

УДК 004.8

Моделирование и прогнозирование потребления тепловой энергии с использованием погодных данных

А. Б. Бирюков, С. В. Гридин
Донецкий национальный технический университет, г. Донецк

e-mail: birukov.donntu@mail.ru

Аннотация

Работа посвящена моделированию и прогнозированию потребления тепловой энергии объектом сферы здравоохранения в зависимости от данных о погодных условиях на основе алгоритмов машинного обучения. В качестве алгоритмов были использованы Random Forest и Support Vector Machines, так как именно эти алгоритмы более всего подходят для решения задач регрессии. Точность разработанных моделей проверялась и сопоставлялась на тестовых данных. Модели, разработанные на основе указанных алгоритмов, показывают достаточно высокую эффективность в условиях ограниченного количества данных для обучения и имеют значительный потенциал для использования.

Введение

Как для отдельно взятого потребителя, так и для системы «генерация-транспортировка-потребление» тепловой энергии в целом, основным фактором, определяющим уровень потребления, являются погодные условия. Техническое состояние потребителя и систем теплоснабжения, эффективность основного и вспомогательного оборудования объектов генерации, степень автоматизации процессов регулирования и распределения теплоносителя влияют на уровень тепловых потерь и класс энергетической эффективности, саму же тепловую нагрузку определяет состояние окружающей среды.

На потребление объектом тепловой энергии на протяжении отопительного сезона влияет не только непосредственно температура наружного воздуха. Достаточно существенными факторами могут быть сила и направление ветра, а также интенсивность солнечного излучения. Сила ветра оказывает влияние на теплопотери с инфильтрацией. В свою очередь, чем интенсивнее солнечное излучение, тем выше теплопоступления. При этом степень влияния ветра и солнечного излучения в значительной степени зависят от конструктивных особенностей здания-потребителя: расположение здания в пространстве, высота, площадь и ориентированность светопрозрачных ограждающих конструкций, уровень герметичности наружных стен и т.п. [1]. В связи с этим, влияние этих факторов необходимо оценивать индивидуально для каждого объекта.

На потребление тепловой энергии может оказать влияние не только среднесуточная температура воздуха, но и температурный перепад, т.к. при резком повышении или снижении температуры воздуха могут возникать

кратковременные перепады тепловой нагрузки, величина которых зависит от инерционности и уровня автоматизации системы генерации.

Имея в своем распоряжении погодные данные за определенный период, а также данные по потреблению тепловой энергии (к примеру, показания теплового счетчика) или непосредственно потребления первичного энергоресурса (природного газа, электроэнергии), можно построить модель потребления энергии на нужды отопления для конкретного объекта. Такая модель позволяет определять и прогнозировать потребление тепловой энергии объектом в зависимости от погодных факторов (температура и температурный перепад, сила ветра, солнечное излучение, влажность, осадки) с учетом их значимости [1].

Постановка задачи моделирования потребления тепловой энергии в зависимости от погодных характеристик

Моделирование и прогнозирование потребления тепловой энергии позволяют решить множество задач на уровне потребителя тепловой энергии, поставщика тепловой энергии и системы энергетического менеджмента, таких как:

- планирование энергетического бюджета. Имея в своем распоряжении достаточно точные погодные данные, можно спрогнозировать затраты на тепловую энергию для определенного периода;

- выявление снижения энергоэффективности объекта. Сопоставляя данные по потреблению тепловой энергии, полученные с помощью модели, с фактическими данными по потреблению (данные узла учета

тепловой энергии), можно выявить отклонения, которые могут свидетельствовать о нарушениях в работе системы отопления объекта. Выявив эти нарушения, потребитель или обслуживающая организация могут их своевременно устранить или запланировать мероприятия по повышению энергоэффективности объекта;

- внедрение моделей потребления в локальные системы погодозависимого регулирования тепловой нагрузки для поддержания параметров теплового комфорта на необходимом уровне независимо от колебаний погодных условий;

- использование моделей потребления для снижения инерционности объектов генерации тепловой энергии за счет прогнозирования резких изменений погодных условий и, соответственно, тепловой нагрузки;

- использование моделей потребления для эффективного распределения энергетических потоков. При совместном использовании традиционных и альтернативных источников тепловой энергии существует необходимость поддержания наиболее эффективного баланса в системе потребления: при соблюдении графика тепловой нагрузки необходимо поддерживать максимально возможную долю энергии от альтернативных источников. Точное прогнозирование потребления тепловой энергии позволяет эффективно регулировать поступление энергии от различных источников.

С учетом вышесказанного, целью статьи является построение модели прогнозирования потребления тепловой энергии с использованием методов машинного обучения (Machine Learning, ML) на примере конкретного объекта, с использованием фактических погодных данных.

Объект исследования

Объектом исследования является комплекс блоков учреждения здравоохранения, расположенный в г. Донецк Донецкой Народной Республики. Здания комплекса 3-х и 4-х этажные, стены выполнены из силикатного кирпича толщиной не менее 0,5 м. Дополнительная теплоизоляция стен отсутствует. В зданиях присутствует техническое подполье и подвальные помещения с расположенными коммуникациями системы теплоснабжения, водоснабжения и канализации, а также неотапливаемые чердачные помещения. Окна в здании частично пластиковые, частично старые деревянные. Пластиковые окна низкоэффективные, однокамерные, без низкоэмиссионного покрытия. Часть старых деревянных окон находится в неудовлетворительном состоянии, что обуславливает увеличение тепловых потерь с инфильтрацией.

Трубопроводы системы отопления здания с нижней разводкой, теплоизоляция трубопроводов в неотапливаемых помещениях либо отсутствует, либо находится в неудовлетворительном состоянии. Система отопления несбалансированная. Радиаторы чугунные и стальные, термостаты и оборудование контроля температурного режима в помещениях отсутствует. Узлы учета тепловой энергии присутствуют, учет не автоматизированный, показания фиксируются вручную.

