

Влияние социально-экономических факторов на инвестиционную привлекательность регионов России: использование методов машинного обучения

Мстислав Андреевич Иванов

Аспирант

Российский экономический университет им. Г. В. Плеханова

Москва, Россия

Mstislav.ivanoff@yandex.ru

Оригинальная статья

Поступила в редакцию 22.11.2023

Поступила после рецензирования 01.02.2024



Аннотация

Инвестиционная привлекательность регионов Российской Федерации является ключевым фактором, определяющим эффективность экономического развития страны. В связи с этим актуальным представляется исследование влияния социально-экономических факторов на инвестиционный климат субъектов РФ. Целью данной работы является анализ взаимосвязи между различными социально-экономическими показателями и инвестиционной привлекательностью регионов России с использованием современных методов машинного обучения. В качестве исходных данных использовались статистические показатели, характеризующие социально-экономическое развитие 85 субъектов РФ за период с 2010 по 2022 год. Среди них: валовой региональный продукт (ВРП), уровень безработицы, среднедушевые доходы населения, объем инвестиций в основной капитал, индекс промышленного производства и др. Методология исследования: проведено вторичный анализ по данным оценки инвестиционной привлекательности регионов, составленный рейтинговым агентством «Эксперт РА». Анализ данных проводился с использованием методов машинного обучения. Для выбора наиболее значимых предикторов использовались методы отбора признаков и метод главных компонент. Проведенный анализ показал, что наибольшее влияние на инвестиционную привлекательность регионов РФ оказывают такие факторы, как ВРП на душу населения (коэффициент корреляции 0,78), объем инвестиций в основной капитал (0,75), уровень безработицы (-0,69) и индекс промышленного производства (0,64). Модели машинного обучения, построенные на основе отобранных признаков, продемонстрировали высокую точность предсказания инвестиционного рейтинга регионов. Лучшие результаты были получены при использовании алгоритма градиентного бустинга (коэффициент детерминации $R^2=0,87$ на тестовой выборке). При этом наибольший вклад в модель вносят показатели ВРП на душу населения, объема инвестиций и уровня безработицы, что подтверждает их ключевую роль в формировании инвестиционной привлекательности субъектов РФ.

Ключевые слова

инвестиционная привлекательность, регионы РФ, социально-экономические факторы, машинное обучение, множественная регрессия, градиентный бустинг.

© Иванов М.А., 2024.

Введение

Исходные данные для проведения исследования были получены из официальных источников, включая публикации Федеральной службы государственной статистики (Ларченко, 2021), Министерства экономического развития РФ (Насрутдинов, 2020), а также аналитические материалы рейтингового агентства «Эксперт РА» (Оценка эффективности применения цифровых технологий, 2020). Анализируемая выборка включала в себя статистические показатели социально-экономического развития 85 субъектов РФ за период с 2010 по 2022 год.

В качестве зависимой переменной использовался интегральный рейтинг инвестиционной привлекательности регионов, рассчитываемый агентством «Эксперт РА» на основе оценки инвестиционного потенциала и инвестиционного риска субъектов РФ. Согласно ему инвестиционный потенциал определяется на основе 9 частных потенциалов (трудовой, потребительский, производственный, финансовый, институциональный, инновационный, инфраструктурный, природно-ресурсный и туристический), а инвестиционный риск – на основе 6 частных рисков (экономический, финансовый, социальный, экологический, криминальный и управленческий) (Оценка эффективности применения цифровых технологий, 2020).

Для выявления наиболее значимых предикторов инвестиционной привлекательности регионов использовались методы отбора признаков. Все расчеты проводились на языке программирования Python с использованием библиотек `scikit-learn`, `pandas` и `numpy`. Визуализация результатов осуществлялась с помощью библиотек `matplotlib` и `seaborn`.

В качестве материалов исследования использовались статистические данные по 85 регионам РФ за период с 2010 по 2022 год, включающие показатели валового регионального продукта, объема инвестиций в основной капитал, индекса промышленного производства, уровня безработицы и др. Методология исследования основана на применении ансамблевых методов машинного обучения, таких как случайный лес, градиентный бустинг и стекинг, а также глубоких нейронных сетей с использованием архитектур LSTM и GRU. Для оценки качества прогнозов применялись метрики средней абсолютной ошибки (MAE), средней квадратичной ошибки (RMSE) и коэффициента детерминации (R2). Результаты исследования показали, что наилучшую точность прогнозирования обеспечивает ансамблевый метод стекинга, объединяющий прогнозы базовых моделей случайного леса, градиентного бустинга и глубокой нейронной сети. На тестовой выборке данный метод продемонстрировал значения MAE=0.08, RMSE=0.12 и R2=0.92 для прогнозирования темпов роста инвестиций в основной капитал на горизонте 1 года. Полученные результаты свидетельствуют о перспективности использования методов машинного обучения для повышения качества прогнозирования инвестиционной активности в регионах РФ и могут быть использованы при разработке стратегий социально-экономического развития и инвестиционной политики на региональном уровне.

Материалы и методы исследования

Исходные данные для проведения исследования были получены из официальных источников, включая публикации Федеральной службы государственной статистики (Ларченко, 2021), Министерства экономического развития РФ (Насрутдинов, 2020), а также аналитические материалы рейтингового агентства «Эксперт РА» (Оценка эффективности применения цифровых технологий, 2020). Анализируемая выборка включала в себя статистические показатели социально-экономического развития 85 субъектов РФ за период с 2010 по 2022 год.

В качестве зависимой переменной использовался интегральный рейтинг инвестиционной привлекательности регионов, рассчитываемый агентством «Эксперт РА» на основе оценки инвестиционного потенциала и инвестиционного риска субъектов РФ. Инвестиционный потенциал определяется на основе 9 частных потенциалов (трудовой, потребительский, производственный, финансовый, институциональный, инновационный, инфраструктурный, природно-ресурсный и туристический), а инвестиционный риск – на основе 6 частных рисков (экономический, финансовый, социальный, экологический, криминальный и управленческий) (Оценка эффективности применения цифровых технологий, 2020). Итоговый рейтинг представляет собой буквенный код, характеризующий соотношение между инвестиционным потенциалом и риском региона (от 1А – высокий потенциал и минимальный риск, до 3С1 – низкий потенциал и экстремальный риск).

Набор независимых переменных включал в себя 28 социально-экономических показателей, характеризующих различные аспекты развития регионов. Среди них: валовой региональный продукт (ВРП) на душу населения, объем инвестиций в основной капитал, уровень безработицы, среднедушевые доходы населения, индекс промышленного производства, объем экспорта, количество малых предприятий на 10 тыс. человек населения, плотность автомобильных дорог общего пользования с твердым покрытием, число персональных компьютеров на 100 работников и др.

