

## ВЫЯВЛЕНИЕ ЗНАЧИМЫХ (РЕШАЮЩИХ) ФАКТОРОВ, ВЛИЯЮЩИХ НА ВЫБОР РАЗМЕЩЕНИЯ ПРОИЗВОДСТВЕННЫХ МОЩНОСТЕЙ КОМПАНИЙ НЕФТЕПЕРЕРАБАТЫВАЮЩЕЙ ОТРАСЛИ

**К.О. Буров, А.Ф. Шуплецов**

*Байкальский государственный университет, г. Иркутск, Российская Федерация*

### Информация о статье

Дата поступления  
10 марта 2022 г.

Дата принятия к печати  
8 апреля 2022 г.

Дата онлайн-размещения  
27 апреля 2022 г.

### Ключевые слова

Размещение производства;  
оптимизация;  
нефтеперерабатывающая  
промышленность; машинное  
обучение; схематическое  
моделирование

### Аннотация

Размещение производственных мощностей — основополагающий элемент построения эффективного и конкурентоспособного бизнеса. Для Российской Федерации стратегически значимой отраслью является нефтедобыча и нефтепереработка. Оптимизация размещения производственных мощностей нефтеперерабатывающих предприятий приобретает особую актуальность в условиях разработки новых месторождений нефти, а также реновации старых месторождений в рамках инновационных проектов крупнейших российских нефтяных корпораций. Для определения оптимальных мест размещения производства необходимо выявить основные факторы, которыми руководствуются предприятия нефтепереработки при дислоцировании своих производственных объектов на территории России. Это и является целью настоящего исследования. В рамках основной цели формируются такие подзадачи, как нахождение ориентировочной схемы размещения компанией (предприятием) своих основных мощностей, валидация сформированной модели на реальных статистических данных социально-экономического положения субъектов РФ с применением методов машинного обучения и оптимизация полученной статистической модели. Основными методами, используемыми в данной работе, являются графический и системный анализ деятельности, а также алгоритм машинного обучения «случайный лес». В рамках статьи сформирована общая для всех нефтеперерабатывающих предприятий схема выбора оптимального места расположения производства, выявлены основные факторы, которыми руководствуются предприятия нефтепереработки в данном процессе. В первую очередь к таким факторам относятся наличие в регионе крупного химического сектора промышленности, большого количества объектов нефтепроводов, высокое внутреннее потребление нефтепродуктов розничными потребителями. Кроме того, в работе обозначена ориентация нефтеперерабатывающей промышленности России на внутреннего потребителя конечной продукции.

Original article

## IDENTIFICATION OF SIGNIFICANT (DECISIVE) FACTORS INFLUENCING THE CHOICE OF THE ALLOCATION OF PRODUCTION CAPACITIES OF THE COMPANIES IN THE REFINING INDUSTRY

**Kirill O. Burov, Alexander F. Shupletsov**

*Baikal State University, Irkutsk, the Russian Federation*

### Article info

Received  
March 10, 2022

### Abstract

The placement of production facilities is a fundamental element of building an efficient and competitive business. Oil production and refining are strategically important industries for the Russian Federation. Optimiza-

Accepted  
April 8, 2022

Available online  
April 27, 2022

### Keywords

Production placement;  
optimization; oil refining industry;  
machine learning; schematic  
modeling

tion of the placement of production capacities of oil refineries is becoming particularly relevant in the context of the development of new oil fields, as well as the renovation of old fields within the framework of innovative projects of the largest Russian oil corporations. In order to develop optimal production locations, it is necessary to identify the main factors that guide oil refining enterprises when deploying their production facilities on the territory of the Russian Federation. The purpose of the work is to identify the main factors that guide the enterprises of the oil refining industry of the Russian Federation when placing their production facilities. Within the framework of the main goal, subtasks are also formed, such as finding an approximate scheme for the placement of the production enterprise of its main capacities, validation of the generated model on real statistical data of the socio-economic situation of the subjects of the Russian Federation with the use of machine learning methods and optimization of the obtained statistical model. The main methods used in this work are graphical analysis, system analysis of activities, as well as the «random forest» machine learning algorithm. Within the framework of this work, a scheme common to all production enterprises for choosing the optimal location of activities was formed, the main factors that guide oil refining enterprises when placing production facilities were identified, such as the presence of a large chemical industry sector in the region, high domestic consumption of petroleum products by retail consumers, as well as the presence of more oil pipeline facilities in the region. The orientation of the oil refining industry of the Russian Federation to the domestic consumer of final products is indicated.

### Предпосылки решения проблем

Оптимальное размещение производственных мощностей является необходимым условием для успешного ведения бизнеса и выживания предприятия в конкурентной среде. С этой целью должны учитываться определенные факторы: близость / удаленность рынков сбыта; доступ к сырью и материалам для производства (в том числе и к рабочей силе); политика государственных и местных властей; наличие необходимой инфраструктуры; эффективность транспортной инфраструктуры.

Данные факторы по-разному влияют на отдельно взятые отрасли промышленности [1]. Так, для производства IT-продуктов существенным будет наличие необходимой инфраструктуры (высокоскоростной Интернет) и подготовленной высококвалифицированной рабочей силы; для игровой индустрии определяющим фактором станет санкционирование федеральным правительством мест, где разрешено размещение игорных зон.

Для российской экономики нефтедобывающая и нефтеперерабатывающая отрасли традиционно являются значимым элементом. Так, по данным Федеральной таможенной службы РФ, за 2020 г. объем экспорта сырой нефти и продуктов нефтепереработки составил 45,24 %, т.е. приблизительно половину от всего экспорта России за исследуемый период<sup>1</sup>. При этом такая высокая

доля нефти и нефтепродуктов в структуре экспорта России не является случайным фактом, характерным только для 2020 г. На рис. 1 показана доля экспортируемой нефти и продуктов нефтепереработки в общем объеме экспорта России за 2011–2020 гг.

Исходя из графика, представленного на рис. 1, можно с уверенностью заключить, что экспорт нефти и нефтепродуктов продолжает оставаться значимым для экономики России. Следует заметить, что с 2014 г. наблюдается снижение экспорта нефти и нефтепродуктов с некоторым застоем в 2016 г. и очередным падением в 2019–2020 гг. Эти колебания вызваны в первую очередь введением экономических санкций в отношении России в 2014 г. и снижением спроса на нефть и нефтепродукты, вызванным пандемией COVID-19 в начале 2020 г., но не структурными изменениями в экономике России. Это подтверждается синхронным колебанием общего объема экспорта России за указанный период. На дополнительной оси (см. рис. 1) изображен валовой объем экспорта за аналогичный период, размер которого позволяет говорить о том, что совокупный объем экспорта сильно зависит от количества вывозимых сырой нефти и продуктов нефтепереработки.

### Механизм реализации задач

В нефтяной отрасли России ведется активная разработка углеводородных месторожде-

<sup>1</sup> Федеральная таможенная служба Российской Федерации. URL: <https://customs.gov.ru/folder/502>.

