

АНО ВО "РОССИЙСКИЙ НОВЫЙ  
УНИВЕРСИТЕТ"

## **Бакалаврская работа**

Разработка хранилища данных аналитической  
информационной системы факультета  
информационных систем и компьютерных  
технологий АНО ВО "РосНОУ"

Автор:

Дергачева Дарья Алексеевна

Научный руководитель:

ст. преподаватель Собетов Кирилл Олегович

Москва  
2016 г.

# Цель бакалаврской работы

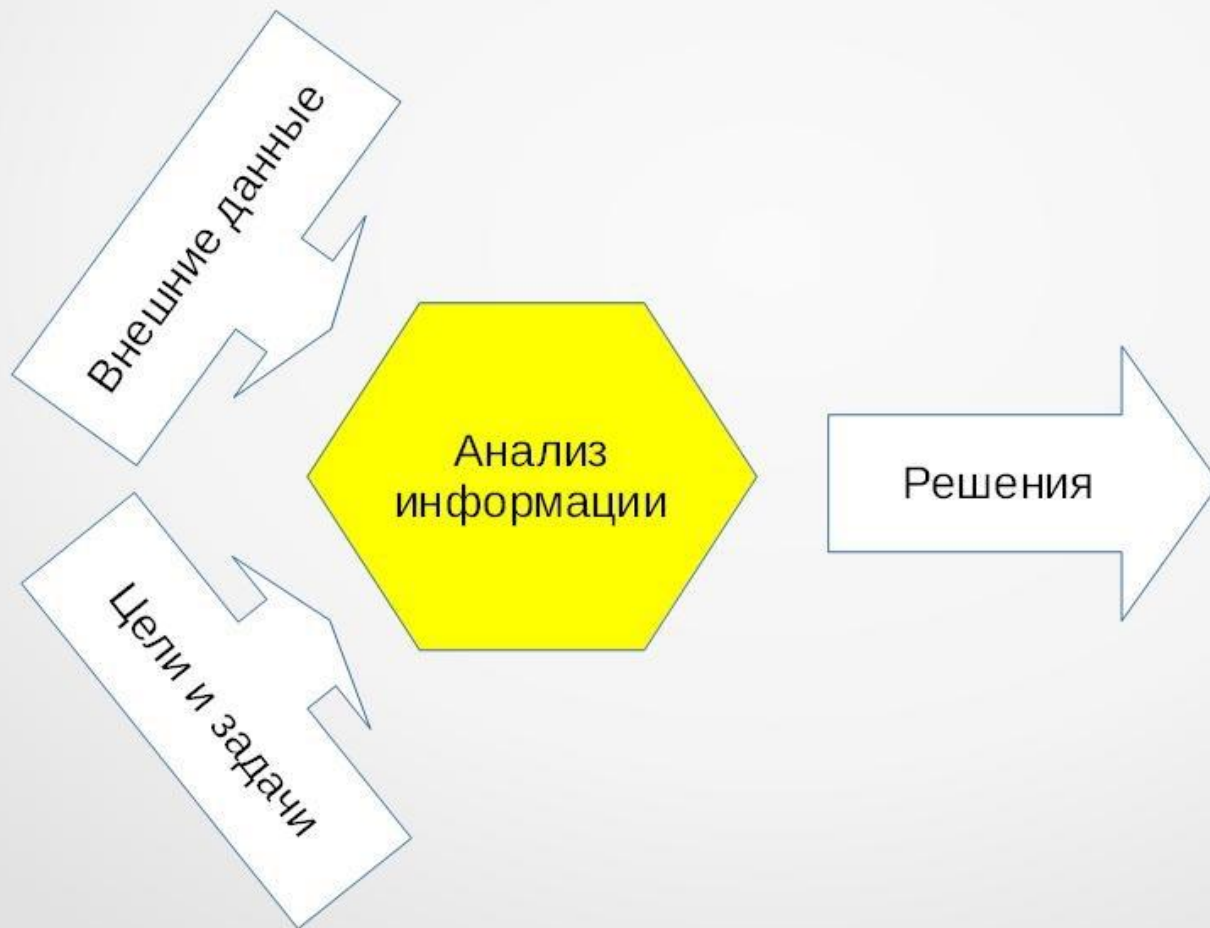
- Цель бакалаврской работы – проектирование и реализация хранилища данных аналитической информационной системы факультета ИСиКТ.