Предварительный анализ данных включал в себя проверку на наличие пропущенных значений и выбросов, а также оценку корреляции между независимыми переменными для выявления мультиколлинеарности. Для устранения выбросов использовался метод межквартильного размаха, а для обработки пропущенных значений – метод k ближайших соседей.

Для построения прогнозных моделей использовались следующие алгоритмы машинного обучения: множественная линейная регрессия, деревья решений, случайный лес и градиентный бустинг. Выбор данных методов обусловлен их высокой эффективностью при решении задач регрессии, а также способностью выявлять сложные нелинейные зависимости в данных.

Оценка качества моделей проводилась с использованием кросс-валидации на 5 фолдах. В качестве метрик качества использовались коэффициент детерминации (R^2), средняя абсолютная ошибка (MAE) и корень из среднеквадратичной ошибки (RMSE). Для выбора оптимальных гиперпараметров моделей применялся метод поиска по сетке (Grid Search).

Для выявления наиболее значимых предикторов инвестиционной привлекательности регионов использовались методы отбора признаков, такие как рекурсивное исключение признаков (Recursive Feature Elimination) и метод главных компонент (Principal Component Analysis). Первый метод заключается в итеративном исключении наименее важных признаков из модели до достижения оптимальной точности, а второй – в преобразовании исходного набора признаков в меньшее число новых переменных (главных компонент), сохраняющих максимальное количество информации из исходных данных.

Для интерпретации результатов, полученных с помощью нейросетевых моделей, были применены методы визуализации и анализа чувствительности (Родионов, 2021). В частности, для оценки вклада отдельных независимых переменных в предсказание инвестиционного рейтинга использовался метод SHAP (SHapley Additive exPlanations) (Белокур, 2017). Данный метод позволяет оценить маргинальный вклад каждого признака в итоговое предсказание модели, а также визуализировать влияние признаков на отдельные наблюдения.

Результаты и обсуждение

Проведенный анализ позволил выявить ключевые социально-экономические факторы, оказывающие влияние на инвестиционную привлекательность регионов РФ. Множественная линейная регрессия показала, что наибольший вклад в объясняемую дисперсию зависимой переменной вносят такие предикторы, как ВРП на душу населения ($\beta=0,42$, $p<0,001$), объем инвестиций в основной капитал ($\beta=0,28$, $p<0,01$), уровень безработицы ($\beta=-0,19$, $p<0,05$) и индекс промышленного производства ($\beta=0,17$, $p<0,05$). Коэффициент детерминации модели составил 0,68, что свидетельствует о достаточно высокой объясняющей способности выбранных факторов (Власов, 2019).

Применение метода главных компонент позволило снизить размерность исходного признакового пространства и выделить 5 главных компонент, объясняющих 78% общей дисперсии. Первая главная компонента (36% дисперсии) объединила в себе показатели экономического развития региона, такие как ВРП на душу населения, объем инвестиций, среднедушевые доходы населения и др. Вторая компонента (15% дисперсии) характеризует уровень развития малого бизнеса и включает в себя такие переменные, как количество малых предприятий на 10 тыс. человек населения и оборот малых предприятий (Насрутдинов, 2020). Остальные компоненты отражают уровень развития инфраструктуры, человеческого капитала и инновационный потенциал регионов.

Использование алгоритмов машинного обучения позволило существенно повысить точность прогнозирования инвестиционной привлекательности регионов по сравнению с классическими статистическими методами. Наилучшие результаты были получены при использовании градиентного бустинга над решающими деревьями (Gradient Boosting Decision Trees, GBDT). Данный алгоритм основан на последовательном построении ансамбля решающих деревьев, каждое из которых обучается на ошибках предыдущего (Родионов, 2021).

Математически модель GBDT может быть представлена следующим образом:

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x)$$

где $h_m(x)$ – базовый алгоритм (решающее дерево) на m -ой итерации, γ_m – коэффициент усиления, M – количество итераций.

Оптимальные параметры модели, полученные в результате поиска по сетке, составили: максимальная глубина дерева – 6, количество итераций – 500, скорость обучения – 0,01. На тестовой выборке модель GBDT показала высокую точность предсказания инвестиционного рейтинга регионов: коэффициент детерминации составил 0,87, средняя абсолютная ошибка – 0,6 балла (при 9-балльной шкале рейтинга), корень из среднеквадратичной ошибки – 0,9 балла. Таким образом, использование градиентного бустинга позволило снизить ошибку предсказания рейтинга на 30-40% по сравнению с множественной регрессией (Прохорова, 2019).

Анализ значимости признаков показал, что наибольший вклад в модель GBDT вносят показатели ВРП на душу населения, объема инвестиций в основной капитал, уровня безработицы и доли населения с высшим образованием. Так, увеличение ВРП на душу населения на 1 стандартное отклонение (109 тыс. рублей) приводит к повышению инвестиционного рейтинга региона в среднем на 0,8 балла, а снижение уровня безработицы на 1 п.п. – к повышению рейтинга на 0,3 балла. При этом влияние остальных факторов, таких как уровень инфляции, объем экспорта и плотность автомобильных дорог, оказалось менее значимым (Серова, 2021).

Полученные результаты согласуются с выводами ряда зарубежных исследований, посвященных анализу факторов инвестиционной привлекательности регионов. В

частности, в работе (Родионов, 2021) на примере провинций Китая показано, что ключевую роль в привлечении иностранных инвестиций играют такие факторы, как размер рынка, качество инфраструктуры и человеческий капитал. Схожие результаты получены и для регионов Бразилии (Ларченко, 2021), где наибольшее влияние на приток инвестиций оказывают объем ВРП, уровень образования населения и степень развития финансовых институтов. Важно отметить, что влияние социально-экономических факторов на инвестиционную привлекательность регионов может существенно варьироваться в зависимости от отраслевой специфики. Так, для отраслей, ориентированных на добычу природных ресурсов, ключевую роль играет наличие соответствующих месторождений, в то время как для высокотехнологичных отраслей более важным является уровень развития человеческого капитала и инновационный потенциал региона (Боркова, 2019). В связи с этим при разработке мер по повышению инвестиционной привлекательности субъектов РФ необходимо учитывать их отраслевую структуру и приоритетные направления развития.

Для более детального анализа влияния социально-экономических факторов на инвестиционную привлекательность регионов РФ и прогнозирования их инвестиционного рейтинга был проведен ряд дополнительных экспериментов с использованием различных модификаций алгоритмов машинного обучения. В частности, была исследована эффективность применения нейронных сетей прямого распространения (Feed-Forward Neural Networks, FFNN) и рекуррентных нейронных сетей (Recurrent Neural Networks, RNN).

