

DOI: 10.26794/2587-5671-2026-30-2-108-120
УДК 336.63(045)
JEL G32

Прогнозирование рисков рынка транспортного страхования в России с использованием машинного обучения

И.В. Филимонова¹, А.В. Комарова², А.П. Саматова³

^{1,2} Институт экономики и организации промышленного производства Сибирского отделения Российской академии наук, Новосибирск, Российская Федерация;

^{1,2,3} Новосибирский национальный исследовательский государственный университет, Новосибирск, Российская Федерация

АННОТАЦИЯ

Цель исследования – разработка и апробация модели прогнозирования возможных убытков по полису ОСАГО с использованием методов машинного обучения. **Актуальность** темы обусловлена значимостью роли, которую играет рынок страхования и, в частности, транспортного страхования в развитии российской экономики. Актуальность исследования подкрепляется высоким коэффициентом убыточности для данного вида страхования и необходимостью оптимизации параметров рынка. **Объектом исследования** является российский рынок транспортного страхования. **Предмет исследования** – методы машинного обучения, позволяющие прогнозировать убытки обязательного страхования автогражданской ответственности (ОСАГО) в зависимости от входных параметров, характеризующих данные о страхователе и его транспортном средстве. В работе проведено сравнение эффективности ансамблевых методов машинного обучения с традиционным методом обобщенной линейной модели при прогнозировании убытка ОСАГО. Используется разделение способов возмещения убытка ОСАГО на традиционный и прямой. Показано, что эффективность применения бустинговых моделей машинного обучения выше, чем при прогнозировании с использованием традиционных методов и моделей бэггинга. Выделены значимые факторы, влияющие на частоту и серьезность страхового случая: количество средней тяжести ДТП на транспортном средстве, количество водителей, коэффициент «бонус-малус»; минимальный стаж водителя в полисе, мощность транспортного средства, тип населенного пункта. Сделаны **выводы** о том, что на рост прогнозируемого убытка положительное влияние оказывает количество ДТП, количество водителей в полисе, низкий стаж и низкий возраст водителей. Выделяется влияние гендерного состава водителей: рост количества женщин-водителей приводит к снижению страховых рисков по полису ОСАГО.

Ключевые слова: ОСАГО; транспортное страхование; прогнозирование убытков; машинное обучение; риски

Для цитирования: Филимонова И.В., Комарова А.В., Саматова А.П. Прогнозирование рисков рынка транспортного страхования в России с использованием машинного обучения. *Финансы: теория и практика*. 2026;30(2):108-120. DOI: 10.26794/2587-5671-2026-30-2-108-120

Forecasting Risks of the Transport Insurance Market in Russia Using Machine Learning

I.V. Filimonova¹, A.V. Komarova², A.P. Samatova³

^{1,2} Institute of Economics and Industrial Engineering, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation;

^{1,2,3} Novosibirsk National Research State University, Novosibirsk, Russian Federation

ABSTRACT

The purpose of the study is to develop and apply a model for predicting potential losses under a CTP policy using machine learning techniques. **The relevance** of the topic is due to the importance of the role played by the insurance market, and, in particular, transport insurance in the development of the Russian economy. This relevance is supported by the high loss ratio for this type of insurance, as well as the need to optimize market conditions. **The object** of this research is the Russian insurance sector. **The subject** of the study is machine learning techniques that allow predicting CTP losses based on input parameters characterizing data about the insured and their vehicle. The paper compares the effectiveness of ensemble machine learning methods with the traditional generalized linear method of predicting compulsory motor insurance losses. The study divides CTP losses into traditional and direct categories. It has been shown that the effectiveness of applying boosting machine learning models for forecasting is higher than using Random

Forest and GLM. Factors that significantly affect the frequency and severity of insured events include: the number of minor accidents on a vehicle, the number of drivers, and the CBM coefficient. Other factors include the minimum length of service for a driver in a policy, the power of a vehicle, and the type of location where the vehicle is used. **It has been concluded** that the growth of the projected loss is positively influenced by the number of accidents, the number of drivers in the policy, low seniority and low age of drivers. The influence of the gender composition of drivers is highlighted: an increase in the number of female drivers leads to a decrease in insurance risks under the CTP policy.

Keywords: CTP; transport insurance; loss forecasting; machine learning; risks

For citation: Filimonova I.V., Komarova A.V., Samatova A.P. Forecasting insurance risks of the transport insurance market in Russia using machine learning. *Finance: Theory and Practice*. 2026;30(2):108-120. (In Russ.). DOI: 10.26794/2587-5671-2026-30-2-108-120

ВВЕДЕНИЕ

Страхование рисков играет значимую роль в экономике и жизни людей. Его важность обуславливается влиянием на стабильность экономического роста, снижение социальной напряженности, повышение инвестиционной активности и др. [1–3]. В работах отечественных авторов согласуются предположения и выводы о значимости страхования как стратегического сектора экономики [4].

Сектор автотранспортного страхования занимает значительную часть страхового рынка [5]. В исследовании С. И. Рыбакова отмечается важность транспортного страхования, которая обуславливается глубиной и плотностью проникновения транспорта в каждый общественный аспект и личную жизнь населения [6].

Проблемы и перспективы развития автотранспортного рынка в России активно обсуждаются в литературе. Во-первых, имеет место несбалансированность страхового рынка. Проблема затрагивает отраслевой и региональный аспекты. Непропорциональное развитие обосновывается преобладанием стандартных и статистически просчитанных видов страхования рисков, территориальной диспропорцией, несбалансированностью структуры клиентов. Также можно выделить превалирование обязательного страхования и низкий спрос на добровольные страховые продукты, особенно в разрезе регионов [7]. Так, например, в 2021 г. на один договор КАСКО приходилось 8 договоров ОСАГО [8].

Во-вторых, на страховом рынке РФ существует тенденция к завышению стоимости добровольного страхования [9]. В контексте автострахования завышение стоимости на добровольные страховые продукты связано с компенсацией убыточности обязательного страхования автогражданской ответственности.

В статьях отечественных авторов обсуждается проблема неэффективности сбытового сегмента транспортного страхования. Значительная часть продаж страховых продуктов идет через посредников — агентскую и банковскую сеть. Авторы показывают, что посреднические продажи плохо регулируются, и наращивание объемов прямых

продаж могло бы повысить эффективность страховых компаний [10, 11].

В российских публикациях, посвященных страхованию, встречаются новые работы, затрагивающие санкционное давление, важность транспортного страхования в условиях изменения цепочек сбыта и весомую роль транспортного страхования на национальном страховом рынке [5].

Другими важными аспектами для рынка автострахования являются: прогнозирование наступления страховых случаев и выплат по ним, финансовые показатели, а также выявление факторов, на них влияющих. Определение факторов, оказывающих значимое влияние на убытки, является ключевым моментом в целом ряде работ [12, 13]. Часто обнаруживают новый фактор, влияющий на результат, чему посвящают отдельные исследования [14].

