



РАЗРАБОТКА МОДЕЛИ ПРОГНОЗОВ ЗАГРУЗКИ УЗЛОВ ЛОГИСТИЧЕСКОЙ СЕТИ

Студент: Чебенева Екатерина

Руководители: д.ф.-м.н., профессор каф . 810Б, Ревизников Дмитрий
Леонидович

ЦЕЛЬ

реализовать прогнозирование количества посылок в узле для последующего принятия решений по регулированию нагрузки на логистический узел.

1

провести анализ существующих подходов к прогнозированию временных рядов

2

выполнить очистку и предобработку данных

3

реализовать и протестировать модели

4

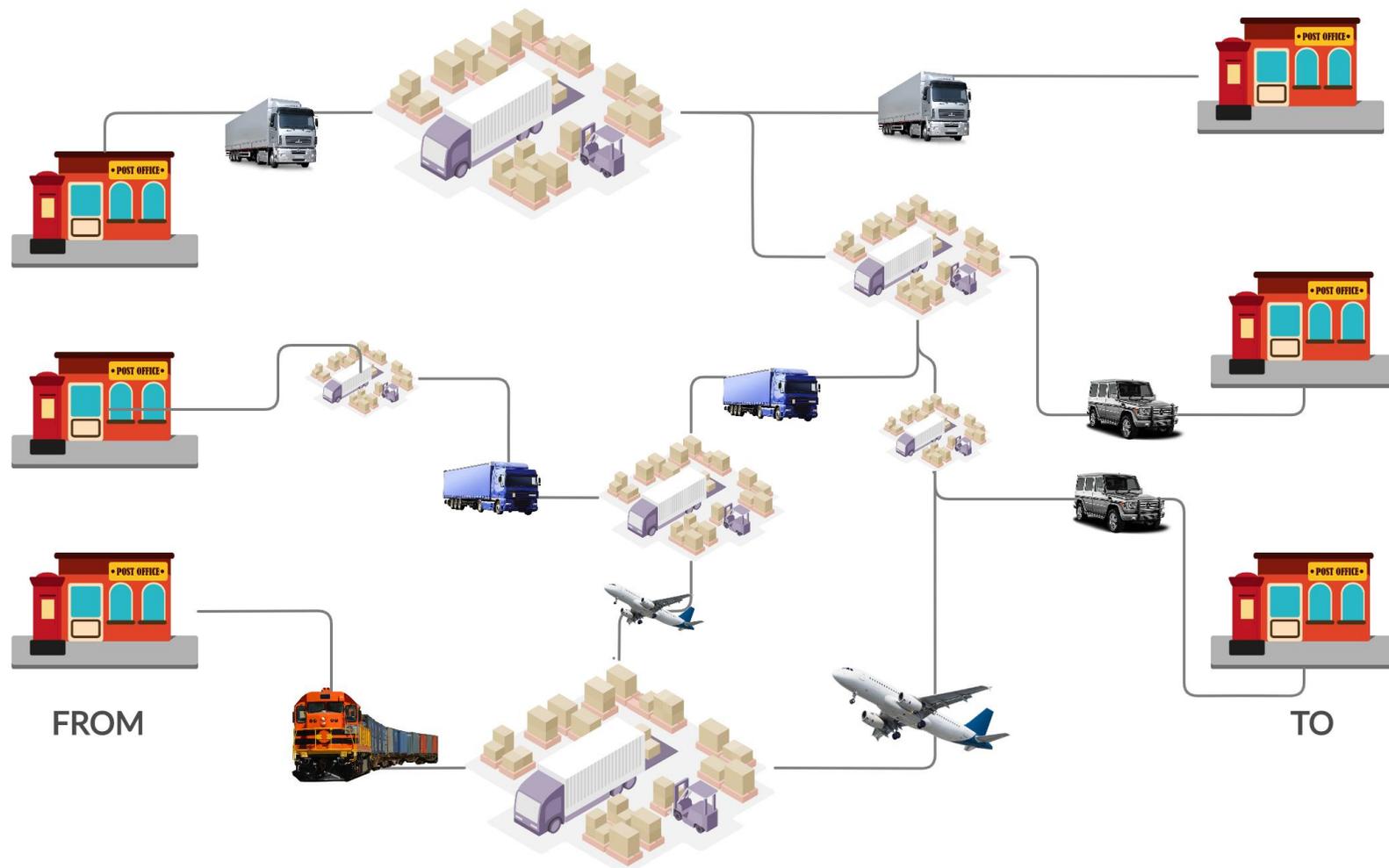
выбрать лучшую модель на основе отобранных метрик

5

интегрировать выбранную модель в производственный процесс

ЛОГИСТИЧЕСКАЯ СЕТЬ

— это большое количество звеньев в логистической системе, находящихся во взаимосвязи между собой по материальным или сопутствующим им информационным и денежным потокам в границах логистической системы.



ТЕРМИНЫ

РПО – регистрируемое почтовое отправление

АОПП – авиационное отделение перевозки почты

ОПС – объект почтовой связи

Плечо – это ребро между двумя ОПС





ОПИСАНИЕ ЦЕЛЕВОЙ ФУНКЦИИ

- Формула расчета целевой переменной *residuals*

$$= \max(left - accepted) + accepted - left$$
- Остатки считаются по временной сетке 8 часов

times_dt	residuals
2019-11-13 16:00:00	19560
2019-11-14 00:00:00	18325
2019-11-14 08:00:00	17990
2019-11-14 16:00:00	18424
2019-11-15 00:00:00	18928
2019-11-15 08:00:00	18252
2019-11-15 16:00:00	18741
2019-11-16 00:00:00	18867
2019-11-16 08:00:00	10485
2019-11-16 16:00:00	8617
2019-11-17 00:00:00	19825

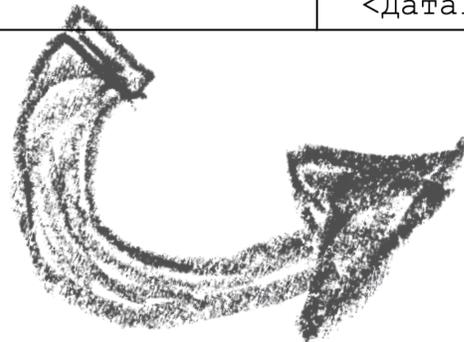
Name: residuals_new, dtype: int64

СТРУКТУРА ДАННЫХ

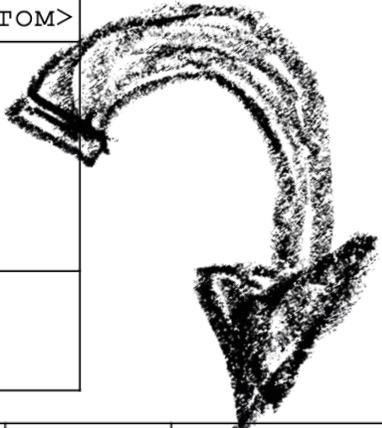
- Домодедово АОПП (57к посылок в день)
- Шереметьево АОПП (23к посылок в день)
- Внуково АОПП (~14к посылок в день)

matreshka_grouped
16-ти значный идентификатор РПО (ШПИ) Время выполнения операции Объект почтовой связи Тип операции Сумма платы за пересылку воздушным транспортом Способ пересылки Почтовый индекс приема ...

