Анализ предметной области.

На ежедневной основе на территории Волгоградской области

возникают непредвиденные происшествий (событий) в различных

социально-экономических сферах жизнедеятельности. В связи с чем,

возникает необходимость фиксации очагов этих событий в разрезе

муниципальных образований с целью оперативного реагирования и

выработки комплексного плана их предотвращения. За последние несколько

лет скопились исторические данные, используя которые можно

проанализировать взаимосвязи происхождения происшествий, а также

прогнозировать будущие.

Анализ данных, отслеживание взаимосвязей и формирование

управленческих выводов необходимо региону, так как ежегодно в регионе

возрастает количество фиксируемых событий.

Проблемы с которыми столкнулся регион в первичной работе с

данными:

- около 2,5 млн строк неструктурированных данных без

сортировки по муниципальным образованиям

- отсутствует понимание очагов локализации происшествий

- отсутствует концепция работы с данными (сбор данных, анализ

и дальнейшее принятие управленческих решений)

Цель задачи — разработка концепции проведения анализа

возникающих полифакторных ЧС, в том числе, в качестве требований по

решению задачи, должны быть произведены:

1. Анализ данных (обработка неструктурированных данных, с целью их

представления в структурированном виде) - анализ имеющегося массива

данных по свершившимся происшествиям на территории Волгоградской

области, в том числе в разрезе муниципальных образований;

2. Построение тепловой карты направлений (категорий)

происшествий (Типы направлений чрезвычайных ситуаций: ЧС на

транспорте, ЧС в сфере ЖКХ, ЧС в сфере ТЭК и Промышленность, ЧС в

сфере здравоохранения, ЧС, связанные с противоправными действиями);

3. Определение:

- зависимостей событий в привязке ко времени и координатам через

слои нескольких типов направлений, выявление областей пересечения

категорий происшествий

(пример: на перекрестке часто встречается смертность от ДТП, при

этом пострадавшие - дети, по координатам напротив школы магазин, при

этом перекресток без светофора. В качестве решения можно предложить

установку светофора, снижение скоростного режима, установку зеркала

для слепых зон)

- точки локализации происшествий в привязке к местности

- распределение событий внутри каждого из 5 направлений (способ

представления данных на усмотрение участника)

4. Составление прогноза вероятности возникновения событий (не

менее 5 событий по каждому направлению):

- гипотезы возникновения событий на основе анализа

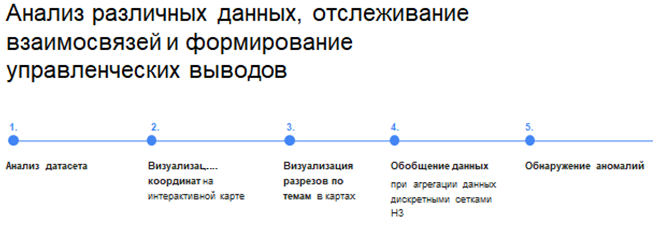
предоставленных данных

- расчет вероятности возникновения происшествия, расчет ущерба,

оценка риска

- управленческие выводы по результатам анализа данных, с

описанием логики либо алгоритма либо механизма принятия решения

Формирование и подготовка данных для анализа.

В качестве исходных данных были даны 6 файлов формата .csv, хранящие в себе 200.000 - 500.000 строк с информацией о записи ЧС, а именно:

* Тип ЧС
* Стадия ЧС
* Время возникновения ЧС
* Уровень ЧС
* Район области
* Район города
* Блокировка
* Рядом ли ЧС
* Широта+долгота
* Статус запроса
* Перечень экстренных служб
* Количество расчётов
* Наименование расчётов
* Опасность
* Прочие, не несущие информации данные

При анализе данных было выявлено, что, например, из всех 38 столбцов в файле 01-09.2019.csv полезных для дальнейшей работы, всего 11.

Таким образом, были выявлены следующие параметры по данным:

* Количество полифакторных ЧС ~ 2,5 млн.
* Количество характеристик ЧС - 16
* Период, в который были собраны данные - 01.01.2019 - 31.12.2021

Определение типа решаемой задачи.

Прогнозированный тип решаемой задачи.

Выбор алгоритма анализа.

