

© 2026 г. В.А. ИВАНЮК, д-р экон. наук (VAIvanyuk@fa.ru)
(Финансовый университет при Правительстве РФ, Москва)

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В АНАЛИЗЕ И ПРОГНОЗИРОВАНИИ МАЛЫХ ВЫБОРОК МАКРОЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ ЭНЕРГЕТИЧЕСКОГО СЕКТОРА РФ

Методологию прогнозирования часто используют для анализа в социально-экономических системах. Применение действенных методик прогнозирования предоставляет организациям возможность предвидеть будущие изменения, оптимизировать распределение ресурсов и уменьшить вероятные риски. Особую значимость эффективное прогнозирование приобретает в области энергетики. Здесь предсказание будущих изменений в спросе и предложении энергии критически важно для поддержания устойчивости энергосети, сокращения эксплуатационных издержек и повышения надежности. Задача данного исследования – оценить результативность методов статистического и нейросетевого моделирования при прогнозировании макроэкономических индикаторов энергетического сектора Российской Федерации.

Ключевые слова: методы машинного обучения, ансамблевый прогноз, энергобаланс, статистика.

DOI: 10.7868/S2413977726050025

1. Введение

В условиях заметных изменений в энергетической сфере и растущей заинтересованности в устойчивом развитии важно изучать вопросы прогнозирования энергетического баланса. Энергобаланс служит отражением взаимозависимости между производственными мощностями и потреблением внутри энергетической системы. Чтобы добиться эффективного анализа колебаний в энергобалансе, необходимо учитывать обширный перечень экономических, экологических и технических аспектов.

Энергетический баланс – это не просто производство и расход энергии. Он включает в себя куда больше факторов. К примеру, важно учитывать применение как возобновляемых источников, так и исчерпаемых. Не стоит забывать про сезонные изменения в потребностях населения, а также про влияние промышленности и других отраслей экономики. Все это имеет решающее значение.

Актуальность выбранной темы продиктована потребностью в преодолении общеэкономической проблемы нехватки электроэнергии.

Статья имеет следующую структуру. В разделе 2 дается описание сбора и обработки данных для построения прогноза. Раздел 3 содержит анализ методов прогнозирования. Раздел 4 содержит результаты ансамблевого прогноза. В разделе 5 представлены заключительные результаты.

2. Сбор и обработка данных для построения прогноза

С целью получения актуальных данных были изучены ежегодные макроэкономические отчеты за периоды 2005–2020 гг. из Единой межведомственной информационно-статистической системы РФ (ЕМИСС) [1, 2]. При этом сводные таблицы данных обладали различными форматами, кодировкой и названиями показателей. В связи с этим было произведено ручное транспонирование, приведение данных к общему формату и унификация названий показателей.

Для получения данных за периоды 2021–2022 гг. были получены ежегодные сводные статистические отчеты по энергобалансу РФ с официального сайта Министерства энергетики [3]. Поскольку эти отчеты содержали данные об относительном (процентном) изменении показателей год к году, был проведен последовательный перевод относительных значений в натуральные показатели в миллионах тонн условного топлива.

На следующем этапе проведен глубокий анализ источников данных, позволивший убедиться в отсутствии в свободном доступе данных об энергетическом балансе РФ за период 2023–2024 гг. После получения подтверждения о максимально возможной полноте полученных макроэкономических данных было произведено объединение данных за периоды 2005–2020 и 2021–2022 гг. и приведение их к единому формату. В результате сформирована таблица показателей, в которой представлены данные по дате, типу энергоресурса и по отрасли потребления.

Получен массив данных из 29 блоков по 8 временным рядам (выборочных значений) с длиной каждого ряда в 18 последовательных значений. При этом два значения в каждом блоке являются агрегаторами, так что общее количество исходных данных в таблице можно определить как 4032 значения. В табл. 1 представлены фрагменты исходных данных.

Создадим на основе методов машинного обучения и статистики прогнозы на следующие 7 лет (2023–2030 гг.).

3. Прогнозирование малых выборок макроэкономических показателей энергетического сектора РФ

3.1. Линейный прогноз на основе методов машинного обучения

Оценка и прогнозирование методом линейной регрессии предполагает ряд условий, как то: нормальность распределения выборки, устойчивость (робастность) прогнозируемого тренда, незначительное количество выбросов и гомоскедастичность ряда. Несмотря на то, что линейный метод является самым быстрым, он крайне чувствителен к выбросам (не робастен) и дает самый большой доверительный интервал (ошибку прогноза). Применять этот метод стоит лишь в тех случаях, когда скорость получения результата гораздо важнее, чем его точность [4–6].

Методика. Для использования линейного метода прогнозирования при помощи машинного обучения необходимо определить допустимые границы значений прогнозируемой величины. Стоит отметить, что такое предварительное определение допустимых границ значений функций и их аргументов необходимо практически во всех методах машинного обучения, использующих градиентную или стохастическую оптимизацию. Обычно такие границы определяются экспертными методами (как, например, граница годового объема газодобычи, зависящая не столько от планирования, сколько от наличия фактических природных, материальных и трудовых ресурсов). В случае невозможности получения экспертных значений границы определяют, как верхний и нижний уровень правдоподобия, соответствующий стандартному диапазону отклонений в два межквартильных интервала дисперсии выборки, или как логически обоснованные значения, соответствующие теоретически достижимым максимуму и минимуму.