# Задачи бакалаврской работы

- Задачи:
  - Изучить и проанализировать проблемы, связанные с хранением и обработкой большого объема данных;
  - Изучить теорию хранилища данных;
  - Реализовать хранилище данных на основе существующей базы данных кафедры;
  - Реализовать методы Интеллектуального Анализа Данных.

# Хранилище данных

Хранилище данных (ХД — data warehouse) является местом складирования собираемых в системе данных и информационным источником для решения задач анализа данных и принятия решений.



# OLAP и OLTP

- OLTP (Online Transaction Processing) — обработка транзакций в реальном времени. Способ организации базы данных, при котором система работает с небольшими по размеру транзакциями, но идущими большим потоком.
- OLAP (Online analytical processing, аналитическая обработка в реальном времени) — технология аналитической обработки информации, которая основана на многомерном представлении данных и которая включает в себя составление и издание отчетов и документов.

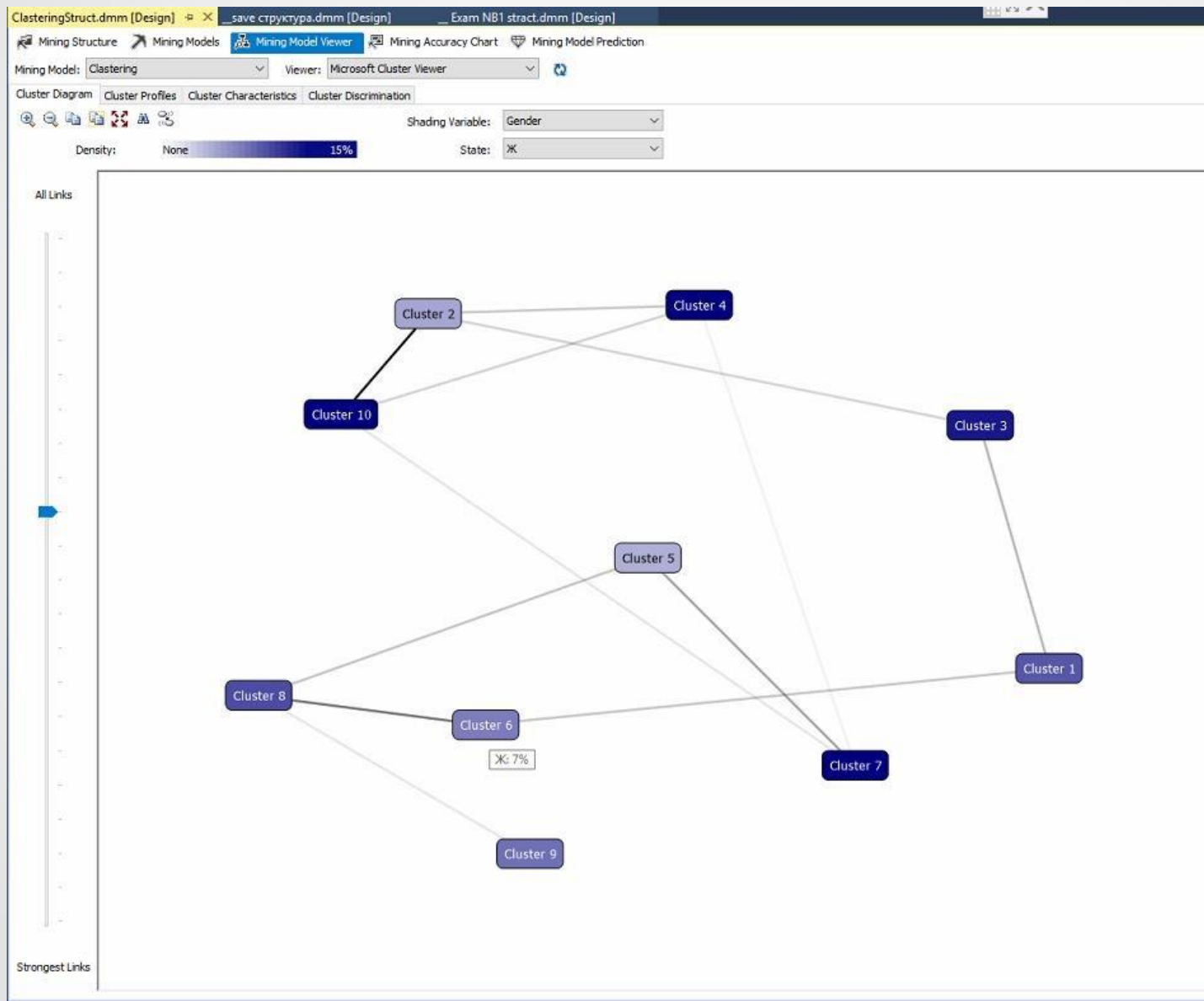
# OLAP и OLTP

| OLAP   | OLTP  |
|--|---|
| Включает в себя внутренние и внешние источники данных  | Требуются внешние источники данных  |
| Объем аналитических данных больше объема оперативных. Для анализа требуются данные за несколько лет.   | Требуются данные за несколько последних месяцев.  |
| Хранилище данных должно содержать единообразно представленную и согласованную информацию, максимально соответствующую содержанию оперативных БД.   | Оперативные БД могут содержать семантически эквивалентную информацию, представленную в разных форматах, с разным указанием времени ее поступления, иногда даже противоречивую |
| ХД создаются для нерегламентированных запросов аналитиков. Информация выбирается не часто, но затрагивает большие объемы данных.   | Системы обработки данных создаются для решения конкретных задач. Информация из БД выбирается часто и небольшими порциями.   |