ШПИ	array<struct<ДатаВремя, Объект, Тип операции>>	struct<Почтовый индекс приема, Способ пересылки, Сумма платы за пересылку воздушным транспортом>
1400159391235660	{<Дата1 10:20, Люберцы 140015, Прием>, <Дата1 10:45, Люберцы 140015, Отправка>, <Дата1 18:20, Домодедово АОПП, Прием> <Дата1 19:20, Домодедово АОПП, Сортировка> <Дата1 22:20, Домодедово АОПП, Отправка> <Дата2 9:05, Иркутск АОПП, Прием>}	<Люберцы 140015, Авиа, 120р.>
3131293500043133	{<Дата1 10:21, Сургут 628402, Прием>, <Дата1 12:45, Сургут 628402, Отправка>, <Дата1 20:20, Сортировочный центр, Прием>}	<Сургут 628402, Наземный, 0>



Дата время	Кол-во РПО ДомодедовоАОПП-ИркутскАОПП
Дата0 16:00	...
Дата1 0:00	...
Дата1 8:00	...
Дата1 16:00	1
Дата2 0:00	...



ШПИ	array<struct<ДатаВремя, Объект, Тип операции>>	struct<Почтовый индекс приема, Способ пересылки, Сумма платы за пересылку воздушным транспортом>
1400159391235660	{<Дата1 10:20, Люберцы 140015, Прием>, <Дата1 10:45, Люберцы 140015, Отправка>, <Дата1 18:20, Домодедово АОПП, Прием> <Дата1 19:20, Домодедово АОПП, Сортировка> <Дата1 22:20, Домодедово АОПП, Отправка> <Дата2 9:05, Иркутск АОПП, Прием>}	<Люберцы 140015, Авиа, 120р.>
3131293500043133	{<Дата1 10:21, Сургут 628402, Прием>, <Дата1 12:45, Сургут 628402, Отправка>, <Дата1 20:20, Сортировочный центр, Прием>}	<Сургут 628402, Наземный, 0>

ШПИ	Объект	Время начала выполнения операций	Время окончания операций	Следующий объект	Предыдущий объект	Почтовый индекс приема	Способ пересылки	Сумма платы за пересылку воздушным транспортом
1400159391235660	Люберцы 140015	Дата1 10:20	Дата1 10:45	Домодедово АОПП	null	Люберцы 140015	Авиа	120р.
1400159391235660	Домодедово АОПП	Дата1 18:20	Дата1 22:20	Иркутск АОПП	Люберцы 140015	Люберцы 140015	Авиа	120р.
1400159391235660	Иркутск АОПП	Дата2 9:05	...	null	Домодедово АОПП	Люберцы 140015	Авиа	120р.
3131293500043133	Сургут 628402	Дата1 10:21	Дата1 12:45	Сортировочный центр	null	Сургут 628402	Наземный	0р.
3131293500043133	Сортировочный центр	Дата1 20:20	...	null	Сургут 628402	Сургут 628402	Наземный	0р.

ШПИ	Объект	Время начала выполнения операций	Время окончания операций	Следующий объект	Предыдущий объект	Почтовый индекс приема	Способ пересылки	Сумма платы за пересылку воздушным транспортом
1400159391235660	Домодедово АОПП	Дата1 18:20	Дата1 22:20	Иркутск АОПП	Люберцы 140015	Люберцы 140015	Авиа	120р.



Дата время	Кол-во РПО ДомодедовоАОПП-ИркутскАОПП
Дата1 0:00	...
Дата1 8:00	...
Дата1 16:00	1
Дата2 0:00	...



МОДЕЛИ

ARIMA

SARIMA

Prophet

eXtreme Gradient Boosting (XGBoost)

Light Gradient Boosting Machine (LightGBM)

Long short-term memory (LSTM)

ARIMA

Autoregressive integrated moving average

- **Авторегрессионная модель** – регрессия ряда на собственные значения в прошлом.
- **Модель скользящего среднего** – линейная комбинация q последних значений ошибки авторегрессионной модели соответствующих лагов.

ARMA(p, q):

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

ARIMA(p, d, q) – модель ARMA(p, q) для d раз продифференцированного ряда

SARIMA

- Ряд имеет сезонный период длины S

Добавим к ARMA(p, q):

- P авторегрессионных компонент с шагом S :

$$+\beta_s Y_{t-s} + \beta_{2s} Y_{t-2s} + \dots + \beta_{ps} Y_{t-ps}$$

- Q компонент скользящего среднего, взятых с шагом S :

$$+\phi_s \epsilon_{t-s} + \phi_{2s} \epsilon_{t-2s} + \dots + \phi_{qs} \epsilon_{t-qs}$$

SARMA

(p, q) \times (P, Q)

