ISSN 1999-4125 (Print) **ISSN 2949-0642 (Online)**

Научная статья
УДК 519.683
DOI: 10.26730/1999-4125-2026-3-195-204

ОПРЕДЕЛЕНИЕ КАЧЕСТВЕННЫХ ХАРАКТЕРИСТИК УГЛЯ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Линник Юрий Николаевич, Линник Владимир Юрьевич*

Государственный университет управления

* для корреспонденции: d0c3n7@gmail.com

Аннотация.
Основная задача угледобывающих предприятий в современных экономических условиях заключается не просто в наращивании валового объема добычи, а в комплексном обеспечении поставок топлива строго заданного качественного состава при одновременной минимизации эксплуатационных и капитальных затрат на весь цикл горных работ. В условиях волатильности рыночных цен и ужесточения экологических требований к сжигаемому топливу именно точное прогнозирование ключевых качественных характеристик угля (таких как зольность, выход летучих веществ, теплотворная способность и содержание серы) выходит на первый план. Это позволяет оптимизировать процессы обогащения, шихтовки и логистики, обеспечивая максимально эффективное использование энергетического потенциала каждой тонны добытого сырья. В связи с этим целью настоящей научной статьи является не только поверхностный обзор, но и глубокое практическое исследование, а также сравнительная оценка группы распространенных алгоритмов машинного обучения и искусственного интеллекта применительно к реальной производственной задаче прогнозирования качества твердого топлива. Эмпирическая база исследования сформирована на основе обширного массива данных лабораторного анализа, которые собирались и верифицировались в течение пятилетнего периода (с 2015 по 2020 год включительно). Выборка включает в себя 33 256 представительных образцов углей, отобранных в различных пластах Кузнецкого угольного бассейна, что обеспечивает высокую статистическую значимость результатов. В ходе экспериментального моделирования анализировались четыре принципиально разных подхода: дерево решений C4.5, метод ближайших соседей IBk, наивный байесовский классификатор и многослойный перцептрон (MLP). Ключевая идея работы заключалась в поиске наиболее адекватной модели: для этого каждый из алгоритмов подвергался тщательной калибровке вероятностных выходов с целью получения несмещенных предсказаний, оценивалось влияние каждого входного параметра на итоговый результат, после чего проводилось итоговое ранжирование методов по критериям точности и устойчивости. По результатам комплексной оценки именно MLP был признан наиболее эффективным инструментом для данной предметной области, а его топология с оптимально подобранным количеством нейронов во входном, скрытом и выходном слоях продемонстрировала наилучшее соотношение скорости обучения и обобщающей способности.

Информация о статье
Поступила:
17 февраля 2026 г.

Одобрена после рецензирования:
15 июня 2026 г.

Принята к публикации:
01 июля 2026 г.

Ключевые слова:
алгоритмы искусственного интеллекта, интеллектуальный анализ данных, классификационная прогнозная модель, качество угля

Для цитирования: Линник Ю.Н, Линник В.Ю. Определение качественных характеристик угля с использованием алгоритмов искусственного интеллекта // Вестник Кузбасского государственного

технического университета. 2026. № 3 (175). С. 195-204. DOI: 10.26730/1999-4125-2026-3-195-204, EDN: EKMGYP

Введение

1. В условиях глобализации, развития техники и технологий уголь, запасы которого превышают запасы всех других ископаемых видов топлива, обеспечивает долгосрочную энергетическую безопасность. Ведущие мировые институты и агентства, используя собственные методики, разрабатывают различные тренды удовлетворения мировых энергетических потребностей. Так, проект долгосрочной программы развития угольной промышленности до 2050 года, представленный Минэнерго России, декларирует, что по ключевым мировым показателям Россия уверенно входит в число лидеров: занимает третье место по объемам экспорта и шестое по объемам добычи, т. е. около 5% от общемировой. По запасам наша страна находится на пятой позиции, обладая примерно 6,9% мировых ресурсов угля [1].

2. В России уголь сохраняет стратегическое значение для энергобезопасности. На его долю приходится около 15% выработки электроэнергии в целом по России, а в Сибири и на Дальнем Востоке этот показатель достигает 50%. Уголь также является основой теплоснабжения: обеспечивает 22% тепла в среднем по стране и до 70% – в восточных регионах [1].

3. В долгосрочной перспективе угольная отрасль сохранит свою важность. Уголь остается ключевым ресурсом в Энергетической стратегии России, чему способствуют огромные неразработанные запасы и его вклад в экономику. При этом акцент делается на внедрение современных технологий, которые позволяют минимизировать экологический след угольной генерации, сохраняя ее доступность и надежность [1].

Так, Управление энергетической информации США в своем ежегодном отчете опубликовало прогноз по потреблению угля до 2050 года. Согласно этому прогнозу, в базовом сценарии, который не учитывает будущие изменения в законах и возможные изменения в геополитике, потребление угля в мире вырастет в период с 2026 по 2050 годы на 5,6% [2].