Для данного сегмента рынка применяется широкий спектр экономико-математических инструментов, таких как: имитационное моделирование [15], линейные регрессионные модели [16, 17], логистические регрессии [18], а также алгоритмы кластеризации [19].

В данном исследовании строится модель прогнозирования убытков с использованием методов машинного обучения. Основной областью применения методов машинного обучения в экономике являются фондовые рынки и банковский сектор. Существует обширное количество работ, посвященных сравнению машинного обучения и других моделей для прогнозирования доходности акций, ВВП и различных показателей. Также машинное обучение широко используется в банковских скорингах, в частности, для оценки заемщика по кредиту [20]. В работах, посвященных соотношению цены-прибыли, машинное обучение тоже показывает хорошие результаты [21].

Для рынка автострахования проводились сравнения в эффективности различных моделей. В большинстве случаев точность моделей машинного обучения оказывается выше, чем у классических эконометрических моделей [22, 23], однако точность и надежность конкретных алгоритмов

прогнозирования варьируется для разных стран и компаний.

Вклад в сложность прогнозирования страховых выплат вносят факторы, связанные с геополитической нестабильностью, поэтому в целях повышения устойчивости рынка страхования в целом необходимо сформировать методы, способные повысить точность предсказания в современных реалиях. Таким образом, целью исследования является разработка и апробация модели для предсказания убытков по полису ОСАГО с помощью методов машинного обучения. Также в исследовании выявляются и анализируются основные влияющие факторы.

МАТЕРИАЛЫ И МЕТОДЫ

На рынке обязательного страхования автогражданской ответственности (ОСАГО) существуют два способа возмещения убытков.

В рамках первого способа пострадавшая сторона обращается к страховой компании виновной стороны за возмещением ущерба, возникшего при ДТП. Этот способ можно назвать традиционным возмещением убытков (ТВУ), так как именно он был первым адаптирован на рынке обязательного автострахования в России в 2009 г.

На рынке обязательного страхования есть возможность перестрахования. Страховые компании могут передать клиента в страховой пул, чтобы распределить возможные убытки между его участниками. Этот процесс включает перераспределение премии для каждого клиента в зависимости от доли компании на рынке. Перестрахование убыточных клиентов снижает риски, так как при ущербе страховая компания несет убытки пропорционально своей доле. Это видно на *рис. 1*.

С 2014 г. в России в дополнение к ТВУ действует прямое возмещение убытков (ПВУ). Пострадавшая сторона вправе обратиться к собственной страховой компании за возмещением ущерба, если выполняется ряд условий. Затем страховая компания пострадавшего обращается к страховщику виновника за компенсацией через Российский союз автостраховщиков.

Более детальная схема формирования прямого убытка для страховой компании представлена на *рис. 2*.

ПВУ возникло как ответ на некоторые недочеты ТВУ и считается более простым и быстрым способом получения страховых выплат для потерпевшего.

В связи с наличием большого количества структурированных наблюдений и независимых влияющих факторов рынок автострахования является хорошим объектом для применения регрессионных моделей и последующего прогнозирования.

В исследовании для сравнения используются классические методы прогнозирования — обо-

бщенные линейные модели (GLM) и модели машинного обучения (Random Forest, XGBoost, CatBoost).

Распространенным методом оценки риска по вероятным страховым случаям для разных страховых продуктов выступали модели GLM, сочетающие хорошую интерпретируемость и гибкость, благодаря функции связи и целевого распределения. Модель позволяет учитывать гетероскедастичность ошибок.

Random Forest (метод случайного леса) является классическим примером ансамблевого метода машинного обучения — бэггинга. Этот подход был предложен Лео Брейманом [24] и Адель Катлер. Принцип работы состоит в разбиении дерева решения на определенное число деревьев и усреднении результатов каждого из деревьев для получения итогового решения. Отдельный результат деревьев обладает низкой точностью прогнозирования, но использование всех результатов позволяет нивелировать ошибки каждого дерева.

XGBoost является эффективной реализацией градиентного бустинга над решающими деревьями. Так же, как и в градиентном бустинге, в модели есть функция потерь, которая принимает в качестве входящего параметра предсказания модели и эталонный ответ. Выводимое значение от функции потерь необходимо минимизировать. Функция потерь должна быть дифференцируема по предсказаниям.

$$\text{sum}_{\text{loss}(y, \hat{y})} = \sum_{i=1}^n \text{loss}(y_i, \hat{y}_i), \quad (1)$$

где y_i — фактическое значение, \hat{y}_i — предсказанное значение.

Модель CatBoost также является более эффективным способом реализации классического градиентного метода над решающими деревьями [25, 26]. Особенности данного метода являются:

1. Использование решающих таблиц.
2. Наличие target-кодирования для категориальных переменных.
3. Упорядоченное обучение примеров. Упорядоченный бустинг позволяет решить проблемы смещенности оценок.

Оценивание качества моделей в исследовании осуществляется с помощью набора метрик, используемых для оценивания качества прогноза: R^2 (коэффициент детерминации), MAE (средняя абсолютная ошибка), RMSE (стандартное отклонение). Способы расчета и пояснения к метрикам представлены в *табл. 1*.

На второй стадии исследования проводилось выявление значимости отдельных факторов для повышения точности прогнозирования. Был ис-

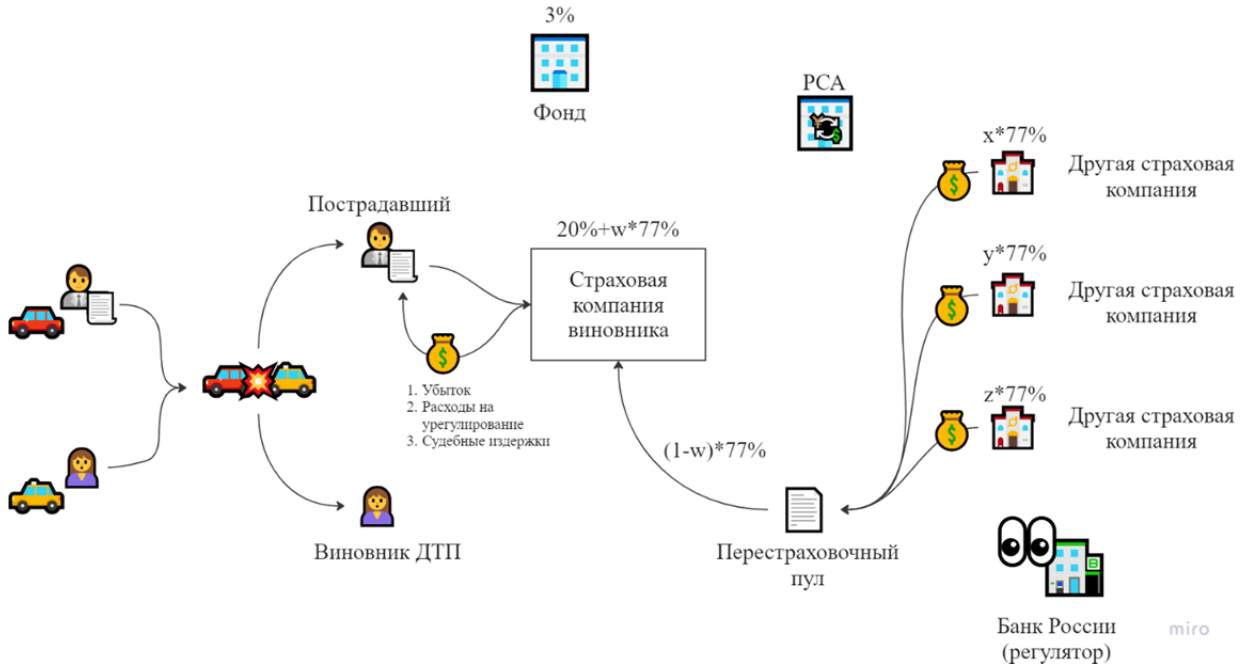


Рис. 1 / Fig. 1. Схема возникновения ТВУ с перестрахованием / The Scheme of Occurrence of TCL Reinsurance

Источник / Source: составлено авторами / Compiled by the authors.