Следует также отметить, что помимо ограничений, обеспечивающих нахождение значений в интервале правдоподобия, к линейным прогнозам также применяют метод регуляризации по Тихонову, заключающийся в устранении выбросов выборки на основе гипотезы о нормальном распределении исходных данных с последующей линейной импутацией средних значений взамен выявленных аномалий. Также широко распространен метод линейного прогнозирования с календарной импутацией, когда выбросы определяют не на основе статистических моментов высших порядков, а исходят из гипотезы их связанности с известными кризисными событиями, в этом случае импутацию значений, соответствующих кризисным моментам, производят смарт-методом (реже – линейным методом).

Таким образом, общий алгоритм линейного прогнозирования выполняют в следующем порядке:

1. Определяют границы допустимых значений прогноза.
2. При необходимости проводят календарную или статистическую регуляризацию.
3. Вычисляют уравнение линейной регрессии.
4. Проводят проверку и замену значений прогноза, выходящих за границы.

Для математической реализации логической составляющей четвертого пункта алгоритма обычно используют кусочно-постоянные функции, такие как:

$$\text{функция Кронекера: } \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

$$\text{функция знака: } \text{sgn } i = \begin{cases} +1, & i > 0, \\ 0, & i = 0, \\ -1, & i < 0, \end{cases}$$

$$\text{функция Хевисайда: } H_i = \begin{cases} 1, & i \geq 0, \\ 0, & i \leq 0. \end{cases}$$

Используя данные функции как элементы произведений, принимающие значения 1 или 0, их можно применить для включения и исключения блоков значений в аддитивных уравнениях.

В целом уравнение линейного прогноза выглядит следующим образом:

$$\begin{aligned}
 y_{n+f} = & H(k_1)H(Y_{+0} - (k_1(n+f) + k_2))(k_1(n+f) + k_2) + \\
 & + H(-k_1)H(-Y_{-0} + (k_1(n+f) + k_2))(k_1(n+f) + k_2) + \\
 & + H(k_1)H(-Y_{+0} + (k_1(n+f) + k_2))Y_{+0} + \\
 & + H(-k_1)H(Y_{-0} - (k_1(n+f) + k_2))Y_{-0},
 \end{aligned}$$

где

$$\begin{aligned}
 k_1 \text{ (коэффициент наклона)} &= \frac{n \sum_{i=1}^n i Y_i - \sum_{i=1}^n i \sum_{i=1}^n Y_i}{\sum_{i=1}^n i^2 - (\sum_{i=1}^n i)^2}, \\
 k_2 \text{ (коэффициент базы)} &= \frac{\sum_{i=1}^n Y_i - \left(\frac{\sum_{i=1}^n i Y_i - \sum_{i=1}^n i \sum_{i=1}^n Y_i}{\sum_{i=1}^n i^2 - (\sum_{i=1}^n i)^2} \right) \sum_{i=1}^n i}{n},
 \end{aligned}$$

y_{n+f} – значение прогноза на заданную дату, Y_i – элементы выборки, Y_{+0} – верхняя граница допустимых значений прогноза, Y_{-0} – нижняя граница допустимых значений прогноза, n – объем выборки, f – прогностический интервал, $y_i = k_1 i + k_2$ – уравнение линейной регрессии.

Объяснение логических составляющих:

$H(k_1) = 1$, если функция возрастает;

$H(-k_1) = 1$, если функция убывает;

$H(Y_{+0} - (k_1(n+f) + k_2)) = 1$, если функция не пересекла верхнюю границу;

$H(-Y_{-0} + (k_1(n+f) + k_2)) = 1$, если функция не пересекла нижнюю границу;

$H(-Y_{+0} + (k_1(n+f) + k_2)) = 1$, если функция пересекла верхнюю границу;

$H(Y_{-0} - (k_1(n+f) + k_2)) = 1$, если функция пересекла нижнюю границу.

3.2. Модель прогнозирования с экспоненциальным сглаживанием ОТС (ETS)

Сокращение ОТС (ошибка/тренд/сезонность или error/trend/seasonality) означает, что при построении прогноза учитываются следующие предиктивные факторы:

1. Общая тенденция роста или спада (тренд).
2. Сезонные (периодические/гармонические) колебания.
3. Возрастание значимости ошибки по мере приближения к точке начала прогноза.

ОТС-прогноз также называют прогнозом с экспоненциальным сглаживанием, что связано с методикой вычисления значимости ошибки (экспоненциальным усилением).

Как и в линейном прогнозе, основным критерием точности аппроксимации в ОТС-методе являются квадраты отклонений, однако при этом вычисляется не сумма квадратов, а сумма произведений квадратов на их веса, экспоненциально увеличивающиеся по мере приближения к точке прогноза [7–9].