В последние годы анализ данных как концепция обработки информации представляет собой вычислительный метод обработки данных, получив широкое распространение во всех отраслях, в т. ч. в угольной промышленности. Анализ данных и машинное обучение состоят из множества методик и алгоритмов, направленных на получение полезных знаний из данных. Основу интеллектуального анализа данных составляет конечный набор данных о процессе, происходящем в реальном мире. При анализе процесса с использованием интеллектуального анализа данных принимаются следующие предпосылки:

- анализируемый процесс четко определен, органичен конкретными рамками и непрерывен;
- цель анализа ясна;
- имеющиеся данные обладают достаточным качеством для описания процесса;
- данные представлены в виде таблицы переменных и сущностей.

Необходимо ограничивать исследуемый процесс от прочих процессов, происходящих при добыче угля, поскольку используемые вычислительные методы требуют значительных ресурсов времени и памяти. Благодаря широкому спектру применения способность этих методов решать сложные задачи может быть адаптирована для решения проблем прогнозирования энергетических характеристик угля. Таким образом, прогнозирование энергетических характеристик является задачей высокой важности. Качество угля – это свойство, зависящее от его химического состава и указывающее на полезную энергетическую ценность, заключенную в угле. В этой связи для его описания мы используем две меры: высшую теплоту сгорания и низшую теплоту сгорания. В настоящее время описан ряд методов определения этих значений, например [3, 4]. На эти показатели оказывают существенное влияние такие характеристики, как относительное содержание влаги, золы, летучих веществ и углерода в угле. С другой стороны, необходимо учитывать содержание химических элементов: углерода (C), водорода (H), кислорода (O), серы (S) и других элементов в образце угля. Химический анализ проводится лабораторным способом, в то время как анализ низшей и высшей температуры сгорания может быть выполнен с помощью бомбового калориметра [4]. В последние годы в исследованиях, посвященных методам прогнозирования теплотворной способности угля, все чаще применяются методы машинного обучения. В работе [5] описывается интерпретируемая модель машинного обучения для прогнозирования высшей теплоты сгорания угля. Для прогнозирования использовались четыре алгоритма машинного обучения: случайный лес, метод опорных векторов, градиентный бустинг и экстремальный градиентный бустинг.

Предложенный в [6] гибридный подход с использованием ансамбля моделей с весами на основе анализа связей позволил в 1.5-2 раза повысить точность прогноза высшей теплоты сгорания угля на новых месторождениях.

В предлагаемой статье сравниваются четыре различных метода анализа данных: дерево решений, метод k-ближайших соседей, наивный байесовский классификатор и нейросеть. Исследование проведено на основе разработанной ранее авторами настоящей статьи базы данных [7]. В качестве объекта исследования использовались данные по Кузнецкому угольному бассейну. Анализ проводился после обучения и тестирования алгоритмов с целью получения оптимального алгоритма для прогнозирования класса качества угля.

Методы

Основные характеристики Кузнецкого угольного бассейна. Кузнецкий угольный бассейн является одним из крупнейших и наиболее значимых угольных бассейнов не только России, но и мира. По структуре и природным характеристикам он представляет собой уникальный и сложный геологический комплекс, обладающий огромными запасами высококачественного угля.

Общие геологические запасы угля в Кузбассе превышают 700 миллиардов тонн, из которых разведанные и оцененные балансовые запасы составляют около 57 миллиардов тонн. Эти запасы представляют собой мощную основу для энергетической безопасности России, обеспечивая сырьем как традиционные тепловые электростанции, так и современные технологии углехимии и глубокой переработки угля. Доля Кузбасса в общероссийской добыче угля устойчиво превышает 55-60%.

Угленосная толща Кузбасса занимает площадь порядка 26 000 км² в пределах обширной Кузнецкой котловины. В пределах бассейна выявлено более 300 угольных пластов суммарной рабочей мощностью до 400 метров. Мощность отдельных пластов варьируется в пределах 1-3 м. Угли отличаются исключительно высоким качеством и разнообразием марок. Значительная часть запасов представлена коксующимися углями (около 30-40% от общих запасов), имеющими критическое значение для металлургической промышленности. Также широко распространены энергетические угли.

Низшая теплота сгорания кузнецких углей колеблется в очень широком диапазоне в зависимости от марки и месторождения: от ~16 000 кДж/кг для высококачественных антрацитов и коксующихся углей до ~5 000 кДж/кг для бурых углей. Одной из ключевых отличительных особенностей углей Кузбасса, помимо высокой теплотворной способности, является относительно низкое содержание серы (часто в пределах 0.3-0.6%), что делает их экологически более предпочтительным топливом.

Уголь добывается как подземным (шахтным), так и открытым (карьерным) способами, при этом доля открытой добычи в последние десятилетия превышает 60-70%. Глубина разработки шахт достигает 500 метров и более, а карьеров – до 300 метров. Основные угледобывающие предприятия (разрезы и шахты) сосредоточены в нескольких промышленных районах: Беловском, Ленинск-Кузнецком, Прокопьевско-Киселевском, Междуреченском и др.

