

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное образовательное учреждение высшего профессионального
образования

УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Факультет информационных систем и технологий

Кафедра «Информатика и вычислительная техника»

Отчет

по практической работе

на тему «Анализ динамики потребления электроэнергии в бытовой сети»

по дисциплине «Прикладные задачи анализа данных»

Выполнил:

студент группы ИВТИИмд-11

Белогубов Андрей

Ульяновск

2024

1. Формулировка актуальности работы.

Актуальность темы "Анализ динамики потребления электроэнергии в бытовой сети" обусловлена несколькими ключевыми аспектами:

Во-первых, это помогает энергетическим компаниям и поставщикам электроэнергии более точно планировать производственные мощности и оптимизировать распределение энергии в сети. Точное предсказание потребления также позволяет избежать перегрузок и улучшить общую надежность энергосистемы.

Во-вторых, предсказание потребления мощности играет важную роль в управлении нагрузкой, особенно в контексте умных сетей. Это позволяет снизить пиковые нагрузки, что способствует более эффективному использованию ресурсов и снижению затрат на производство энергии.

Также предсказание потребления мощности имеет значение для потребителей, поскольку позволяет им более эффективно управлять своими энергетическими потребностями, оптимизировать расходы и даже участвовать в программах снижения нагрузки.

Наконец, такие прогнозы играют важную роль в разработке стратегий расширения инфраструктуры и планировании будущих инвестиций в энергетический сектор.

2. Разработка проектных решений

Для анализа данных и выполнения кода используется популярное решение в виде блокнотов Jupyter – среда разработки, где можно пользоваться результатами выполнения отдельных фрагментов кода без необходимости полностью выполнять затратные вычисления вновь и вновь.

