

Ю.К. Зайцев

Российская академия народного хозяйства и государственной службы, Москва

## Использование элементов машинного обучения для оценки результатов маркетинговых кампаний в розничных банках<sup>1</sup>

**Аннотация.** Геополитическая напряженность и финансовая нестабильность меняют требования к капитализации розничных банков во многих странах. В банковской отрасли оптимизация проведения кампаний телемаркетинга является одним из ключевых вопросов операционной деятельности розничных банков, учитывая их растущую потребность в увеличении прибыли и сокращении затрат. Методы машинного обучения с использованием программы Python позволяют оценить результаты кампании розничных банков, направленной на привлечение клиентов к новым продуктам, и тем самым снизить издержки, связанные с поиском новых и удержанием старых клиентов. Регрессионный анализ показал, что наиболее влиятельными факторами для приобретения нового продукта банка – открытие депозитного вклада – являются продолжительность контакта с клиентом, а также дата контакта с клиентом. Статистический анализ данных позволил определить портрет потенциального клиента банка.

**Ключевые слова:** *розничный банк, машинное обучение, Python, маркетинговая кампания.*

Классификация JEL: F650, F410.

Для цитирования: **Зайцев Ю.К.** (2025). Использование элементов машинного обучения для оценки результатов маркетинговых кампаний в розничных банках // *Журнал Новой экономической ассоциации*. № 4 (69). С. 191–205.

DOI: 10.31737/22212264\_2025\_4\_191-205

EDN: DUPQON

### 1. Введение

Настоящее исследование посвящено анализу маркетинговой кампании продвижения банковских продуктов в розничном банке. Под *розничным банком* в настоящей работе будет пониматься банковский институт, предоставляющий широкий спектр финансовых услуг массового потребления, ориентированных на различные типы клиентов. Маркетинговые кампании позволяют розничным банкам повысить не только эффективность ведения бизнеса, но и его прибыльность. Анализ результатов маркетинговой кампании позволяет правильно определить потенциальных клиентов банка, выявить перспективные направления для инвестиций и открытия новых сфер развития бизнеса, а также грамотно спланировать будущие маркетинговые кампании (Liu, Zhang, 2008; Sing'oei, Wang, 2013).

Дистанционные формы взаимодействия с клиентами в рамках маркетинга посредством использования интернет- и телекоммуникационных технологий принято называть *телемаркетингом* (Kotler, Keller, 2012). Такой подход позволяет построить долгосрочные отношения с клиентом, правильно оценивая его потребности, предлагая ему подходящие банковские продукты (Nobibon, Leus, Spieksma, 2011). Это в существенной мере превосходит традиционный способ массового маркетинга (путем продвижения информации в широкие

<sup>1</sup> Статья подготовлена в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС.

массы населения) (Elsalamony, 2014). По сравнению с традиционным маркетингом телемаркетинг точно продвигает услуги отдельным клиентам. Это, в свою очередь, позволяет розничным банковским институтам повышать свою прибыль. Однако неграмотно спланированная маркетинговая кампания может привести к росту недовольства клиентов и к их оттоку (Mogo, Laugano, Cortez, 2012b).

В настоящей статье на основе открытых данных для тренинга в сфере машинного обучения будет предпринята попытка оценить факторы, влияющие на решение клиентов приобрести новый продукт на примере «срочного депозита». Данный продукт является наименее рискованным для розничного банка, поскольку не связан с такими дополнительными издержками, как, например, проверка платежеспособности клиента (в случае открытия потребительского кредита). Данные включают информацию о клиентах, которым в рамках телемаркетинга розничным банком предлагается оформить подписку на продукт «срочный депозит». Когда клиент подписывается на этот продукт, он помещает определенное количество денег на отдельный счет и не сможет получить доступа к этим средствам до истечения срока действия договора. Взамен клиент получает выплаты по процентам от банка по истечении этого срока.

В данной работе будет предпринята попытка определить *наиболее значимые факторы*, влияющие на поведение целевого клиента банка<sup>2</sup>, который с высокой вероятностью откроет депозит в рамках проведения кампании банка. Для достижения этой цели будет проведен эконометрический анализ. Анализ должен позволить объяснить связь между характеристиками клиента и целевой переменной, определяющей, был ли им открыт срочный вклад или нет. Структура работы отражает решение упомянутых выше задач и будет включать описание данных, анализ стилизованных фактов и визуализацию данных, моделирование и эмпирическую оценку, выводы и практические рекомендации.

## 2. Обзор литературы

На сегодняшний день существует недостаток исследований в сфере изучения поведения клиентов розничных банков. Однако все же представляется возможным выделить ряд уже опубликованных работ в этой сфере. К примеру, в статье (Lau, Chow, Liu, 2004) авторы анализируют применимость методов интеллектуального анализа данных (DM) в маркетинге в банковском секторе Гонконга, однако модель с использованием реальных данных авторами не тестировалась. В своем исследовании авторы (Martens, Provost, 2012) провели анализ клиентов с помощью псевдосоциальных сетей на основе переводов между заинтересованными сторонами. Также в своих работах авторы (Mogo, Laugano, Cortez, 2012a) проводили моделирование результатов кампании банковского телемаркетинга. В частности, в одной из работ экономисты предложили персональную и интеллектуальную систему поддержки принятия решений (Decision support system, DSS), которая может автоматически предсказывать результат телефонного контакта с клиентом с предложением приобрести депозитные контракты. Такой подход полезен для того, чтобы помочь менеджерам расставить приоритеты и выбрать клиентов, с которыми необходимо связываться во время банковских маркетинговых кампаний. Как следствие – издержки роз-

<sup>2</sup> Под целевым клиентом понимается клиент, который участвует в маркетинговой кампании банка.