Подготовка набора данных

Данные, использованные в статье для прогнозирования теплоты сгорания угля, были получены из упомянутой выше базы данных [7], а также предоставлены угольными предприятиями Кузбасса. Угли с содержанием золы более 50% были исключены из выборки. В общей сложности было проанализировано более 30 000 записей базы данных по 4 шахтам. Часть анализируемых данных, отобранных случайным образом из каждых 300 тонн добытого угля, представлена в Таблице 1. Выбранное для прогноза подмножество данных содержит 7 атрибутов. Строки базы данных с пустыми значениями опущены, поскольку восстановить отсутствующие значения крайне сложно. Показатель, на основе которого можно спрогнозировать энергетические характеристики угля, основан на его низшей теплоте сгорания и может принадлежать к одному из семи классов: A (>12.000кДж/кг), B (11.000–12.000 кДж/кг), C (10.000–11.000 кДж/кг), D (9.000–10.000 кДж/кг), E (8.000–9.000 кДж/кг), F (7.000–8.000 кДж/кг) и G (<7.000кДж/кг).

Оценка прогнозной модели

В задачах машинного обучения для оценки

Таблица 1. Перечень входных переменных для построения прогноза
Table 1. Set of variables

№	Показатель	Единица измерения	Примечание
1	Зольность	%	≤ 50%
2	Содержание влаги	%	
3	Выход летучих	%	
4	Марка угля	-	
5	Сопротивляемость угля резанию	Н/мм	
6	Тип угольной складки	-	Синклиналь/антиклиналь
7	Способ добычи	-	Открытый/подземный

качества моделей и сравнения различных алгоритмов используются метрики, а их выбор и анализ – непременная часть работы аналитика. В работах [8, 9] утверждается, что метрика, подходящая для оценки модели в одной предметной области, может быть непригодна в другой, и наоборот. Выбор метрики оценки зависит от области применения и поставленной задачи. Каждая из них имеет специфические характеристики, подчеркивающие различные аспекты оценки алгоритмов [8, 9]. Такими метриками являются доля корректных классификаций (Accuracy), частота ошибок (Error rate), точность (Precision), полнота (Recall) и F-меру (F-Measure). Формулы расчета этих метрик известны и представлены в подавляющем большинстве справочных и учебных изданий по анализу данных, в частности в [8–10].

Выбор способа создания выборки

Далее решалась задача разделения набора данных. Основой всех мер для оценки прогнозной модели является измерение ее эффективности, т. е. оценка способности корректно классифицировать как можно больше образцов, которые ранее не участвовали в процессе создания модели. Поэтому в процессе генерации моделей не используются все доступные данные с известными характеристиками. Вместо этого набор данных делится на две части:

- обучающая выборка (training set), которая используется для генерации моделей;
- тестовая выборка (test set), которая используется для оценки полученной модели.

Необходимо, чтобы наборы данных были случайными и независимыми. Для обучения модели использовано четыре алгоритма, описанных ниже. Тесты проводились для каждого из выбранных алгоритмов с использованием следующих техник разделения модели: 5-кратная кросс-валидация, 10-кратная кросс-валидация, 15-кратная кросс-валидация, разделение 50/50% и стратифицированное разделение. Результаты, полученные для конкретных конфигураций тестов, различались в зависимости от выбранного алгоритма и не позволили однозначно выбрать наилучший. Тем не менее, как показали результаты, 10-кратная кросс-валидация и стратифицированная выборка продемонстрировали некоторое преимущество перед другими протестированными техниками оценки модели. В работе [9] утверждается, что для повышения качества обучения следует применять n-кратную кросс-валидацию, поэтому для итоговой оценки была выбрана 10-кратная кросс-валидация.

Прогнозные модели для классификации

Термин «интеллектуальный анализ данных» описывает процесс анализа больших объемов данных с целью обнаружения в них полезной информации [9]. Обнаружение знаний в больших объемах данных включает применение конкретных алгоритмов для выявления закономерностей в данных. В прогнозной модели одна из переменных определяется как функция других, что позволяет прогнозировать значение зависимой переменной по заданным значениям других (независимых или предикторных переменных) [11]. Вся суть в том, чтобы научить алгоритм находить скрытые закономерности, которые связывают известные параметры (например, химический состав угля) с неизвестным, но существующим свойством. После обучения модель сможет определять это скрытое свойство самостоятельно для новых образцов данных. Это проблема классификации, где модель подгоняется к набору данных, состоящему из $k \in \{1,\ldots,N\}$ примеров, каждый из которых отображает входной вектор $\{x_1^k,\ldots,x_n^k\}$ на заданную цель yk. Модель прогнозирования включает два этапа: обучение модели и ее использование. Первый этап связан с получением набора предопределенных классов на основе обучающего набора в результате обучения на этом наборе данных. Предполагается, что каждый образец в обучающем наборе принадлежит к предопределенному классу, определяемому меткой атрибута класса. Модель представляется в виде классификационных правил, деревьев решений или математических формул. Второй этап подразумевает использование модели для прогнозирования или классификации неизвестных объектов на основе закономерностей, наблюдаемых в обучающем наборе [12].

Для того, чтобы идентифицировать (спрогнозировать) класс качества угля в настоящем исследовании, авторы проанализировали четыре различных алгоритма машинного обучения. На основе этих алгоритмов были разработаны классификационные модели, целью которых являлось прогнозирование класса качества угля, к которому будет принадлежать новый образец угля.