Несмотря на то, что имеется больше информации, желательно выбрать несколько соответствующих характеристик, которые оказывают более сильное влияние на аварию и ее тяжесть. Это может снизить потребность в вычислениях и повысить точность прогнозов. В основном выбирались время, место и погодные условия.

Был выбран алгоритм определения ограничений и требований к алгоритму по типу исходных данных.

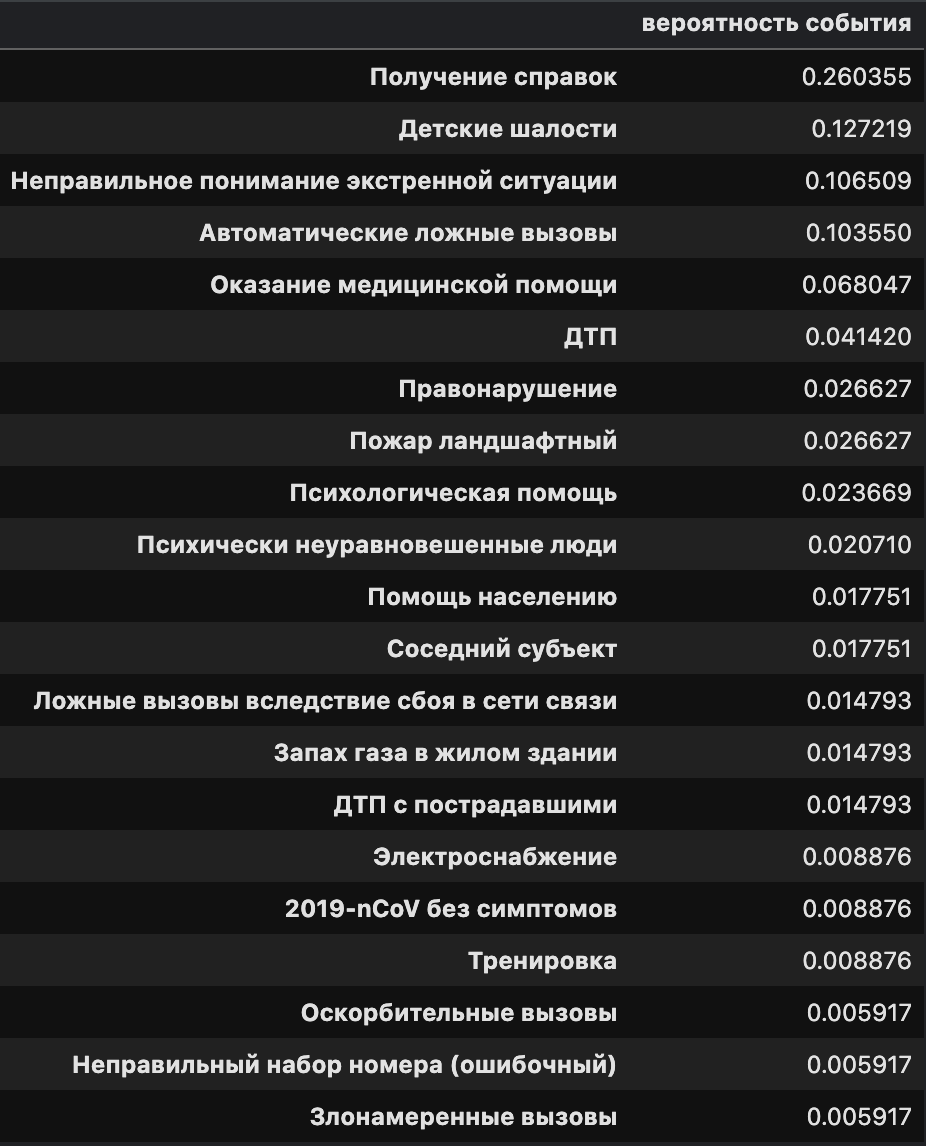
Сам “Data mining”.