Теплоснабжение комплекса объекта здравоохранения осуществляется за счет собственной котельной, находящейся непосредственно на территории комплекса. В котельной установлены 3 одноконтурных водогрейных газовых котла с КПД не ниже 0,9. На котлах установлены автоматизированные панели управления, осуществляющие регулировку работы горелки по заданной температуре теплоносителя. Вмешательство персонала котельной в работу котла минимально. Котельная обслуживает исключительно рассматриваемый объект сферы здравоохранения. Фиксация расхода природного газа осуществляется за счет установленного в котельной счетчика природного газа и корректора расхода газа. Съём показаний корректора расхода газа осуществляется как в ручном, так и в автоматизированном режиме.

Теплотрассы системы теплоснабжения подземной укладки в каналах, теплоизоляция присутствует. Все теплотрассы располагаются непосредственно на территории рассматриваемого объекта.

Данные о потреблении объектом энергетических ресурсов

На данный момент сбор данных о потреблении тем или иным объектом энергетических ресурсов может быть затруднителен. Далеко не все объекты теплоснабжения, особенно в жилищно-коммунальном секторе, оборудованы узлами учета тепловой энергии. Сбор данных не автоматизирован, оборудование для съема показаний на электронные носители отсутствует. Ручной сбор данных может носить несистемный характер и иметь значительную погрешность. В то же время, для проведения исследований, связанных с анализом потребления энергоресурсов, нужны надежные и стабильные данные для рассматриваемого временного интервала.

Для исследуемого объекта имеются частичные данные суточного потребления тепловой энергии за период с 01.01.2018 г. по 02.01.2019 г., а также с 11.12.2019 г. по 22.04.2020 г (рис. 1, рис. 2).

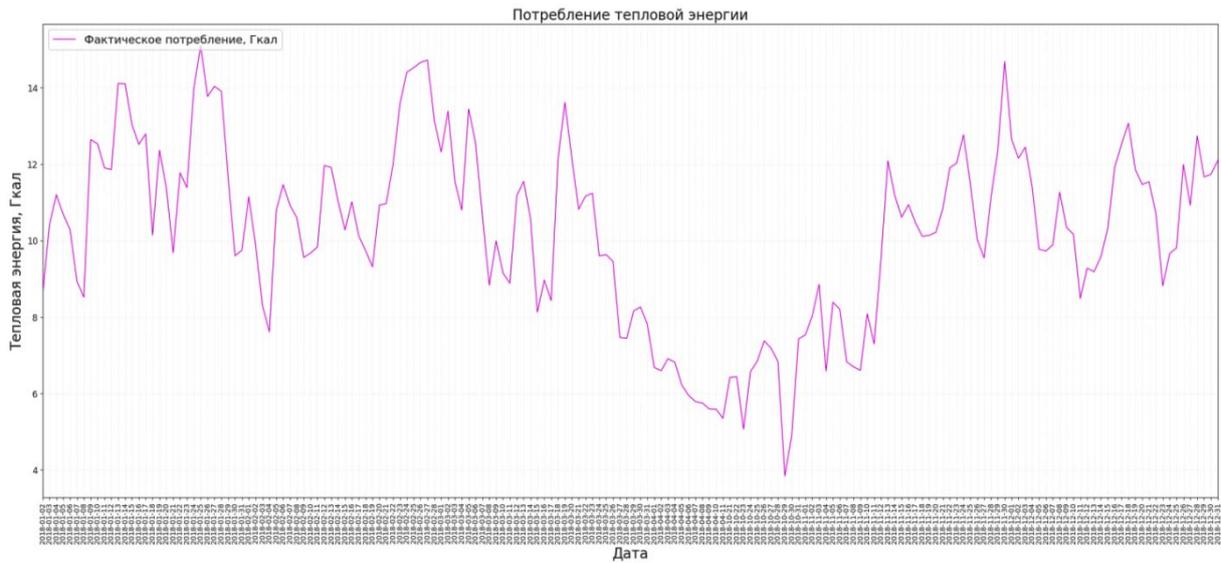


Рисунок 1 – Фактическое потребление тепловой энергии в 2018 г.

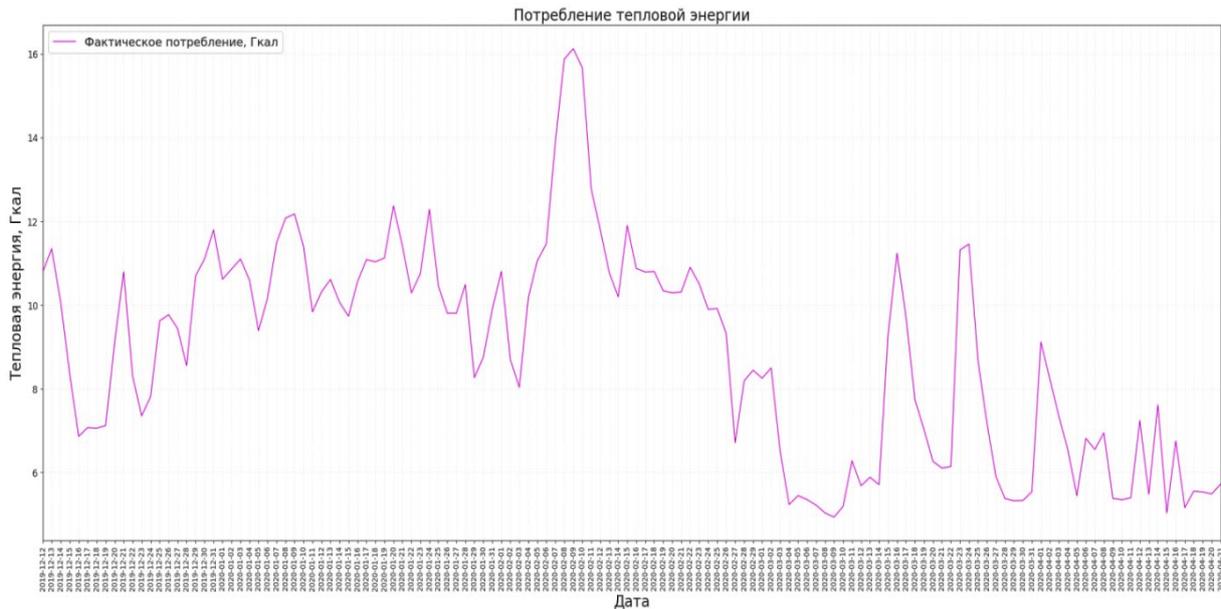


Рисунок 2 – Фактическое потребление тепловой энергии в 2019-2020 г.г.