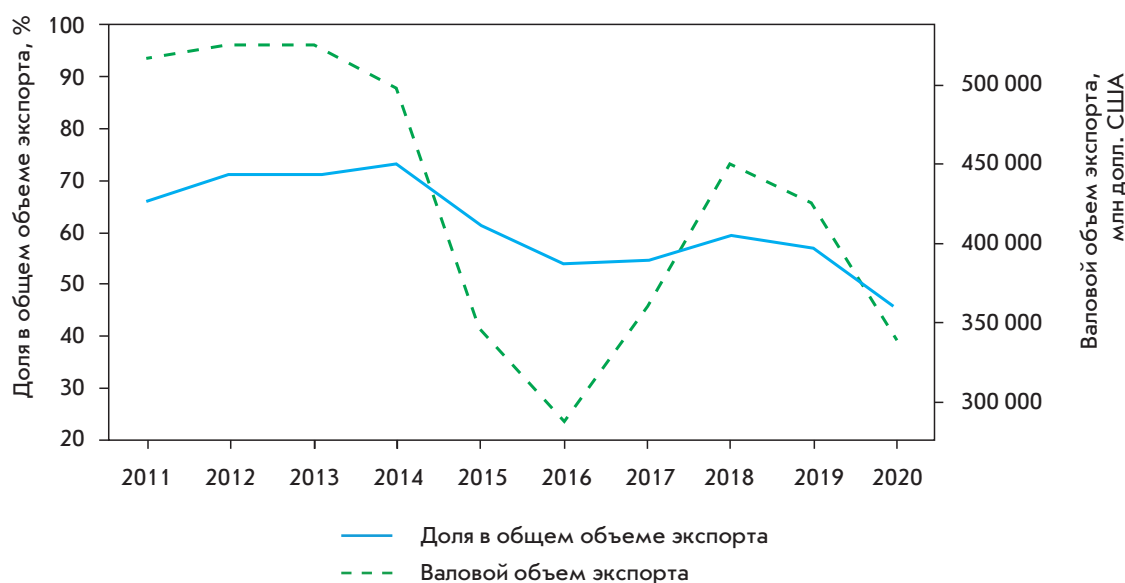


Рис. 1. Нефтяная составляющая экспорта и валовой объем экспорта России в 2011–2020 гг.

ний арктического шельфа<sup>2</sup>, нефтяных месторождений на севере Красноярского края<sup>3</sup>. Развитие нефтедобычи продвигает и нефтепереработку, для которой требуются большие ресурсы, а значит, должны быть увеличены и производственные мощности отрасли. Следует отметить, что заметно снизившиеся цены на нефть вынуждают экспортеров сырья все чаще обращать внимание на производство нефтепродуктов с более высокой добавленной стоимостью, что увеличивает потенциал нефтеперерабатывающей отрасли [2].

Выбор места размещения производства нефтеперерабатывающих заводов (НПЗ) должен быть технически и экономически обоснованным. Актуальность данной работы заключается в систематизации процесса выбора таких мест и классификации влияния факторов по степени их значимости на выбор региона производственной деятельности в условиях увеличения количества действующих и потенциальных месторождений нефти.

Производственные мощности нефтедобывающей промышленности размещаются непосредственно в местах добычи нефти. Там же обычно строятся нефтеочистительные заводы [3], которые производят первичную дегидратацию добытого ресурса. Однако для нефтеперерабатывающей промышленности вопрос размещения производственных мощностей представляется более сложной задачей.

<sup>2</sup> Приразломное месторождение // Газпром. URL: <https://www.gazprom.ru/projects/prirazlomnoye>.

<sup>3</sup> Восток Ойл // Роснефть. URL: [https://vostokoil.rosneft.ru/about/Glance/OperationalStructure/Dobicha\\_i\\_razrabotka/Vostochnaja\\_Sibir/vostokoil](https://vostokoil.rosneft.ru/about/Glance/OperationalStructure/Dobicha_i_razrabotka/Vostochnaja_Sibir/vostokoil).

Перед тем как начать непосредственное производство независимо от специфики отрасли, необходимо ответить на несколько вопросов, которые позволят сформировать четкое понимание особенностей производства и его структуры в целом, получить стратегическое видение бизнеса, масштабы его расширения, обозначить проблемы, с которыми придется столкнуться, и наметить возможные пути их решения.

Для полноты понимания представим вопросы, на которые необходимо ответить при принятии окончательного управленческого решения о выборе региона деятельности и размещении производственных мощностей, в виде нескольких последовательных блоков (этапов).

Для упрощения рассмотрения основных факторов, которые влияют на выбор оптимального места размещения производства, будем использовать стандартную производственную функцию следующего вида:

$$y = F(L, K),$$

где  $y$  — объем производства;  $L$  — количество трудовых ресурсов;  $K$  — количество капитальных ресурсов (инвестиции).

Допустим, что компания не является вертикально интегрированным образованием (т.е. собственным поставщиком продукции) и не имеет в управлении агентов, которые осуществляют сбыт продукции.

В первую очередь необходимо разрешить вопросы, связанные с производственным блоком деятельности. На данном этапе исследуются условия, касающиеся поставки

сырья для производства, выбора оптимальных тарифов и транспортных компаний, минимизации времени между отгрузкой сырья и началом производства для избежания простоя в производственном цикле. Здесь нужно рассмотреть следующие аспекты:

- основные поставщики сырья, имеющиеся на рынке;
- стоимость сырья с учетом доставки;
- существующие способы транспортировки сырья;
- время доставки сырья до места переработки;
- предложение на рынке трудовых ресурсов и влияние профсоюзных организаций.

Входными данными для модели применительно к российской экономике будут являться регионы РФ, для каждого из которых вычисляются соответствующие показатели стоимости сырья и доставки его до производства, время между датой отправки заказа на поставку сырья и датой его поступления на предполагаемое место размещения производства компании. В этих условиях становится возможным проранжировать регионы исходя из конечной стоимости сырья и времени его поставки.

Для того чтобы выявить потенциал рынка труда, используем индексы макроэкономической среды (ME-index) и бизнес-среды (BS-index), которые позволяют получить достоверную информацию об экономическом состоянии региона исходя из ожиданий бизнеса и макроэкономических показателей. Становится возможным оценить рынок труда, основываясь на таких статистических показателях, как количество рабочей силы, уровень безработицы на территории и среднемесячная заработная плата для потенциального региона размещения производственных мощностей.

Применительно к нефтеперерабатывающей и ряду других отраслей промышленности следует учитывать наличие дистрибьютеров, которые будут осуществлять конечный сбыт продукции. Наилучшим вариантом для компании будет, если в регионе размещения деятельности (строительства нефтеперерабатывающих объектов) присутствует много компаний-дистрибьютеров с целью предотвращения выхода на монопсонический рынок сбыта.

Следующим этапом для определения эффективного местоположения производственных мощностей будет решение вопросов, связанных со структурой сбыта конечной продукции компании. Сюда включаются следующие аспекты:

- основной сегмент сбыта продукции — внешний или внутренний<sup>4</sup>;
- емкость потенциального рынка сбыта;
- плотность распределения потенциальных потребителей продукции;
- количество потенциальных потребителей конечной продукции производства, находящихся в регионе.

Далее исследуется конкурентная среда, в которой потенциально будет существовать компания со своими производственными мощностями. Выявляются основные конкуренты, их доли на рынке, вероятное поведение при выходе на рынок нового агента (возможность объявления ценовой войны, демпинга). Значимым подкреплением к данному этапу будет вычисление индекса концентрации или индекса Херфиндала — Хиршмана (HHI) [4], которые помогут смоделировать более объемную и полную картину гипотетического уровня конкуренции для отрасли на предполагаемом месте размещения производственных мощностей.

Косвенным показателем, отображающим состояние конкурентной среды, может быть показатель развития конкуренции в субъектах РФ, публикуемый Федеральной антимонопольной службой, либо количество инвестиций в основной капитал, который характеризует инвестиционную привлекательность региона [5].

Следующим этапом будет оценка инфраструктурных возможностей региона, в которые включаются институциональные ограничения, накладываемые федеральными, региональными и местными властями. Сюда входят оценка транспортного и логистического потенциала предполагаемого региона размещения производственной деятельности и оценка экологического аспекта ведения деятельности, так как имеются существенные ограничения на размещение производства вблизи водоемов, заповедников, санитарно-защитных зон и иных особо охраняемых природных территорий [6].

Для оценки транспортного аспекта следует учесть стоимость доставки готового продукта от места производства до конеч-

<sup>4</sup> Вводя в нашу модель поиска наилучшего месторасположения предприятия выбор между внутренним и внешним рынком, вполне можно предположить, что не вся продукция идет только на экспорт или, наоборот, на удовлетворение внутреннего спроса. Распределение сбыта продукции может осуществляться в каком-либо ином соотношении, что также должно учитываться в рамках более полной и адекватной модели. Однако для упрощения можно допустить, что выбор распределения продукции между внешним и внутренним рынком осуществляется в пользу либо одного, либо другого рынка.

