УДК 004.8:368.03
DOI: 10.26102/2310-6018/2025.50.3.046

Оценка точности и производительности моделей машинного обучения для прогнозирования удовлетворенности клиентов страховой компании

К.Г. Мадияров✉

Новосибирский государственный университет экономики и управления «НИНХ», Новосибирск, Российская Федерация

Резюме. В работе представлено исследование по прогнозированию удовлетворенности клиентов страховой компании на основе методов машинного обучения. Актуальность темы обусловлена высокой конкуренцией на страховом рынке и необходимостью удержания клиентов за счет повышения их удовлетворенности обслуживанием. Цель исследования – оценить точность и производительность моделей, способных предсказать уровень удовлетворенности клиента страховой услугой по данным о взаимодействии клиента с компанией. В качестве методов использованы алгоритмы классификации. Проведена оценка точности и производительности моделей на реальных данных опросов клиентов страховых компаний. Лучшими оказались ансамблевые методы – случайный лес и градиентный бустинг, продемонстрировавшие точность прогноза удовлетворенности до 85 %, существенно превосходя более простые модели. Показано, что градиентный бустинг позволяет учитывать нелинейные зависимости факторов, например, наличие эскалации обращения, и тем самым более точно выявлять «неудовлетворенных» клиентов. В настоящее время подобное прогнозирование в страховых компаниях либо не осуществляется, либо существенно опирается на случайные факторы. Это приводит либо к слишком частым жалобам, либо к низкой удовлетворенности клиентов с их последующим оттоком. Материалы статьи представляют практическую ценность для страховых организаций: внедрение разработанных моделей позволит оперативно идентифицировать клиентов с риском неудовлетворенности и обоснованно применять превентивные меры, например, дополнительные сервисные меры или компенсации для повышения их удовлетворенности.

Ключевые слова: удовлетворенность клиентов, страховая компания, машинное обучение, прогнозирование, градиентный бустинг, точность модели.

Для цитирования: Мадияров К.Г. Оценка точности и производительности моделей машинного обучения для прогнозирования удовлетворенности клиентов страховой компании. *Моделирование, оптимизация и информационные технологии.* 2025;13(3). URL: https://moitvivt.ru/ru/journal/pdf?id=2041 DOI: 10.26102/2310-6018/2025.50.3.043

Evaluation of accuracy and performance of machine learning models for prognosis customer satisfaction in insurance companies

K.G. Madiyarov✉

Novosibirsk State University of Economics and Management, Novosibirsk, the Russian Federation

Abstract. The paper presents a study on predicting customer satisfaction of an insurance company based on machine learning methods. The relevance of the topic is due to the high competition in the insurance market and the need to retain customers by increasing their satisfaction with the service. The purpose of the study is to evaluate the accuracy and performance of models capable of predicting the level of customer satisfaction with an insurance service based on data on customer interaction with the company.

Classification algorithms are used as methods. The accuracy and performance of the models were evaluated based on real data from surveys of insurance company clients. The best ensemble methods turned out to be random forest and gradient boosting, which demonstrated the accuracy of predicting satisfaction up to 85 %, significantly surpassing simpler models. It is shown that gradient boosting makes it possible to take into account non-linear dependences of factors, for example, the presence of escalation of treatment, and thereby more accurately identify "dissatisfied" customers. Currently, such forecasting in insurance companies is either not carried out, or relies heavily on random factors. This leads either to too frequent complaints, or to low customer satisfaction with their subsequent churn. The materials of the article are of practical value for insurance companies: the implementation of the developed models will allow them to quickly identify customers at risk of dissatisfaction and reasonably apply preventive measures, for example, additional service measures or compensation to increase their satisfaction.

Keywords: customer satisfaction, insurance company, machine learning, prediction, gradient boosting, model accuracy.

For citation: Madiyarov K.G. Evaluation of accuracy and performance of machine learning models for prognosis customer satisfaction in insurance companies. *Modeling, Optimization and Information Technology*. 2025;13(3). (In Russ.). URL: https://moitvivt.ru/ru/journal/pdf?id=2041 DOI: 10.26102/2310-6018/2025.50.3.043

Введение

Высокий уровень удовлетворенности клиентов является ключевым фактором успеха в страховом бизнесе. На сегодняшний день проблемы прогнозирования удовлетворенности клиентов в страховании изучена недостаточно. Так, авторы [1] предложили гибридный подход на стыке интеллектуального анализа процессов и машинного обучения для предсказания исходов клиентского опыта в онлайн-страховании. Их модель продемонстрировала высокую точность – порядка 99 % при прогнозировании уровня удовлетворенности клиентов. Подобных наилучших показателей авторам удалось достичь, комбинируя методы интеллектуального анализа процессов и машинного обучения, им также удалось предсказать исходы клиентского опыта в онлайн-страховании с точностью порядка 99 % [2]. В [3] была показана эффективность анализа эмоциональной окраски речи для оценки удовлетворенности: интеграция технологий распознавания эмоций в речи с алгоритмами машинного обучения позволила более точно определять удовлетворенность звонящих клиентов качеством обслуживания в колл-центрах. Схожую работу провел автор, используя алгоритмы градиентного бустинга при извлечении признаков и прогнозировании итоговой удовлетворенности с точностью до 80 % [4]. Также были проведены исследования для выявления нежелательного контента с помощью рекуррентных искусственных нейронных сетей [5]. В смежной задаче прогнозирования оттока клиентов ансамблевые методы градиентного бустинга зачастую обеспечивают более высокое качество классификации в задачах, связанных с поведением клиентов, чем традиционные статистические подходы [6].