| Малая изменчивость аналитических БД (только при загрузке данных). Упорядоченность массивов, более быстрые методы индексации при массовой выборке, хранение заранее агрегированных данных | Системы обработки данных являются сильно изменчивыми, что учитывается в используемых СУБД (нормализованная структура БД, строки хранятся неупорядоченно)                      |

# Интеллектуальный Анализ Данных

- Технология Интеллектуального Анализа Данных (Data mining) это процесс, реализованный для поиска в больших объемах данных закономерностей, которые объективны, полезны на практике, но не очевидны на первый взгляд.

# Кластерный анализ (1/2)





# Кластерный анализ(2/2)

Drill Through

Cases Classified to:

Population

| Mark Value Full ... | __ Exam   |
|---------------------|-----------|
| зачтено             | + __ Exam |
| незачтено           | + __ Exam |
| неудовлетвор...     | - __ Exam |

| Family      | Mark Value Full... | Structure.First ... | Structure.Sec ... | Structu... | Structure.Subject Name                    | Structure.Testing Date |
|-------------|--------------------|---------------------|-------------------|------------|---|------------------------|
| Давыдкин    | неудовлетвор...    | Вячеслав            | Владимирович      | 9          | Вычислительные системы, сети и телеко...  | 6/13/2006 12:00:00 AM  |
| Кобзев      | неудовлетвор...    | Дмитрий             | Игоревич          | 9          | Бухгалтерский учет                        | 6/30/2007 12:00:00 AM  |
| Зоидов      | неудовлетвор...    | Зафар               | Кабилджонов...    | 9          | Математическая логика и дискретная мат... | 1/5/2005 12:00:00 AM   |
| Ципенюк     | неудовлетвор...    | Михаил              | Александрович     | 7          | Бухгалтерский учет                        | 6/24/2008 12:00:00 AM  |
| Тихановский | неудовлетвор...    | Александр           | Владимирович      | 9          | Предметно-ориентированные экономиче...    | 5/26/2008 12:00:00 AM  |
| Пирогов     | неудовлетвор...    | Владислав           | Алексеевич        | 9          | Теория вероятностей и математическая с... | 1/24/2007 12:00:00 AM  |