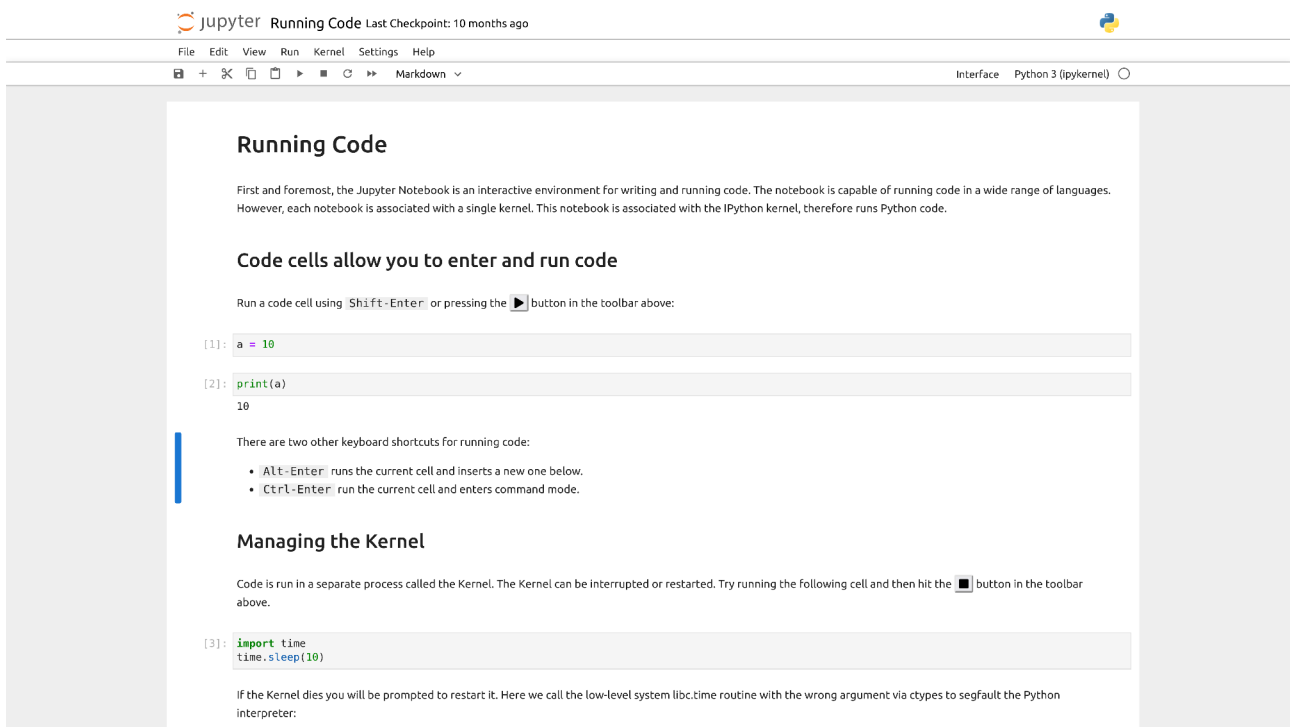


Рис. 2.1. Пример Jupyter-блокнота.

В качестве исходных данных для анализа был выбран датасет «Hourly Energy Consumption» (почасовое потребление энергии) от American Electric Power – крупного поставщика электроэнергии в Америке. Он охватывает временной диапазон от 2004-10-01 до 2018-08-03 и содержит всего 2 колонки:

- Datetime – дата и время измерения с разрешением в 1 час
- Usage – мгновенная мощность в МВт в момент времени Datetime

Этих данных недостаточно для предсказания потребления электроэнергии, поэтому было принято решение дополнить существующий набор данных. Для этого был использован датасет содержащий среднюю за день температуру для большинства городов по всему миру. Поскольку потребление электроэнергии измерялось только в одной стране, остальные страны были отфильтрованы из

датасета. После добавления отфильтрованных данных по температуре датасет принял следующий вид:

- Datetime – дата и время измерения с разрешением в 1 час
- Usage – мгновенная мощность в МВт в момент времени Datetime
- AvgTemperature – средняя температура за день

Для улучшения точности предсказаний модели было принято решение разбить поле Datetime на составляющие компоненты. Итоговый датасет имеет следующий вид:

- Datetime – дата и время измерения с разрешением в 1 час
- Usage – мгновенная мощность в МВт в момент времени Datetime
- Year – год
- Month – месяц
- Day – день
- Hour – час
- Dayofweek – день недели
- Dayofmonth – день месяца
- AvgTemperature – средняя температура за день

	Usage	Year	Month	Day	Hour	Dayofweek	Dayofmonth	AvgTemperature
Datetime								
2004-10-01 01:00:00	1621.0	2004	10	275	1	4	1	57.928571
2004-10-01 02:00:00	1536.0	2004	10	275	2	4	1	57.928571
2004-10-01 03:00:00	1500.0	2004	10	275	3	4	1	57.928571
2004-10-01 04:00:00	1434.0	2004	10	275	4	4	1	57.928571
2004-10-01 05:00:00	1489.0	2004	10	275	5	4	1	57.928571
...
2018-08-02 20:00:00	2554.0	2018	8	214	20	3	2	72.500000
2018-08-02 21:00:00	2481.0	2018	8	214	21	3	2	72.500000
2018-08-02 22:00:00	2405.0	2018	8	214	22	3	2	72.500000
2018-08-02 23:00:00	2250.0	2018	8	214	23	3	2	72.500000
2018-08-03 00:00:00	2042.0	2018	8	215	0	4	3	75.214286