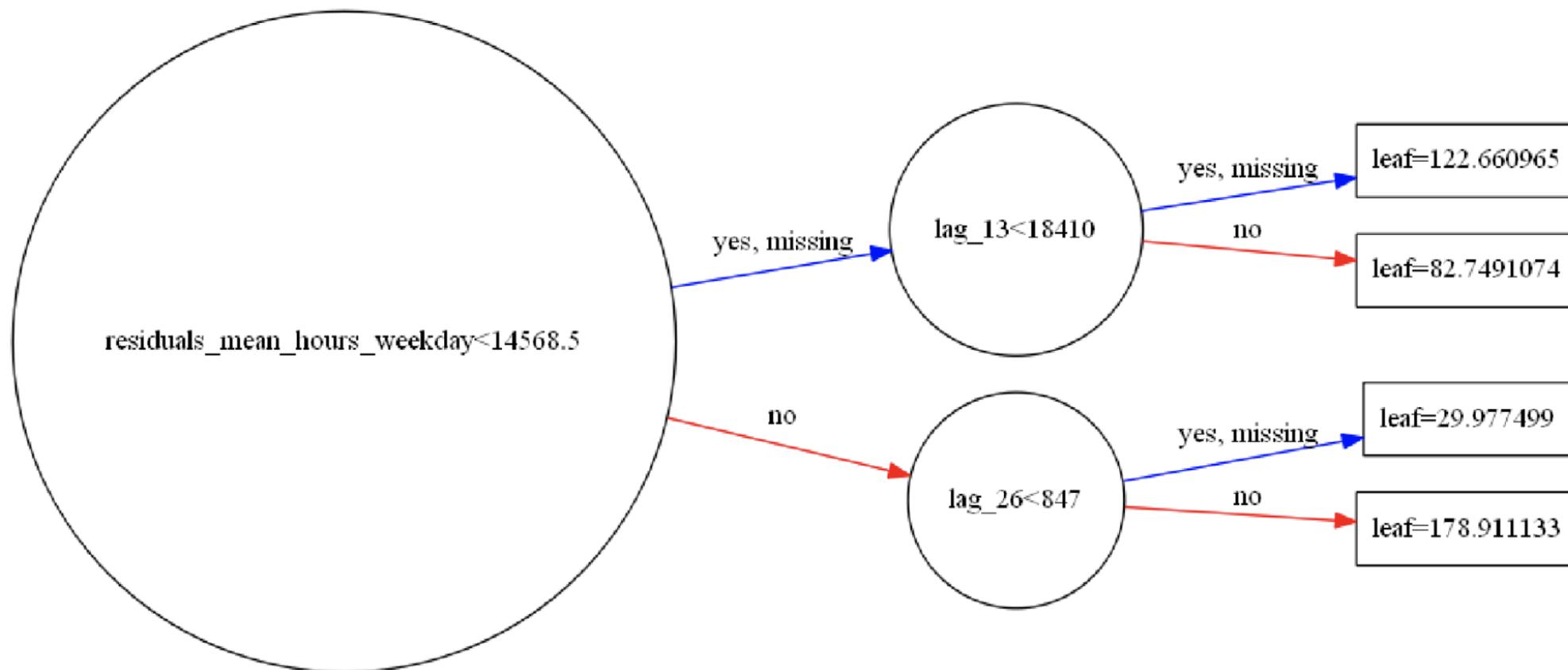
SARIMA(p, d, q) \times (P, D, Q) – модель SARMA(p, q) \times (P, Q) для ряда, к которому d раз применили обычное дифференцирование и D раз – сезонное

Бустинг

☞ это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений b_n .

$$a_N(x) = \sum_{n=1}^N b_n$$
$$\sum_{i=1}^l L(y_i, a_N(x_i)) \rightarrow \min$$

- Деревья добавляются по одному в ансамбль и выполняют исправления ошибок предсказания, сделанных предыдущими моделями



XGBoost



LightGBM

Prophet

☞ представляет собой аддитивную регрессионную модель с четырьмя основными компонентами:

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t)$$

- $g(t)$ – трендовая компонента
- $s(t)$ – сезонные компоненты, отвечают за моделирование периодических изменений
- $h(t)$ отвечает за заданные пользователем **аномальные дни**
- $\epsilon(t)$ охватывает специфические изменения, не учитываемые моделью



LSTM

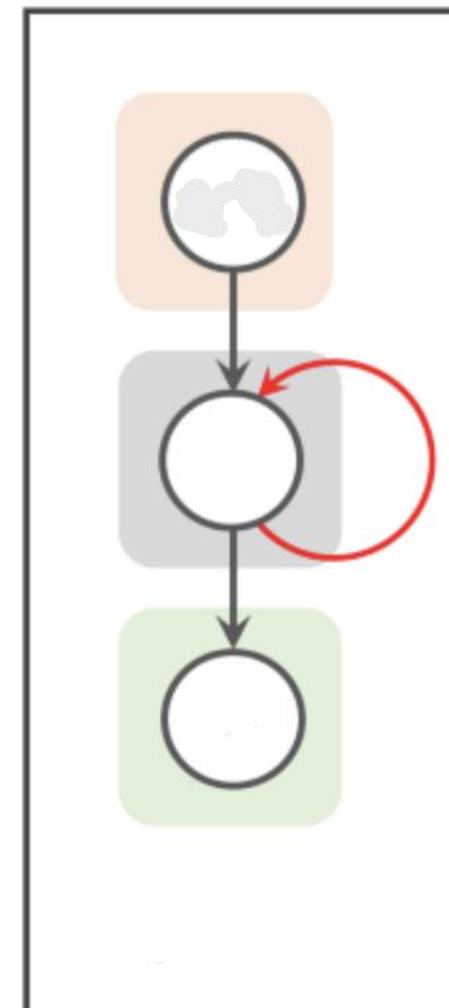
```
# create and fit the LSTM network
model = Sequential()
model.add(LSTM(32, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error',
              optimizer='adam')
model.fit(trainX, trainY, epochs=1000,
          batch_size=1, verbose=2)
```