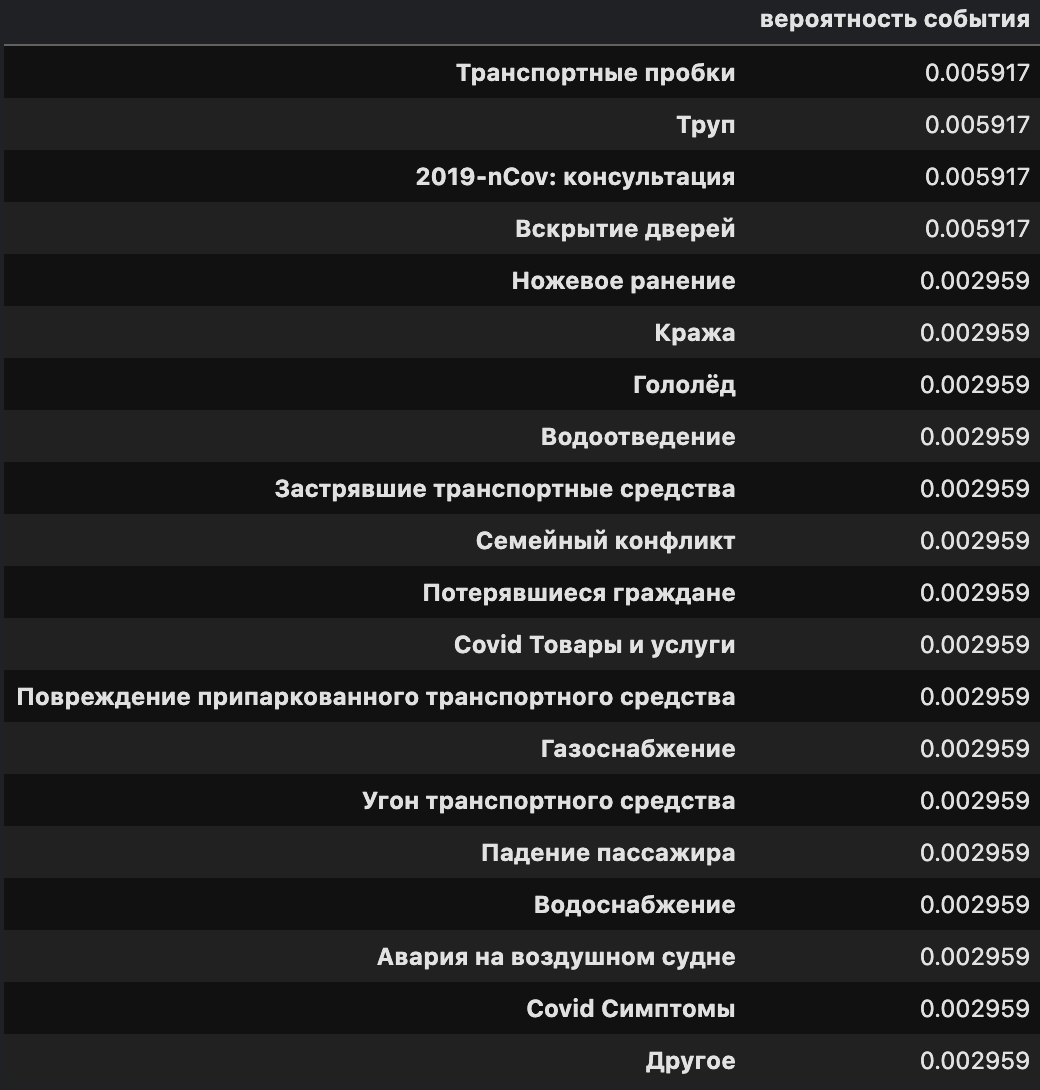
Методы Data Mining помогают решить многие задачи, с которыми сталкивается аналитик. Из них основными являются: задача классификации и регрессии, задача поиска ассоциативных правил и задача кластеризации.

Проверка моделей и представление результатов анализа.

Прогнозирование направлено на определение параметров динамики конкретного объекта или процесса на основе ретроспективных данных, т. е. анализа его состояния в прошлом.

Прогноз вероятности возникновения событий





Управленческие выводы по результатам анализа данных

Перевести часть сбора и структурирования информации по чрезвычайным ситуациям на телефонного робота.