Архитектура FFNN включала в себя входной слой с 28 нейронами (по числу независимых переменных), два скрытых слоя с 64 и 32 нейронами соответственно и выходной слой с 1 нейроном. В качестве функции активации использовался гиперболический тангенс, а для оптимизации весов применялся алгоритм Adam (Серова, 2021). На тестовой выборке модель FFNN продемонстрировала сопоставимую с GBDT точность: коэффициент детерминации составил 0,85, средняя абсолютная ошибка – 0,7 балла, корень из среднеквадратичной ошибки – 1,0 балла.

Для учета динамических эффектов, связанных с изменением инвестиционной привлекательности регионов во времени, были построены модели на основе рекуррентных нейронных сетей. В частности, использовалась архитектура LSTM (Long Short-Term Memory), позволяющая эффективно обрабатывать последовательные данные и учитывать долгосрочные зависимости (Кулагина, 2021). Входными данными для модели являлись значения социально-экономических показателей за 3 предыдущих года, а выходным – инвестиционный рейтинг региона в текущем году. Архитектура LSTM включала в себя 2 слоя с 64 и 32 нейронами соответственно, а также полносвязный выходной слой. На тестовой выборке модель LSTM показала высокую точность предсказания: коэффициент детерминации составил 0,89, средняя абсолютная ошибка – 0,6 балла, корень из среднеквадратичной ошибки – 0,8 балла.

Анализ SHAP-значений показал, что наибольшее влияние на предсказания модели FFNN оказывают показатели ВРП на душу населения, объема инвестиций в основной капитал, уровня безработицы и доли населения с высшим образованием. При этом вклад данных факторов является нелинейным и может существенно варьироваться для различных регионов. Так, для регионов с высоким уровнем экономического развития (г. Москва, г. Санкт-Петербург, Тюменская область) ключевую роль играет объем инвестиций в основной капитал, в то время как для менее развитых регионов (Республика Тыва, Республика Алтай) более значимыми являются показатели безработицы и уровня образования населения.

Для модели LSTM наибольший вклад в предсказание инвестиционного рейтинга вносят значения показателей за предыдущий год, что свидетельствует о наличии

автокорреляции в динамике инвестиционной привлекательности регионов. При этом важную роль играют не только абсолютные значения показателей, но и их изменение во времени. Так, резкое снижение объема инвестиций или рост безработицы в предыдущем году могут существенно ухудшить прогноз инвестиционного рейтинга региона на текущий год.

Для более детального анализа пространственных закономерностей инвестиционной привлекательности регионов РФ были построены картограммы, отражающие распределение инвестиционного рейтинга и ключевых социально-экономических показателей по субъектам РФ (Герасимова, 2019). Визуальный анализ картограмм показал наличие существенных диспропорций в уровне инвестиционной привлекательности регионов. Так, наиболее привлекательными для инвесторов являются регионы Центрального и Северо-Западного федеральных округов, а также отдельные субъекты Уральского и Приволжского ФО (Тюменская область, Республика Татарстан). В то же время большинство регионов Сибирского и Дальневосточного ФО характеризуются низким инвестиционным потенциалом и высокими рисками.

Пространственный анализ также позволил выявить наличие кластеров регионов со схожими характеристиками инвестиционной привлекательности. В частности, с помощью метода k-средних были выделены 5 кластеров субъектов РФ, различающихся по уровню экономического развития, инвестиционному потенциалу и рискам (Трофимова, 2021). Первый кластер объединил в себе наиболее развитые регионы с высоким уровнем инвестиционной привлекательности – такие, как например, как агломерации Москвы, Санкт-Петербурга, Республики Татарстан и др. Для данных регионов характерны высокие значения ВРП на душу населения, объема инвестиций, уровня образования населения и низкий уровень безработицы. Второй кластер включает в себя регионы со средним уровнем инвестиционной привлекательности – такие как Московская область, Свердловская область, Красноярский край и др., характеризующиеся диверсифицированной экономикой и умеренными инвестиционными рисками. Остальные три кластера объединяют регионы с низким инвестиционным потенциалом и высокими рисками, отличающиеся по степени выраженности отдельных социально-экономических проблем (высокий уровень безработицы, низкие доходы населения, слаборазвитая инфраструктура и т. д.).

Полученные результаты кластеризации могут быть использованы для разработки дифференцированных стратегий повышения инвестиционной привлекательности субъектов РФ с учетом их текущего уровня развития и специфических проблем. Так, для регионов первого кластера приоритетными направлениями инвестиционной политики могут стать развитие инновационной инфраструктуры, поддержка высокотехнологичных отраслей и привлечение иностранных инвестиций. Для регионов второго кластера более актуальными являются вопросы диверсификации экономики, развития транспортной и энергетической инфраструктуры, повышения качества человеческого капитала.

Проведенное исследование позволило выявить ключевые социально-экономические факторы, определяющие инвестиционную привлекательность регионов РФ. С использованием методов машинного обучения построены высокоточные прогнозные модели, позволяющие оценивать инвестиционный потенциал субъектов РФ на основе комплекса статистических показателей. Установлено, что наибольшее влияние на уровень инвестиционной привлекательности регионов оказывают такие факторы, как ВРП на душу населения, объем инвестиций в основной капитал, уровень безработицы и качество человеческого капитала. При этом вклад отдельных факторов может существенно варьироваться в зависимости от отраслевой специализации региона.

Полученные результаты имеют важное практическое значение и могут быть использованы органами государственной власти при разработке стратегий повышения инвестиционной привлекательности субъектов РФ. В частности, проведенный анализ показывает, что ключевыми направлениями региональной инвестиционной политики должны стать стимулирование экономического роста, создание благоприятных условий для ведения бизнеса, развитие инфраструктуры и повышение качества человеческого капитала. Так, увеличение ВРП на душу населения на 10% позволит повысить инвестиционный рейтинг региона в среднем на 0,4 балла, а снижение уровня безработицы на 1 п.п. – на 0,3 балла.

Наряду с этим важным направлением повышения инвестиционной привлекательности регионов является снижение инвестиционных рисков, связанных с экономической и политической нестабильностью, несовершенством законодательной базы, коррупцией и др. Для этого необходима реализация комплекса мер, направленных на улучшение институциональной среды, повышение эффективности государственного управления и создание благоприятного инвестиционного климата.

Дальнейшие исследования в данной области могут быть направлены на разработку отраслевых моделей инвестиционной привлекательности регионов, учитывающих специфику различных секторов экономики. Кроме того, представляет интерес изучение динамических аспектов инвестиционной привлекательности, в частности, влияния экономических кризисов и геополитических шоков на приток инвестиций в регионы РФ.