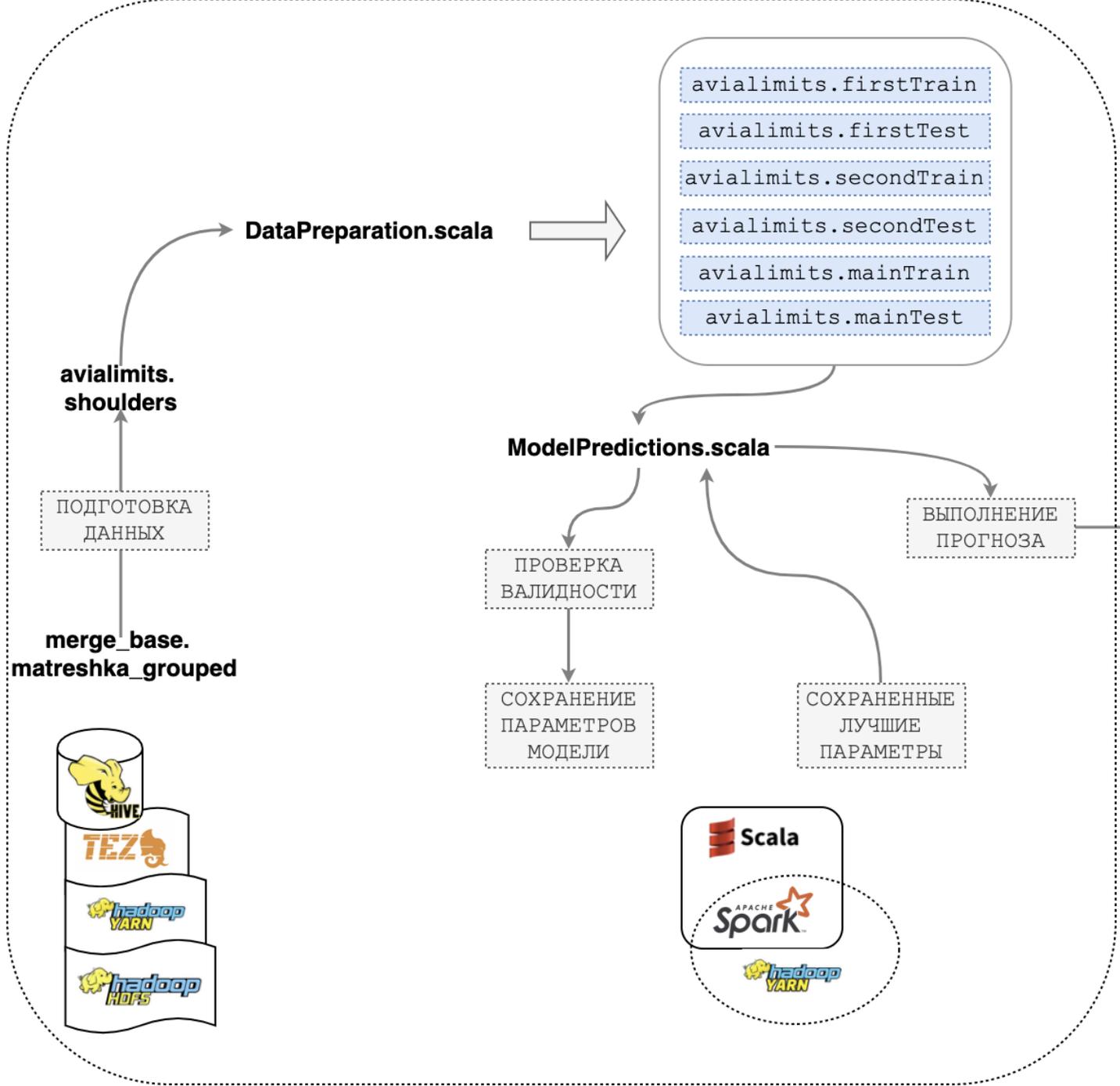
```
# make predictions
trainPredict = model.predict(trainX)
testPredict = model.predict(testX)
```