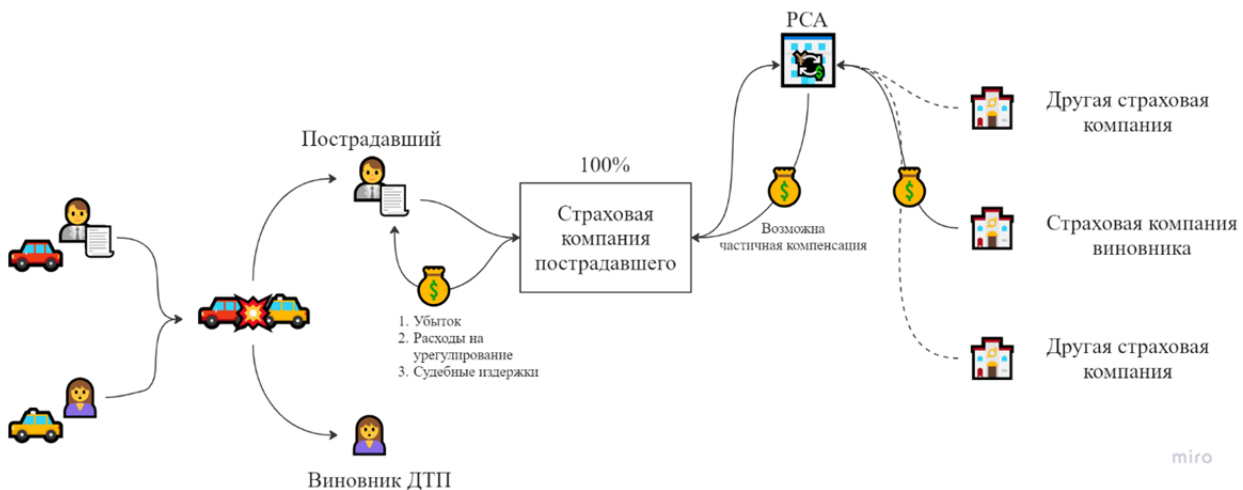


Рис. 2 / Fig. 2. Схема возникновения ПВУ без перестрахования / The Scheme of Occurrence of DCLs Without Reinsurance

Источник / Source: составлено авторами / Compiled by the authors.

пользован подход «аддитивное объяснение Шепли» (shape value, SHAP). Данный метод является наиболее популярным для объяснения работы моделей машинного обучения, он показывает влияние факторов с точки зрения положительного или отрицательного влияния на прогнозируемую переменную.

В данном исследовании проводится оценка риска по ОСАГО на примере крупной российской страховой компании. Также в работе учтены особенности российского рынка страхования, а именно — разделение на ТВУ и ПВУ, что позволит более комплексно исследовать рассматриваемую тему.

Информационную базу исследования составили данные Росстата, Банка России, рассматриваемой страховой компании, Министерства транспорта и ГИБДД. Размер базы данных, используемой в исследовании, составляет 773 тысячи строк. Для анализа используются данные о клиенте и его транспортном средстве за 2019–2023 гг. Для получения корректных результатов из базы данных были удалены полисы, принадлежащие юридическим лицам, а также полисы, при использовании которых подозревалось злоупотребление или иные мошеннические действия. Дополнительно из базы данных были

Метрики, применяемые в исследовании / Metrics Used in the Study

Метрика / Metric	Формула / Formula	Пояснение / Explanation
Коэффициент детерминации, R^2	$R^2 = \frac{S_q^2}{S^2} = 1 - \frac{S_e^2}{S^2}$	Показывает, какую долю в общей дисперсии занимает объясненная дисперсия
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - y_i^c)^2}{N}}$	Корень из среднеквадратической ошибки. Применяют, чтобы метрика имела размерность исходных данных
Средняя абсолютная ошибка, MAE	$MAE = \frac{1}{N} \sum_{i=1}^N y_i - y_i^c $	Менее чувствительна к выбросам по сравнению с RMSE

Источник / Source: составлено авторами / Compiled by the authors.

Примечание / Note: S_q^2 – объясненная дисперсия; S^2 – полная дисперсия; S_e^2 – остаточная дисперсия; y_i – фактическое i -е наблюдение; y_i^c – предсказанное моделью i -е наблюдение; N – число наблюдений.

удалены «выбросы», которые могут внести искажения в расчеты — полисы с высокими убытками.

Зависимой или целевой переменной в работе выступают частота и серьезность страхового случая. Предсказание частоты страхового случая является аналогом вероятности попадания в ДТП. Однако в отличие от вероятности, принимающей значения от 0 до 1, частота не ограничена сверху, т.е. может принимать значения от 0 до бесконечности. Серьезность претензии — среднее значение ущерба в одном ДТП. Для получения итогового результата вероятностного ущерба по страховому полису предсказанные значения перемножаются. Такое разделение позволяет более точно определить риски по страховым полисам.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ рынка транспортного страхования

Динамика сборов премий в сегменте страхования ином, чем страхование жизни (non-life страхования) за последние семь лет (рис. 3), показывает, что транспортное страхование, включающее ОСАГО и добровольное страхование автотранспорта (КАСКО), занимает стабильно высокую долю. Сборы премий транспортного страхования за рассматриваемый период не опускались ниже 30%, а доля ОСАГО занимает пятую часть рынка non-life страхования.

Однако транспортное страхование сохраняет и стабильно высокую долю (на уровне 50%) в структуре выплат по страховым случаям в non-life сегменте (рис. 4). Это свидетельствует о большом влиянии, которое риски по транспортному стра-

хованию оказывают на результаты рынка non-life страхования в целом.

Таким образом, транспортное страхование занимает весомую долю на рынке страхования. Значительный вклад в сборы премий и убыточность вносит именно ОСАГО.