Алгоритм C4.5. Наиболее распространенным и, вероятно, наиболее широко используемым в настоящее время алгоритмом дерева решений является C4.5 [13, 14], который строит модель на основе древовидной структуры. Дерево решений – это иерархическая структура данных для реализации стратегии «разделяй и властвуй», распространенной в компьютерных науках. Это эффективный непараметрический метод, который можно использовать для обучения модели. В непараметрических моделях входное множество данных делится на локальные области, определяемые метрикой расстояния. В дереве решений локальная область идентифицируется в последовательности

рекурсивных разбиений за наименьшее количество шагов. Дерево решений состоит из внутренних узлов решений и конечных листьев. Каждый узел m реализует тестовую функцию $f_m(x)$ с дискретными исходами, обозначающими ветви. Этот процесс начинается от корня и повторяется до достижения конечного узла (листа). Значение конечного узла является результатом классификации [13].

Многослойный перцептрон (MLP). Алгоритм многослойного перцептрона является одной из наиболее широко используемых и популярных искусственных нейронных сетей. Искусственные нейронные сети — это сети элементов, называемых нейронами, которые обмениваются информацией в форме числовых значений друг с другом через синаптические связи. Многослойный персептрон [15, 16] – это базовый вид алгоритмов глубокого обучения, моделирующих процесс человеческого восприятия, на основе модели искусственного нейрона, другое название которого – узел нейронной сети. Обработка информации начинается от входного слоя сети и распространяется через скрытые слои к выходному слою. Нейроны вычисляют взвешенную сумму входных значений с помощью весовых коэффициентов, после чего к взвешенной сумме применяется нелинейная функция активации (f). Нелинейные преобразования, выполняемые нейронами, обеспечивают персептрону и другим глубоким нейронным сетям возможность моделирования нелинейных процессов, встречающихся во многих природных явлениях и промышленных процессах. Подобные сети можно обучать несколькими способами, однако оптимальным выбором для обучения сетей MLP является метод обратного распространения ошибки (backpropagation). MLP чаще всего используется для аппроксимации классификационной функции, которая соотносит пример, определяемый значениями векторных атрибутов, с одним или несколькими классами [17].

Алгоритм IBk. Алгоритм IBk является одной из реализаций алгоритма k-ближайших соседей [18]. Метод k-ближайших соседей – это непараметрический алгоритм, основанный на мере сходства (например, функции расстояния, такой как Евклидово расстояние между двумя образцами или расстояние Хэмминга между бинарными векторами). Алгоритм использует меру сходства для поиска k ближайших соседних точек в данных выборки, и на основе наиболее часто встречающегося класса среди этих k соседей предсказывает выходное значение.

Одним из входных параметров для алгоритма k-ближайших соседей является количество соседей, которые ищутся вблизи точки выборки с неизвестной категорией. Примерная реализация и результаты работы этого алгоритма демонстрируют приемлемую производительность. Однако метод не демонстрирует достаточной точности по сравнению с алгоритмами на основе деревьев. В своей простейшей форме алгоритм k-ближайших соседей может заключаться в присвоении объекту класса его ближайшего соседа или большинства его ближайших соседей. Кроме того, благодаря своей простоте алгоритм легко модифицировать для более сложных задач классификации. Алгоритм IB1 является простейшей реализацией алгоритма IBk [18].

Наивный байесовский классификатор (NB). Данный алгоритм представляет собой простой метод классификации, основанный на теории вероятностей, а именно на теореме Байеса [19]. Использование термина «наивный» обусловлено тем, что алгоритм упрощает решаемую задачу, полагаясь на два важных предположения: прогностические атрибуты условно независимы при известной классификации, и нет скрытых атрибутов, которые могли бы повлиять на процесс прогнозирования. Такой метод классификации анализирует взаимосвязь между экземпляром каждого класса и каждым атрибутом для получения условной вероятности взаимосвязей между значениями атрибутов и классом. Классификатор NB предполагает, что значения входных признаков независимы и не влияют на класс. Такое предположение необходимо для упрощения вычислений. NB может применяться для классификации путем вычисления вероятности появления каждого класса (априорной вероятности) и вычисления вероятности появления экземпляра в заданном классе [19].

Результаты и обсуждение.

Сегодня объемы данных, получаемые на промышленных предприятиях, настолько объемны, что для их анализа, а тем более для корректной прогнозной оценки требуется применение вычислительных систем и языков программирования. Однако прежде, чем применять методы машинного обучения к набору данных, необходимо подготовить этот набор, очистить его от некорректных данных, исключить лишнюю информацию. Такая базовая подготовка данных в рассматриваемом случае направлена в том числе и на снижение размерности данных, т. е. на исключение лишних, незначимых атрибутов. Подход, примененный авторами статьи для выбора оптимального количества атрибутов, состоит из двух фаз:

-оценка значимости входных атрибутов по отношению к прогнозируемому атрибуту;

-итоговое сравнение алгоритмов с учетом их производительности для сокращения размерности входного набора данных.

