

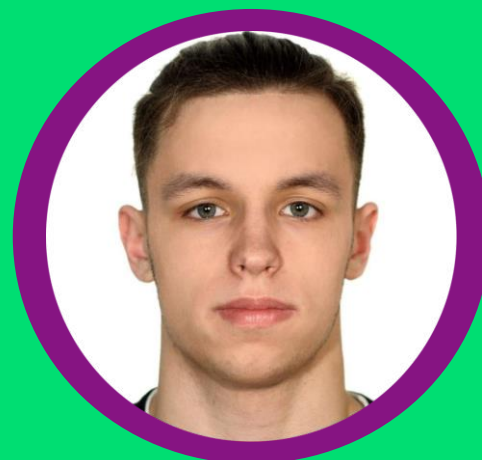
Предсказание длительности перебоев электроснабжения на базовых станциях



Камаев Богдан



Мышлянов Алексей



Куш Артём



Синяев Максим

Построить модель, предсказывающую длительность перерывов электроснабжения базовых станций

Обработка данных

- Кластеризация по геоданным на 50 кластеров
- Генерация значимых предикатов: group_id, hour_start

Модель

- Сверточная-рекуррентная нейронная сеть с показателем по MAPE ~67,5%
- Использование батч-нормализации и сверточных слоев

Финансовый доход

- Выручка от внедрения технологии составит 309 млн. руб.
- Внедрение окупится за 1 месяц
- Дальнейшая автоматизация процесса принятия решений

Решение данной задачи разделено на 3 основных этапа

Обработка данных

- Метод кластеризации
- Тип кластеризации
- Генерация дополнительных признаков
- Поиск наиболее важных признаков

Построение модели

- Обучение одиночных моделей
- Обучение ансамблей моделей
- Обучение рекуррентных нейронных сетей
- Обучение сверточных нейронных сетей

Применение решения

- Расчет стоимости внедрения
- Анализ рисков
- Расчет доходов

Оптимальная кластеризация станций является наиболее важным этапом предобработки данных

Методы кластеризации:

Плюсы

Минусы

По геолокации

Географическая
связь
отключений

Не учитывает
время

По времени

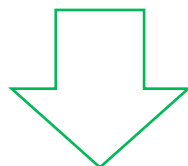
Связь временных
признаков

Масштабируемость

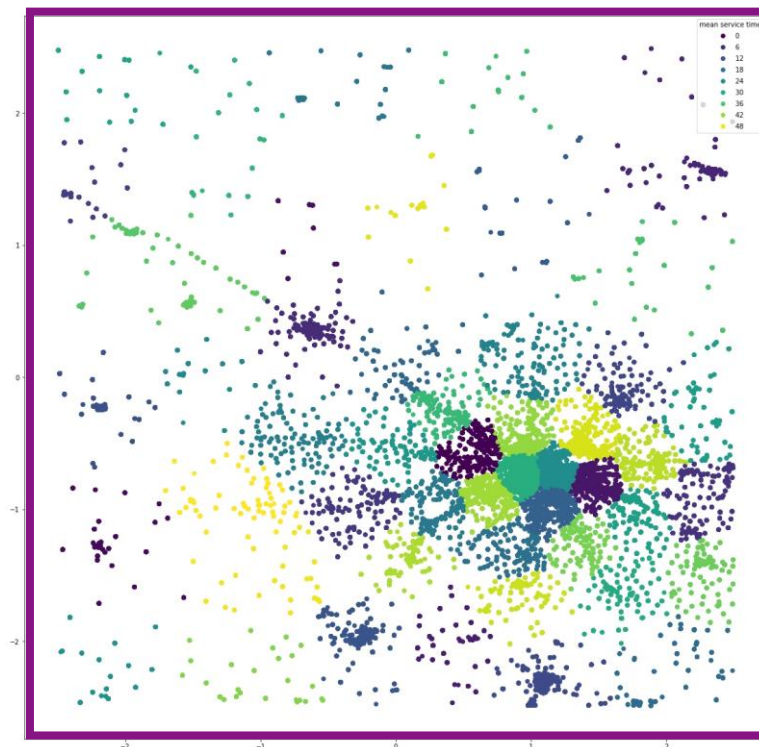
Комплексная

Учитывает оба
фактора

Разная природа
данных



Для кластеризации по геолокации была выбрана модель KMeans на 50 кластеров, при этом временные признаки учитываются в виде сгенерированных показателей.