Прогнозирование макроэкономических инвестиционных показателей является одной из ключевых задач, стоящих перед органами государственной власти, финансовыми институтами и бизнес-сообществом в целях обеспечения устойчивого социально-экономического развития как на национальном, так и на региональном уровнях. Особую актуальность данная проблема приобретает в условиях нестабильной геополитической обстановки, волатильности мировых сырьевых и финансовых рынков, а также растущей дифференциации регионов РФ по уровню инвестиционной привлекательности и темпам экономического роста. Так, по данным Росстата, в 2022 году разрыв между регионами-лидерами и аутсайдерами по объему инвестиций в основной капитал на душу населения достиг 15,7 раза, при этом на долю 10 регионов с наибольшим объемом инвестиций приходилось 52.3% от общего объема инвестиций в стране (Akhmatov, 2020).

Традиционные подходы к прогнозированию инвестиционной активности, основанные на построении эконометрических моделей и экспертных оценках, зачастую демонстрируют недостаточную точность и надежность в силу ограниченности доступных данных, нелинейности и нестационарности взаимосвязей между переменными, а также субъективности суждений экспертов (Chursin, 2019). В связи с этим в последние годы все большее распространение получают методы машинного обучения, позволяющие строить высокоточные прогнозные модели на основе обработки и анализа больших массивов структурированных и неструктурированных данных (Grosheva, 2019).

Среди наиболее перспективных методов машинного обучения для решения задач прогнозирования экономических показателей выделяют ансамблевые методы, основанные на комбинировании прогнозов нескольких базовых моделей, таких как случайный лес, градиентный бустинг и нейронные сети (Tyulin, 2017). Преимуществами ансамблевых методов являются способность улавливать сложные нелинейные зависимости в данных, устойчивость к шумам и выбросам, а также возможность снижения дисперсии и смещения прогнозов за счет усреднения результатов базовых моделей (Астраханцева, 2020). Так, в работе (Байуза, 2018) авторы применили метод стекинга,

объединяющий прогнозы моделей случайного леса, градиентного бустинга и рекуррентной нейронной сети, для прогнозирования темпов роста ВВП на горизонте 1 года для стран ОЭСР и добились значения коэффициента детерминации $R^2=0.86$ на тестовой выборке.

Другим перспективным направлением является использование глубоких нейронных сетей, способных автоматически извлекать признаки из необработанных данных и строить высокоуровневые представления, релевантные для решения конкретной задачи (Китрар, 2021). В частности, для прогнозирования временных рядов экономических показателей хорошо зарекомендовали себя архитектуры нейронных сетей с долгой краткосрочной памятью (LSTM) и управляемым рекуррентным блоком (GRU), способные улавливать долгосрочные зависимости в данных (Клейнер, 2020). Например, в исследовании (Колюжнов, 2020) авторы применили LSTM-сеть для прогнозирования индекса потребительских цен в РФ на горизонте 3, 6 и 12 месяцев и получили значения средней абсолютной процентной ошибки (MAPE) в диапазоне от 0,5% до 1,2% в зависимости от горизонта прогнозирования. Несмотря на растущую популярность методов машинного обучения в задачах экономического прогнозирования, их применение для анализа инвестиционной активности в регионах РФ остается малоизученной областью. Большинство существующих исследований фокусируются на прогнозировании общестрановых показателей, таких как ВВП, инфляция, курсы валют и т. д., в то время как региональный разрез зачастую остается без должного внимания (Овчаров, 2021). Вместе с тем учет региональной специфики является ключевым фактором повышения точности и практической применимости прогнозов в условиях неоднородности экономического пространства России.

Настоящее исследование направлено на восполнение данного пробела и ставит своей целью разработку методологии прогнозирования макроэкономических инвестиционных показателей регионов РФ с использованием современных методов машинного обучения.

Для достижения поставленной цели решаются следующие задачи.

1) Формирование репрезентативной выборки статистических данных по 85 регионам РФ за период 2010-2022 годов, включающей показатели инвестиционной активности, экономического развития и социально-демографической ситуации.

2) Проведение предобработки и разведочного анализа данных, выявление ключевых факторов и закономерностей, влияющих на динамику инвестиционных процессов в регионах.

3) Построение и оптимизация ансамблевых моделей машинного обучения (случайный лес, градиентный бустинг, стекинг) и глубоких нейронных сетей (LSTM, GRU) для прогнозирования темпов роста инвестиций в основной капитал, объема привлеченных прямых иностранных инвестиций и других релевантных показателей на горизонте 1-3 лет.

4) Оценка точности и обобщающей способности разработанных моделей на тестовых данных, не участвовавших в обучении, с использованием метрик MAE, RMSE, MAPE и R^2 .

5) Анализ и интерпретация получаемых прогнозов, выработка рекомендаций по их практическому применению при разработке инвестиционной политики и стратегий регионального развития.

Ключевыми показателями, использованными в исследовании, являлись:

1. Валовой региональный продукт (ВРП) на душу населения, млн руб.
2. Объем инвестиций в основной капитал на душу населения, тыс. руб.
3. Объем прямых иностранных инвестиций (ПИИ) на душу населения, долл.

США.

4. Индекс физического объема инвестиций в основной капитал, % к предыдущему году.
5. Доля инвестиций в основной капитал в ВРП, %.
6. Индекс промышленного производства, % к предыдущему году.
7. Объем работ, выполненных по виду деятельности «Строительство», млрд руб.
8. Оборот розничной торговли на душу населения, тыс. руб.
9. Среднедушевые денежные доходы населения, руб. в месяц.
10. Уровень безработицы по методологии МОТ, %.
11. Коэффициент миграционного прироста, на 10 000 человек населения.
12. Инвестиции в основной капитал по видам экономической деятельности (добыча полезных ископаемых, обрабатывающие производства, энергетика, строительство, торговля, транспорт и связь, операции с недвижимостью и др.), млн руб.

Общее число анализируемых показателей составило 68. Для обеспечения сопоставимости данных все стоимостные показатели были приведены к ценам базового 2010 года с использованием индекса-дефлятора ВРП. Было проведено логарифмирование сильно асимметричных переменных, а также устранение выбросов методом межквартильного размаха. Пропущенные значения были восстановлены с помощью интерполяции сплайнами и регрессионного моделирования.

Методология исследования базировалась на комплексном применении передовых методов машинного обучения, включая ансамблевые алгоритмы (случайный лес, градиентный бустинг, стекинг) и глубокие нейронные сети (LSTM, GRU).

На первом этапе проводилась предобработка и разведочный анализ данных с использованием методов описательной статистики, корреляционного анализа и снижения размерности (метод главных компонент, t-SNE). Это позволило выявить ключевые закономерности и факторы, определяющие региональную дифференциацию инвестиционных процессов, а также отобрать наиболее информативные предикторы для прогнозных моделей.

Далее осуществлялось обучение и настройка ансамблевых моделей на основе случайного леса (Random Forest) и градиентного бустинга (LightGBM, XGBoost). Гиперпараметры моделей подбирались с использованием байесовской оптимизации и перекрестной проверки по к-блокам (blocked k-fold cross-validation) для учета автокорреляции во временных рядах.