INPUT LAYER

HIDDEN LAYER

OUTPUT LAYER

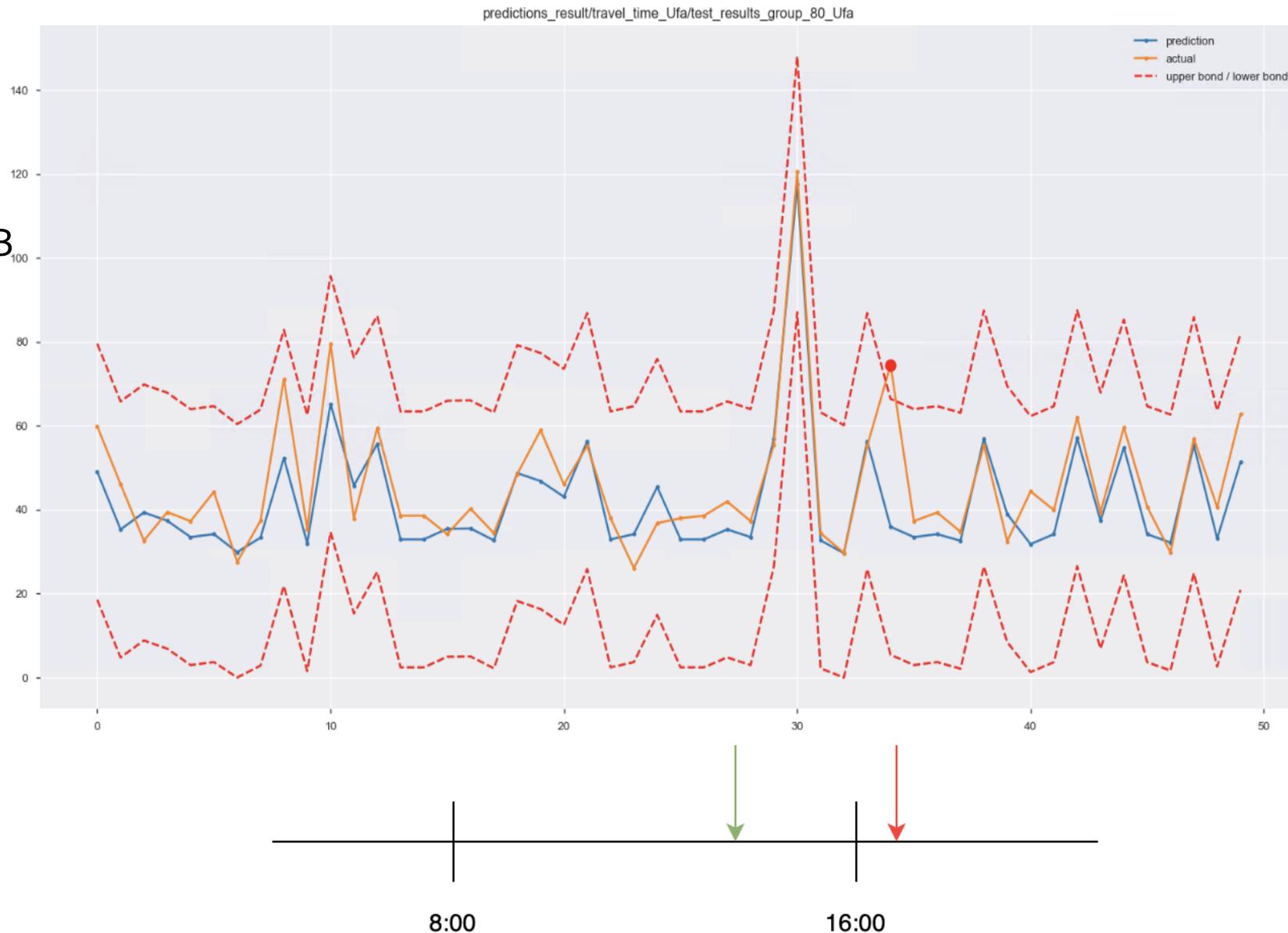




АРХИТЕКТУРА РЕШЕНИЯ

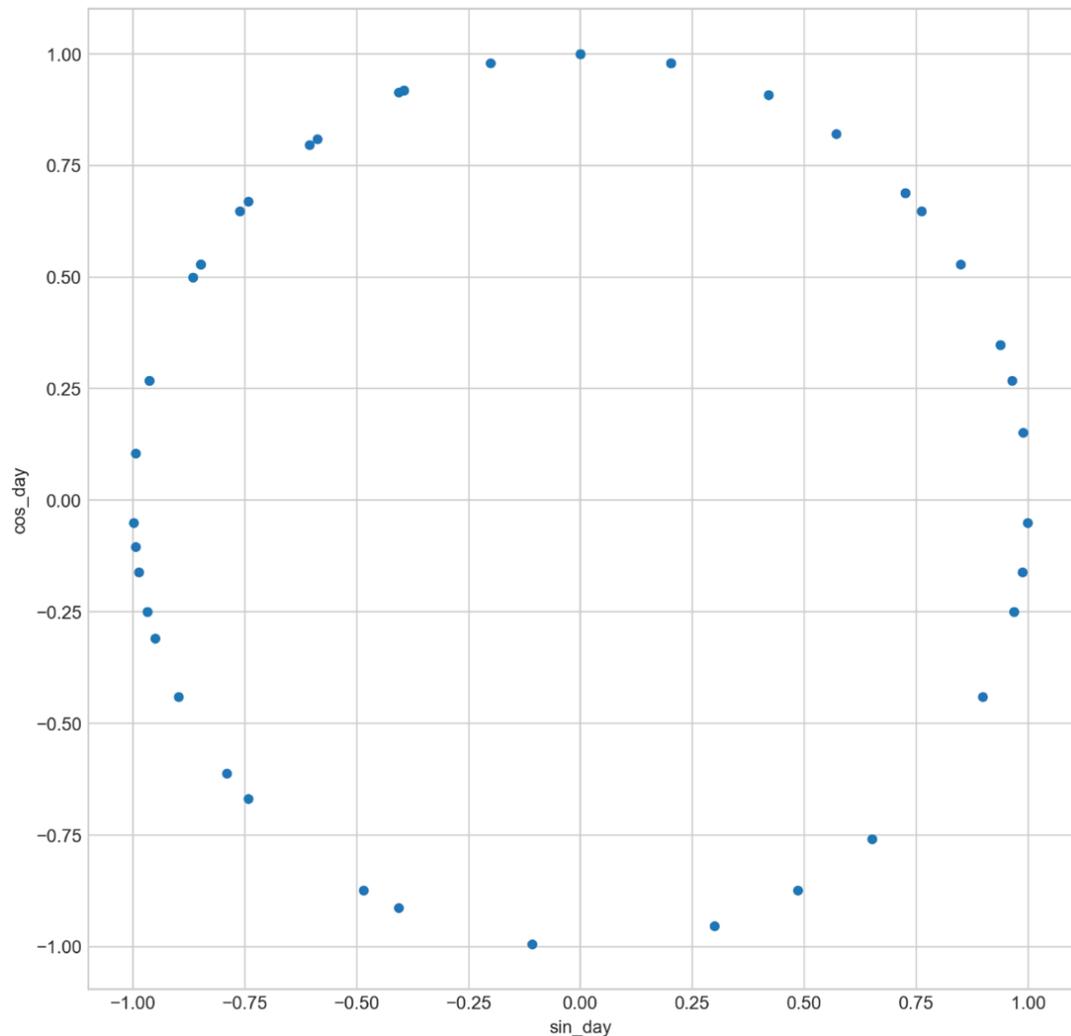
ВОЗНИКШИЕ СЛОЖНОСТИ

- Предсказание времени в пути
- Слишком большое количество признаков
- Погодные условия
- Разделение международной и внутренней почты
- Проблемы с обработкой данных в HDFS



