

30-31 января 2018 г.

XVIII международная научно-практическая конференция

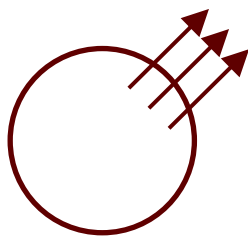
НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В
ОБРАЗОВАНИИ

Интеллектуальный анализ данных Финансово- аналитической системы университета. Задача кластеризации

Чуков Антон Максимович,
студент НИТУ МИСиС
Фомичева Ольга Евгеньевна,
доцент кафедры АСУ НИТУ МИСиС

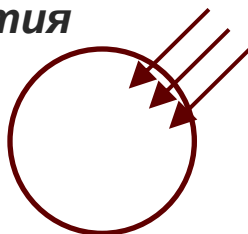


Пути развития



Образование – конкурентный рынок

Внешние: совершенствование образовательной, научной деятельности



Внутренние: оптимизация ресурсов

Подразделения

Отделы ПФУ

Руководство и внешние потребители

*Выгрузка через файл
Полуавтоматическая обработка*

Системы управления
контингентом

Системы бухгалтерского
учета

Другие учетные системы

Файлы отделов
ПФУ (Excel)

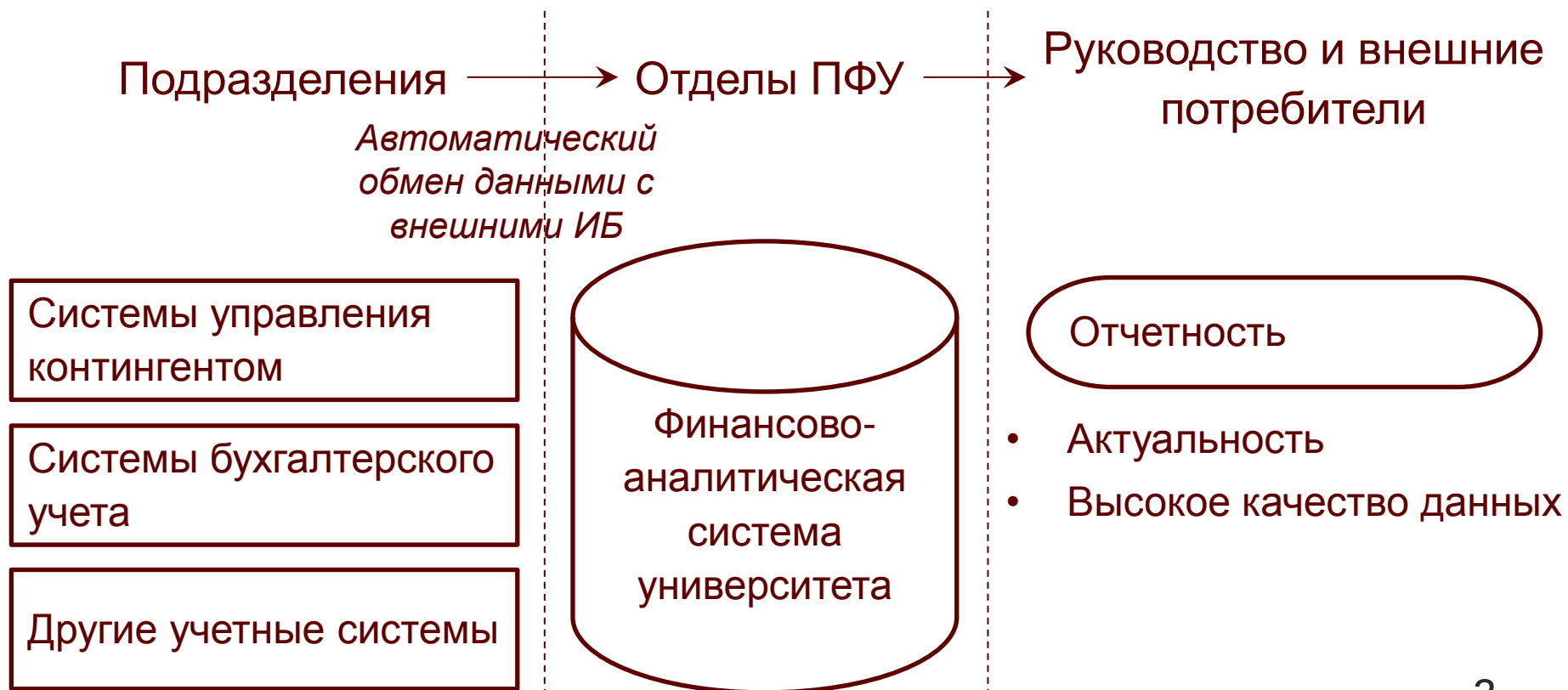
Отчетность

- Высокая трудоемкость
- Неактуальные данные
- Намеренное “загрубление” аналитики



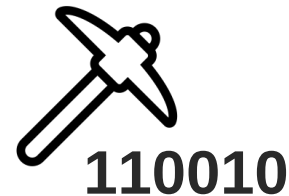
Финансово-аналитическая система университета

- Интеграция с внешними ИБ
- Обеспечение инструментария планирования, контроля исполнения и управления бюджетами





Интеллектуальный анализ данных (Data Mining)



Ряд методов, задачами которых является построение моделей, позволяющих выявлять **новую, нетривиальную, потенциально полезную для практики** информацию из уже имеющегося большого числа данных.