Для получения детального представления о значимости атрибутов обычно проводится анализ входных переменных с целью исключения нерелевантных атрибутов из обучающего набора, что позволило бы повысить качество прогноза. Именно по причине негативного влияния нерелевантных атрибутов перед обучением проводится отбор, задача которого – исключить все, кроме тех, которые в наибольшей степени оказывают влияние на изучаемые характеристики. Наиболее приемлемый способ выбрать релевантные атрибуты – экспертный, базирующийся на мнении специалистов и глубоком понимании задачи обучения. Тесты проводились с использованием трех алгоритмов для оценки значимости входных переменных [8]: тест χ^2, gain ratio и критерий прироста информации (information gain). В качестве окончательного результата ранжирования атрибутов было принято среднее значение по всем алгоритмам. По результатам анализа атрибут, характеризующий зольность угля, показал наибольшее влияние на результат. За ним следуют атрибуты: влажность, содержание летучих, марка угля. Остальные показатели, представленные в Таблице 1 (сопротивляемость резанию, тип угольной складки, способ добычи) оказали, как и ожидалось, наименьшее влияние на результат.

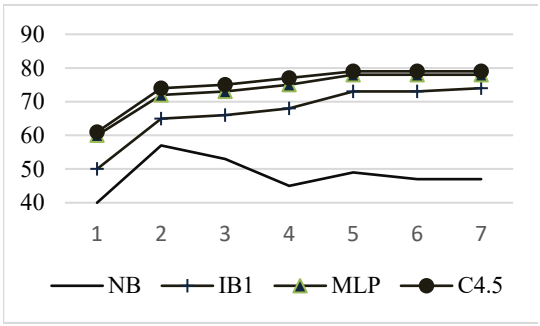

Рис. 1. Сравнение алгоритмов прогнозирования целевого атрибута – энергетических характеристик угля
Fig. 1. Overview of the percentage of correctly predicted coal quality class for each of the classification prediction models

На этапе второй фазы из входной базы данных удалялось по одному атрибуту для каждого из четырех выбранных алгоритмов классификации. Оценка и сравнение результатов каждого из четырех алгоритмов анализа данных проводились с помощью 10-кратной кросс-валидации.

Таким образом было проведено 7 экспериментов для каждой из четырех выбранных методик интеллектуального анализа данных или в общей сложности 28 экспериментов с соответствующим классификационным анализом. Сравнение алгоритмов в прогнозировании целевого показателя – энергетических характеристик угля – представлено на Рис. 1. На рисунке показан рейтинг каждой модели в зависимости от количества входных атрибутов; также видно, что наибольшая доля корректных прогнозов достигается при использовании моделей с 5-7 атрибутами.

Далее был разработан ряд прогнозных сценариев на языке программирования Python с применением библиотек машинного обучения (PyTorch, Scikit-learn). Разработка заключалась в построении прогнозной модели на основе данных, полученных на конкретной шахте Кузбасса, что позволило бы прогнозировать энергетические характеристики угля для новых образцов. В ходе эксперимента для прогнозирования энергетической ценности угля использовались все описанные выше алгоритмы. В качестве метрик производительности для оценки классификационной прогнозной модели использовались доля корректных классификаций, частота ошибок, точность, полнота и F-мера. Отобранные алгоритмы обучались с использованием техники разделения 10-кратной кросс-валидации. Результаты эксперимента сведены в Таблицу 2, где представлены результаты сравнения четырех выбранных тестовых алгоритмов для моделей с наиболее значимыми входными атрибутами.

Таблица 2. Сравнение производительности тестируемых алгоритмов
Table 2. Comparison of performance of tested algorithms

Мера \ Алгоритм	C4.5	IB1	MLP	NB
Acc (%)	83.26	80.33	86.44	49.07
Err (%)	16.74	19.67	13.56	50.93
Precision	0.832	0.803	0.864	0.536
Recall	0.833	0.803	0.864	0.491
F-Measure	0.832	0.803	0.864	0.445

Как показывает анализ данных, представленных в Таблице 2, наилучшие результаты были получены при использовании многослойного перцептрона (MLP), который показал наивысший процент правильного прогноза (86.44%). Второе место по производительности занимает алгоритм C4.5, за которым следует алгоритм IB1 с довольно приемлемой производительностью. Худшая производительность (ниже 50%) была получена у наивного байесовского классификатора. Далее

