимодействующий с основным backend-контуром платформы. В этом случае backend передает в GenAI-модуль структурированный контекст, сформированный на основе данных CRM, результатов сегментации и параметров кампании, а в ответ получает сгенерированный текст или набор контентных вариантов. Такой подход соответствует общей модульной логике платформы и позволяет в дальнейшем подключать различные внешние API или внутренние языковые модели без перестройки всей системы.

Использование генеративного ИИ требует обязательного контроля качества контента. В практическом CRM-маркетинге недопустимо, чтобы сгенерированные сообщения автоматически отправлялись клиентам без проверки их содержания, соответствия бренду, юридическим ограничениям и бизнес-целям кампании. В проектной логике рассматриваемого решения генеративный ИИ следует рассматривать прежде всего как инструмент подготовки черновиков и вариантов, которые затем проходят редактуру и утверждение специалистом. Такая модель особенно важна на этапе апробации платформы, когда необходимо показать прикладную ценность технологии, не создавая избыточных рисков для качества клиентских коммуникаций.

В таблице 1 приведен сравнительный анализ LLM-моделей для модуля генерации контента.

С учетом ограничений для MVP-этапа проекта в качестве базовой модели генерации выбрана Qwen3 14B. Данная модель занимает промежуточное положение между малыми и крупными LLM: с одной стороны, она обеспечивает более высокое качество генерации, чем модели 7–8B-класса, с другой – остается существенно более реалистичной для локального запуска на видеокарте с 12 ГБ VRAM, чем модели класса 30B+. Данная модель обеспечивает необходимый уровень качества генерации при допустимой вычислительной нагрузке и соответствует имеющимся ресурсным ограничениям. При этом архитектура платформы остается расширяемой: в дальнейшем, по мере увеличения вычислительных ресурсов, возможен переход к более крупным моделям или использование гибридной схемы локального и внешнего инференса.

Апробация платформы и оценка эффективности проекта

Апробация платформы (первая версия платформы) для сегментации клиентов выполнялась на базе российской маркетинговой компании на исторических данных о заказах и коммуникациях компании.

Экономический эффект оценивался через сокращение трудозатрат сотрудников и перераспределение времени с технических операций на стратегические задачи. Показатели эффективности внедрения платформы представлены в таблице 2. Если в исходной модели значительная часть времени аналитиков и CRM-специалистов расходовалась на механические действия по переводу бизнес-логики сегмента в фильтр CDP, ручной проверке результатов и повторной корректировке условий, то в целевой модели эти операции частично автоматизируются. Следовательно, экономический эффект проявляется не только в прямом сокращении человеко-часов, но и в повышении производительности маркетинговой функции в целом.

Таблица 1
Сравнение LLM-моделей для модуля генерации контента

Модель	Размер	Требования к запуску	Преимущества	Ограничения	Вывод для проекта
Qwen3 14B	**14,8 млрд параметров**	Для локального запуска на GPU 12 ГБ требуется квантование Q4/Q5	Хороший баланс качества и ресурсоёмкости; поддержка длинного контекста;	Требует оптимизации инференса; при большом контексте возрастает нагрузка на VRAM	**Оптимальный вариант для MVP**
Qwen3 8B	8 млрд параметров	Запускается проще, чем 14B; менее требовательна к VRAM	Быстрая и легкая модель для базовой генерации	Ниже качество сложных инструкций и длинных сценариев	Подходит как резервная модель
Llama 3.1 8B Instruct	8 млрд параметров	Возможен локальный запуск в квантованном виде	Хорошее следование инструкциям; развитая экосистема	Может требовать дополнительной проверки качества русскоязычных текстов	Возможная альтернатива, но неосновная модель
Mistral 7B Instruct	7,3 млрд параметров	Низкие требования к ресурсам	Быстрая генерация, простота локального запуска	Ограниченное качество при сложной персонализации и длинном контексте	Подходит для простых черновиков
DeepSeek-R1-Distill-Qwen-14B	14 млрд параметров	Для GPU 12 ГБ требуется квантование	Хорошо подходит для аналитических рассуждений и объяснений	Избыточна для простых CRM-текстов; может быть медленнее	Может использоваться как дополнительная модель
Модели 30B–70B	30–70+ млрд параметров	Требуют существенно большего объема VRAM или серверных GPU	Более высокое качество генерации и устойчивость к сложным инструкциям	Не соответствуют ресурсам MVP; высокие инфраструктурные затраты	Не подходят для текущего MVP

