

## РЕГИОНАЛЬНЫЕ АСПЕКТЫ РАЗВИТИЯ ЭКОНОМИКИ

УДК 336.6

### ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ПРОЦЕССАХ УПРАВЛЕНИЯ ИНВЕСТИЦИОННОЙ ПРИВЛЕКАТЕЛЬНОСТЬЮ РЕГИОНА

*Хомякова Анна Александровна (xotakova@mail.ru)*

*Мизгирев Лев Сергеевич*

*Шергин Владимир Владимирович*

*Ивановский государственный химико-технологический университет*

В статье отражены результаты использования различных алгоритмов интеллектуального анализа данных в качестве инструмента поддержки принятия решения в процедурах управления инвестиционной привлекательностью региона с целью обеспечения качества и обоснованности выбора основных мер её повышения. Предложена функциональная модель процесса управления инвестиционной привлекательностью региона, включающая подпроцессы анализа факторов инвестиционной привлекательности, разработки положений стратегии развития инвестиционной привлекательности региона и формирования региональной инвестиционной программы. С целью анализа факторов и оценки инвестиционной привлекательности региона предложено использовать алгоритм «случайный лес».

**Ключевые слова:** интеллектуальный анализ данных, нейронная сеть, управление инвестиционной привлекательностью региона, экономика региона.

Управление инвестиционной привлекательностью направлено на выработку мер по повышению эффективности региональной экономики. При этом решения, принятые в ходе управления инвестиционной привлекательностью региона, находят своё отражение в стратегии его социально-экономического развития и в формируемой региональной инвестиционной программе. Так, принимая объективные решения о наиболее целесообразных мерах поддержки того или иного сектора экономики региона, выбирая наиболее перспективные точки экономического роста, включая тот или иной проект в региональную инвестиционную программу, уполномоченные органы государственного управления решают задачу выбора наиболее привлекательной альтернативы из ряда имеющихся. В рассматриваемом случае, из ряда альтернатив выбор необходимо делать в пользу той, которая обеспечивает наибольший рост инвестиционной привлекательности региона. В контексте проводимого нами исследования интерес представляет обоснование возможности использования методов интеллектуального анализа данных как инструмента поддержки принятия управленческих решений, обеспечивающего прогнозирование изменения инвестиционной привлекательности региона в результате реализации того или иного варианта управленческого решения.

Инвестиционная привлекательность региона отражает уровень его развития и степень соответствия интересам инвесторов и зависит от большого количества разнообразных внешних и внутренних факторов. Отметим, что в экономической литературе наблюдаются

различные подходы к группировке факторов, влияющих на инвестиционную привлекательность региона. Так, в своей научной статье Белокур А.А. [1] выделяет следующие группы факторов:

1) Факторы, определяющие экономический потенциал региональной экономической системы (обеспеченность региона ресурсами, биоклиматический потенциал, наличие свободных земель для производственного инвестирования, уровень обеспеченности энерго- и трудовыми ресурсами, развитость научно-технического потенциала и инфраструктуры);

2) Факторы, характеризующие общие условия хозяйствования (экологическая безопасность, развитие отраслей материального производства, объем незавершенного строительства, развитие строительной базы);

3) Факторы, свидетельствующие о зрелости рыночной среды в регионе (развитость рыночной инфраструктуры, емкость местного рынка сбыта, экспортные возможности);

4) Политические факторы (степень доверия населения к региональной власти, взаимоотношения федерального центра и властей региона, состояние национально-религиозных отношений);

5) Социальные и социокультурные факторы (уровень жизни населения, распространенность алкоголизма и наркомании, уровень преступности, величина реальной заработной платы, условия работы для иностранных специалистов);

6) Организационно-правовые факторы;

7) Финансовые факторы (доходы бюджета, обеспеченность средствами внебюджетных

фондов на душу населения, уровень банковского процента, развитость межбанковского сотрудничества).

Таким образом, актуальность разработки теоретических и методических аспектов использования методов интеллектуального анализа данных в процедурах принятия решений по управлению инвестиционной привлекательностью региона связана с возможностью учесть влияние на изменение инвестиционной привлекательности широкого круга разнообразных ситуационных факторов, состояние которых отражается большим объёмом структурированных и неструктурированных данных.

Управление инвестиционной привлекательностью региона представляет собой многоплановый комплекс взаимосвязанных и взаимообусловленных

действий, охватывающих все основные факторы, обеспечивающие инвестиционную привлекательность региона в текущем и перспективном периоде. В сети процессов функционирования региона (рис. 1), рассматриваемого как социально-экономическая система, процесс управления инвестиционной привлекательностью может быть классифицирован как процесс развития.

Как таковые, процессы развития реализуются с целью выработки управленческих воздействий на систему, формирующих новые импульсы для её развития. В ходе проводимого исследования была построена функциональная модель процесса управления инвестиционной привлекательностью региона (рис. 2).