Наиболее распространенные кластеры

Рекуррентная нейронная сеть со сверточными слоями является наиболее точной – MAPE 67,5%

Критерии сравнения методов

Метод	Интерпретируемость	Скорость обучения	Точность
Одиночная модель	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★
Ансамбль моделей	★ ★ ★ ★	★ ★ ★ ★	★
Рекуррентная нейронная сеть	★ ★	★	★ ★ ★
Сверточно-рекуррентная нейронная сеть	★ ★	★ ★	★ ★ ★ ★ ★

Плюсом в такой модели является то, что она универсальна на этапе обучения для векторов различной длины, также сверточные слои увеличивают скорость обучения.

Предлагаемый алгоритм решает ряд финансовых проблем, которые видит заказчик

Проблемы	Аренда автомобильных дизельных генераторов	Поддержка аккумуляторов	Использование автономных мобильных базовых станций	Устранение последствий сбоя электропитания	Недовольство клиентов в зоне действия базовой станции
Эффекты	Уменьшение количества вызовов	Увеличение работоспособности аккумулятора	Отказ от использования	Увеличение эффективности обслуживающего персонала	Уменьшение нерабочего времени станции
	7 т.р. С каждого вызова	Увеличение срока службы	Уменьшение расходов оборудования	Ускорение восстановления потерь	Повышение лояльности клиентов

Данная технология принесет 309 млн. рублей

Внедрение

500 тыс.

Оборудование

4,8 млн.

Годовая зарплата
4 сотрудникам

5 млн.

Непредвиденные
ситуации

10,3 млн.

Риски

490 млн.

Ложный вызов

2,4 млн.

Поддержка модели
при ее деградации

1 млн.

Усложнение модели
при ошибках GPS

493,4 млн.

Доход

800 млн.

Снижение количества
Ложных вызовов

8,9 млн.

Предотвращение износа
батарей

3 млн.

Выгодная транспортировка и
использование топлива

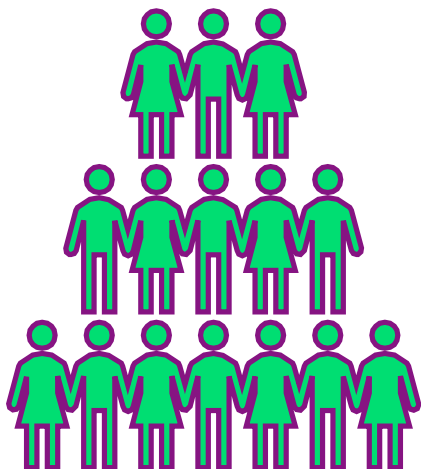
812,9 млн.

В дальнейшем предложенная технология поможет заменить менеджеров решающих вопрос о дальнейших действиях, связанных с перебоем электроэнергии

Через год после внедрения:



Алгоритм принесет 121 млн. руб. при сокращении штата менеджеров в 4 раза



Автоматизированный алгоритм принятия решения позволит:



Сократить штат менеджеров



Уменьшить время принятия решений



Повысить вариативность целей алгоритма



Прозрачность принятия решений

Спасибо за внимание!

Команда «The Boys»



Камаев Богдан

Data Scientist (ML,DL)

Москва, МГУ

kamae98@gmail.com

+7(963) 670-16-88



Мышлянов Алексей

Business Analytic

Москва, МИСИС

ananasnya@gmail.com

+7(999) 132-02-16



Куш Артем

Data Scientist (ML,DL)