Несмотря на то, что данные о потреблении тепловой энергии ограничены, тем не менее, для построения достаточно эффективной модели прогнозирования с использованием современных методов машинного обучения их может быть достаточно, как будет показано далее. При дальнейшем сборе данных и использовании их для обучения разрабатываемой модели эффективность и точность модели будет только возрастать.

Из используемых в исследовании данных потребления тепловой энергии были выделены

для тестирования модели данные за февраль – март 2020 г. Остальные данные использовались для обучения регрессионной модели.

Погодные данные для анализа

Для исследования были использованы фактические погодные данные для г. Донецк за исследуемый период. Фрагмент полного набора данных отображен на рис. 3. Набор данных, используемый для обучения модели, отображен на рис. 4.

name	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	...	solarenergy	uvindex	severerisk	sunrise	sunset	moonphase	conditions	description	icon	
0	Donetsk	2018-01-02	4.8	0.5	2.6	1.5	-4.6	-0.9	1.8	94.5	...	3.4	2	NaN	02.01.2018 7:18	02.01.2018 15:46	0.50	Partially cloudy	Partly cloudy throughout the day.	partly-cloudy-day
1	Donetsk	2018-01-03	2.6	-2.7	-0.4	-2.8	-8.7	-6.3	-2.1	89.1	...	3.9	3	NaN	03.01.2018 7:18	03.01.2018 15:47	0.54	Partially cloudy	Partly cloudy throughout the day.	partly-cloudy-day
2	Donetsk	2018-01-04	1.3	-3.1	-0.9	-3.9	-9.3	-6.5	-2.4	89.7	...	2.6	2	NaN	04.01.2018 7:18	04.01.2018 15:49	0.58	Snow, Rain, Overcast	Cloudy skies throughout the day with rain or s...	rain
3	Donetsk	2018-01-05	2.7	-1.5	0.0	-0.6	-6.5	-4.3	-1.9	87.3	...	3.9	2	NaN	05.01.2018 7:18	05.01.2018 15:50	0.61	Snow, Rain, Partially cloudy	Partly cloudy throughout the day with early mo...	rain
4	Donetsk	2018-01-06	3.9	-0.9	1.0	3.4	-3.3	-0.9	-0.2	91.4	...	3.1	2	NaN	06.01.2018 7:18	06.01.2018 15:51	0.64	Partially cloudy	Partly cloudy throughout the day.	partly-cloudy-day
...
300	Donetsk	2020-04-17	18.4	6.8	11.7	18.4	2.0	9.9	4.5	64.3	...	18.5	7	NaN	17.04.2020 5:35	17.04.2020 19:22	0.81	Partially cloudy	Partly cloudy throughout the day.	partly-cloudy-day
301	Donetsk	2020-04-18	17.2	5.6	10.8	17.2	3.2	10.1	0.6	53.2	...	23.0	8	NaN	18.04.2020 5:33	18.04.2020 19:23	0.84	Partially cloudy	Partly cloudy throughout the day.	partly-cloudy-day
302	Donetsk	2020-04-19	15.2	3.8	9.8	15.2	2.0	9.0	0.2	55.5	...	15.6	5	NaN	19.04.2020 5:31	19.04.2020 19:25	0.88	Rain, Partially cloudy	Partly cloudy throughout the day with late aft...	rain
303	Donetsk	2020-04-20	13.0	3.1	7.6	13.0	-0.8	6.0	1.9	71.0	...	15.5	7	NaN	20.04.2020 5:29	20.04.2020 19:26	0.91	Rain, Partially cloudy	Clearing in the afternoon with a chance of rai...	rain
304	Donetsk	2020-04-21	10.1	0.1	5.6	10.1	-3.9	2.4	-0.5	68.0	...	17.4	6	NaN	21.04.2020 5:27	21.04.2020 19:28	0.94	Partially cloudy	Partly cloudy throughout the day.	partly-cloudy-day

305 rows × 33 columns

Рисунок 3 – Погодные данные для г. Донецка

	date	temp	tempmin	tempmax	humidity	precip	windspeed	cloudcover	solarenergy
0	2018-01-02	2.6	0.5	4.8	94.5	0.0	19.4	75.1	3.4
1	2018-01-03	-0.4	-2.7	2.6	89.1	0.0	29.2	69.6	3.9
2	2018-01-04	-0.9	-3.1	1.3	89.7	3.3	24.8	98.1	2.6
3	2018-01-05	0.0	-1.5	2.7	87.3	0.2	18.7	83.0	3.9
4	2018-01-06	1.0	-0.9	3.9	91.4	0.0	16.2	84.6	3.1
...
300	2020-04-17	11.7	6.8	18.4	64.3	0.0	38.5	81.5	18.5
301	2020-04-18	10.8	5.6	17.2	53.2	0.0	23.4	50.3	23.0
302	2020-04-19	9.8	3.8	15.2	55.5	0.8	14.4	87.8	15.6
303	2020-04-20	7.6	3.1	13.0	71.0	13.2	24.8	57.5	15.5
304	2020-04-21	5.6	0.1	10.1	68.0	0.0	22.3	25.8	17.4

305 rows × 9 columns

Рисунок 4 – Погодные данные, используемые для обучения модели

При разработке модели были использованы следующие данные (рис. 4): temp – среднесуточная температура воздуха; tempmin – минимальная температура за рассматриваемые сутки; tempmax – максимальная температура для рассматриваемых суток; humidity – относительная влажность воздуха; precip – количество выпавших осадков; windspeed – средняя скорость ветра; cloudcover – облачность; solarenergy – солнечная энергия. Основопологающим фактором, влияющим на

тепловую нагрузку, является среднесуточная температура, однако суточный температурный перепад (разница между значениями максимальной и минимальной температурами воздуха) также оказывает влияние на потребление тепловой энергии, т.к. с ним связано изменение режимов работы объекта тепловой генерации (котельной установки). Теплопоступления от солнечного излучения являются одним из факторов, определяющих расчетную тепловую нагрузку объекта, как и

сила и повторяемость ветра, влияющие на теплотери с инфильтрацией. Относительная влажность и осадки являются, скорее, косвенными факторами, однако также были использованы для построения модели. Следует отметить, что указанные параметры могут влиять на ощущение теплового комфорта и для помещений с возможностью контроля параметров микроклимата могут способствовать увеличению тепловой нагрузки.

