**Опишите стандарты жизненного цикла Big Data.**

Общепризнанный стандарт по ЖЦ пока отсутствует

Известные методологии, являющиеся мировыми

стандартами: CRISP-DM, Two Crows, SEMMA, а также для

аналитики данных можно применить свои собственные

методологии.

• Стандарт CRISP-DM включает:

1 Средства предварительной обработки информации на местах

её хранения

2 Возможность запуска алгоритмов анализа прямо в этих местах

Изображение выглядит как диаграмма

Автоматически созданное описание

Этапы построения модели анализа данных

•Формулирование и проверка гипотез

•Разработка численных критериев качества

решений

•Просчет эффективности: время/качество

**Что называется когнитивным анализом данных?**

Когнитивный анализ данных направлен на изучение методов, с помощью которых человек извлекает закономерности из информации об окружающем мире и затем использует их для получения новых знаний, для более эффективного принятия решений. Эти исследования служат основой для построения компьютерных систем анализа данных, которые, имитируя простейшие когнитивные способности человека, позволяют решать задачи больших объемов, извлекать закономерности тогда, когда человеческих ресурсов на это не хватает.

**Назовите этапы интеллектуального анализа данных.**

Интеллектуальный анализ включает в себя следующие этапы: понимание и формулировка задачи анализа, подготовка данных для автоматизированного анализа, применение методов Data Mining и построение моделей, проверка построенных моделей, интерпретация моделей человеком.

Перед применением методов Data Mining исходные данные должны быть преобразованы.

Вид преобразований зависит от применяемых методов.

**30. Что такое статистическое обучение?**

Статистическое обучение - это подраздел машинного обучения, который использует статистические методы для изучения данных. Он используется для создания моделей и прогнозирования результатов на основе данных.

Статистическое обучение может быть разделено на две категории: контролируемое и неконтролируемое. В контролируемом обучении учитель предоставляет модели данные и правильные ответы, чтобы модель могла научиться предсказывать правильный ответ. В неконтролируемом обучении учитель не предоставляет модели правильные ответы, а модель должна сама найти закономерности в данных.

Изображение выглядит как текст

Автоматически созданное описание

**31.В чем разница между описательными и предсказательными задачами DM? Какие методы анализа лучше приспособлены для описательных, а какие для предсказательных задач?**

Решение предсказательных (predictive) задач разбивается на два этапа: 1) На первом этапе на основании набора данных с известными результатами строится модель.

2) На втором этапе она используется для предсказания результатов на

основании новых наборов данных. К данному виду задач относят задачи классификации и регрессии.

Описательные (descriptive) задачи (статистический вывод ) уделяют внимание улучшению понимания анализируемых данных. Ключевой момент в таких моделях — легкость и прозрачность результатов для восприятия человеком. Возможно, обнаруженные закономерности будут специфической чертой именно конкретных

исследуемых данных и больше нигде не встретятся, но это все равно может быть полезно и потому должно быть известно. К такому виду задач относятся кластеризация и поиск ассоциативных правил.

Для описательных задач можно использовать методы анализа данных, такие как гистограммы, диаграммы рассеяния и корреляционный анализ. Для предсказательных задач можно использовать методы машинного обучения, такие как линейная регрессия и классификация.

**32.В каких случаях лучше использовать линейные, а когда нелинейные модели анализа данных?** Изображение выглядит как текст

Автоматически созданное описание

Линейные модели обычно используются в случаях, когда зависимость между переменными линейна. Они хорошо работают в случаях, когда данные имеют линейную структуру.

Нелинейные модели используются в случаях, когда зависимость между переменными нелинейна2. Они могут быть более точными, чем линейные модели, если данные имеют нелинейную структуру.

Однако выбор модели зависит от конкретной задачи и данных. Некоторые задачи могут быть решены только с помощью нелинейных моделей.

**33.Приведите математическое выражение для параметрической модели статистического обучения, для каких задач анализа данных их лучше использовать?**

Изображение выглядит как диаграмма

Автоматически созданное описание

**34. В чем состоят преимущества и недостатки непараметрических моделей анализа данных, как осуществить выбор между параметрической и непараметрической моделью?**

Параметрические модели анализа данных предполагают наличие заранее определенных параметров, которые могут быть оценены на основе данных. Непараметрические модели не предполагают наличие таких параметров и могут быть использованы в случаях, когда данные не имеют явной структуры или когда необходимо изучить свойства распределения данных.

Преимущества параметрических моделей заключаются в том, что они могут быть более точными и эффективными при наличии хорошо определенных параметров. Однако непараметрические модели могут быть более гибкими и могут использоваться в случаях, когда данные не имеют явной структуры или когда необходимо изучить свойства распределения данных.

Выбор между параметрической и непараметрической моделью зависит от конкретной задачи анализа данных и от того, какие свойства данных вы хотите изучить.

**35.Как выполняется измерение качества модели анализа данных?**

Изображение выглядит как текст

Автоматически созданное описание

**36. В чем состоит фундаментальное свойство статистического обучения?**

Изображение выглядит как текст

Автоматически созданное описание