Результатом прогноза является аддитивная функция, состоящая из линейной и периодической составляющих [10–14].

Алгоритм построения прогноза:

Шаг 1. Определение границ допустимых значений прогноза.

Шаг 2. Выделение линейного тренда классическим диагональным методом.

Шаг 3. Аппроксимация остатков тренда периодической функцией.

Шаг 4. Экспертное определение значения «инерции данных».

Шаг 5. Аппроксимация выборки линейной функцией с учетом весов ошибки.

Шаг 6. Суммирование линейной и периодической функций.

Шаг 7. Формирование прогноза суммарным трендом.

Шаг 8. Проверка и замена значений прогноза, выходящих за границы.

Формализуем реализацию ОТС-прогноза как часть пунктов (шагов) математически:

Шаг 2. Выделение линейного тренда классическим диагональным методом:

$$y_{i1} = \frac{n \left(n \sum_{i=1}^n i y_i - \sum_{i=1}^n i \sum_{i=1}^n y_i \right)}{\sum_{i=1}^n i^2 - \left(\sum_{i=1}^n i \right)^2} + \frac{\sum_{i=1}^n y_i - \left(\frac{\sum_{i=1}^n i y_i - \sum_{i=1}^n i \sum_{i=1}^n y_i}{\sum_{i=1}^n i^2 - \left(\sum_{i=1}^n i \right)^2} \right) \sum_{i=1}^n i}{n},$$

где y_{i1} – начальный линейный тренд для получения сезонных остатков, y_i – элементы выборки, i – индексы элементов, n – объем выборки,

$$k_1 \text{ (коэффициент наклона)} = \frac{n \sum_{i=1}^n i Y_i - \sum_{i=1}^n i \sum_{i=1}^n Y_i}{\sum_{i=1}^n i^2 - \left(\sum_{i=1}^n i \right)^2},$$

$$k_2 \text{ (коэффициент базы)} = \frac{\sum_{i=1}^n Y_i - \left(\frac{\sum_{i=1}^n i Y_i - \sum_{i=1}^n i \sum_{i=1}^n Y_i}{\sum_{i=1}^n i^2 - \left(\sum_{i=1}^n i \right)^2} \right) \sum_{i=1}^n i}{n},$$

где Y_i – элементы выборки.

Шаг 3. Аппроксимация остатков тренда периодической функцией:

$$\sum_{i=1}^n (k_3 \varphi(k_4 i) - (k_1 i + k_2 - Y_i))^2 \rightarrow \min,$$

где k_1 – коэффициент наклона, k_2 – коэффициент базы, k_3 – подбираемый амплитудный коэффициент периодической функции, k_4 – подбираемый частотный коэффициент периодической функции.

Шаг 5. Аппроксимация выборки линейной функцией с учетом весов ошибки:

Значение коэффициента инерции данных $\alpha \in [0; 1]$ обычно определяется по рекомендациям экспертов, в случае же отсутствия таковых рекомендуют брать значение в пределах 0,25–0,30. Значимость «свежих» данных возрастает по мере увеличения коэффициента, а значимость «устаревших» данных снижается.

Шаг 6. Суммирование линейной и периодической функций:

$$\sum_{i=1}^n \left(\frac{e^{\alpha i} (k_5 i + k_6 - Y_i)}{e^{\alpha n}} \right)^2 \rightarrow \min,$$

где α – коэффициент «инерции (устаревания) данных», k_5 – подбираемый коэффициент наклона, k_6 – подбираемый коэффициент базы, Y_i – элементы выборки.

Шаг 7. Итоговое уравнение прогноза без ограничений выглядит как

$$y_{n+f} = k_5(n + f) + k_6 + k_3\varphi(k_4(n + f)).$$

Шаг 8. Проводим проверку и замену значений прогноза, выходящих за границы.

Соответственно прогноз с ограничениями:

$$\begin{aligned} y_{n+f} = & H(k_5)H\left(Y_{+0} - \left(k_5(n + f) + k_6 + k_3\varphi(k_4(n + f))\right)\right) \left(k_5(n + f) + k_6 + \right. \\ & \left. + k_3\varphi(k_4(n + f))\right) + \\ & + H(-k_5)H\left(-Y_{-0} + \left(k_5(n + f) + k_6 + k_3\varphi(k_4(n + f))\right)\right) \left(k_5(n + f) + k_6 + \right. \\ & \left. + k_3\varphi(k_4(n + f))\right) + \\ & + H(k_5)H\left(-Y_{+0} + \left(k_5(n + f) + k_6 + k_3\varphi(k_4(n + f))\right)\right) Y_{+0} + \\ & + H(-k_5)H\left(Y_{-0} - \left(k_5(n + f) + k_6 + k_3\varphi(k_4(n + f))\right)\right) Y_{-0}. \end{aligned}$$

3.3. Прогноз на основе нейронной сети

Нейронная сеть функционирует через взаимосвязанные узлы, именуемые нейронами. Эти узлы организованы в многослойную структуру, которая напоминает устройство человеческого мозга.