Используемые инструменты данные

При проведении исследования использовался язык программирования Python совместно с рядом библиотек: *sklear*, *matplotlib*, *Pandas*, *sqlite3*. Выбор Python связан с тем, что этот язык достаточно прост в изучении и обладает рядом библиотек и модулей, разработанных специально для выполнения ряда задач из области инженерного эксперимента, математического анализа, работы с данными, статистики и др. Это позволяет использовать язык для научных разработок специалистам, чей профиль непосредственно не связан с программированием.

Для первоначальной обработки данных о потреблении энергетических ресурсов и погодных данных использовался язык работы с базами данных *SQL*.

Для построения регрессионной модели прогнозирования потребления тепловой энергии рассматриваемого объекта использовались методы машинного обучения (*Random forest* и *Support Vector Machines*).

Машинное обучение (*Machine Learning, ML*) – совокупность методов, позволяющая создавать самообучающиеся компьютерные системы. Помимо более глобальной задачи по созданию нейросетей и искусственного интеллекта, методы машинного обучения используются для более узких задач: классификация, регрессия, кластеризация и др. [3-5].

Существует ряд преимуществ использования алгоритмов машинного обучения для решения задач регрессии: возможность анализировать очень большие объемы данных, возможность постоянного повышения эффективности модели за счет обновления входных данных (к примеру, погодных), высокая скорость осуществления расчетов и точность разработанных моделей.

В нашем случае, для построения модели прогнозирования потребления энергетических ресурсов, использовались наиболее подходящие алгоритмы для задач, связанных с регрессией – *Random forest* (дерево решений) и *Support Vector Machines* (метод опорных векторов) [3, 4].

Алгоритм *Random forest* достаточно точный, для построения модели с

использованием библиотеки *sclearn* необходимо незначительное число параметров. В общем виде, работа алгоритма основывается на изучении иерархии «Если – то...», что в результате приводит к определенному решению [3, 4].

Алгоритм *Support Vector Machines* является наиболее универсальным и используется во многих направлениях. Работа алгоритма основывается на создании в n -мерном пространстве гиперплоскости, которая разделяет данные на классы. Точность предсказания зависит от расстояния от гиперплоскости до объектов – чем оно больше, тем выше точность [3, 4].

На основе используемых погодных данных и данных о потреблении тепловой энергии будет сопоставлена эффективность этих двух алгоритмов в условиях ограниченного количества данных для обучения модели.

Для определения значимости используемых для обучения модели переменных (в нашем случае – погодных данных) использовался коэффициент корреляции Пирсона. В общем виде, выражение для расчета коэффициента корреляции Пирсона имеет вид [2, 3]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

где x_i , y_i – числовые значения рассматриваемых переменных,
 n – объем выборки.

Для анализа эффективности разработанной модели прогнозирования использовались следующие параметры: коэффициент детерминации R^2 и среднеквадратическая ошибка MSE [2, 3].

Коэффициент детерминации R^2 показывает, насколько близки результаты, спрогнозированные моделью, к реальным данным. В общем виде, выражение для определения коэффициента детерминации R^2 имеет вид:

$$R^2 = 1 - \frac{\sigma^2}{\sigma_y^2}, \quad (2)$$

где σ^2 – условная дисперсия зависимой переменной (дисперсия ошибки модели);
 σ_y^2 – дисперсия случайной величины y .

Среднеквадратическая ошибка измеряет среднее значение квадратов ошибок, то есть

среднеквадратичную разницу между оценочными значениями и фактическим значением. Она имеет положительное значение, при этом, чем меньше её значение, тем более точной является рассматриваемая модель.

В общем виде выражение для определения среднеквадратической ошибки MSE имеет вид:

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$$

где n - количество наблюдений;
 Y_i – фактическое значение;
 \hat{Y}_i - предсказанное значение.

Разработка регрессионной модели прогнозирования потребления тепловой энергии в зависимости от погодных данных

Итоговые данные для обучения модели отображены на рис. 5, где heat_cons – суточное потребление рассматриваемым объектом тепловой энергии в Гкал.

Значения коэффициента корреляции Пирсона отображены на рис. 6.

Ожидаемо, наибольшее влияние на потребление оказывают температуры наружного воздуха. Остальные факторы имеют умеренную и слабую корреляцию.

	date	temp	tempmin	tempmax	humidity	precip	windspeed	cloudcover	solarenergy	heat_cons
0	2018-01-02	2.6	0.5	4.8	94.5	0.0	19.4	75.1	3.4	8.5811
1	2018-01-03	-0.4	-2.7	2.6	89.1	0.0	29.2	69.6	3.9	10.4148
2	2018-01-04	-0.9	-3.1	1.3	89.7	3.3	24.8	98.1	2.6	11.1996
3	2018-01-05	0.0	-1.5	2.7	87.3	0.2	18.7	83.0	3.9	10.6862
4	2018-01-06	1.0	-0.9	3.9	91.4	0.0	16.2	84.6	3.1	10.2795
...
240	2020-04-17	11.7	6.8	18.4	64.3	0.0	38.5	81.5	18.5	5.1600
241	2020-04-18	10.8	5.6	17.2	53.2	0.0	23.4	50.3	23.0	5.5579
242	2020-04-19	9.8	3.8	15.2	55.5	0.8	14.4	87.8	15.6	5.5358
243	2020-04-20	7.6	3.1	13.0	71.0	13.2	24.8	57.5	15.5	5.4902
244	2020-04-21	5.6	0.1	10.1	68.0	0.0	22.3	25.8	17.4	5.7256