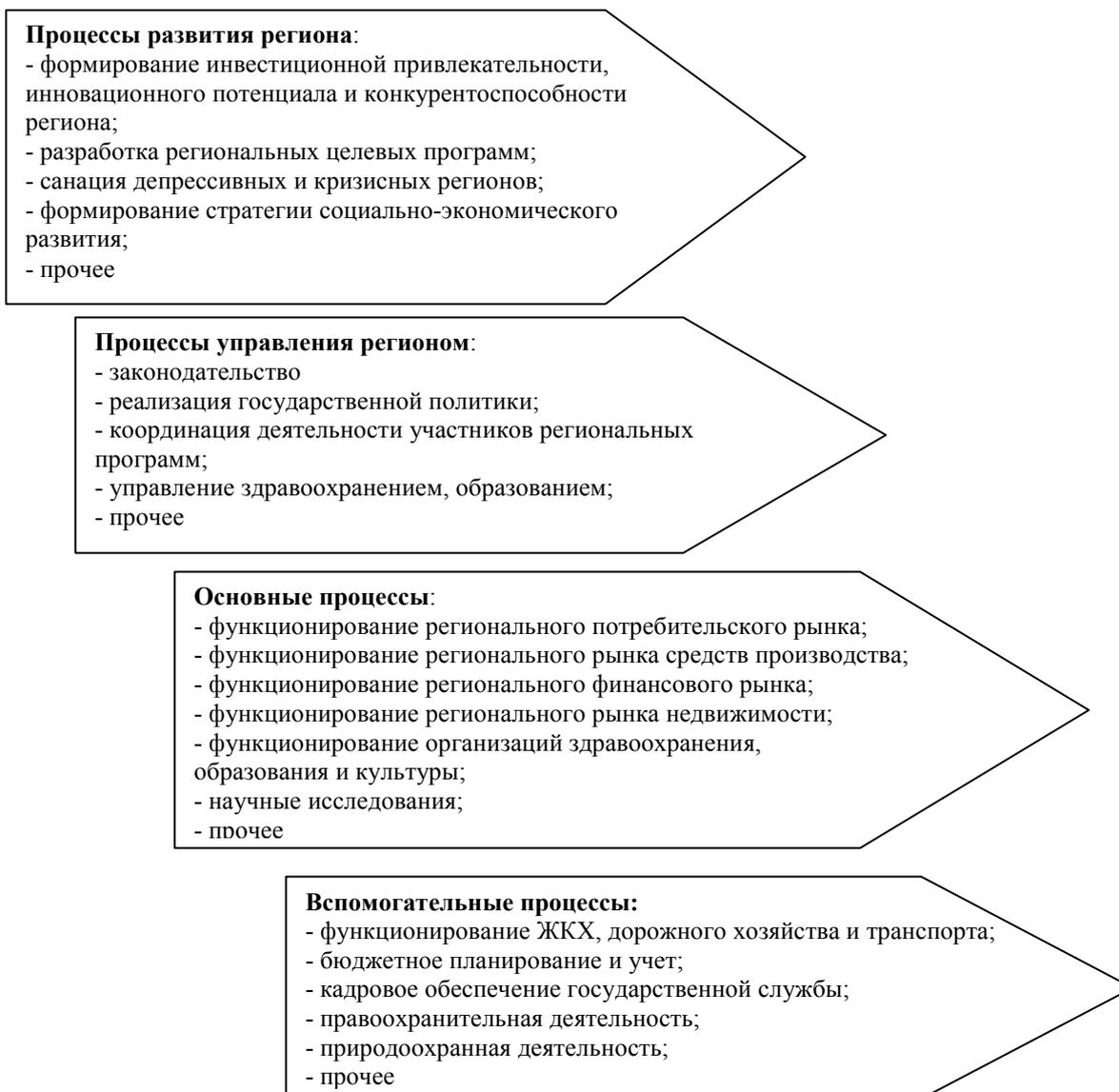
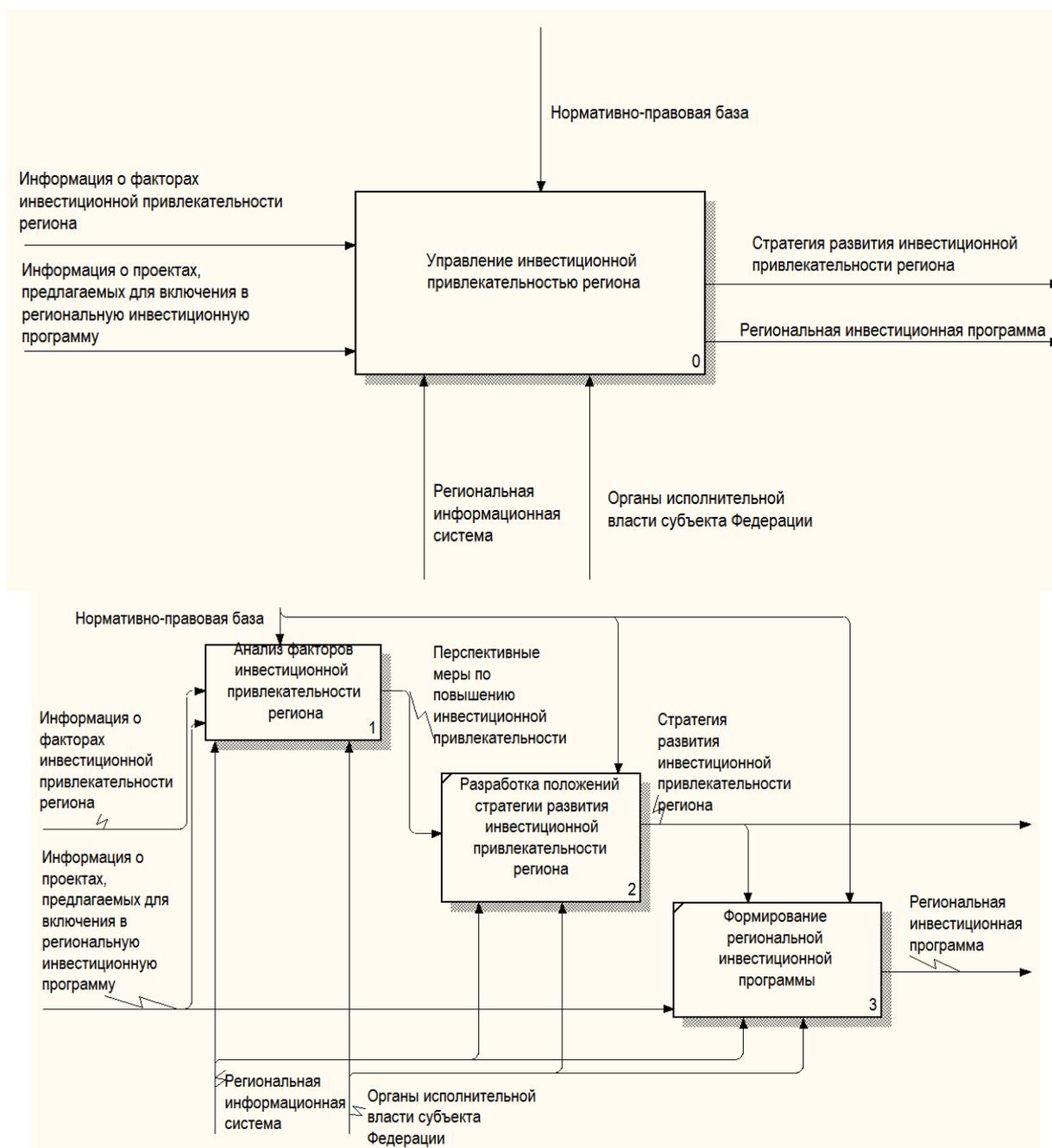