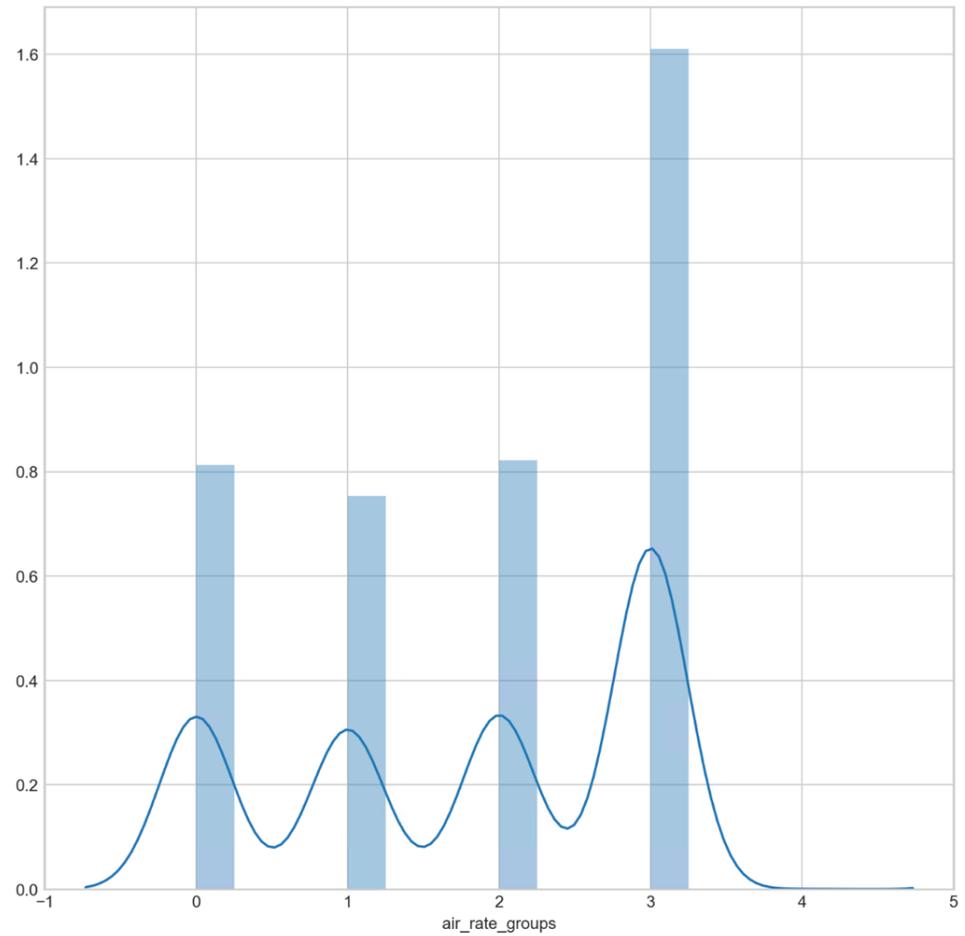
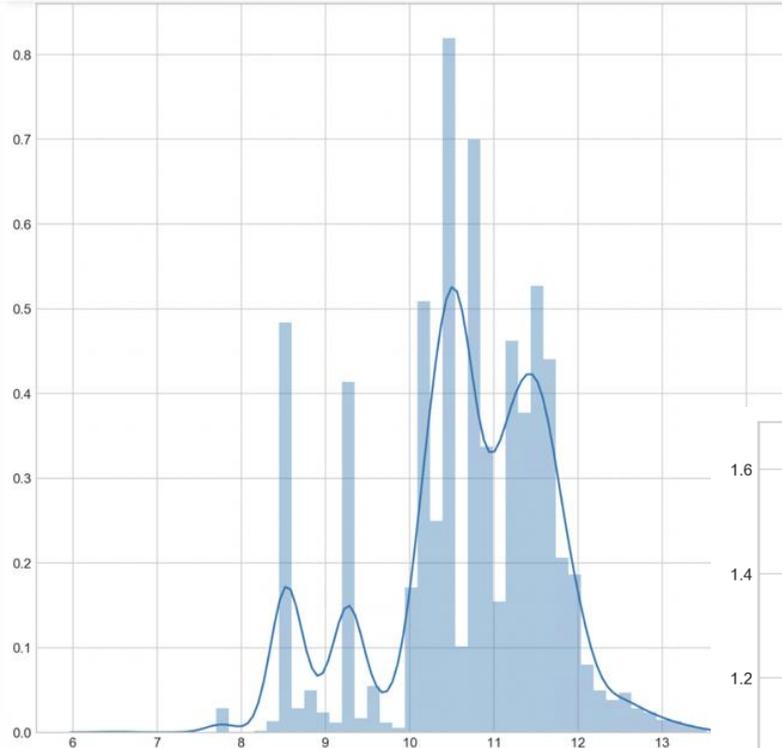
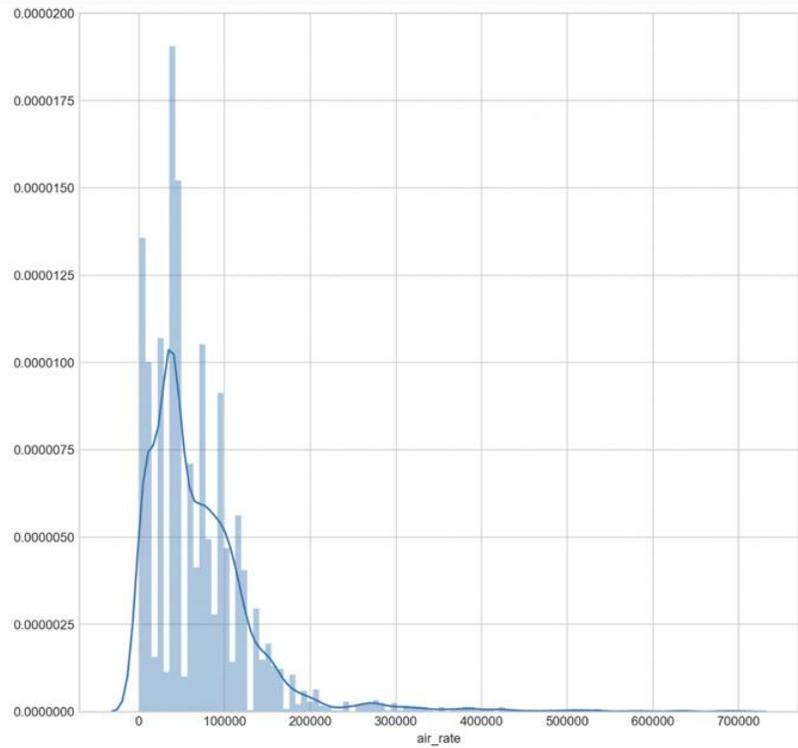