ничного банка на проведение такого рода кампаний могут быть сокращены (Мого, Cortez, Rita, 2014).

В то время как в (Javaheri, Sepehri, Teimourpour, 2014) на примере иранского розничного банка авторы проанализировали, каким образом маркетинговая кампания в средствах массовой информации (СМИ) может повлиять на покупку нового банковского продукта.

Существенного внимания заслуживает также изучение работ российских исследователей в сфере машинного обучения для корпоративного управления и банковского бизнеса. Так, к примеру, А. А. Дулев в своей статье объясняет возможности снижения издержек кредитных организаций в банковском секторе за счет использования технологий искусственного интеллекта (Дулев, 2018). К. И. Лихоузов анализирует платформы для обработки данных в банковском секторе, которые позволяют финансовым организациям повысить свою прибыль, улучшить систему риск-менеджмента, а также усовершенствовать механизмы оценки кредитоспособности банковских институтов (Лихоузов, 2015). И. Ю. Литвин и А. Ю. Литвин показывают, каким образом инструменты машинного обучения могут помочь сформировать новые банковские продукты на основе изучения предпочтений клиентов (Литвин И., Литвин А., 2021). В статьях освещаются общие вопросы по перечисленным направлениям. Однако все же существует определенный недостаток изучения конкретных примеров и эмпирических исследований в заявленной сфере.

### 3. Описание данных

В исследовании будут использоваться данные, собранные розничным банком о клиентах. В работе были использованы данные, предоставленные платформой Kaggle для учебно-методических целей. Они включают наблюдения по розничным банкам с 2017 г. по 2022 г., в общей сложности по 52 944 контактам клиентов<sup>3</sup>. В целях моделирования первоначально было выполнено разделение по респондентам, при котором записи были разделены на данные для целей обучения (три года) и тестовые данные (два года). Данные для обучающей модели включили 50 651 респондента. Данные для тестовой модели включили наиболее последние 2293 контакта респондентов.

Данные включают такие сведения, как тип занятости, семейное положение, уровень образования, баланс счета, жилищный кредит, кредитное бремя (наличие кредита), способ контактирования, день контакта с клиентом (день), месяц контакта с клиентом (месяц), продолжительность взаимодействия с клиентом, информация о том, в скольких маркетинговых компаниях участвовал клиент (маркетинговая кампания), предыдущий опыт участия клиентов, открытие депозита (была ли кампания успешной) и т. д. Полный набор объясняющих переменных и их описание представлены в табл. 1. Набор переменных включает как числовые (количественные), так и категориальные (качественные) переменные.

<sup>3</sup> В целях недопущения рекламы продуктов розничных банков данные о рассматриваемых банках Kaggle (система организации конкурсов для исследования данных, а также социальная сеть специалистов для обработки данных и машинного обучения. Принадлежит корпорации Google) не приводятся.

Таблица 1.

Описание переменных, использованных в исследовании

Название переменной	Описание	Единица измерения	Тип
Тип занятости	От типа занятости зависит, сколько денег клиент может держать на банковском счете	Полная/ частичная	Категориальная
Семейное положение	Состоит ли клиент в браке	В браке/ не в браке	Категориальная
Уровень образования	Наличие у клиента определенного уровня образования	Наличие высшего образования и/ или выше	Категориальная
Дефолт	Наличие у клиента просроченного кредита	Да/ нет	Категориальная
Баланс на счете	Среднегодовой баланс счета	Сумма в денежных единицах	Количественная
Кредит	Объем обязательств по существующему кредиту	Сумма в денежных единицах	Категориальная
Контакт	Легко ли связаться с клиентом	Да / Нет	Категориальная
Дата	Дата контакта с клиентом	Дата	Категориальная
Месяц	Месяц контакта с клиентом	Месяц	Категориальная
Продолжительность контакта	Продолжительность последнего контакта	Число дней	Количественная
Маркетинговая кампания	Информация о том, в скольких маркетинговых кампаниях участвовал клиент	Число участия клиента	Количественная
Число дней	Сколько дней прошло после последнего контакта с клиентом	Число дней	Количественная
Предыдущий опыт обращения к клиенту	Сколько раз с клиентом устанавливался контакт во время одной кампании	Число обращений	Количественная
Приобретение депозита	Информация о том, была ли кампания успешной	Да / Нет	Категориальная
Возраст	Информация о возрасте клиентов	Число полных лет	Количественная
Пол	Информация о поле клиентов	Мужской / Женский	Количественная
Наличие жилья	Информация о наличии недвижимого имущества	Да / Нет	Категориальная

Источник: составлено автором.

#### 4. Стилизованные факты и анализ статистики

В данной части работы будет проведен статистический анализ данных существующих и новых клиентов розничного банка, участвующих в его кампании. Для розничного банка при проведении маркетинговой кампании большое значение имеет привлечение новых клиентов, поскольку они расширяют его клиентскую базу и, как следствие, повышают прибыльность бизнеса. Тем не менее, в настоящем исследовании клиенты банка – участники маркетинговой кампании – не подразделяются на «новых» и «старых» в силу ограниченности имеющихся данных для анализа. В этом контексте под *целью маркетинга* будет пониматься как привлечение новых, так и удержание старых клиентов банка.

В первую очередь была проведена оценка корреляции между переменными, связанными с фактическим участием клиентов непосредственно в кам-

пании банка (дата, продолжительность контакта, предыдущий опыт обращения к клиенту, маркетинговая кампания, количество дней). Дополнительно были рассмотрены две переменные – баланс на счете и возраст клиента – которые, по нашему мнению, в наибольшей степени могут быть связаны с вышеперечисленными переменными, определяющими участие клиента в кампании банка. Матрица корреляции показывает корреляцию на уровне 0,6 между числом дней после последнего контакта с клиентом и предыдущим опытом участия в кампании (рис. 1). Это означает, что в кампании принимают участие потенциально заинтересованные клиенты. Также существует небольшая корреляция между возрастом и остатком на счетах клиентов, что связано с возможностью клиентов накапливать в более продолжительном периоде времени.