Рисунок 1. Процессы функционирования региональной социально-экономической системы



**Рисунок 2. Функциональная модель процесса управления инвестиционной привлекательностью региона**

В результате управления инвестиционной привлекательностью с учетом социально-экономических, географических, климатических, инфраструктурных и других региональных особенностей разрабатывается стратегия развития инвестиционной привлекательности региона, включающая в себя программу привлечения инвестиционных ресурсов. Кроме того принимаются решения о целесообразности реализации того или иного регионального инвестиционного проекта. Совокупность таких взаимосвязанных по ресурсам, срокам и исполнителям проектов представляет собой региональную инвестиционную программу.

Согласно предложенной модели, процесс управления инвестиционной привлекательностью региона предполагает выполнение следующих подпроцессов:

- анализ изменения факторов инвестиционной привлекательности региона, в результате которого выявляется комплекс перспективных мер по повышению инвестиционной привлекательности. В ходе проводимого анализа выявляются наиболее значимые для данного региона факторы, обуславливающие рост инвестиционной привлекательности, и оценивается влияние

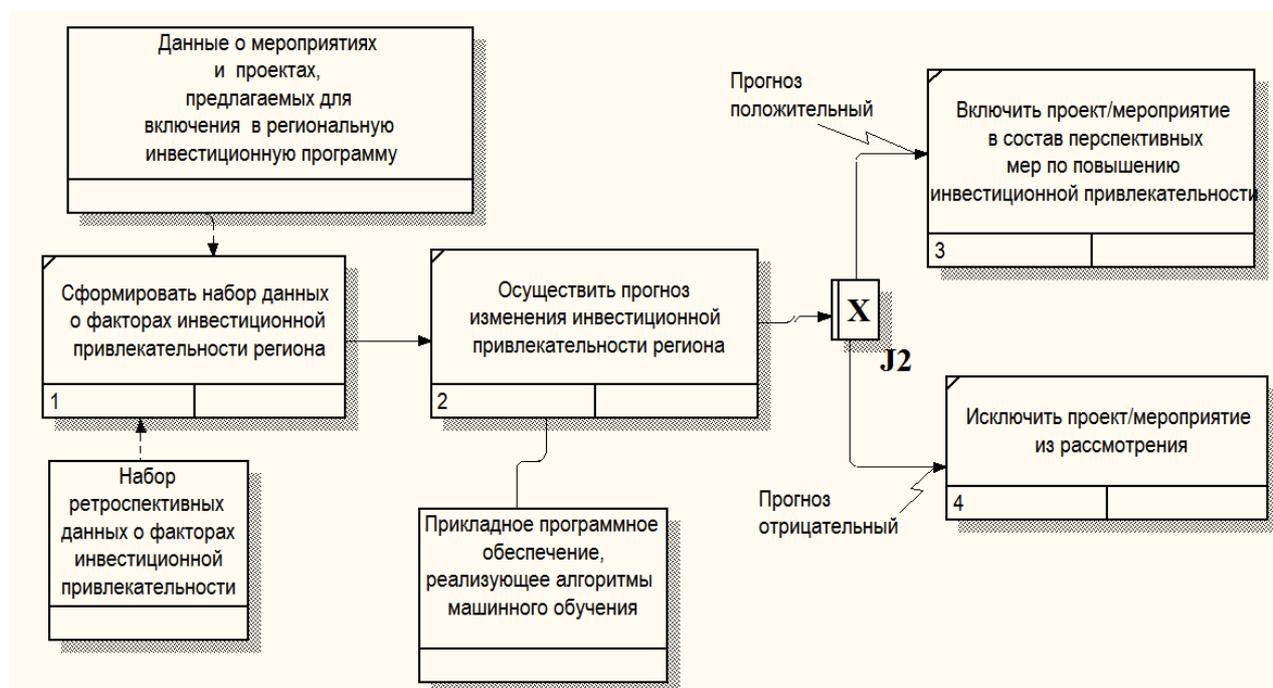
последствий реализации альтернативных мер и проектов на изменение данных факторов;