245 rows × 10 columns

Рисунок 5 – Входные данные для обучения модели

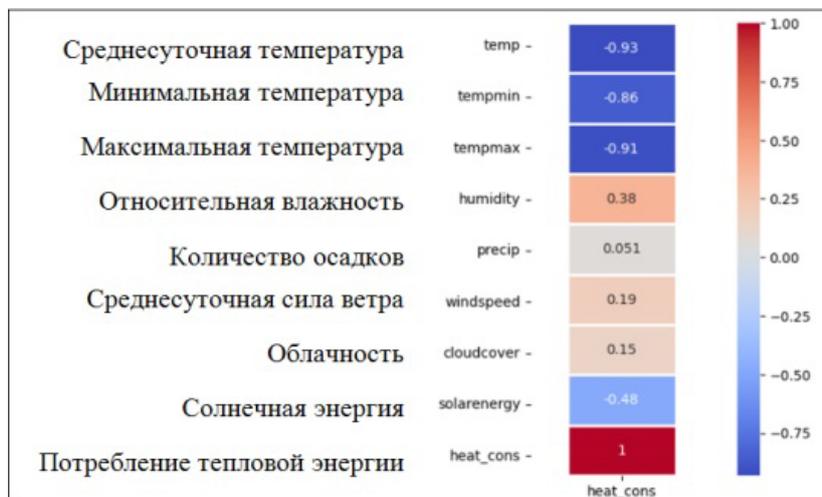


Рисунок 6 – Значения коэффициента корреляции Пирсона

Для обучения модели будут использованы все указанные погодные параметры. Обучение модели для прогнозирования потребления тепловой энергии при использовании алгоритма Random Forest.

Фрагмент кода, с помощью которого

```
#Импортирование библиотек
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
from sklearn.ensemble import RandomForestRegressor

#Задание параметров для модели
RFreg_daily= RandomForestRegressor(n_estimators = 500, max_features = 'log2', max_depth = 7, random_state=0)
RFreg_daily.fit(x_train,y_train)
Predicted_train = RFreg_daily.predict(x_train)

#Анализ эффективности модели
print(r2_score(y_train,Predicted_train))
print(mean_squared_error(y_train,Predicted_train))

0.9733644444893533
0.14894040036277562
```

Рисунок 7 – Значения коэффициента корреляции Пирсона

Значения коэффициента детерминации и среднеквадратической ошибки для тренировочных данных отображены в табл. 1.

Таблица 1 – Критерии эффективности модели

Параметр	Значение
Коэффициент детерминации R^2	0,97
Среднеквадратическая ошибка MSE	0,15

импортируется необходимый алгоритм, вносятся некоторые параметры, вносятся тренировочные данные, а также рассчитывается коэффициент детерминации R^2 и среднеквадратическая ошибка отображен на рис. 7.

Фактическое и прогнозируемое потребление тепловой энергии для тренировочных данных отображено на рис. 8. Исходя из данных табл. 1 и данных графика на рис. 8, разработанная модель показывает достаточную эффективность на тренировочных данных. Далее необходимо проверить модель на тестовых данных. Визуализация фрагмента дерева принятия решений созданной модели изображена на рис. 9.

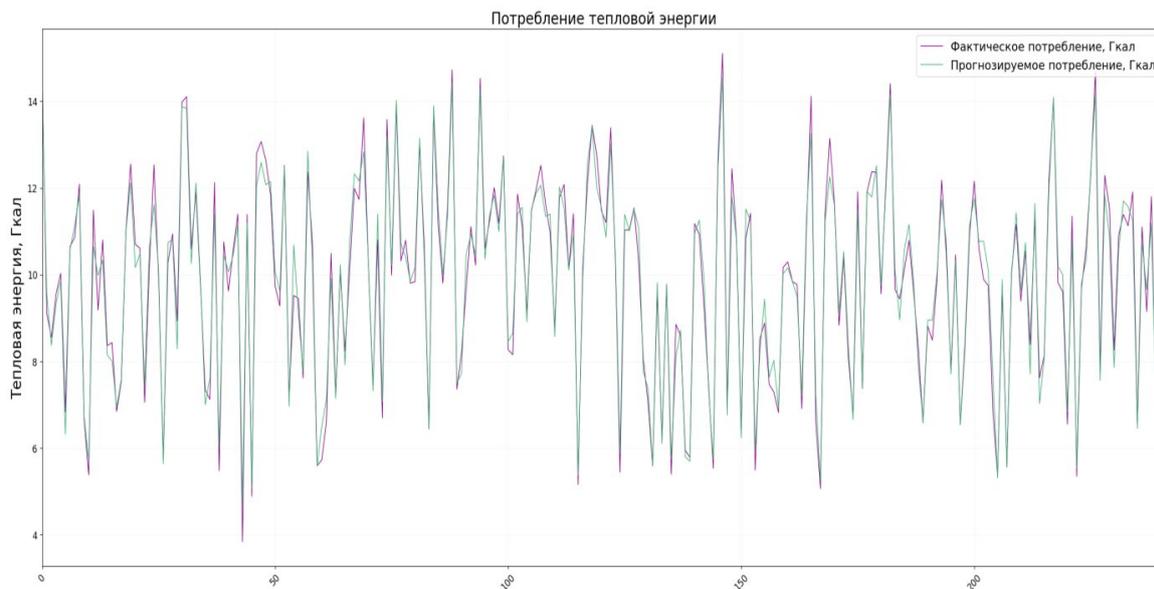


Рисунок 8 – Фактическое и прогнозируемое потребление для тренировочных данных (Random Forest)

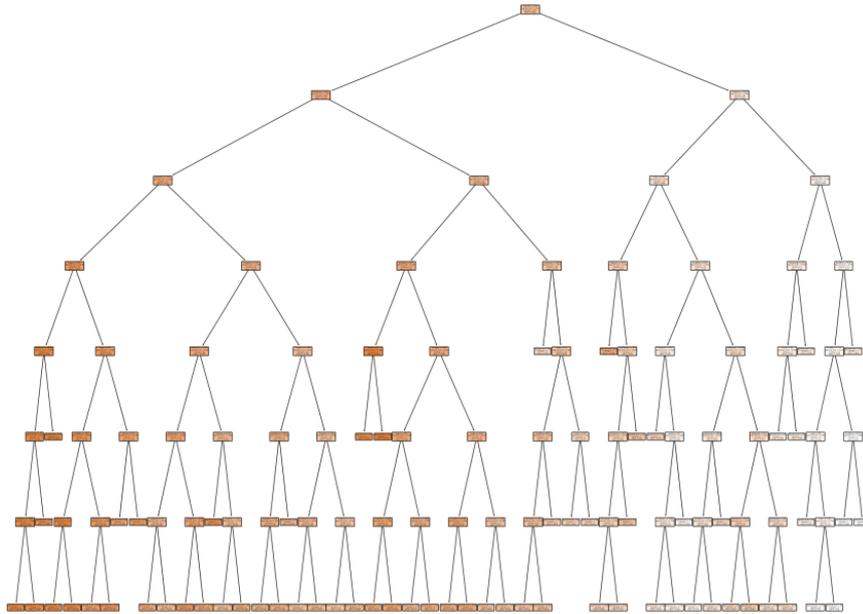


Рисунок 9 – Фрагмент дерева принятия решений для рассматриваемой модели

Анализ эффективности модели на основе алгоритма Random Forest на тестовых данных

Погодные данные и данные о потреблении тепловой энергии, используемые для тестирования, отображены на рис. 10.