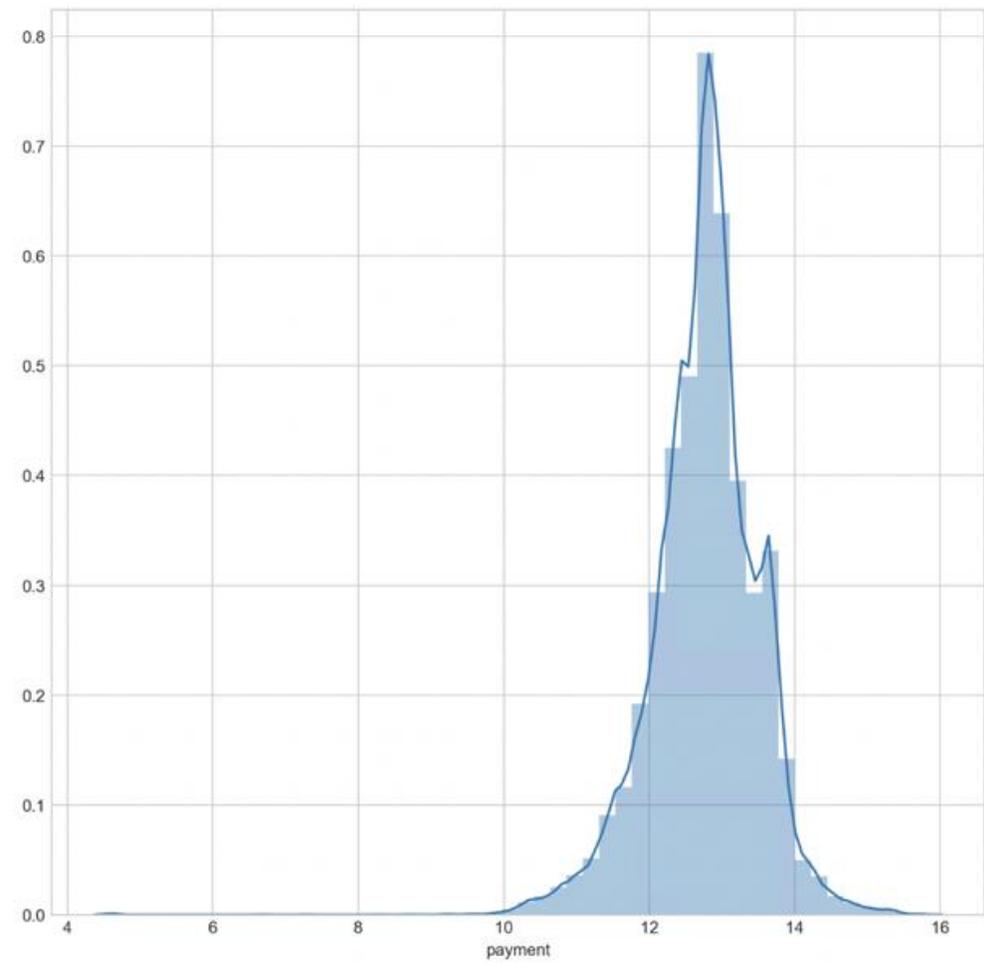
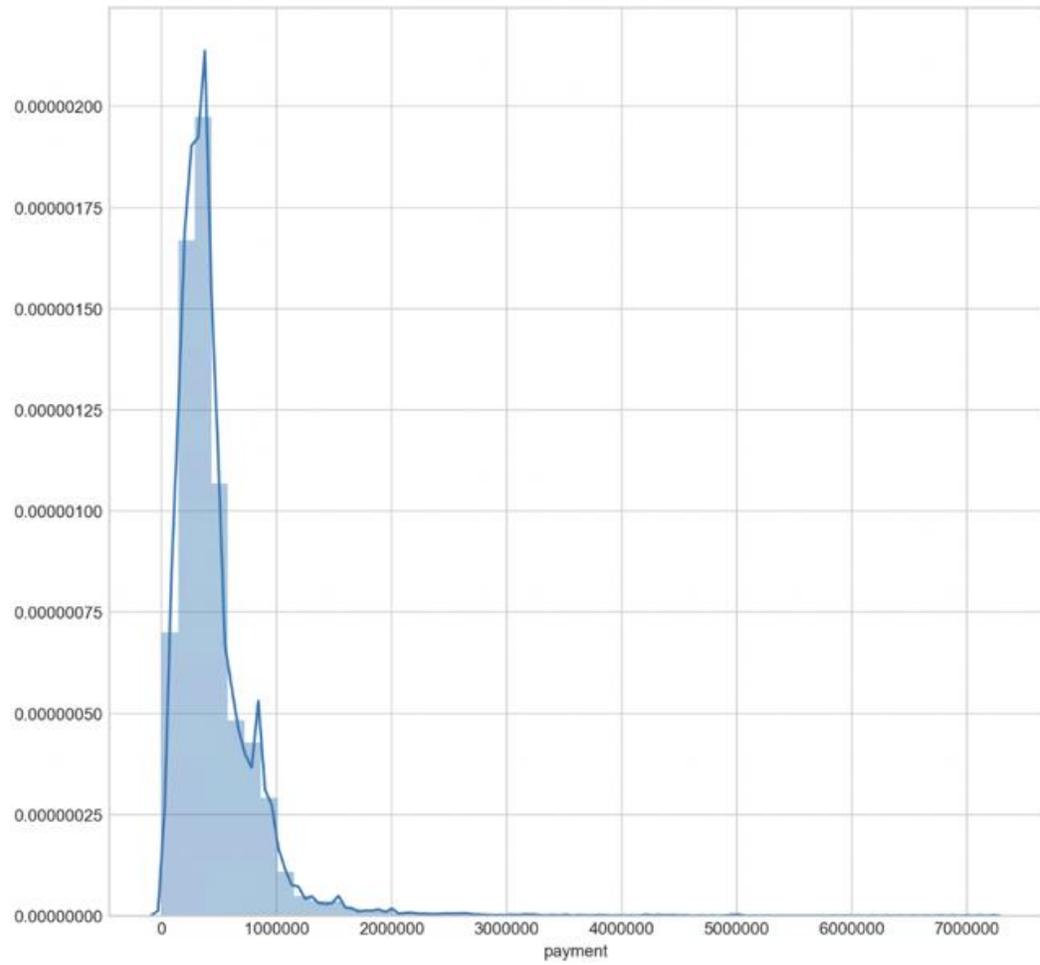
ГЕНЕРИРОВАНИЕ ПРИЗНАКОВ