Было выявлено строго отрицательное значение корреляции между балансом на счете и продолжительностью последнего контакта с клиентом. Большая часть наблюдений проводилась среди розничных клиентов, имеющих небольшие денежные средства на балансе и обратившихся к услугам банка относительно недавно (рис. 2).

<i>age</i>	1	0,11	-0,00076	0,00019	-0,0053	0,0028	0,02
<i>balance</i>	0,11	1	0,01	0,022	-0,014	0,017	0,031
<i>day</i>	-0,00076	0,01	1	-0,019	0,14	-0,077	-0,059
<i>duration</i>	0,00019	0,022	-0,019	1	-0,042	-0,027	-0,027
<i>campaign</i>	-0,0053	-0,014	0,14	-0,042	1	-0,1	-0,05
<i>pdays</i>	0,0028	0,017	-0,077	-0,027	-0,1	1	0,51
<i>previous</i>	0,02	0,031	-0,059	-0,027	-0,05	0,51	1
	<i>age</i>	<i>balance</i>	<i>day</i>	<i>duration</i>	<i>campaign</i>	<i>pdays</i>	<i>previous</i>

Рис. 1

Матрица корреляции

Источник: составлено автором.

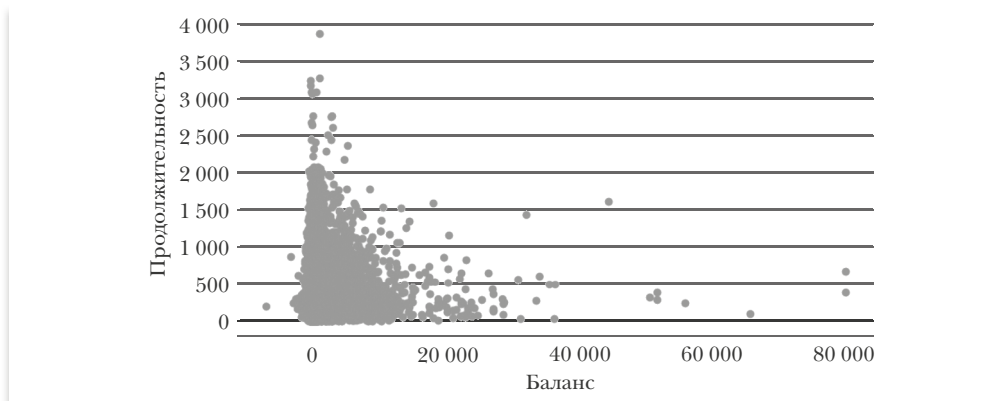
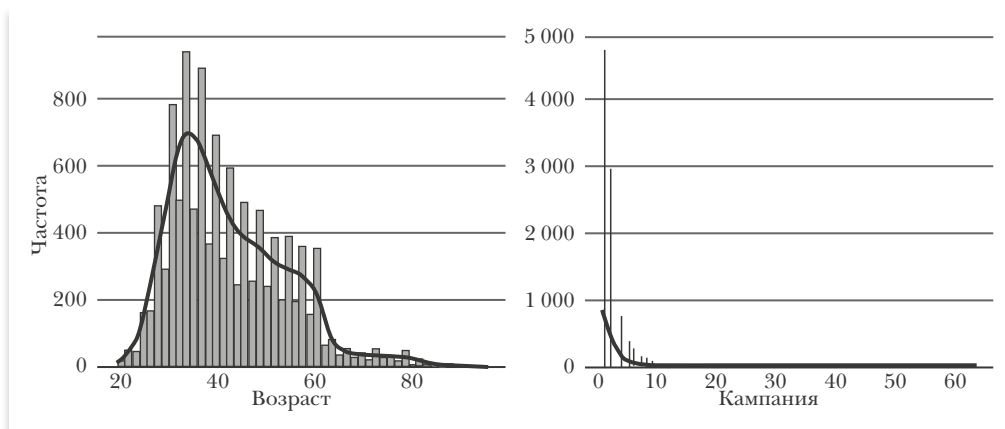


Рис. 2

Диаграмма рассеяния по переменным «Баланс счета» и «Продолжительность контакта»

Источник: составлено автором.

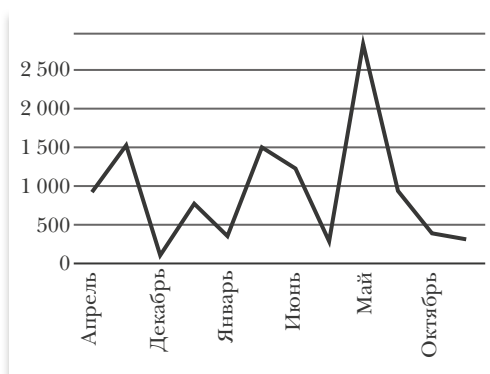
**Рис. 3.**

Распределение переменных «Возраст» и «Участие в кампании»

Источник: составлено автором.