Самая элементарная нейронная сеть, называемая перцептроном, представляет собой упрощенную имитацию деятельности биологической сети, состоящей из нейронов и связей, соединяющих их.

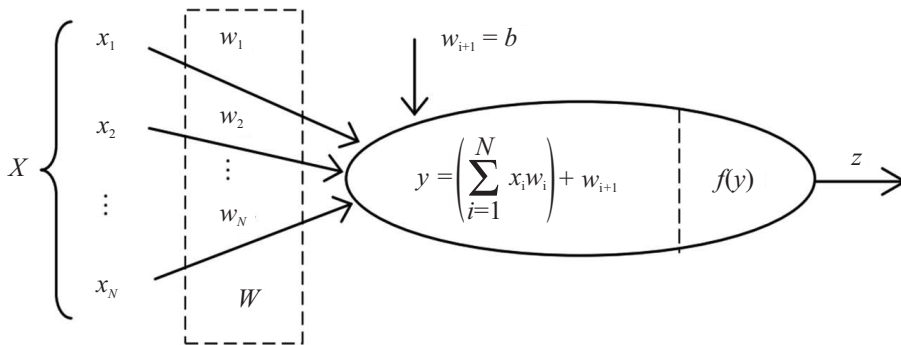


Рис. 1. Схематическое изображение искусственной нейронной сети.

Структура формального нейрона показана на рис. 1.

На вход поступает набор сигналов. Каждый сигнал подвергается процедуре взвешивания, т.е. умножается на заранее заданный коэффициент:

$$net_j = \sum_{i=1}^N x_i w_{ij},$$

где net_j является итоговым значением, получаемым в результате обработки всех поступающих на нейрон j сигналов (суммарный вход нейрона), N – суммарное количество элементов, чьи входные параметры оказывают воздействие на входной сигнал j ; w_{ij} – вес, определяющий силу связи, соединяющей нейрон i с нейроном j .

Суммировав все входные сигналы, предварительно помноженные на соответствующие веса, получаем общий входной сигнал нейрона.

Каждый нейрон функционирует в соответствии со своим внутренним алгоритмом, определяющим его входной сигнал. Этот алгоритм вычисляет результат на основе суммарного значения, полученного нейроном. Данное правило именуется функцией активации $f(x)$.

Обратимся к правилу обучения Уидроу–Хоффа, также называемому дельта-правилом. Суть его заключается в стремлении к минимизации среднеквадратической ошибки, возникшей в нейронной сети. Вычисление этой ошибки для входных данных производится по формуле

$$E = \frac{1}{2}(Y - d)^2,$$

где d – искомая выходная величина.

Каждый нейрон осуществляет вычисления взвешенной суммы, применяя следующую формулу:

$$S = w_1 X_1 + w_2 X_2 - b,$$

где b определяет критический уровень (порог).

Рассмотрим для примера линейную функцию активации, заданную как $Y = x$. В этом случае функционал ошибки будет определяться следующей формулой:

$$\begin{aligned}\frac{\partial E}{\partial w_1} &= (X - d)X_1, \\ \frac{\partial E}{\partial w_2} &= (X - d)X_2, \\ \frac{\partial E}{\partial b} &= -(X - d).\end{aligned}$$

Весовые коэффициенты и смещение нейрона определяются следующими формулами:

$$\begin{aligned}w_i(t + 1) &= w_i(t) - \alpha(Y - d)X_i, \\ b(t + 1) &= b(t) + \alpha(Y - d).\end{aligned}$$

В данном конкретном сценарии функционал ошибки определится как:

$$E = \frac{1}{2} \sum_j (Y_j - d_j)^2.$$

Коэффициенты, определяющие веса и смещения нейронов, рассчитываются по следующим формулам:

$$\begin{aligned}w_{ij}(t + 1) &= w_{ij}(t) - \alpha(Y_j - d_j)X_i, \\ b_j(t + 1) &= b_j(t) - \alpha(Y_j - d_j).\end{aligned}$$

Алгоритм обратного распространения ошибки служит основой для обучения многослойных нейронных сетей, построенных по принципу прямого распространения сигнала. Обучение по сути сводится к двум последовательным стадиям, охватывающим каждый слой сети: прямой и обратный проходы. Синаптические веса, обозначаемые как $w_{i,j}^k$ (i – номер веса, j – номер нейрона, k – номер слоя), подстраиваются для достижения наибольшего соответствия выдаваемого сетью сигнала желаемому. Значение выхода j -го нейрона, расположенного в k -ом слое, определяется следующим образом:

$$Y_j^k = F \left(\sum w_{i,j}^k Y_i^{k-1} - b_j^k \right).$$

Входной сигнал для j -го нейрона последнего слоя вычисляется следующим образом:

$$Y_j = F \left(\sum w_{i,j} Y_i^{n-1} - b_j \right).$$

Функционал ошибки нейронной сети определяется как

$$E = \frac{1}{2} \sum_j (\gamma_j)^2,$$

где $\gamma_j = Y_j - d_j$ – ошибка j -го нейрона, находящегося в выходном слое.