121275 rows × 8 columns

Рис 2.2. Исходный датасет

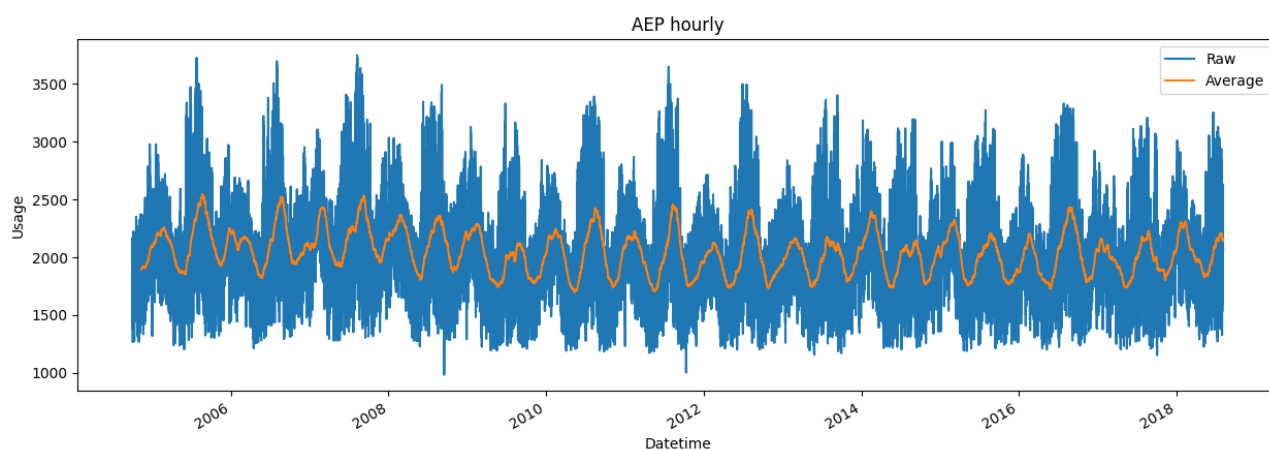


Рис. 2.3. Исходный датасет (Raw) и Усредненные значения(Average) в виде графика

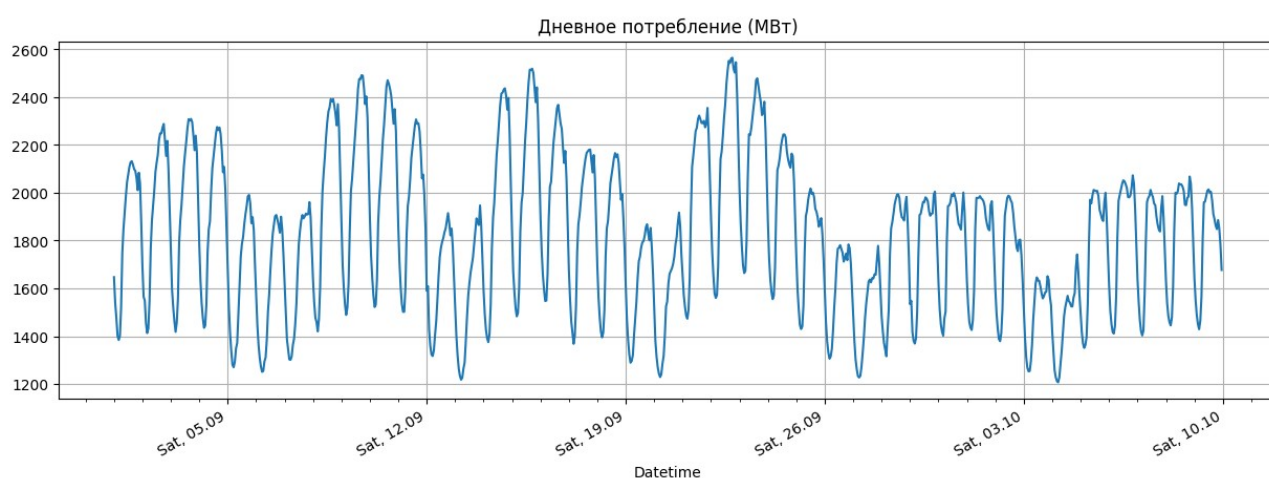


Рис. 2.4. Дневное потребление с 01.09.2009 по 10.10.2009

По графикам потребления можем заметить закономерности как в годовых масштабах, так и в пределах нескольких недель. Для более точной оценки построим диаграммы размаха (Boxplot, ящик с усами) для потребления в зависимости от часа, дня недели и месяца:

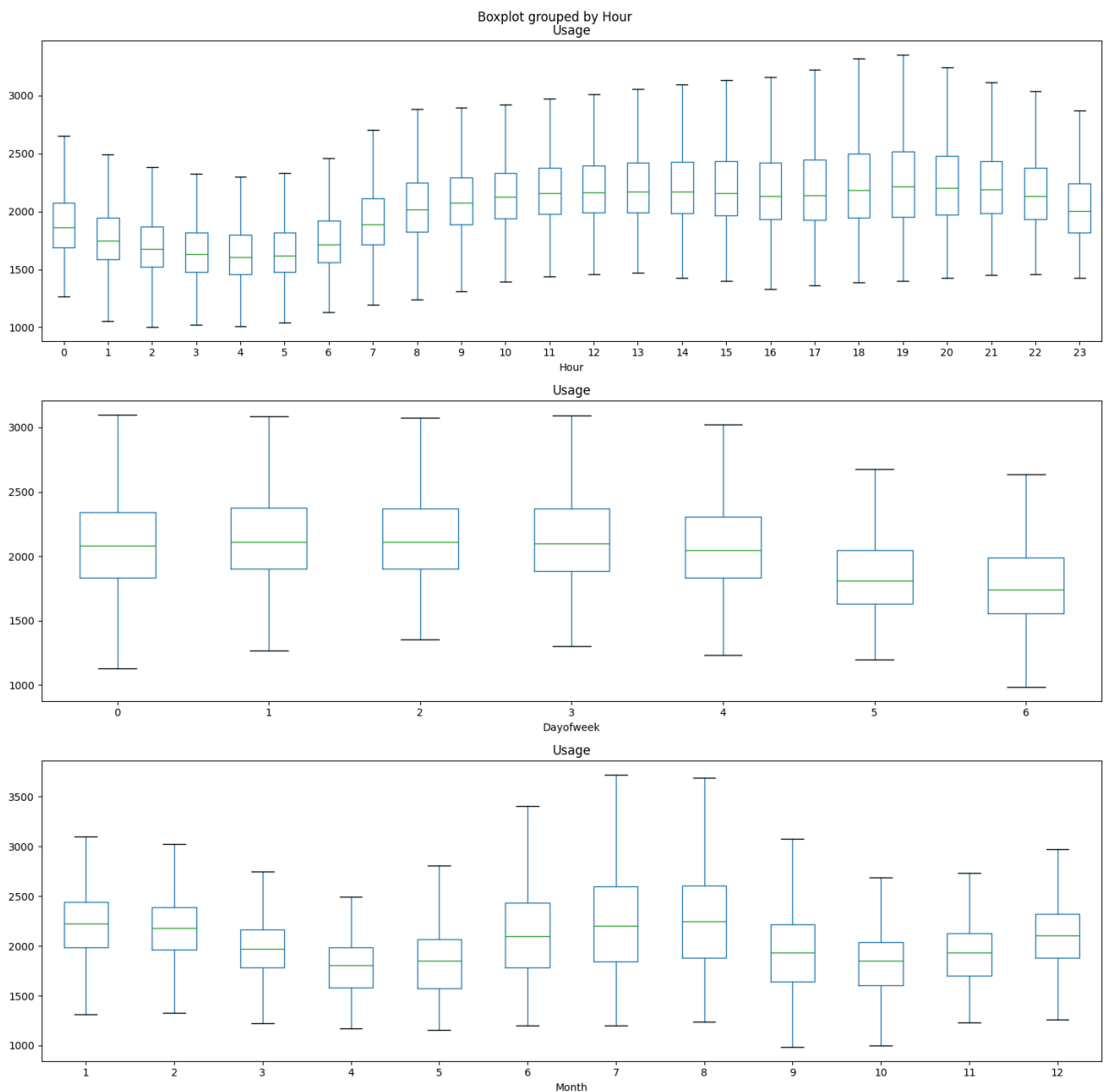


Рис. 2.5. Диаграмма размаха для потребления в зависимости от часа, дня недели и месяца

По диаграммам можем заметить, что потребление имеет тенденцию снижаться в ночные часы, выходные дни и осенний/весенний период, когда не нужны обогреватели и кондиционеры.