- *разработка положений стратегии развития инвестиционной привлекательности региона*, которые являются регламентирующим документом в процессе формирования региональной инвестиционной программы, а также учитываются при разработке стратегии социально-экономического развития региона;

- *формирование региональной инвестиционной программы*, включающей комплекс взаимосвязанных по содержанию и используемым ресурсам проектов, имеющих

целевую ориентацию на улучшение социально-экономического положения региона и увеличение его инвестиционной привлекательности (данное определение предложено в исследовании А.В. Кизеева [4]).

Для анализа факторов инвестиционной привлекательности и последующей оценки целесообразности включения того или иного проекта в региональную инвестиционную программу с точки зрения его влияния на инвестиционную привлекательность региона нами предложено использовать методы интеллектуального анализа данных (рис. 3).



**Рисунок 3. Модель процесса анализа факторов инвестиционной привлекательности региона**

На первом этапе анализа факторов инвестиционной привлекательности региона необходимо сформировать набор данных по указанным факторам, содержащий информацию об изменении факторов инвестиционной привлекательности в предшествующих периодах, а также информацию о предлагаемых к осуществлению мероприятиях социально-экономического развития и региональных инвестиционных проектах.

На втором этапе необходимо осуществить прогноз изменения инвестиционной привлекательности региона, учитывающий как сложившиеся ранее тенденции социально-экономического развития, так и результаты реализации предлагаемых альтернативных мероприятий и проектов. Учитывая большое разнообразие факторов, определяющих инвестиционную привлекательность региона, и

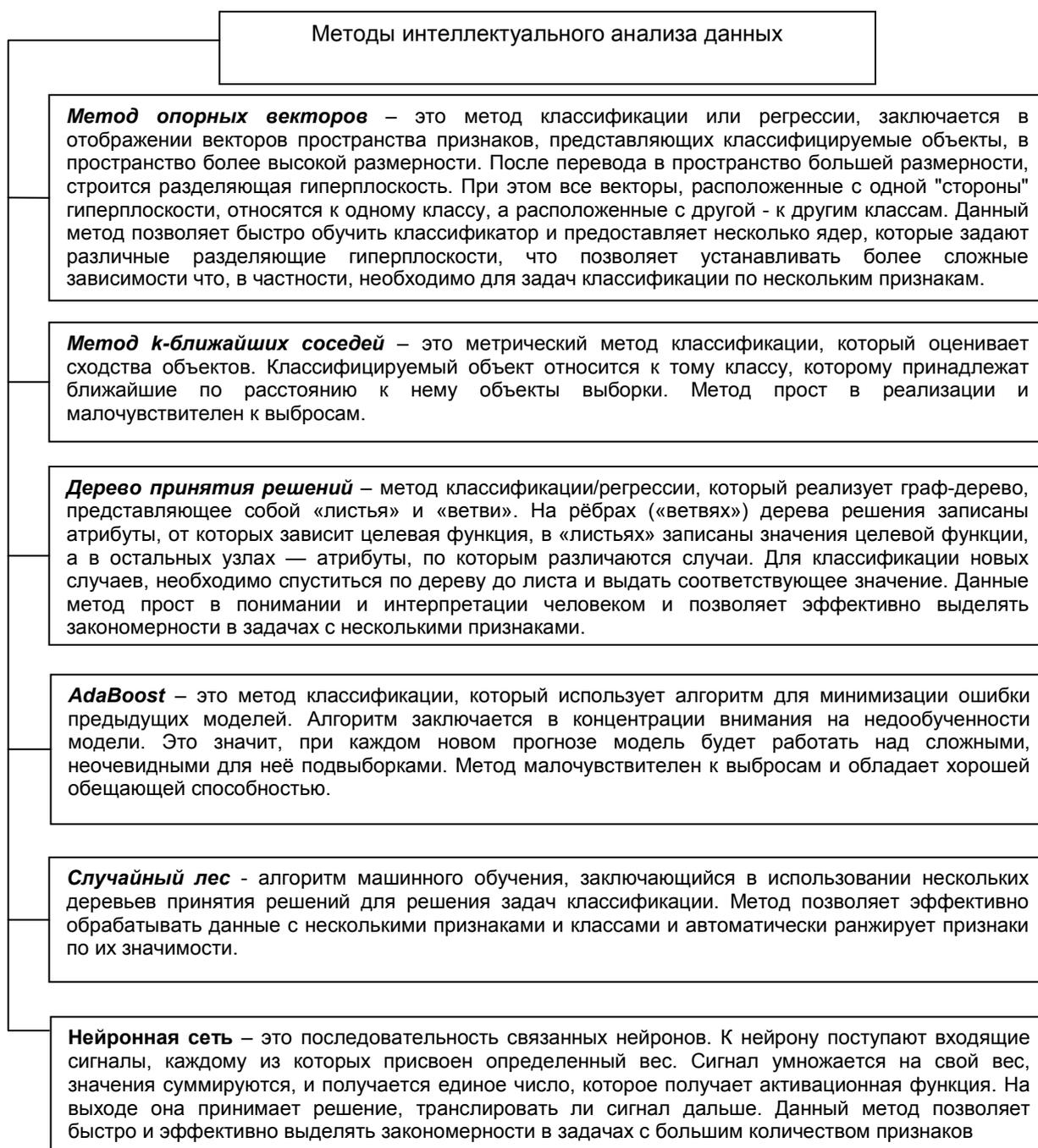
неструктурированность данных о них, представляется целесообразным для осуществления подобного прогноза использовать методы интеллектуального анализа данных. Результатом прогнозирования является ожидаемая классификация региона как инвестиционно привлекательного (положительный прогноз) или инвестиционно не привлекательного (отрицательный прогноз). Если прогноз положительный, то рассматриваемую альтернативу мероприятия или проекта следует включить в состав перспективных мер по повышению инвестиционной привлекательности региона, в противном случае альтернативу следует отклонить.