Ошибка j -го элемента, расположенного на k -ом скрытом слое:

$$\begin{aligned}\gamma_j^k &= \frac{\partial E}{\partial Y_j^k} = \sum_j \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial S_j} \frac{\partial S_j}{\partial Y_j^k} = \sum_j \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial S_j} w_{i,j} = \\ &= \sum_j (Y_j - d_j) F'(S_j) w_{i,j} = \sum_j \gamma_j F'(S_j) w_{i,j}.\end{aligned}$$

Градиенты ошибок:

$$\begin{aligned}\frac{\partial E}{\partial w_{i,j}} &= \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial S_j} \frac{\partial S_j}{\partial w_{i,j}} = \gamma_j F'(S_j) Y_j^k, \\ \frac{\partial E}{\partial b_j} &= \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial S_j} \frac{\partial S_j}{\partial b_j} = -\gamma_j F'(S_j), \\ \frac{\partial E}{\partial w_{i,j}^k} &= \sum_j \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial S_j} \frac{\partial S_j}{\partial Y_j^{k-1}} \frac{\partial Y_j^{k-1}}{\partial S_j^{k-1}} \frac{\partial S_j^{k-1}}{\partial w_{i,j}^k} = \gamma_j F'(S_j^k) Y_j^k.\end{aligned}$$

Весовые коэффициенты и смещения нейронов определяются следующими формулами:

$$\begin{aligned}w_{ij}^k(t+1) &= w_{ij}^k - \alpha \gamma_j^k F'(S_j^k) Y_j^k, \\ b_j^k(t+1) &= b_j^k - \alpha \gamma_j^k F'(S_j^k),\end{aligned}$$

где $\alpha (0 < \alpha < 1)$ – шаг обучения.

Алгоритм будет выполняться до тех пор, пока $E > E_m$, E_m – это желаемое значение среднеквадратичной ошибки нейронной сети.

Создадим простую, не рекуррентную, перцептронную нейросеть для авто-регрессии каждого ряда данных:

$$y_{n+f} = k_{n0} \operatorname{th} k_{n1} \operatorname{th} \left(\sum_{i=1}^{i=m} k_{ni} \operatorname{th} \left(\sum_{j=i}^{i=1} k_{nij} \operatorname{th}(Y_{ij}) \right) \right),$$

где y_{n+f} – прогноз на момент времени $n + f$, Y_{ij} – элементы выборки, $\operatorname{th}(x)$ – функция гиперболического тангенса, k_{nij} – вычисляемые при обучении коэффициенты входных весов нейрона, f – период упреждения прогноза, m – количество входных нейронов, n – количество входов каждого нейрона первого слоя.

Обучим нейронную сеть стандартным стохастическим методом Хопфилда или методом градиентного спуска (ОРО), минимизируя сумму квадратов отклонений.

3.4. Ансамблевый прогноз

На рис. 2 построены графики для всех трех вышеперечисленных типов прогнозов (линейный прогноз, OTS-прогноз, нейронный прогноз).



Рис. 2. Линейный прогноз, OTS-прогноз, нейронный прогноз.

Как видим, прогнозы значительно отличаются:

1. *Линейный* прогноз показывает значительное снижение.
2. *OTS-прогноз* показывает снижение, рост, а затем снова снижение.
3. *Нейронный* прогноз показывает рост.

Столь значительные расхождения в прогнозах появляются обычно при прогнозировании малых и сверхмалых временных рядов (до 30 элементов), поскольку отклонения их статистик очень велики и, например, на выборке в 20 элементов могут составлять до 2σ .

В таких случаях из набора разрозненных прогнозов создают прогностический ансамбль (совокупный прогноз), основная задача которого – дать консенсуальный прогноз, зависящий от степени доверия к той или иной методике:

$$y_{n+f} = \sum_{i=1}^m d_i y_{n+f_i}; \quad \sum_{i=1}^m d_i = 1,$$

где y_{n+f} – прогноз на момент времени $n + f$, y_{n+f_i} – прогноз по i -ой методике, d_i – коэффициент доверия к i -ой методике, m – количество методик прогнозирования.

Простейшим методом ансамблирования является метод Байеса, предполагающий равную степень доверия ко всем методикам прогнозирования [15–17]:

$$y_{n+f} = \sum_{i=1}^m \frac{y_{n+f_i}}{m}.$$



Рис. 3. Прогноз энергобаланса РФ.



Рис. 4. Прогноз. Добыча и генерация. Импорт энергоресурсов.