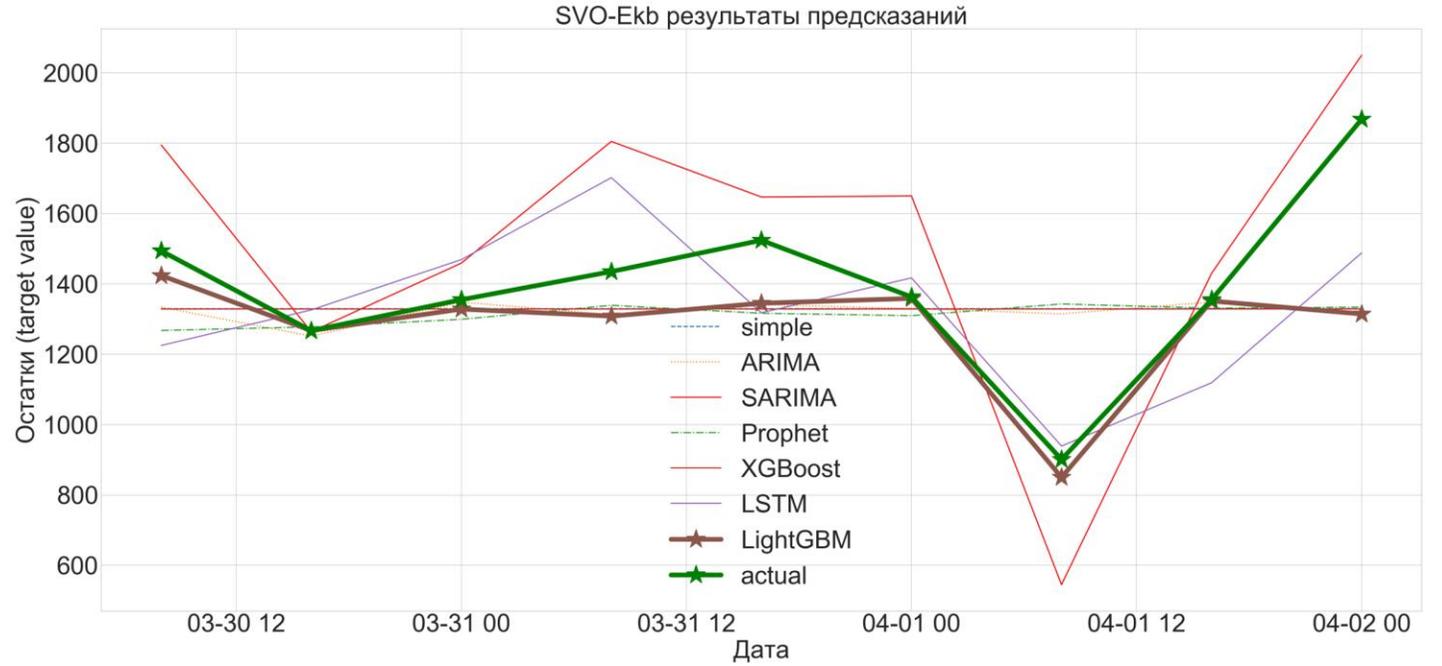
$$val_sin = \sin\left(\frac{val \cdot 2 \cdot \pi}{period}\right)$$

$$val_cos = \cos\left(\frac{val \cdot 2 \cdot \pi}{period}\right)$$

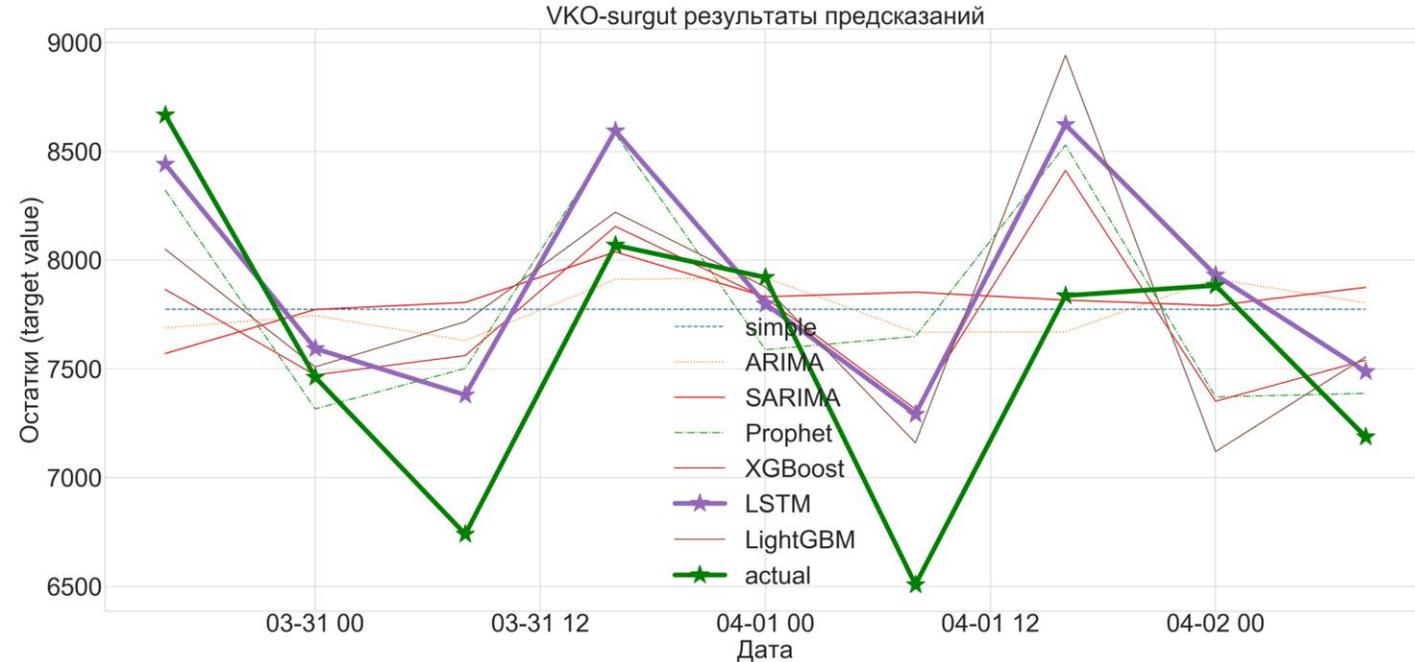




Плечо Шереметьево АОПП Екатеринбург АОПП



Плечо Внуково АОПП — Сургут АОПП



РЕЗУЛЬТАТЫ (Mean Absolut Error (MAE))

Плечо \ Модель	ARIMA	SARIMA	XGBoost	LightGBM	Prophet	LSTM	Mean target value
SVO – Владивосток (Б)	1435	1852	3120	1423	2096	1916	16871
SVO – Иркутск (С)	560	728	601	558	577	466	4704
SVO – Екб (М)	166	165	176	131	183	160	1329
DME – Новосибирск (Б)	1932	2259	2559	1930	2321	2541	38616
DME – Челябинск (С)	589	501	561	344	473	763	2255
DME – Красноярск (М)	791	596	747	859	1178	890	15367
VKO – Сургут (С)	476	526	454	525	516	396	7774
VKO – Казань (М)	235	260	477	396	532	353	3095



ЗАКЛЮЧЕНИЕ

В работе «разработка модели прогнозов загрузки узлов логистической сети»:

- проведен анализ существующих подходов к прогнозированию временных рядов
- выполнена очистка и предобработка данных
- модели протестированы и отобрана лучшая
- выбранная модель интегрирована в производственный процесс



Публикация тезисов:

- Чебенева Е.А. «Разработка модели прогнозов загрузки узлов логистической сети» XLV Международная молодёжная научная конференция «Гагаринские чтения – 2019»