Тем не менее, в контексте прогнозирования удовлетворенности клиентов страховой компании прямых сравнительных исследований моделей пока недостаточно. В банковском секторе, электронной коммерции и телекоммуникациях алгоритмы машинного обучения уже давно применяются для оценки оттока и удовлетворенности клиентов [7, 8]. Также, в [9] был проведен сравнительный анализ различных алгоритмов искусственного интеллекта для прогнозирования удовлетворенности клиентов. Однако страховые компании в большинстве случаев не все могут позволить себе иметь огромный штат инженеров, которые могут заниматься просто прогнозированием оттока или исследовать эмоции клиентов. Зачастую подобные функции ложатся на плечи

актуариев и рядовых андеррайтеров, которые должны заниматься не только актуарной практикой, но и работать с данными, которые носят либо слишком абстрактный характер, либо для обработки, которых уходит слишком много времени. Исследования подтверждают, что рост удовлетворенности клиентов способствует увеличению их удержания и финансовых показателей фирм [10]. Однако у каждой методики есть слабые стороны, как, например, обучение BERT-модели занимает много времени и требует множество вычислительных ресурсов, а точность достигает 80 % [11]. В [12] описан бюджетный вариант, который подходит для автоматической отработки жалоб клиентов. При наличии соответствующих разметок, возможно применение для прогнозирования удовлетворенности. Таким образом, для подобной бюджетной модели необходимо 24 часа при обучении на одной GPU (NVIDIA 100) и в среднем 150 долларов по текущим ценам для обучения одной модели. Это тоже немаловажный фактор, учитывать в начале все преимущества и недостатки машинного обучения путем постоянного замера технических характеристик моделей.

Целью данного исследования является сравнение моделей машинного обучения для прогнозирования удовлетворенности клиентов страховой компании по данным обслуживания. Для достижения данной цели был проведен сбор и анализ отраслевых данных, обучение ряда моделей классификации, оценка их точности и производительности, а также выбор лучшей модели и формулировка рекомендаций по применению.

Материалы и методы

Схема проведения экспериментального исследования и используемый инструментарий. Экспериментальное исследование включает анализ и подготовку исходных данных, формализацию задачи классификации, выбор и обучение моделей машинного обучения, а также оценку их качества. Входные данные для прогнозной модели представлены разнообразными характеристиками клиента и предоставленной ему услуги. Рассматриваются как социально демографические сведения (например, возраст, пол) и параметры страхового продукта (тип полиса, срок действия, размер страховой премии), так и показатели взаимодействия клиента с компанией: количество обращений в поддержку, наличие страховых случаев и скорость выплат, использование онлайн-сервисов и т. д. Дополнительно могут учитываться результаты послепродажных опросов или анализ текстов обращений клиентов. Предполагается, что эти признаки доступны в информационной системе страховой компании и могут быть агрегированы в единый вектор для каждого клиента. Для предварительного анализа влияния факторов был выполнен корреляционный анализ числовых признаков (Рисунок 1), показывающий взаимосвязи между основными показателями качества обслуживания.

Рисунок 1 иллюстрирует матрицу коэффициентов корреляции Пирсона между сервисными метриками и итоговой удовлетворенностью: например, заметны высокие корреляции между оценками решения проблемы, профессионализма, приветствия / прощания и финальной оценкой клиента. Это свидетельствует о значимости указанных характеристик для уровня удовлетворенности и позволяет выявить потенциальную мультиколлинеарность признаков, что важно для последующего моделирования. Таким образом, предварительный корреляционный анализ подтверждает ожидаемые связи и служит обоснованием выбора ключевых признаков, влияющих на удовлетворенность, что согласуется с целью исследования – построением точной модели прогнозирования.