Для чего ИАД в ФАСУ?

Улучшить понимание имеющихся данных и происходящих процессов

Упростить анализ большого числа данных с помощью методов визуального представления

Перейти к планированию, построенному на прогнозных данных высокого качества

Повысить точность уже используемых в ПФУ прогнозных моделей

Повысить качество хранимой информации в целом

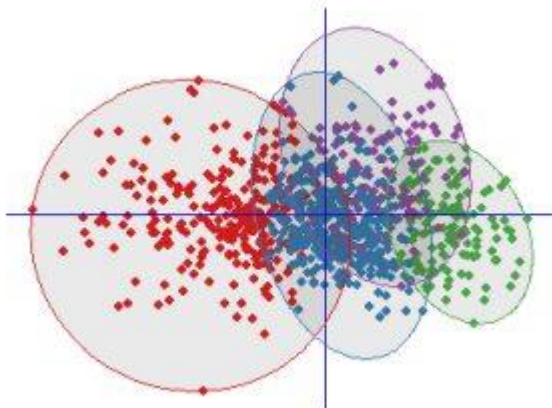


Кластеризация -

разбиение объектов (подразделений) на непересекающиеся и однородные группы (кластеры) объектов так, чтобы объекты были более похожи на другие объекты внутри кластера, чем на объекты из других кластеров.

Решение задачи кластеризации может использоваться для:

- понимание данных, выявление кластерной структуры. Разбиение подразделений на кластеры позволяет упростить дальнейшую обработку данных и принятие решений. Например, выработку стратегий;
- выявление новых знаний, например, выявление нетипичных объектов.





Процесс кластеризации

1. Отбор выборки объектов

Подразделения ВУЗа, являющиеся ЦФО в процессе бюджетирования (около 200 объектов).

2. Выбор меры близости

Квадрат Евклидова расстояния

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2$$

3. Выбор алгоритма кластеризации

Метод k-means (метод k-средних). Основан на стремлении минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров.