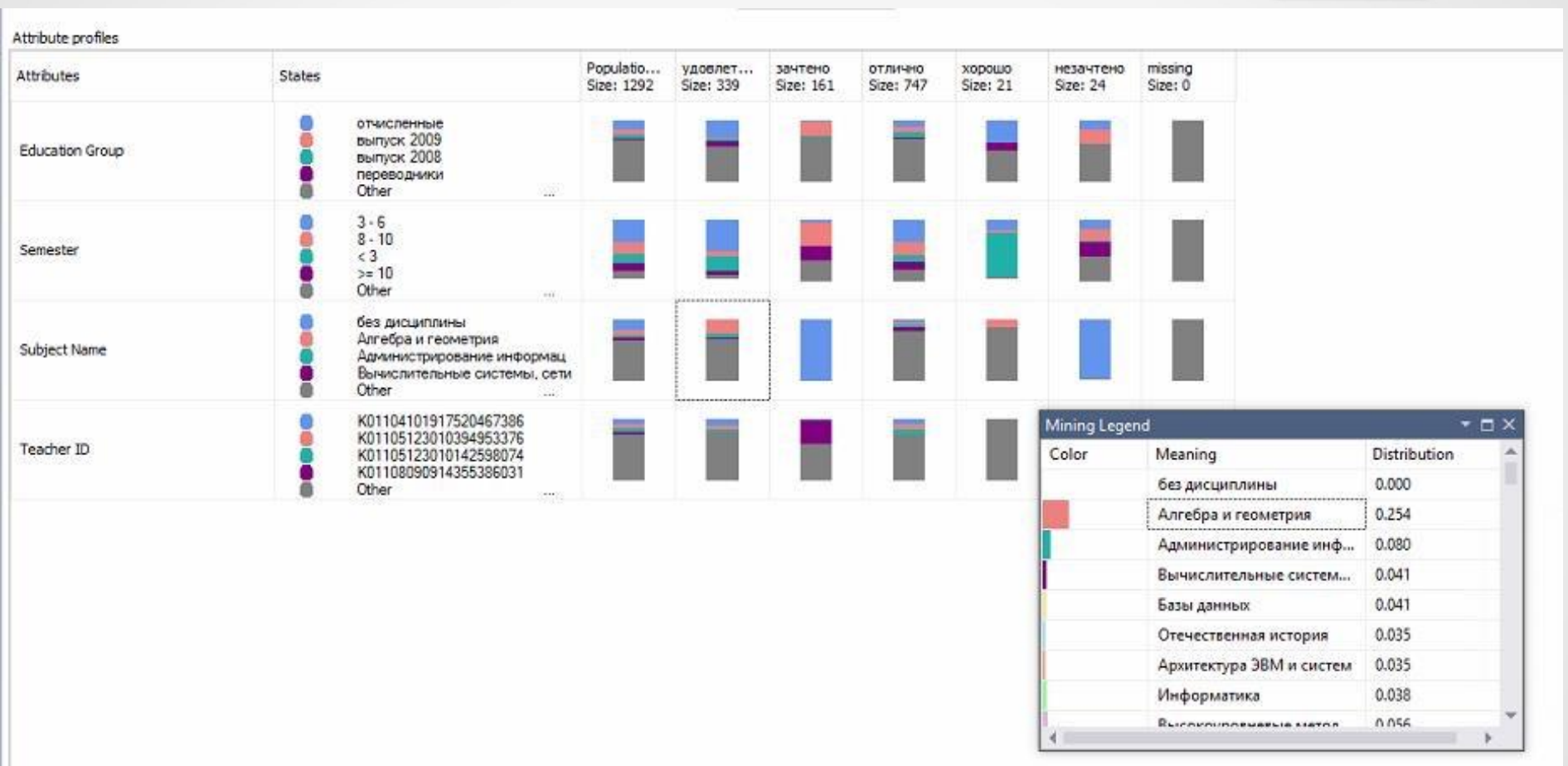
| отлично          | + __ Exam |
|------------------|-----------|
| удовлетворите... | + __ Exam |
| хорошо           | - __ Exam |