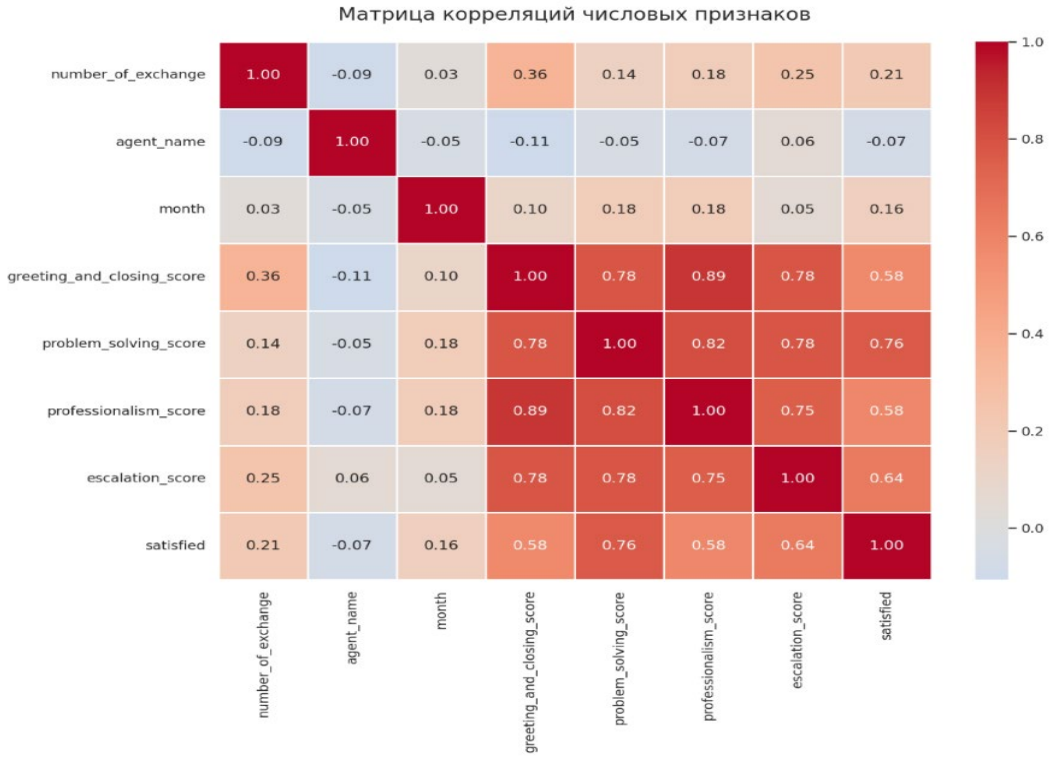

Рисунок 1 – Анализ корреляций числовых признаков
Figure 1 – Analysis of correlations of numerical features

Формализация постановки задачи. Формально задача прогнозирования удовлетворенности клиентов формулируется как проблема бинарной классификации. Пусть для каждого клиента i задан вектор признаков $X_i = \{x_{i1}, x_{i2}, ..., x_{in}\}$, описывающих его характеристики и опыт обслуживания и имеется соответствующая метка удовлетворенности $y_i \in \{0, 1\}$, отражающая факт неудовлетворенности (0) либо удовлетворенности (1) данного клиента. Требуется построить такую функцию $f(X_i) \rightarrow y_i$, которая на основании новых данных X_j смогла бы предсказать y_j, – степень удовлетворенности нового клиента j. Цель моделирования – максимизировать качество этой предсказательной функции (максимизировать точность классификации по выбранным метрикам). Важным ограничением является дисбаланс классов: как правило, доля удовлетворенных клиентов существенно превосходит долю неудовлетворенных. В нашем наборе данных положительный класс (satisfed = 1) составляет около 68 % случаев, отрицательный (satisfed = 0) – 32 %. Дисбаланс может приводить к смещению модели в сторону большинства, поэтому при обучении применялись меры балансировки (например, стратифицированное разбиение данных, настройка весов классов либо синтетическое дополнение выборки менее представленного класса) для корректного учета редких примеров. Кроме того, с практической точки зрения желательна интерпретируемость предсказаний модели для доверия бизнеса к системе, однако в данной работе основной упор делается на достижение высокой точности прогноза.

Подход к решению задачи основан на методах машинного обучения. В литературе предложен ряд различных подходов к прогнозированию удовлетворенности: от интеграции технологий интеллектуального анализа процессов с моделированием на данных страховых кейсов [1, 2] до анализа эмоциональной окраски речи звонков клиентов колл-центров [3]. В смежных исследованиях успешно применялись ансамблевые алгоритмы градиентного бустинга для близких задач оценки клиентского опыта [4, 6]. С учетом этого, в настоящей работе в качестве модели $f(X)$ рассматриваются

несколько алгоритмов классификации, включающих как базовые методы (логистическая регрессия), так и современные ансамблевые методы (случайный лес и различные реализации градиентного бустинга). Такое сочетание позволяет сравнить эффективность традиционных и более сложных алгоритмов на общей задаче.

Исходные данные и предобработка. В исследовании использованы реальные данные опросов клиентов службы поддержки крупной страховой компании[1]. Полученный датасет содержит 100 записей, каждая из которых соответствует отдельному диалогу (обращению клиента) и описывается набором признаков, характеризующих качество обслуживания. Признаковое пространство сформировано показателями, приведенными в Таблице 1.

Таблица 1 – Описание признаков
Table 1 – Features describe

№	Название признака	Описание
1.	number_of_exchange	количество обменов сообщениями между клиентом и оператором (длительность диалога)
2.	agent_name	идентификатор оператора, обработавшего обращение (категориальный признак)
3.	month	месяц обращения (категориальный временной признак)
4.	greeting_and_closing_score	интегральная оценка приветствия и прощания клиенту
5.	problem_solving_scor	оценка решения проблемы
6.	professionalism_score	оценка профессионализма оператора
7.	escalation_score	оценка необходимости эскалации обращения на вышестоящий уровень поддержки

Все указанные атрибуты качества сервиса представлены числовыми значениями по результатам клиентского анкетирования (например, по шкале от 1 до 10 баллов). Кроме перечисленных признаков, в данных присутствует итоговая оценка удовлетворенности сервисом по десятибалльной шкале, которая использована для формирования бинарной целевой переменной satisfed. Считается, что если клиент поставил высокую итоговую оценку (не ниже 8 из 10), то он удовлетворен обслуживанием (satisfed = 1); в противном случае (оценка 1–7) клиент условно считается неудовлетворенным (satisfed = 0).