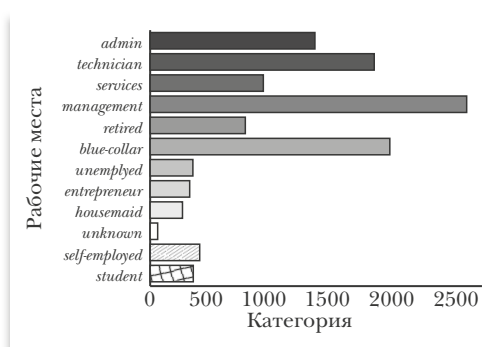
Диаграмма распределения клиентов по возрастным группам показывает, что большинству клиентов, участвующих в кампании, от 30 до 40 лет (рис. 3). Более того, большинство из них участвовали в одной или двух кампаниях. При этом предусматривается, что данная группа является наиболее экономически активной частью населения, обладающей достаточными ресурсами для участия в кампании розничного банка и являющейся наиболее вероятной его целевой группой.

Другой спектр анализа – распределение категориальных переменных, указывающих на занятость клиентов. На рис. 4–5 видно, что большинство клиентов заняты в сфере управления или являются рабочими, или техническими специалистами. Общая численность безработных, предпринимателей, самозанятых и студентов не превышает 2500 человек.

**Рис. 4.**

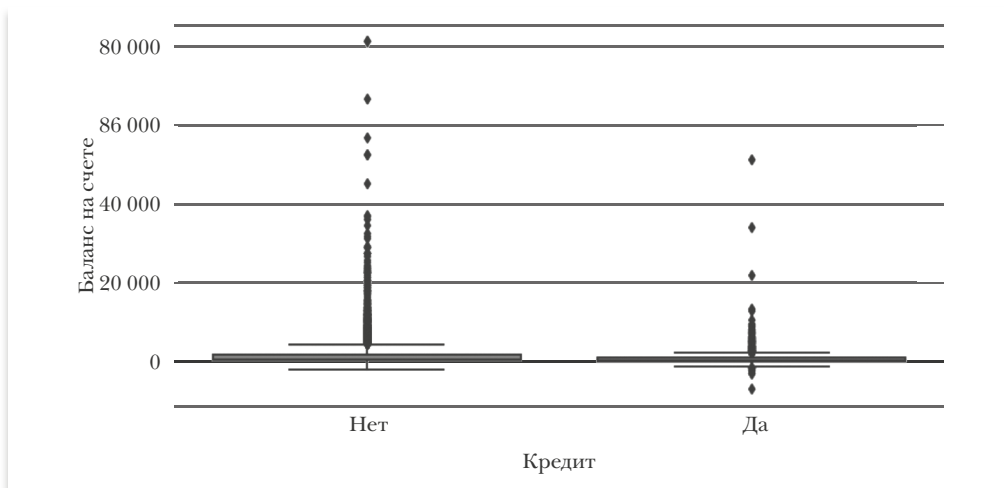
Линейный график числа записей по месяцам

Источник: составлено автором.

**Рис. 5.**

Распределение категориальных переменных, характеризующих рабочие места

Источник: составлено автором.



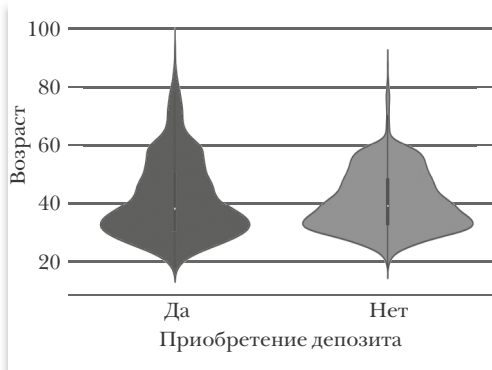
**Рис. 6.**

Диаграмма столбца «Баланс на счете», сгруппированная по столбцу «Кредит»

Источник: составлено автором.

На рис. 6 показана корреляция между кредитами и остатками на счетах клиентов. Те клиенты, у которых меньше остатков на счетах, имеют больше кредитов. Эта группа клиентов с меньшей вероятностью имеет депозит в банке. Однако большинство клиентов, не имеющих кредитов, потенциально в большей степени готовы открывать депозиты.

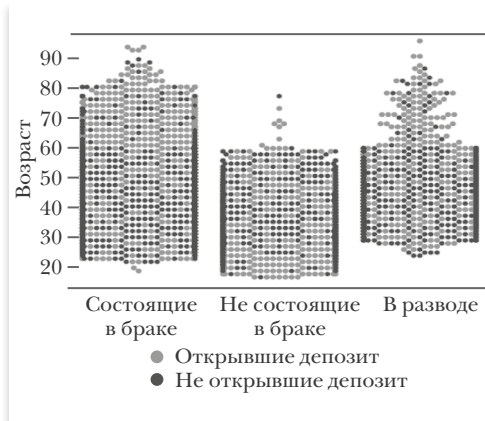
На рис. 7 показано распределение клиентов по возрасту, открывших депозитный счет в банке. Большинство клиентов, открывших депозит, моложе 40 лет. Это соответствует тенденции, связанной с клиентами, не открывшими депозитных счетов по программе банка. Данные на рис. 8 подтверждают приведенный



**Рис. 7.**

Диаграмма по столбцу «Возраст», сгруппированному по переменной «Приобретение депозита»

Источник: составлено автором.



**Рис. 8.**

Семейное положение клиентов

Источник: составлено автором.

выше аргумент о том, что состоящие в браке клиенты с большей вероятностью депонируют свои сбережения.

Визуальный статистический анализ данных позволяет сделать некоторые предварительные выводы о целевом клиенте кампании банка. Во-первых, наиболее вероятный клиент имеет семью. Во-вторых, целевой клиент — в возрасте от 30 до 40 лет. В-третьих, у потенциального клиента нет кредита. В-четвертых, целевой клиент скорее всего работает в сфере управления или является служащим или техническим специалистом. В-пятых, с целевым клиентом сотрудники банка общались недавно, но не так часто.