ного потребителя. В случае ориентации производства на внутренний рынок и при использовании автомобильного транспорта<sup>5</sup> для доставки товара нужно брать в расчет стоимость 1 машино-часа того транспорта, с помощью которого будет производиться доставка, либо стоимость 1 л топлива. При необходимости размещения готовой продукции и запаса материалов на складских территориях важным показателем будет стоимость 1 м<sup>3</sup> складского помещения. Также учитывается плотность путей дорожного сообщения, включая автодороги, железные дороги, трубопроводный транспорт, пути внутреннего речного сообщения и др.

Следующим институциональным фактором, который может повлиять на выбор месторасположения производственной деятельности отрасли, является наличие или отсутствие налоговых преференций и льгот как для конкретной отрасли, так и для всего региона в целом. Представим все указанные блоки (этапы) в виде схемы (рис. 2).

Схема, представленная на рис. 2, иллюстрирует, по каким параметрам должен осуществляться выбор места производства и дистрибуции конечной продукции. Эту схему можно применить к любой отрасли, в этом

случае в ней меняются только конкретные вопросы, присущие каждому блоку, отражающие специфику определенной отрасли. Последовательность реализации блоков остается той же.

Представленный алгоритм (см. рис. 2) имеет общий характер и не отражает специфику какой-либо сферы деятельности. Полезность данной схемы заключается в наличии последовательной структуры реализации этапов (блоков), которая систематизирует процесс выбора и принятия управленческого решения.

Для валидации представленной схемы размещения производственных мощностей испытаем ее на реальных данных социального и экономического положения субъектов РФ за 2019 г., используя для их классификации целевой признак, который будет показывать, есть или нет в данном регионе нефтеперерабатывающий завод.

Выделим исходные признаки, по которым будем классифицировать субъекты РФ:

- количество месторождений;
- число объектов нефтепроводов;
- рабочая сила, тыс. чел.;
- среднемесячная заработная плата, р.;
- плотность населения, чел/км<sup>2</sup>;
- количество автозаправочных станций

(АЗС);

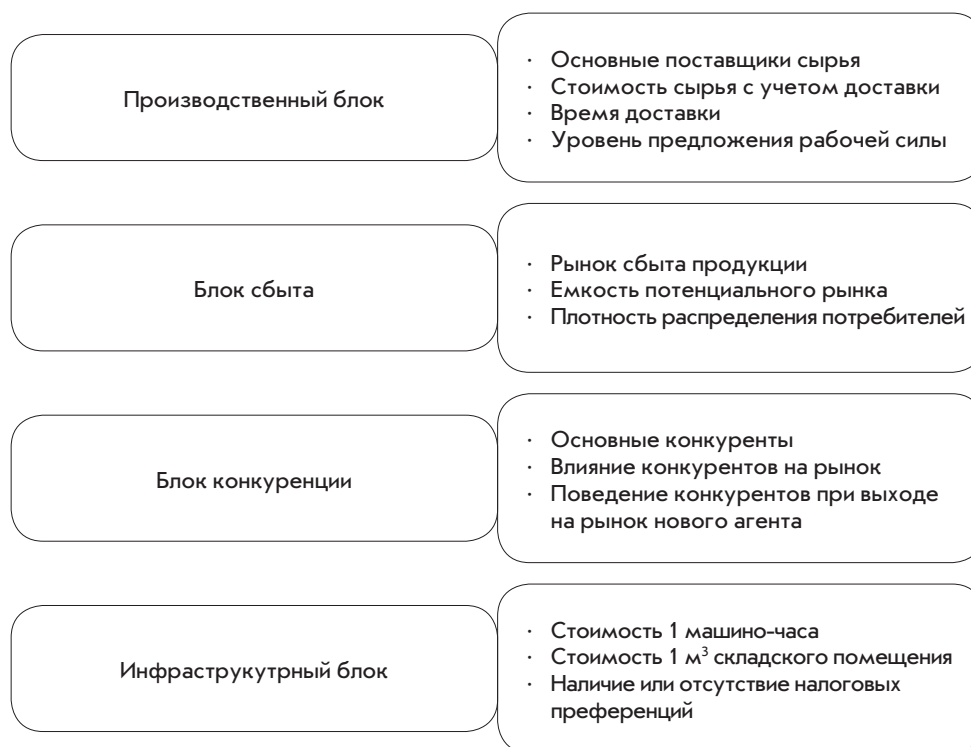


Рис. 2. Схема выбора оптимального размещения производства

- доля химической промышленности в общем объеме промышленности региона;
- инвестиции в основной капитал, млн р.;
- плотность железнодорожных путей сообщения, км/10 000 км<sup>2</sup>;
- мощность электростанций, млн кВт (табл. 1).

Первые четыре признака характеризуют производственный блок, пятый, шестой и седьмой — блок сбыта, восьмой — блок конкуренции, а девятый и десятый признаки отсылают к инфраструктурному блоку.

Данные о количестве месторождений в разрезе субъектов РФ взяты из реестра Росгеолфонда. Учитывались только введенные в эксплуатацию месторождения<sup>6</sup> с действующими техническими паспортами. Информация по количеству объектов взята из единой энергетической базы. Принимались в расчет объекты, относящиеся к нефтяному и нефтепродуктовому трубопроводному транспорту: линейные производственно-диспетчерские станции; компрессорные станции; нефтеперекачивающие станции; пункты сдачи приемки сырья и установки подготовки нефти. Остальные сведения получены из данных региональной статистической службы (Росстат).

#### Дискуссионные моменты

В формализованном виде модель будет выглядеть следующим образом:

$$Y = f(p_1, p_2, \dots, p_n, a),$$

где  $Y$  — бинарная переменная со значениями (0; 1), указывающая на наличие / отсутствие в данном регионе НПЗ (0 — НПЗ отсутствует, 1 — присутствует);  $p_1, p_2, \dots, p_n$  — факторы, влияющие на размещение НПЗ в регионах;  $a$  — коэффициент точности, который определяет качество модели.

Задача сводится к тому, чтобы подобрать оптимальную функцию  $f$  так, чтобы коэффициент точности был максимально высоким. Тогда выходной моделью будет следующая функция:

$$Y = f(p_1, p_2, \dots, p_k, a_{max}),$$

где  $k \leq n$ , так как условия оптимизации предполагают уменьшение количества влияющих факторов;  $a_{max}$  — максимально возможный коэффициент «точности классификации».

Для классификации регионов России используем алгоритм «случайного леса», который является одним из методов машинного

<sup>6</sup> Без учета месторождений, находящихся в морских акваториях.

обучения. В его основе лежат два базовых алгоритма — «bagging» и «дерево решений», позволяющих решать задачи регрессионного анализа и классификации «путем разбиения множества данных на более мелкие по объему выборки», выявляя особо значимые признаки и агрегируя полученные результаты.

В алгоритме «случайного леса» выбор классифицирующих признаков на ряде более мелких совокупностей данных происходит случайным образом [7], что позволяет улучшить итоговые результаты обработки информации.

Обработав данные и получив промежуточный результат анализа, переходим к оптимизации полученных сведений с помощью метода оптимизации гиперпараметров. Суть этого метода состоит в выборе таких оптимальных надстроек исходной модели, как количество деревьев в «случайном лесу»; количество признаков для разбиения исходной совокупности данных; максимальная глубина деревьев; максимальный конечный узел; повторение данных в подвыборках; минимальное число объектов в «листьях».

Оценивать качество модели будем с помощью коэффициента точности предсказаний классификации (currency), который определим при заполнении матрицы ошибок. Одновременно получим распределение переменных параметров исходной выборки по степени значимости.

Значимость параметров модели находится с помощью нормализованного коэффициента Джини (Gini's coefficient) (среднее уменьшение примеси, MDI), определяемого как время, в течение которого параметр используется для разбиения дерева пропорционально количеству выборок, на которые этот параметр разделяет сжатую подвыборку [8]. Обработку и обучение модели проведем с помощью языка Python с подключением таких библиотек, как Pandas, ScikitLearn, Matplotlib, Seaborn и NumPy.