Для подготовки данных к моделированию были выполнены стандартные процедуры. Категориальные признаки (имя агента, месяц обращения) преобразованы в числовую форму посредством порядкового кодирования (LabelEncoder). Числовые показатели (количество сообщений, различные оценки) предварительно нормализованы методом стандартизации (приведение к нулевому среднему и единичному стандартному отклонению) для обеспечения сопоставимости шкал. После очистки и трансформации данных исходная выборка объемом 100 наблюдений была случайным образом разделена на обучающую (60 записей), валидационную (20 записей) и тестовую (20 записей) подвыборки. Разбиение осуществлялось стратифицированно, то есть с сохранением исходной пропорции классов 68 % / 32 % в каждой части. Этапы предварительной обработки и разбиения данных проиллюстрированы на Рисунке 2.

[1] aibabyshark. insurance_customer_support_QA_result. Hugging Face. URL: https://huggingface.co/datasets/aibabyshark/insurance_customer_support_QA_result?library=datasets (дата обращения: 16.02.2025).

Рисунок 2 – Работа с данными (схема предобработки данных и разбиения выборки)
Figure 2 – Data analysis (data preprocessing and sample splitting scheme)

Построение и обучение моделей. Для прогнозирования удовлетворенности клиентов выбраны несколько алгоритмов машинного обучения, широко применяемых в задачах классификации. В качестве относительно простых базовых моделей использованы логистическая регрессия и многослойный перцептрон (полносвязная нейронная сеть). Также рассмотрены современные ансамблевые методы на основе деревьев решений: случайный лес и градиентный бустинг. В частности, реализованы три популярные версии бустинговых алгоритмов – XGBoost, LightGBM и CatBoost, которые показывают высокую эффективность на различных типах данных.

Гиперпараметры моделей подбирались эмпирически с учетом небольшого объема данных. Для всех ансамблевых алгоритмов число деревьев (итераций бустинга) установлено равным 200; скорость обучения (learning rate) для бустингов выбрана 0,1. Логистическая регрессия обучалась с L2-регуляризацией (коэффициент $C = 0{,}1$) для предотвращения переобучения.

Обучение всех моделей проводилось на тренировочной подвыборке, после чего качество классификаторов оценивалось на валидационной выборке, не участвовавшей в обучении. Помимо этого, была проведена перекрестная проверка на обучающей части: использована 5-кратная стратегия (5-fold cross-validation), по результатам которой вычислялось среднее значение F1-метрики для каждого алгоритма. Данный подход позволил убедиться в стабильности обучения моделей и снизить влияние случайности единичного разбиения данных на обучающую / проверочную выборки.

Критерии эффективности моделей прогнозирования. Для всестороннего сравнения качества моделей применялось несколько распространенных метрик классификации. Доля верно классифицированных случаев (Accuracy[2]) вычисляется как отношение числа правильных предсказаний ко всему количеству примеров:

[2] accuracy_score – scikit-learn 1.7.1 documentation. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html#sklearn.metrics.accuracy_score (дата обращения: 21.06.2025).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (1)$$

где *TP (True Positive)* – число верно предсказанных положительных случаев, *TN (True Negative)* – верно предсказанных отрицательных, *FP (False Positive)* – ложных тревог (неудовлетворенных клиентов, ошибочно классифицированных как довольные), а *FN* (False Negative) – пропущенных положительных случаев (удовлетворенных клиентов, отнесенных моделью к неудовлетворенным). Данная метрика отражает общую точность, однако может быть недостаточно информативной при сильном перекосе классов.

Точность позитивного прогноза (Precision[3]) определяется как доля действительно удовлетворенных клиентов среди всех, кого модель отнесла к категории удовлетворенных:

$$Precision = \frac{TP}{TP+FP}. \quad (2)$$

Этот показатель характеризует вероятность того, что «предсказанный довольный» клиент на самом деле удовлетворен сервисом. Низкое значение Precision указывает на большое число ложных срабатываний, когда модель ошибочно предсказывает удовлетворенность там, где в реальности клиент недоволен.

Полнота обнаружения положительного класса (Recall, или чувствительность) вычисляется как доля удовлетворенных клиентов, которых модель сумела правильно идентифицировать:

$$Recall = \frac{TP}{TP+FN}. \quad (3)$$

Иными словами, Recall[3] показывает, насколько хорошо алгоритм покрывает все случаи положительного класса. Если Recall низкий, это означает, что модель пропускает значительную часть «довольных» клиентов (или, в терминах задачи, не выявляет многих неудовлетворенных клиентов, сигнализируя о проблемах не во всех необходимых случаях).

F1-мера[4] представляет собой гармоническое среднее Precision и Recall:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision \cdot Recall}. \quad (4)$$

Значение F1 будет высоким только если и точность, и полнота модели высоки одновременно. Таким образом, F1-мера позволяет объединить обе характеристики в одном показателе и удобно применять для оценки алгоритмов на несбалансированных данных, учитывая как ложноположительные, так и ложноотрицательные ошибки.

ROC AUC (Area Under Curve)[5] – это площадь под кривой ROC (Receiver Operating Characteristic), отражающей зависимость доли верных положительных срабатываний от доли ложных тревог при варьировании порога решения классификатора. Показатель AUC численно равен вероятности того, что случайно выбранный удовлетворенный клиент получит от модели более высокий прогнозируемый «рейтинг» (например, вероятность принадлежности к классу 1), чем случайно выбранный неудовлетворенный клиент. Иными словами, AUC характеризует способность алгоритма различать по вероятностным оценкам объекты разных классов и не зависит от выбранного порога классификации.