Москва, МИСИС

artemkush1@gmail.com

+7(903) 246-93-31



Синяев Максим

Team Lead/Data Engineer

Москва, НГТУ

sinyaevmaxim@gmail.com

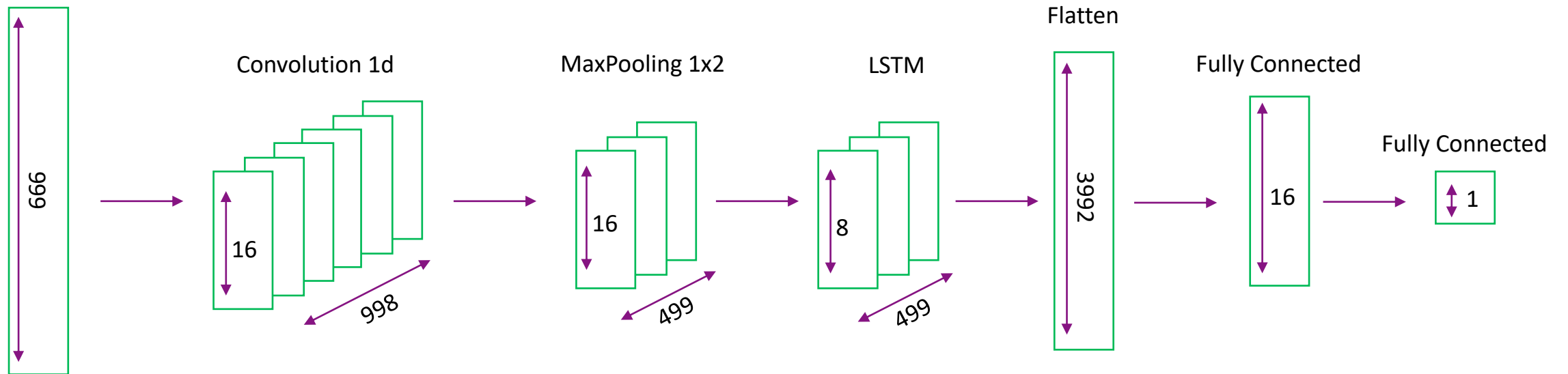
+7(913) 914-63-94

Описание параметров

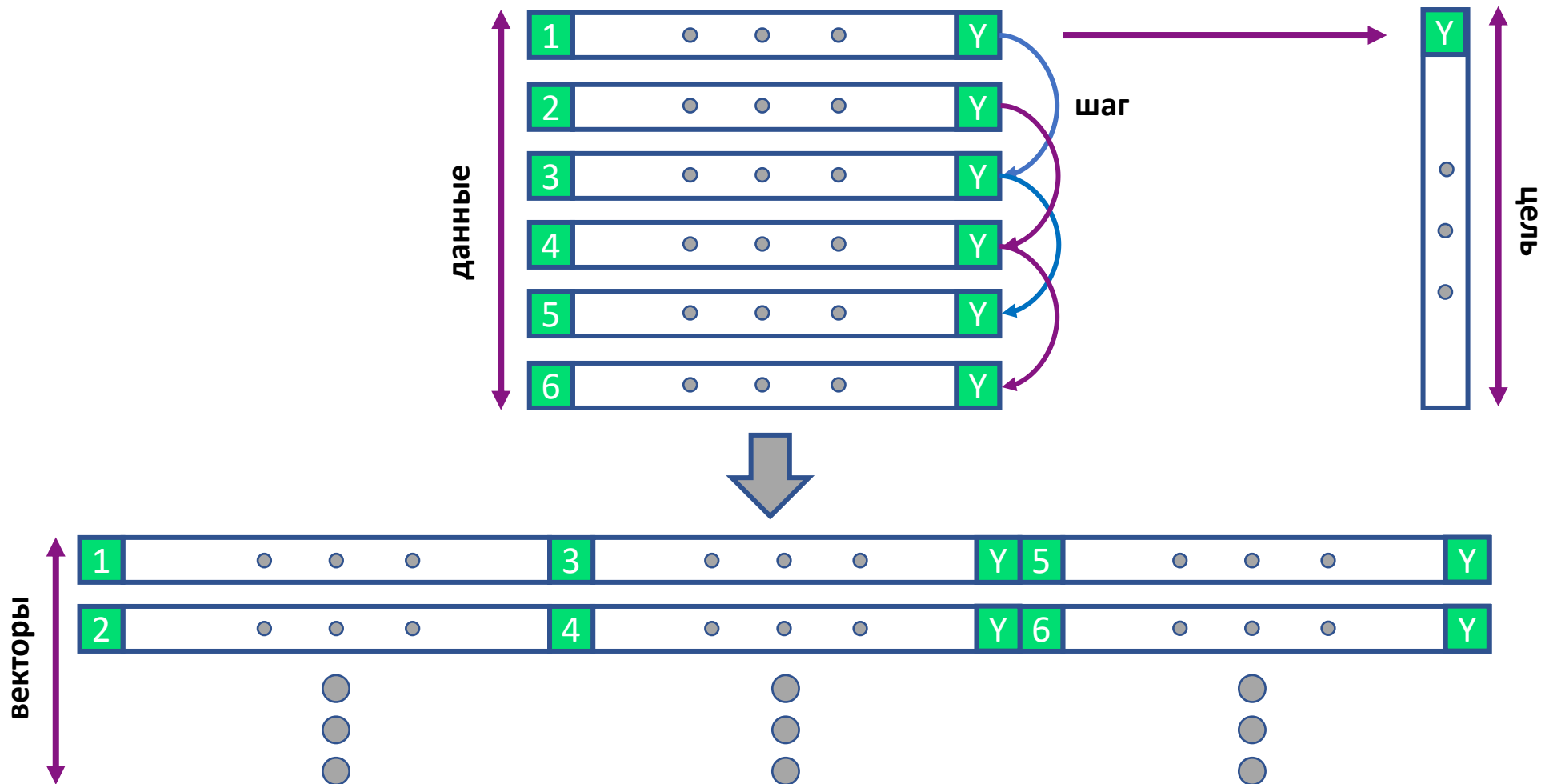
Параметр	Описание
service_time_minutes	Целевая переменная
time_start	Время начала инцидента
time_end	Время конца инцидента
place_latitude	Широта
place_longitude	Долгота
season	Время года
night_time	Метка происшествия в ночное время
hour_start	Час начала происшествия
holiday	Праздничный день или выходной
fails_count	Количество отключений в этот день
cluster_id	Принадлежность к кластеру
group_id	Объединение по группам станций, отключившихся в одно и тоже время



