

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»
(Университет ИТМО)

Факультет программной инженерии и компьютерной техники

Отчет по лабораторной работе №1

по дисциплине « База данных »

Тема: Data Mining на основе модели CRIPS-DM

Выполнила:

студентка гр. № Р33212 Ян Цзяфэн

Санкт-Петербург

2020

Data Mining

Интеллектуальный анализ данных (Data Mining), также известный как обнаружение знаний (Knowledge Discover in Database, KDD) в базах данных, в настоящее время является актуальной проблемой в области искусственного интеллекта и исследований баз данных. Интеллектуальный анализ данных относится к нетривиальному процессу выявления скрытой, ранее неизвестной и потенциально ценной информации из большого количества данных в базе данных.

Интеллектуальный анализ данных - это процесс поддержки принятия решений, который в основном основан на искусственном интеллекте, машинном обучении, распознавании образов, статистике, базе данных, технологиях визуализации и т. д. Путём высокоавтоматизированного анализа данных предприятия, создания индуктивных рассуждений и из них выявления потенциальных закономерностей, помогает лицам, принимающим решения, корректировать рыночные стратегии, снижать риски и принимать правильные решения.

В этой предметной области будет анализироваться интеллектуальный анализ данных на основе модели CRISP-DM.

CRISP-DM

В проекте CRISP-DM (CRoss Industry Standard Process for Data Mining) была предложена комплексная модель процесса для выполнения проектов интеллектуального анализа данных. Модель процесса не зависит ни от отрасли, ни от используемой технологии.

Модель процесса CRISP-DM направлена на то, чтобы сделать крупные проекты интеллектуального анализа данных менее затратными, более надежными, более воспроизводимыми, более управляемыми и более быстрыми.

Модель CRISP-DM предоставляет полное описание процесса для проекта KDD. Эта модель делит проект KDD на 6 различных этапов (как показано на рисунке 1), но последовательность не остается полностью неизменной:

- Понимание бизнес-целей (Business Understanding)
- Начальное изучение данных (Data Understanding)
- Подготовка данных (Data Preparation)
- Моделирование (Modeling)
- Оценка (Evaluation)
- Внедрение (Deployment)

Конечно, эти шесть этапов можно разделить на несколько подпроцессов.

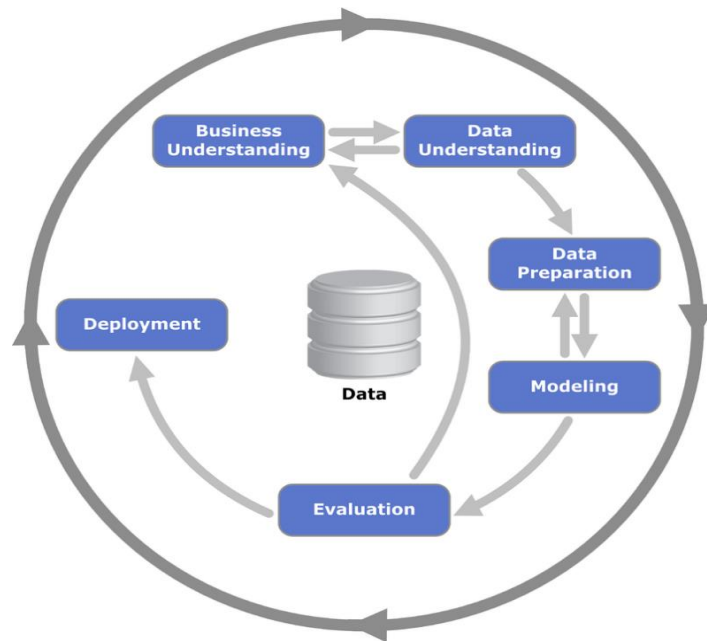


рис. 1 Жизненный цикл исследования данных в модели CRIPS-DM

Таблицы

Наверное, она включает несколько таблиц:

- 1) Таблица о описании процессов Data Mining на основании модели CRISP-DM.
- 2) Таблица о описании подпроцессов.
- 3) Таблица о описании алгоритмов Data Mining. (Каждый процесс использует соответствующие алгоритмы.)
- 4) Таблица о описании задач Data Mining. (Каждый процесс решает соответствующие задачи.)
- 5) Таблица о описании характеров данных. (Существуют ли Noisy value, missing value и т.д.).
- 6) Таблица о параметрах алгоритмов.
- 7) Таблица о компонентах алгоритмов. (Similarity measure, loss function, и т.д.)
- 8) Таблица о выводных моделях.