Результаты прогнозирования убытков рынка транспортного страхования

Основные метрики качества полученных с использованием рассматриваемых моделей прогнозов представлены в табл. 2. Можно заметить, что наименьшей надежностью прогноза обладает GLM, являющаяся классическим инструментом для прогнозирования на non-life рынке страховых продуктов. Так, коэффициент детерминации равен 0,624, что является наименьшим из полученных, а значение RMSE составляет 26 972 руб. — наибольшее из полученных. Аналогичная ситуация прослеживается для ПВУ: наиболее низкое значение коэффициента детерминации у GLM. Величина стандартного отклонения для ПВУ составляет 18 161 рубль, что является максимальным значением из всех относительных ошибок.

При этом Random Forest (RF) показал схожие результаты по каждому из видов возмещения убытка. Такие значения по основным метрикам качества можно обосновать отсутствием технологических мощностей для полноценного применения модели. Однако в сравнении с GLM величина ошибок меньше, а R^2 больше, что указывает на большую точность при прогнозировании. Величина RMSE составляет 26 583 и 18 099 руб. для ТВУ и ПВУ соответственно.

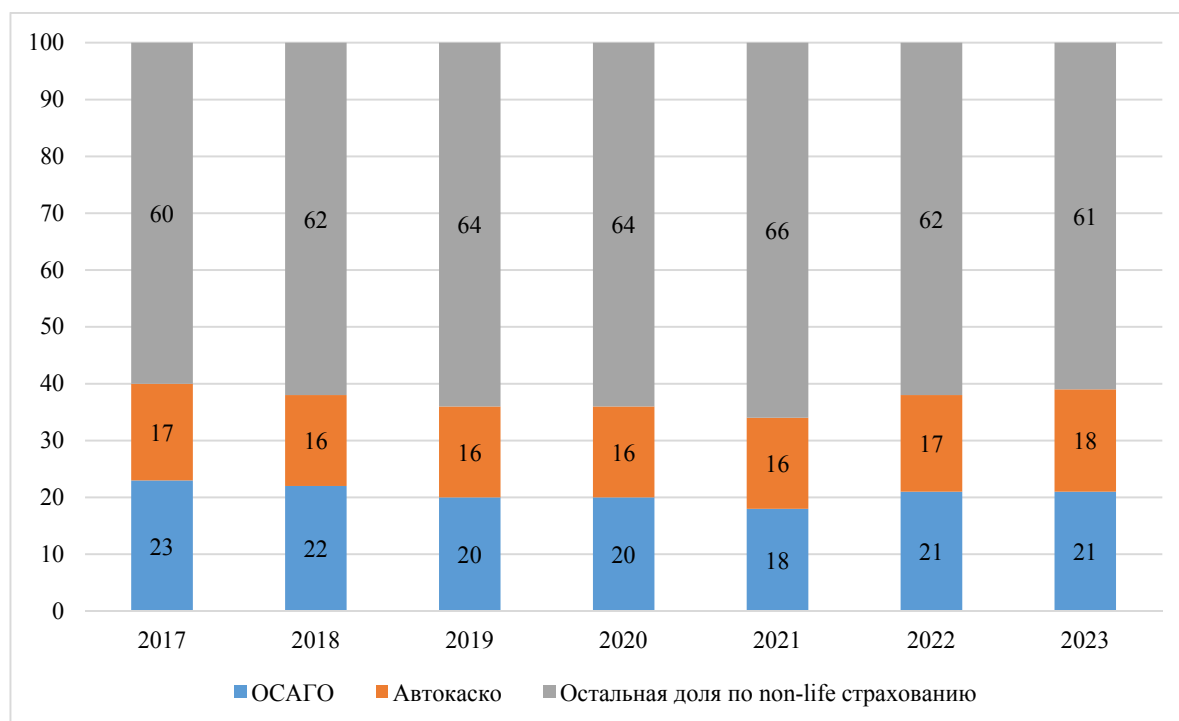


Рис. 3 / Fig. 3. Доли премий транспортного страхования на non-life рынке за 2017–2023 гг., % /
The Share of Transport Insurance Premiums in the Non-Life Insurance Market for the Period
2017–2023, in %

Источник / Source: составлено авторами по данным Банка России / Compiled by the authors according to the data of the Bank of Russia

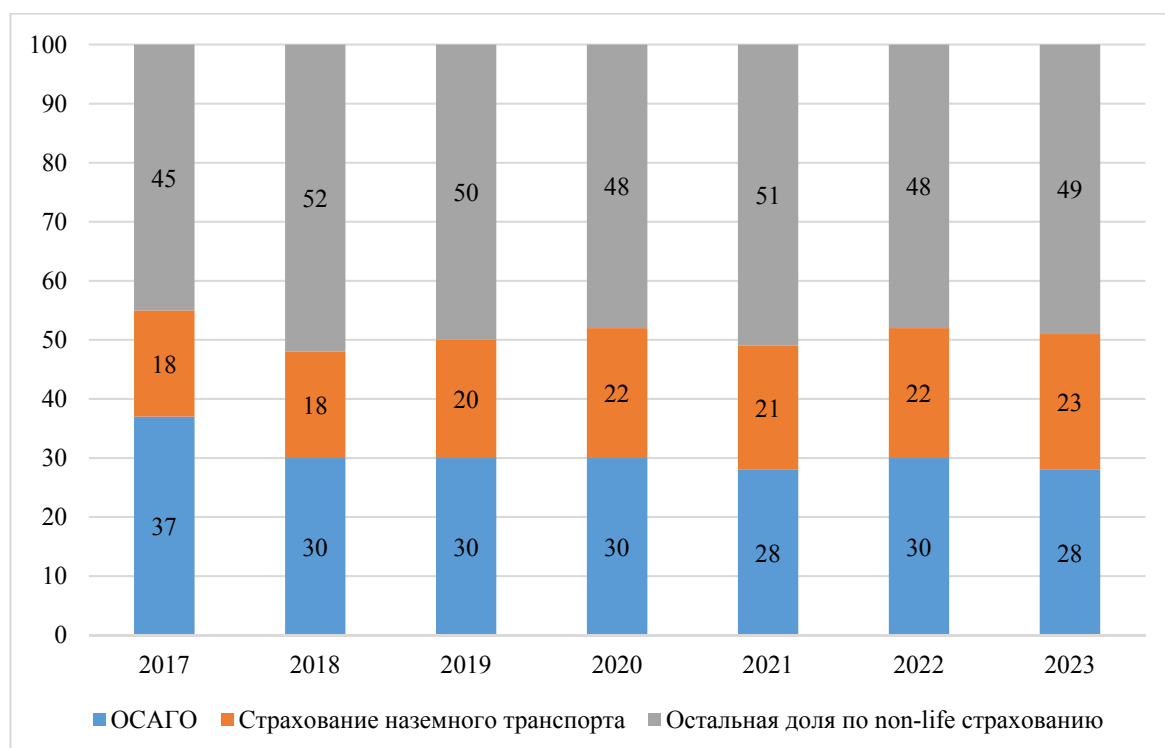


Рис. 4 / Fig. 4. Доли выплат по транспортному страхованию на non-life рынке за 2017–2023 гг., % /
The Share of Transport Insurance Payouts in the Non-Life Insurance Market for the Period
2017–2023, in %

Источник / Source: составлено авторами по данным Банка России / Compiled by the authors according to the data of the Bank of Russia.