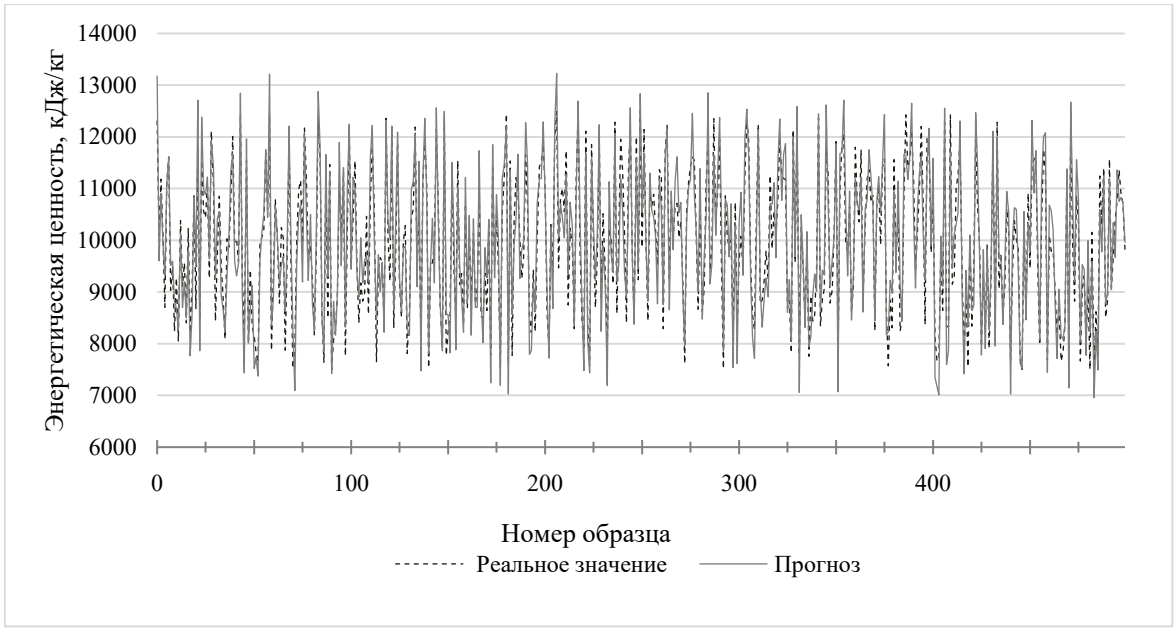

Рис. 2. Результаты тестирования модели
Fig. 2. Comparison between exact and predicted values of coal quality

авторы провели дополнительную настройку алгоритма MLP в целях его оптимизации. Для получения наилучшей сконфигурированной модели были настроены четыре параметра алгоритма MPL: количество полных циклов прохождения алгоритма через весь обучающий набор данных (количество эпох), скорость обучения и количество нейронов в скрытом слое. Это дало возможность построить оптимальную нейронную сеть, дающую наилучшие результаты. Структура оптимизированной нейронной сети составляет 14/24/7 нейронов для входного, скрытого и выходного слоев соответственно, скорость обучения равна 0,2. Такая настройка модели позволила повысить ее точность до 87,26%. Тестирование оптимизированной нейросети проводилось на 500 новых образцах угля. Реальные и прогнозные значения для каждого нового образца сетью MLP показаны на Рис. 2.

Заключение

Целью данного исследования было построить на основе описанных данных прогнозную модель, которая сможет прогнозировать, к какому классу качества принадлежит продукт для неизвестных образцов. Это исследование проводилось с использованием различных классификационных прогнозных моделей (C4.5, IBk, наивный байесовский классификатор и нейронная сеть MLP) для прогнозирования качества угля. Результаты, полученные в ходе анализа, показывают, что нейронная сеть MLP показала хорошую производительность и разумную прогностическую точность этой модели. Прогнозная модель для данной предметной области достигла оптимальной структуры (14-24-7) для входного, скрытого и выходного слоев. Результаты позволяют предположить, что эта нейронная сеть может стать важным инструментом в прогнозировании качества угля. Полученные знания могут оказать ценную помощь в принятии решений и управлении сложными системами, обеспечивая качество продукции и производства. В будущей работе исследуется возможность внедрения полученной прогнозной модели в систему онлайн-мониторинга фактических значений содержания компонентов в угле, поступающем по конвейерной ленте, что может предоставлять информацию о качестве угля в реальном времени.

СПИСОК ЛИТЕРАТУРЫ

1. Лежнева Л. Тепловой контур: Минэнерго подготовило проект стратегии развития угольной отрасли. [Электронный ресурс] // Известия. 2025. URL: https://iz.ru/1994681/liubov-lezhneva/teplovoi-kontur-minenergo-podgotovilo-proekt-strategii-razvitiia-ugolnoi-otrasli (дата обращения: 03.02.2026)

2. International Energy Outlook 2024. World coal consumption by region [Электронный ресурс] // U.S. Energy Information Administration URL: https://www.eia.gov/outlooks/aeo/data/browser/#/?id=7-IEO2023&cases=Reference&sourcekey=0 (дата обращения: 03.02.2026)

3. ГОСТ 147-2013. Топливо твердое минеральное. Определение высшей теплоты сгорания и вычисление низшей теплоты сгорания. М. : Стандартинформ, 2019. 32 с.

4. Техническое описание и руководство по эксплуатации. Калориметр бомбовый АБК-1. М. : Изд. стандартов, 2009. 38 с.

5. Lawal A. I., Ajeboriogbon A. F., Onifade M., Bada S. O., Mulenga F. A novel grey relational analysis-based committee of machine learning methods for enhanced prediction of coal calorific value // Fuel.

2026. T. 406. Art. 137070.

6. Unveiling the predictive power of machine learning in coal gross calorific value estimation: An interpretability perspective – ScienceDirect. URL: https://www.sciencedirect.com/science/article/abs/pii/S0360544225004232 (дата обращения: 03.02.2026).

7. Линник В. Ю. Формирование баз данных для прогнозирования развития сырьевой базы угольной промышленности. М. : Государственный университет управления, 2011. 102 с. ISBN 978-5-215-02370-9. EDN QMZFHT.