3. Обучение модели

Для предсказания данных используется популярная python-библиотека XGBoost. В основе XGBoost лежит алгоритм градиентного бустинга деревьев решений. Градиентный бустинг — это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Обучение ансамбля проводится последовательно в отличие, например от бэггинга. На каждой итерации вычисляются отклонения предсказаний уже обученного ансамбля на обучающей выборке. Следующая модель, которая будет добавлена в ансамбль будет предсказывать эти отклонения. Таким образом, добавив предсказания нового дерева к предсказаниям обученного ансамбля мы можем уменьшить среднее отклонение модели, которое является таргетом оптимизационной задачи. Новые деревья добавляются в ансамбль до тех пор, пока ошибка уменьшается, либо пока не выполняется одно из правил "ранней остановки".

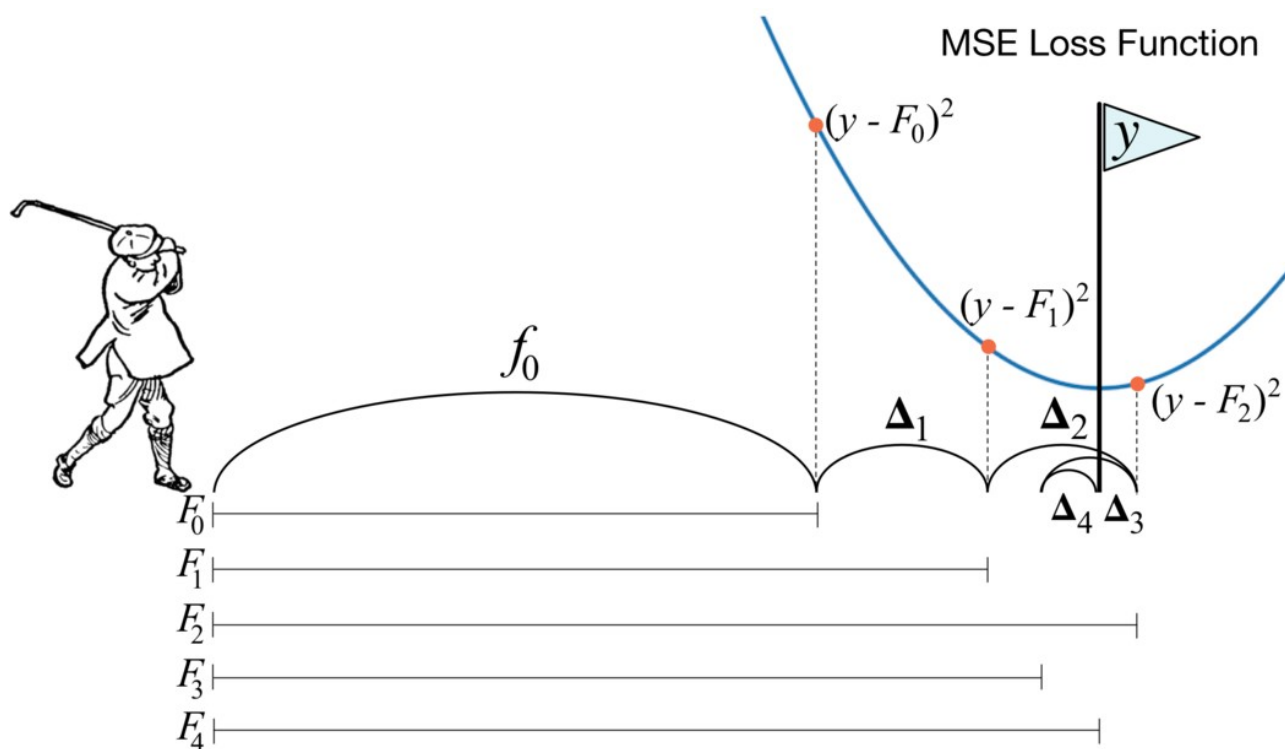


Рис. 3.1. Иллюстрация бустинга

Рассмотрим иллюстрацию бустинга. На ней рассматривается поведение модели на одной точке абстрактной задачи линейной регрессии. Предположим, что первая модель ансамбля F всегда выдает выборочное среднее предсказываемой величины f_0 . Такое предсказание довольно грубое, поэтому среднеквадратичное отклонение на выбранной нами точке будет довольно большим. Мы попробуем это исправить обучив модель Δ_1 , которая будет "корректировать" предсказание предыдущего ансамбля F_0 . Таким образом мы получим ансамбль F_1 , предсказание которого будет суммироваться из предсказаний моделей f_0 и Δ_1 . Продолжая такую последовательность мы приходим к ансамблю F_4 предсказание которого суммируется из предсказаний f_0 , Δ_1 , Δ_2 , Δ_3 , Δ_4 и предсказывает в точности значение заданного таргета.

Для обучения модели разделим данные на обучающую и тестовую выборку по дате 01-01-2017:

