

Анализ поведения пользователей на базе платформы бизнес-аналитики

Д. П. Плахотников¹, Е. Е. Котова²

СПбГЭТУ «ЛЭТИ»

¹dimapl21@yandex.ru, ²apu_kotova@mail.ru

Аннотация. В процессе обучения традиционным способом представления результатов учащихся является оценка. По мере образовательной деятельности у преподавателей образуется огромный массив оценочной информации. Чтобы проанализировать и визуализировать этот массив данных существуют специальные системы обработки данных – системы класса Business Intelligence (BI) или интеллектуального анализа данных.

Ключевые слова: большие данные; наука о данных; обработка данных; интеллектуальный анализ данных; бизнес-аналитика

I. СИСТЕМЫ КЛАССА BI?

Чаще всего для хранения различных оценочных данных и их визуализации, полученных из автоматизированных систем обработки информации, используется Microsoft Excel. Но данный инструмент требует большой ручной обработки данных. К тому же, средства визуализации в нём не настолько информативны, как это требуется.

На сегодняшний день появились автоматизированные и более информативные средства обработки данных – системы класса Business Intelligence (BI) или интеллектуального анализа данных. В основе BI лежит организация доступа конечных пользователей и анализ структурированных количественных по своей природе данных.

A. История термина Business Intelligence

Впервые термин «Business Intelligence» был употреблён в 1958 году в журнале IBM System Journal. В нём американский ученый Ханс Петер Лун опубликовал статью «A Business Intelligence System» [1]. И только в конце 1980-х годов аналитики компании Gartner дали трактовку термину Business Intelligence – «ориентированный на пользователя процесс, который включает доступ и исследование информации, ее анализ, выработку интуиции и понимания, которые ведут к улучшенному и неформальному принятию решений» [2]. Позднее, в 1996 году, появилось уточнение определению: «Business Intelligence – это инструменты для анализа данных, построения отчетов и запросов, которые могут помочь бизнес-пользователям преодолеть море данных для того, чтобы синтезировать из них значимую информацию» [3].

B. Основные возможности BI-систем

Основные возможности BI-систем развиваются по следующим направлениям: хранение данных, интеграция данных, анализ данных и представление данных. Хранение данных происходит с помощью специальных хранилищ – «data warehouse» [4]. Информация в хранилище данных, включая исторические данные, собирается из различных операционных (транзакционных) систем и структурируется специальным образом для более эффективного анализа и обработки запросов (в отличие от обычных баз данных, где информация организована таким образом, чтобы оптимизировать время обработки текущих транзакций). Для решения более узких, конкретных задач из общего хранилища могут вычленяться подмножества данных – так называемые витрины данных («data marts»).

Структура BI-системы изображена на рис. 1.



Рис. 1. Структура BI-систем

C. Основные преимущества BI-систем

По мнению аналитиков компании MiPro Consulting, внедрение в организации самостоятельной BI-системы обеспечивает целый ряд преимуществ перед использованием аналитических инструментов, встроенных в другие корпоративные информационные системы [5].

Среди таких преимуществ BI-системы:

- наглядность с помощью инструментов визуализации;
- масштабируемость и низкие трудозатраты;

- извлечение, анализ и консолидация любых данные практически из различных источников;
- наличие встроенных инструментов для решения любых аналитических задач;
- легкий централизованный доступ к данным и аналитическим инструментам.

II. ВЫБОР ПЛАТФОРМЫ BI

Компания Gartner ежегодно определяет требования, которым должны удовлетворять современные Business Intelligence и иные аналитические платформы, оценивает сильные и слабые стороны продуктов и предлагает свой рейтинг издателей рынке BI-платформ [6].

A. Типы платформ

Все BI-платформы делятся на два типа:

1. ИТ-ориентированные платформы (IT-centric platforms) – например, Microsoft, IBM, SAS, SAP, Oracle.
2. Платформы, ориентированные на бизнес пользователей (business-user-driven analytics platforms). Например – Qlik, Tableau, TIBCO.

Различия типов платформ приведены в табл. 1.

ТАБЛИЦА 1 Различия типов платформ

ИТ-ориентированные платформы	Бизнес-ориентированные платформы
Для создания и эксплуатации аналитических решений требуют высокого уровня компетенции в IT сфере и специалиста в области	Пользователь напрямую «общается» с данными, не обращаясь к специалисту в области. Встроенные инструменты интуитивно понятны даже новичку
Сложность создания аналитических приложений. Строятся на базе On-Line Analytical Processing (OLAP) кубов и централизованного хранилища данных, в которое грузятся предварительно очищенные данные из систем обработки данных оперативных транзакций (Online Transaction Processing)	Высокие скорость и простота создания аналитических приложений. Не требуют традиционных реляционных хранилищ данных и специально созданных OLAP-кубов
В обработке данных доминирует технология регламентированной отчетности в табличной форме (reporting) и набора ключевых показателей (scoreboards).	В обработке данных доминирует идеология обнаружения/открытия информации (data discovery, data mining) – произвольный интерактивный анализ (поиск, фильтрация, закладки, графическая визуализация данных). Интерактивность обеспечивается вычислениями в оперативной памяти (in memory)

Исходя из потребностей сбора, анализа и визуализации данных, получаемых в учебном процессе, была выбрана бизнес-ориентированная платформа Qlik в качестве платформы для разработки аналитических приложений.