Таким образом, задача прогнозирования инвестиционной привлекательности региона связана с анализом влияющих факторов и дальнейшим выводом информации о

последующем долго- и краткосрочном поведении величины показателя инвестиционной привлекательности. Изучение ряда научных работ по данной тематике показало, что задачу оценки инвестиционной привлекательности региона с помощью методов интеллектуального анализа данных можно решать как задачу установления

рейтинга объекта, либо как задачу квантования (кластеризации).

В рамках проведённого нами исследования была оценена возможность использования различных методов интеллектуального анализа данных для решения задачи классификации регионов по уровню их инвестиционной привлекательности (рис.4).



**Рисунок 4. Краткая характеристика методов интеллектуального анализа данных, применимых для решения задачи классификации инвестиционной привлекательности региона (составлено авторами на основе Руководство пользователя и документация библиотеки scikit-learn 0.22.2 [7])**

В качестве зависимой переменной в наборе данных об инвестиционной привлекательности регионов использован рейтинг инвестиционной привлекательности региона [2, 3] который отражает мнение Национального Рейтингового

Агентства (НРА) об уровне инвестиционной активности регионов и качества организации условий для развития инвестиционной деятельности (табл.1).

**Таблица 1**

**Классификация регионов по уровню инвестиционной привлекательности по методике НРА**

Уровень	Категория	Характеристика
регионы с высоким уровнем инвестиционной привлекательности	IC1	Высокая инвестиционная привлекательность - первый уровень.
	IC2	Высокая инвестиционная привлекательность - второй уровень.
	IC3	Высокая инвестиционная привлекательность - третий уровень.
регионы со средним уровнем инвестиционной привлекательности	IC4	Средняя инвестиционная привлекательность - первый уровень.
	IC5	Средняя инвестиционная привлекательность - второй уровень.
	IC6	Средняя инвестиционная привлекательность - третий уровень.
регионы с умеренным уровнем инвестиционной привлекательности	IC7	Умеренная инвестиционная привлекательность - первый уровень.
	IC8	Умеренная инвестиционная привлекательность - второй уровень.
	IC9	Умеренная инвестиционная привлекательность - третий уровень.

Методика НРА основана на совокупности качественных и количественных показателей, объединенных в семь факторов: природные ресурсы и географическое положение, трудовые ресурсы региона, региональная инфраструктура, внутренний рынок (потенциал регионального спроса), производственный потенциал региональной экономики, институциональная среда и социально-политическая стабильность, финансовая устойчивость регионального бюджета и предприятий региона.

В целях упрощения интерпретации полученных в ходе прогнозирования инвестиционной привлекательности результатов, используемые национальным рейтинговым агентством, классы предлагается агрегировать следующим образом:

- регионы с высоким уровнем инвестиционной привлекательности (IC1- IC3);
- регионы со средним уровнем инвестиционной привлекательности (IC4- IC6);
- регионы с умеренным уровнем инвестиционной привлекательности (IC7- IC9).

В качестве меток классификации выбрано 3 класса «Нет», «Среднее», «Привлекательно».

В качестве независимых переменных в наборе данных об инвестиционной

привлекательности регионов нами использованы следующие показатели:

- Численность населения, тыс. чел;
- Среднегодовая численность занятых, тыс. чел;
- Уровень безработицы, %;
- Среднедушевые денежные доходы, руб.;
- Оборот малых предприятий, млрд. руб.;
- Добыча полезных ископаемых, млн. руб.;
- Обрабатывающие производства, млн. руб.;
- Продукция сельского хозяйства, млн. руб.;
- Строительство, млн. руб.;
- Оборот розничной торговли, млн. руб.;
- Оборот оптовой торговли, млн. руб.;
- Объём платных услуг населению, млн. руб.;
- Используемые передовые производственные технологии, ед.;
- Сальдированный финансовый результат, млн.руб.

Источником информации о значениях указанных показателей послужили данные Росстата о социально-экономическом положении регионов за 2016-2017 гг. [5, 6].