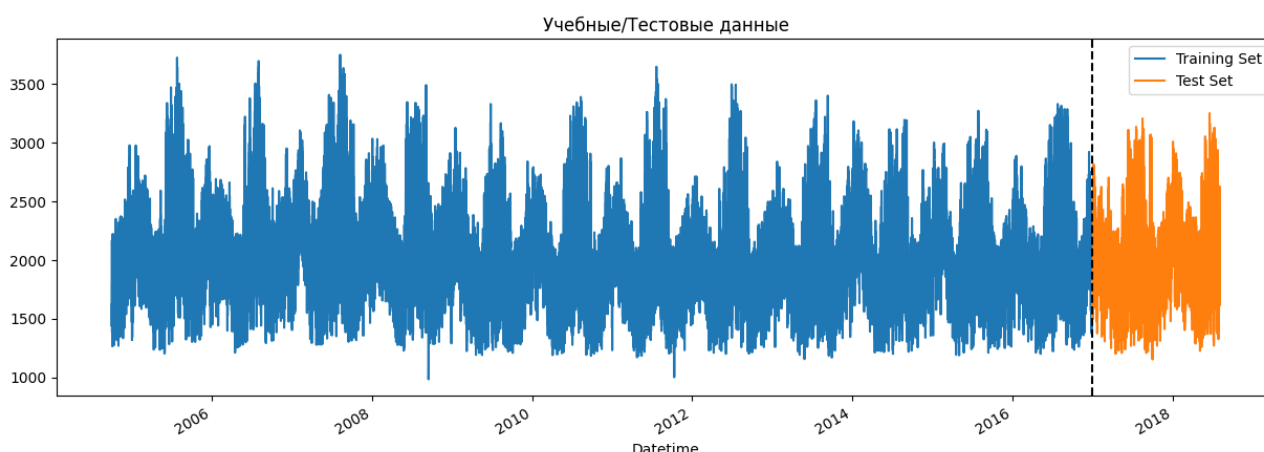


Рис 3.2. Учебные и тестовые данные

Код для разделения данных и обучения модели представлен в листинге 1. Он принимает в качестве параметра датафрейм (`df`), содержащий данные в виде датафрейма библиотеки `pandas` и создает 4 колонки — `X_train`, `X_test`, содержащие признаки для обучения для обучающей и тестовой выборки, и `y_train` и `y_test`, содержащие потребление для обучающей и тестовой выборки. После этого данные передаются модели для обучения. После обучения модель готова к использованию.


```

split_point = "01-01-2017"
FEATURES = ["Year", "Month", "Day", "Hour", "Dayofweek", "Dayofmonth",
"AvgTemperature"]
TARGET = "Usage"

train = df.loc[df.index < split_point]
test = df.drop(train.index)

X_train = train[FEATURES]
y_train = train[TARGET]

X_test = test[FEATURES]
y_test = test[TARGET]

reg = xgb.XGBRegressor(
    n_jobs=8,
    booster="gbtree",
    n_estimators=1000,
    early_stopping_rounds=50,
    objective="reg:squarederror",
    max_depth=5,
    learning_rate=0.01,
)
reg.fit(X_train, y_train, eval_set=[(X_train, y_train), (X_test,
y_test)], verbose=100)

```

Код для использования модели представлен в листинге 3.2. Он использует модель, полученную на предыдущем шаге для построения графиков реального и предсказанного потребления и измерения важности признаков.

Листинг 3.2

```

test["Prediction"] = reg.predict(X_test)

fi = pd.DataFrame(
    data=reg.feature_importances_,
    index=reg.feature_names_in_,
    columns=["importance"],
)
fi.sort_values("importance").plot(kind="barh", title="Важность
признаков")
plt.show()

```

```

fig, ax = plt.subplots(figsize=(15, 20), nrows=3)
combined = test[["Usage", "Prediction"]]
comb_zoomed = combined[
    (combined.index >= "03.01.2018") & (combined.index < "04.01.2018")
]
mean_avg = combined.rolling(window=1000).mean()

combined.plot(ax=ax[0], title="Реальные и предсказанные значения")
comb_zoomed.plot(ax=ax[1], title="Дневные значения")
mean_avg.plot(ax=ax[2], title="Усредненные значения")

fig.tight_layout()
plt.show()

```



Рис. 3.1. Важность признаков (в долях)