B. Преимущества платформы

Преимущества обусловлены следующими факторами:

- Ориентированность платформы на широкий рынок – требуемая масштабируемость и производительность для любых объемов разнородных данных позволяет пользователям относительно просто и легко выполнять сложные виды анализа.
- Скорость обработки массивов данных и простота в работе. Скорость и простота в работе обеспечивается ассоциативной моделью в оперативной памяти с визуализацией данных «Green-White-Gray» (зеленый-белый-серый). Это технология быстрой загрузки необходимых данных.
- Низкая стоимость владения аналитической платформой. Согласно анализу Gartner на основе оценок, полученных от 1700 компаний пользователей BI платформ, Qlik имеет самую низкую общую стоимость владения в течение трех лет. К тому же в образовательных целях платформа предоставляется бесплатно.

C. Возможности платформы

Платформа содержит множество различных видов диаграмм и таблиц, комбинированных и специализированных элементов для расширенного анализа. Элементы типа «список» позволяют быстро перемещаться по всем значениям измерений. На каждый графический элемент можно кликнуть мышью, чтобы сделать запрос на исследование данных. Существуют следующие виды аналитики: информационные панели («dashboards»), многомерный интерактивный анализ в срезах, кросс-таблицы («pivoting»), анализ «что, если» («what-if»), анализ групп данных, статистический анализ: отслеживание трендов, выявление корреляций, проверка гипотез и др., моделирование и прогнозирование [3].

D. Безопасность платформы

В платформе был реализован механизм защиты информации, который может быть задан двумя различными способами: встроен в скрипт документа, а также задан посредством специального средства Qlik Sense Publisher. Связь между сервером и конечным клиентом шифруется, что делает информацию недоступной для средств просмотра, отладчиков и иных средств перехвата информации [7].

III. ВНЕДРЕНИЕ В ОБРАЗОВАТЕЛЬНЫЙ ПРОЦЕСС

При внедрении BI-технологии в образовательный процесс в первую очередь встает задача организации единого информационного пространства, единой базы

данных. В хранилище Business Intelligence может поступать информация из любых доступных источников, в том числе, обычная электронная таблица Microsoft Excel, базы данных (например, Microsoft SQL или Oracle SQL), ERP, CRM или другие корпоративные системы, а также многомерные базы данных.

Для хранения данных в системе используется ассоциативная модель данных. Она представляет собой таблицы, связанные по ключевым полям. Данная модель создается динамически в момент загрузки данных. Поля с полностью идентичными названиями ассоциируются друг с другом. При выборе какого-либо значения в одной таблице автоматически выбираются все ассоциированные значения в других таблицах, что позволяет упростить процесс обработки информации.

IV. ПРИМЕР ВНЕДРЕНИЯ BI-ПЛАТФОРМЫ

Для внедрения BI-платформы в образовательный процесс было разработано приложение для анализа данных результатов обучения студентов (оценок, баллов) [8].

Исходные данные были получены в ходе учебного процесса. Данные (фрагменты) представлены в табл. 2. Исходные данные для приложения.

ТАБЛИЦА 2

№	Код студента	Итог (балл)
1	ВаИ	4,00
2	ГаА	3,00
3	ДаД	5,00
4	ДмА	5,00
5	ДоА	4,00
...

Данные хранятся в обычном файле *.xlsx в виде информация по группам обучения (в данном примере «Код студента» и «Итог (балл)» по двум группам).

Для загрузки исходных данных в приложение был написан скрипт загрузки, представленный на рис. 2. Функция Num при загрузке была использована, чтобы преобразовать текстовые данные в числовые, а переменная group используется для указания номера группы.

```
[Группа 1]:
LOAD
1 as group,
  [№],
  [Код студента],
  Num([Итог (балл)]) as [Итог]
FROM [lib://AF/Сравнение.xlsx]
(ooxml, embedded labels, table is [Группа 1]);

[Группа 2]:
LOAD
2 as group,
  [№],
  [Код студента],
  Num([Итог (балл)]) as [Итог]
FROM [lib://AF/Сравнение.xlsx]
(ooxml, embedded labels, table is [Группа 2]);
```

Рис. 2. Текст скрипта загрузки данных

После загрузки информации в приложение был разработан интерфейс приложения. Он представлен на рис. 3. В нём отображены показатели среднего балла по группам (в исходных данных имеются 2 группы), график среднего балла по группам, показатели количества отлично успевающих и неуспевающих, круговая диаграмма успеваемости и таблицы с оценками групп.