Разобьем данные на тренировочную и тестовую подвыборки, где первая будет составлять 40 % от исходного объема информации. Предварительно очистим исходные данные от зависимых признаков, составив для этого матрицу корреляций (рис. 3). Следует избавиться от признаков «рабочая сила» и «плотность населения», которые, в свою очередь, зависят от таких признаков, как «инвестиции в основной капитал» и «плотность железнодорожных путей сообщения».

Проведем нормализацию данных для улучшения качества классификации и упрощения процедуры обучения модели.

## Социально-экономические характеристики субъектов РФ

Регион	Наличие нефтеперерабатывающего завода	Плотность населения, чел./км <sup>2</sup>	Количество месторождений	Количество объектов нефтепроектирования	Среднемесячная заработная плата, р.	Инвестиции в основной капитал, млн р.	Рабочая сила, тыс. чел.	Плотность железно-дорожных путей сообщения, км/10 000 км <sup>2</sup>	Количество автостоянок	Мощность электростанций, млн кВт	Доля химической промышленности в общем объеме промышленности региона
Белгородская область	1	57,17	1	1	34 615	167 367	826	258	275	0,3	0,047
Брянская область	0	34,17	1	4	29 853	63 944	595	280	160	0,04	0,068
Владимирская область	0	46,68	0	4	33 076	90 085	721	316	288	0,7	0,177
Воронежская область	0	44,52	0	6	33 690	298 024	1 182	229	397	4,6	0,127
Ивановская область	0	46,59	0	4	27 553	37 992	517	161	171	0,9	0,069
Калужская область	1	33,64	0	0	41 442	108 251	537	288	155	0,2	0,126
Костромская область	0	10,52	0	0	31 421	26 194	310	107	84	3,8	0,082
Курская область	0	36,80	1	3	32 709	144 906	569	352	221	4,7	0,222
Липецкая область	0	47,48	0	5	34 312	155 038	598	315	212	1,2	0,034
Московская область	0	173,61	0	9	55 555	1 044 870	4 189	490	1 123	7,3	0,184
Орловская область	1	29,70	0	4	29 683	55 877	347	241	116	0,4	0,064
Рязанская область	1	28,0	0	8	34 488	69 191	535	238	250	3,9	0,260
Смоленская область	1	18,77	0	3	31 269	69 358	483	226	229	4,0	0,259
Тамбовская область	0	29,18	0	9	28 697	120 579	500	214	191	0,4	0,065
Тверская область	1	14,97	0	6	33 524	86 252	676	215	311	6,9	0,144

Продолжение табл. 1

Регион	Наличие нефтеперерабатывающего завода	Плотность населения, чел/км <sup>2</sup>	Количество месторождений	Количество объектов нефтепроектирования	Среднемесячная заработная плата, р.	Инвестиции в основной капитал, млн р.	Рабочая сила, тыс. чел.	Плотность железных дорог, км/10 000 км <sup>2</sup>	Количество автостанций, шт.	Мощность электростанций, млн кВт	Доля химической промышленности в объеме промышленности региона
Тульская область	0	57,05	0	4	38 151	177 724	793	367	279	1,7	0,208
Ярославская область	1	34,62	0	11	36 016	90 094	650	180	157	2,0	0,327
Москва	1	4 876,19	0	0	94 294	2 856 935	7 308	1 921	1 122	16,5	0,558
Республика Карелия	0	3,40	1	4	42 964	48 159	305	123	118	1,1	0,030
Республика Коми	1	1,97	0	28	53 416	114 630	428	41	91	3,0	0,313
Ненецкий автономный округ	0	0,25	1	0	88 027	95 692	23	0	4	0,9	0,0
Архангельская область	0	2,64	1	7	49 435	94 040	535	30	158	2,7	0,028
Вологодская область	0	8,03	0	16	39 116	197 130	566	53	176	2,2	0,016
Калининградская область	0	67,05	1	1	35 637	103 037	537	442	176	1,8	0,050
Ленинградская область	1	22,36	1	16	46 387	419 126	969	304	358	8,6	0,365
Мурманская область	0	5,12	1	5	63 715	171 361	421	60	101	3,6	0,017
Новгородская область	0	10,94	0	5	32 174	44 733	303	210	145	0,5	0,366
Псковская область	0	11,30	1	4	29 441	33 503	316	197	171	0,5	0,028
Санкт-Петербург	0	3 855,7	0	0	65 872	690 722	3 073	3 082	453	4,6	0,286
Республика Адыгея	1	59,37	1	2	30 192	42 381	201	205	118	0,05	0,099



Продолжение табл. 1

Регион	Наличие нефтеперерабатывающего завода	Плотность населения, чел./км <sup>2</sup>	Количество месторождений	Количество объектов нефтепроектирования	Среднемесячная заработная плата, р.	Инвестиции в основной капитал, млн р.	Рабочая сила, тыс. чел.	Плотность железных дорог, км/10 000 км <sup>2</sup>	Количество автостанций, шт.	Мощность электростанций, млн кВт	Доля химической промышленности в общем объеме промышленности региона
Республика Калмыкия	0	3,63	0	2	28 617	16 035	133	22	99	0,09	0,159
Республика Крым	0	73,28	1	0	32 748	199 821	918	255	465	1,8	0,207
Краснодарский край	1	75,17	10	24	36 133	439 635	2 807	297	980	2,8	0,313
Астраханская область	0	20,53	1	6	36 093	95 658	507	128	171	1,1	0,648
Волгоградская область	1	22,06	1	15	33 371	199 490	1 243	143	425	4,3	0,605
Ростовская область	1	41,56	1	17	33 757	282 970	2 097	188	843	7,4	0,244
Севастополь	0	499,11	0	0	34 621	43 023	222	0	61	1,7	0,036
Республика Дагестан	1	61,84	44	5	26 835	224 773	1 382	101	1 123	1,9	0,038
Республика Ингушетия	0	140,86	0	0	27 410	23 884	259	108	183	1,3	0,295
Кабардино-Балкарская Республика	0	69,47	0	0	27 466	44 095	446	107	214	0,2	0,028
Карачаево-Черкесская Республика	0	32,55	0	0	26 955	22 586	204	35	142	0,6	0,215
Республика Северная Осетия — Алания	1	87,11	0	1	28 751	32 832	311	180	232	0,1	0,029
Чеченская Республика	0	94,79	0	0	27 757	82 653	634	195	268	0,4	0,065

Продолжение табл. 1

Регион	Наличие нефтеперерабатывающего завода	Плотность населения, чел./км <sup>2</sup>	Количество месторождений	Количество объектов нефтепро- водов	Средне- месячная заработная плата, р.	Инвестиции в основной капитал, млн р.	Рабочая сила, тыс. чел.	Плотность железнодо- рожных путей сообщения, км/10 000 км <sup>2</sup>	Количество автоса- правочных станций, шт.	Мощ- ность электро- станций, млн кВт	Доля химической промышлен- ности в общем объеме про- мышленности региона
Ставропольский край	1	42,35	0	9	31 836	180 017	1 383	139	528	4,5	0,368
Республика Башкортостан	1	28,26	1	27	36 465	322 104	1 896	102	507	5,7	0,652
Республика Марий Эл	1	29,03	0	5	30 152	26 667	334	65	126	0,3	0,158
Республика Мордовия	0	30,28	0	1	28 826	52 189	438	208	109	0,4	0,037
Республика Татарстан	1	57,56	0	17	37 418	640 837	2 036	129	747	8,2	0,528
Удмуртская Республика	1	35,65	0	11	34 052	100 516	763	185	267	0,8	0,031
Чувашская Республика	0	66,55	0	5	29 671	63 622	608	230	165	2,2	0,154
Пермский край	1	16,23	15	16	39 210	291 099	1 225	98	362	8,0	0,624
Кировская область	0	10,49	0	0	30 213	71 005	637	91	205	1,0	0,290
Нижегородская область	1	41,81	0	12	35 212	292 587	1 754	158	485	2,8	0,381
Оренбургская область	1	15,82	1	9	32 883	211 726	930	118	323	4,1	0,333
Пензенская область	0	30,08	0	6	30 765	90 426	656	191	270	0,4	0,024
Самарская область	1	59,32	0	29	36 431	287 615	1 683	256	605	6,1	0,261
Саратовская область	1	23,93	1	20	30 717	167 900	1 203	226	400	6,7	0,280
Ульяновская область	1	33,06	0	1	30 677	72 019	611	187	224	1,1	0,071