Анализ стилизованных фактов показал, что статистические данные имеют нормальное распределение. Во временном ряду имеются все необходимые данные. Таким образом, нет необходимости удалять экстерналии или создавать новые фиктивные переменные. Статистический обзор данных показал, что около 47,4% рассмотренных клиентов имеют действующие депозитные банковские вклады. Анализ корреляции данных на основе результатов логит-регрессии показывает, что предыдущий опыт размещения депозитов и текущий баланс влияют на окончательное решение клиента об открытии депозита.

## 6. Методология оценки

Регрессионный анализ предполагал использование программы Python и включал несколько этапов. Сначала были подготовлены данные. Были определены зависимые и независимые переменные, которые в дальнейшем были использованы в модели. Для этого были созданы соответствующие коды. Независимыми переменными являются «возраст», «баланс», «маркетинговая кампания», «предыдущий опыт обращения», «уровень образования», «дефолт», «наличие жилья», «кредит», «контакт», «месяц», «тип занятости», «семейное положение», «продолжительность контакта», «дата». Зависимой переменной является «депозит», которая является дамми-переменной и принимает значения 1 («да») и 0 («нет»). Описание гипотез для переменных представлены в табл. 2.

Переменные делятся на категориальные и числовые (см. табл. 1). Обе модели для числовых и категориальных переменных были протестированы и обучены с помощью соответствующих кодов.

Для повышения эффективности моделирования данные были случайно разделены на две части. Первая часть данных была использована для обучения модели<sup>4</sup>, вторая часть данных — для тестирования (оценки)<sup>5</sup> модели. Таким образом, числовые и категориальные данные были разделены для обучающей и тестовой части модели с долей тестирования 20% (не были использованы при обучении модели). Это позволило получить более реальную оценку модели.

Для дальнейшего обучения модели и корректировки данных был применен преобразователь столбцов (*clean.compose*). Этот инструмент позволяет отдельно преобразовывать разные столбцы или подмножества столбцов входных данных, а признаки, сгенерированные каждым преобразователем, которые могут объединяться для формирования единого пространства признаков. Это

<sup>4</sup> Обучающий набор данных — часть базы данных, на которой модель будет учиться и тренироваться. Поэтому на этом этапе будут оценены параметры модели.

<sup>5</sup> Тестовый набор данных — та часть базы данных, которая никогда не изучалась моделью, поэтому именно по этой части данных проходит оценка.

Таблица 2.

## Проверяемые гипотезы для оценки переменных

Объясняющая переменная	Проверяемая гипотеза для переменной
Возраст	Возраст в диапазоне от 30 до 40 лет положительно воздействует на решение клиента участвовать в кампании банка
Баланс	Рост значения среднегодового баланса счета положительно влияет на решение клиента участвовать в кампании банка
Маркетинговая кампания	Продолжительный опыт участия клиента в прошедших кампаниях банка положительно влияет на решение клиента участвовать в текущей кампании банка
Предыдущий опыт обращения	Большее число обращений представителей банка к клиенту положительно влияет на решение клиента участвовать в текущей кампании банка
Уровень образования	Наличие высшего образования положительно влияет на решение клиента участвовать в текущей кампании банка
Дефолт	Наличие у клиента просроченного кредита/ кредитов отрицательно влияет на решение клиента участвовать в текущей кампании банка
Наличие жилья	Наличие недвижимого имущества у клиента положительно влияет на решение клиента участвовать в текущей кампании банка
Кредит	Наличие текущих кредитных обязательств у клиента отрицательно влияет на его решение участвовать в текущей кампании банка
Продолжительность контакта	Многочисленное обращение к клиенту со стороны банка положительно влияет на его решение участвовать в текущей кампании банка
Контакт	Легкость и быстрота связи с клиентом положительно влияют на его решение участвовать в текущей кампании банка
Семейное положение	Если клиент находится в браке, то это положительно влияет на его решение участвовать в текущей кампании банка
Тип занятости	Полная занятость клиента положительно влияет на его решение участвовать в текущей кампании банка.
Дата / месяц	Последняя дата / месяц контакта положительно влияет на решение клиента участвовать в текущей кампании банка

Источник: составлено автором.

полезно для гетерогенных данных, чтобы объединить несколько механизмов извлечения признаков или преобразований в один механизм.

Все категориальные (нечисловые) переменные необходимо перекодировать, поскольку алгоритмы машинного обучения используют только количественные переменные в качестве входящих данных для модели. Формат данных важен для качества модели.

Как и другие переменные, целевую переменную также необходимо перекодировать. Для этого был использован специальный преобразователь OneHotEncoder. Недостающие данные для новой модели были заполнены с помощью метода SimpleImputer. Функция конвейера использовалась для сбора всех этих преобразований в одном объекте. Конвейер обучил модель на обучающем наборе данных и сделал оценку на тестовом наборе. Для стандартизации данных и преобразования к нашим данным был использован преобразователь StandardScaler<sup>6</sup>. Используя StandardScaler, мы оценили параметры обучающего набора и применили их к тестовому набору.

<sup>6</sup> Метод StandardScaler вычисляет эмпирическое среднее  $\mu$  и эмпирическое стандартное отклонение  $\sigma$  каждой переменной в наборе данных, а затем сохраняет их в атрибутах экземпляра StandardScaler.

## 7. Оценка моделей

Основной гипотезой для оценки модели является предположение о том, что клиент банка станет участником кампании банка (открытие депозита), если он соответствует критериям целевого клиента, которые были выявлены во второй части настоящей статьи (наличие семьи, возраст от 30 до 40 лет, отсутствие текущего кредита, работа по обозначенным специальностям, наличие недавнего контакта со стороны банка). В свою очередь, предполагается, что существует статистически значимая связь между перечисленными выше объясняющими факторами и объясняемой дамми-переменной (открытие депозита).