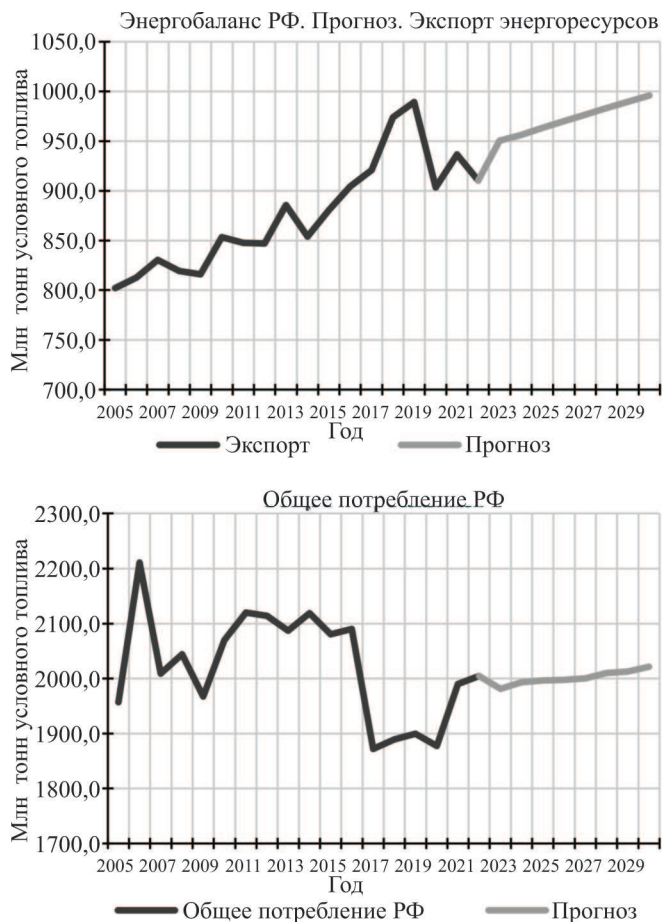


Рис. 5. Прогноз. Экспорт энергоресурсов. Общее потребление РФ.

Используя этот метод, рассчитаем ансамблевые прогнозы и построим графики ансамблевых прогнозов как отдельных показателей, так и энергобаланса в целом.

Рассмотрим общий ансамблевый прогноз энергобаланса (рис. 3).

Как видим, при сохранении текущих тенденций в период 2023–2030 гг. ожидается возникновение энергодефицита в размере от 20 до 120 млн. тонн условного топлива, что при общем объеме текущей добычи составляет от 0,7% до 4,2%. Такой дефицит может быть легко покрыт плановым увеличением добычи на 0,6–0,7% в год.

Теперь выясним причины возникновения возможного дефицита. На рис. 4 видно, что общая добыча останется неизменной, а импорт энергоресурсов вырастет незначительно.

На рис. 5 видно, что экспорт и внутреннее потребление значительно возрастут.

Таким образом, даже на основании интуитивного анализа графиков можно сделать вывод, что рост внутреннего потребления и экспорта энергоресурсов и будут являться основными причинами возможного энергодефицита.

4. Результат. Оценка качества прогноза

Оценка возможного отклонения прогнозируемого значения – это предполагаемые границы колебаний прогнозируемой величины, определяемые как доверительный интервал прогноза. Обычно их определяют как «пессимистическую» и «оптимистическую» границы прогноза. Величина доверительного интервала основывается на предположении, что прогнозируемая величина имеет гауссовское нормальное распределение, а также на том, какая степень неопределенности допускается в прогнозе.

Значение доверительного интервала для каждого прогнозируемого периода вычисляется по формуле

$$\Delta X_n = nZ_\alpha \text{RMSE},$$

где n – прогностический интервал, Z_α – множитель уровня значимости, RMSE – стандартное отклонение тенденции.

Таким образом, каждое последующее значение прогнозируемой величины расширяет свои вероятностные границы как минимум на одно стандартное отклонение, а максимальный уровень значимости прогноза соответствует отклонению границ в 68%.

Оценим доверительный интервал компонентов линейного, экспоненциального, нейронного и ансамблевого прогнозов энергетического баланса, к ним относятся экспорт, импорт, производство и потребление энергоресурсов.

Для этой оценки сначала необходимо найти стандартную выборочную ошибку аппроксимации для линейного, экспоненциального и нейронного трендов этих компонентов. Во всех трех случаях оценить ее можно как

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_i - x_i)^2}{n - 1}},$$

где X_i – фактические значения прогнозируемой величины, x_i – значения прогностического тренда, n – объем выборки фактических значений.

В таком случае, границы доверительного интервала для прогнозируемых значений определяются, как:

$$D_{i>n} = x_i \pm Z_\alpha (i - n) \text{RMSE},$$

где x_i – значения прогностического тренда, n – объем выборки фактических значений, Z_α – множитель уровня значимости, RMSE – стандартное отклонение тенденции.