**Сравнение основных метрик качества алгоритмов /
Comparison of the Main Metrics for Algorithm Quality**

Показатель / Indicator	ТВУ / Traditional loss			ПВУ / Direct loss		
	R ²	MAE, руб.	RMSE, руб.	R ²	MAE, руб.	RMSE, руб.
GLM	0,624	8201	26 972	0,644	6300	18 161
RF	0,631	8049	26 467	0,661	6229	18 099
XGBoost	0,670	7625	25 467	0,718	5894	17 388
CatBoost	0,685	7217	25 220	0,733	5393	16 622

Источник / Source: составлено авторами / Compiled by the authors.

Более надежными метриками обладают бустинговые алгоритмы машинного обучения — XGBoost и CatBoost. Значения коэффициента детерминации значительно выше, а ошибки ниже. Так R^2 у XGBoost равняется 0,670 и 0,718, а относительные ошибки примерно на тысячу меньше, чем у Random Forest.

Близкие метрики получил CatBoost. Данная модель показала лучшие результаты по R^2 по каждому убытку: 0,685 для ТВУ и 0,733 для ПВУ. CatBoost имеет наименьшие ошибки среди используемых моделей, так, стандартное отклонение — 7217 и 5393 руб. для ТВУ и ПВУ соответственно.

Таким образом, использование бустинговых методов машинного обучения позволило повысить точность прогноза.

**Оценка влияния переменных на страховые риски
транспортного страхования**

Важный результат исследования — выявление наиболее значимых факторов, что также позволит повысить точность прогноза.

Рассмотрим полученный набор значимых факторов для модели XGBoost и их оценки. Полученные результаты свидетельствуют о высоком влиянии на частоту для ТВУ таких факторов, как коэффициент территории, коэффициент «бонус-малус» (КБМ), категория транспортного средства, максимальный возраст водителя, период использования транспортного средства (ТС), тип населенного пункта (рис. 5А). Значение коэффициента территории зависит от места регистрации автовладельца, так, например, в крупных городах с большим транспортным потоком и интенсивным движением значение коэффициента будет выше. КБМ тем больше, чем больше количество страховых возмещений было у водителя в предыдущий период страхования.

Если рассматривать серьезность для ТВУ, то значимыми переменными являются количество женщин-водителей, тип ТС, возраст ТС, период использования (рис. 5Б).

При рассмотрении значимости переменных при прогнозировании частоты и серьезности прямого убытка для ПВУ можно выделить аналогичные факторы: количество ДТП, количество водителей, КБМ, коэффициент территории, переменные, связанные с опытом водителей и периодом активного использования, а также гендерный состав водителей в полисе ОСАГО, как видно на рис. 5В и 5Г. Полученные результаты позволяют сделать вывод о том, что опыт вождения, состояние ТС и территория использования являются наиболее значимыми факторами, влияющими на страховой риск.

Далее был применен подход SHAP для выявления как значимости, так и направления влияния основных факторов (рис. 6).

Из рис. 6А видно, что большие значения КБМ приводят к увеличению частоты ТВУ, а более низкие — к снижению. К росту прогнозной частоты также приводит увеличение показателей коэффициента территории, количества ДТП на транспортном средстве со средней степенью повреждений, мощности и региона собственника. К уменьшению частоты приводит большее количество женщин-водителей и максимальное значение стажа водителей.

Средний убыток повышают: высокое значение показателя региона (характеризует убыточность региона в общем рейтинге по портфелю: чем менее убыточен регион — тем ниже значение показателя), КБМ, возраста автомобиля. К уменьшению убытка приводит большее число женщин-водителей, больший максимальный возраст водителя (рис. 6В).

Рассмотрение влияния факторов для ТВУ методом CatBoost показывает схожие результаты с XGBoost. Рост количества женщин среди водителей транспортного средства снижает частоту и убыток по ДТП. Большая мощность транспортного средства положительно влияет на каждую из переменных как в традиционном, так и в прямом способе возмещения убытка; аналогичное влияние с методом XGBoost показывает и наличие средней тяжести ДТП

name_features	Sign_features	name_features	Sign_features
policy_coefficient_kbm	0.132479	drivers_gender_female_count	0.079770
policy_coefficient_kt	0.066321	vehicle_type	0.047534
vehicle_category	0.061483	policy_coefficient_ks	0.043117
vehicle_count_dtp	0.042126	policy_coefficient_ko	0.036276
insurant_locality_type	0.038454	vehicle_count_medium_dtp	0.032102
vehicle_type	0.033186	owner_region	0.030840
drivers_age_max	0.032485	owner_locality_type	0.028746
policy_coefficient_km	0.027878	vehicle_age	0.026377
vehicle_count_medium_dtp	0.027871	policy_start_year	0.025500
policy_period_active	0.025210	vehicle_power_kwt	0.025046
drivers_age_min	0.025165	policy_coefficient_kt	0.023170
drivers_gender_female_count	0.024937	policy_coefficient_kvs	0.022552
drivers_experience_max	0.021553	owner_age	0.022013
owner_region	0.020742	drivers_experience_max	0.021710
policy_aim_use_taxi	0.019583	policy_period_active	0.021642
vehicle_power_kwt	0.019566	vehicle_model	0.021472
policy_coefficient_ks	0.019338		
drivers_experience_min	0.019069		
vehicle_count_medium_points_dtp	0.018275		
insurant_age	0.018010		
А) Частота, ТБУ		Б) Серьезность, ТБУ	
name_features	Sign_features	name_features	Sign_features
policy_coefficient_kt	0.076566	vehicle_power_hp	0.081134
vehicle_power_hp	0.062220	vehicle_power_kwt	0.060909
vehicle_make	0.058040	drivers_count	0.060345
drivers_age_max	0.047239	policy_coefficient_ks	0.053398
vehicle_category	0.045503	vehicle_manufacture_country	0.043917
insurant_age	0.038312	vehicle_count_medium_points_dtp	0.034532
insurant_locality_type	0.038123	policy_start_year	0.031480
vehicle_count_dtp	0.033822	vehicle_make	0.029554
vehicle_power_kwt	0.029714	vehicle_power_hp_group	0.028178
vehicle_count_medium_points_dtp	0.028332	vehicle_count_serious_points_dtp	0.025731
policy_coefficient_km	0.027194	owner_region	0.024118
policy_period_active	0.025914	owner_locality_type	0.023684
policy_coefficient_kbm	0.024952	policy_coefficient_ko	0.023165
vehicle_type	0.023690	policy_period_active	0.021034
vehicle_count_medium_dtp	0.022618	vehicle_model	0.020762
drivers_gender_female_count	0.022216	drivers_gender_female_count	0.020411
vehicle_age	0.021840	is_insurant_equal_owner	0.019579
owner_locality_type	0.020841	vehicle_category	0.018960
vehicle_model	0.018066		
В) Частота, ПБУ		Г) Серьезность, ПБУ	

Рис. 5. / Fig. 5. **Значимость признаков при прогнозе параметров убытка (XGBoost) / The Significance of Features in Predicting Loss Metrics (XGBoost)**

Источник / Source: составлено авторами / Compiled by the authors.

