Вопросы к экзамену по теме «Предиктивная аналитика в ИБ»

|  |  |
| --- | --- |
| № | Вопрос |
| 1 | Выберите тип аналитики, который наиболее точно решает следующую задачу «В компании, занимающейся аналитикой и предоставляющей облачный сервис аналитики данных случилась неприятность: хранилище данных оказалось переполнено и по этой причине сервис был недоступен около 4 часов, что привело к оттоку пользователей. Руководство компании поставило задачу, что о скором переполнении хранилища оно должно узнавать минимум за неделю»:   1. Описательная аналитика 2. Диагностическая аналитика 3. **Предиктивная аналитика** 4. Предписывающая аналитика |
| 2 | Насколько важным является определение задачи, проблемы или бизнес-цели в жизненном цикле аналитики данных:   1. **Очень важно практически на любом этапе жизненного цикла** 2. Важно только на этапе выбора метода аналитики 3. Важно, но не принципиально, сам процесс аналитики даст понимание проблемы 4. Не важно, так как аналитика является самодостаточным инструментом |
| 3 | Что **НЕ** является источником данных для аналитики:   1. База данных транзакций пользователей 2. Показания загрузки ЦПУ сервера 3. Веб-сервис, предоставляющий данные о температуре уличного воздуха 4. Сетевой коммутатор |
| 4 | Вам поставлена задача отобразить динамику изменения количества пользователей вашей системы за последний год, чтобы понять в какой момент произошел основной рост, визуализация будет продемонстрирована топ-менеджеру, у вас есть данные о количестве пользователей за каждый день в течение года. Какой тип визуализации вы выберете:   1. Таблицу с отображением количества пользователей за каждый день года, так будет понятен объем обработанных данных, возможно на основании объема руководитель поймет мою ценность 2. Круговую диаграмму, где секторы – это количество пользователей в месяц, так будет понятен общий объем пользователей 3. **Линейный график количества пользователей за год, для наглядности добавлю линию основного тренда и линию тренда для периода бурного роста** 4. Гистограмму количества пользователей по месяцам, для наглядности добавлю показания по дисперсии и математическому ожиданию количества пользователей за год |
| 5 | Выберите наиболее правильную расшифровку аббревиатуры «RCA»:   1. Причинно-следственный метод оценки зависимостей 2. **Анализ корневых причин** 3. Анализ корреляционного взаимодействия 4. Причинный анализ |
| 6 | Могут ли являться логи журнала событий Windows временным рядом?   1. Да, как и любой тип данных 2. **Да, но понадобится преобразование логов с учетом временных меток событий** 3. Нет, логи – это текстовые данные, и они не могут быть преобразованы во временной ряд 4. Нет, отсчеты во временном ряде обязаны идти с равным интервалом, а логи в журнал записываются без какой-либо периодичности |
| 7 | Выберите основные характеристики временного ряда (несколько):   1. Амплитуда 2. **Тренд** 3. **Цикличность** 4. Количество отсчетов |
| 8 | Какой метод (фреймворк, библиотека) **НЕ** применяется для прогнозирования временных рядов:   1. Prophet 2. ARIMA 3. Рекуррентные нейронные сети 4. **Байесовские сети** |
| 9 | Что означает коэффициент *p* в модели ARIMA:  :   1. Коэффициент авторегрессии 2. **Порядок авторегрессии** 3. Порядок разности временного ряда 4. Количество отсчетов в анализируемом временном ряду |
| 10 | Чем отличается архитектура сети LSTM от других нейронных сетей и почему их используют для прогнозирования:   1. Каждая ячейка сети LSTM обладает уникальной структурой, которая «подстраивается» под временной ряд. 2. **Каждая ячейка сети LSTM имеет встроенные механизмы, которые реализуют долгосрочную зависимость и исключают проблему градиентного взрыва и обнуления градиента**. 3. Сеть LSTM в отличие от других нейронных сетей может работать с многомерными временными рядами. 4. Для прогнозирования можно использовать любые разновидности рекуррентных сетей, поскольку они сохраняют всю информацию, когда-либо поступившую в сеть. |
| 11 | Какого метода выявления аномалий **НЕ** существует:   1. Proximity-based 2. Reconstruction-based 3. Prediction-based 4. **Stochastic-based** |
| 12 | Алгоритм LOF является разновидностью методов:   1. Классификации 2. **Кластеризации** 3. Регрессии 4. Локализации |
| 13 | Отличительной особенностью автоэнкодеров является:   1. **Одинаковое количество нейронов на входном и выходном слое** 2. Наличие скрытого слоя 3. Использование метода обратного распространения ошибки 4. Высокая точность прогнозирования |
| 14 | Возможно ли использовать метод LOF для решения задачи RCA с учителем:   1. Нет, поскольку LOF решает задачу кластеризации, являющейся задачей обучения без учителя 2. Да, поскольку LOF решает задачу классификации, являющейся задачей обучения с учителем 3. Да, LOF может применяться непосредственно на этапе поиска корневых причин 4. **Да, LOF может применяться на этапе выявления аномалий** |
| 15 | Почему SOM – это «самоорганизующаяся» карта:   1. Поскольку инициализация весов происходит произвольно и не влияет на качество обучения 2. **Поскольку нейроны SOM организуют кластеры вокруг векторов входного пространства** 3. Поскольку веса нейронов SOM вычисляются в строго организованном порядке, который выбирается на этапе инициализации 4. Поскольку это алгоритм кластеризации, т.е. обучение без учителя, т.е. самообучение |