[3] precision_recall_curve – scikit-learn 1.7.1 documentation. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_curve.html#sklearn.metrics.precision_recall_curve (дата обращения: 21.06.2025).
[4] f1_score – scikit-learn 1.7.1 documentation. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html (дата обращения: 21.06.2025).
[5] roc_auc_score – scikit-learn 1.7.1 documentation. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html (дата обращения: 21.06.2025).

Наконец, интегральным показателем качества бинарной классификации является коэффициент корреляции Маттеуса (MCC[6]). Он вычисляется по формуле:

$$MCC = \frac{TP \times TN - FP \times FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \quad (5)$$

Значение *MCC* интерпретируется как корреляция между предсказаниями модели и истинными классами (принимает значения от –1 до +1). В отличие от простой точности, данный коэффициент учитывает сбалансированность классификации, принимая во внимание все четыре исхода (*TP*, *TN*, *FP*, *FN*) одновременно. Благодаря этому *MCC* считается надежным критерием даже на несбалансированных выборках: высокое значение *MCC* указывает, что модель одинаково хорошо справляется с предсказанием как положительного, так и отрицательного классов, минимизируя перекос в сторону одного из них.

Вычислительная среда и замеры времени. Все эксперименты выполнялись на вычислительной системе с многоядерным CPU (44 ядра) и 173 ГБ оперативной памяти; тем не менее, благодаря небольшому объему данных обучение моделей занимало доли секунды. Для каждой модели замерялось фактическое время обучения (в секундах) – от запуска процедуры fit до получения готового классификатора. Результаты по длительности обучения приведены в составе сравнительной Таблицы 2 с основными метриками качества. В частности, алгоритмы LightGBM и логистическая регрессия оказались самыми быстродействующими (обучение ~0,04–0,05 с); немногим больше потребовалось для XGBoost (~0,05 с). В то же время Random Forest и CatBoost обучались дольше (порядка 0,3 с каждый на том же объеме данных). Эта разница обусловлена различной вычислительной сложностью алгоритмов: случайный лес строит сотни независимых деревьев, что увеличивает затраты времени при большом их числе, CatBoost выполняет трудоемкую процедуру кодирования категориальных признаков и последовательное обучение деревьев, тогда как LightGBM оптимизирован по скорости за счет специального алгоритма разбиения. В результате модель XGBoost можно рассматривать как оптимальный компромисс – она обеспечивает высокое качество прогноза при сравнительно малом времени обучения. Отметим, что даже самая «медленная» модель (CatBoost) обучилась менее чем за секунду ввиду небольшого размера выборки, поэтому вычислительная сложность не являлась ограничивающим фактором в данном исследовании.

Результаты

Экспериментальные данные представлены в обработанном виде в Таблице 2 и на Рисунках 3, 4. На основании полученных результатов можно сделать несколько выводов. Как видно из Таблицы 2, ансамблевые модели на основе деревьев решений продемонстрировали наивысшую точность прогнозирования удовлетворенности на валидационной выборке: Random Forest, XGBoost и CatBoost достигли Accuracy порядка 0,95, тогда как LightGBM и логистическая регрессия – лишь около 0,85 и 0,75 соответственно. На отложенном тестовом наборе лучшие модели сохранили высокий результат (точность около 85 %), превосходя логистическую регрессию.

[6] matthews_corrcoef – scikit-learn 1.7.1 documentation. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html (дата обращения: 21.06.2025).

Таблица 2 – Метрики качества моделей машинного обучения на задаче прогнозирования удовлетворенности клиентов (валидационная выборка, 20 случаев)
Table 2 – Quality metrics of machine learning models on the task of predicting customer satisfaction (validation sample, 20 cases)

Модель	Accu-racy	Preci-sion	Re-call	F1	AUC	MCC	Время обучения, с	F1 (CV)
Logistic Regression	0,7500	0,8000	0,8571	0,8276	0,8929	0,3780	0,0490	0,9165
Random Forest	0,9500	1,0000	0,9286	0,9630	1,0000	0,8921	0,3298	0,9631
XGBoost	0,9500	0,9333	1,0000	0,9655	0,9881	0,8819	0,0520	0,9644
LightGBM	0,8500	1,0000	0,7857	0,8800	0,9881	0,7237	0,0390	0,9025
CatBoost	0,9500	0,9333	1,0000	0,9655	0,9286	0,8819	0,2973	0,9644

Все три алгоритма (Random Forest, XGBoost, CatBoost) смогли без ошибок выявить всех неудовлетворенных клиентов на валидации: Recall по отрицательному классу составил 1,0, т. е. ни один случай неудовлетворенности не остался непризнанным (см. матрицу ошибок на Рисунке 3). Для сравнения, модели LightGBM и логистическая регрессия допускали пропуски негативных случаев (Recall = 0,79 и 0,86 соответственно). Precision положительного класса достиг 1,0 у Random Forest и LightGBM (ни одного ложноположительного прогноза неудовлетворенности), а у XGBoost и CatBoost был ~0,93, что означает лишь незначительную долю ошибочно отнесенных к «неудовлетворенным» клиентам. Благодаря сочетанию максимального Recall и высокого Precision, суммарные F1-меры у XGBoost и CatBoost достигли ~0,966, несколько превысив показатель Random Forest (~0,963) и превышая значения LightGBM (~0,880) и логистической регрессии (~0,828), согласно Рисунку 4.