8. Trauth M. H. MATLAB® Recipes for Earth Sciences. Springer Cham. 2021. 517 p.

9. De S., Dey S., Bhattacharyya S., Bhatia S. Advanced Data Mining Tools and Methods for Social Computing. Elsevier Science & Technology Books, 2022. 270 p.

10. Бослаф С. Статистика для всех [Текст]. Пер. с англ. Волкова П. А., Флямер И. М., Либерман М. В., Галицына А. А. М. : ДМК Пресс, 2015. 586 с. : ил..; ISBN 978-5-94074-969-1.

11. Han J., Tong H. Data Mining. Concepts and Techniques. 4nd ed. Morgan Kaufmann, 2022. 752 p.

12. Kumar P.-N. Introduction to data mining. 2nd ed. Pearson India, 2021. 1000 p.

13. Quinlan J. R. Improved Use of Continuous Attributes in C4.5 (англ.) // Journal of Artificial Intelligence Research. 1996. Vol. 4. Pp. 77–90. ISSN 1076-9757. DOI: 10.1613/jair.279.

14. 1. Pandey Y. N. Machine Learning in the Oil and Gas Industry: Including Geosciences, Reservoir Engineering, and Production Engineering with Python. Machine Learning in the Oil and Gas Industry. Berkeley, CA : Apress L.P., 2020.

15. Галушкин А. И. Синтез многослойных систем распознавания образов. М. : «Энергия», 1974. 367 с.

16. Bao Q., Li B., Wang L., Liao L. SeMi_Detector: Multilayer Perceptron-Based Selfish Mining Detection // Peer-to-Peer Networking and Applications. 2025. Vol. 18. № 6. Art. 327.

17. Pardoe I. Applied regression modeling. John Wiley & Sons, 2020.

18. Yuan H., Xu W. A novel method of feature selection and information fusion for multi-source ordered information systems based on k-nearest neighbor rough sets // Information Sciences. 2026. T. 734. Art. 122997.

19. Berrar D. Bayes' Theorem and Naive Bayes Classifier // Encyclopedia of Bioinformatics and Computational Biology. Elsevier. 2025. Pp. 483–494.

Авторы заявляют об отсутствии конфликта интересов.

Об авторах:
Линник Юрий Николаевич, профессор, ФГБОУ ВО «Государственный университет управления» (109542, Российская Федерация, г. Москва, ул. Рязанский проспект, 99), доктор технических наук, профессор, ORCID: 0000-0003-3968-0026, e-mail: ylinnik@rambler.ru
Линник Владимир Юрьевич, профессор, ФГБОУ ВО «Государственный университет управления» (109542, Российская Федерация, г. Москва, ул. Рязанский проспект, 99), доктор экономических наук, доцент, ORCID: https://orcid.org/0000-0001-5130-8222, e-mail: d0c3n7@gmail.com

Заявленный вклад авторов:
Линник Юрий Николаевич – постановка исследовательской задачи, концептуализация исследования, научный менеджмент.
Линник Владимир Юрьевич – сбор и анализ данных, выводы, написание текста, обзор соответствующей литературы, написание текста.

Все авторы прочитали и одобрили окончательный вариант рукописи.

Original article

DETERMINATION OF COAL QUALITY CHARACTERISTICS USING ARTIFICIAL INTELLIGENCE ALGORITHM

Yuri N. Linnik, Vladimir Yu. Linnik *

State University of Management

* for correspondence: d0c3n7@gmail.com

Abstract.
The primary objective of coal producers in today's economic environment is not merely to increase gross output volumes, but to ensure the comprehensive supply of fuel that strictly meets specified quality parameters, while simultaneously minimizing operating and capital expenditures across the entire mining cycle. Given the volatility of market prices and the tightening of environmental standards for combusted fuel, accurate prediction of key coal quality attributes—such as ash content, volatile matter yield, calorific value, and sulfur content—becomes paramount. This enables the optimization of beneficiation, blending, and logistics processes, thereby maximizing the energy utilization potential of every ton of extracted raw material. Consequently, the aim of this paper is not only to provide a superficial review, but to conduct an in-depth practical investigation and comparative evaluation of a set of common machine learning and artificial intelligence algorithms applied to a real-world production task—forecasting coal quality. The empirical foundation of the study is built upon an extensive dataset of laboratory analyses, collected and verified over a five-year period (2015–2020). The sample comprises 33,256 representative coal specimens sourced from various seams of the Kuznetsk coal basin, ensuring high statistical significance of the findings. In the experimental modelling, four fundamentally different approaches were analyzed: the C4.5 decision tree, the IBk nearest neighbor method, the naïve Bayes classifier, and the multilayer perceptron (MLP). The core idea of the work was to identify the most suitable model: each algorithm underwent careful probabilistic output calibration to obtain unbiased predictions, the influence of each input parameter on the final result was assessed, and a final ranking of the methods was performed based on accuracy and robustness criteria. Based on the comprehensive evaluation, the MLP was recognized as the most effective tool for this domain, with its topology—featuring optimally tuned numbers of neurons in the input, hidden, and output layers—demonstrating the best trade-off between training speed and generalization capability.