Таким образом, в набор независимых переменных включены показатели, характеризующие основные факторы инвестиционной привлекательности региона (природные ресурсы и географическое положение, трудовые ресурсы, региональную инфраструктуру, внутренний рынок, производственный потенциал региональной экономики, институциональную среду и социально-политическую стабильность, финансовую устойчивость регионального бюджета и предприятий региона).

Для решения задачи данные были преобразованы: нормализованы в диапазоне [0, 1], разделены на обучающую (80%) и контрольную выборки (20%), отобраны значимые признаки по критерию хи-квадрат:

- Среднедушевые денежные доходы (income\_per\_capita), руб.;
- Добыча полезных ископаемых (mining), млн. руб.;
- Обрабатывающие производства (refining), млн. руб.;
- Строительство (construction), млн. руб.;
- Используемые передовые производственные технологии (new\_tech), ед.

На рисунке 5 представлен фрагмент обучающей выборки. Первая таблица состоит из 132 строк и 5 столбцов по каждому, выделенному выше, критерию. Вторая таблица - это список из 132 элементов, меток классификации (1 - «Нет», 2 - «Средне», 3 - «Привлекательно»).

```
Index(['income_per_capita', 'mining', 'refining', 'construction', 'new_tech'], dtype='object')
      0      1      2      3      4
0  0.986107  0.012886  0.000117  0.002636  0.009936
1  0.194478  0.000707  0.032522  0.062421  0.100616
2  0.185788  0.000149  0.023791  0.032151  0.044443
3  0.522382  0.515276  0.083005  0.214653  0.094654
4  0.149869  0.076879  0.054369  0.050468  0.055154
...
127 0.065647  0.000309  0.000998  0.006806  0.004944
128 0.616629  0.119383  0.009782  0.092996  0.035719
129 0.066410  0.000102  0.028677  0.037874  0.143750
130 0.093016  0.005561  0.061533  0.058730  0.364125
131 0.018783  0.000289  0.000267  0.011729  0.000388

132 rows x 5 columns
0      3
1      2
2      2
3      2
4      1
..
127    3
128    3
129    1
130    3
131    3
Name: investor_appeal, Length: 132, dtype: int64
```

**Рисунок 5. Фрагмент обучающей выборки**

В ходе исследования нами были опробованы 6 алгоритмов машинного обучения: метод опорных векторов, нейронная сеть, k-ближайших соседей, дерево принятия решений, алгоритм бустинга AdaBoost и случайный лес. У каждой модели есть ряд гиперпараметров:

Метод опорных векторов:

- kernel – вид разделяющей плоскости (linear – линейная; rbf – задана гауссовой радиальной базовой функцией)
- C – параметр регуляризации [8] k-ближайших соседей:
- n\_neighbors – количество точек, по расстоянию от которых, определяется класс новой точки

Дерево принятия решений:

- `max_depth` – максимальная глубина дерева (количество уровней дерева);
- `min_weight_fraction_leaf` - доля входных элементов выборки, которые должны находиться на листовом узле;
- `criterion` – критерий расщепления по признаку (`gini` – вероятность неправильной классификации случайно выбранного элемента в выборке, если ему была присвоена случайным образом метка в соответствии с распределением классов выборки; `entropy` - энтропия)

AdaBoost (на основе дерева принятия решений):

- `n_estimators` - количество базовых моделей (в нашем случае - деревьев решений);
- `learning_rate` – скорость обучения моделей.

Случайный лес:

- `n_estimators` – количество деревьев решений в модели;
- `max_depth` - максимальная глубина дерева (количество уровней дерева);
- `criterion` – критерий расщепления по признаку (см. Дерево принятия решений)

Нейронная сеть:

- `hidden_layer_sizes` определяет архитектуру сети (кол-во нейронов в слое, количество слоёв);

- `solver` – алгоритм градиентного спуска (`lbfgs` – Алгоритм Бройдена — Флетчера — Гольдфарба — Шанно [10]; `sgd` - стохастический градиентный спуск; `adam` – модифицированный стохастический градиентный спуск [9];
- `activation` – функция активации нейронов сети (`relu` – “выпрямитель”  $f(x) = \max(0, x)$ ; `tanh` - гиперболический тангенс);
- `learning_rate` – скорость изменения скорости изучения (`constant` – постоянная, `adaptive` – адаптивный, изменяется в зависимости от истории изменений значений функции ошибки; `invscaling` – скорость постепенно уменьшается с каждой новой эпохой обучения);
- `learning_rate_init` – начальная скорость обучения.

Архитектура нейронной сети подбиралась вручную. Подбор гиперпараметров моделей, кроме нейронной сети осуществлялся с помощью жадного алгоритма (Greed Search). Основные диапазоны поиска для гиперпараметров представлены в таблице 2.

В результате подбора были получены следующие гиперпараметры для моделей (табл. 3).

