

Секция «Экономика природопользования, энергетики и биотехнологий»

## Структура жилого фонда в России: дооценка с помощью применения методов машинного обучения

Научный руководитель – Колпаков Андрей Юрьевич

*Тюрин Артем Андреевич*

*Аспирант*

Институт народнохозяйственного прогнозирования Российской Академии Наук, Москва,  
Россия

*E-mail: tyurin.aa@phystech.edu*

В условиях современных тенденций, направленных на снижение выбросов парниковых газов в целях борьбы с изменением климата, задача оптимизации использования ресурсов в строительстве и, что особенно важно, в эксплуатации зданий приобретает особую актуальность.

Энергоэффективность зданий играет ключевую роль в снижении затрат на их эксплуатацию и минимизации воздействия на окружающую среду. В этой связи точные прогнозы потребления энергии и грамотная оценка потенциала для модернизации жилого фонда становятся важными инструментами стратегического планирования и оптимизации.

Однако при попытке построить такого рода прогнозы исследователи сталкиваются с проблемой малого количества данных по жилому фонду РФ. В этом смысле прогнозные оценки будут иметь довольно низкую точность, что сильно снижает вероятность применимости таких оценок. Для улучшения сценарного прогнозирования спроса на тепло требуется наличие большего количества данных о жилых домах. Сам по себе этот процесс довольно трудоемкий и, вероятно, потребует больших временных затрат.

В целом, в экономике сравнительно мало сюжетов, в которых можно с относительно большой точностью применить методы машинного обучения. Тем не менее, этот сюжет является одним из них, ввиду возможности классификации (например, по сериям домов или году постройки) неучтенного жилого фонда на основании данных, которые есть в наличии. Благодаря этому дооценка структуры жилого фонда РФ с помощью методов машинного обучения выглядит вполне обоснованной.

Таким образом, в данной работе представлена модель машинного обучения, предназначенная для прогнозирования структуры жилого фонда РФ в разрезе классов энергоэффективности зданий. Система позволяет осуществлять сценарное прогнозирование, а также предоставляет практическую основу для разработки решений, направленных на повышение устойчивости и оптимизацию потребления энергии в жилом фонде.

В рамках исследования использовались открытые данные с сайта Фонда развития территорий (Открытые данные, 2024). В данном случае нас интересовали реестры домов по регионам РФ.

Набор данных всего содержит 60 переменных (столбцов) и содержит информацию о 1,03 млн жилых домах. Как упоминалось ранее, на практике значительная часть данных, которые представляются ценными для формирования дальнейших выводов, отсутствуют: так, класс энергоэффективности присвоен только 13% домов, в то время как данные об общей площади домов – присутствуют в 83% строк. В среднем, отсутствует около 30% данных по каждой из переменных.

Алгоритм, описанный ниже, применяется в качестве предиктивной модели, результатом работы которого является класс энергоэффективности. В качестве входных данных используется та информация по дому, которая есть в базе данных. Обучающей же выборкой является тот набор данных по домам, где был присвоен тот или иной класс энергоэффективности.

В рамках исследования были использован метод случайного леса. Для обучения моделей машинного обучения использовался язык программирования Python и библиотека scikit-learn. Предварительно столбцы с категориальными данными были закодированы с использованием «one-hot» схемы кодирования. Данное кодирование создает двоичный столбец для каждой уникальной категории в столбце и возвращает разреженную бинарную матрицу. В таком виде данные уже подаются на вход классификатора для обучения.

Количество деревьев в ансамбле было установлено в параметре классификатора равное 100. В рамках проверки работы модели на тестовой выборке её точность достигла уже 60,2% при длительности обучения модели в 3,5 часа. Результаты моделирования представлены на рисунке 1.

В целом, результаты модели соответствуют изначальным представлениям о предполагаемом распределении домов. Различия между моделью и данными, по которым известен класс энергоэффективности, вероятно, объясняются тем, что дома с присвоенным классом, как правило, находятся в европейской части РФ, а также в регионах с крупными городами. В этих субъектах наблюдается более благоприятная ситуация с состоянием и энергоэффективностью жилого фонда. Однако оценка площадей домов с тем или иным классом энергоэффективности уже не согласуется с изначальными представлениями: вероятно, из-за большого количества пропусков (около 17% домов не содержат информации об общей площади).

Для дооценки выпадающих площадей было принято решение экстраполировать полученное распределение домов по классам энергоэффективности на оставшийся жилой фонд. Из данных Росстата за 2023 год (Росстат, 2024) был получен общий объем жилого фонда РФ на конец 2023 года – он составил 4215,1 млн. кв. м. Таким образом, экстраполировалось около 1458,9 млн. кв. м. неучтенной площади. Результаты распределения площадей по классам энергоэффективности представлены на рисунке 2.