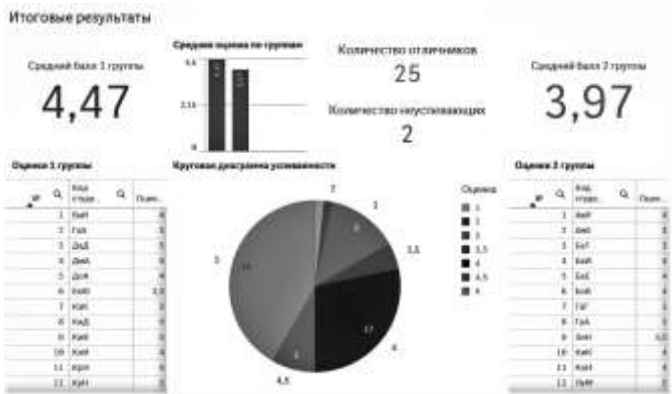


Рис. 3. Интерфейс разработанного приложения

При изменении исходных данных, приложение не нужно пересоздавать, достаточно лишь заново загрузить данные, что позволяет облегчить обработку данных.

Платформа позволяет использовать различные выборки, то есть добавлять или исключать какие-либо данные (рис. 4).

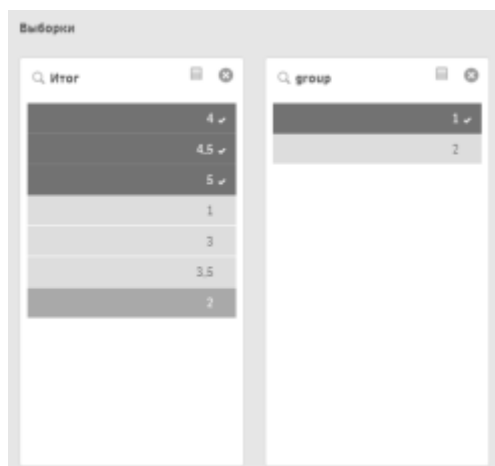


Рис. 4. Инструмент выборки

В результате исходные данные изменяются в соответствии с применёнными выборками (рис. 5).

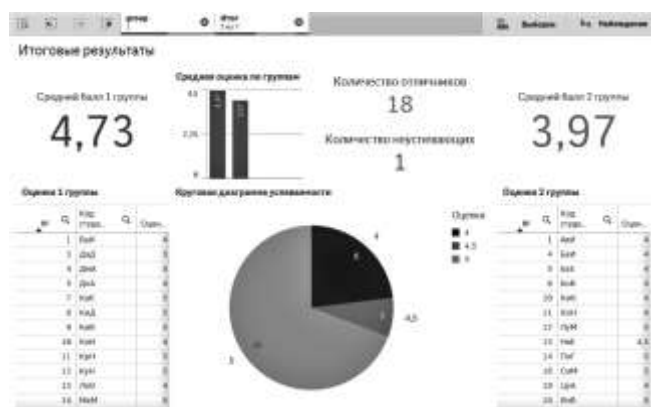


Рис. 5. Применения выборки

V. ЗАКЛЮЧЕНИЕ

Системы BI являются эффективной поддержкой операционных и управленческих решений.

В результате внедрения платформы удалось сократить трудозатраты преподавателя на анализ эффективности процесса обучения, а также улучшить восприятие оценочной информации за счёт её визуализации.

Изучение моделей использования BI в образовательных организациях представляет интерес для дальнейших исследований. На практике новая структура BI может использоваться образовательными организациями для планирования бизнес-среды обучения.

СПИСОК ЛИТЕРАТУРЫ

- [1] Н. Р. Luhn. A Business Intelligence System. IBM Journal (Октябрь 1958).
- [2] Что такое Business Intelligence? [Электронный ресурс]. – Интернет-сайт. – URL: <http://citforum.ru/consulting/BI/whatis/> (дата обращения: 25.09.2019).
- [3] Business Intelligence [Электронный ресурс]. – Интернет-сайт. – URL: <https://www.forrester.com/Business-Intelligence> (дата обращения: 25.09.2019).
- [4] Ralph Kimball, Margy Ross. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling - 3rd edition // Wiley. 2013. 601 p.
- [5] Преимущества использования BI-системы [Электронный ресурс]. – Интернет-сайт. – URL: http://www.tadviser.ru/index.php/Статья:Преимущества_использования_BI-системы (дата обращения: 25.09.2019).
- [6] Magic Quadrant for Analytics and Business Intelligence Platforms [Электронный ресурс]. – Интернет-сайт. – URL: <https://www.gartner.com/en/documents/3900992-magic-quadrant-for-analytics-and-business-intelligence-p> (дата обращения: 25.09.2019).
- [7] Безопасность Qlik // FBconsult. URL: <https://fbconsult.ru/bi-tekhologii/security-qlik> (дата обращения: 25.09.2019)
- [8] Котова Е.Е., Плахотников Д.П. Анализ процесса обучения на базе платформы QLIK SENSE. Современное образование: содержание, технологии, качество. Материалы XXV международной научно-методической конференции. СПб.: Изд-во СПбГЭТУ «ЛЭТИ». 2019. 666 с.