Продолжение табл. 1

Регион	Наличие нефтеперерабатывающего завода	Плотность населения, чел./км <sup>2</sup>	Количество месторождений	Количество объектов нефтепроектирования	Среднемесячная заработная плата, р.	Инвестиции в основной капитал, млн р.	Рабочая сила, тыс. чел.	Плотность железных дорог, км/10 000 км <sup>2</sup>	Количество автостанций, шт.	Мощность электростанций, млн кВт	Доля химической промышленности в общем объеме промышленности региона
Курганская область	1	11,57	1	9	30 632	40 238	365	104	175	0,7	0,094
Свердловская область	0	22,19	0	11	41 110	450 397	2 125	181	612	10,7	0,061
Ханты-Мансийский автономный округ	0	3,13	2	33	75 087	953 516	915	20	227	15,5	0,893
Ямало-Ненецкий автономный округ	0	0,71	1	29	101 012	864 881	315	6	83	4,2	0,950
Тюменская область	1	9,60	0	30	48 613	302 945	727	55	535	6,4	0,860
Челябинская область	0	39,17	1	21	37 433	299 051	1 875	203	724	6,4	0,062
Республика Алтай	0	2,37	0	0	33 387	21 587	95	0	85	0,1	0,067
Республика Тыва	0	1,94	0	0	39 673	17 493	117	0	133	0,05	0,0
Республика Хакасия	0	8,67	0	0	40 548	34 675	246	108	220	7,2	0,023
Алтайский край	1	13,79	1	0	27 962	115 369	1 140	93	541	1,6	0,255
Красноярский край	1	1,21	1	11	49 932	426 491	1 482	9	833	18,6	0,054
Иркутская область	1	3,09	0	16	46 387	359 197	1 167	32	547	13,7	0,201
Кемеровская область	1	27,77	0	2	41 770	297 946	1 291	175	473	5,6	0,362

Окончание табл. 1

Регион	Наличие нефтеперерабатывающего завода	Плотность населения, чел./км <sup>2</sup>	Количество месторождений	Количество объектов нефтепроектирования	Среднемесячная заработная плата, р.	Инвестиции в основной капитал, млн р.	Рабочая сила, тыс. чел.	Плотность железных дорог, км/10 000 км <sup>2</sup>	Количество автостанций, шт.	Мощность электростанций, млн кВт	Доля химической промышленности в общем объеме промышленности региона
Новосибирская область	1	15,74	1	10	39 076	248 078	1 430	85	415	3,1	0,156
Омская область	1	13,65	1	12	35 368	172 195	1 016	52	231	1,7	0,792
Томская область	1	3,43	0	6	45 526	96 214	541	11	190	1,2	0,344
Республика Бурятия	0	2,81	1	0	39 115	71 128	433	35	203	1,5	0,009
Республика Саха (Якутия)	1	0,32	0	17	73 402	381 147	501	3	292	3,1	0,186
Забайкальский край	0	2,45	1	0	43 896	86 667	525	56	331	1,7	0,305
Камчатский край	0	0,67	1	0	80 448	47 241	182	0	69	0,7	0,001
Приморский край	1	11,51	1	9	46 867	179 356	999	95	325	2,8	0,032
Хабаровский край	1	1,67	1	11	50 213	161 488	702	27	220	2,6	0,173
Амурская область	1	2,18	1	10	47 234	339 934	408	81	197	4,2	0,011
Магаданская область	0	0,30	1	0	94 856	35 797	86	0	30	1,6	0,007
Сахалинская область	0	5,61	1	3	87 418	231 701	275	96	83	1,6	0,031
Еврейская автономная область	0	4,36	1	6	42 400	15 490	0	0	49	0,2	0,027
Чукотский автономный округ	0	0,07	1	0	107 107	25 906	0	0	19	0,3	0,0

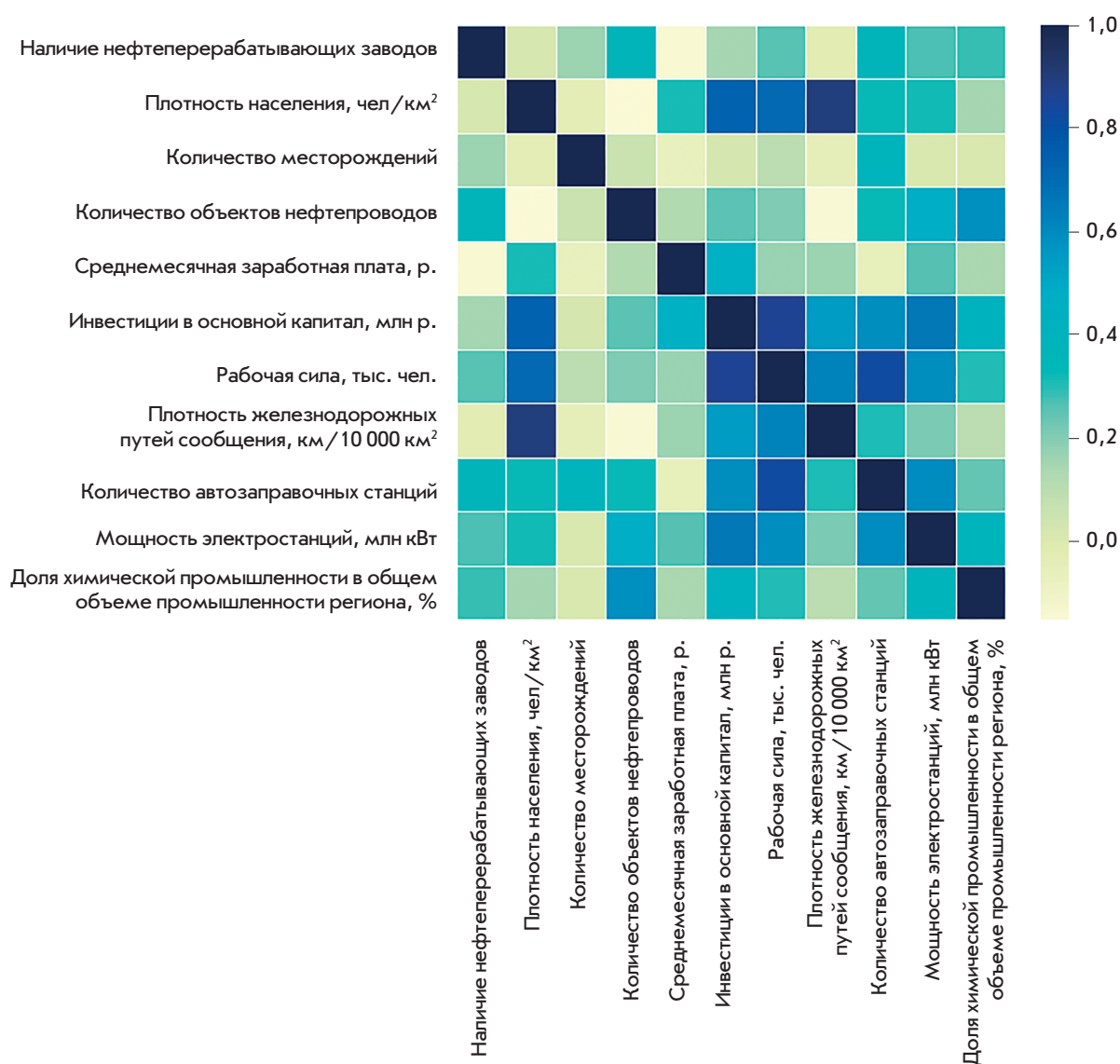


Рис. 3. Матрица корреляций исходных признаков

При первом запуске исходной модели необходимо указать первоначальные параметры, а именно количество «деревьев в лесу» и псевдослучайное число, которое будет гарантировать случайное разбиение исходного множества на более мелкие подвыборки. Для того чтобы оптимизировать значение первого параметра, запустим цикл, в котором модель будет обучаться, предсказывать данные, а затем сравнивать их с верными значениями. Количество итераций цикла зададим равным 100, и каждый раз количество «деревьев в лесу» будет принимать значение данного итератора. Получим распределение ошибок в зависимости от указанного значения количества деревьев. Представим это распределение в виде графика (рис. 4).