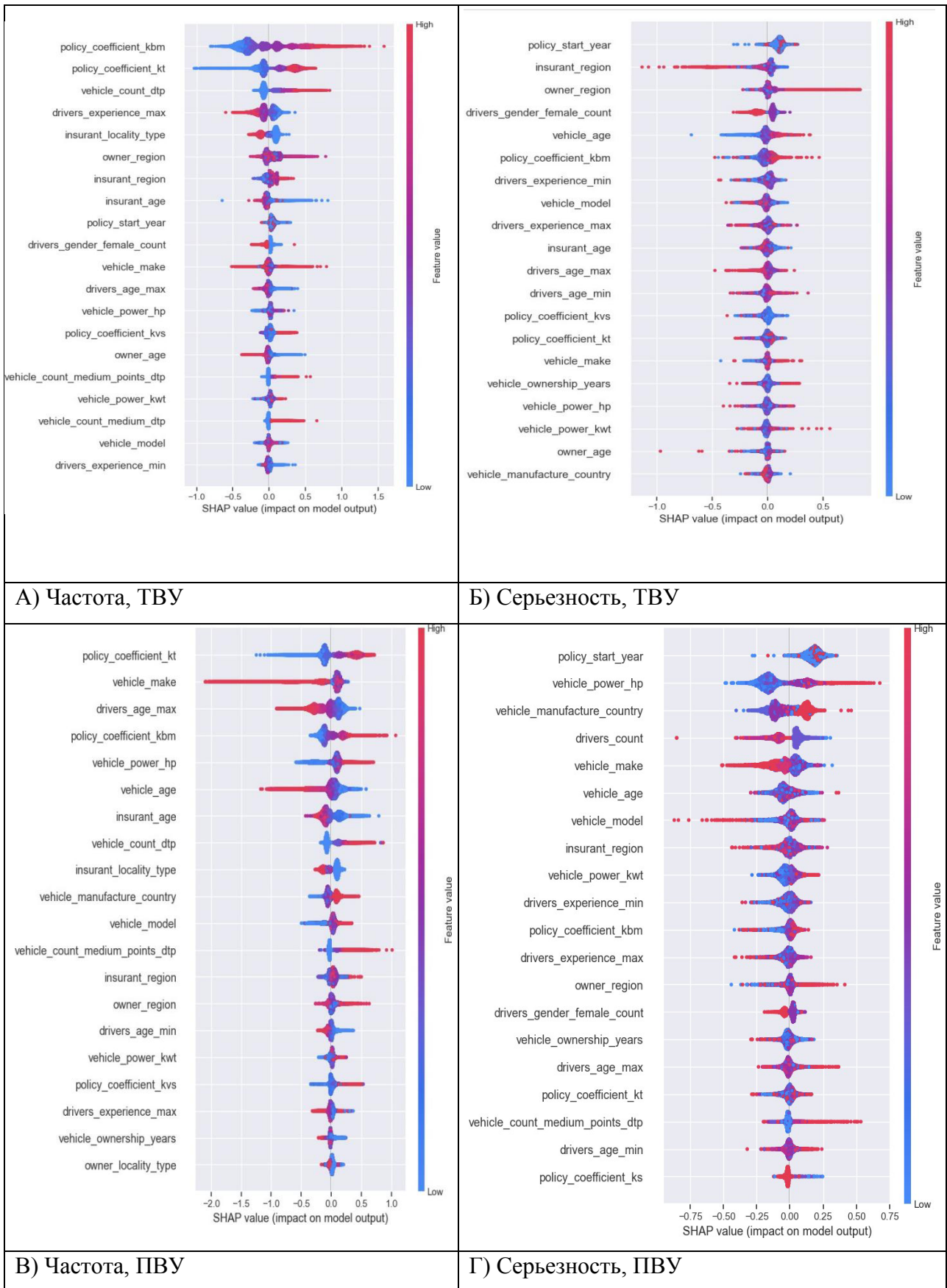


Рис. 6. / Fig. 6. Влияние переменных в модели XGBoost (SHAP) / The Influence of Variables in the XGBoost Model (SHAP)

Источник / Source: составлено авторами / Compiled by the authors.

Влияние наиболее значимых факторов на страховые риски / The Impact of the Most Significant Factors on Insurance Risks

Увеличивают риск / Increase the risk	Снижают риск / Reduce the risk
Мощность ТС	Максимальный возраст водителя
Количество ДТП на машине, в особенности средней тяжести	Количество вписанных в полис женщин-водителей
Возраст ТС	Количество водителей
Коэффициент КБМ	Максимальный опыт водителей
Срок владения ТС	Возраст страхователя
Коэффициент территории	Возраст собственника

Источник / Source: составлено авторами / Compiled by the authors.

на ТС до момента начала действия полиса. А наличие опытных водителей снижает вероятность страховых случаев и их повреждений в денежном выражении. Данные по значимым переменным и их влиянию на страховые риски приведены в табл. 3.

Полученные оценки влияния на целевые переменные, частоту и серьезность для ТВУ и ПВУ говорят о надежности построенных моделей и значимости отобранных факторов. Так, наличие опыта у водителей позволяет лучше среагировать на аварийную ситуацию и способствует повышению понимания ситуации на дороге и ожидаемых последствий. А с увеличением возраста у водителей может снижаться склонность к риску в целом. Наличие ДТП на автомобиле может свидетельствовать о возможных технических неисправностях, которые влекут за собой отрицательные последствия.

ВЫВОДЫ

В ходе исследования было рассмотрено несколько регрессионных моделей, прогнозирующих убытки по полису ОСАГО. При прогнозировании были учтены особенности данного типа страхования — прогнозируемый убыток разделялся по способу возмещения на традиционный и прямой. Также было задано две целевых переменных — частота и серьезность, итоговый убыток получается путем

их перемножения. Подобный подход был обусловлен особенностями формирования итогового значения ущерба.

Полученные результаты показали, что при прогнозировании убытков по полису ОСАГО эффективность использования бустинговых моделей машинного обучения выше по сравнению с классическим методом прогнозирования GLM. Рассмотренные бустинговые модели, XGBoost и CatBoost, показали лучшие результаты по всем рассматриваемым метрикам, что подтверждает целесообразность их использования.

Важным результатом исследования является выделение наиболее значимых факторов, влияющих на частоту и серьезность убытка. Показано, что значимыми факторами выступают количество ДТП средней тяжести на ТС, количество водителей, коэффициент КБМ, минимальный стаж водителя в полисе, мощность транспортного средства и тип населенного пункта использования ТС. В целом, к высоким значениям прогнозируемого убытка приводят высокие значения мощности ТС, количества ДТП на машине, возраста ТС, коэффициента КБМ. Снижают риск убытка высокий максимальный возраст водителя в полисе, большее количество женщин-водителей, вписанных в полис, больший опыт водителей, а также возраст страхователя и собственника.