Далее тестируемые данные были подставлены в созданную ранее модель, проведен расчет коэффициента детерминации и среднеквадратической ошибки (табл. 2).

График фактического и прогнозируемого потребления для тестовых данных отображен на рис. 11.

	date	temp	tempmin	tempmax	humidity	precip	windspeed	cloudcover	solarenergy	heat_cons
0	2020-02-01	1.4	-1.1	4.9	94.5	1.6	27.4	91.8	3.8	10.8047
1	2020-02-02	2.6	-1.4	6.5	94.6	2.2	36.7	86.1	3.8	8.6980
2	2020-02-03	3.7	0.2	6.6	93.7	10.5	27.4	88.5	3.6	8.0367
3	2020-02-04	0.3	-3.1	4.9	83.7	0.1	30.2	51.8	7.4	10.1735
4	2020-02-05	0.4	-1.3	1.7	96.1	30.1	32.4	93.9	1.9	11.0592
5	2020-02-06	-1.2	-4.2	0.4	93.3	22.5	41.0	91.0	3.8	11.4630
54	2020-03-26	5.4	-1.4	13.3	63.4	0.0	31.7	6.5	16.5	7.1886
55	2020-03-27	8.5	2.0	15.9	54.2	0.0	40.7	57.1	18.0	5.9058
56	2020-03-28	9.2	3.5	16.2	49.6	0.0	38.2	84.6	11.7	5.3792
57	2020-03-29	8.7	3.0	14.7	59.3	0.0	29.5	90.4	14.5	5.3247
58	2020-03-30	11.3	6.0	17.9	73.4	0.0	20.5	62.3	15.4	5.3313
59	2020-03-31	7.8	2.6	13.7	72.2	0.0	24.5	18.3	18.4	5.5425

Рисунок 10 – Данные для тестирования

Таблица 2 – Критерии эффективности модели

Параметр	Значение
Коэффициент детерминации R^2	0,89
Среднеквадратическая ошибка MSE	0,93

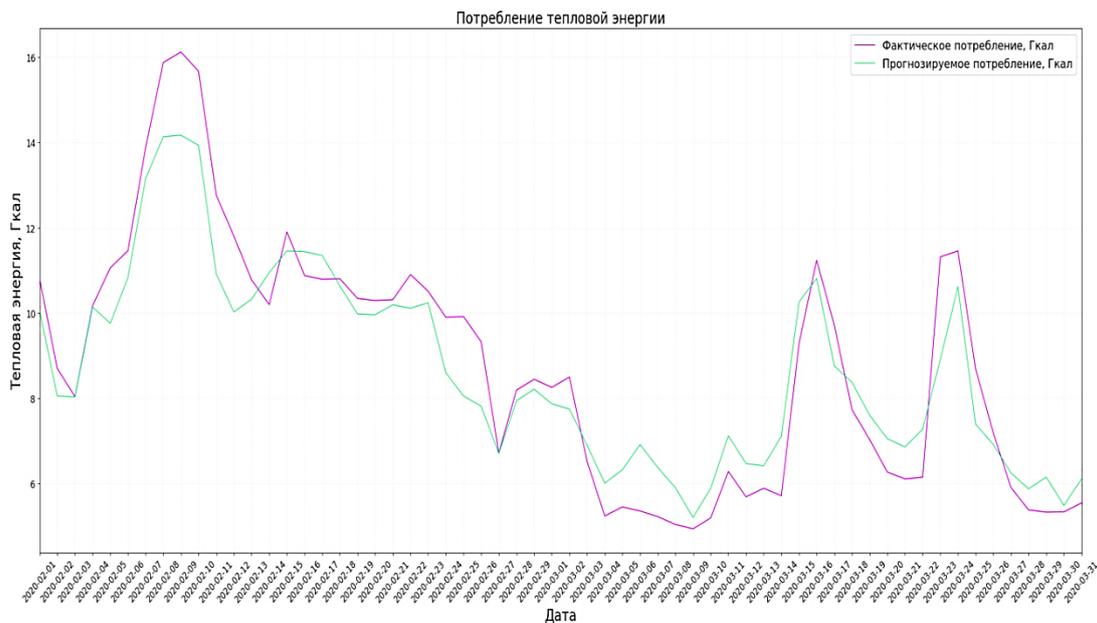


Рисунок 11 – Фактическое и прогнозируемое потребление для тестовых данных (Random Forest)

Несмотря на некоторое снижение эффективности для тестовых данных, разработанная модель всё равно обладает достаточно высокой эффективностью. Некоторое снижение коэффициента детерминации и повышение среднеквадратической ошибки указывают, скорее, не на недостаток модели, а на ограниченное количество данных, используемых для тренировки модели.

При увеличении тренировочных данных эффективность модели на тестовых данных будет увеличиваться.