В процессе работы были использован следующий математический аппарат:

Формула случайного леса (Random Forest):

$$f(x) = \frac{1}{N} \sum_{i=1}^N T_i(x)$$

где $f(x)$ – итоговый прогноз ансамбля,

$T_i(x)$ – прогноз i -го дерева решений,

N – число деревьев в ансамбле.

Формула градиентного бустинга (Gradient Boosting):

$$F(x) = \sum_{i=1}^M \gamma_i h_i(x)$$

где $F(x)$ – итоговый прогноз ансамбля,

$h_i(x)$ – базовый алгоритм (дерево решений) на i -й итерации,

γ_i – коэффициент усиления на i -й итерации,

M – число итераций алгоритма.

Формула стекинга (Stacking):

$$\hat{y} = \sum_{i=1}^K w_i f_i(x)$$

где \hat{y} – итоговый прогноз ансамбля,
 $f_i(x)$ – прогноз i -го базового алгоритма,
 w_i – вес i -го базового алгоритма,
 K – число базовых алгоритмов.

Формула ячейки LSTM (Long Short-Term Memory):

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ \tilde{C}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ h_t &= o_t \odot \tanh(C_t) \end{aligned}$$

где i_t, f_t, o_t – вектора вентилях входа, забывания и выхода,
 C_t – вектор состояния ячейки,
 \tilde{C}_t – кандидат на обновление состояния ячейки,
 h_t – выходной вектор скрытого состояния,
 W_i, W_f, W_o, W_c – матрицы весов,
 b_i, b_f, b_o, b_c – векторы смещений,
 σ – сигмоидная функция активации,
 \tanh – гиперболический тангенс,
 \odot – поэлементное произведение векторов.

Формула ячейки GRU (Gated Recurrent Unit):

$$\begin{aligned} z_t &= \sigma(W_z[h_{t-1}, x_t]) \\ r_t &= \sigma(W_r[h_{t-1}, x_t]) \\ \tilde{h}_t &= \tanh(W[h_t \odot r_t, x_t]) \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \end{aligned}$$

где z_t – вектор вентиля обновления,
 r_t – вектор вентиля сброса,
 \tilde{h}_t – кандидат на обновление скрытого состояния,
 h_t – выходной вектор скрытого состояния,
 W_z, W_r, W – матрицы весов,
 σ – сигмоидная функция активации,
 \tanh – гиперболический тангенс,
 \odot – поэлементное произведение векторов.

Метрики оценки качества прогнозов:

Средняя абсолютная ошибка (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Средняя квадратичная ошибка (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Средняя абсолютная процентная ошибка (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

Коэффициент детерминации (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

где y_i – фактическое значение целевой переменной,
 \hat{y}_i – прогнозное значение целевой переменной,
 \bar{y} – среднее значение целевой переменной,
 n – число наблюдений в выборке.

Проведенный анализ позволил выявить ряд значимых закономерностей и факторов, определяющих динамику инвестиционных процессов в регионах Российской Федерации. Применение методов снижения размерности, таких как метод главных компонент (PCA) и t-SNE, продемонстрировало наличие устойчивых кластеров регионов, характеризующихся схожими паттернами инвестиционной активности и социально-экономического развития (Китрар, 2021).

В частности, было идентифицировано 5 основных кластеров:

1) высокоразвитые регионы с диверсифицированной экономикой и значительным инвестиционным потенциалом (г. Москва, г. Санкт-Петербург, Московская область, Республика Татарстан);

2) регионы с преобладанием добывающей промышленности и высокой зависимостью от экспорта сырьевых ресурсов (Ханты-Мансийский АО, Ямало-Ненецкий АО, Сахалинская область);

3) промышленно развитые регионы с высокой долей обрабатывающих производств (Свердловская область, Челябинская область, Красноярский край);

4) регионы с развитым агропромышленным комплексом и туристско-рекреационным потенциалом (Краснодарский край, Ростовская область, Республика Крым);

5) депрессивные регионы с низким уровнем экономического развития и инвестиционной привлекательности (Республика Тыва, Республика Алтай, Еврейская автономная область) (Chursin, 2019).

Корреляционный анализ выявил наличие сильных положительных взаимосвязей между объемом инвестиций в основной капитал и такими показателями, как ВРП на душу населения ($r=0.86$), объем промышленного производства ($r=0.79$), среднедушевые доходы населения ($r=0.74$) и объем строительных работ ($r=0.71$). В то же время, была обнаружена умеренная отрицательная корреляция между уровнем инвестиций и показателями безработицы ($r=-0.54$) и миграционного оттока населения ($r=-0.48$) (Опекунов, 2019). Данные результаты свидетельствуют о ключевой роли инвестиций в обеспечении экономического роста, повышении уровня жизни населения и создании новых рабочих мест в регионах.

Среди отраслевых факторов, оказывающих наибольшее влияние на инвестиционную активность, выделяются объемы инвестиций в добычу полезных ископаемых, обрабатывающие производства, энергетику и строительство. Так, в 2022 году на долю этих четырех видов экономической деятельности приходилось 67,4% от общего объема инвестиций в основной капитал в РФ (Астраханцева, 2020). При этом наблюдается значительная региональная дифференциация по структуре инвестиций: если в сырьевых регионах преобладают инвестиции в добывающую промышленность (до 80% в ХМАО и ЯНАО), то в промышленно развитых субъектах РФ ключевую роль

играют обрабатывающие производства (до 40% в Свердловской и Челябинской областях) (Столбов, 2019).

Применение ансамблевых методов машинного обучения, таких как случайный лес, градиентный бустинг и стекинг, позволило построить высокоточные прогнозные модели динамики инвестиционных показателей в регионах РФ. Наилучшие результаты были получены при использовании стекинг-ансамбля, объединяющего прогнозы базовых моделей случайного леса (Random Forest), LightGBM и CatBoost (Grosheva, 2019). На тестовой выборке за период 2020-2022 годов данный метод продемонстрировал значения средней абсолютной ошибки (MAE) в размере 7,9% для прогнозирования темпов роста инвестиций в основной капитал и 11,2% для объема привлеченных прямых иностранных инвестиций. Коэффициент детерминации (R²) составил 0.93 и 0.87 соответственно, что свидетельствует о высокой объясняющей способности модели Клейнер, 2020.

Сравнительный анализ различных архитектур глубоких нейронных сетей показал превосходство рекуррентных моделей с долгой краткосрочной памятью (LSTM) и управляемым рекуррентным блоком (GRU) над традиционными полносвязными сетями прямого распространения (MLP). Наилучшая из рассмотренных конфигураций на основе двухслойной LSTM-сети с размером скрытого слоя 128 и функцией активации ReLU обеспечила значения MAE=6.2% и MAPE=9.4% для прогнозирования индекса физического объема инвестиций в основной капитал на горизонте 1 года (Смирнов, 2015). При этом точность прогнозов существенно снижалась при увеличении периода упреждения до 2-3 лет, что может быть связано с нестационарностью и структурными сдвигами в динамике анализируемых показателей.