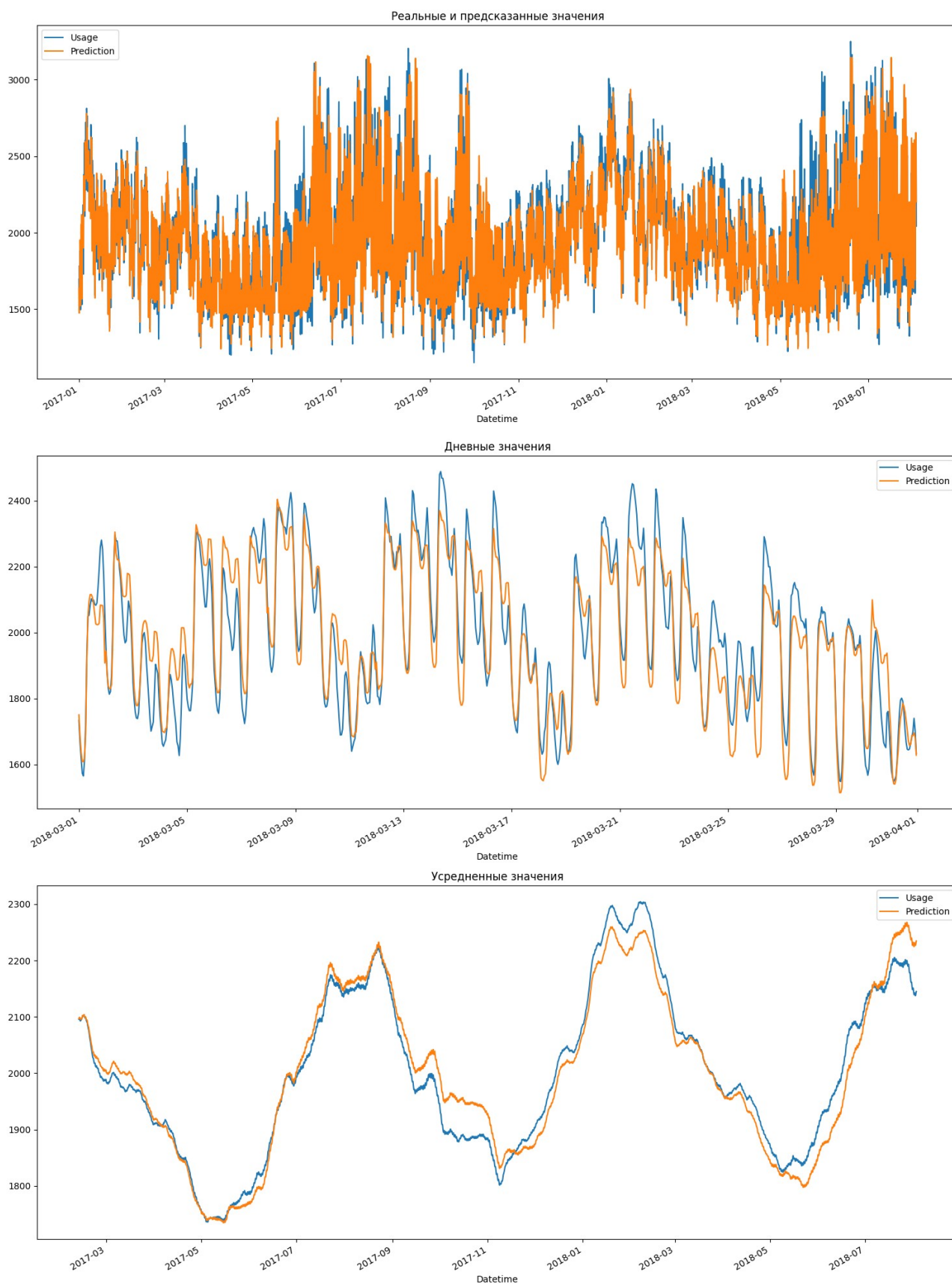


Рис 3.3 Реальные (Usage) и предсказанные (Predicted) значения

По результирующим графикам можем видеть, что модель неплохо предсказывает потребление в рамках одного дня, нескольких недель или года. Усредненные значения хорошо повторяют форму сигнала, но не учитывают неизвестные факторы, ввиду которых предсказанные значения смещаются относительно реальных.

Для дальнейшего улучшения результатов можно учитывать потребление за предыдущий час, день недели или месяц. Другим способом увеличения точности является поиск параметров влияющих на использование энергии таких как влажность, осадки, праздники и подобные события, влияющие на увеличение или уменьшение потребления энергии.

Заключение

В рамках данной практической работы был проведен анализ динамики потребления электроэнергии в бытовой сети. В ходе исследования была обучена и протестирована модель на базе деревьев принятия решений XGBoost. Выявлена сильная зависимость потребления электроэнергии в зависимости от времени и температуры.

Использование современных моделей машинного обучения для предсказания помогают снизить затраты, повысить надежность сетей, снизить нагрузки в пиковые часы и способствуют более эффективному использованию ресурсов. Это также открывает новые возможности для участия потребителей в управлении своими энергетическими потребностями. Развитие технологий предсказания потребления мощности будет продолжать играть важную роль в переходе к более устойчивой и инновационной энергетике в будущем.

Список источников

1. Hourly Energy Consumption/[Электронный ресурс] // kaggle.com: [сайт].— URL: <https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption> (Дата обращения: 02.05.2024).
2. XGBoost Documentation/[Электронный ресурс] // readthedocs.io: [сайт].— URL: <https://xgboost.readthedocs.io/en/stable/> (Дата обращения: 02.05.2024).
3. Temperature of All Countries (1995-2020) // kaggle.com: [сайт].—URL: <https://www.kaggle.com/datasets/subhamjain/temperature-of-all-countries-19952020> (Дата обращения: 02.05.2024).