Обучение модели для прогнозирования потребления тепловой энергии при использовании алгоритма Support Vector Machines

Фрагмент кода, с помощью которого импортируется необходимый алгоритм, вносятся некоторые параметры и тренировочные данные, а также рассчитываются коэффициент детерминации R^2 и среднеквадратическая ошибка, отображен на рис. 12.

```
[88] from sklearn.svm import SVR
      SVReg = SVR(kernel = 'rbf',C=40.0, epsilon = 0.03, gamma = 'scale') #C=1.0, epsilon = 0.1, gamma = 'scale' - default
      SVReg.fit(x_train_SVR, y_train_SVR)

SVR(C=40.0, epsilon=0.03)

Predicted_train_SVR = SVReg.predict(x_train_SVR)
Predicted_train_SVR

[17] from sklearn.metrics import r2_score
      from sklearn.metrics import mean_squared_error

[90] print(r2_score(y_train_SVR,Predicted_train_SVR))
      print(mean_squared_error(y_train_SVR,Predicted_train_SVR))

0.9019015633705991
0.548545736945683
```

Рисунок 12 – Фрагмент кода с обучением модели алгоритма Support Vector Machines

Значения коэффициента детерминации и среднеквадратической ошибки для тренировочных данных отображены в табл. 3. Фактической и прогнозируемое потребление тепловой энергии для тренировочных данных отображены на рис.13.

Таблица 3 – Критерии эффективности модели

Параметр	Значение
Коэффициент детерминации R^2	0,90
Среднеквадратическая ошибка MSE	0,55

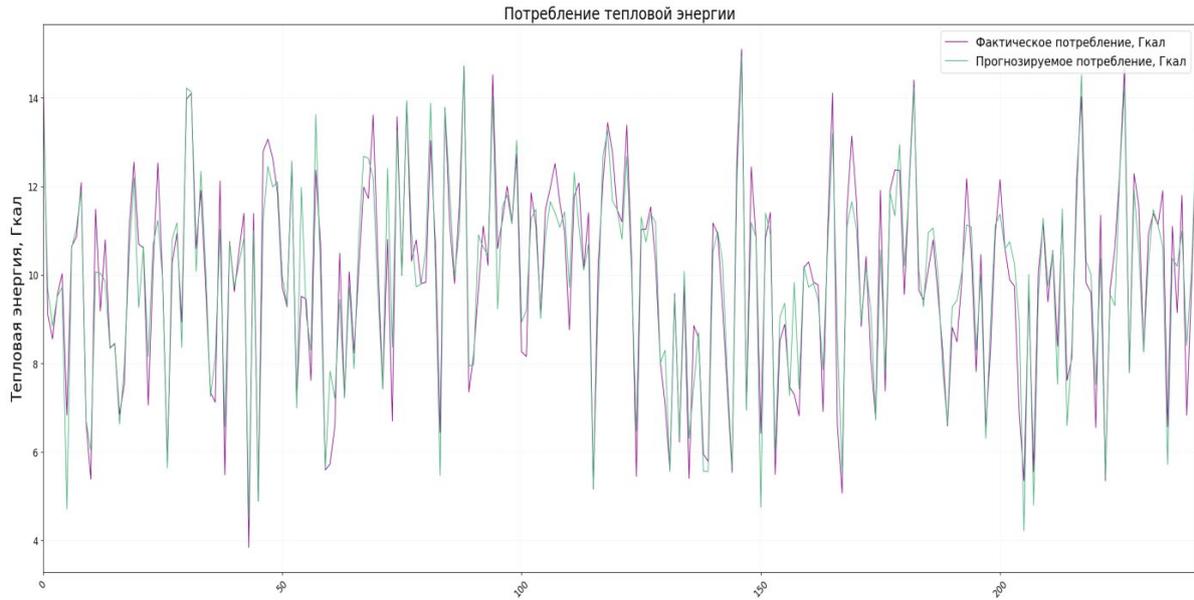


Рисунок 13 – Фактическое и прогнозируемое потребление для тренировочных данных (*Support Vector Machines*)

Исходя из результатов таблицы 3 и графика на рис. 13, разработанная модель показывает достаточно высокую эффективность на тренировочных данных. Однако величина среднеквадратической погрешности для тренировочных данных несколько выше, чем для алгоритма *Random Forest*. Далее необходимо проверить модель на тестовых данных.

Анализ эффективности модели на основе алгоритма *Support Vector Machines* на тестовых данных

Погодные данные и данные о потреблении тепловой энергии, которые используются для

тестирования, отображены на рис. 10. Далее тестируемые данные были подставлены в созданную модель, проведен расчет коэффициента детерминации и среднеквадратической ошибки (табл. 4). График фактического и прогнозируемого потребления для тестовых данных отображен на рис. 14.

Таблица 4 – Критерии эффективности модели

Параметр	Значение
Коэффициент детерминации R^2	0,90
Среднеквадратическая ошибка MSE	0,8

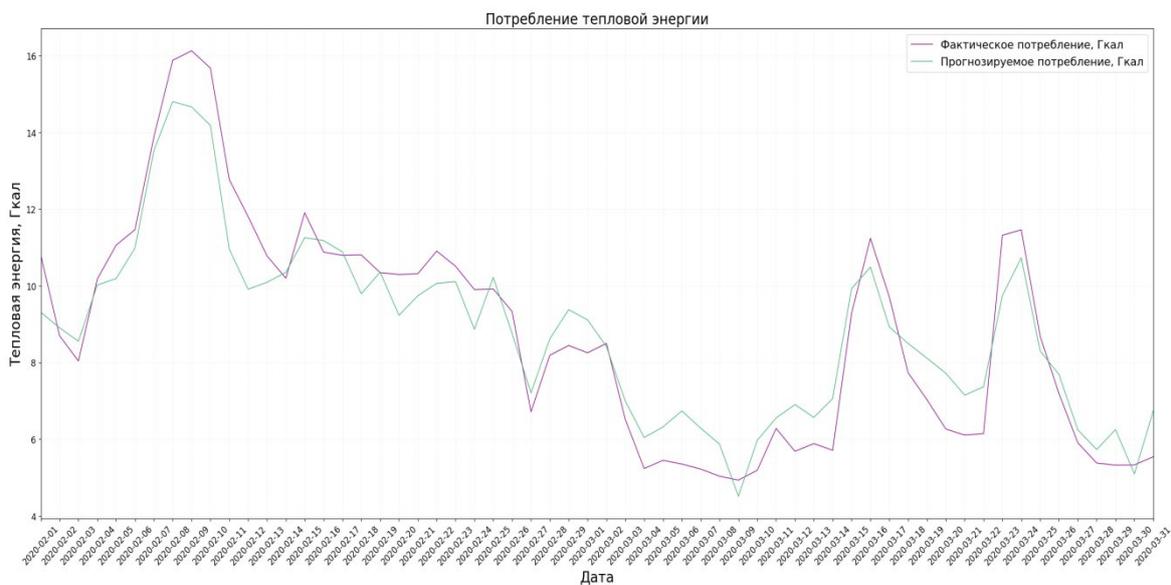


Рисунок 14 – Фактическое и прогнозируемое потребление для тестовых данных (*Support Vector Machines*)

В отличие от модели на основе алгоритма *Random Forest*, модель на основе алгоритма *Support Vector Machines* показывает несколько большую точность на тестовых данных. В любом случае, как и для модели на основе алгоритма *Random Forest*, точность предсказаний будет повышаться при увеличении количества данных для обучения модели.