Article info
Received:
17 February 2026

Accepted for publication:
15 June 2026

Accepted:
01 July 2026

Keywords: *artificial intelligence algorithms, intelligent data analysis, classification predictive model, coal quality*

For citation: Linnik Yu.N., Linnik V.Yu. Determination of coal quality characteristics using artificial intelligence algorithm. *Vestnik Kuzbasskogo gosudarstvennogo tekhnicheskogo universiteta*=Bulletin of the Kuzbass State Technical University. 2026; 3(175):195-204. (In Russ., abstract in Eng.). DOI: 10.26730/1999-4125-2026-3-195-204, EDN: EKMGYP

REFERENCES

1. Lezhneva L. Thermal contour: Minenergo has prepared a draft strategy for the development of the coal industry. [Electronic resource] // Izvestia. 2025. URL: https://iz.ru/1994681/liubov-lezhneva/teplovoi-kontur-minenergo-podgotovilo-proekt-strategii-razvitiia-ugolnoi-otrasli (accessed on 03.02.2026).

2. International Energy Outlook 2024. World Coal Consumption by Region [Electronic resource]. // U.S. Energy Information Administration. URL: https://www.eia.gov/outlooks/aeo/data/browser/#/?id=7-IEO2023&cases=Reference&sourcekey=0 (accessed on 03.02.2026).

3. GOST 147-2013. Solid mineral fuel. Determination of higher calorific value and calculation of lower calorific value. M.: Standartinform; 2019. 32 p.

4. Technical Description and Operating Instructions. Bomb Calorimeter ABK-1. M.: Publishing House of Standards, 2009. 38 p.

5. Lawal A.I., Ajeboriogbon A.F., Onifade M., Bada S.O., Mulenga F. A novel grey relational analysis-based committee of machine learning methods for enhanced prediction of coal calorific value. *Fuel.* 2026; 406. Art. 137070.

6. Unveiling the predictive power of machine learning in coal gross calorific value estimation: An interpretability perspective – ScienceDirect. [Electronic resource] URL: https://www.sciencedirect.com/science/article/abs/pii/S0360544225004232 (accessed on 03.02.2026).

7. Linnik V.Yu. Formation of databases for forecasting the development of raw material base in coal industry. M.: State University of Management, 2011. 102 p. ISBN 978-5-215-02370-9. EDN QMZFHT.

8. Trauth M.H. MATLAB® Recipes for Earth Sciences. Springer Cham. 2021. 517 c.

9. De S., Dey S., Bhattacharyya S., Bhatia S. Advanced Data Mining Tools and Methods for Social Computing. Elsevier Science & Technology Books, 2022. 270 c.

10. Boslaff S. Statistics for Everyone. Translated from English by Volkov P.A., Flyamer I.M., Liberman M.V., Galitsyna A.A. M.: DMK Press; 2015. 586 p. ISBN 978-5-94074-969-1.

11. Han J., Tong H. Data Mining. Concepts and Techniques. 4nd ed. Morgan Kaufmann, 2022. 752 p.

12. Kumar P.-N. Introduction to data mining. 2nd ed. Pearson India, 2021. 1000 p.

13. Quinlan J.R. Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research.* 1996; 4:77-90. ISSN 1076-9757. DOI: 10.1613/jair.279.

14. 1. Pandey Y.N. Machine Learning in the Oil and Gas Industry: Including Geosciences, Reservoir Engineering, and Production Engineering with Python. Machine Learning in the Oil and Gas Industry. Berkeley, CA: Apress L.P.; 2020.

15. Galushkin A.I. Synthesis of Multilayer Pattern Recognition Systems. M.: Energiya Publishers; 1974. 367 p.

16. Bao Q., Li B., Wang L., Liao L. SeMi_Detector: Multilayer Perceptron-Based Selfish Mining Detection. *Peer-to-Peer Networking and Applications.* 2025; 18(6). Art. 327.

17. Pardoe I. Applied regression modeling. John Wiley & Sons, 2020.

18. Yuan H., Xu W. A novel method of feature selection and information fusion for multi-source ordered information systems based on k-nearest neighbor rough sets. *Information Sciences.* 2026; 734. Art. 122997.

19. Berrar D. Bayes' Theorem and Naive Bayes Classifier. Encyclopedia of Bioinformatics and Computational Biology. Elsevier. 2025. Pp. 483-494.

The authors declare no conflict of interest.

About the authors:
Yuriy N. Linnik, professor, «State University of Management» (109542, Russian Federation, Moscow, Ryazansky prospect, 99, Ryazansky str.), Doctor of Technical Sciences, Professor, ORCID: https://orcid.org/0000-0003-3968-0026, e-mail: ylinnik@rambler.ru
Vladimir Yu. Linnik, professor, «State University of Management» (109542, Russian Federation, Moscow, 99, Ryazansky Prospekt St., Moscow), Doctor of Economic Sciences, Associate Professor, ORCID: https://orcid.org/0000-0001-5130-8222, e-mail: d0c3n7@gmail.com.

Contribution of the authors:
Yuriy N. Linnik – research problem statement; conceptualisation of research, scientific management.
Vladimir Yu. Linnik – data collection and analysis, drawing the conclusions, reviewing relevant literature, writing the text.

All authors have read and approved the final manuscript.