Таблица 2

### Основные диапазоны поиска оптимальных значений для гиперпараметров

Модель	Диапазон параметров
Метод опорных векторов	{ <code>kernel</code> :('linear', 'rbf'), <code>C</code> : [1, 10]}
k-ближайших соседей	{ <code>n_neighbors</code> : [3, 30]}
Дерево принятия решений	{ <code>max_depth</code> : [3, 50], <code>min_weight_fraction_leaf</code> : [0, 0.5], <code>'criterion'</code> : ('gini', 'entropy')}
AdaBoost (на основе дерева принятия решений)	{ <code>n_estimators</code> : [10, 100], <code>learning_rate</code> : [0.001, 1]}
Случайный лес	{ <code>n_estimators</code> : [10, 1000], <code>max_depth</code> : [10, 100], <code>'criterion'</code> : ('gini', 'entropy')}
Нейронная сеть	{ <code>hidden_layer_sizes</code> : [10, 200], <code>solver</code> : ('lbfgs', 'sgd', 'adam'), <code>activation</code> : ('tanh', 'relu'), <code>learning_rate</code> : ('constant', 'invscaling', 'adaptive'), <code>learning_rate_init</code> : [0.001, 0.05]}

Таблица 3

### Модели с оптимальными гиперпараметрами

Модель	Параметры
Метод опорных векторов	{ <code>C</code> : 3, <code>kernel</code> : 'linear'}
k-ближайших соседей	{ <code>n_neighbors</code> : 8}
Дерево принятия решений	{ <code>'criterion'</code> : 'gini', <code>max_depth</code> : 50, <code>n_estimators</code> : 100}
AdaBoost	{ <code>learning_rate</code> : 0.8, <code>n_estimators</code> : 200}
Случайный лес	{ <code>'criterion'</code> : 'entropy', <code>max_depth</code> : 10, <code>n_estimators</code> : 100}
Нейронная сеть	{ <code>activation</code> : 'tanh', <code>hidden_layer_sizes</code> : 190, <code>learning_rate</code> : 'invscaling', <code>learning_rate_init</code> : 0.01, <code>solver</code> : 'lbfgs'}

В ходе работы проведена оценка эффективности моделей, которая показала, насколько хорошо модели способны различать

классы с помощью матриц ошибок и значения метрик на тестовой выборке. Матрица ошибок - это таблица, которая позволяет

визуализировать эффективность алгоритма классификации путем сравнения прогнозируемого значения целевой переменной с ее фактическим значением. Столбцы матрицы представляют наблюдения в прогнозируемом классе, а строки — наблюдения в фактическом

классе (или наоборот), чем лучше выражена главная диагональ матрица, тем эффективнее работает классификатор. Матрицы ошибок по моделям представлены на рисунке 6.