4. Представление результатов анализа

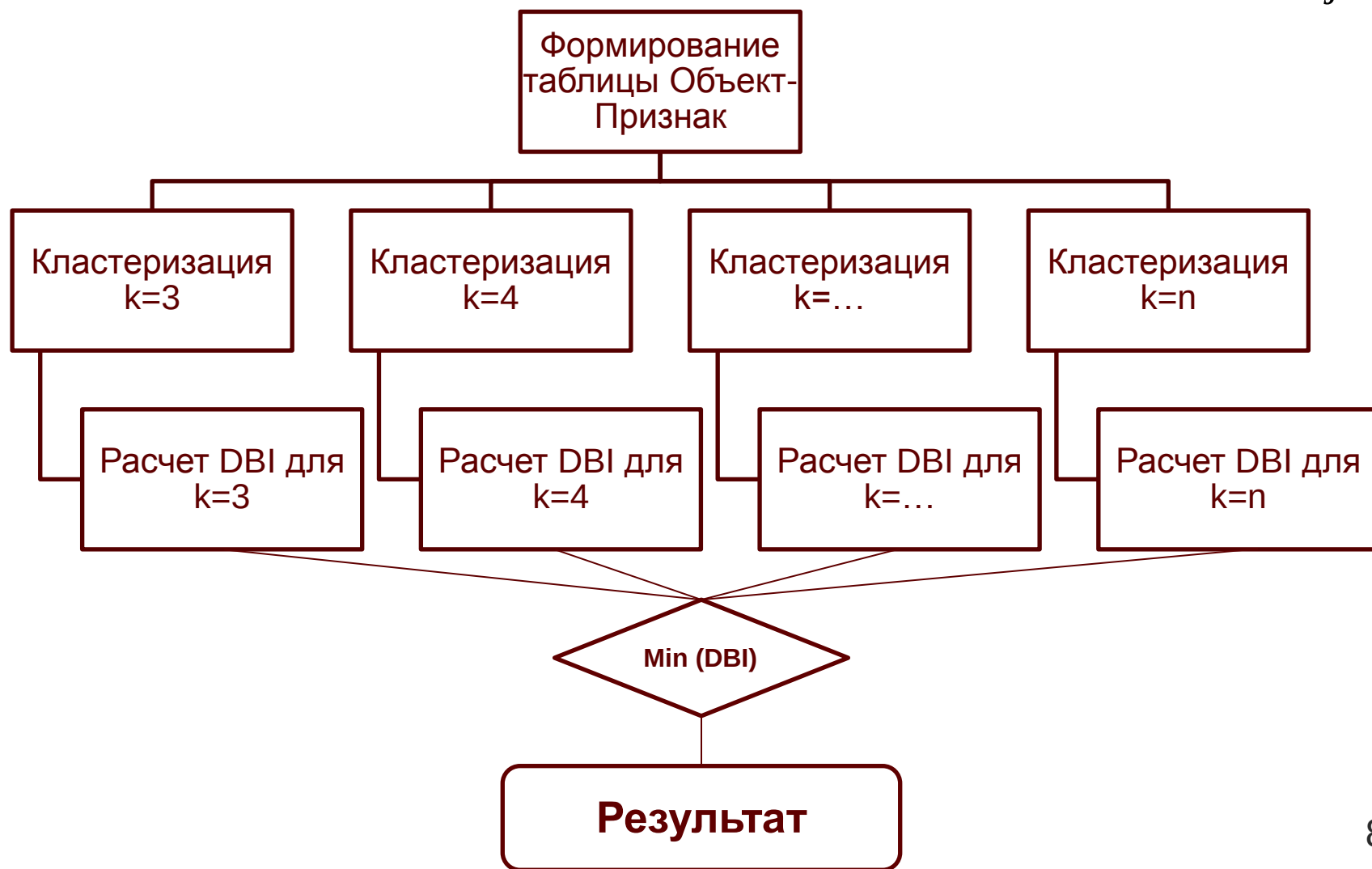
Признаки объектов

Признак	Ед. изм.
Образовательная деятельность	
Численность студентов (бакалавриат, магистратура, специалитет, аспирантура)	человек
Средний бал ЕГЭ принятых студентов	баллы
Численность «олимпиадников», принятых без вступительных испытаний	человек
Научно-исследовательская деятельность	
Количество статей в научной периодике, индексируемой в системах Web of Science, Scopus, РИНЦ	единиц
Общий объем НИОКР	тыс. руб.
Доходы от НИОКР в расчете на одного НПР	тыс. руб.
Удельный вес численности НПР, имеющих ученую степень	%
Количество научных журналов, издаваемых образовательной организацией	единиц
Количество грантов за отчетный период в расчете на 100 НПР	единиц
Международная деятельность	
Численность иностранных студентов (кроме СНГ)	человек
Численность иностранных граждан из числа НПР	человек
Объем средств, полученных на выполнение НИОКР от иностранных граждан или юридических лиц	тыс. руб.
Финансово-экономическая деятельность	
Доходы	тыс. руб.
Расходы	тыс. руб.



Индекс Дэвиса-Болдина

$$DBI = \frac{1}{k} \sum_{i=1, j=1, i \neq j}^k \max\left(\frac{S_i + S_j}{M_{ij}}\right),$$





Реализация

Финансово-аналитическая система университета реализована на конфигурации **1С Управление холдингом**.

Подсистема анализа данных (ПАД) – это прикладное решение, встраиваемое в любую конфигурацию для реализации методов интеллектуального анализа данных.



Параметры анализа

Количество искоемых кластеров:
Стандартизация:
Мера расстояния:
Метод кластеризации:

Колонки источника данных

Входные колонки

Имя колонки	Тип данных	Вес
КоличествоРозничныхТочек	Непрерывный	1
КоличествоАвтомобилей	Непрерывный	1
ВремяРаботыОрганизации	Дискретный	1
ВремяЗаключенияДоговора	Дискретный	1
ПрекращениеОтношений	Дискретный	1

Информация о данных

Количество объектов: 9

Результат анализа

Найдено кластеров: 3

Кластеры

№ кластера	Количество объектов	Процент
1	6	66,67
2	2	22,22
3	1	11,11

Центры кластеров

Поле \ Кластер	1	2	3
КоличествоРозничныхТочек	1,3333	11,0000	1,0000
КоличествоАвтомобилей	1,1667	3,5000	10,0000
ВремяРаботыОрганизации = Свыше десяти лет	0,3333		
ВремяРаботыОрганизации = Менше года	0,5000		
ВремяРаботыОрганизации = От года до трех	0,1667		
ВремяРаботыОрганизации = От трех до десяти лет		1,0000	1,0000
ВремяЗаключенияДоговора = От трех до десяти лет	0,3333		
ВремяЗаключенияДоговора = Менше года	0,6667	0,5000	
ВремяЗаключенияДоговора = От года до трех		0,5000	1,0000
ПрекращениеОтношений = Не прекращены	0,6667		
ПрекращениеОтношений = Несоблюдение договора	0,1667		
ПрекращениеОтношений = Прекращение контрагентом	0,1667	1,0000	1,0000

Расстояния между кластерами

Кластеры	1	2	3
1		7,6073	12,7857
2	7,6073		9,5370
3	12,7857	9,5370	

30-31 января 2018 г.

XVIII международная научно-практическая конференция

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В
ОБРАЗОВАНИИ

Спасибо за внимание

Чуков Антон Максимович,
студент НИТУ МИСиС
amchukov@ya.ru