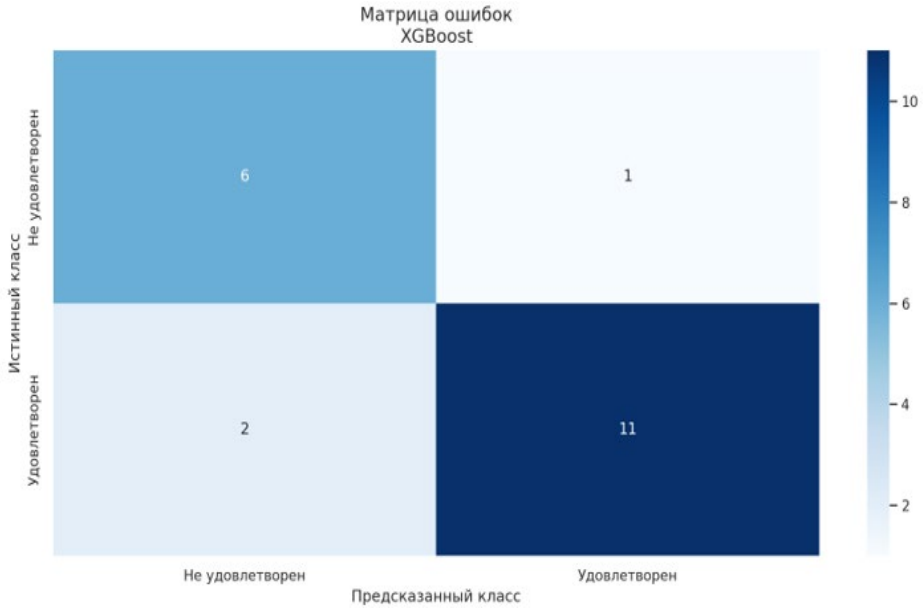

Рисунок 3 – Матрица ошибок XGBoost
Figure 3 – XGBoost confusion matrix

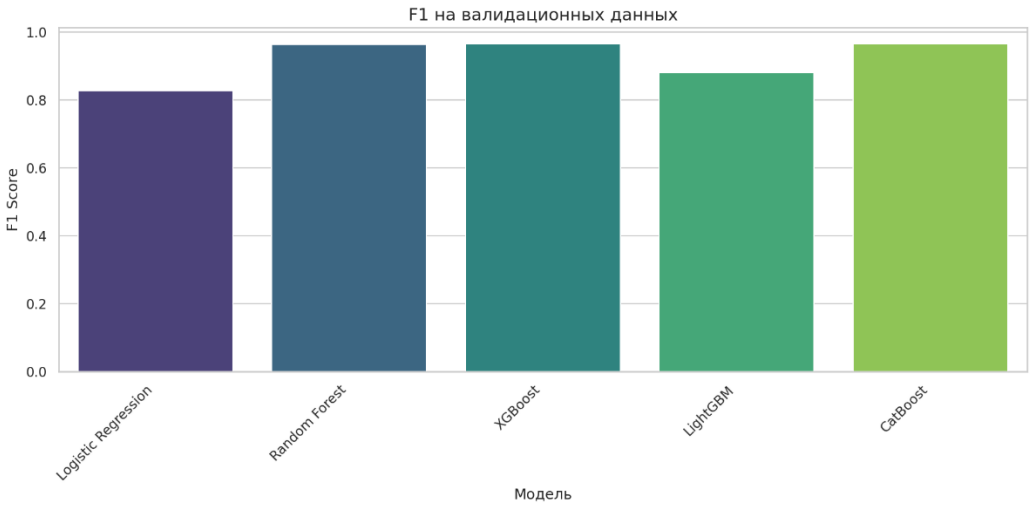

Рисунок 4 – F1-метрики моделей
Figure 4 – F1-metrics of models

Помимо этого, зафиксированы различия в показателях площади под ROC-кривой. Модель Random Forest достигла AUC = 1,0 на валидационных данных, XGBoost и LightGBM – около 0,99, CatBoost – ~0,93, тогда как у логистической регрессии AUC составил ~0,89. Интегральная метрика сбалансированной точности MCC (коэффициент Маттьюса) у ансамблевых моделей также высока (~0,88–0,89), что существенно превышает MCC логистической регрессии (~0,38). Среднее качество по кросс-валидации (F1 (CV) в Таблице 2) для трех лучших моделей оказалось примерно одинаковым (~0,96), подтверждая стабильность их обучения. Примечательно, что для логистической регрессии средняя F1 по кросс-валидации (0,917) заметно превышала разовый показатель на валидации (0,828); тем не менее на тестовых данных эта модель все равно показала более слабый результат по сравнению с ансамблями.

Анализ вычислительной производительности обучения моделей (Таблица 2) показал, что самым быстрым алгоритмом является LightGBM: обучение заняло порядка 0,04 с. Логистическая регрессия и XGBoost также обучились очень быстро (~0,05 с), тогда как Random Forest и CatBoost потребовали на порядок больше времени (≈0,33 с и 0,30 с соответственно) на том же объеме данных. Таким образом, среди моделей с наивысшим качеством XGBoost отличился минимальным временем обучения (~0,05 с при F1≈0,96 и AUC≈0,99), продемонстрировав выгодное сочетание эффективности и точности прогноза.