Таблица 2
Показатели эффективности проекта разработки платформы интеллектуальной сегментации

Показатель	AS-IS (до внедрения)	TO-BE (после внедрения)	Эффект
Время создания сегмента	До 2 часов	До 30 минут	Сокращение до 75%
Количество кампаний в месяц	4-6 кампаний	15-20 кампаний	Рост в 3-4 раза
Точность таргетинга	Базовый rule-based	ML-предсказание отклика	Максимизация эффекта за счет данных, а не за счет экспертной оценки
Проверка качества данных	Ручная	DQ – автоматизированная проверка	Автоматизированная проверка пригодности данных уже на этапе подключения к данным

На основе данных таблицы 2 разработка и внедрение платформы интеллектуальной сегментации приводит к сокращению времени подготовки сегментов до 75%, росту числа кампаний в 3–4 раза и повышению точности таргетинга за счет обучения модели машинного обучения на клиентских данных и генерации контента с ИИ.

Заключение

В рамках проведенного исследования получилось разработать архитектуру и программную реализацию цифровой платформы, которая впервые объединяет автоматический контроль качества клиентских данных с формированием DQ-отчетов, прогнозирование временного интервала до следующей покупки на основе ансамблевых ML-моделей, генерацию персонализированного контента с использованием LLM Qwen3 14B. Предложенное решение отличается от существующих аналогов целостностью пайплайна и ориентацией на российский рынок MarTech-решений в условиях импортозамещения.

Автоматизация CRM-сегментации с использованием ML-прогнозирования и LLM-генерации контента позволила изменить структуру труда и логику формирования сегментов. Вместо последовательности «контент-план → экспертное описание → настройка фильтра → проверка → корректировка» в новой модели операции интегрируются в единый интерфейс и автоматизируются. Дальнейшее развитие системы предполагает расширение ML-функционала (модели оттока, LTV-прогнозирование), углубленную интеграцию генеративного ИИ для персонализации контента, подключение дополнительных источников данных и развитие self-service аналитики для маркетологов.

Разработанная платформа может рассматриваться как практически значимое ИТ-решение, ориентированное на автоматизацию клиентской аналитики и поддержку управленческих решений в контуре маркетинговых коммуникаций.

Литература

1. Ремизова А.А., Самигулин Т.Р. Поведенческая аналитика: анализ современного состояния и ее применение для решения задач бизнеса// Научный результат. Информационные технологии. – 2022. – Т. 7. – № 3. – С. 63-73. EDN: KJULXR
2. Демченко С.А., Макарчук Т.А., Чачис Д.Ю. Методология построения аналитической системы клиентской базы в условиях импортозамещения// Экономика строительства. – 2025. – №9. – С. 94-96. EDN: ZYJOVW
3. Макарчук Т.А., Анастасьев Д.А. Управление продуктовыми предпочтениями клиентов с использованием технологии PROCESS MINING// Современные проблемы цивилизации и устойчивого развития в информационном обществе: Сборник материалов VIII Международной научно-практической конференции. – Москва, 2022. – С. 119-122. EDN: XNTIZG
4. Ойнер О.К., Казаков С.Л. Поведенческая и психографическая сегментация как инструмент улучшения потребительского опыта// ЭКОНОМИКА. БИЗНЕС. БАНКИ. – 2019. – №12(38). – С. 46-53. EDN: YBJNXQ
5. Хлынин Э.В., Коровкина Н.И., Васин А.С., Быковская Е.В. Клиентоориентированный подход в процессе управления взаимоотношениями с клиентами и формирования клиентской базы //Финансы и кредит. – 2025. – Т. 31. № 7. – С. 34-51. EDN: FWVYXX
6. Бериков В.Б., Кутненко О.А., Пестунов И.А. Слабо-контролируемая групповая классификация// Вычислительные технологии. – 2024. – Т. 29. – № 1. – С. 45-58. EDN: LPNRCI
7. Ма С., Сенько О.В. Некоторые подходы к повышению точности прогнозирования с использованием ансамблевых методов// Электронные библиотеки. – 2025. – Т. 28. – № 6. – С. 1415-1434. EDN: CIXSIW
8. Aazmaan Ahmed Khan R.A. Customer segmentation using machine learning techniques// TUIJIN JISHU. – 2023. – Т.44. – №3. – С. 2051-2061. EDN: UXVMDI
9. Talaat F.M., Aljadani A., Alharthi B., Farsi M.A., Badawy M., Elhosseini M. A mathematical model for customer segmentation leveraging Deep Learning, Explainable AI and RFM analysis in targeted marketing// Mathematics. – 2023. – Т. 11. – № 18. – С. 3930-3933. EDN: PXNBGD
10. Староверов Б.А., Хамитов Р.Н. Реализация глубокого обучения для прогнозирования при помощи ансамбля нейронных сетей// Известия Тульского государственного университета. Технические науки. – 2023. –№ 4. – С. 185-189. EDN: MNHGOP