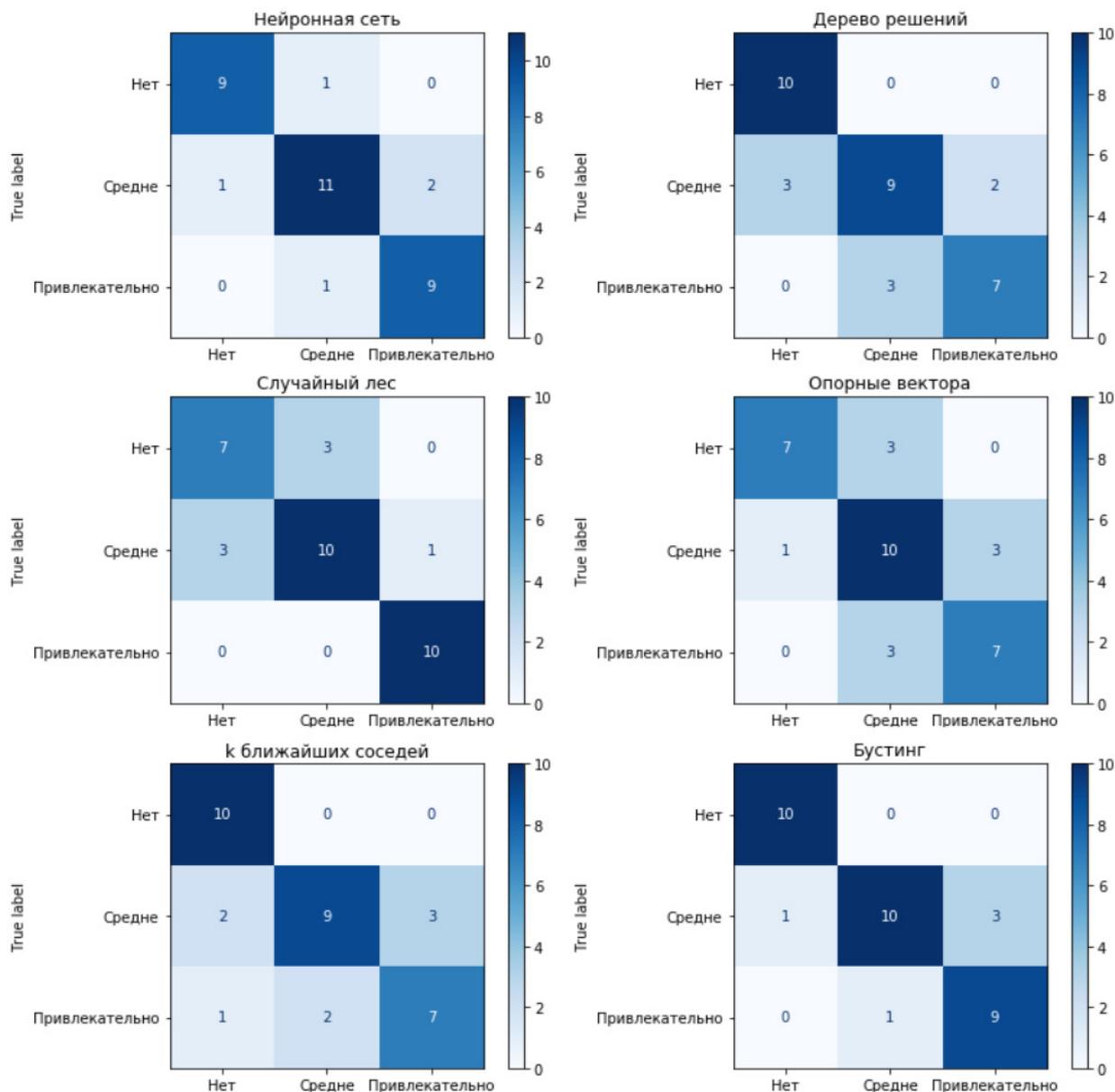


Рисунок 6. Матрицы ошибок для моделей машинного обучения

Оценка качества классификатора проводилась с помощью основных метрик качества классификации:

Accuracy - доля правильно предсказанных меток модели;

Precision - доля элементов в тестовой выборке действительно принадлежащих данному классу относительно всех элементов, которые система отнесла к этому классу;

Recall - это доля найденных классификатором элементов, принадлежащих классу относительно всех элементов этого класса в тестовой выборке.

Сравнительная таблица по основным метрикам представлена в таблице 4.

Таблица 4

## Основные метрики классификации по моделям

Модель	Accuracy↑,%	Precision↑ (по классам)			Recall↑ (по классам)			F-мера↑ (по классам)		
		1	2	3	1	2	3	1	2	3
Метод опорных векторов	71%	0.88	0.62	0.70	0.70	0.71	0.7	0.78	0.67	0.70
к-ближайших соседей	76%	0.77	0.82	0.70	1,00	0.64	0.70	0.87	0.72	0.70
Дерево принятия решений	76%	0,77	0,75	0,78	1,00	0,64	0,69	0,87	0,69	0,74
AdaBoost	85%	0,91	0,91	0,75	1,00	0,71	0,90	0,95	0,80	0,82
Нейронная сеть	85%	0,90	0,85	0,82	0,90	0,79	0,90	0,90	0,81	0,86
Случайный лес	94%	0,91	1,00	0,91	1,00	0,86	1,00	0,95	0,92	0,95

Самой лучшей моделью для задачи оказалась модель на основе алгоритма «случайный лес» показавшая точность 94% на тестовой выборке. Нейронная сеть и AdaBoost показали точность 85%, однако нейронная сеть смогла чуть лучше различать классы, чем AdaBoost. Модели на основе алгоритма «дерево принятия решений» и алгоритма «к-ближайших соседей» показали точность 76% и, в целом, одинаковую способность различать классы. Метод опорных векторов показал точность 71% и не смог точно различить классы.

В заключении отметим, что повышение инвестиционной привлекательности региона является одной из важнейших задач государственного управления региональным развитием, а процессы принятия решений в этой сфере деятельности предполагают необходимость выбора наиболее эффективных инвестиционных проектов, способных существенно повышать уровень социально-экономического развития региональной экономики. Проведённое исследование показало целесообразность использования методов интеллектуального анализа данных в процедурах принятия решений в области управления инвестиционной привлекательностью с целью обеспечения качества и обоснованности выбора основных мер повышения инвестиционной привлекательности региона.

## Литература

1. Белокур, А. А. Экономическая сущность инвестиционной привлекательности региона / А. А. Белокур // Экономика и социум. – 2017. – №3. – С. 44-50.
2. Итоги V ежегодного Рейтинга инвестиционной привлекательности регионов России (по итогам 2017 года) [Электронный ресурс] – URL: [http://www.rational.ru/sites/default/files/analytic\\_article/IPR-6-06112018.pdf](http://www.rational.ru/sites/default/files/analytic_article/IPR-6-06112018.pdf) (дата обращения 01.06.2020).
3. Итоги VI ежегодного Рейтинга инвестиционной привлекательности регионов России (по итогам 2018 года) [Электронный ресурс] – URL: [http://www.rational.ru/sites/default/files/analytic\\_article/IPR-6-06112018.pdf](http://www.rational.ru/sites/default/files/analytic_article/IPR-6-06112018.pdf) (дата обращения 01.06.2020).
4. Кизеев А.В. Оценка результативности и эффективности региональных инвестиционных программ в условиях дотационного региона // Известия высших учебных заведений. Серия «Экономика, финансы и управление производством». – 2011. – №02(08). – С. 122 - 128.
5. Регионы России. Социально-экономические показатели. 2017: Стат. сб. / Росстат. – М., 2017. –1402 с.
6. Регионы России. Социально-экономические показатели. 2018: Стат. сб. / Росстат. – М., 2018. –1162 с.
7. Руководство пользователя и документация библиотеки scikit-learn 0.22.2 [Электронный ресурс] – URL: [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html) (дата обращения 01.06.2020).
8. Метод опорных векторов для стандартной задачи классификации. URL: [http://www.machinelearning.ru/wiki/images/2/25/SMAIS11\\_SVM.pdf](http://www.machinelearning.ru/wiki/images/2/25/SMAIS11_SVM.pdf) (дата обращения: 01.06.2020).
9. Adam: A Method for Stochastic Optimization [Электронный ресурс] – URL: <https://arxiv.org/abs/1412.6980>.
10. Nocedal, George; Wright, Stephen J. Numerical Optimization. — 2nd edition. — USA: Springer, 2006.