Обсуждение

Анализ факторов, влияющих на прогноз удовлетворенности. Результаты экспериментов показывают, что выбор алгоритма влияет на точность прогноза удовлетворенности. Ансамблевые модели на основе деревьев решений (Random Forest, градиентный бустинг) превзошли линейную модель (логистическую регрессию) по качеству классификации. Вероятная причина заключается в способности деревьев решений учитывать нелинейные зависимости и сложные взаимодействия между признаками. Например, модель XGBoost фиксирует резкое падение удовлетворенности клиента при эскалации обращения на вышестоящий уровень поддержки, тогда как линейная логистическая регрессия не может напрямую уловить такой эффект, что ведет к снижению ее точности.

Для выявления наиболее значимых факторов, влияющих на предсказание удовлетворенности, была проведена интерпретация лучшей модели (XGBoost) с

помощью метода SHAP (Shapley Additive Explanations)[7]. Выбор SHAP обусловлен его строгой теоретической основой и удобством для сложных моделей: этот метод опирается на концепцию значений Шепли из теории кооперативных игр и позволяет количественно оценить вклад каждого признака в итоговый прогноз. Иными словами, SHAP рассчитывает, как изменяется предсказанная вероятность удовлетворенности при учете того или иного фактора, распределяя влияние между признаками справедливым образом. На Рисунке 5 представлен результат SHAP-анализа для модели XGBoost, на котором визуализирована относительная значимость всех использованных признаков.

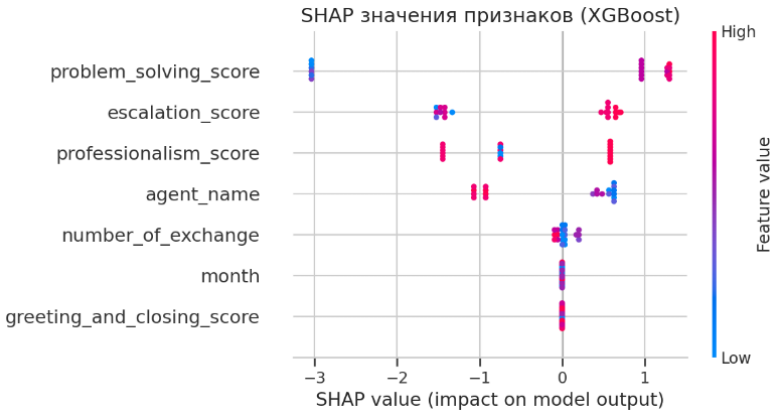

Рисунок 5 – SHAP-анализ признаков (XGBoost)
Figure 5 – SHAP feature analysis (XGBoost)

По результатам SHAP-анализа ключевым фактором риска неудовлетворенности оказался признак «эскалация обращения». Если решение проблемы требовало эскалации на более высокий уровень обслуживания, вероятность высокой удовлетворенности клиента резко снижалась (что согласуется с наблюдением в работе [8]). Иными словами, даже единичный факт эскалации служит сильным индикатором потенциальной неудовлетворенности клиента.

Помимо эскалации, существенный вклад в прогноз вносят оценки качества решения проблемы и профессионализма агента. Высокие баллы по этим параметрам повышают вероятность того, что клиент останется доволен, тогда как низкие оценки, напротив, ведут к негативному прогнозу.

Например, если клиент невысоко оценил качество решения своего вопроса, модель в большинстве таких случаев относит его к неудовлетворенным, даже при благоприятных прочих показателях. Также выявлено, что увеличенное число обменов сообщениями (длительный диалог с поддержкой) коррелирует с уменьшением удовлетворенности: затянутый процесс общения снижает вероятность положительной оценки со стороны клиента.

В то же время некоторые из рассматриваемых признаков оказались наименее значимыми для модели. Согласно SHAP-анализу, ни конкретная личность обслуживающего агента, ни месяц обращения практически не влияют на итоговый прогноз удовлетворенности. Это означает, что градиентный бустинг успешно обобщает паттерны удовлетворенности независимо от индивидуальных характеристик сотрудника или сезонных факторов – модель выявляет общие закономерности, существенные для формирования удовлетворенности, не привязываясь к отдельным персоналиям или времени года.

[7] Welcome to the SHAP documentation – SHAP latest documentation. URL: https://shap.readthedocs.io/en/latest/ (дата обращения: 11.05.2025).

Заключение

В рамках данной работы выполнена разработка и оценка нескольких моделей машинного обучения для прогноза удовлетворенности клиентов страховой компании. На реальных данных показано, что современные методы МО эффективно решают задачу классификации «удовлетворен / не удовлетворен» по метрикам обслуживания. Ансамблевые модели на основе деревьев решений (особенно градиентный бустинг) продемонстрировали наивысшую точность прогнозирования (~85 % на тесте), существенно превзойдя базовые подходы – логистическую регрессию и простой перцептрон. Лучшая модель (XGBoost) достигла F1-меры ~0.96 и безошибочно выявляла всех неудовлетворенных клиентов, учитывая ключевые факторы риска (такие как эскалация проблемы), при минимальном времени обучения. Результаты подтверждают, что внедрение подобных алгоритмов может принести практическую пользу страховым компаниям, позволяя им заблаговременно обнаруживать недовольство клиентов и принимать меры для его снижения.