Архитектура нейронной сети



Преобразование данных в вектор для модели RNN



Расчет стоимостей

Константные величины	Значение
Социальные выплаты сотрудникам, руб	961.7 млн.
Расходы на обучение, руб	143.4 млн
Средняя заработная плата в РФ, руб	47.6 тыс.
Штат сотрудников	39.1 тыс.
Аренда генератора, руб/сутки	5000
Средняя заработная плата в РФ в сутки, руб	1588.56
Стоимость батареи, руб	30 тыс.
Стоимость батарей на 1 базовой станции, руб	300 тыс.
Срок службы батареи, лет	10

Расчетные величины	Значение
Итоговые затраты на команду из двух человек (подрядчики), руб	15 тыс.
Средние затраты на одного сотрудника в год, руб	600.12 тыс.
Стоимость батарей на 50000 базовых станциях, руб	15 млрд.
Потери компании от порчи батареи, р/день	168 тыс.
Возможное улучшение аккумулятора на 30%, при условии, что за день может сломаться 2%, р/час	1008.04
Уменьшение расходов на износ батарей за день, р	24192.03
Потери при ложных вызовах в год	2.49 млрд.
Доход при устранении ложных вызовов	4.073 млрд.



Расчет вероятности ошибки

Предсказание	Нижний интервал	Верхний интервал	Надо, не отпр	Не надо, отпр
1	0,35	1,65		
1,25	0,4375	2,0625		
1,5	0,525	2,475		
1,75	0,6125	2,8875		
2	0,7	3,3		
2,25	0,7875	3,7125		
2,5	0,875	4,125	0,05	
2,75	0,9625	4,5375	0,195454545	
3	1,05	4,95	0,316666667	
3,25	1,1375	5,3625	0,419230769	
3,5	1,225	5,775	0,507142857	
3,75	1,3125	6,1875	0,583333333	
4	1,4	6,6	0,65	0,65
4,25	1,4875	7,0125		0,591176
4,75	1,6625	7,8375		0,492105
5	1,75	8,25		0,45
5,25	1,8375	8,6625		0,411905
5,5	1,925	9,075		0,377273
5,75	2,0125	9,4875		0,345652
6	2,1	9,9		0,316667
6,25	2,1875	10,3125		0,29
6,5	2,275	10,725		0,265385
6,75	2,3625	11,1375		0,242593
7	2,45	11,55		0,221429

Предсказание	Нижний интервал	Верхний интервал	Надо, не отпр	Не надо, отпр
7,25	2,5375	11,9625		0,201724
7,5	2,625	12,375		0,183333
7,75	2,7125	12,7875		0,166129
8	2,8	13,2		0,15
8,25	2,8875	13,6125		0,134848
8,5	2,975	14,025		0,120588
8,75	3,0625	14,4375		0,107143
9	3,15	14,85		0,094444
9,25	3,2375	15,2625		0,082432
9,5	3,325	15,675		0,071053
9,75	3,4125	16,0875		0,060256
10	3,5	16,5		0,05
10,25	3,5875	16,9125		0,040244
10,5	3,675	17,325		0,030952
10,75	3,7625	17,7375		0,022093
11	3,85	18,15		0,013636
11,25	3,9375	18,5625		0,005556
11,5	4,025	18,975		
11,75	4,1125	19,3875		
12	4,2	19,8		
12,25	4,2875	20,2125		
Описание	Вероятность			
P(FP)	0,38			
P(FN)	0,224			