Полученные результаты подтверждаются Стратегией социально-экономического развития Российской Федерации с низким уровнем выбросов парниковых газов до 2050 года (Правительство Российской Федерации, 2021): так, согласно данной стратегии, высокие классы энергетической эффективности (А, В и С) на данный момент имеют не более 30 процентов ежегодно вводимых в эксплуатацию многоквартирных домов, что совпадает с полученными результатами моделирования.

В качестве дальнейших действий, которые могли бы быть направлены на повышение точности модели, можно выделить следующие:

- производство масштабирования признаков, присвоение весов столбцам на основе их влияния на целевой показатель;
- агрегация уникальных значений столбцов там, где их количество достаточно весомое;
- повышение качества входных данных, уменьшение числа пропусков значений в датасете.

Однако даже при такой точности и качестве входных данных результаты совпадают с теми, которые отражены в правительственных документах. Это позволяет использовать данную модель уже для расчета потребления тепла и сценарного прогнозирования структуры жилого фонда в РФ.

## Источники и литература

- 1) Открытые данные. (16 августа 2024 г.). Получено из Фонд развития территорий: <http://витрина.фрт.рф/opendata>

- 2) Правительство Российской Федерации. (2021). Стратегию социально-экономического развития Российской Федерации с низким уровнем выбросов парниковых газов до 2050 года. Москва. Получено из <http://static.government.ru/media/files/ADKkCzp3fWO32e2yA0BhtIpyzWfHaiUa.pdf>
- 3) Росстат. (16 августа 2024 г.). Получено из <https://rosstat.gov.ru/search?q=%D0%B6%D0%B8%D0%BB%D0%BE%D0%B9+%D1%84%D0%BE%D0%BD%D0%B4>

### Иллюстрации

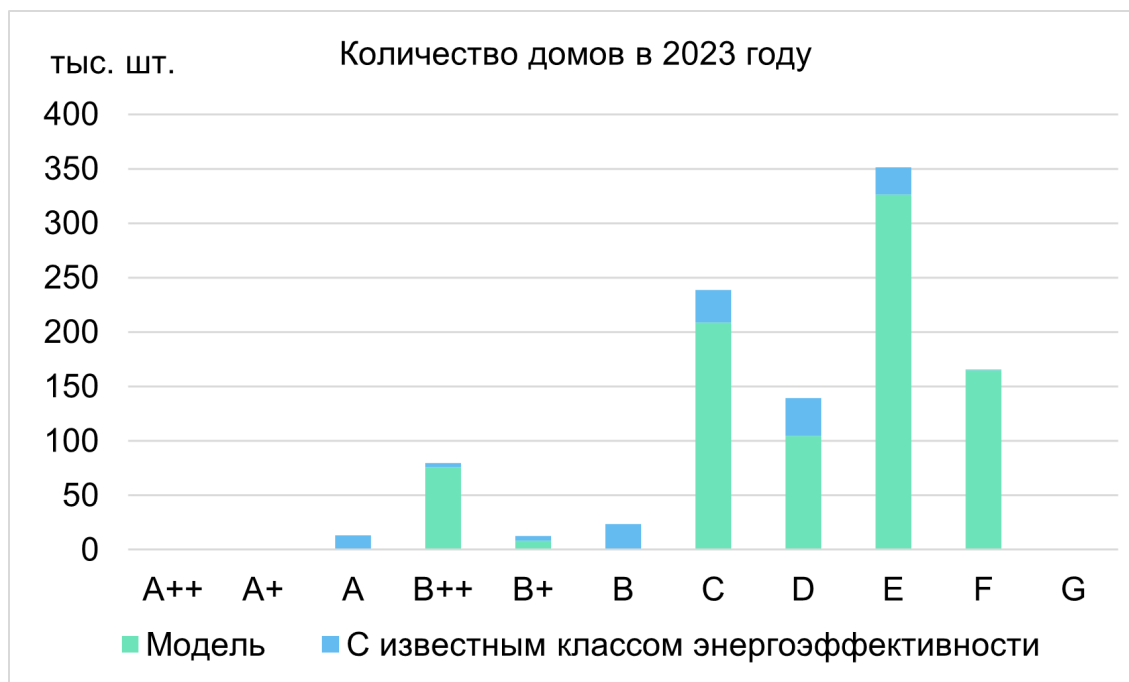


Рис. : Общее распределение домов по классам энергоэффективности в 2023 году. Источник: составлено автором



Рис. : Общее распределение площадей жилого фонда РФ по классам энергоэффективности в 2023 году, с проведенной экстраполяцией. Источник: рассчитано автором с помощью данных Росстата (Росстат, 2024)