БЛАГОДАРНОСТИ

Работа выполнена по проекту «Методология и методика разработки и обоснования приоритетов инвестиционной, кредитно-денежной и фискальной политики структурной трансформации российской экономики в новых геополитических условиях» № 126020516495-3. Институт экономики и организации промышленного производства Сибирского отделения Российской академии наук, Новосибирск, Российская Федерация. Авторы благодарят Андрея Владимировича Малкова за участие в проведении исследования.

ACKNOWLEDGEMENTS

The work was carried out under the project “Methodology and methods for developing and substantiating priorities of investment, monetary and fiscal policies for the structural transformation of the Russian economy in new geopolitical conditions” No. 126020516495-3. Institute of Economics and Industrial Engineering, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation.

The authors thank Andrey V. Malkov for his participation in the study.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Щербанин Ю.А. Транспорт и экономический рост: взаимосвязь и влияние. *Евразийская экономическая интеграция*. 2011;(3):65-78.
Shcherbanin Yu.A. Transport and economic growth: Interconnection and impact. *Evraziiskaya ekonomicheskaya integratsiya = Journal of Eurasian Economic Integration*. 2011;(3):65-78. (In Russ.).
2. Нижегородцев Р.М., Горидько Н.П., Рослякова Н.А. Взаимосвязь между объемом ВРП и развитием транспортной инфраструктуры: опыт кластеризации регионов России. *Региональная экономика: теория и практика*. 2013;(33):19-24.
Nizhegorodtsev R. M., Gorid'ko N.P., Roslyakova N. A. The relationship between the volume of GRP and the development of transport infrastructure: The experience of clustering regions of Russia. *Regional'naya ekonomika: teoriya i praktika = Regional Economics: Theory and Practice*. 2013;(33):19-24. (In Russ.).
3. Васильева Т.С. Актуальные вопросы развития регионального рынка страхования и его роль в экономике. *Проблемы развития территории*. 2020;(3):103-115. DOI: 10.15838/ptd.2020.3.107.7
Vasil'eva T. S. Current issues of regional insurance market development and its role in the economy. *Problemy razvitiya territorii = Problems of Territory's Development*. 2020;(3):103-115. (In Russ.). DOI: 10.15838/ptd.2020.3.107.7
4. Прокопьева Е.Л. Современный страховой рынок России: проблемы и потенциал развития. *Финансы и кредит*. 2019;25(1):177-195. DOI: 10.24891/fc.25.1.177
Prokop'eva E. L. The modern Russian insurance market: Problems and potential for development. *Finansy i kredit = Finance and Credit*. 2019;25(1):177-195. (In Russ.). DOI: 10.24891/fc.25.1.177
5. Рыбаков С.И. Оценка экономической эффективности транспортного страхования в Российской Федерации. *ЭТАП: экономическая теория, анализ, практика*. 2022;(4):135-167. DOI: 10.24412/2071-6435-2022-4-135-167
Rybakov S. I. The assessment of the economic efficiency of transport insurance in the Russian Federation. *ETAP: ekonomicheskaya teoriya, analiz, praktika = ETAP: Economic Theory, Analysis, and Practice*. 2022;(4):135-167. (In Russ.). DOI: 10.24412/2071-6435-2022-4-135-167
6. Шанихина Н.Н., Окунева А.А. Проблемы и перспективы развития автострахования в РФ. *International Journal of Advanced Studies*. 2018;8(1):169-179. DOI: 10.12731/2227-930X-2018-1-169-179
Shanikhina N. N., Okuneva A. A. Problems and prospects of development of auto insurance in Russia. *International Journal of Advanced Studies*. 2018;8(1):169-179. (In Russ.). DOI: 10.12731/2227-930X-2018-1-169-179
7. Комарова Н.В. Основные тенденции развития современного российского страхового рынка. *Вестник Санкт-Петербургского университета. Экономика*. 2013;(1):91-99.
Komarova N. V. The mainstream of the modern Russian insurance market development. *Vestnik Sankt-Peterburgskogo universiteta. Ekonomika = St. Petersburg University Journal of Economic Studies*. 2013;(1):91-99. (In Russ.).
8. Туленты Д.С. Эффективность автострахования в Российской Федерации. *Имущественные отношения в Российской Федерации*. 2022;(9):54-65. DOI: 10.24412/2072-4098-2022-9252-54-65
Tulenty D. S. Efficiency of auto insurance in the Russian Federation. *Imushchestvennyye otnosheniya v Rossiiskoi Federatsii = Property Relations in the Russian Federation*. 2022;(9):54-65. (In Russ.). DOI: 10.24412/2072-4098-2022-9252-54-65
9. Цыганов А.А., Брызгалов Д.В. Цифровизация страхового рынка: задачи, проблемы и перспективы. *Экономика. Налоги. Право*. 2018;11(2):111-120. DOI: 10.26794/1999-849X-2018-11-2-111-120
Tsyganov A. A., Bryzgalov D. V. Digitalization of the insurance market: Tasks, problems and prospects. *Ekonomika. Nalogi. Pravo = Economics, Taxes & Law*. 2018;11(2):111-120. (In Russ.). DOI: 10.26794/1999-849X-2018-11-2-111-120
10. Шарапова С.А. Концептуальные вопросы функционирования и проблемы развития вертикально интегрированных сбытовых систем страхования. *Вестник МГПУ. Серия: Экономика*. 2018;(2):43-53.
Sharapova S. A. Conceptual issues of functioning and problems of development of vertically integrated marketing insurance systems. *Vestnik MGPU. Seriya: Ekonomika = MCU Journal of Economic Studies*. 2018;(2):43-53. (In Russ.).
11. Котлововский И.Б., Саган А.И. Государственное регулирование страхового рынка России в условиях вступления в ВТО. *Финансы*. 2012;(12):47-53.
Kotlovovskii I. B., Sagan A. I. State regulation of the Russian insurance market in the context of accession to the WTO. *Finansy = Finance*. 2012;(12): 47-53. (In Russ.).
12. Бекетнова Ю.М. Сравнительный анализ методов машинного обучения при идентификации признаков вовлеченности кредитных организаций и их клиентов в сомнительные операции. *Финансы: теория и практика*. 2021;25(5):186-199. DOI: 10.26794/2587-5671-2020-25-5-186-199