Выводы

Прогнозирование потребления энергетических ресурсов имеет значительный потенциал применения в сферах энергоснабжения, энергосбережения, учета потребления энергетических ресурсов [6-9]. Современные ресурсы позволяют анализировать значительные объемы данных и создавать модели, которые с достаточно высокой точностью способны прогнозировать потребление энергетических ресурсов для отдельных объектов или комплексов [10].

В данной работе были построены модели прогнозирования потребления тепловой энергии объектом сферы здравоохранения в зависимости от погодных данных на основе алгоритмов машинного обучения. При этом набор входных данных для обучения моделей был весьма ограничен (непосредственно для обучения модели были использованы данные 246 суток).

В качестве алгоритмов были использованы *Random Forest* и *Support Vector Machines* т.к. именно эти алгоритмы более всего подходят для решения задач регрессии. Точность разработанных моделей проверялась и сопоставлялась на тренировочных и тестовых данных.

Было установлено, что обе модели показывают достаточно высокую эффективность в условиях ограниченного количества данных для обучения. При этом для алгоритма *Support Vector Machines* показывает себя несколько лучше именно на тестовых данных (значение среднеквадратической ошибки ниже, отличие между критериями эффективности для тренировочных и тестовых данных также ниже).

В любом случае, оба алгоритма имеют значительный потенциал для построения моделей прогнозирования потребления энергетических ресурсов.

Литература

1. Копейка, Д. В. Определение параметров влагопередачи и расположения плоскости возможной конденсации в системах навесных фасадов с вентилируемой воздушной прослойкой / Д. В. Копейка, С. В. Гридин // Современное промышленное и гражданское

строительство. – Макеевка: ДонНАСА, 2019. – Т. 15. - №1. – С. 5–11.

2. Елисеева, И. И., Юзбашев М. М. Общая теория статистики: учебник / Под ред. И.И. Елисеевой. - 5-е изд., перераб. и доп. - М.: Финансы и статистика, 2004. — 656 с.

3. Larry Wasserman. All of statistics: a concise course in statistical inference / Larry a. Wasserman. 1959.

4. Andreas C. Mueller and Sarah Guido. Introduction to Machine Learning with Python. 2006.

5. Sebastian Raschka. Python Machine Learning. 2015.

6. Сергеев, Н. Н. Теоретические аспекты энергосбережения и повышения энергетической эффективности промышленных предприятий / Н. Н. Сергеев // Вестник АГТУ. Сер. : Экономика, 2013. - №1. - С. 29-36.

7. Колосов, М. В. Л. Н. Энерго- и ресурсосбережение в системах централизованного теплоснабжения / М. В. Колосов, Л. Н. Борисов // 11-я Межд. ИНК «Проблемы энергосбережения и экологии в промышленном и жилищно-коммунальном комплексах». - Пенза, 2010. - С. 128-130.

8. Копейка, Д. В. Повышение энергоэффективности жилых и административных зданий типовых серий / Д. В. Копейка, С.В. Гридин // Энергия – 2016: материалы XI-й межд. науч.-техн. конф. студ., асп. и мол. учёных, 5-7 апреля 2016 г., РФ – Иваново: Ивановский ГЭУ, 2016.– С. 125-131.

9. Копейка, Д. В. Определение параметров теплового комфорта в помещении при регулировании тепловой нагрузки централизованного теплоснабжения // Д. В. Копейка // Материалы 6-ой международной научно-практической конференции молодых ученых и студентов в рамках 12-ой международной конференции по проблемам горной промышленности, строительства и энергетики «Опыт прошлого – взгляд в будущее». - Тула: ТулГУ, 2016. - С. 255-258.

10. Бирюков, А. Б. Методика оперативного сбора данных для анализа энергоэффективности теплоснабжения общественных зданий / А. Б. Бирюков, А. Ю. Харитонов // Энергетические, управляющие и информационные системы: сб. докладов I-ой межд. научно-техн. конф. – Белгород: Изд-во БГТУ им. В.Г. Шухова, 2016. – С. 40-45.

Бирюков А. Б., Гридин С. В. Моделирование и прогнозирование потребления тепловой энергии с использованием погодных данных. Работа посвящена моделированию и прогнозированию потребления тепловой энергии объектом сферы здравоохранения в зависимости от данных о погодных условиях на основе алгоритмов машинного обучения. В качестве алгоритмов были использованы *Random Forest* и *Support Vector Machines*, так как именно эти алгоритмы более всего подходят для решения задач регрессии. Точность разработанных моделей проверялась и сопоставлялась на тестовых данных. Модели, разработанные на основе указанных алгоритмов, показывают достаточно высокую эффективность в условиях ограниченного количества данных для обучения и имеют значительный потенциал для использования.

Ключевые слова: моделирование, тепловая энергия, алгоритм, машинное обучение.

Biryukov A. B., Gridin S. V. Modeling and forecasting of thermal energy consumption using weather data. The work is devoted to modeling and forecasting the consumption of thermal energy by a healthcare facility depending on weather data based on machine learning algorithms. *Random Forest* and *Support Vector Machines* were used as algorithms, since these algorithms are most suitable for solving regression problems. The accuracy of the developed models was checked and compared on test data. The models developed on the basis of these algorithms show sufficiently high efficiency in conditions of a limited amount of training data and have significant potential for use.

Keywords: modeling, thermal energy, algorithm, machine learning.

Статья поступила в редакцию 12.04.2024
Рекомендована к публикации профессором Павлышом В. Н.