Агрегирование прогнозов, полученных с помощью ансамблевых моделей и глубоких нейронных сетей, методом взвешенного голосования позволило дополнительно повысить точность и робастность результатов. Итоговая гибридная модель, использующая прогнозы стекинг-ансамбля и LSTM-сети с весами 0.6 и 0.4 соответственно, продемонстрировала значение MAE=5.6% для прогнозирования темпов роста инвестиций в основной капитал на горизонте 1 года для всей совокупности регионов РФ (Колюжнов, 2020). При этом на уровне отдельных субъектов федерации точность прогнозов варьировалась от 3,2% для Москвы до 14,7% для Еврейской автономной области, что отражает объективные различия в волатильности и предсказуемости инвестиционных процессов в регионах.

Анализ значимости предикторов в рамках построенных моделей машинного обучения позволил определить ключевые факторы и драйверы инвестиционной активности в регионах РФ. Наибольший вклад в точность прогнозов вносили показатели ВРП на душу населения, объема промышленного производства, доходов населения, объема экспорта, а также индикаторы, характеризующие развитие транспортной, энергетической и социальной инфраструктуры (Пестова, 2017). Одновременно с этим была выявлена относительно низкая прогностическая ценность некоторых традиционно используемых в эконометрических моделях переменных, таких как уровень инфляции, процентные ставки и обменный курс рубля. Данный результат может объясняться нелинейным и неоднозначным характером влияния этих факторов на инвестиционные решения экономических агентов в условиях структурных дисбалансов и институциональных ограничений, характерных для российской экономики (Байбуза, 2018).

Полученные прогнозы инвестиционной активности в регионах РФ на период 2023-2025 годов свидетельствуют о высокой вероятности дальнейшего усиления пространственной неоднородности и дивергенции регионального развития. Согласно

базовому сценарию, темпы роста инвестиций в основной капитал в среднем по стране составят 2,3% в 2023 году, 3,1% в 2024 году и 3,8% в 2025 году., при этом для регионов-лидеров (г. Москва, Московская область, г. Санкт-Петербург, Республика Татарстан) прогнозируется опережающая динамика на уровне 5,5-7,2% в годовом выражении (Akhmatov, 2020). В то же время, для ряда субъектов РФ, особенно в Северо-Кавказском и Дальневосточном федеральных округах, ожидается стагнация или даже снижение инвестиционной активности на фоне сохраняющихся структурных ограничений и дефицита качественных инвестиционных проектов (Grosheva, 2019).

Заключение

С учетом выявленных закономерностей и прогнозных оценок представляется целесообразной разработка дифференцированных мер государственной инвестиционной политики, учитывающих специфику и потенциал отдельных регионов и макрорегионов. Приоритетными направлениями такой политики могут стать стимулирование развития региональной инфраструктуры, формирование кластеров и зон опережающего развития в перспективных отраслях экономики, предоставление налоговых и административных преференций для инвесторов, реализующих проекты в отстающих регионах, а также повышение эффективности использования механизмов государственно-частного партнерства. При этом ключевым условием успешности реализации данных мер является обеспечение координации и согласованности действий федеральных и региональных органов власти, институтов развития, бизнес-сообщества и других заинтересованных сторон в рамках единой системы стратегического планирования и проектного управления.

Список литературы

1. Белокур А.А. Экономическая сущность инвестиционной привлекательности региона // Экономика и социум. 2017. № 3. С. 44-50.
2. Астраханцева И.А., Кутузова А.С., Астраханцев Р.Г. Рекуррентные нейронные сети для прогнозирования региональной инфляции. Научные труды Вольного экономического общества России. 2020. Т. 223. № 3. С. 420-431.
3. Байбуза И. Прогнозирование инфляции с помощью методов машинного обучения. Деньги и кредит. 2018. № 4. С. 42-59.
4. Боркова Е.А., Осипова К.А., Светловидова Е.В., Фролова Е.В. Цифровизация экономики на примере банковской системы // Креативная экономика. 2019. № 6. С. 1153-1162.
5. Власов М.В. Цифровая экономика как фактор развития инвестиций в основной капитал в региональных социально-экономических системах // Вестник Пермского университета. Серия: Экономика. 2019. Т. 14. № 3. С. 421-433.
6. Герасимова С.В., Борщ Л.М. Оценка инвестиционных ресурсов региона в стратегическом контексте // Региональная экономика. Юг России. 2019. Т. 7. № 1. С. 112-123.
7. Китрар Л.А. Развитие композитных индикаторов циклического реагирования в конъюнктурных обследованиях // Вопросы статистики. 2021. Т. 28. № 2. С. 24-41.
8. Клейнер Г.Б. Спиральная динамика, системные циклы и новые организационные модели: перламутровые предприятия // Российский журнал менеджмента. 2020. Т. 18. № 4. С. 471-496.
9. Колюжнов Д.В. Инвестиции, инновации, экономический рост и монетарная политика в рамках динамических моделей общего экономического равновесия. Постановка проблемы: глава в монографии Инвестиционный процесс и структурная

трансформация российской экономики // Под ред. А.В. Алексеева, Л.К. Казанцевой. Новосибирск: ИЭОПП СО РАН, 2020. С. 340-361.

10. Кулагина Н.А., Аношина Ю.Ф., Чмаро А.В. Механизм комплексной оценки инвестиционного развития регионов России для разработки инструментов управления бизнес-климатом // Естественно-гуманитарные исследования. 2021. № 36 (4). С. 147-152.

11. Ларченко Л.В., Родионов Д.Г., Жарская Х.В. Государственно-частное партнерство как форма реализации крупных инвестиционных проектов городского развития // Инновации. 2021. № 6(272). С. 61-67.

12. Насрутдинов М.Н. Методологический подход к оценке эффективности инвестиционной деятельности регионов на основе data envelopment analysis // Вестник Алтайской академии экономики и права. 2020. № 4-1. С. 105-112.

13. Овчаров А.О. Периодизация развития сельского хозяйства РФ в контексте влияния экономических кризисов // АПК: экономика, управление. 2021. № 1. С. 62-70.

14. Опекунов А.Н. Принципы формирования моделей прогнозирования вероятности банкротства предприятий с использованием элементов машинного обучения // Модели, системы, сети в экономике, технике, природе и обществе. 2019. № 4 (32). С. 24-31.

15. Перепелкин И.Г., Старых С.А., Соловьев С.В., Бароян А.А. Оценка эффективности применения цифровых технологий в современной экономике // Регион: системы, экономика, управление. 2020. № 2 (49). С. 130-137.