Как видно из графика (см. рис. 4), наименьшее число ошибок достигается при

количестве «деревьев», равном 20. Укажем этот параметр при запуске модели. Для выбора второго параметра обозначим случайный генератор целых чисел из модуля «random», предварительно установив и передав в него абсолютно любое целое число. В нашем случае мы передадим число 10.

Первоначальная модель, обученная с помощью алгоритма «случайного леса», показала точность 65 %, а матрица ошибок выглядит следующим образом (табл. 2).

Таблица 2

Матрица ошибок неоптимизированной модели

	Истина	Ложь
Истина	9	7
Ложь	5	13

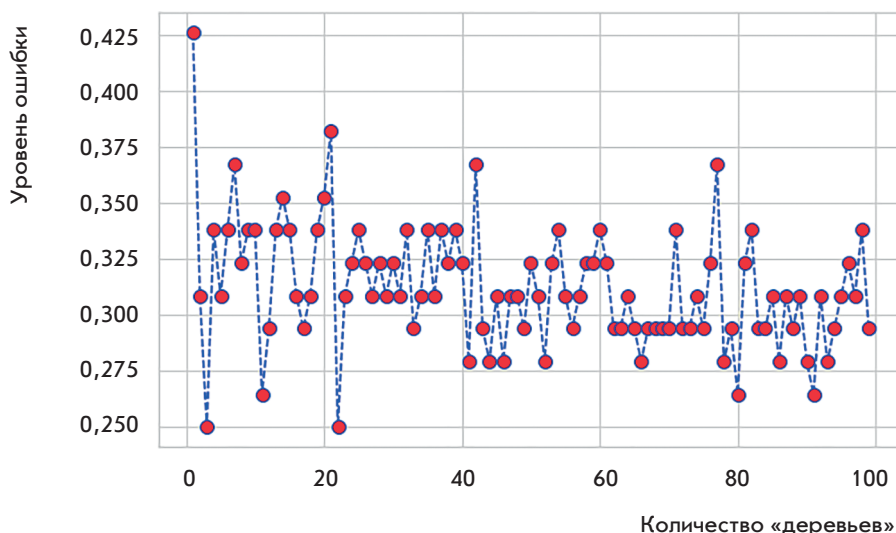


Рис. 4. Распределение уровня ошибок в зависимости от параметра модели

Наиболее важно количество истинно положительных и истинно отрицательных результатов, составляющих для модели соответственно 9 и 13. Интерпретация этих показаний состоит в следующем: в 9 случаях модель правильно определила, что в данном регионе находится НПЗ, а в 13 случаях верно определила, что в указанном субъекте РФ отсутствует НПЗ.

Исходя из анализа (рис. 5), можно отметить, что наиболее значимым признаком для модели является количество объектов нефтепроводов, а наименее значимым — количество месторождений нефти в регионе. Из чего следует, что для размещения НПЗ на территории какого-либо субъекта РФ нужно учитывать скорее не близость / удаленность от основных месторождений нефти, а степень развития трубопроводной инфраструктуры региона.

Важными элементами обученной модели являются количество АЗС, представленных в регионе, мощность электростанций, доля химической промышленности в структуре производства региона и плотность железных дорог.

Можно предположить, что в регионах, в которых присутствуют НПЗ, высокую значимость имеют такие показатели, как количество потенциальных потребителей продукции, энергетическая мощность региона, характеризующая инфраструктурный блок схемы (см. рис. 1), уровень среднемесячной заработной платы, которая является одним из элементов производственного блока.

Поскольку уровень точности первоначальной модели составляет 69 %, то ее необходимо оптимизировать, используя для этого корректировки гиперпараметров исходной модели. Для этого возьмем алгоритм

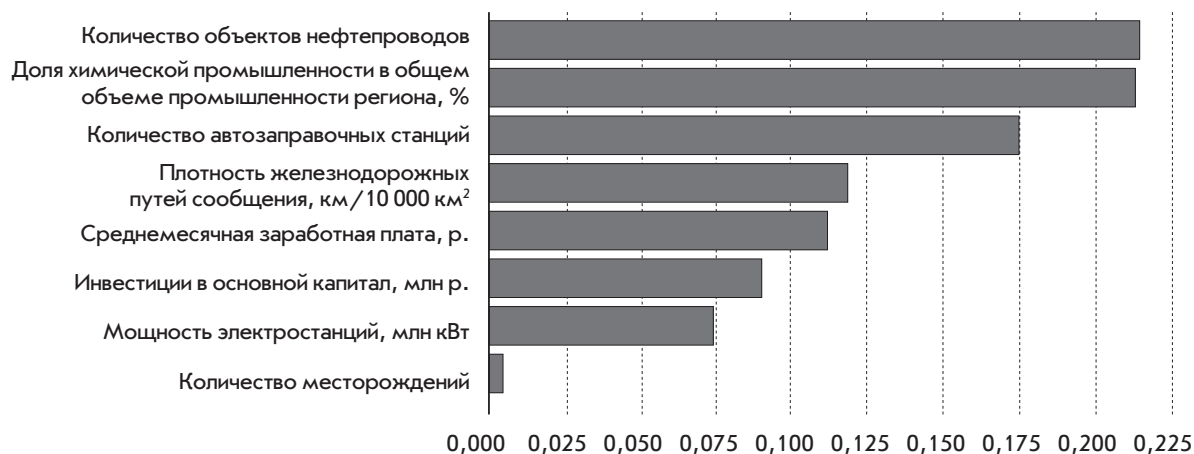


Рис. 5. Распределение значимости признаков неоптимизированной модели

RandomizedSearchCV, который позволит подобрать улучшенные гиперпараметры [9].

Получив оптимальные параметры и применив их к исходной модели, проведем те же манипуляции с данными для характеристики как качества классификации, так и значимости классифицирующих признаков.

Матрица ошибки оптимизированной модели выглядит следующим образом (табл. 3).

Таблица 3  
Таблица ошибок оптимизированной модели

	Истина	Ложь
Истина	13	5
Ложь	5	11

Уровень точности оптимизированной модели составляет 72 %, что является средним значением точности классификации [10]. Такой уровень объясняется рядом факторов, где само количество регионов играет решающую роль. Недостаток данных мешает получить высокую точность предсказаний, и, следовательно, увеличить ее возможно путем перехода от макроуровня к более мелкому, а именно к городам РФ, количество которых значительно больше. В этом случае точность предсказаний будет существенно выше. Обилие социально-экономических и промышленных показателей субъектов РФ не позволяет одновременно улучшить модель и при этом не упустить из виду ее интерпретируемость.

Увеличение количества классификаторов, во-первых, повысит вероятность переобучить модель [11], что сделает ее непригодной для последующих испытаний, во-вторых, может улучшить предсказательную силу модели, но при этом ее интерпретация станет более сложной из-за наложения множества показателей друг на друга.

Как следует из полученных данных (см. табл. 3), увеличение данного показателя вызвано ростом количества истинно отрицательных результатов. Распределение степени значимости классифицирующих признаков представлено на рис. 6.

Значимость признаков оптимальной модели выводит на первый план те же факторы, которые были и в неоптимизированной модели. Исходя из распределения (см. рис. 5), можно утверждать, что размещение НПЗ на территории РФ ориентировано в основном на «как можно большее приближение к потребителю» с целью минимизации транспортных и временных издержек и увеличения объема продаж. Большую роль при выборе места размещения производственных мощностей НПЗ играет и мощность электростанций<sup>7</sup>.

Отметим, что одним из направлений повышения (улучшения) качества «модели классификации» становится использование вероятностных алгоритмов, например самого распространенного — байесовского классификатора, который опирается на классическую теорию вероятностей, но при этом качественно не уступает в классификации нейронным сетям.