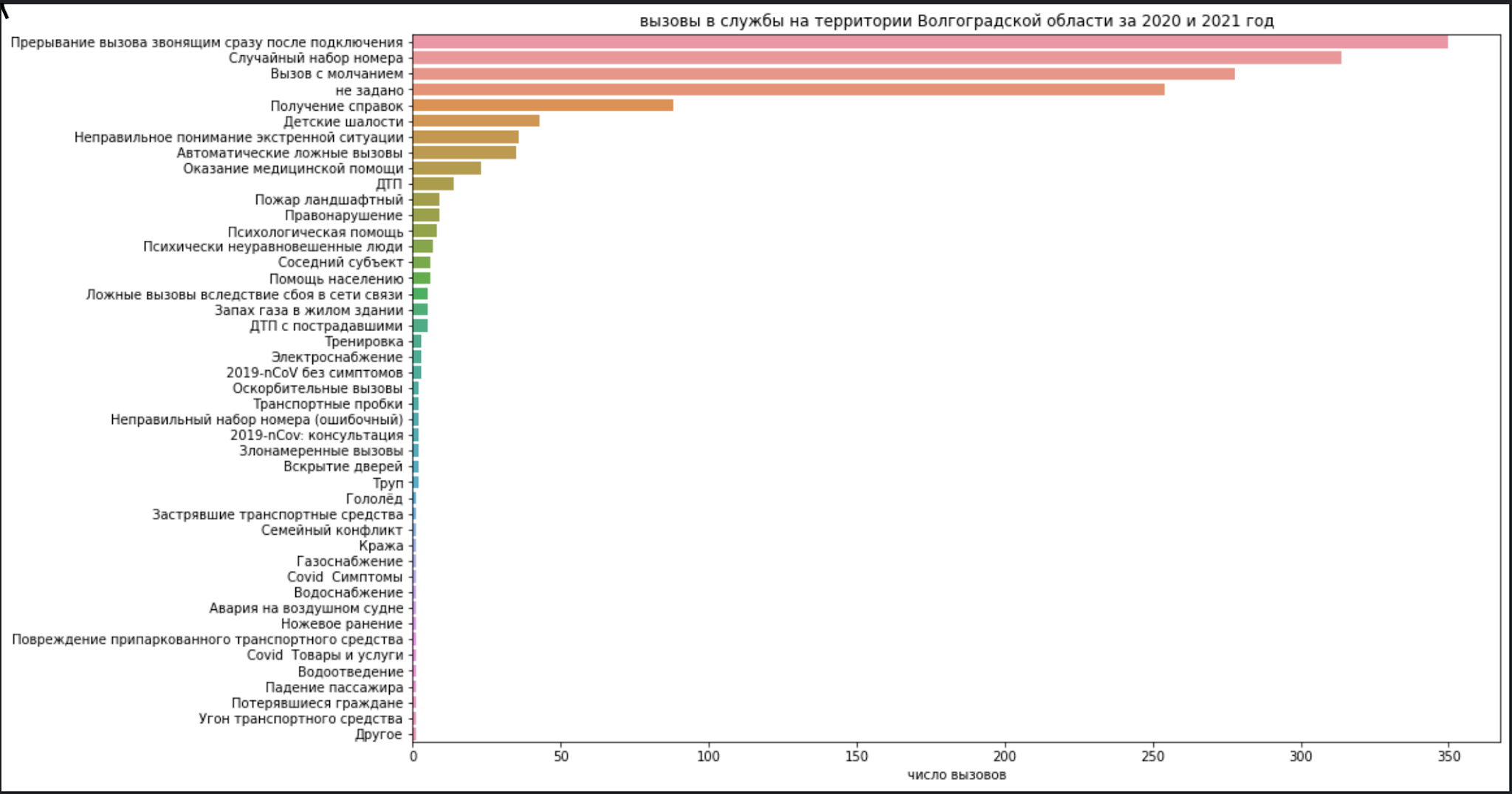
Для уменьшения загруженности call-центра ложными звонками можно перейти

на телефонного робота, который будет принимать вызовы, и в случае опасной ситуации для звонящего, будет переводить вызов на оператора call-центра.

С помощью теории вероятности можно просчитать как вероятность наступления

Применение построенных моделей.

Анализ имеющегося массива данных по свершившимся происшествиям на территории Волгоградской области, в том числе в разрезе муниципальных образований;



Происшествия в Ворошиловском районе