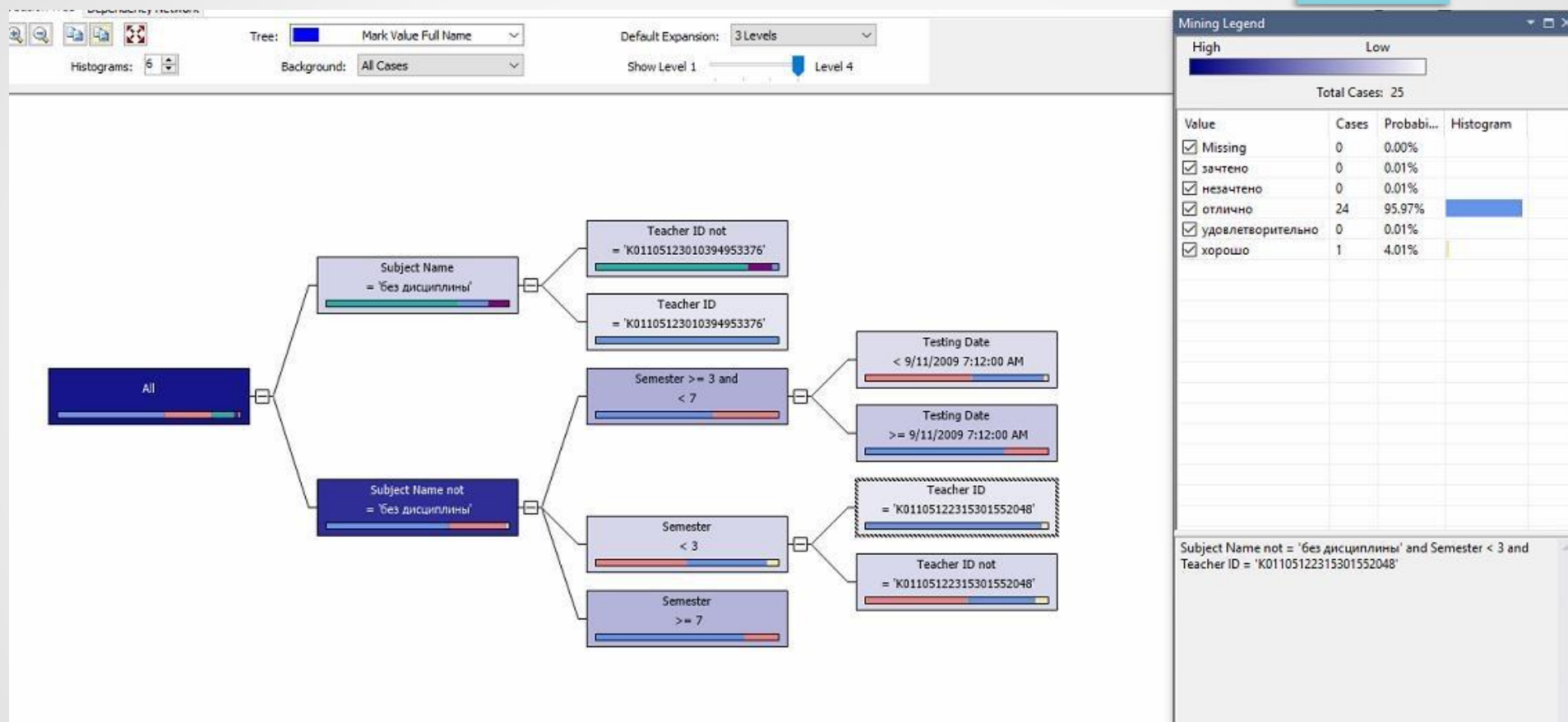
| Family    | Mark Value Full... | Structure.First ... | Structure.Sec ... | Structur... | Structure.Subject Name    | Structure.Testing Date |
|-----------|--------------------|---------------------|-------------------|-------------|---------------------------|------------------------|
| Водолазов | хорошо             | Виталий             | Сергеевич         | 9           | Базы данных               | 3/26/2010 12:00:00 AM  |
| Исаев     | хорошо             | Александр           | Александрович     | 3           | Математический анализ     | 1/21/2010 12:00:00 AM  |
| Ходяков   | хорошо             | Андрей              | Владимирович      | 3           | Архитектура ЭВМ и систем  | 1/22/2010 12:00:00 AM  |
| Евсеев    | хорошо             | Антон               | Евгеньевич        | 6           | Алгебра и геометрия       | 3/26/2010 12:00:00 AM  |
| Вдовченко | хорошо             | Владимир            | Александрович     | 3           | Информационные технологии | 1/26/2010 12:00:00 AM  |
| Никифоров | хорошо             | Алексей             | Михайлович        | 6           | Отечественная история     | 1/18/2008 12:00:00 AM  |