Решение практической задачи оценить факторы, влияющие на решение клиента открыть депозит, связано с оценкой трех моделей. Во-первых, после ряда подготовительных шагов мы рассмотрели первую модель логистической регрессии (Accuracy Score of the Logistic Regression Model) в качестве базовой, поскольку её можно быстро обучить и легко оценить. В наборе данных очень мало выбросов, поэтому обучение модели прошло без дополнительных трудностей. Вторая модель, обученная для регрессионного анализа, — это модель дерева решений (Accuracy score of the Decision Tree Model), которая также очень быстро обучается. Древовидные модели устойчивы к выбросам. Наконец, третья модель, модель случайного классификатора (Random Forest Model), является более сложной, но требует более точной настройки. Она также устойчива к выбросам. Оценка перечисленных моделей дала следующие результаты (табл. 3), близкие по значению к единице. Это означает, что оценки, согласно всем трем моделям, близки к верным.

**Таблица 3.**  
Результаты оценок регрессионных моделей

Модель оценки	Оценка
Accuracy score of the Logistic Regression Model	0,78
Accuracy score of the Decision Tree Model	0,78
Accuracy score of the Random Forest Model	0,84

*Источник:* составлено автором.

Далее были опробованы другие методы. Мы закодировали модальности целевой переменной с помощью LabelEncoder, оценив кодировку на обучающем наборе данных и применив её к обучающему и тестовому наборам данных. Мы попытались также оценить новый набор данных, используя метод Decision Tree Regressor Model. Полученный результат оказался близок к нулю (0,11), что говорит о несостоятельности модели для интерпретации данных.

На следующем этапе мы продолжаем оценивать регрессию, используя различные инструменты классификации, такие как MLP Classifier<sup>7</sup>, K Neighbors Classifier<sup>8</sup>, Gradient Boosting Classifier<sup>9</sup>, Gaussian NB<sup>10</sup>.

<sup>7</sup> Класс MLP Classifier реализует алгоритм многоуровневого перцептрона (MLP), который обучается с использованием обратного распространения ошибки. В настоящее время MLP Classifier поддерживает только функцию перекрестных энтропийных потерь, которая позволяет оценивать вероятность, запуская метод Predict\_proba.

<sup>8</sup> K Neighbours Classifier ищет пять ближайших соседей. Классификатор использует евклидово расстояние для определения близости между соседними точками.

<sup>9</sup> Gradient Boosting Classifier строит аддитивную модель поэтапно; это позволяет оптимизировать произвольные дифференцируемые функции потерь. На каждом этапе дерева регрессии n\_classes\_ подбираются по отрицательному градиенту функции потерь, например потеря двоичного или многоклассового журнала. Бинарная классификация — особый случай, когда создается только одно дерево регрессии.

<sup>10</sup> Gaussian NB — метод классификации, используемый в машинном обучении (ML), основанный на вероятностном подходе и распределении Гаусса. Гауссово-наивный байесовский подход предполагает, что каждый параметр (также называемый признаками, или предикторами) имеет независимую способность оценивать выходную переменную.

**Таблица 4.**

Результаты регрессии  
с использованием классификации

Индикатор	Оценка
Accuracy score of the K Neighbors Classifier	0,73
Accuracy score of the MLP Classifier	0,67
Accuracy score of the Gradient Boosting Classifier	0,83
Accuracy score of the Gaussian NB	0,72

Источник: составлено автором.

Все методы представляют собой разные подходы к классификации данных (табл. 4).

Из приведенной выше части статьи мы видим, что лучший метод оценки – Random Forest Classifier с оценкой 0,84; второй метод по эффективности показателей – Gradient Boosting Classifier с оценкой 0,83.

Мы будем рассматривать модель логистической регрессии в качестве базовой, поскольку ее можно быстро обучить и легко оценить. В наборе данных очень мало выбросов. Это значит,

что тренировка модели должна пройти удачно (табл. 5). В модель были включены оценка средневзвешенных значений, оценка макро/средних величин, точность модели (*accuracy*).

**Таблица 5.**

Результаты оценки логистической регрессии

Переменная	<i>precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>support</i>
<i>no</i>	0,78	0,80	0,79	1173
<i>yes</i>	0,78	0,75	0,77	1060
<i>accuracy</i>	0,78	2233		
<i>macro avg</i>	0,78	0,78	0,78	2233
<i>weighted avg</i>	0,78	0,78	0,78	2233
Decision Tree Classifier				
<i>no</i>	0,80	0,78	0,79	1173
<i>yes</i>	0,76	0,78	0,77	1060
<i>accuracy</i>	0,78	2233		
<i>macro avg</i>	0,78	0,78	0,78	2233
<i>weighted avg</i>	0,78	0,78	0,78	2233
Random Forest Classifier				
<i>no</i>	0,88	0,81	0,84	1173
<i>yes</i>	0,80	0,88	0,84	1060
<i>accuracy</i>	0,84	2233		
<i>macro avg</i>	0,84	0,84	0,84	2233
<i>weighted avg</i>	0,84	0,84	0,84	2233
Gradient Boosting Classifier				
<i>no</i>	0,62	0,95	0,75	1173
<i>yes</i>	0,87	0,35	0,50	1060
<i>accuracy</i>	0,67	2233		
<i>macro avg</i>	0,75	0,65	0,62	2233
<i>weighted avg</i>	0,74	0,67	0,63	2233

Источник: составлено автором.