Automation of CRM segmentation using ML-prediction and LLM-based content generation

Makarchuk T.A., Mikhaylov A.N., Staver E.Yu.

Saint Petersburg State University of Economics, LLC "KOKOC GROUP", LLC Veim

JEL classification: C10, C50, C60, C61, C80, C87, C90

The article presents an approach to automating customer base segmentation for CRM marketing based on an intelligent web platform. The relevance of the study is driven by the growing volume of customer data and the need to transition from manual, expert-dependent segmentation methods to managed analytical processes. The aim of the work is to develop and test an intelligent segmentation system that integrates data loading and preparation modules, a customer base segmentation module, a machine learning module, and an AI-based content generation module. The article describes architectural solutions, justifies the choice of an LLM model, outlines the implementation of key platform modules, and presents the results of testing on data from a marketing services company. It is shown that implementing the proposed solution reduces campaign preparation time by up to 75% and increases the number of testable hypotheses by 3–4 times. The results confirm the practical significance of intelligent segmentation for improving marketing communication efficiency.

Keywords: CRM marketing, customer segmentation, machine learning, predictive analytics, generative artificial intelligence, CDP platform, marketing automation.

References

1. Remizova A.A., Samigulin T.R. Behavioral analytics: analysis of the current state and its application for solving business problems // Scientific Result. Information Technologies. – 2022. – Vol. 7. – No. 3. – P. 63-73.
2. Demchenko S.A., Makarchuk T.A., Chachis D.Yu. Methodology for building an analytical customer base system in the context of import substitution // Economics of Construction. – 2025. – No. 9. – P. 94-96.
3. Makarchuk T.A., Anastasyev D.A. Management of customer product preferences using PROCESS MINING technology // Modern Problems of Civilization and Sustainable Development in the Information Society: Proceedings of the VIII International Scientific and Practical Conference. – Moscow, 2022. – P. 119-122.
4. Oiner O.K., Kazakov S.L. Behavioral and psychographic segmentation as a tool for improving customer experience // ECONOMICS. BUSINESS. BANKS. – 2019. – No. 12(38). – P. 46-53.
5. Khlynin E.V., Korovkina N.I., Vasin A.S., Bykovskaya E.V. Customer-oriented approach in the process of customer relationship management and customer base formation // Finance and Credit. – 2025. – Vol. 31. – No. 7. – P. 34-51.
6. Berikov V.B., Kutnenko O.A., Pestunov I.A. Weakly supervised group classification // Computational Technologies. – 2024. – Vol. 29. – No. 1. – P. 45-58.
7. Ma S., Senko O.V. Some approaches to improving forecasting accuracy using ensemble methods // Electronic Libraries. – 2025. – Vol. 28. – No. 6. – P. 1415-1434.
8. Aazmaan Ahmed Khan R.A. Customer segmentation using machine learning techniques // TUIJIN JISHU. – 2023. – Vol. 44. – No. 3. – P. 2051-2061.
9. Talaat F.M., Aljadani A., Alharthi B., Farsi M.A., Badawy M., Elhosseini M. A mathematical model for customer segmentation leveraging Deep Learning, Explainable AI and RFM analysis in targeted marketing // Mathematics. – 2023. – Vol. 11. – No. 18. – P. 3930-3933.
10. Staroverov B.A., Khamitov R.N. Implementation of deep learning for forecasting using an ensemble of neural networks // News of Tula State University. Technical Sciences. – 2023. – No. 4. – P. 185-189.