Поскольку фактический энергобаланс является величиной, сведенной в ноль, оценить его отклонение от любых прогностических трендов невозможно. В этом случае можно воспользоваться вычисляемыми оценками ошибки

Энергобаланс РФ. Прогноз 2023–2030 гг. Ансамбль Байеса, $\alpha = 0,25$

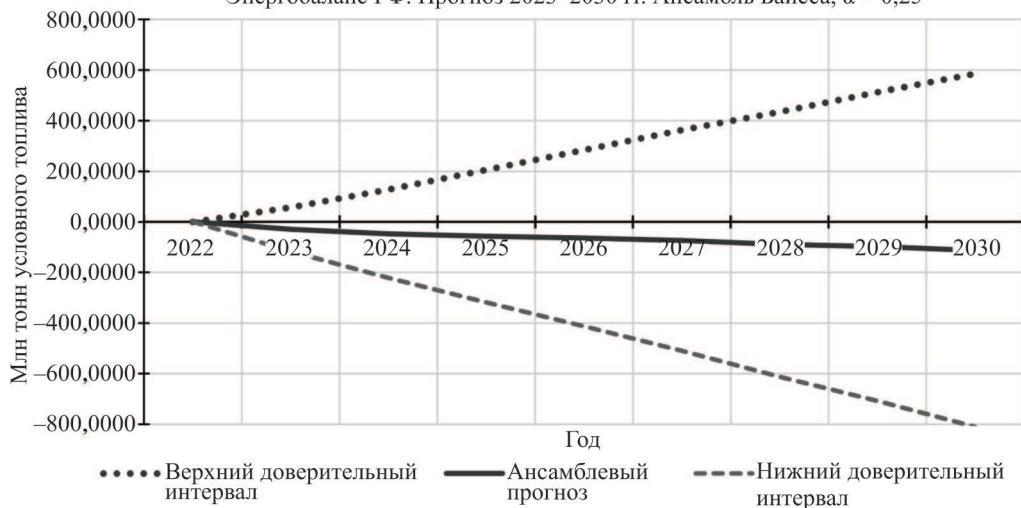


Рис. 6. Ансамблевый прогноз энергобаланса РФ с доверительным интервалом.

компонентов энергобаланса, при этом независимо от того, отрицательным или положительным влиянием на энергобаланс обладает компонент, привносимая им доля ошибки должна учитываться как увеличивающая совокупную ошибку.

Для каждого из типов прогнозов энергобаланса получим значения совокупной RMSE, используя ансамблирование по Байесу, оценим ошибку ансамблевого прогноза как среднюю арифметическую величину (см. табл. 2).

Таблица 2. Сравнение прогнозов

Прогноз энергобаланса	Абсолютная совокупная RMSE	Относительная совокупная RMSE
Линейный	74,5223	5,14%
ОТС (экспоненциальный)	75,4225	5,22%
Нейронный	77,5736	5,67%
Ансамблевый	75,8395	5,23%

Построим график результирующего ансамблевого прогноза с доверительными интервалами (см. рис. 6).

Следует заметить, что доверительный интервал отражает внешние границы, за которые при указанном уровне значимости α прогнозируемая величина не выйдет [18–20]. Однако изображение доверительного интервала, особенно на длительных прогностических периодах, вызывает ощущение низкой точности прогноза и значительного разброса возможных значений. В этом случае на графиках рекомендуется использовать прогностический коридор, являющийся визуальным отражением стандартного отклонения или результатов частотного анализа. В этом случае отображение границ прогноза будет

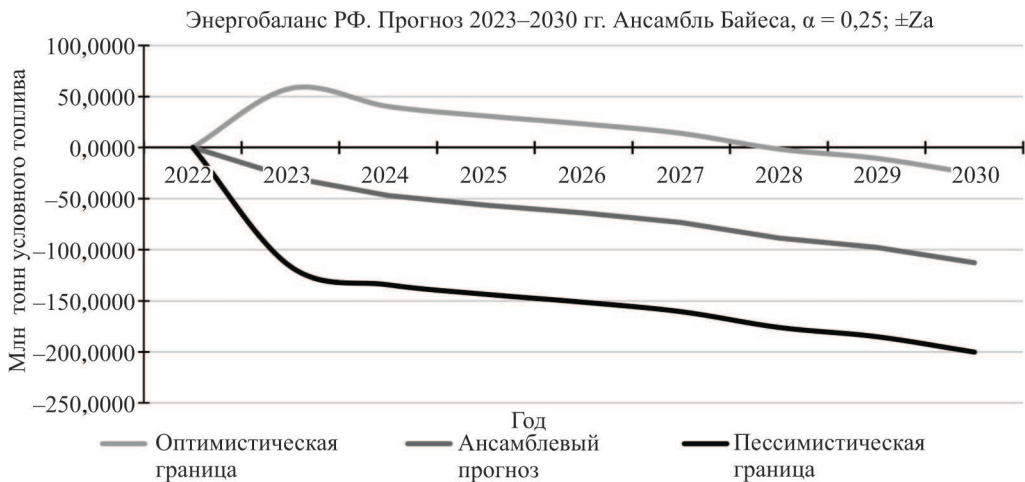


Рис. 7. Ансамблевый прогноз энергобаланса РФ с прогностическим коридором.



Рис. 8. Трехмерный график отраслевых прогнозов.

более плавным, узким и нерасширяющимся. Рассмотрим границы прогностического коридора для отображения ансамблевого прогноза энергобаланса.

В таком случае границы прогностического коридора для прогнозируемых значений определяются как:

$$F_{i>n} = x_i \pm Z_{\alpha}RMSE,$$

где x_i – значения прогностического тренда, n – объем выборки фактических значений, Z_α – множитель уровня значимости, RMSE – стандартное отклонение тенденции.