**Таблица 6.**  
Результаты оценки

Переменная	Оценка	Характеристика переменной
Тип занятости	(0,068005)	Object
Семейное положение	(0,068793)	Object
Уровень образования	(0,062902)	Object
Дефолт	(0,054642)	Object
Баланс	(0,054387)	int64
Наличие жилья	(0,053448)	object
Кредит	(0,049144)	object
Контакт	(0,042083)	object
Дата	(0,041017)	int64
Месяц	(0,331534)	object
Продолжительность контакта	(0,221839)	int64
Маркетинговая кампания	(0,019038)	int64
Предыдущий опыт обращения	(0,011140)	int64

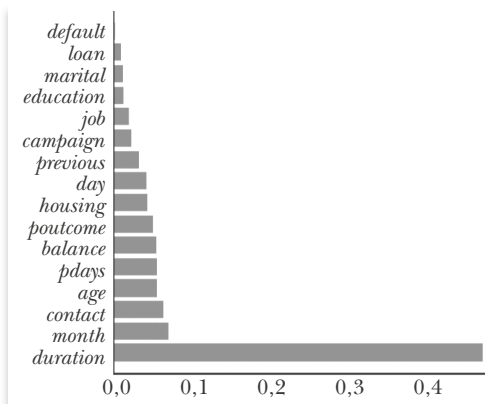
Источник: составлено автором.

Моделирование продолжалось перебором заданных значений параметров для оценки. Метод Grid Search CV применялся для подгонки и оценки. Также были использованы функции «*transform*» и «*inverse\_transform*». Для Random Forest Classifier были выбраны параметры с наилучшими значениями, использовался Grid Search CV.

Параметры<sup>11</sup>, имеющие лучшую оценку, были использованы для оценки и определения наиболее важных переменных с помощью метода *Features\_importances*.

Была создана новая зависимая переменная *y* под названием «*target\_name*» и независимые переменные в подмножестве *X*. Подмножество было разделено на две части – *X\_train* и *X\_test*. Была также проведена классификация по четырем информативным признакам, которые были определены методом Grid Search CV. Результаты этой оценки представлены в табл. 6. Регрессионный анализ показал, что наиболее влиятельными факторами для открытия вклада является продолжительность контакта и месяц контакта.

На рис. 9 была проведена попытка определить данные, связанные с объектами, с помощью случайного



**Рис. 9.**

Определение данных, связанных с объектами, с помощью случайного классификатора

Источник: составлено автором.

<sup>11</sup> К этим параметрам относятся: { 'max\_length': 10, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'n\_estimators': 100 }.

классификатора (Random Forest Classifier). Согласно классификатору наиболее важными факторами являются продолжительность сотрудничества (*duration*), месяц (*month*) и установленный контакт с клиентом (*contact*). Остальные факторы – возраст (*age*), баланс (*balance*), недвижимостью (*housing*), заем (*loan*), семейное положение (*marital*), наличие работы (*job*), предыдущий опыт контакта с клиентом (*previous*), образование (*education*), опыт участия в других кампаниях (*campaign*), дата (*data*), неучастие (*default*), предыдущий результат (*poutcome*), длительность контакта в предыдущий раз (*pdays*) – являются менее важными.

## 8. Заключение

Основной целью исследования было выявление факторов, влияющих на решение клиента об открытии депозита. Для этого был проведен статистический визуальный и регрессионный анализ. На основе статистического анализа удалось составить портрет целевого клиента рекламной программы банка. Во-первых, наиболее вероятный клиент имеет семью. Во-вторых, возраст целевого клиента – от 30 до 40 лет. В-третьих, у потенциального клиента нет действующих кредитов. В-четвертых, целевой клиент скорее всего работает в управленческой сфере или является служащим или техническим специалистом. В-пятых, к целевому клиенту сотрудники банка обращались недавно, но делали это нечасто.

Регрессионный анализ показал, что наиболее влиятельными факторами для открытия вклада являются продолжительность контакта, время (месяц) контакта.

Проведенное нами исследование показало возможности использования методов машинного обучения, которые позволяют оптимизировать процессы будущих маркетинговых кампаний розничных банков с точки зрения выявления факторов поведения клиентов, определяющих их результативность и позволяющих повысить прибыль финансовых институтов. На примере регрессионного анализа были выявлены характеристики потенциальных клиентов розничного банка. Это, в свою очередь, позволяет определить перспективные направления для развития банковского бизнеса, а также спланировать будущие маркетинговые кампании.

Предложенный подход оценки маркетинговой кампании может быть использован как на уровне отдельно взятых банковских отделений, так и масштабирован в отношении всей сети банковского института. Тем не менее, микромаркетинг может повысить результативность бизнеса отдельных сегментов.

Несмотря на то что из проведенного анализа сложно сформировать универсальные рекомендации для всех маркетинговых кампаний розничных банков, поскольку финансовые организации работают с разными продуктами, ориентированными на различных потребителей, представляется возможным обозначить универсальные направления для успешной маркетинговой кампании. Во-первых, необходимо учитывать время проведения кампании. Во-вторых, нужно увеличивать время контакта с потенциальными клиентами банка (включая диверсификацию каналов коммуникации). В-третьих, требуется фокусироваться на конкретных категориях клиентов (в рассмотренном случае в возрасте от 30 до 40 лет). В-четвертых, желательно формировать целевые группы клиентов по социально-экономическим категориям. Таким образом, представляется возможным концентрироваться на тех группах клиентов, которые потенциально более интересны для банка. Концентрация

позволит банковскому институту более эффективно распределить ресурсы кампании на главных факторах – времени и продолжительности контакта банка с клиентом.

Предложенный подход может быть использован не только в отношении маркетинговой кампании для открытия срочного депозита, но и распространен на другие банковские депозитные продукты.

