

УДК 004.42, 519.254

РАЗРАБОТКА ПРОГРАММНОГО ИНСТРУМЕНТАРИЯ ДЛЯ АНАЛИЗА ДАННЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

Вашлаев А.В., студент гр. ПИб-161, IV курс

Научный руководитель: Славолюбова Я. В., к.ф.-м.н., доцент

Кузбасский государственный технический университет
имени Т.Ф. Горбачёва,
г. Кемерово

Анализ данных представляет собой нахождение общих закономерностей и расчет возможных последствий путем различных методов, выбираемых в зависимости от источника данных и конечной цели работы с ними.

На данный момент можно выделить ряд программ, направленных на статистический анализ информации:

- SPSS Statistic. Продукт развивается более 50 лет. Обладает простым интерфейсом, разнообразной графической визуализацией и широким выбором методов статистики. Распространяется на платной основе;
- STATA. Отличительной чертой является поддержка проектов, разработанных на более старых версиях программы. Распространяется на платной основе;
- SAS. Набор статистических программ, разделенные на конкретные отрасли (например, банк, здравоохранение, транспорт, энергетика). Распространяется на платной основе;
- Jamovi. позиционирует себя как статистическая таблица третьего поколения и бесплатная альтернатива SPSS и SAS;
- RStudio. Представляет среду разработки, объединяющую все необходимые инструменты для комфорtnого написания на языке программирования R.

Общей чертой вышеперечисленных программ являются большое количество методов обработки информации. Многообразие методов обусловлено широким спектром источников данных и возможных вариантов прогнозируемого поведения в зависимости от выбранного метода. Именно поэтому необходим многосторонний подход к изучению данных. Ниже приведен ряд методов анализа:

- Кластерный анализ. Включает в себя набор различных алгоритмов классификации. Анализ производится по некоторым общим параметрам, количество которых может достигать нескольких десятков;
- Факторный анализ. Используется, когда количество параметров превышает сто единиц. Параметры группируются в факторы по каким-то общим признакам и после этого производится работа с ними;

- Регрессионный анализ. Основной целью является определение наличия и характера связи между переменными;
- Дискриминантный анализ. Используется для принятия решения о том, какие переменные разделяют возникающие наборы данных;
- Нейронные сети. Из-за универсальности в построении структуры сети, может быть применено к любой области данных. Главным минусом данного метода является то, что процесс обучения нейронной сети и процесс принятия решений абсолютно неконтролируемы;
- Корреляционный анализ. Позволяет судить о том, насколько похоже ведут себя различные переменные. При изучении корреляций стараются установить, существует ли какая-то связь между двумя показателями в одной выборке.

Разрабатываемый программный инструментарий должен отвечать следующим требованиям:

- Простой и интуитивно понятный интерфейс;
- Разнообразие методов анализа данных;
- Графическая, табличная визуализация результатов;
- Импорт данных из Excel-документов;
- Сохранение результатов обработки;
- Отличительная особенностью продукта будет визуализация результатов на карте, путем выделения используемых областей.

Продукт будет представлять из себя десктопное приложение. Разработка производится на платформе Microsoft Visual Studio с использованием языка C# и подключаемых сторонних модулях. Ниже представлены модель бизнес-процесса (рис. 1) и модель декомпозиции (рис. 2) обработки данных.

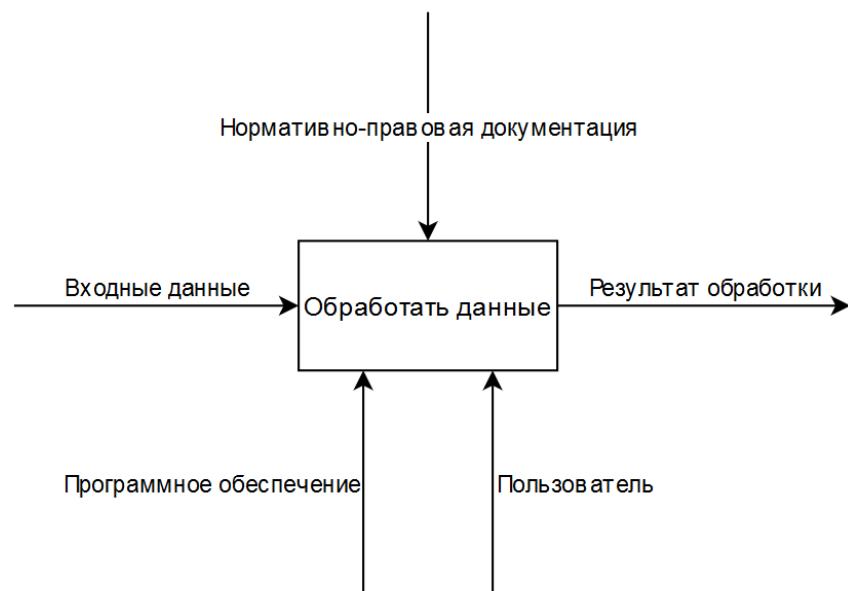


Рисунок 1. Модель бизнес-процесса «обработка данных»



Рисунок 2. Модель декомпозиции «обработка данных»

Список литературы:

1. RStudio Open source & professional software for data science teams – RStudio [Электронный ресурс]. – Режим доступа: <https://rstudio.com/> (дата обращения 13.03.20).
2. PSPP - GNU Project - Free Software Foundation [Электронный ресурс]. – Режим доступа: <https://www.gnu.org/software/pspp/> (дата обращения 13.03.20).
3. Jamovi - Stats. Open. Now [Электронный ресурс]. – Режим доступа: <https://www.jamovi.org/> (дата обращения 13.03.20).
4. Аналитика, бизнес-аналитика и управление данными [Электронный ресурс]. – Режим доступа: https://www.sas.com/ru_ru/home.html (дата обращения 13.03.20).
5. Stata: Software for Statistics and Data Science [Электронный ресурс]. – Режим доступа: <https://www.stata.com/> (дата обращения 13.03.20).
6. Современные методы анализа данных [Электронный ресурс]. – Режим доступа: <http://riep.ru/upload/iblock/031/031173bb40e099800b248497db44cb88.pdf> (дата обращения 13.03.20).
7. А.П. Кулаичев. Методы и средства комплексного статистического анализа данных: учеб. пособие / М.: ИНФРА-М, 2017г. – 484 с.
8. Козлов А.Д., Лекае В.А., Шаповалова М.С. Методы анализа предметных областей: учебное пособие / РГГУ, 2018г. – 201 с.
9. Нестеров С. Интеллектуальный анализ данных: базовые понятия [Электронный ресурс]. – Режим доступа: <https://www.intuit.ru/studies/courses/2312/612/lecture/13260> (дата обращения 13.03.20).