Так, для $\alpha = 0,25$ предполагаемое отклонение составит $1,15\sigma$, что отражено на графике рис. 7.

Как видно из рисунка, такое представление прогноза психологически воспринимается гораздо благоприятнее.

Как завершающий итог построен трехмерный график отраслевых прогнозов (рис. 8) с целью выделить отрасли промышленности, энергоснабжение которых требует особого внимания.

Из графика (зеленая зона околонулевых значений справа) сразу становится понятно, развитию каких отраслей в РФ уделяется недостаточно внимания. С целью лучшей диверсификации экономики эти отрасли необходимо стимулировать.

5. Заключение

На основе анализа временных рядов показателей энергобаланса РФ были реализованы и исследованы как классические статистические модели, так и методы глубокого обучения, а также их комбинация в виде ансамбля.

Полученные результаты позволили выявить преимущества и ограничения каждой из моделей прогнозирования. Показано, что использование ансамблевой модели, объединяющей предсказания различных методов, обеспечивает более высокую устойчивость и точность результатов по сравнению с отдельными моделями.

Экономическая ценность разработанной прогнозной системы подтверждается возможностью ее применения для стратегического и оперативного планирования, управления рисками и формирования конкурентных преимуществ на рынке энергоресурсов. Практическая реализация моделей и построенного ансамбля на реальных данных показала их применимость для задач, стоящих перед энергетическими компаниями, трейдерами и государственными регуляторами.

СПИСОК ЛИТЕРАТУРЫ

1. ЕМИСС – Государственная статистика [Электронный ресурс]: Режим доступа: <https://www.fedstat.ru>
2. Росстат – Федеральная служба государственной статистики [Электронный ресурс]: Режим доступа: <https://rosstat.gov.ru/>
3. Министерство энергетики [Электронный ресурс]: Режим доступа: <https://minenergo.gov.ru>
4. *Васильев М.В., Дранко О.И.* Двухуровневая модель прогнозирования доходов крупномасштабной энергетической системы // Датчики и системы. 2023. № 2 (267). С. 71–78.
5. *Дранко О.И.* О прогнозировании финансирования конверсии предприятий // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2020. Т. 20. № 4. С. 74–82.

6. Дранко О.И., Благодарный Е.В. Моделирование разрушения стоимости российских энергетических компаний // Информационные и математические технологии в науке и управлении. 2022. № 3 (27). С. 104–112.
7. Дранко О.И., Резчикова А.Ф., Степановская И.А. и др. Сценарное моделирование развития страны на основе индикативного планирования // Проблемы управления. 2024. № 5. С. 25–41.
8. Дранко О.И., Тароян К.К. Прогнозирование выручки быстрорастущей компании с использованием логистической кривой // Автоматизация и моделирование в проектировании и управлении. 2024. № 2 (24). С. 84–92.
9. Дранко О.И., Тароян К.К. О модели прогнозирования выручки организации с быстрым ростом // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2023. Т. 23. № 4. С. 66–75.
10. Иванюк В.А. Методика долгосрочного прогнозирования на основе мультитрендового прогноза // Мягкие измерения и вычисления. 2023. Т. 73. № 12–1. С. 128–138.
11. Brockwell P.J., Davis R.A. Introduction to Time Series and Forecasting. 3rd ed. New York: Springer, 2016.
12. Brown R.G. Smoothing forecasting and prediction of discrete time series. New York: Prentice Hall, 1963.
13. Chen W., Xu H., Liu Z., et al. Hybrid modeling of energy price forecasting combining temporal features and economic indicators // Energy. 2022. V. 243. P. 123–135.
14. Hamid A., Islam M.S., Hasan M.R. A comprehensive review of deep recurrent neural networks for energy price forecasting // Renewable and Sustainable Energy Reviews. 2020. V. 135. Article 110354.
15. Ivanyuk V. The method of residual-based bootstrap averaging of the forecast ensemble // Financial Innovation. 2023. V. 9. No. 1. P. 37.
16. Ivanyuk V. Forecasting of digital financial crimes in Russia based on machine learning methods // Journal of Computer Virology and Hacking Techniques. 2024. V. 20. No. 3. P. 349–362.
17. Li K., Liu X., Zhu J., et al. Electricity price forecasting using gradient boosting and random forests // Applied Energy. 2019. V. 253. P. 113–122.
18. Smith J., Reeve D. Forecasting oil and natural gas prices using ARIMA and exponential smoothing models // Energy Economics. 2021. V. 93. P. 105–120.
19. Wang L., Zhou Y., Yang G. Ensemble forecasting model combining ARIMA, XGBoost, and LSTM for short-term natural gas price prediction // Energy Reports. 2022. V. 8. P. 1430–1440.
20. Zhang G.P. Time series forecasting using hybrid ARIMA and neural network models // Neurocomputing. 2003. V. 50. P. 159–175.

Статья представлена к публикации членом редколлегии А.А. Галеевым.

Поступила в редакцию 08.07.2025

После доработки 14.10.2025

Принята к публикации 20.11.2025