Таким образом, цели исследования достигнуты: выполнена оценка точности и производительности моделей машинного обучения, и выбрана оптимальная модель. Перспективой дальнейших исследований является расширение набора данных, применение текстовых анализаторов для учета мнений клиентов и исследование моделей для прогнозирования оценок клиентов, а также интеграция предложенной модели в систему поддержки принятия решений страховой компании. Полученные результаты могут быть полезны не только в страховании, но и в других сервисных отраслях, где важно отслеживать удовлетворенность клиентов и оперативно реагировать на ее снижение.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Akhavan F., Hassannayebi E. A Hybrid Machine Learning with Process Analytics for Predicting Customer Experience in Online Insurance Services Industry. *Decision Analytics Journal*. 2024;11. https://doi.org/10.1016/j.dajour.2024.100452
2. Berrada Chakour O., Ettaoufik A., Aissaoui Kh., Maizate A. Artificial Intelligence Algorithms to Predict Customer Satisfaction: A Comparative Study. *IAES International Journal of Artificial Intelligence.* 2025;14(2):1654–1662. https://doi.org/10.11591/ijai.v14.i2.pp1654-1662
3. Рудакова П.А., Семенов Т.А., Сычугов А.А., Котов В.В. Определение уровня удовлетворенности клиента в call-центрах. *Известия Тульского государственного университета. Технические науки.* 2024;(10):346–352.
   Rudakova P.A., Semenov T.A., Sychugov A.A., Kotov V.V. Determining the Level of Customer Satisfaction Based on Machine Learning in Call-Centers. *News of the Tula State University. Technical Sciences.* 2024;(10):346–352. (In Russ.).
4. Airlangga G. Comparative Study of XGBoost, Random Forest, and Logistic Regression Models for Predicting Customer Interest in Vehicle Insurance. *Sinkron: Jurnal Dan Penelitian Teknik Informatika.* 2024;8(4):2542–2549. https://doi.org/10.33395/sinkron.v8i4.14194
5. Никифоров А.А. Разработка модуля распознавания эмоций разговора колл-центра с использованием рекуррентных искусственных нейронных сетей, для выявления нежелательного контента. *Вестник науки*. 2023;3(7):226–232.
   Nikiforov A.A. Development of a Call Center Conversation Emotion Recognition Module Using Recurrent Artificial Neural Networks to Identify Unwanted Content. *Science Bulletin.* 2023;3(7):226–232. (In Russ.).

6. Boozary P., Sheykhan S., GhorbanTanhaei H., Magazzino C. Enhancing Customer Retention with Machine Learning: A Comparative Analysis of Ensemble Models for Accurate Churn Prediction. *International Journal of Information Management Data Insights*. 2025;5(1). https://doi.org/10.1016/j.jjimei.2025.100331
7. Hossain M.Sh., Rahman M.F. Customer Sentiment Analysis and Prediction of Insurance Products' Reviews Using Machine Learning Approaches. *FIIB Business Review*. 2022;12(4):386–402. https://doi.org/10.1177/23197145221115793
8. Wang Z., Abolarin E., Wu K., et al. Beyond Charging Anxiety: An Explainable Approach to Understanding User Preferences of EV Charging Stations Using Review Data. [Preprint]. arXiv. URL: https://arxiv.org/abs/2507.03243 [Accessed 21st June 2025].
9. Wangkiat P., Polprasert Ch. Machine Learning Approach to Predict E-Commerce Customer Satisfaction Score. In: *Proceedings of the 2023 8th International Conference on Business and Industrial Research (ICBIR), 18–19 May 2023, Bangkok, Thailand.* IEEE; 2023. P. 1176–1181. https://doi.org/10.1109/ICBIR57571.2023.10147542
10. Zaghloul M., Barakat Sh., Rezk A. Predicting E-Commerce Customer Satisfaction: Traditional Machine Learning vs. Deep Learning Approaches. *Journal of Retailing and Consumer Services*. 2024;79. https://doi.org/10.1016/j.jretconser.2024.103865
11. Le H.-S., Huynh Do Th.-V., Nguyen M.H., et al. Predictive Model for Customer Satisfaction Analytics in E-Commerce Sector Using Machine Learning and Deep Learning. *International Journal of Information Management Data Insights*. 2024;4(2). https://doi.org/10.1016/j.jjimei.2024.100295
12. Izsak P., Berchansky M., Levy O. How to Train BERT with an Academic Budget. arXiv. URL: https://arxiv.org/abs/2104.07705 [Accessed 21st June 2025].

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Мадияров Куан Галымович, аспирант, Новосибирский государственный университет экономики и управления «НИНХ», Новосибирск, Российская Федерация.
e-mail: kuan.mad@mail.ru
ORCID: 0009-0002-1447-4357

Kuan G. Madiyarov, Postgraduate, Novosibirsk State University of Economics and Management, Novosibirsk, the Russian Federation.

Статья поступила в редакцию 02.08.2025; одобрена после рецензирования 28.08.2025; принята к публикации 10.09.2025.

The article was submitted 02.08.2025; approved after reviewing 28.08.2025; accepted for publication 10.09.2025.