Для выявления и уточнения направлений увеличения возможностей размещения НПЗ в субъектах РФ воспользуемся графическим анализом факторов. Выведем график взаимных парных корреляций для таких факторов, как плотность железнодорожных путей сообщения; количество объектов нефтепроводов; количество АЗС; мощность электростанций; доля химической промышленности в общем объеме промышленности региона; количество автозаправочных станций; Плотность железнодорожных путей сообщения, км / 10 000 км<sup>2</sup>; Инвестиции в основной капитал, млн р. Среднемесячная заработная плата, р. Количество месторождений

<sup>7</sup> Данный фактор указывает на то, что в основном выбираются регионы, развитые в энергетическом отношении.

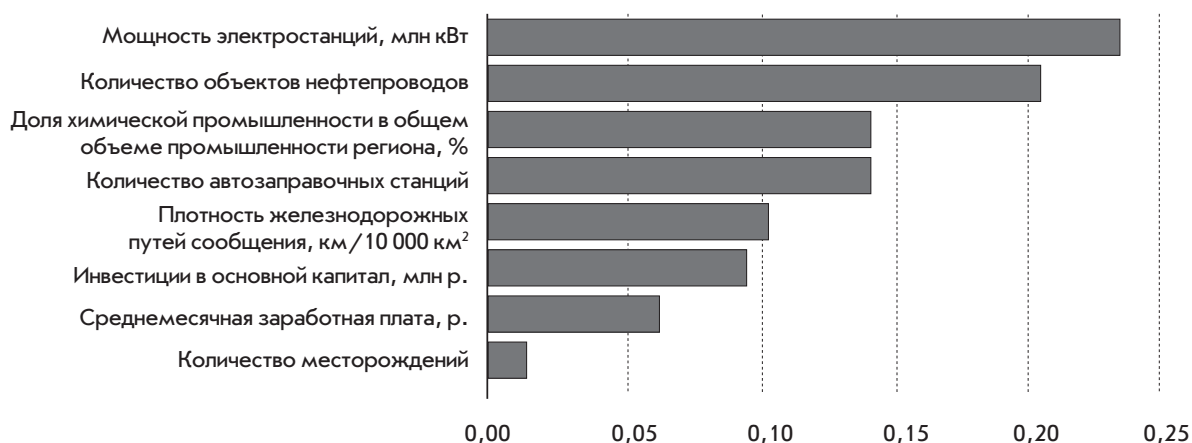


Рис. 6. Распределение значимости признаков оптимизированной модели

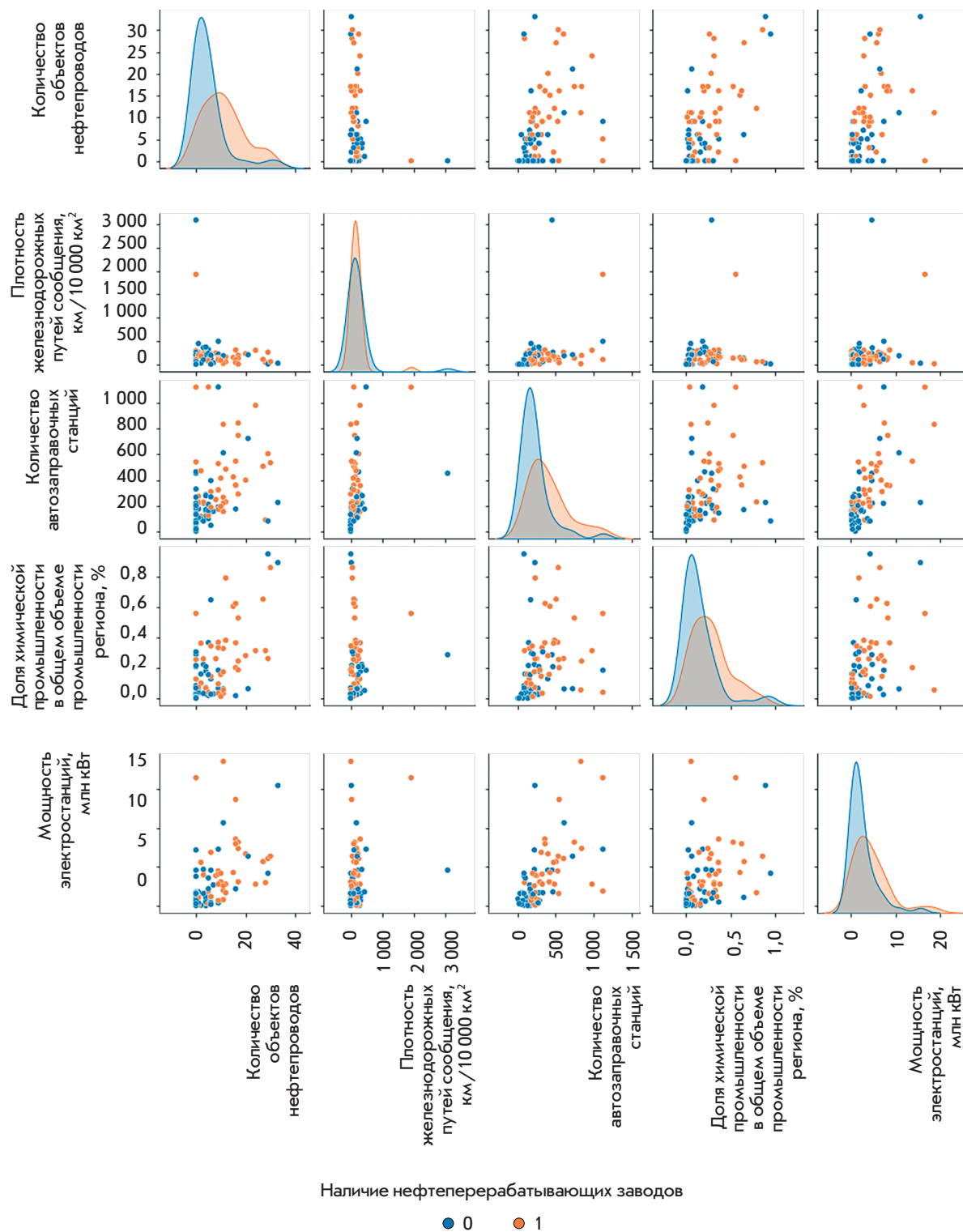


Рис. 7. Диаграммы рассеяния в разрезе наличия нефтеперерабатывающих заводов

В первую очередь нас интересует диагональ графика. На нем расположены диаграммы рассеяния факторов в разбиении по факту наличия / отсутствия НПЗ в регионе.

#### Выводы

В конечном итоге была получена «приблизительная схема планирования распо-

ложения производственных мощностей», прошедшая валидацию на реальных данных социально-экономического состояния субъектов РФ. Первоначальная модель, классифицирующая регионы России в зависимости от наличия НПЗ, полученная с помощью алгоритма машинного обучения «случайный лес», предсказывала конечный результат с



точностью 65 % и ранжировала факторы по степени их значимости для модели. После оптимизации модели были получены уточненные данные: точность конечного результата составила 72 % с учетом значимости факторов.

Размещение НПЗ на территории России в первую очередь ориентировано на потребителей продуктов переработки нефти, предприятия химической промышленности, основным ресурсом которых являются продукты нефтепереработки, а также потребителей автомобильного топлива, что объясняется большим количеством АЗС в тех регионах, где расположены НПЗ.

Знание этого факта позволит оптимизировать процесс размещения НПЗ на территории субъектов РФ, выявляя наиболее плотные места в потребительском сегменте, поможет распределить пути размещения магистральных нефтепроводов, учитывая количество потенциальных потребителей нефтепродуктов.

При размещении НПЗ учитываются такие производственные факторы, как количество объектов нефтепроводов, уровень заработной платы. Отметим при этом, что количество месторождений мало влияет на расположение производственных мощностей. Это свидетельствует о высокой степени важности трубопроводного транспорта для промышленности и экономики РФ.

### СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Дадалова М.В. Влияние производственной мощности предприятия на его экономическую устойчивость / М.В. Дадалова // Вестник ИргТУ. — 2015. — № 1 (96). — С. 137–141.
2. Пиголева И.В. Мировая нефтегазохимия: основные тенденции развития в условиях трансформации мировой энергетики / И.В. Пиголева, К.Н. Миловидов // Микроэкономика. — 2018. — № 5. — С. 22–30.
3. Гончаренко С.Н. Модели и методы оптимизации плана добычи и первичной переработки нефти / С.Н. Гончаренко, З.А. Сафронова // Горный информационно-аналитический бюллетень. — 2008. — № S10. — С. 221–229.
4. Титкова И.К. Индекс Херфиндаля и индекс энтропии как показатели диверсификации экономики / И.К. Титкова // Мировой опыт и экономика регионов России: материалы XV Всерос. студ. науч. конф. с междунар. участием, Курск, 7 апр. 2017 г. / под ред. Л.А. Дремовой. — Курск, 2017. — С. 375–377.
5. Шкодинский С.В. Развитие промышленных предприятий: инвестиции в основной капитал и региональная асимметрия / С.В. Шкодинский, И.Н. Рыкова, А.Г. Назаров. — DOI 10.18384/2310-6646-2020-2-116-130 // Вестник Московского государственного областного университета. Сер.: Экономика. — 2020. — № 2. — С. 116–130.
6. Кичигин Н.В. Приостановление, ограничение, прекращение экологически опасной хозяйственной деятельности: мера юридической ответственности или способ предупреждения причинения экологического вреда? / Н.В. Кичигин // Конституционно-правовые основы ответственности в сфере экологии: сб. материалов Междунар. науч. конф., Москва, 20 дек. 2018 г. — 14 марта 2019 г. — Москва, 2019. — С. 305–309.
7. Пальмов С.В. Случайный лес: основные особенности / С.В. Пальмов, А.О. Денискова // Наука сегодня: теоретические и практические аспекты: материалы Междунар. науч.-практ. конф., Вологда, 28 дек. 2016 г. — Вологда, 2017. — С. 51–53.
8. Perrier A. Feature Importance in Random Forests / A. Perrier // Data Science. — URL: <https://alexisperrier.com/datascience/2015/08/27/feature-importance-random-forests-gini-accuracy.html>.
9. Автоматизированное машинное обучение: обзор возможностей современных платформ анализа данных / И.П. Болодурина, Д.И. Парфенов, А.Е. Шухман, Л.С. Забродина // Системная инженерия и информационные технологии. — 2021. — Т. 3, № 1 (5). — С. 50–57.
10. Бринк Х. Машинное обучение / Х. Бринк, Д. Ричардс, М. Феверолов; пер. с англ. И. Рузмайкина. — Санкт-Петербург: Питер, 2017. — 336 с.
11. Деревья решений в задачах классификации: особенности применения и методы повышения качества классификации / Я.А. Полин, Т.В. Зудилова, И.В. Ананченко, Т.Е. Войтюк. — DOI 10.17513/snt.38215 // Современные наукоемкие технологии. — 2020. — № 9. — С. 59–63.

### REFERENCES

1. Dadalova M.V. Effect of Enterprise Productive Capacity on its Economic Stability. *Vestnik Irkutskogo gosudarstvennogo tehnikeskogo universiteta = Proceedings of Irkutsk State Technical University*, 2015, no. 1, pp. 137–141. (In Russian).
2. Pigoleva I.V., Milovidov K.N. World Oil and Gas Chemistry: Major Development Trends in Conditions of Transformation of World Energy. *Mikroekonomika = Microeconomics*, 2018, no. 5, pp. 22–30. (In Russian).
3. Goncharenko S.N., Sazonova Z.A. Models and Methods for Optimizing the Plan for Production and Primary Oil Refining. *Gornyi informatsionno-analiticheskii byulleten' = Mining Informational and Analytical Bulletin*, 2008, no. S10, pp. 221–229. (In Russian).
4. Titkova I.K. Herfindal Index and Entropy Index as Economic Diversification Indicators. In Dremova L.A. (ed.). *World Experience and Economics of Russian Regions. Materials of the XV All-Russian Student Scientific Conference with International Participation, Kursk, April 7, 2017*. Kursk, 2017, pp. 375–377. (In Russian).

5. Shkodinsky S.V., Rykova I.N., Nazarov A.G. Development of Industrial Enterprises: Fixed Capital Investment and Regional Asymmetry. *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Ekonomika = Bulletin of the Moscow Region State University. Series: Economy*, 2020, no. 2, pp. 116–130. (In Russian). DOI: 10.18384/2310-6646-2020-2-116-130.

6. Kichigin N.V. Suspension, Restriction, Termination of Environmentally Hazardous Economic Activities: a Measure of Environmental Liability or a Way to Prevent Environmental Damage? *Constitutional and Legal Basis of Responsibility in the Field of Ecology. Collection of Materials of International Scientific Conferences, Moscow, December 20, 2018 — March 14, 2019*. Moscow, 2019, pp. 305–309. (In Russian).

7. Palmov S.V., Deniskova A.O. Random Forest Key Features. *Science Today: Theoretical and Practical Aspects. Materials of International Scientific and Practical Conference, Vologda, December 28, 2016*. Vologda, 2017, pp. 51–53. (In Russian).

8. Perrier A. Feature Importance in Random Forests. *Data Science*. Available at: <https://alexisperrier.com/datascience/2015/08/27/feature-importance-random-forests-gini-accuracy.html>.

9. Bolodurina I.P., Parfenov D.I., Shukhman A.E., Zabrodina L.S. Automated Machine Learning: Overview of the Capabilities of Modern Data Analysis Platforms. *Sistemnaya inzheneriya i informatsionnye tekhnologii = Systems Engineering and Information Technology*, 2021, vol. 3, no. 1, pp. 50–57. (In Russian).

10. Brink H., Richards J., Fetherolf M. *Real-World Machine Learning*. New York, Manning Publ., 2016. 264 p. (Russ. ed.: Brink H., Richards J., Fetherolf M. *Real-World Machine Learning*. Saint Petersburg, Piter Publ., 2017. 336 p.).

11. Polin Ya.A., Zudilova T.V., Ananchenko I.V., Voytiuk T.E. Decision Trees in Classification Problems: Application Features and Methods for Improving the Quality of Classification. *Sovremennye naukoemkie tekhnologii = Modern High Technologies*, 2020, no. 9, pp. 59–63. (In Russian). DOI: 10.17513/snt.38215.

#### Информация об авторах

Буров Кирилл Олегович — аспирант, кафедры экономики предприятия и предпринимательской деятельности, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: burovkirill1995@gmail.com.

Шуплецов Александр Федорович — доктор экономических наук, профессор, заведующий кафедрой экономики предприятия и предпринимательской деятельности, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: ShupletsovAF@bgu.ru.

#### Вклад авторов

Все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

#### Для цитирования

Буров К.О. Выявление значимых (решающих) факторов, влияющих на выбор размещения производственных мощностей компаний нефтеперерабатывающей отрасли / К.О. Буров, А.Ф. Шуплецов. — DOI 10.17150/2500-2759.2022.32(1).39-56 // Известия Байкальского государственного университета. — 2022. — Т. 32, № 1. — С. 39–56.

#### Authors

Kirill O. Burov — Ph.D. Student, Department of Enterprise Economics and Entrepreneurship, Baikal State University, Irkutsk, the Russian Federation, e-mail: burovkirill1995@gmail.com.

Alexander F. Shupletsov — D.Sc. in Economics, Professor, Head of the Department of Enterprise Economics and Entrepreneurship, Baikal State University, Irkutsk, the Russian Federation, e-mail: ShupletsovAF@bgu.ru.

#### Contribution of the authors

The authors contributed equally to this article. The authors declare no conflicts of interests.

#### For Citation

Burov K.O., Shupletsov A.F. Identification of Significant (Decisive) Factors Influencing the Choice of the Allocation of Production Capacities of the Companies in the Refining Industry. *Izvestiya Baikal'skogo gosudarstvennogo universiteta = Bulletin of Baikal State University*, 2022, vol. 32, no. 1, pp. 39–56. (In Russian). DOI: 10.17150/2500-2759.2022.32(1).39-56.