Query execution completed with 6 rows fetched

# Реализация Байесовской классификации



# Реализация деревьев в решениях

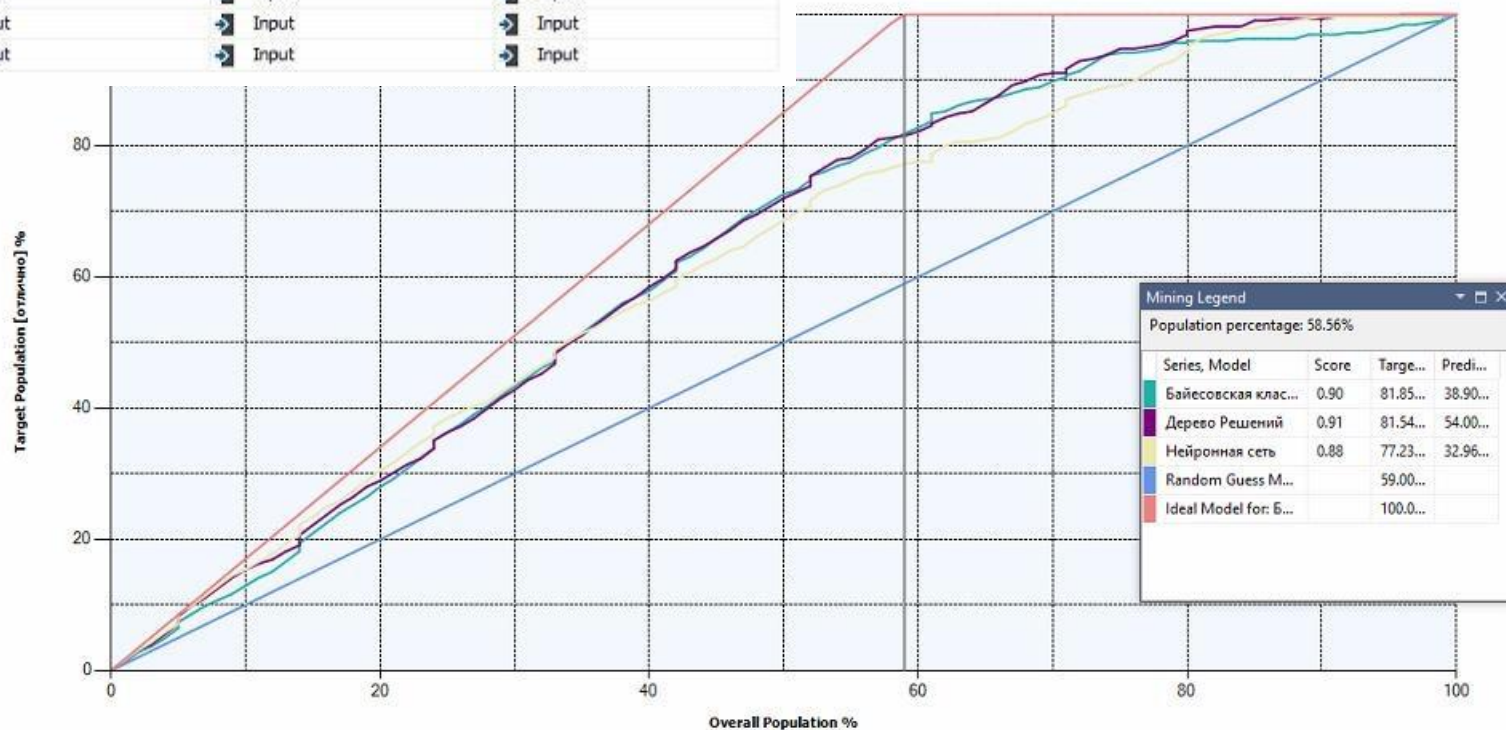


Воробьева Светлана Евгеньевна( TeacherID K011051223153011552048)

# Сравнение моделей

| Structure            | Байесовская классификация | Дерево Решений           | Нейронная сеть           |
|----------------------|---------------------------|--------------------------|--------------------------|
|                      | Microsoft_Naive_Bayes     | Microsoft_Decision_Trees | Microsoft_Neural_Network |
| Testing Date         | Input                     | Input                    | Input                    |
| Teacher ID           | Input                     | Input                    | Input                    |
| Subject Name         | Input                     | Input                    | Input                    |
| Semester             | Input                     | Input                    | Input                    |
| Person ID            | Key                       | Key                      | Key                      |
| Mark Value Full Name | PredictOnly               | PredictOnly              | PredictOnly              |
| Gender               | Input                     | Input                    | Input                    |
| Family               | Input                     | Input                    | Input                    |
| Education Group      | Input                     | Input                    | Input                    |
| Birth Place          | Input                     | Input                    | Input                    |

Байесовская классификация



# Выводы

- Цель бакалаврской работы достигнута
- Все задачи выполнены
- Описаны функциональные возможности систем;
- Реализовано хранилище данных;
- Обнаружены закономерности в ходе

## проведения анализа данных

Данная работа может быть успешно использована для встраивания в информационную систему факультета, что позволит значительно снизить время, затрачиваемое сотрудниками факультета, на анализ данных.