Ограничения исследования связаны прежде всего с отсутствием полных данных, включая информацию о новых клиентах банка. Очевидно, существуют и другие показатели, характеризующие поведение клиентов банка и их решение принять участие в программе банка и открыть депозит. Однако использование случайных данных из других источников испортило бы выборку имеющихся данных. Именно поэтому было принято решение сосредоточиться только на данных Kaggle. Чем больше разнообразие характеристик клиентов, тем больше шансов экстраполировать полученные данные на другие случаи.

Разумеется, использованные в исследовании подходы не являются исчерпывающим инструментарием. Дальнейшие исследования могут опираться на другие исследовательские инструменты, в том числе связанные с выявлением причинно-следственных связей между переменными. Именно в этом направлении выявляются наибольшие трудности. Существуют некоторые наборы данных, которые коррелируют друг с другом. Чтобы продолжить дальнейшее моделирование, необходимо проводить такие дополнительные тесты на причинно-следственные связи, как тест Грейнджера и т.д. Учитывая ограниченные рамки текущего исследования, другие тесты могут быть проведены в рамках будущих работ.

#### ЛИТЕРАТУРА / REFERENCES

- Дулев А.А.** (2018). Внедрение искусственного интеллекта в деятельность кредитных организаций // *Хроноэкономика*. № 5 (13). С. 27–30. [**Dulev A.A.** (2018). Implementation of artificial intelligence in the activities of credit institutions. *ChronoEconomics*, 5 (13), 27–30 (in Russian).]
- Литвин И.Ю., Литвин А.Ю.** (2021). Искусственный интеллект как одно из направлений инновационного развития банковского сектора // *Инновационное развитие экономики*. № 2–3 (62–63). С. 42–46. [**Litvin I.Yu., Litvin A.Yu.** (2021). Artificial intelligence as one of the directions of innovative development of the banking sector. *Innovative Development of Economy*, 2–3 (62–63), 42–46 (in Russian).]
- Лихоузов К.И.** (2015). Применение задач машинного обучения на платформе распределенных вычислений больших данных в банковской сфере. *Scientific Cooperation Center “Interactive plus”* [сайт]. Режим доступа: <https://interactive-plus.ru/e-articles/858/Action858-559532.pdf> [**Likhouзов K.I.** (2015). Application of machine learning tasks on the platform of distributed big data computing in the banking sector. *Scientific Cooperation Center “Interactive plus”* [website]. Available at: <https://interactive-plus.ru/e-articles/858/Action858-559532.pdf> (in Russian).]
- Elsalamony H.A.** (2014). Bank direct marketing analysis of data mining techniques. *International Journal of Computer Applications*, 85, 12–22. DOI: 10.5120/14852-3218
- Javaheri S.H., Sepehri M.M., Teimourpour B.** (2014). Response modeling in direct marketing: A data mining based approach for target selection. In: Y. Zhao, Y. Cen (eds.). “*Data mining applications with R*”. Chapter 6. Elsevier, 153–178.

- Kotler P., Keller K.L.** (2012). *Framework for marketing management*. 5<sup>th</sup> ed. Boston: Pearson Education Limited, Prentice Hall.
- Lau K.N., Chow H., Liu C.** (2004). A database approach to cross selling in the banking industry: Practices, strategies and challenges. *Journal of Database Marketing and Customer Strategy Management*, 11 (3), 216–234.
- Liu F.-J., Zhang S.** (2008). Discussion of financial product innovation and marketing mode. *China Soft Science*, 2, 65–75.
- Martens D., Provost F.** (2011). Pseudo-social network targeting from consumer transaction data. *NYU Working Papers Series*, CeDER-11-05.
- Moro S., Cortez P., Rita P.** (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22–31. DOI: 10.1016/j.dss.2014.03.001
- Moro S., Laureano R., Cortez P.** (2012a). Enhancing bank direct marketing through data mining. *Decision Support Systems*, 62 (2014), 22–31.
- Moro S., Laureano R., Cortez P.** (2012b) Enhancing bank direct marketing through data mining. *Proceedings of the Forty-First International Conference of the European Marketing Academy, European Marketing Academy*, 1–8.
- Nobibon F.T., Leus R., Spieksma F.** (2011). Optimization models for targeted offers in direct marketing: Exact and heuristic algorithms. *European Journal of Operational Research*, 210 (3), 670–683.
- Sing'oei L., Wang J.** (2013). Data mining framework for direct marketing: A case study of bank marketing. *International Journal of Computer Science Issues (IJCSI)*, 10, 198.

Поступила в редакцию 04.08.2024

Received 04.08.2024

**Iu.K. Zaicev**

Russian Academy of National Economy and Public Administration, Moscow, Russia

## Using machine learning elements to assess the results of marketing campaigns in retail banks<sup>12</sup>

**Abstract.** Geopolitical tensions and financial instability are changing the capitalization requirements of retail banks in many countries. In the banking industry, optimizing telemarketing campaigns is one of the key operational issues for retail banks, given their growing need to increase profits and reduce costs. Machine learning methods using the “Python” program make it possible to assess the results of a retail banking company in attracting customers to new products, and thereby reduce the costs of searching new and retaining existing customers. Regression analysis showed that the most influential factors for purchasing a new bank product – opening a deposit account – are the duration of contact with the client, as well as the month of contact. Moreover, statistical analysis made possible to determine the portrait of a potential bank client.

**Keywords:** retail bank, machine training, Python, marketing campaign.

JEL Classification: F650, F410.

For reference: **Zaicev Iu.K.** (2025). Using machine learning elements to assess the results of marketing campaigns in retail banks. *Journal of the New Economic Association*, 4 (69), 191–205 (in Russian).

DOI: 10.31737/22212264\_2025\_4\_191-205

EDN: DUPQON

<sup>12</sup> The article was prepared as part of the research work of the state assignment of the Russian Presidential Academy of National Economy and Public Administration.