- Beketnova Yu.M. Comparative analysis of machine learning methods to identify signs of suspicious transactions of credit institutions and their clients. *Finance: Theory and Practice*. 2021;25(5):186-199. DOI: 10.26794/2587-5671-2020-25-5-186-199
13. Abdelhadi S., Elbahnasy K., Abdelsalam M. A proposed model to predict auto insurance claims using machine learning techniques. *Journal of Theoretical and Applied Information Technology*. 2020;98(22):3428-3437.
 14. Hsu Y.-C., Chou P.-L., Shiu Y.-M. An examination of the relationship between vehicle insurance purchase and the frequency of accidents. *Asia Pacific Management Review*. 2016;21(4):231-238. DOI: 10.1016/j.apmr.2016.08.001
 15. Яркова О.Н., Федорова П.С. Математическое моделирование финансовых показателей страховой компании по договорам автострахования. Удалов Ф.Е., Бондаренко В.В., ред. Повышение управленческого, экономического, социального и инновационно-технического потенциала предприятий, отраслей и народно-хозяйственных комплексов. Сб. ст. XIII Междунар. науч.-практ. конф. (Пенза, 19–20 мая 2022 г.). Пенза: РГАУ; 2022: 334-337.
Yarkova O.N., Fedorova P.S. Mathematical modeling of financial indicators of an insurance company under motor insurance contracts. In: Udalov F.E., Bondarenko V.V., eds. Increasing the managerial, economic, social and innovative-technical potential of enterprises, industries and national economic complexes. Proc. 13th Int. sci.-pract. conf. (Penza, May 19-20, 2022). Penza: Penza State Agrarian University; 2022:334-337. (In Russ.).
 16. Макаренко Е.А. Использование инструментария моделирования для прогнозирования банкротства страховых компаний. *Известия высших учебных заведений. Серия: Экономика, финансы и управление производством*. 2021;(2):43-50. DOI: 10.6060/ivecofin.2021482.533
Makarenko E.A. Modeling tools to predict the bankruptcy of insurance companies. *Izvestiya vysshikh uchebnykh zavedenii. Seriya: Ekonomika, finansy i upravlenie proizvodstvom = News of Higher Educational Institutions. Series: Economy, Finance and Production Management*. 2021;(2):43-50. (In Russ.). DOI: 10.6060/ivecofin.2021482.533
 17. Цыганов А.А., Языков А.Д., Яненко Е.А., Грызенкова Ю.В. Опыт обоснования изменения коэффициента возраста и стажа водителя в ОСАГО. *Финансы: теория и практика*. 2020;24(4):31-46. DOI: 10.26794/2587-5671-2020-24-4-31-46
Tsyganov A.A. Yazykov A.D., Yanenko E.A., Gryzenkova Yu.V. Interpreting the change of the age and experience coefficient in Motor Third-Party Liability Insurance. *Finance: Theory and Practice*. 2020;24(4):31-46. DOI: 10.26794/2587-5671-2020-24-4-31-46
 18. Харламов А.В. Модели прогнозирования в страховании. Трансформация страховой индустрии для устойчивого развития в новых условиях. Сб. тр. XXII Междунар. науч.-практ. конф. (Москва, 2–3 июня 2021 г.). М.: Всероссийский союз страховщиков; 2021:231-236.
Kharlamov A.V. Forecasting models in insurance. In: Transforming the insurance industry for sustainable development in the new environment. Proc. 22nd Int. sci.-pract. conf. (Moscow, June 2–3, 2021). Moscow: All-Russian Union of Insurers; 2021:231-236. (In Russ.).
 19. Yeo A.C., Smith K.A., Willis R.J., Brooks M. Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry. *Intelligent Systems in Accounting, Finance and Management*. 2001;10(1):39-50. DOI: 10.1002/isaf.196
 20. Babii A., Ghysels E., Striaukas J. Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*. 2022;40(3):1094-1106. DOI: 10.1080/07350015.2021.1899933
 21. Ball R.T., Ghysels E. Automated earnings forecasts: Beat analysts or combine and conquer? *Management Science*. 2018;64 (10):4936-4952. DOI: 10.1287/mnsc.2017.2864
 22. Hanafy M., Ming R. Machine learning approaches for auto insurance Big Data. *Risks*. 2021;9(2):42. DOI: 10.3390/risks9020042
 23. Zanke P., Sontakke D. Leveraging machine learning algorithms for risk assessment in auto insurance. *Journal of Artificial Intelligence Research*. 2021;1(1):21-39.
 24. Breiman L. Random forests. *Machine learning*. 2001;45(1):5-32. DOI: 10.1023/A:1010933404324
 25. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: Unbiased boosting with categorical features. In: Bengio S., et al., eds. Proc. 32nd Conf. on neural information processing systems (NeurIPS 2018). (Montreal, December 3-8, 2018). San Diego, CA: Neural Information Processing Systems Foundation; 2018:75-90. (Advances in neural information processing systems. Vol. 31).
 26. Dorogush A.V., Ershov V., Gulin A. CatBoost: Gradient boosting with categorical features support. arXiv preprint. 2018. DOI: 10.48550/arXiv.1810.11363

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS



Ирина Викторовна Филимонова — доктор экономических наук, профессор, старший научный сотрудник, Институт экономики и организации промышленного производства Сибирского отделения Российской академии наук, Новосибирск, Российская Федерация; заведующая кафедрой политической экономии, Новосибирский национальный исследовательский государственный университет, Новосибирск, Российская Федерация

Irina V. Filimonova — Dr. Sci. (Econ.), Prof., Senior Researcher, Institute of Economics and Industrial Engineering, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation; Head of the Department of Political Economy, Novosibirsk National Research State University, Novosibirsk, Russian Federation

<https://orcid.org/0000-0003-4447-6425>

Автор для корреспонденции / Corresponding author:

filimonovaiv@list.ru



Анна Владимировна Комарова — кандидат экономических наук, доцент, старший научный сотрудник, Институт экономики и организации промышленного производства Сибирского отделения Российской академии наук, Новосибирск, Российская Федерация; доцент кафедры политической экономии, Новосибирский национальный исследовательский государственный университет, Новосибирск, Российская Федерация

Anna V. Komarova — Cand. Sci. (Econ.), Assoc. Prof., Senior Researcher, Institute of Economics and Industrial Engineering, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation; Assoc. Prof. of the Department of Political Economy, Novosibirsk National Research State University, Novosibirsk, Russian Federation

<http://orcid.org/0000-0002-5844-1648>

a.komarova@g.nsu.ru



Анастасия Павловна Саматова — студентка, Новосибирский национальный исследовательский государственный университет, Новосибирск, Российская Федерация

Anastasia P. Samatova — student, Novosibirsk National Research State University, Novosibirsk, Russian Federation

<https://orcid.org/0009-0002-8742-8669>

a.samatova@g.nsu.ru

Заявленный вклад авторов:

И.В. Филимонова — постановка проблемы, разработка концепции статьи, формирование выводов исследования.

А.В. Комарова — анализ литературы, проблемы, подходы и методы анализа страховых рисков.

А.П. Саматова — оформление, табличное представление и описание результатов исследования.

Authors' declared contributions:

I. V. Filimonova — statement of the problem, development of the concept of the article, formation of the conclusions of the study.

A. V. Komarova — literature analysis, problems, approaches and methods of insurance risk analysis.

A. P. Samatova — design, tabular presentation and description of the results of the study.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила в редакцию 26.09.2024; после рецензирования 29.10.2024; принята к публикации 22.11.2024.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was submitted on 26.09.2024; revised on 29.10.2024 and accepted for publication on 22.11.2024.

The authors read and approved the final version of the manuscript.

Переводчик Н.И. Соколова