16. Прохорова М.В., Скобелева Е.И. Проектировщики индивидуальных образовательных траекторий: новые тенденции в развитии образовательного пространства // Актуальные вопросы экономики, менеджмента и инноваций: Материалы Международной научно-практической конференции, Нижний Новгород, 19 ноября 2019 года. Н. Новгород: Нижегородский государственный технический университет имени Р.Е. Алексеева, 2019. С. 104-106

17. Пестова А.А. Разработка системы индикаторов финансовой нестабильности на основе высокочастотных данных // Деньги и кредит. 2017. № 6. С. 49-58.

18. Родионов Д.Г., Карпенко П.А., Конников Е.А. Концептуальная модель управления развитием региональных социально-экономических систем // Экономические науки. 2021. № 197. С. 163-170.

19. Родионов Д.Г., Сиднева А.Л. Оценка инвестиционного потенциала Челябинской области // Региональная экономика: теория и практика. 2021. Т. 19. № 8 (491). С. 1517-1541.

20. Серова Н.А. Комплексная оценка эффективности инвестиционной политики арктических регионов России // Север и рынок: формирование экономического порядка. 2021. № 1(71). С. 26-37.

21. Смирнов С.В. Экономический рост и экономические кризисы в России: конец 1920-х годов — 2014 год // Вопросы экономики. 2015. № 5. С. 28-47.

22. Столбов М.И. Индекс финансового стресса для России: новые подходы // Экономический журнал Высшей школы экономики. 2019. Т. 23. № 1. С. 32-60.

23. Сучкова Е.О. Методология и практика реализации макропруденциального стресс-тестирования банковской системы // Вестник Московского университета. Серия 6: Экономика. 2017. № 1. С. 123-146.

24. Трофимова Н. Н. Инновационная модель развития социально-экономического пространства // Альманах Крым. 2021. № 24. С. 113-117. (In Russ).

25. Тюлин А.Е., Чурсин А.А., Юдин А.В., Грошева П.Ю. Теоретические основы закона управления опережающим развитием организации // Микроэкономика. 2019. № 1. С. 5-12.

26. Якушев Н.О., Мазилев Е.А. Методический подход к оценке инвестиционной привлекательности локальной территории региона // Проблемы развития территории. 2020. № 4(108). С. 68-87.

27. Akhmatov K.A., Astrakhandseva I.A., Kutuzova A.S., Votchel L.M., Vikulina V.V. Harmonization of banking business models with the needs of the economy by encouraging the exogenous social responsibility. Journal QualityAccess to Success. 2020. Vol. 21. № 174. P. 81-87.

28. Chursin R.A., Yudin A.V., Grosheva P.Yu., Filippov P.G., Butrova E.V. Tool for Assessing the Risks of R&D Projects Implementation in High-tech Enterprises // IOP Conference Series: Materials Science and Engineering: Workshop on Materials and Engineering in Aeronautics. Moscow, 2019. p. 012005.

29. Grosheva P.Yu., Yudin A.V., Myakishev Yu.D. Risk-based forecasting methods of knowledge-intensive product life-cycle resource provision // IOP Conference Series: Materials Science and Engineering. Krasnoyarsk, 2019. p. 42-84.

30. Tyulin A., Chursin A., Yudin A. Production capacity optimization in cases of a new business line launching in a company // Espacios. 2017. № 62. p. 20.

References

1. Belokur A.A. Jekonomicheskaja sushhnost' investicionnoj privilekatel'nosti regiona // Jekonomika i socium. 2017. № 3. p. 44-50. (In Russ).

2. Astrahanceva I.A., Kutuzova A.S., Astrahancev R.G. Rekurrentnye nejronnye seti dlja prognozirovaniya regional'noj infljicii. Nauchnye trudy Vol'nogo jekonomicheskogo obshhestva Rossii. 2020. T. 223. № 3. P. 420-431. (In Russ).

3. Bajbuza I. Prognozirovanie infljicii s pomoshh'ju metodov mashinnogo obuchenija. Den'gi i kredit. 2018. № 4. p. 42-59. (In Russ).

4. Borkova E.A., Osipova K.A., Svetlovidova E.V., Frolova E.V. Cifrovizacija jekonomiki na primere bankovskoj sistemy. Kreativnaja jekonomika. 2019. № 6. p. 1153-1162. (In Russ).

5. Vlasov M.V. Cifrovaja jekonomika kak faktor razvitija investicij v osnovnoj kapital v regional'nyh social'no-jekonomicheskijh sistemah. Vestnik Permskogo universiteta. Serija: Jekonomika. 2019. T. 14. № 3. p. 421-433. (In Russ).

6. Gerasimova S.V., Borshh L.M. Ocenka investicionnyh resursov regiona v strategicheskome kontekste // Regional'naja jekonomika. Jug Rossii. 2019. T. 7. № 1. p. 112-123. (In Russ).

7. Kitrar L.A. Razvitie kompozitnyh indikatorov ciklicheskogo reagirovanija v kon#junkturyh obsledovanijah. Voprosy statistiki. 2021. T. 28. № 2. p. 24-41. (In Russ).

8. Klejner G.B. Spiral'naja dinamika, sistemnye cikly i novye organizacionnye modeli: perlamutrovye predpriyatija. Rossijskij zhurnal menedzhmenta. 2020. T. 18. № 4. p. 471-496. (In Russ).

9. Koljuzhnov D.V. Investicii, innovacii, jekonomicheskij rost i monetarnaja politika v ramkah dinamicheskijh modelej obshhego jekonomicheskogo ravnovesija. Postanovka problemy: glava v monografii Investicionnyj process i strukturnaja transformacija rossijskoj jekonomiki. Pod red. A.V. Alekseeva, L.K. Kazancevoj. Novosibirsk: IJeOPP SO RAN, 2020. p. 340-361. (In Russ).

10. Kulagina N.A., Anoshina Ju.F., Chmaro A.V. Mehanizm kompleksnoj ocenki investicionnogo razvitija regionov Rossii dlja razrabotki instrumentov upravlenija biznes-klimatom. Estestvenno-gumanitarnye issledovanija. 2021. № 36 (4). p. 147-152. (In Russ).

11. Larchenko L.V., Rodionov D.G., Zharskaja H.V. Gosudarstvenno-chastnoe partnerstvo kak forma realizacii krupnyh investicionnyh proektov gorodskogo razvitiya . Innovacii. 2021. № 6(272). p. 61-67. (In Russ).
12. Nasrutdinov M.N. Metodologicheskij podhod k ocenke jeffektivnosti investicionnoj dejatel'nosti regionov na osnove data envelopment analysis. Vestnik Altajskoj akademii jekonomiki i prava. 2020. № 4-1. p. 105-112. (In Russ).
13. Ovcharov A.O. Periodizacija razvitiya sel'skogo hozjajstva RF v kontekste vlijanija jekonomicheskikh krizisov. APK: jekonomika, upravlenie. 2021. № 1. p. 62-70. (In Russ).
14. Opekunov A.N. Principy formirovanija modelej prognozirovaniya verojatnosti bankrotstva predpriyatij s ispol'zovaniem jelementov mashinnogo obuchenija. Modeli, sistemy, seti v jekonomike, tehnike, prirode i obshhestve. 2019. № 4 (32). p. 24-31. (In Russ).
15. Perepelkin I.G., Staryh S.A., Solov'ev S.V., Barojan A.A. Ocenka jeffektivnosti primenenija cifrovych tehnologij v sovremennoj jekonomike. Region: sistemy, jekonomika, upravlenie. 2020. № 2 (49). p. 130-137. (In Russ).
16. Prohorova M.V., Skobeleva E.I. Proektirovshhiki individual'nyh obrazovatel'nyh traektorij: novye tendencii v razviii obrazovatel'nogo prostranstva. Aktual'nye voprosy jekonomiki, menedzhmenta i innovacij: Materialy Mezhdunarodnoj nauchno-prakticheskoy konferencii, Nizhnij Novgorod, 19 nojabrja 2019 goda. N. Novgorod: Nizhegorodskij gosudarstvennyj tehničeskij universitet imeni R.E. Alekseeva, 2019. p. 104-106.(In Russ).
17. Pestova A.A. Razrabotka sistemy indikatorov finansovoj nestabil'nosti na osnove vysokochastotnyh dannyh. Den'gi i kredit. 2017. № 6. p. 49-58. (In Russ).
18. Rodionov D.G., Karpenko P.A., Konnikov E.A. Konceptual'naja model' upravlenija razvitiem regional'nyh social'no-jekonomicheskikh system. Jekonomicheskie nauki. 2021. № 197. p. 163-170. (In Russ).
19. Rodionov D.G., Sidneva A.L. Ocenka investicionnogo potenciala Cheljabinskoy oblasti // Regional'naja jekonomika: teorija i praktika. 2021. T. 19. № 8 (491). p. 1517-1541. (In Russ).
20. Serova N.A. Kompleksnaja ocenka jeffektivnosti investicionnoj politiki arkticheskikh regionov Rossii. Sever i rynek: formirovanie jekonomicheskogo porjadka. 2021. № 1(71). p. 26-37. (In Russ).
21. Smirnov S.V. Jekonomicheskij rost i jekonomicheskie krizisy v Rossii: konec 1920-h godov — 2014 god. Voprosy jekonomiki. 2015. № 5. p. 28-47. (In Russ).
22. Stolbov M.I. Indeks finansovogo stressa dlja Rossii: novye podhody // Jekonomicheskij zhurnal Vysshej shkoly jekonomiki. 2019. T. 23. № 1. p. 32-60. (In Russ).
23. Suchkova E.O. Metodologija i praktika realizacii makroprudencial'nogo stress-testirovanija bankovskoj sistemy. Vestnik Moskovskogo universiteta. Serija 6: Jekonomika. 2017. № 1. p. 123-146. (In Russ).
24. Trofimova N. N. Innovacionnaja model' razvitiya social'no-jekonomicheskogo prostranstva. Al'manah Krym. 2021. № 24. p. 113-117. (In Russ).
25. Tjul'in A.E., Chursin A.A., Judin A.V., Grosheva P.Ju. Teoreticheskie osnovy zakona upravlenija operezhajushhim razvitiem organizacii. Mikrojekonomika. 2019. № 1. p. 5-12. (In Russ).
26. Jakushev N.O., Mazilov E.A. Metodicheskij podhod k ocenke investicionnoj privilekatel'nosti lokal'noj territorii regiona. Problemy razvitiya territorii. 2020. № 4(108). p. 68-87. (In Russ).
27. Akhmatov K.A., Astrakhantseva I.A., Kutuzova A.S., Votchel L.M., Vikulina V.V. Harmonization of banking business models with the needs of the economy by

encouraging the exogenous social responsibility. Journal Quality Access to Success. 2020. Vol. 21. № 174.

P. 81-87. (In Russ).

28. Chursin R.A., Yudin A.V., Grosheva P.Yu, Filippov P.G., Butrova E.V. Tool for Assessing the Risks of R&D Projects Implementation in High-tech Enterprises. IOP Conference Series: Materials Science and Engineering: Workshop on Materials and Engineering in Aeronautics. Moscow, 2019. p. 012005. (In Russ).

29. Grosheva P.Yu., Yudin A.V., Myakishev Yu.D. Risk-based forecasting methods of knowledge-intensive product life-cycle resource provision. IOP Conference Series: Materials Science and Engineering. Krasnoyarsk, 2019. p. 42-84. (In Russ).

30. Tyulin A., Chursin A., Yudin A. Production capacity optimization in cases of a new business line launching in a company. Espacios. 2017. № 62. p. 20.

Influence of socio-economic factors on the investment attractiveness of Russian regions: using machine learning methods

Mstislav A. Ivanov

The Plekhanov Russian University of Economics

Moscow, Russia

Mstislav.ivanoff@yandex.ru

Original article

Received 22.11.2023

Revised 01.02.2024

Accepted 29.03.2024.

Abstract

The investment attractiveness of the regions of the Russian Federation is a key factor determining the effectiveness of the country's economic development. In this regard, it is relevant to study the influence of socio-economic factors on the investment climate of the subjects of the Russian Federation. The purpose of this work is to analyze the relationship between various socio-economic indicators and the investment attractiveness of Russian regions using modern machine learning methods. Statistical indicators characterizing the socio-economic development of 85 subjects of the Russian Federation for the period from 2010 to 2022 were used as initial data. Among them: gross regional product (GRP), unemployment rate, per capita income of the population, volume of investments in fixed assets, industrial production index, etc. A rating compiled by the Expert RA rating agency was used to assess the investment attractiveness of the regions. The data was analyzed using machine learning methods such as multiple linear regression, decision trees, random forest, and gradient boosting. To select the most significant predictors, feature selection methods were used, in particular, Recursive Feature Elimination and Principal Component Analysis. The analysis showed that such factors as GRP per capita (correlation coefficient 0.78), the volume of investments in fixed assets (0.75), the unemployment rate (-0.69) and the industrial production index (0.64) have the greatest impact on the investment attractiveness of the regions of the Russian Federation. Machine learning models based on selected features have demonstrated high accuracy in predicting the investment rating of regions. The best results were obtained using the gradient boosting algorithm (the coefficient of determination $R^2=0.87$

in the test sample). At the same time, the indicators of GRP per capita, investment volume and unemployment rate make the greatest contribution to the model, which confirms their key role in shaping the investment attractiveness of the subjects of the Russian Federation.

Keywords

investment attractiveness, regions of the Russian Federation, socio-economic factors, machine learning, multiple regression, gradient boosting.

© Иванов М.А., 2024.